Particle-in-Cell Simulations and their Applications to Magnetospheres of Neutron Stars

Yuran (Alexander) Chen

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017
Abstract

Particle-in-Cell Simulations and their Applications to Magnetospheres of Neutron Stars

Yuran (Alexander) Chen

Neutron stars are surrounded by dense magnetospheres with nontrivial magnetic field structure. They are sources of multi-band emission from radio waves to very high energy gamma-rays. Pulsar wind nebulae observations also show that a large number of $e^\pm$ pairs flow from the neutron star, which are produced in the magnetosphere. The structure of the magnetosphere, the mechanism of pair production and particle acceleration in the magnetosphere, and how magnetic energy is converted to kinetic energy is a complex problem that only recently has started to be addressed fully from first principles. In this dissertation I describe how I developed a numerical code tailored to study this problem. A detailed description of the code and method is given, then it is used to study the pair discharge mechanism in the magnetosphere of rotating neutron stars whose rotating axis is aligned with the magnetic axis. It was found that to form an active magnetosphere it is necessary to have pair creation all the way towards the light cylinder. In the dissertation I classify the pulsars into two classes, and describe their differences.

The magnetospheres of magnetars are believed to be different from ordinary pulsars, in that they are sustained not by the rotation of the star, but by a twist launched from the stellar surface due to some sudden breakdown of the crust. I apply the same numerical tool to study the particle acceleration and pair creation mechanism in the twisted magnetosphere of the magnetar, showing where the gap is, and how the magnetosphere evolves over time. The magnetic twist was found to live much longer than the Alfvén time of the system, and slowly dissipates through developing a cavity in the inner magnetosphere. This not only explains the long term evolution of the magnetar lightcurve after an outburst, but also explains the observed evolution hotspots on the stellar surface.
Contents

List of Figures v
List of Tables viii

1 Introduction 1
   1.1 Pulsars .................................................. 2
       1.1.1 Early Observations ................................... 2
       1.1.2 High-energy radiation from pulsars ..................... 4
       1.1.3 Theoretical models of the pulsar magnetosphere ....... 5
   1.2 Magnetars .................................................. 14
       1.2.1 Early Observations ................................... 14
       1.2.2 Observational Puzzles ................................. 16
       1.2.3 Theoretical Models .................................... 18
   1.3 This Dissertation ........................................ 22

2 Particle-in-Cell Method 23
   2.1 Introduction .............................................. 23
   2.2 The Particle-in-Cell Method ............................... 24
       2.2.1 Discretization and Spatial Grid ....................... 27
       2.2.2 Current Deposition .................................... 32
2.2.3 Particle Pusher ................................................. 39
2.2.4 Integrating the Maxwell Equations ......................... 43
2.2.5 Coordinate Systems and Boundary Conditions .......... 44
2.2.6 Radiative Transfer ........................................... 50
2.2.7 Units .......................................................... 55
2.3 Aperture ......................................................... 59
2.4 Test Problems .................................................... 62
  2.4.1 Relativistic Two-stream instability ....................... 63
  2.4.2 Cylindrical Waveguide .................................... 64
  2.4.3 Monopole magnetosphere .................................. 67
2.5 Discussions and Remarks ..................................... 69

3 Polar-Cap Discharge ............................................. 71
  3.1 Introduction ................................................... 71
  3.2 Steady-state solution for a charge-separated flow ......... 73
    3.2.1 Basic equations ........................................... 73
    3.2.2 Cold and warm solutions ................................. 76
    3.2.3 Stability of the flow .................................... 77
  3.3 Numerical setup ............................................... 79
  3.4 Results ........................................................ 81
    3.4.1 Steady state tests ........................................ 81
    3.4.2 Time-dependent state with warm particle injection .... 86
  3.5 Mixed ion flow and two stream instability ................ 91
  3.6 Discussion ..................................................... 94

4 Global Pulsar Magnetosphere .................................. 97
  4.1 Introduction ................................................... 97
4.2 Problem formulation and simulation setup ........................................... 99
4.3 Rotators of type I .................................................................................. 101
4.4 Rotators of type II ................................................................................ 105
4.5 Discussion ............................................................................................. 107

5 Twisted Magnetar Magnetosphere ......................................................... 109

5.1 Introduction .......................................................................................... 109
5.2 Sustaining Currents in the Twisted Magnetosphere .............................. 112
  5.2.1 Voltage without pair discharge ...................................................... 113
  5.2.2 Voltage with pair discharge ........................................................... 116
  5.2.3 Characteristic timescales and energy scales ................................. 117
  5.2.4 Mechanism of untwisting ............................................................... 119
5.3 Setup of the simulation ....................................................................... 120
  5.3.1 Implanting the twist ...................................................................... 120
  5.3.2 Surface atmospheric layer ............................................................. 122
  5.3.3 Creation of $e^\pm$ pairs ................................................................. 123
  5.3.4 Rescaling of large numbers in the problem ................................... 123
  5.3.5 Evolving the fields and the plasma: Aperture ............................... 125
  5.3.6 Units ............................................................................................... 127
5.4 Results .................................................................................................. 127
  5.4.1 Initial relaxation .............................................................................. 127
  5.4.2 The equatorial gap ........................................................................ 129
  5.4.3 Dependence on the threshold voltage ......................................... 133
  5.4.4 Expanding cavity .......................................................................... 134
5.5 Discussion ............................................................................................. 138

6 Further Explorations in Extension of PIC ............................................. 141
List of Figures

1.1 A record of pulsating radio source ................................................. 2
1.2 Gamma-ray lightcurves of Geminga .............................................. 4
1.3 Charge density distribution of the electrosphere of an aligned rotator. .... 6
1.4 The force-free magnetosphere. ..................................................... 9
1.5 Breakdown of the polar gap ........................................................... 11
1.6 Space-charge limited polar gap and vacuum polar gap. ....................... 12
1.7 Outer gap .................................................................................. 13
1.8 Slot gap geometry ................................................................. 14
1.9 Light curve for the outburst from transient magnetar XTE J1810-197 ....... 16
1.10 Evolution of the fitted area of the hot component in the X-ray spectrum of XTE J1810-197 ............................................................ 17
1.11 The evolution of hotspots observed on transient magnetars following outbursts ........................................................... 18
1.12 X-ray spectrum of 1E 2259+586 ...................................................... 19
1.13 Formation of the equatorial current sheet in an over-twisted magnetar magneto-
sphere. .................................................................................. 20
1.14 Sketch of a twist magnetic loop ...................................................... 21

2.1 Staggered Yee cell for a 2D spherical grid ......................................... 28
2.2 Staggered Yee cell for a 3D Cartesian grid ......................................... 29
2.3 Three cases for Buneman cell crossing in 2D ........................................ 34
5.1 Electric gap in magnetar magnetosphere ........................................... 130
5.2 Charge density in magnetar magnetosphere ........................................... 131
5.3 Magnetar magnetospheric pair multiplicity ........................................... 132
5.4 Average hydrodynamic momenta in magnetar magnetosphere .............. 133
5.5 Evolution of energy in the magnetar magnetosphere ........................... 135
5.6 Evolution of the poloidal current density in magnetar magnetosphere .... 136
5.7 Time evolution of the integrated current in magnetar magnetosphere ....... 137
5.8 Evolution of the magnetar magnetospheric twist angle ........................... 137

A.1 Positions where the function is evaluated and its derivative is evaluated in 1D .................................................. 164
A.2 Several scenarios for evaluating one-sided derivatives ........................... 167

B.1 Difference in GPU and CPU internal structure ................................. 169
B.2 Memory hierarchy of Nvidia GPUs .................................................... 172
B.3 4-color scheme for current deposition to avoid overlapping of guard cells .................................................................. 174
B.4 Thematic illustration for parallel pair creation ..................................... 175
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Scale functions for coordinate systems implemented in Aperture</td>
<td>45</td>
</tr>
<tr>
<td>A.1</td>
<td>Summary of 1D finite difference operator coefficients for evaluating $f''(x)$</td>
<td>165</td>
</tr>
<tr>
<td>A.2</td>
<td>One-sided derivatives part 1</td>
<td>166</td>
</tr>
<tr>
<td>A.3</td>
<td>One-sided derivatives part 2</td>
<td>167</td>
</tr>
</tbody>
</table>
Acknowledgments

There are many people without whose advice, help and support I would never have been able to finish this thesis.

I can never overstate my gratitude towards my thesis advisor Andrei Beloborodov, who has been ever supportive and patient, even in the darkest of times. An ancient Chinese proverb says “A teacher for a day is a father for a lifetime.” I look up to Andrei as a fatherly figure, both academically and personally, who taught me not only how to think about physics problems, but how to cope with problems in life as well.

I would like to express my gratitude towards Professor Anatoly Spitkovksy, from whom I learned about numerical methods and their potential pitfalls. He also helped me tremendously in the process of postdoc application. Thanks also to Professor Lorenzo Sironi, who helped me a lot with my job applications too, and kindly agreed to serve on my thesis committee.

I would like to thank the colleagues whom I shared many inspiring discussions with. I want to thank Rui Hu, who invested so much time working together with me on this challenging coding project. Thank you to Kyle Parfrey, Romain Hascoët, Indrek Vurm, Christoffer Lundman, Jonathan Zrake, and Xinyu Li; I greatly enjoyed my discussions with you. Thank you to Mickey McDonald and Professor Jeremy Dodd, it was a great pleasure working with you as a preceptor. I would also like to thank my fellow grad students Zhe Liu, Xiao Xiao, Lian Liu, Daiqian Zhang, Shuo Zhang, Xinyuan Ai, Lei Zhou, Jonghee Kang, Carlos Forsythe, Sathish Thiyagarajan, for the fellowship we shared on our common paths towards the PhD.

There are many friends whose company I cherish over this long journey. To Richard Xie, thank
you for always being there for me. To Iris Chen, thank you for always listening to my incessant rants and offering moral support. To Hanfei Mao, thank you for your companionship over the years. To Annabel Huang, I am grateful for the time spent with you in my last year, which made it the most wonderful one among all these years. There are many more who I owe my gratitude to: Chong and Nannan Wang, Sarah Xie, Tim Zhao, Fanlong Zeng, Luchang Jin, Yihan Zhang, Daniel Fleischer, Junpu Wang, Hantao Yin ...

Finally I would like to thank my parents. You have been unconditionally supportive to whatever choices I make in life, whatever dreams I pursue, and however far they take me from home. Without your support and encouragement I would not have been able to make this far. This thesis is dedicated to you.
To the my parents, Bei Lin and Zhiping Chen
Chapter 1

Introduction

Astrophysics builds upon observations of astronomical objects at various radiation bands. Some of the most interesting objects in the sky are sources of high energy radiation, including but not limited to pulsars, supernovae, magnetars, super-massive black holes, etc. The high-energy spectra of these objects up to hard X-ray and gamma-rays are often highly non-thermal, suggesting that it is from high-energy particles. The study of particle acceleration and conversion of other forms of energy into particle kinetic energy is therefore vital in the study of high-energy astrophysics.

In many of the sources, the most notable source of energy for nonthermal particles is the magnetic energy. For example in pulsars, magnetic field plays an intermediary role, acting as a channel that converts the rotational energy of the pulsar into the particle energy, which then is radiated away eventually and produce the observed radio/X-ray/gamma ray emission from the pulsar or in the surrounding pulsar wind nebula. In other sources such as magnetars it is the magnetic energy itself that is directly converted to particle energy that is radiated.

Dissipation of magnetic energy and acceleration of particles is inherently a highly nonlinear process, and a direct calculation is very difficult. The recent development in computational power has enabled direct simulations of various physical scenarios that can lead to particle acceleration, such as collisionless shocks and magnetic reconnection, allowing rapid development in these fields.
Simulations have become a vital tool in understanding high energy plasma astrophysics. With the advent of new architectures like the Graphics Processing Unit (GPU), larger scale simulations are becoming more accessible, enabling direct simulations of global systems like the magnetospheres of neutron stars.

This dissertation will be focusing on the physics of isolated neutron stars. In the following sections we will give a broad overview of the history and basic physics of rotation-powered pulsars, and the exotic magnetars which have extremely high magnetic field. Finally we will give an outline of this thesis.

1.1 Pulsars

1.1.1 Early Observations

The first observational discovery of the rotation-powered pulsars (RPP) dates back to late 1960s (Hewish et al. 1968). In this paper a periodic radio source was reported to be emitting regular pulsation at a frequency of 81.5 MHz, with a period of 1.337 s at extreme accuracy. The pulse width of the object gives an upper bound on its physical size, which should not exceed $4.8 \times 10^3$ km. The extreme constancy of the intrinsic period suggests that the source is a massive object, rather than some astrophysical plasma configuration. It was therefore conjectured that this pulsed radio emission was coming from a compact star: a white dwarf or a neutron star; the extreme regular pulsation is a result of its rapid rotation (Gold 1968).

![Figure 1.1: A record of the pulsating radio source discovered in 1967 (Hewish et al. 1968)](image)

More pulsars were soon discovered, e.g. the Vela with a period of 89 ms (Large et al. 1968) and
the Crab with a period of 33 ms \cite{Lovelace1968}. Slowing down of the periods were also discovered in the known pulsars. The spin-down luminosity can be easily estimated given the period and period derivative of the pulsar:

\[ L_d = -I\Omega\dot{\Omega} = 4\pi^2 I \frac{\dot{P}}{P^3} \tag{1.1} \]

which works out to be \( \sim 10^{39} \) erg/s for the Crab pulsar, which has \( P = 33 \) ms and \( \dot{P} = 4.2 \times 10^{-13} \) s s\(^{-1}\). This matches the observed luminosity of the Crab nebula, and a model was soon proposed by Gold \cite{Gold1969} that gas was liberated from the star and accelerated to relativistic energies, forming a corotating magnetosphere around the star up to the radius where corotation speed becomes equal to the speed of light, \( R_{LC} = c/\Omega \). This radius is called the light-cylinder radius, beyond which corotation is not possible. In this model, the relativistic gas carries away most of the spin-down luminosity, and make its contribution to the luminosity of the nebula.

The spin-down of the pulsar was typically modeled by a spinning magnetic dipole in vacuum. A magnetic dipole of strength \( \mu \) will lose energy at a rate:

\[ L_d = \frac{2\mu^2\Omega^4}{3c^3} \tag{1.2} \]

therefore one can naively estimate the surface magnetic field by equating this with the spin-down luminosity (1.1), which gives:

\[ B_0 = 3.2 \times 10^{19}\sqrt{P\dot{P}} \text{ G} \tag{1.3} \]

For typical pulsar parameters, this gives a polar magnetic field of the order \( \sim 10^{12} \) G. The strong magnetic field requirement rules out the possibility of a white dwarf, and since then a rapid rotating neutron star has been the standard model for a rotation-powered pulsar. Equation (1.3) remains the standard formula for estimating the surface magnetic field of a newly discovered pulsar.
1.1.2 High-energy radiation from pulsars

Although radio emission has been the primary wavelength at which rotation powered pulsars are discovered, only a tiny fraction ($\sim 10^{-5}$) of their spin-down power actually goes into radio emission. For young pulsars and millisecond pulsar (MSPs), a significant fraction of their spin-down power goes into gamma-rays in the 100 MeV–30 GeV band (see e.g. Abdo et al., 2010a).

![Gamma-ray lightcurves of Geminga in five energy ranges](image)

Figure 1.2: Gamma-ray lightcurves of Geminga in five energy ranges, from *Fermi* observations (Abdo et al., 2010b). Two prominent peaks are seen in each rotation period.

Take the pulsar Geminga for example. It is the second brightest non-variable GeV gamma-ray source in the sky. Its gamma-ray emission was first discovered in the 1970s by the SAS-2 satellite (Fichtel et al., 1975; Kniffen et al., 1975). In contrast to the Crab pulsar, Geminga is observed to be radio quiet, and is the first representative of the class of radio quiet gamma-ray pulsars. As can be
seen from figure 1.2, the gamma-ray peaks come in pairs in every period, in contrast to the typical radio peaks in rotation powered pulsars, suggesting that the gamma-ray emission comes from a very different region than the radio emission.

After the launch of the *Fermi* satellite, the catalog of gamma-ray pulsars exploded from 7 to well over 130 (Abdo et al., 2010c, 2013). The gamma-ray pulsars are evenly divided into 3 groups: millisecond pulsars, young radio-loud pulsars, and young radio-quiet pulsars. This discovery revolutionized the way pulsars were studied. The pulsed gamma-ray emission typically carries the highest fraction of the spin-down power $L_d$, therefore it can reveal the most information about the particle acceleration and field structure in the pulsar magnetosphere, much more so than the radio emission.

Many pulsars are also found to be X-ray sources, with pulsations detected in many of them. The emission is usually made up of two components: a thermal component from surface cooling or heated polar caps, and a non-thermal component that is most likely magnetospheric due to synchrotron radiation (Kaspi et al., 2006).

### 1.1.3 Theoretical models of the pulsar magnetosphere

Despite the success of a simple vacuum dipole model, it proves to be extremely difficult to obtain a more detailed self-consistent model of the pulsar magnetosphere. The vacuum dipole model has a few problems. First, the electric field from uni-polar induction effect gives a voltage that can accelerate particles to $\sim 10^{14}$ V. This voltage has two effects: it far exceeds the binding energy of electrons and ions at the surface of the neutron star and can extract charged particles from the stellar surface; it also accelerates the extracted particles to energies that produce high-energy photons that are capable of interacting with the intense magnetic field and convert into $e^\pm$ pairs, inducing a pair cascade (Erber, 1966).

As a result, it is not possible for the pulsar magnetosphere to be near vacuum. A minimum corotating charge density is guaranteed around the pulsar which is conventionally called the
Goldreich-Julian density (Goldreich & Julian, 1969):

\[ \rho_{\text{GJ}} = \frac{\Omega \cdot B}{2\pi c} \frac{1}{1 - (\Omega r/c)^2 \sin^2 \theta} \]  

(1.4)

where \( \theta \) is the angle between magnetic axis and the rotation axis.

**Electrosphere**

![Figure 1.3: Charge density distribution of the electrosphere of an aligned rotator. Blue indicates negative charge and red indicates positive charge. Snapshot taken at time of 47.5 \( R_\ast /c \) which is about 1.5 rotations of the star, and quasisteady state has been achieved.](image)

The problem of the Goldreich-Julian model is that charge lifted from the surface alone is not sufficient to fill the magnetosphere with \( \rho_{\text{GJ}} \). For an aligned rotator (magnetic axis aligned with
rotation axis), there exists an electrostatic equilibrium solution for the lifted charge (Jackson, 1976; Krause-Polstorff & Michel, 1985b,a). The surface charges are lifted to form a dome around both poles of the star, and a torus near the equator. In both the dome and the torus \( \mathbf{E} \cdot \mathbf{B} = 0 \), whereas an unscreened vacuum gap exists between them where parallel \( \mathbf{E} \) field is nonzero (figure 1.3). This equilibrium solution has no outgoing Poynting flux, therefore no spin-down at all. It is a dead pulsar.

It was speculated by Spitkovsky (2004) that oblique rotators will be able to escape this fate due to diocotron instability developing inside the torus, which leads to its slow expansion and eventually reaching the light cylinder. This might jump-start the global current circulation and allow the pulsar to start operating, but no simulation or analytical results have been able to confirm this hypothesis. Furthermore, Pétri (2007) showed that this instability is suppressed by relativistic effects that become important near the light cylinder.

Another problem with the electrosphere is that a huge region with unscreened parallel electric field exists between the dome and torus capable of accelerating stray particles to very high energies. These particles will be able to produce curvature photons that are able to interact with the magnetic field to produce \( e^\pm \) pairs. Any cosmic ray particle can initiate this process and produce a pair avalanche: the electrosphere solution is unstable to pair creation.

Pair creation in the pulsar magnetosphere has been studied extensively since the first theoretical models of the pulsar. There are several motivations for the pulsars to produce abundant \( e^\pm \) outflow. One stems from models attempting to explain the radio emission which was central to pulsar studies for decades. Most models for radio emission involve some plasma instability causing the clumping of charges, which requires a dense quasi-neutral plasma, not the charge-separated dome and torus in the electrosphere solution. Another motivation is that the observed synchrotron radiation in the pulsar wind nebulae calls for large numbers of \( e^\pm \) pairs. In the case of the Crab, the multiplicity of pairs (number of \( e^\pm \) pairs over the minimum Goldreich-Julian density) is estimated to be \( M \gtrsim 10^6 \) (de Jager et al., 1996), thus requiring abundant pair creation in the magnetosphere.
Therefore, particle acceleration and pair creation have been a major topic in theoretical pulsar research for decades. One of the challenges for any theoretical model is that pair creation is naturally a self-limiting process: the creation of abundant neutral plasma tends to screen the accelerating electric field, thus reducing its efficiency or even turning off the process altogether. Therefore it is difficult to have large regions in the magnetosphere where there is unscreened parallel electric field: all pulsar particle acceleration models involve a somewhat local “gap” where unscreened electric field keeps accelerating particles, and pairs are created outside the “gap”, unable to screen it. It is not clear \textit{a priori} whether the gap (or gaps) would be static or periodically turning on and off, or whether the position would be static in time.

\textbf{MHD and force-free models}

Although the existence of gaps is crucial to fill the magnetosphere with the required amount of plasma, it is instructive to study what happens to the magnetosphere when plasma supply is not an issue. It makes sense to study the approximation where plasma is so abundant as to screen all the parallel electric field, $\mathbf{E} \cdot \mathbf{B} \approx 0$. The model assumes that the inertial mass of the particles is much less than the magnetic field energy $B^2/8\pi c$. This limit is called “Force-free electrodynamics” (FFE), defined by the equation

$$\rho \mathbf{E} + \frac{\mathbf{J} \times \mathbf{B}}{c} = 0 \tag{1.5}$$

The FFE condition is basically an equation for the current $\mathbf{J}$, which closes the Maxwell equations. Force-free electrodynamics has no constraints on the distribution of the plasma, other than the obvious requirement that $\rho = \nabla \cdot \mathbf{E}/4\pi$, and that $\mathbf{J}$ satisfies the force-free equation \textit{(1.5)}. It assumes that enough plasma is always supplied to maintain these two conditions as demanded by the electromagnetic field.

The force-free equation together with Maxwell equations form a closed system and can be solved with the boundary condition of a rotating neutron star. The equations were first solved numerically by Contopoulos et al. \textit{(1999)} (see also e.g. Goodwin et al. \textit{2004} Gruzinov \textit{2005}.)
The magnetosphere is split into an open zone with outward Poynting flux and a closed zone with no energy flux. The separatrix between these zones forms a Y-shaped current sheet.

- An open zone where magnetic field lines extend to infinity, $B_\phi \neq 0$ and current flows along the field lines. Poynting flux $\mathbf{S} = c \mathbf{E} \times \mathbf{B} / 4\pi$ points outward along the poloidal $\mathbf{B}$ field.

- A closed zone where $B_\phi = 0$ and poloidal $j = 0$, filled by plasma with density equal to $\rho_{GJ}$. The closed zone corotates with the star, similar to the torus in the electrosphere solution. Poloidal magnetic field remains close to dipole configuration.
• A thin Y-shaped current sheet that separates the close and open zones, with the Y-point close to the light cylinder.

These features are nicely summarized in figure 1.4.

However, FFE remains an approximation that glaringly breaks down in some regions of the magnetosphere, namely inside the current sheets and in some inevitable gaps where \( \mathbf{E} \cdot \mathbf{B} \neq 0 \) and particles are accelerated to produce the plasma required by FFE. One can introduce finite resistivity and use a full resistive MHD approach to study the magnetosphere (e.g. Kalapotharakos et al., 2014). However the issue of particle acceleration and formation of localized gaps remains impossible to tackle in this framework.

**Gaps and pair cascade**

Several mechanisms of forming and maintaining localized gaps where particles are accelerated have been proposed over the decades of pulsar research. The classical picture was that the gap exists at the pulsar polar cap (Sturrock, 1971). The motivation here is that, for open field lines that penetrate the light cylinder, the plasma corotating on the field lines can’t move faster than the speed of light, therefore they have to lag behind the corotating field lines at the light cylinder, bending the field lines backwards to create a spiral-like pattern. This creates non-zero \( \nabla \times \mathbf{B} \) therefore nonzero current along the open field lines: current flowing from the polar cap region. Sturrock (1971) considered space-charge limited flow (Pierce, 1954) from the polar cap and estimated the voltage drop from the surface of the star, and it was enough for particles to emit energetic gamma-rays that can convert into \( e^\pm \) pairs in the strong magnetic field. The polar caps operate as “guns” shooting electron-positron pairs along the open field lines and serve as the source of radio emission. From the pulse profile, the width of the pulse is very small, indicating that radio emission is coming from close to the star. Therefore the accelerating gap must be close to the star as well.

Instead of space-charge limited flow, Ruderman & Sutherland (1975) considered the case where
positive charges need to be extracted from the star (e.g. anti-aligned rotator) but the electric field is not enough to overcome the binding energy. In this case a vacuum gap near the polar cap is developed and a pair discharge is initiated (figure 1.5). In the originally vacuum gap, any stray electron or positron from say cosmic rays can initiate this process of pair avalanche, where the stray particle is accelerated and produces highly energetic curvature radiation. The photon propagates a short distance before converting to an $e^\pm$ pair, which then serve as seeds for the same process. The gap height $h$ is self-regulated such that this process does not shut down, and keeps generating fresh plasma as it flows out to the light cylinder to conduct the required current. The difference between these two models lies in whether primary particles are extracted from the star, or recycled from the previous generation of created pairs (figure 1.6).

Another gap model is the outer gap, proposed by Cheng et al. (1986), which originate from the region in the magnetosphere where $\rho_{GJ} = 0$, or in other words, $\Omega \cdot B = 0$. A charge-separated flow from the star cannot penetrate this boundary since the space charge density changes sign here. Therefore, if the current from the star is carried by particles extracted from the surface, this
is where an unscreened gap will develop, leading to a pair discharge in the outer magnetosphere.

Figure 1.6: Space-charge limited polar gap vs vacuum polar gap. \( T_s > T_{ei} \) for \( \Omega \cdot B > 0 \) and \( T_s < T_{ei} \) for \( \Omega \cdot B < 0 \).

This outer gap model predicts a double pulse gamma-ray light curve that is similar to that observed in Crab and Vela, which is one of the reasons that it is popular among observers in explaining the emission from gamma-ray pulsars. However, the outer gap picture assumes a vacuum dipole field configuration even up to the outer magnetosphere, whereas we know from FFE simulations that the field structure is significantly altered by the presence of plasma. Bai & Spitkovsky (2010) used the more realistic field configuration from 3D FFE calculations and showed that in fact this traditional outer gap model is capable of producing only one peak under general conditions because a large fraction of open field lines do not cross the null surface.

Another gap model is the slot gap which is an extension of the polar gap along the last closed field line (figure 1.8). It also aims at explaining the high energy gamma-ray emission from pulsars. The original idea was proposed by Arons & Scharlemann (1979) who realized that the acceleration potential varies significantly across the polar cap, resulting in an extended pair creation front almost parallel to the last closed field line. This gap is capable of slowly accelerating particles along the field lines to high energies at a much higher altitude than the polar cap gap, producing...
CHAPTER 1. INTRODUCTION

Figure 1.7: Outer gap, from Cheng et al. (1986).

gamma-ray emission from curvature radiation.

Despite the decades of effort, a global picture of the pulsar magnetosphere with both a realistic FFE field structure and a gap that self-consistently generates the required plasma is still lacking. The physics of the pulsar magnetosphere is surprisingly rich and it is important to take into account the interplay between plasma physics and radiative transfer at very high energies. A global simulation from first principles is needed to fully understand how pulsars work.
1.2 Magnetars

Magnetars are a class of neutron stars with very drastic variability in X-ray and soft $\gamma$-ray bands. They exhibit recurrent bursts, flares, and sometimes giant bursts that can briefly outshine entire galaxies in their X-ray luminosity. Their activity is powered by the decay of their strong magnetic field, which is typically 100 times higher than ordinary rotation-powered pulsars.

1.2.1 Early Observations

Historically “magnetar” was not the name given upon its discovery. The first report of magnetar activity can be traced back to 1979. The Venera 11 and Venera 12 space probes recorded 3 repeated soft gamma-ray bursts from a single source B1900+14 \cite{Mazets1979}. It was initially believed to share similar origins with other short gamma-ray bursts. During the same year, the space probes also recorded hard X-ray bursts from a different source FXP 0520-66 in
Dorado, which was apparently an X-ray pulsar from the beginning (Mazets et al. 1979b). These sources were designated "Soft Gamma-ray Repeaters" (SGRs).

Several years later another SGR 1806-20 was found in our galaxy, and it underwent repeated bursts on the order of 100 times over less than 10 years (Kouveliotou et al. 1987; Laros et al. 1987). It was also the first SGR that had a measured spin-down rate (Kouveliotou et al. 1998). The dipolar magnetic field computed by the simple spin-down formula (1.3) gives a surface field strength of $8 \times 10^{14}$ G, much higher than typical rotation-powered pulsars yet discovered. The ultrastrong magnetic field of SGR 1806-20 was in line with the magnetar model proposed by Duncan & Thompson (1992). In their paper they coined the term "magnetar" to describe young neutron stars with high magnetic field ($10^{14} \sim 10^{15}$ G), and argued that magnetars are the sources of SGRs, where the bursts are powered by spontaneous decay of the strong magnetic field of the magnetar. Being young objects with age $10^3 \sim 10^4$ years, magnetars are still dynamically evolving internally and building up stress that can lead to breaking of the crust, releasing a significant amount of energy into the magnetosphere in the form of Alfvén waves, which then powers the X-ray and soft gamma-ray emission.

Another class of magnetars was discovered separately and recognized as “Anomalous X-ray Pulsars” (AXPs). In 1980 it was reported that in the region CTB 109 there was “an extraordinary new celestial X-ray source” which was a supernova remnant (Gregory & Fahlman 1980). Soon it was found that this source displays very strong pulsation with a period of $\sim 3.5$ s (Fahlman & Gregory 1981). More X-ray pulsars were subsequently discovered that also have similar few-second period, and share similar soft X-ray spectra.

Thompson and Duncan speculated that AXPs may be related to SGRs, and predicted that long-term observations of AXPs might lead to bursting behavior (Thompson & Duncan 1996). This was then confirmed by observation. Now both AXPs and SGRs are recognized under the same umbrella known as magnetars. Among the $\sim 30$ known magnetars, a more intrinsic characterization is their quiescent luminosity. Transient magnetars are typically only detected during their bursts,
while *persistent magnetars* are bright even in their quiescent state. They are usually what used to be associated with AXPs, showing strong and persistent pulsed X-ray emission.

### 1.2.2 Observational Puzzles

![Light curve for the outburst from transient magnetar XTE J1810-197](image)

Figure 1.9: Light curve for the outburst from transient magnetar XTE J1810-197 (Gotthelf & Halpern, 2007).

For transient magnetars, apart from short and irregular bursts that are the signature of SGR activity, they also show large outbursts with short rise time and long decay. A classic example is the transient magnetar XTE J1810-197, which underwent an outburst in early 2013 (Ibrahim et al., 2004), and its luminosity slowly decayed to quiescent levels over several years (figure 1.9). Gotthelf & Halpern (2007) found that the X-ray spectrum of the magnetar during outburst can be fitted by 2 blackbody components, one with lower temperature and larger area, which subsequently cooled but expanded to cover almost the entire star, and another with smaller area and higher temperature, *shrinking* over time (figure 1.10). This has very important implications for modeling the magnetar outburst: the model should be able to explain the origin of the hotspot, as well as
the reason and timescale of its shrinking.

Figure 1.10: Evolution of the fitted area of the hot component in the X-ray spectrum of XTE J1810-197. The area shrunk by a factor of more than 8 over the course of a few years.

The shrinking hotspot feature was not only seen in one transient magnetar. Figure 1.11 shows the evolution of luminosity versus the area of the fitted hotspot for 7 of the known transient magnetars for which a hotspot has been identified. All seem to follow a similar trajectory over the $A-L$ plane.

Persistent magnetars also pose an important observational puzzle. Previously AXPs were observed to have a soft X-ray spectrum that can be described by a blackbody plus power law. The X-ray spectrum turns down and was predicted to be not detectable above 10 keV. However Kuiper et al. (2006) discovered hard spectral tails for 3 persistent magnetars 1RXS J1708–4009, 4U 0142+4009, and 1E 2259+586. This turned out to be a great surprise. Furthermore the X-ray component above 10 keV is extraordinarily hard and extends up to and beyond 150 keV. Later Enoto et al. (2010a) reported the identification of the hard X-ray tail in 7 magnetars including 2 transient magnetars in outburst states, namely SGR 0501+4516 (Enoto et al., 2010c) and 1E 1547.0–5408 (Enoto et al., 2010b). By now it has been recognized that this soft emission plus hard tail is a common feature for magnetars. This high-energy nonthermal component dominates the
Figure 1.11: The evolution of hotspots observed on transient magnetars following outbursts. The hotspots shrink and become dimmer over time, tracking a similar trajectory on the $A$-$L$ plane. (Beloborodov & Li, 2016)

emission energy, and is thought to be of magnetospheric origin. This hard X-ray component is another interesting puzzle posed by the magnetar observations.

### 1.2.3 Theoretical Models

Magnetars are all very young neutron stars that have very high surface magnetic field. The interior of a neutron star is an excellent conductor, and magnetic field is practically frozen into the material. Field evolution is a result of the evolution of electron fluid coupled to the ion and neutron fluids, which can be described by two processes: ambipolar diffusion and Hall drift (Goldreich & Reisenegger, 1992). These processes move magnetic field lines and can build up magnetic stress inside a neutron star, which could lead to a sudden failure of the crust and ejection of this energy into the magnetosphere (Thompson & Duncan, 1995, 2001). An alternative way of triggering this release is a slow build up of the stress due to a gradual deformation of the magnetosphere, as long as this process is faster than the rate it is damped (Lyutikov, 2006).
In any case, the result of surface shear is the deformation of the external magnetosphere from the quasisteady configuration, twisting the magnetic field lines and launching Alfvén waves into the magnetosphere, creating regions where $\nabla \times \mathbf{B} \neq 0$. Current will flow along the twisted field line bundle, $\mathbf{j} = (c/4\pi) \nabla \times \mathbf{B}$. A strongly twisted magnetosphere is prone to global instability \cite{Uzdensky2002}, and will result in the formation of an equatorial current sheet, where magnetic reconnection happens and energy is released violently in a short time scale. This was seen in force-free simulations carried out by \cite{Parfrey2013}.

When the twist is not as dramatic, the current bundle can be long-lived. This should feed the observed long decay of the X-ray luminosity after an outburst event. \cite{Beloborodov2007} studied the dynamics of the current loop, concluding that electric field will be induced to extract particles from the surface of the star, accelerate them, and initiate a pair avalanche similar to that in pulsar magnetosphere. The current loop acts as a “corona” of the magnetar. Particles are lost to the stellar surface over one light crossing time of the system and are constantly replenished from pair creation. The main channel for pair creation is from photons upscattered resonantly: electrons moving at Lorentz factor of $\sim 1000$ will see a background sea of soft X-ray photons of a
CHAPTER 1. INTRODUCTION

Figure 1.13: Formation of the equatorial current sheet in over twisted magnetar magnetosphere. Color shows toroidal current density. Time is indicated in units of light crossing time of the star. From (Parfrey et al., 2013)

few keV, which when boosted into the rest frame of the electron matches the energy to excite the particle from the ground Landau level to the first excited level. When this resonance condition is satisfied, the electron will be able to absorb the photon and re-emit it. The re-emitted photon will see an energy boost of $\sim \gamma^2$ in the lab frame, and capable of creating $e^\pm$ pairs by means of magnetic conversion.

The lifetime of the magnetar twist is determined by the acceleration voltage induced in the $j$-bundle, which is in turn governed by the threshold of pair discharge. However the self-regulation of this process is a non-trivial problem. It was also discovered by Beloborodov (2009) that the process of untwisting proceeds in an interesting manner. A current cavity develops first in the inner magnetosphere, which subsequently expands and erases the current carrying part of the twisted field lines. This effectively means that the cross-sectional area of the current bundle shrinks as the magnetosphere untwists over time. If the hotspot seen in the blackbody spectrum after an outburst is to be mapped to the footprint of the $j$-bundle, then this provides a natural explanation of the shrinking hotspot.
Resonant scattering not only provides a means to generate high-energy photons that can convert to pairs, it also affects the relativistic flow of the pair plasma along the twisted field lines. Beloborodov (2013) proposed a mechanism for the hard X-ray component seen in the magnetar spectrum. Particles are extracted from the star and undergo acceleration to $\gamma \gg 10$. In this region the resonantly scattered photons interact with the local $B$ field and quickly convert to $e^\pm$ pairs, effectively pair loading the plasma with multiplicity $M \sim 100$. When the flow gets to a region with weaker magnetic field, the resonantly scattered photon will be able to escape to form the observed X-ray spectrum (figure 1.14). The resonant up-scattering of the target photons will act as an effective drag on the plasma flow, reducing its hydrodynamic speed. Finally at the tip of the magnetic loop near the equator, the radiative drag will be strong enough to stop the particles to $\gamma \sim 1$. The particles suspended near the equator will serve as a reflector for the X-ray photons from the star. They might also be able to generate radio emission which is seen in some of the magnetars.

Figure 1.14: Sketch of a twist magnetic loop. Particles are accelerated in the blue region and resonantly scatter photons reflected from the pink region. The upscattered photons convert to pairs near the star, but can escape in the white region to form the observed X-ray spectrum (Beloborodov, 2013).

The picture described above was successfully used to fit the phase-resolved spectra of several
magnetars (Hascoët et al., 2014). The narrowness of the fitted parameters space is a strong indication for the correctness of the model. It would be a decisive confirmation if the model can be reproduced in a first-principle kinetic simulation.

1.3 This Dissertation

In this dissertation we will attempt to study the problem of particle acceleration and global structure of the magnetosphere of pulsars and magnetars using numerical experiments. Most of the work is done using a computer code named Aperture that I developed in the course of the PhD.

Chapter 2 will be devoted to a detailed exposition of the particle-in-cell method, which will be the basic numerical tool for our study. We will introduce the Aperture code, and explain its novel features, as well as providing some tests to demonstrate its correctness.

Chapter 3 will focus on a local study of the pulsar polar cap. We will look at the region well within the pulsar magnetosphere, and approximate the geometry as 1D. We will discuss implications of this approximation, and what we can and can not learn from this local study.

Chapter 4 will study the global pulsar magnetosphere, motivated by the study of the polar cap particle acceleration. We will discuss the result from global PIC simulations and contrast it with force-free results, and study the condition under which a pulsar can sustain itself through pair discharge.

Chapter 5 will study the twisted magnetosphere of magnetars and attempt to answer how particles are accelerated in the current bundle and what regulates the overall voltage.

Chapter 6 will contain some exploratory extensions of the PIC code into the general relativistic regime.
Chapter 2

Particle-in-Cell Method

2.1 Introduction

The Particle-in-Cell (PIC) technique is a powerful tool to study the kinetic properties of plasma from first principles. The strength of a PIC code is that it can resolve the plasma skin depth, faithfully reproducing the microscopic interaction between particles and fields, making them invaluable in the study of particle acceleration processes in plasmas. Also it is relatively straight-forward to implement and easily parallelizable. It has been very successfully applied to collisionless shocks and reconnection processes in astrophysics. However, because a PIC code has to resolve the plasma skin depth which is usually many orders of magnitude smaller than the scale of a realistic astrophysical problem, this kind of simulation is extremely expensive and very often only applicable to local problems to study the microphysics in a relatively large system.

This chapter introduces the PIC technique in detail, and then introduces Aperture, a versatile PIC code designed and developed from scratch as part of my PhD thesis. The original and main purpose of the code was to simulate the global structure of the pulsar magnetosphere from first principles. However, we designed the code to be general enough to be applied to many different problems in Astrophysics, especially in problems where particles are accelerated in strong magnetic fields and capable of producing pair cascade.
In Section 2.2 we will present the numerical algorithms and techniques employed in PIC codes, including the standard Yee staggered grid, Boris and Vay pusher, and charge conserving current deposition. We will also explain some of the novel features of the Aperture code, including the treatment of general curvilinear coordinate systems, boundary conditions, and radiative processes. Section 2.3 gives a general overview of the Aperture code, including its architecture and highlighting some implementation details. Section 2.4 presents some test cases to demonstrate the validity of the code. Finally Section 2.5 discusses the strengths and weaknesses of the code, and remark on the outstanding challenges and potential future applications.

2.2 The Particle-in-Cell Method

The PIC method is essentially a way to solve the coupled Maxwell-Vlasov equations by approximating the plasma distribution function using the sum of a large number of discrete macro-particle distributions, each resembling a physical particle. The system of equations under question is simply the Maxwell equations combined with the Vlasov equation:

\[
\frac{\partial f_s}{\partial t} + \mathbf{u} \cdot \frac{\partial f_s}{\partial \mathbf{x}} + \frac{q_s}{m_s} \left( \mathbf{E} + \frac{\mathbf{u} \times \mathbf{B}}{c} \right) \cdot \frac{\partial f_s}{\partial (\gamma \mathbf{u})} = 0 \quad (2.1)
\]

\[
\nabla \cdot \mathbf{E} = 4\pi \rho \quad (2.2)
\]

\[
\nabla \cdot \mathbf{B} = 0 \quad (2.3)
\]

\[
\nabla \times \mathbf{E} = -\frac{1}{c} \partial_t \mathbf{B} \quad (2.4)
\]

\[
\nabla \times \mathbf{B} = \frac{1}{c} \partial_t \mathbf{E} + \frac{4\pi}{c} \mathbf{j} \quad (2.5)
\]

where \( s \) denotes the particle species (electrons, positrons, ions, ...). We assume a collisionless plasma by setting the right hand side of equation (2.1) to zero, which is applicable for many astrophysical applications including collisionless shock acceleration, magnetic reconnection, and pulsar wind nebulae. In these astrophysical systems the effective free path of electrons and
positrons are much larger than the size of the system itself, rendering the collision term negligible. All plasma interactions are mediated by the electromagnetic field.

The charge and current densities appearing in the Maxwell equations as source terms are found by taking the moments of the distribution function $f^s$ over the momentum space

$$\rho = \int d\mathbf{u} \sum_s q^s f^s(\mathbf{x}, \mathbf{u})$$  \hspace{1cm} (2.6)

$$\mathbf{j} = \int d\mathbf{u} \sum_s q^s \mathbf{u} f^s(\mathbf{x}, \mathbf{u})$$  \hspace{1cm} (2.7)

Macro particles are introduced to sample the distribution function $f^s$ in both position and momentum space. We approximate the distribution function of each species by sampling it with a finite but large number of macro particles each with a smeared out distribution in space but unique momentum $\mathbf{p}$:

$$f^s(\mathbf{x}, \mathbf{u}) = \sum_p f^s_p(\mathbf{x}, \mathbf{u}) = \sum_p \delta(ym_p \mathbf{u} - \mathbf{p})S(\mathbf{x} - \mathbf{x}_p)$$  \hspace{1cm} (2.8)

where $S$ is a function that describes the shape of the macro particle, with the property that $S$ has finite support, and that the integral of $S$ over all space is normalized to 1. Since the Vlasov equation is linear, if each individual macro particle satisfies the Vlasov equation, then the linear superposition of a large number of them still satisfy the Vlasov equation, and should provide a good approximation for the dynamics of the plasma.

The dynamic equations for the macro particles can be derived by taking the moments of the Vlasov equation with the single particle distribution function (2.8). Plugging the single particle distribution function into the Vlasov equation and taking the zeroth moment by integrating over
\(\mathbf{u}\), we get

\[
\int d\mathbf{u} \left\{ \left[ \partial_t \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \right] S(\mathbf{x} - \mathbf{x}_p) + \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \partial_t S(\mathbf{x} - \mathbf{x}_p) \\
+ \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \mathbf{u} \cdot \nabla S(\mathbf{x} - \mathbf{x}_p) \\
+ S(\mathbf{x} - \mathbf{x}_p) \frac{q_p}{m_p} \left( \mathbf{E} + \frac{\mathbf{u}}{c} \times \mathbf{B} \right) \cdot \nabla_{\gamma u} \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \right\} = 0
\]

(2.9)

The first term is an integral of the derivative of a delta function, which should give zero since \(S\) is independent of the integration variable \(\mathbf{u}\). The second and third term can be integrated trivially to get

\[
\frac{d\mathbf{x}_p}{dt} = \mathbf{u}_p = \frac{\mathbf{p}_p}{\gamma_p m_p}
\]

(2.10)

since \(\mathbf{x}_p\) depends only on \(t\), the partial derivative becomes a total derivative, \(\partial_t S(\mathbf{x} - \mathbf{x}_p) = -\nabla S(\mathbf{x} - \mathbf{x}_p) \cdot d\mathbf{x}_p/dt\). The last term which involves the electromagnetic force requires a bit more attention. The \(\mathbf{E}\) field term again integrates to zero since it is an integral of the derivative of a delta function. The Lorentz force term needs an integration by parts, but then it would become zero because \(\nabla_{\gamma u} \cdot (\mathbf{u} \times \mathbf{B})\) is zero.

Taking the first moment of the Vlasov equation yields

\[
\int d\mathbf{u} \left\{ \mathbf{u} \left[ \partial_t \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \right] S(\mathbf{x} - \mathbf{x}_p) + \mathbf{u} \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \partial_t S(\mathbf{x} - \mathbf{x}_p) \\
+ \mathbf{u} \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \mathbf{u} \cdot \nabla S(\mathbf{x} - \mathbf{x}_p) \\
+ \mathbf{u} S(\mathbf{x} - \mathbf{x}_p) \frac{q_p}{m_p} \left( \mathbf{E} + \frac{\mathbf{u}}{c} \times \mathbf{B} \right) \cdot \nabla_{\gamma u} \delta(\gamma m_p \mathbf{u} - \mathbf{p}_p) \right\} = 0
\]

(2.11)

The second and third terms give the same equation of motion as above, and they are proportional to the gradient of \(S\), independent of the other two terms, therefore we ignore them. The last two terms can be reduced to the simple single-particle equation of motion:

\[
\frac{d\mathbf{p}_p}{dt} = q_p \left( \mathbf{E} + \frac{\mathbf{u}_p}{c} \times \mathbf{B} \right)
\]

(2.12)
These are simply equations of motion for ordinary particles in an electromagnetic field. Therefore it is justified to treat macro particles as their name suggests: simply as physical particles. A PIC code simply traces their motion in the electromagnetic field as described by the above dynamic equations.

### 2.2.1 Discretization and Spatial Grid

So far the only approximation we have introduced is using an ensemble of macro particles to approximate the real distribution function of a plasma system. To solve the Maxwell equations and particle dynamic equations numerically, one needs to discretize the continuous field quantities onto a finite grid consisting of cells, hence the name Particle-in-Cell. The discretization is done on space and time using the Finite Difference Time Domain (FDTD) method. Fields $E$ and $B$ are sampled on a finite grid, as well as the current and charge densities $J$ and $\rho$. One evolves the equations using a given time evolution scheme step by step starting from the initial condition. At each step, one updates the positions and momenta of all particles according to the fields on the grid, computes the current density due to particle motion, and uses this current density to evolve the fields themselves.

A PIC code can operate in either 2D or 3D\footnote{In 1D, a discretization is not necessary, since there is no magnetic field, and electric field at any given particle location can be found exactly by integrating Gauss’s law. See chapter 3.}. In the former case, although a grid of lower dimension is used, all 3 vector components of $E$ and $B$ fields need to be evolved\footnote{A code like this is sometimes called “2.5D” due to full 3D vector quantities defined on a 2D grid.} This is applicable when the problem has inherent symmetry, such as axisymmetry or translational invariance in one direction. It is typical to use the classical staggered\footnote{Yee (1966)} grid for electric and magnetic fields (figures 2.1 and 2.2).

The reason for staggering the fields this way is that all numerical derivatives that arise in Maxwell equations will be centered naturally and have at least second order accuracy. For example
CHAPTER 2. PARTICLE-IN-CELL METHOD

The electric field values needed show up exactly where they are defined on the Yee lattice, and due to the symmetric structure of the finite difference, the $\Delta x^2$ term in the Taylor expansion naturally cancels:

$$\partial_x E_y = \frac{E_y(x + \Delta x/2) - E_y(x - \Delta x/2)}{\Delta x} + O(\Delta x^3) \quad (2.14)$$

This naturally applies to all derivative terms in the Maxwell equation.

Another strength of the Yee grid is that the differential relations $\nabla \cdot (\nabla \times F) = 0$ and $\nabla \times \nabla f = 0$ are satisfied by construction, so long as Maxwell equations are used for the evolution of fields, $\nabla \cdot B = 0$ should be preserved to numerical precision without extra work.

This construction of symmetric finite difference scheme allows the extension to higher orders naturally. Equation (2.14) is accurate to second order in $\Delta x$. To achieve 4th order accuracy, we...
need to use more terms to cancel out the $\Delta x^3$ term in the Taylor expansion:

$$\partial_x E_y = \frac{9}{8} \frac{E_y(x + \Delta x/2) - E_y(x - \Delta x/2)}{\Delta x} - \frac{1}{24} \frac{E_y(x + 3\Delta x/2) - E_y(x - 3\Delta x/2)}{\Delta x} + O(\Delta x^5) \quad (2.15)$$

Further discussion of the implementation of higher order finite difference schemes can be found in appendix A.

Macro-particles stream freely in the grid, and their positions and momenta are not discretized. To convert between local "particle" quantities and grid variables, an interpolation scheme is required. Fortunately we already have a function $S(x - x_p)$ which describes the shape of the macro particle smeared in space. Integrating equation (2.6) using a macro particle distribution function over the volume of a cell gives:

$$Q_c = \rho_c \Delta V = \sum_p q_p S(x_p - x_c) \quad (2.16)$$

where $q_p$ is the charge of an individual macro particle, $x_c$ is the position of the center of the cell,
and $S$ is defined as the integral of $S$:

$$S(x_p - x_c) = \int_{x_c - \Delta/2}^{x_c + \Delta/2} S(x_p - x) \, dx$$

(2.17)

and $\Delta$ is the size of the cell. The assumption that $S$ has finite support implies that $S$ will be nonzero for only a few cells with $x_c$ close to $x_p$, and since the integration of $S$ over the whole volume is constrained to be 1, the sum of all nonzero $S$ near a certain particle is guaranteed to be 1. This is another way of saying total charge is conserved. Since in the PIC code we will only be using the discrete interpolation functions $S$, not the original $S$, we will call $S(x_p - x_c)$ the “shape function”. Typical shape functions used in PIC codes are derived from so-called “B-spline” functions, which are piecewise polynomial functions with minimal support. Following are the shape functions often used in PIC codes, in ascending polynomial order:

**CIC: Cloud in Cell**

$$S^1(x) = \begin{cases} 
1 - |\delta| & \text{if } |\delta| < 1, \\
0 & \text{otherwise}
\end{cases}$$

(2.18)

**TSC: Triangular Shaped Cloud**

$$S^2(x) = \begin{cases} 
\frac{3}{4} - \delta^2 & \text{if } |\delta| < 1/2, \\
\frac{1}{2} \left( \frac{3}{2} - |\delta| \right)^2 & \text{if } 1/2 \leq |\delta| < 3/2, \\
0 & \text{otherwise}
\end{cases}$$

(2.19)

**PCS: Piecewise Cubic Spline**

$$S^3(x) = \begin{cases} 
\frac{1}{6} (4 - 6\delta^2 + 3|\delta|^3) & \text{if } |\delta| < 1, \\
\frac{1}{6} (2 - |\delta|)^3 & \text{if } 1 \leq |\delta| < 2, \\
0 & \text{otherwise}
\end{cases}$$

(2.20)
These functions are the integration of the B-spline functions of the corresponding polynomial order ([Haugboelle et al., 2012]). In 2D or 3D problems, the weight of the particle is given by multiplication of these shape functions, e.g. in 3D Cartesian coordinates
\[ S(x_p - x_c)S(y_p - y_c)S(z_p - z_c). \]
In general, higher order shape functions will have better noise properties for the result, but are computationally more intensive, not only because they involve more multiplications (higher order polynomial), but each particle can influence more grid points and one needs to sum over more terms of nonzero \( S \).

Equation (2.16) can be taken as the definition of the discretized charge density in a given cell. In other words, it is the average charge contained in the cell assuming the cell is uniformly filled with charges from the macro particles.

The particle shape function also serves as a way to interpolate the grid quantities such as \( E \) and \( B \) fields to the particle location:

\[
E(x_p) = \sum_c E(x_c)S(x_p - x_c)
\]

\[
B(x_p) = \sum_c B(x_c)S(x_p - x_c)
\]

where the summation is over the grid points where \( S \neq 0 \).

One could also define and compute the current density in the same way as charge density, from equation (2.7):

\[
j(x_c) = \sum_p \frac{q_p}{\Delta V} u_p S(x_p - x_c)
\]

However, simple application of this equation will lead to charge conservation issues and violation of the continuity equation (see e.g. [Hockney & Eastwood, 1981, Birdsall & Langdon, 1991]):

\[
\partial_t \rho + \nabla \cdot j = 0
\]

To enforce charge conservation at every timestep, instead of interpolating on the current,
it is desirable to solve the continuity equation directly at every timestep. This is done with the so-called charge-conserving current deposition. We will discuss various techniques to achieve this in section 2.2.2.

2.2.2 Current Deposition

There are various ways to achieve charge conservation numerically. The simplest way is to use the naive current deposition (2.23), being aware of the fact that it does not conserve charge according to the continuity equation. As a result, $\nabla \cdot E/4\pi$ will slowly deviate from the charge density $\rho$. This will turn up in the simulation as artifacts in the electric field as if there is spurious charge density dispersed in the plasma distribution which are not tracked by the code. Depending on the application, this might or might not be an issue, and some PIC codes decide to ignore this problem in favor of faster computation speed.

A typical way to help alleviate this problem is to employ divergence cleaning. There is no unique way of doing this, and many methods exist in literature. We only outline a simple method here. The goal is to solve the equation $\nabla \cdot E = 4\pi \rho$ every few timesteps. Since the deviation built up over a short time should be small as long as the time step is small, one can solve a diffusion equation of the electric potential $\phi$

$$\frac{\partial \phi}{\partial t} = -\nabla^2 \phi - 4\pi \rho$$

Since the diffusion equation has the property that an initial distribution with nonzero right hand side will relax towards an equilibrium where the right hand side becomes zero, we only need to apply some relaxation iterations to let it converge. A typical method is the Gauss-Seidel method (see e.g. Press et al., 2007) which applies a filter to the field consecutively. Since the initial deviation should be small, it only takes several relaxation iterations to achieve a reasonable result.

One difficulty of divergence cleaning lies in parallelization, since the relaxation filter is usually a global operation which involves communication between nodes every time (see section 2.3 for
parallelization). Another problem is that it inevitably introduces tiny fluctuations to the electric field, which will couple to particle dynamics and introduce heating to the plasma. A more efficient way is to directly solve the continuity equation \((2.24)\) for \(j\) at every timestep, which ensures that it is satisfied to numerical precision. Since charge density \(\rho\) never really shows up in the evolution equations in the first place, it only serves as a constraint for electric field, therefore we can actually avoid evaluating the charge density and deposit \(j\) directly. As long as the condition \(\nabla \cdot E = 4\pi \rho\) is satisfied for the initial condition, the continuity equation will guarantee that it is satisfied in subsequent times to numerical precision, thus side-stepping the problem of charge-conservation.

There are two main ways to solve the continuity equation numerically at each timestep. The classical way was proposed by Villasenor & Buneman (1992), which uses exact solutions for charge fluxes across cell boundaries for each kind of particle movement pattern. The Buneman deposit assumes first order shape functions (equation \((2.18)\)), and particles are effectively squares (cubes in 3D). When particles move, their shapes will overlap with some cell boundaries, creating charge fluxes across these boundaries and thus creating current. It is possible to solve exactly for the amount of current induced on each cell surface, effectively solving the continuity equation to machine precision.

The difficulty of this method lies in the many ways the cell boundary can be crossed. In 2D, there are 3 different cases where the moving particle can cross 4, 7, or 10 cell boundaries in a particular timestep (figure \(2.3\)). For the simplest case where only 4 boundaries are crossed, the current involved can be easily found to be:

\[
\begin{align*}
J_{x1} &= q \Delta x \left( \frac{1}{2} - y - \frac{1}{2} \Delta y \right), & J_{x2} &= q \Delta x \left( \frac{1}{2} + y + \frac{1}{2} \Delta y \right) \\
J_{y1} &= q \Delta y \left( \frac{1}{2} - x - \frac{1}{2} \Delta x \right), & J_{y2} &= q \Delta y \left( \frac{1}{2} + x + \frac{1}{2} \Delta x \right)
\end{align*}
\]

(2.26)

For the 7-boundary case, it can be decomposed into 2 4-boundary moves, and similarly the 10-boundary case can be decomposed into 3 4-boundary moves. Care must be taken when
decomposing the moves, since the second move can involve 4 different set of 4 boundaries, depending on the direction the particle moves. After decomposing the particle path into segments, each segment is treated as an ordinary move and equations (2.26) and (2.27) are used for each segment. The final current due to the complete particle path is the sum of the current from each segment.

The situation becomes worse in 3D. The simplest case involves current through 12 faces (4 faces on each coordinate plane), and the more complex cases are treated as follows: each time the particle crosses a cell face, the displacement is split into two segments, each consists of only movement within a single cell, which only needs to deal with 12 faces. This involves many branching and many if-else statements in the actual code. When implemented properly the difference between divergence of $E$ field and the charge density stays within truncation error of the floating point precision used in the calculation. Note that in this algorithm, it was implicitly assumed that the particle can not move more than one cell spacing. This condition is usually satisfied due to the
CHAPTER 2. PARTICLE-IN-CELL METHOD

Courant condition (section 2.2.4) being a more strict requirement.

![Figure 2.4: Zigzag scheme for current deposition. Mid points are chosen to be a cell vertex if the particle crosses two boundaries, or the midpoint of the cell face of perpendicular position on the crossed boundary if only there is only one cell crossing. Umeda et al. (2003)](image)

This method was later improved by Umeda et al. (2003) to cut down the number of different cases and increasing performance by splitting the particle path in the case of cell-crossing into a zigzag pattern. They noticed that the branching conditions for Buneman pusher depend mostly on whether the particle in question crosses a cell boundary. The idea is simply to find a universal midpoint for all kinds of trajectories to reduce the number of necessary segments to the minimum. The cases in 2D are shown in figure 2.4. This way, the particle path in one timestep is split to at most 2 segments, saving a lot of branching cases. Each segment is then deposited using the standard equations (2.26) and (2.27). The improvement is even more significant in 3D, where again only 4 cases are needed (figure 2.5), and in all cases only up to two segments in one timestep is needed to complete the deposition, greatly simplifying the amount of computations needed compared to the original Buneman scheme. Another strength of this method is that higher order shape functions (2.19) or (2.20) can be used since after segmenting the particle path, the only case one need to consider is one where particle does not cross cell boundary, and the current due to this movement can be solved exactly, whereas the original Buneman algorithm will need to consider
many more different cases due to one particle has influence on many more cell faces.

Figure 2.5: Zigzag scheme for 3D. Only up to 2 segments are needed for all cases of cell-crossing, reduced from the up to 8 cases of the original Buneman scheme, greatly reducing the number of operations. Umeda et al. (2003)

Another way to solve the continuity equation is proposed by Esirkepov (2001). It decomposes the motion of charged particles into motions along individual axes since they are independent. Then the change of charge density $\Delta \rho$ is split into components that correspond to components of $\nabla \cdot \mathbf{j}$. This is simplest to write in Cartesian coordinates:

$$\Delta \rho = \Delta \rho_x + \Delta \rho_y + \Delta \rho_z = -\Delta t (\partial_x j_x + \partial_y j_y + \partial_z j_z)$$  \hspace{1cm} (2.28)

Due to the fact that coordinate directions are independent, motion of particles in $x$ direction for example does not generate current in the $y$ and $z$ direction. This means that we can identify the $\partial_x j_x$ term with $\Delta \rho_x$, similarly for the $y$ and $z$ terms. For the simplest case where $\partial_x$ is simply the
2-term symmetric finite difference operator \[2.14\] then we can write

\[
j_x(x + \Delta x/2) = j_x(x - \Delta x/2) - \frac{\Delta x}{\Delta t} \Delta \rho_x(x) \tag{2.29}
\]

Given a boundary condition of \( j \) at one boundary (usually \( j = 0 \)), one can do a prefix sum over \( \Delta \rho \) to get the local \( j_x \) of each cell. This may sound like a very expensive operation, especially when parallelization is considered, since each node down the row will need to wait for the previous node to finish the prefix sum and take the final result to start the accumulation of this node. However practically since all particle shape functions have finite support, one simply needs to pass \( \Delta \rho \) in the guard cells to the neighboring nodes, then each node can start the prefix sum from \( j = 0 \) independently. See section \[2.3\] for discussion on domain decomposition and guard cells.

Now the task is simply to find \( \Delta \rho \) for each direction. Esirkepov (2001) found a unique solution for them in terms of particle shape functions \( S \):

\[
\frac{\Delta \rho_x(i, j, k)}{q} = \frac{1}{3} S(x + \Delta x, y + \Delta y, z + \Delta z) - \frac{1}{3} S(x, y + \Delta y, z + \Delta z) + \frac{1}{6} S(x + \Delta x, y + \Delta y, z) - \frac{1}{6} S(x, y + \Delta y, z) + \frac{1}{6} S(x + \Delta x, y, z - \frac{1}{3} S(x, y, z) \tag{2.30}
\]

where \( x, y, z \) are the original position of the particle, \( \Delta x, \Delta y, \Delta z \) are the particle displacements during one timestep, and \( S(x + \Delta x, y + \Delta y, z + \Delta z) \) is the shape function evaluated for the particle at the final position, with respect to the cell for which we are evaluating \( \rho_x \). For other directions simply permute the coordinate indices. Figure \[2.6\] shows a geometric interpretation of this solution as an average over 3 parallel paths.

For 2D, the current in the third direction does not enter the continuity equation since usually symmetry is implied in the third dimension, and the partial derivative vanishes in that direction. The Esirkepov solution becomes (in Cartesian coordinates where \( z \) axis is taken to be translational
CHAPTER 2. PARTICLE-IN-CELL METHOD

\[ j_x(i+1,j) - j_x(i,j) = -q \frac{\Delta x}{\Delta t} \left[ S(x+\Delta x, y+\Delta y) - S(x, y) \right] \]

(2.31)

\[ j_y(i,j+1) - j_y(i,j) = -q \frac{\Delta y}{\Delta t} \left[ S(x+\Delta x, y+\Delta y) - S(x+\Delta x, y) \right] \]

(2.32)

\[ j_z(i,j) = -qv_z \left[ \frac{1}{3} S(x+\Delta x, y+\Delta y) + \frac{1}{6} S(x+\Delta x, y) + \frac{1}{6} S(x, y+\Delta y) + \frac{1}{3} S(x, y) \right] \]

(2.33)

In other words, \( j_z \) is simply the velocity times the charge density averaged over the 3 intermediate positions similar to the 3 paths outlined in figure 2.6.

There are several advantages of the Esirkepov current deposition algorithm. First, no branching condition is necessary at all. All particles are assumed to undergo 3 segments of displacement in every timestep, and every segment is treated uniformly. This has huge implications on SIMD platforms like GPUs where branching incurs significant loss of parallelization, often doubling or tripling the effective computational load (more on this in appendix B). Secondly, it is trivial to
extend to higher order shape functions, since one can explicitly insert equations (2.19) or (2.20) into the above solutions and evaluate on the cells where \( S \) is nonzero. It is also possible to extend it to higher order spatial derivatives, whereas it is not apparently doable for the Buneman pusher. This will be discussed in appendix A.

Although the original paper by Esirkepov explicitly stated that the method is limited to Cartesian geometry, it is actually simple to extend it to other coordinate systems, approximating each cell as locally Cartesian. The only modification required is to use the correct divergence operator. In general curvilinear coordinates we have (see section 2.2.5 for more detailed discussion)

\[
\nabla \cdot \mathbf{j} = \frac{1}{h_1 h_2 h_3} \left[ \partial_1 (j_1 h_2 h_3) + \partial_2 (j_2 h_3 h_1) + \partial_3 (j_3 h_1 h_2) \right] \quad (2.34)
\]

therefore, one only needs to modify the equation (2.29) into:

\[
(j_1 h_2 h_3)(x_1 + \Delta x_1/2) = (j_1 h_2 h_3)(x - \Delta x_1/2) - \frac{\Delta x}{\Delta t} (h_1 h_2 h_3 \Delta \rho_1)(x) \quad (2.35)
\]

where \( h_1 h_2 h_3 \Delta \rho \) is found using the Esirkepov solution (2.30). After the current prefix sum, the resulting quantity will become \( j_1 h_2 h_3 \) instead of simply \( j_1 \), and one needs to divide by the appropriate scale functions to obtain the correct current.

### 2.2.3 Particle Pusher

From the discretization scheme we outlined in the previous section, the task of solving the Maxwell-Vlasov system reduces to solving the Maxwell equations coupled with the particle equations of motion (2.10) and (2.12), with the above-described current deposition scheme to translate from particle motion to the current on the grid. In this section and the next, we will outline the ways of solving these finite difference equations.

For particle equations of motion, we use the standard leap-frog scheme, meaning that \( \mathbf{x} \) and \( \mathbf{p} \) are always evaluated at a time difference of exactly a half timestep \( \Delta t / 2 \) (here \( i \) is the time step...
CHAPTER 2. PARTICLE-IN-CELL METHOD

label):

\[
x_{p}^{i+1} = x_p^i + u_p^{i+1/2} \Delta t
\]

\[
y^{i+1/2} u_p^{i+1/2} = y^{i-1/2} u_p^{i-1/2} + \frac{q\Delta t}{m} \left( E^i + \frac{u^i}{c} \times B^i \right)
\]

This scheme is stable as long as the time step satisfies the constraint \( \Delta t < \frac{2}{\omega_p} \) (Tajima, 1989).

Notice that in equation (2.37), an average velocity \( u^i \) appears on the right hand side, meaning that this is an implicit equation. The simplest way is to invert a \( 3 \times 3 \) matrix, but it is both slow and prone to numerical errors. Boris (1970) proposed an algorithm which uses geometric rotations to simplify the solution of this equation, which became widely used in PIC codes, and the algorithm was named the Boris pusher.

The Boris pusher assumes the following form of the average velocity on the right hand side of equation (2.37):

\[
u^i = \frac{p^{i-1/2} + p^{i+1/2}}{2\bar{y}^i m}, \quad \text{where} \quad \bar{y}^i = \sqrt{1 + \left( \frac{p^{i-1/2} + \frac{q\Delta t}{2} E^i}{mc} \right)^2}
\]

(2.38)

One can define intermediate momenta \( p^- \) and \( p^+ \)

\[
p^{i-1/2} = p^- - \frac{q\Delta t}{2} E^i, \quad p^{i+1/2} = p^+ + \frac{q\Delta t}{2} E^i
\]

(2.39)

then the equation for momentum update becomes

\[
\frac{p^+ - p^-}{\Delta t} = \frac{q}{2\bar{y}^i c} (p^+ + p^-) \times B^i
\]

(2.40)

From geometry, this means that the angle to rotate from \( p^+ - p^- \) to \( p^+ + p^- \) is related to \( B \). An
illustration is given in figure 2.7. The gist is that we have

\[ p' = p^- + p^- \times t, \quad t = \frac{q\Delta t}{2\gamma_i c} B^i \] (2.41)

and we can compute \( p^+ \) from \( p' \) by

\[ p^+ = p^- + p' \times s, \quad s = \frac{2t}{1 + t^2} \] (2.42)

This method greatly improves the numerical accuracy and speed of the particle momentum update in PIC simulations, and has become the \textit{de facto} particle pusher in many plasma simulation codes.

However one can see that an approximation was made in Boris pusher, namely the average velocity vector between timesteps \( i + 1/2 \) and \( i - 1/2 \) is given by equation (2.38). This is an approximation since Boris pusher was derived with non-relativistic particle dynamics in mind. Vay (2008) pointed out that in the special case where \( E + u \times B / c = 0 \) the Boris pusher may introduce a spurious force on the particles. A better average is simply given by

\[
    u^i = \frac{u^{i-1/2} + u^{i+1/2}}{2} = \frac{p^{i-1/2} / \gamma^{i-1/2} + p^{i+1/2} / \gamma^{i+1/2}}{2m}
\] (2.43)

Then the simple Boris scheme is no longer applicable. Vay outlined a new scheme which is slightly more complicated but more accurate when the Lorentz force of the particles is almost balanced by the electric force. The procedure is as follows. One first defines \( p' \) similar to that in Boris pusher

\[
p' = p^{i-1/2} + q\Delta t \left( E^i + \frac{u^{i-1/2}}{2c} \times B^i \right)
\] (2.44)
then one can find $p^{i+1/2}$ and $\gamma^{i+1/2}$ as follows:

\begin{align}
\gamma^{i+1/2} &= \sqrt{\frac{\sigma + \sqrt{\sigma^2 + 4(\tau^2 + p^2_\ast)}}{2}} \quad (2.45) \\

p^{i+1/2} &= s [p' + (p' \cdot t)t + p' \times t] \quad (2.46)
\end{align}

where $\tau = (q\Delta t/2c)B^i, p_\ast = p' \cdot \tau/c, \sigma = \gamma'^2 - \tau^2, \gamma' = \sqrt{1 + p' \gamma^2/c^2}, t = \tau/\gamma^{i+1/2}$, and $s = 1/(1 + t^2)$. 

Figure 2.7: Illustration of Boris velocity rotation, from (Boris, 1970)
Since in the simulations of our interest, much of the plasma is both relativistic and close to force-free, meaning that $E + u \times B / c$ is close to zero, it is important to use Vay pusher to avoid spurious forces. Therefore in Aperture we use the Vay pusher exclusively.

### 2.2.4 Integrating the Maxwell Equations

In this section we outline how we use the fields $E^i$ and $B^i$ to compute their values at the next time step. The most natural way is to use the traditional leapfrog method again

\[
E^{i+1} = E^i + \Delta t (c \nabla \times B^{i+1/2} - 4\pi j^{i+1/2}) 
\]

\[
B^{i+1/2} = B^{i-1/2} - \Delta t (c \nabla \times E^i) 
\]

This scheme is stable as long as the time step $\Delta t$ satisfies the Courant condition $\Delta t < \Delta x / c$ where $\Delta x$ is the smallest grid spacing. However, one difficulty is that in this scheme $E$ and $B$ fields are not defined at the same time step, whereas in our particle pusher (2.36) and (2.37) they are implicitly assumed to be sampled at the same time step.

An improved method we employ is the semi-implicit field solver used by Haugboelle et al. (2012). We start from the following finite difference equations:

\[
\frac{E^{i+1} - E^i}{\Delta t} = c \nabla \times (\alpha B^{i+1} + \beta B^i) - 4\pi j^{i+1/2} 
\]

\[
\frac{B^{i+1} - B^i}{\Delta t} = -c \nabla \times (\alpha E^{i+1} + \beta E^i) 
\]

where $\alpha$ and $\beta$ are numerical parameters that determines the “implicitness” of the scheme, with the constraint $\alpha + \beta = 1$. If $\alpha \gtrsim 0.5$ then the scheme is unconditionally stable. Bigger $\alpha$ leads to the damping of grid-scale waves, which helps smoothing the solution.

To solve these equations, we can plug the first equation into the second one to eliminate $E^{i+1}$,
explicitly assuming $\nabla \cdot \mathbf{B} = 0$:

\[
(1 - \alpha^2 \Delta t^2 \nabla^2) \mathbf{B}^{i+1} = \mathbf{B}^i + \alpha \beta \Delta t^2 \nabla^2 \mathbf{B}^i - \Delta t \nabla \times \mathbf{E}^i + \alpha \Delta t^2 \nabla \times \mathbf{j}^{i+1/2}
\]  

(2.51)

Now this is an equation of $\mathbf{B}^{i+1}$ only. We can solve it by inverting the operator $(1 - \alpha^2 \Delta t \nabla^2)$ and apply it to the right hand side

\[
(1 - \alpha^2 \Delta t^2 \nabla^2)^{-1} = 1 + \alpha^2 \Delta t^2 \nabla^2 + (\alpha^2 \Delta t^2 \nabla^2)^2 + \ldots
\]  

(2.52)

Provided $\Delta t$ is small enough, this series converges rapidly. Practically we found that it is sufficient to terminate the series at around 4 terms, giving a relative numerical error less than $10^{-8}$. Thus, although the Laplace operator is a global operation and a global exchange of guard cell values is required, the field solver is still significantly faster than the particle pusher, rendering the overhead negligible. Again, see section 2.3 for more detail on domain decomposition and the usage of guard cells.

One particular note is that if higher order numerical differential operators were used in current deposition, then the same order of differential operators should be used here in the field evolution, otherwise the exact charge conservation will not be imposed. This is because we need the curl of $\mathbf{B}$ to be divergence-free to numerical precision, which is only true when the divergence and curl are evaluated to the same order of accuracy with the same staggering structure (see section 2.2.1). This will be discussed in more detail in appendix A.

2.2.5 Coordinate Systems and Boundary Conditions

Aperture supports the use of orthogonal curvilinear coordinate systems, which basically means any parametrization of the Euclidean $\mathbb{R}^3$ that has a diagonal metric. The coordinate system can be specified by three scale functions $h_i = \sqrt{g_{ii}}$. Since the metric is diagonal, this contains all the
information of the metric.

The vector operations that we use in solving the Maxwell equations can be written in terms of these scale functions:

\[
\nabla f = \frac{1}{h_i} \partial_i f \quad (2.53)
\]

\[
\nabla \cdot \mathbf{V} = \frac{1}{\prod_k h_k} \partial_i (V_i \prod_{j \neq i} h_j) \quad (2.54)
\]

\[
\nabla \times \mathbf{V} = \frac{1}{\prod_i h_i} \mathbf{e}_i \epsilon_{ijk} h_i \partial_j (h_k V_k) \quad (2.55)
\]

where Einstein summation convention is assumed and repeated indices are summed over (except in the product sign). Due to the general form of these equations, it is straightforward to implement a general framework that works with vector derivatives, that can take in any scale functions \(h_i\).

<table>
<thead>
<tr>
<th>Coordinate System</th>
<th>(h_1)</th>
<th>(h_2)</th>
<th>(h_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartesian ((x, y, z))</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cylindrical ((\rho, \theta, z))</td>
<td>1</td>
<td>(\rho)</td>
<td>1</td>
</tr>
<tr>
<td>Spherical ((r, \theta, \phi))</td>
<td>1</td>
<td>(r)</td>
<td>(r \sin \theta)</td>
</tr>
<tr>
<td>Log spherical ((x, \theta, \phi))</td>
<td>(e^x)</td>
<td>(e^x)</td>
<td>(e^x \sin \theta)</td>
</tr>
</tbody>
</table>

Table 2.1: Scale functions for coordinate systems implemented in Aperture

Although the code was designed with flexibility, the main coordinate systems we support in the actual code are Cartesian, cylindrical, spherical coordinates and some of their variants. The scale functions of these coordinate systems are listed in table 2.1. One particular challenge for each curvilinear coordinate system is the treatment of particle movement, since technically one needs to solve the geodesic equation, since the orthonormal coordinate basis is position dependent. However our curvilinear coordinates are always a parametrization of \(\mathbb{R}^3\), so instead of trying to solve for the geodesic equation, we always convert the position and velocity of the particle to the corresponding Cartesian values before moving it, where we can simply use the equation (2.10) to update particle position.

An additional challenge in implementation of the coordinate systems has to do with how
boundary singularities are treated, which varies for different coordinate systems. We will outline the treatment for the axis boundary in the spherical coordinates here since it is the one implemented first, and most relevant for the physics applications detailed in the later part of the thesis.

**Coordinate boundary condition**

For 2D spherical coordinates, axisymmetry means $\partial_\phi = 0$ in all Maxwell equations. By symmetry, we automatically have on the axis

$$E_\theta = E_\phi = B_\theta = B_\phi = 0$$  \hspace{1cm} (2.56)

simply because there is no preferred direction of these field components. Furthermore, since we have $B_\phi = E_\phi = 0$ at the boundary for all times, their time derivatives should vanish, which makes the following conditions on the axis:

$$\partial_r (r B_\theta) - \partial_\theta B_r = 0, \quad \partial_r (r E_\theta) - \partial_\theta E_r = 0$$  \hspace{1cm} (2.57)
Since $B_\theta = E_\theta = 0$ on the axis, so are their $r$ derivatives, therefore we have instead

$$\partial_\theta B_r = \partial_\theta E_r = 0$$  \hspace{1cm} (2.58)

which is akin to a “symmetric” boundary condition for both $E_r$ and $B_r$.

For the actual boundary conditions we have to impose on the fields, we need to refer to the staggered grid configuration (figure 2.1). Our simulation domain boundary coincides with the spherical axis $\theta = 0, \pi$; then 4 components will be defined exactly on the boundary: $B_\phi, E_\theta, E_r,$ and $j_\theta$. From the discussion above, we immediately have $B_\phi = E_\theta = 0$. We impose this condition simply as a Dirichlet boundary condition every timestep, by setting these field components to zero for all the cells on the axis.

Similarly, $j_\theta$ should also be zero. This is handled by slightly modifying the current deposition step: every particle that moves in the cells adjacent to the axis boundary is assigned a ghost particle that is its mirror image across the axis. The resulting $\Delta \rho$ is taken to be the average from the movement of both particles. This symmetric construction automatically guarantees that $j_\theta = 0$ on the boundary. This also serves as a suitable starting point for the current prefix sum for $j_\theta$ since we know that it always starts from 0 at $\theta = 0$.

Finally for $E_r$, its update equation is:

$$\partial_t E_r = \frac{1}{r \sin \theta} \partial_\theta \sin \theta B_\phi$$  \hspace{1cm} (2.59)

The difficulty here is that $\sin \theta \to 0$ as $\theta \to 0$ or $\theta \to \pi$ at the axis. Our way to solve this issue is to take the limit:

$$\lim_{\theta \to 0} \partial_t E_r = \lim_{\theta \to 0} \frac{\sin \theta \partial_\theta B_\phi + \cos \theta B_\phi}{r \sin \theta} = \frac{1}{r} \partial_\theta B_\phi + \lim_{\theta \to 0} \frac{\partial_\theta (\cos \theta B_\phi)}{\partial_\theta (r \sin \theta)} = \frac{2}{r} \partial_\theta B_\phi$$  \hspace{1cm} (2.60)

The only remaining complication here is that to evaluate $\partial_\theta B_\phi$ we can’t use the information across
the axis and need to rely on the cells within the physical domain. Therefore one cannot use the
symmetric finite difference operator (equation [2.14]). One can however use the information that
\( B_\phi \) is zero on the boundary. To get the same 2nd order accuracy in \( \Delta \theta \), we can use:

\[
\partial_\theta B_\phi = 2B_\phi(\Delta \theta) - \frac{1}{2}B_\phi(2\Delta \theta) + O(\Delta \theta^3)
\]

(2.61)

We will discuss this in more detail and the extension to higher order in appendix A.

**Stellar boundary condition**

In magnetospheric simulations, the inner boundary of the box is always the neutron star itself.
The stellar boundary is treated as a perfect conductor: \( E = -v_{\text{rot}} \times B/c \) and \( B = B_{\text{dipole}} \) below
the surface. \( v_{\text{rot}} = r \times \Omega \) is the rotation velocity of the star; it determines \( E \) at the surface and
this is how the simulation knows about the rotation of the star. Typically at the beginning of the
simulation \( v_{\text{rot}} \) is taken to be identically zero, and the star is spun up smoothly over a relatively
short time frame. For magnetar simulations, it is the differential rotation of the star that we impose,
which is done through an latitude dependent \( v_{\text{rot}}(\theta) \), and the rotation of the star is ignored (see
chapter 5).

Particles are injected into the simulation domain below the surface in pairs of electrons and
ions. In order to not introduce artificial charges into the simulation, they are injected at exactly
the same location with only a small random thermal velocity. Particles entering the star are erased
once they penetrate deep enough such that all cells they contribute the current to are below
the surface. This is again to avoid violating the Gauss’s law inside the simulation domain, since
erasing charges in general introduce errors to the Gauss’s law.

For magnetospheric simulations we also implemented gravity for all particles to control the
injection layer. Injected particles have a thermal momentum distribution with mean momentum
\( \bar{p}_0 \), and gravity is implemented with a force profile of \( g = -\hat{r}mg_0/r^2 \). The parameter \( g_0 \) controls the
strength of gravity, as well as the thickness of the atmospheric layer, \( h \sim v_0^2/2g_0 \). In the beginning
of the simulation, the star starts at rest to allow the formation of a dense atmosphere. The use of this atmospheric layer ensures that there is always an abundant supply of particles from the surface, and decouples it from the detail of actual particle injection at each timestep.

**Free escape boundary condition**

The outer boundary uses a free escape condition. To achieve this we place a damping layer at the outer boundary as described by Umeda et al. (2001). The damping layer is placed near \( r_{\text{max}} \) in which the field values at every time step is multiplied by a masking function that depends on the radial position \( r \):

\[
E^{n+1} = f_M(r) \left[ E^i + c\Delta t \nabla \times (\alpha B^{i+1} + \beta B^i) - 4\pi \Delta t j^{i+1/2} \right] \tag{2.62}
\]

\[
B^{n+1} = f_M(r) \left[ B^i - c\Delta t \nabla \times (\alpha E^{i+1} + \beta E^i) \right] \tag{2.63}
\]

and we adopt the masking function used in Umeda et al. (2001):

\[
f_M(r) = \begin{cases} 
1, & \text{for } r < r_D \\
1 - \eta \left( \frac{r - r_D}{r_{\text{max}} - r_D} \right)^2, & \text{for } r \geq r_D 
\end{cases} \tag{2.64}
\]

where \( r_D \) is where the damping layer begins, and \( \eta \) is a numerical coefficient that controls the effectiveness of the damping. In the simulations, the damping layer is usually chosen to be 10 to 15 cells, and \( \eta \) is chosen to be small such that the characteristic skin depth of waves into the medium is similar to the damping layer thickness. In other words, the coefficient is chosen that the waves are damped slowly enough to not generate much reflection. In practice, the wave damping is very efficient, and wave reflection is very weak for all production runs.

It is very common for PIC codes to implement a perfectly-matched layer (PML) as an absorbing boundary condition (Berenger 1994). We have also implemented it and compared it with the effectiveness of the simple damping layer. We found that the result is almost identical, but since
the damping layer is more efficient in terms of computation and memory, we decided to keep this implementation.

Particles streaming into the damping layer decouples from the field and are allowed to freely escape from the box. When they enter the guard cells located at the outer boundary, they are erased.

2.2.6 Radiative Transfer

A significant feature of Aperture compared to traditional PIC codes is the inclusion of various radiative transfer mechanisms that are important in high-energy astrophysics. In this section we list the supported physical processes and outline how we model them in the code.

Production of high energy photons

In the magnetosphere of neutron stars, energetic electrons or positrons can induce vacuum breakdown and create $e^\pm$ pairs. This is usually a 2-step process. First, a high energy photon ($E_{\text{ph}} > 2m_e c^2$) is emitted from the energetic particle, and the photon will then convert into a pair of $e^\pm$. In pulsar magnetospheres, the first process is typically from curvature radiation from electrons accelerated to Lorentz factors $\gamma \gtrsim 10^6$ (see discussion in section [1.1.3]). Another channel for producing highly energetic photons capable of converting into $e^\pm$ pairs is from resonant upscattering of background X-ray photons. The electron absorbs photons that have the energy to excite it from the lowest Landau level to the first excited level, and subsequently drop down to the lowest level, re-emitting the photon. The photon energy in the lab frame is boosted by a factor of $\gamma^2$ during this process, similar to ordinary inverse Compton scattering.

Aperture models both channels of high-energy photon emission with a threshold condition and an emission rate. For curvature radiation, we compute the local radius of curvature of the particle trajectory using

$$R_c = \frac{|p|^3 \Delta t}{\gamma |p \times \Delta p|}$$ (2.65)
and then we set the threshold for emitting a curvature photon that is capable of converting to a pair to be

$$\gamma_{\text{th}} = K \left( R_c / R_* \right)^{1/3} \quad (2.66)$$

where $K$ is a numerical parameter that we can tune to fit the parameter space for the simulation. In real pulsars this coefficient can be $K \gtrsim 10^6$. The photon energy emitted is artificially set to be a certain multiple of $m_e c^2$ which is another numerical parameter that affects the final pair multiplicity.

For resonant scattering we can set the threshold condition to be exactly the resonance condition for the local $B$ field:

$$\gamma_{\text{th}} = \gamma_0 \frac{B}{B_Q} \quad (2.67)$$

where $B_Q$ is the critical field for QED effects, $B_Q = m^2 c^3 / e \hbar = 4.4 \times 10^{13} \, G$. It was found that in the realistic magnetar outflow the threshold scales as $\gamma_{\text{th}} \sim 100 B / B_Q$ (Beloborodov 2013), but this coefficient can be tuned for the particular range of parameters used in a simulation.

The energy of the emitted photon is set by the local magnetic field as well. The energy of the emitted photon for a given resonant scattering event is given by (Beloborodov & Thompson 2007)

$$E_{\text{ph}}(\theta_{\text{em}}) = \frac{E_B}{\sin^2 \theta_{\text{em}}} \left[ 1 - \left( \cos^2 \theta_{\text{em}} + \frac{m_e^2 c^4}{E_B^2} \sin^2 \theta_{\text{em}} \right)^{1/2} \right] \quad (2.68)$$

where $\theta_{\text{em}}$ is the emission angle of the de-excitation photon with respect to the $B$ field in the frame where the parallel momentum of the electron/positron vanishes. $E_B$ is the energy of the first Landau level:

$$E_B = \left( \frac{2B}{B_Q} + 1 \right)^{1/2} m_e c^2 \quad (2.69)$$

In the actual code, we use the angle-averaged version of equation (2.68) (Beloborodov 2013)

$$E_{\text{ph}} = \gamma m_e c^2 \left( 1 - \frac{1}{1 + 2B / B_Q} \right) \quad (2.70)$$
and in the very narrow parameter range when this energy is greater than \((\gamma - 1)m_{e}c^{2}\) we cap the emitted photon energy at \((\gamma - 1)m_{e}c^{2}\).

**Conversion of high energy photon into pairs**

It is impossible for a photon to convert into \(e^{\pm}\) spontaneously due to inability to conserve both energy and momentum. There are two main channels for energetic photons to convert into an \(e^{\pm}\) pair: through interaction with the local \(B\) field, or through collision of two energetic photons.

Magnetic conversion has strict requirement on the strength of \(B\) field. The rate decreases exponentially with respect to \(B/B_{Q}\) [Erber, 1966]. In addition, the photon propagating at an angle \(\theta\) with respect to the \(B\) field has to satisfy the threshold condition

\[
E_{ph} > \frac{m_{e}c^{2}}{\sin \theta}
\]  

(2.71)

We model the exponential cross section for the magnetic conversion as a sharp cut-off \(B_{th}\) below which magnetic conversion simply turns off. When local \(B\) field is larger than \(B_{th}\) then the energy of the photon is compared to the threshold energy (2.71) and pair creation is triggered when the condition is satisfied.

In the outer regions of the magnetosphere \(B\) quickly falls far below \(B_{Q}\) and magnetic conversion is prohibited. The only channel becomes \(\gamma-\gamma\) collision. It is extremely difficult to compute the background target field for the high energy photons to collide with, since it is produced by both thermal emission from the star and synchrotron radiation from the particles in the magnetosphere. Instead of calculating the target field, we assume a random photon free-path:

\[
\ell_{ph} \sim N(aR_{*}, (bR_{*})^{2})
\]  

(2.72)

where \(N(\mu, \sigma^{2})\) denotes a normal distribution with mean \(\mu\) and variance \(\sigma^{2}\), and \(a\) and \(b\) are parameters we choose (see chapter [4]). The normal distribution is truncated at \(\ell_{ph} = 0\) and any
negative value is discarded. Every photon is produced with a free-path drawn from distribution \ref{eq:free_path}, and if it is not converted magnetically or escapes the domain, then it is converted to a pair of \( e^\pm \) at the end of its free path. The energy of the photon is evenly shared by the resulting electron and positron, while the momenta of these particles are along the same direction as the original photon.

Although unphysical, this approximation captures the qualitative behavior of a pair cascade in the pulsar magnetosphere, producing the required amount of plasma in the correct regions. A more sophisticated modeling of the pair creation process will definitely improve the calculation of local pair density in the magnetosphere, but the overall structure of the system will remain unchanged, since it is close to the force-free limit already.

**Synchrotron and general radiative loss**

When energetic particles gyrate in a magnetic field they emit synchrotron radiation. Even when the synchrotron (or curvature) photons are not energetic enough to produce pairs, their emission can result in significant energy loss of particles, which may affect the global dynamics of plasma and fields. In Aperture we implemented a general radiative damping force to account for the synchrotron and curvature radiation loss that is not explicitly tracked by photon production. The radiation power by a moving particle in an electromagnetic field is (see e.g. \cite{Longair11}):

\[
P_{\text{rad}} = \frac{2e^2}{3c^3} \gamma_e^4 \left( a_\perp^2 + \gamma_e^2 a_\parallel^2 \right)
\]

(2.73)

where \( \perp \) and \( \parallel \) are with respect to the direction of motion of the particle. The radiative damping force acting on relativistic particles is given by:

\[
F_{\text{rad}} = -\frac{P_{\text{rad}}}{c^2} \mathbf{p}
\]

(2.74)
Radiative drag due to resonant scattering

In the magnetospheres of magnetars, resonant scattering plays a crucial role in the particle dynamics. The relativistic outflow of $e^\pm$ pairs is bathed in a background of X-ray photons that come from the thermal emission from the star; as the particles flow along the field lines, they sample different part of the photon spectrum.

We simplify the problem by ignoring the contribution to the X-ray spectrum from the flow itself, and only consider thermal photons from the star. This way one can compute analytically the effect of resonant scattering on the particle flow. The central blackbody radiation on average applies a force on the particles (Beloborodov 2012):

$$ F(\beta) = \frac{\alpha^2 R^2_r m_e c^2}{4r^2} \Theta^3 g(y)(\beta_\ast - \beta) $$

where $\beta_\ast = \mu = B_r/B$, $\Theta = kT/m_e c^2$ is the surface temperature in units of electron rest mass, and

$$ g(y) = \frac{y^3}{e^y - 1}, \quad y = \frac{\hbar \omega_{\text{res}}}{kT} = \frac{B/B_Q}{\gamma(1 - \beta \mu) \Theta} $$

The effect of the force (2.75) is pushing the flow velocity towards the “radiatively locked flow” velocity $\beta_\ast$. Numerically we implement the force with two parameters:

$$ \frac{dp}{dt} = F(p) = D \frac{g(y)}{(r/R_\ast)^2}(\gamma \mu - p) $$

and the direction of this force is along the $B$ field. The two numerical parameters are $D$ and $\Theta$. The former determines the overall strength of the drag, and the second determines which part of the thermal spectrum interacts with the particles. Depending on these two parameters, a certain part of the field line bundle will have particles stopping at the equator since $\beta_\ast = \mu = 0$ at the equator. In the inner magnetosphere where $B$ field is too strong ($B \gtrsim B_Q$) only very high energy particles resonantly scatter thermal photons, and the drag is not enough to stop particles near the
equator. The scenario described in chapter 5 applies to this region. For very extended field lines the drag is too weak since it scales with $r^{-2}$, and particles are not stopped at the equator either.

2.2.7 Units

In a computer simulation it is convenient to deal with dimensionless quantities, expressed in some units of choice. It is to our benefit to choose a unit system that is most natural to the problem and in which the dynamic equations take the simplest dimensionless form. Our choice of units is described in this section. This process also highlights the relevant scales of the problem.

The primary goal of the code is to simulate an isolated neutron star, therefore the natural length scale is the stellar radius $R_*$, and there is no other length scale a priori. Naturally, the time scale of the problem would be the light-crossing time of the star. Therefore we define the dimensionless length and time using the following equations (in all the following equations, a tilde on the symbol means the dimensionless version):

$$ r = \tilde{r} R_*, \quad t = \tilde{t} \frac{R_*}{c}, \quad \omega = \tilde{\omega} \frac{c}{R_*} $$

The radius $\tilde{r} = 1$ shows up mainly as the lower boundary of most of our simulations.

Once we add rotation of the star and a magnetic field, two important frequencies show up in this problem that can then be written in dimensionless units:

$$ \tilde{\Omega} = \frac{\Omega R_*}{c}, \quad \tilde{\omega}_B = \frac{\omega_B R_*}{c} = \frac{eB_0 R_*}{mc^2} $$

(2.79)

where $m$ is the electron rest mass, $\Omega$ is the rotation angular frequency, equal to $2\pi/P$ where $P$ is the period of rotation, and $B_0$ is the surface magnetic field at the pole. Given a pulsar with certain radius, these two are the dimensionless parameters that will govern the simulation, and hence the physical behavior of the system.

Next, we need to define a unit for charge density. This can be done by relating the plasma
frequency, a third frequency in the problem, with the local corotation charge density \( \rho_{\text{GJ}} = -\frac{\mathbf{\Omega} \cdot \mathbf{B}}{2\pi c} \). We define the reference plasma frequency at every point as

\[
\omega_{p}^{2} = \frac{4\pi \rho_{\text{GJ}} e}{m}
\]  

(2.80)

then we can define a version of dimensionless Goldreich-Julian density, or any charge density

\[
\rho_{\text{GJ}} = \frac{m\tilde{\omega}_{p}^{2}e^{2}}{4\pi e R_{e}^{2}} = \tilde{\rho}_{\text{GJ}} \frac{mc^{2}}{4\pi e R_{e}^{2}}
\]

(2.81)

Note that the real plasma frequency will depend on the actual plasma number density at the point of interest and will in general be different from what we call \( \omega_{p} \) above. However, it is still useful to determine such a characteristic plasma frequency for reference. What the above equation says is that, by choosing these units we make \( \tilde{\rho} = \tilde{\omega}_{p}^{2} \). Especially, there is a simple relationship between the three characteristic frequencies in the problem, if we define \( \rho_{\text{GJ}} = \frac{\Omega B_{0}}{2\pi c} \):

\[
\frac{m\tilde{\omega}_{p}^{2}}{4\pi e} = \frac{\Omega B_{0}}{2\pi c} \implies \tilde{\omega}_{p}^{2} = 2\tilde{\Omega}\tilde{\omega}_{B}
\]

(2.82)

Therefore, specifying the numerical values of \( \tilde{\Omega} \) and \( \tilde{\omega}_{B} \) automatically determines the characteristic plasma frequency in the problem.

With the above choices, we can work out the dimensionless electric and magnetic fields. Gauss’s law gives

\[
\nabla \cdot \mathbf{E} = 4\pi \rho \implies \frac{\Delta \mathbf{E}}{\Delta \mathbf{R}_{e}} = \tilde{\rho} \frac{mc^{2}}{e R_{e}^{2}} \implies \mathbf{E} = \tilde{E} \frac{mc^{2}}{e R_{e}}
\]

(2.83)

and similarly

\[
\mathbf{B} = \tilde{B} \frac{mc^{2}}{e R_{e}}
\]

(2.84)

Note that this means that \( B \) and \( \omega_{B} \) have the same dimensionless form, therefore \( \tilde{B}_{0} = \tilde{\omega}_{B} \).
In these units, Maxwell equations look as follows:

\[
\frac{\partial \tilde{E}}{\partial t} = \tilde{\nabla} \times \tilde{B} - \tilde{j}, \quad \frac{\partial \tilde{B}}{\partial t} = -\tilde{\nabla} \times \tilde{E}
\]  

(2.85)

and the particle equations of motion become

\[
\frac{d \tilde{p}}{dt} = \tilde{E}, \quad \frac{d \tilde{r}}{dt} = \tilde{u} = \frac{\tilde{p}}{\sqrt{1 + \tilde{p}^2}}
\]

(2.86)

Note that all numerical coefficients disappear.

Up to now, the only numerical parameters we have introduced are \(\tilde{\Omega}\) and \(\tilde{\omega}_B\), and they completely specify the problem. For simulation purposes, one needs to rescale both of these parameters to bring the size of the system into computable regime. However, one more numerical parameter is required for interpolating between the particle charge and the charge density on the grid, namely the charge of an individual macro particle. Since in PIC there is no way to simulate realistic plasma of order \(10^{20}\) particles, one needs to use one single macro particle to represent a collection of physical particles. A macro particle has the same charge-to-mass ratio as a physical particle. The way we choose the charge per particle is to normalize it so that our characteristic charge density \(\rho_{GJ}\) would correspond to a certain manageable number of particles \(N_p\) per grid cell.

The way we compute charge density is to iterate over all particles and deposit onto a single cell:

\[
\rho_c = \frac{\sum eS(r_p, r_c)}{\Delta V}
\]

(2.87)

where \(e\) is the charge per macro particle, \(S\) is a shape factor depending on particle position \(r_p\) and grid point position \(r_c\), and \(\Delta V\) is the cell volume. In 2.5D simulations, \(\Delta V\) involves the arbitrary size of the third dimension which is the symmetric direction, so we usually take the size in that direction to be unity. Assuming every particle has the same charge \(e\), and there are \(N_p\) particles in
the cell, up to a form factor, we have
\[ \rho = \frac{eN_p}{\Delta V} = \frac{eN_p}{\Delta V R_s^3} = \tilde{\rho} \frac{mc^2}{4\pi eR_s^2}, \quad \tilde{\rho} = \frac{\tilde{e}N_p}{\Delta V} \tag{2.88} \]

so we can identify the dimensionless charge per particle
\[ \tilde{e} = e \frac{4\pi e}{mc^2 R_s} \tag{2.89} \]

Note that \( e \) appears twice on the right hand side. This is because we treat \( e/m \) as a physical constant, therefore \( 4\pi e/mc^2 R_s \) is a physical constant which is fixed as soon as we choose the physical size of the star. We would like to choose \( \tilde{e} \) so that \( N_p \) particles per cell corresponds to the characteristic charge density \( \rho_{GJ} \):
\[ \tilde{\rho} \sim \frac{\tilde{e}N_p}{\Delta V} \sim \tilde{\rho}_{GJ} \quad \Rightarrow \quad \tilde{e} = \frac{2\tilde{\Omega}\tilde{B}\Delta \tilde{V}}{N_p} \tag{2.90} \]

One technical aspect about the numerical value \( \tilde{e} \) in the code is that, every time we use \( \tilde{e} \), we actually use the combination \( \tilde{e}/\Delta \tilde{V} \), which shows up in charge/current deposition. Therefore it is not necessary to carry \( \Delta \tilde{V} \) around. We simply remove it from the definition of \( \tilde{e} \), and do not divide by \( \Delta \tilde{V} \). Also a factor of 2 is not that important, so in the code the actual value of \( \tilde{e} \) is taken to be
\[ \tilde{\tilde{e}} = \frac{\tilde{\Omega}\tilde{B}_0}{N_p} \tag{2.91} \]

To summarize, the numerical parameters in our simulation are \( \tilde{\Omega}, \tilde{\omega}_B \) (or equivalently \( \tilde{B}_0 \)), and \( \tilde{\tilde{e}} \) (or equivalently \( N_p \) after specifying cell size). This completely specifies the computational problem. In order to translate the numerical simulation to a real magnetosphere, we need to specify \( R_s \) in physical units (e.g. \( 10^6 \) cm), and multiply the dimensionless units by their physical units to get the realistic values. The dimensionful units are summarized below (subscript * means
that these are the dimensionful units for the physical quantity in question, e.g. $E = \bar{E} E_*$)

\begin{align*}
r_* &= R_\ast = R_6 10^6 \text{ cm} \quad (2.92) \\
t_* &= R_\ast / c = R_6 3.33 \times 10^{-5} \text{ s} \quad (2.93) \\
E_* &= B_* = \frac{mc^2}{eR_*} = R_6^{-1} 2.08 \times 10^{-5} \text{ statV/cm(G)} \quad (2.94) \\
\rho_* &= \frac{mc^2}{4\pi eR_*^2} = R_6^{-2} 1.66 \times 10^{-12} \text{ statC/cm}^3 \quad (2.95) \\
j_* &= \frac{mc^3}{4\pi eR_*^2} = R_6^{-2} 4.96 \times 10^{-2} \text{ statC/cm}^2 \cdot \text{s} \quad (2.96) \\
e_* &= \frac{mc^2 R_*}{4\pi e} = R_6 1.66 \times 10^6 \text{ statC} \quad (2.97) \\
g_* &= R_\ast c^2 = R_6 9.0 \times 10^{26} \text{ m}^3\text{s}^{-2} \quad (2.98) \\
p_* &= mc \quad (2.99)
\end{align*}

\section{2.3 Aperture}

Aperture is a PIC code that I designed and developed initially on my own, and later with the help of Rui Hu. The name is a recursive acronym which stands for “Aperture is a code for Particles, Electrodynamics and Radiative Transfer at Ultra-Relativistic Energies”. There are two versions of Aperture, one written for Nvidia GPUs (Graphics Processing Units), another written for ordinary CPUs and targets large clusters. The basic algorithms and designs of the two versions are very similar, therefore they can be considered to be the same code. The GPU version was developed using the CUDA programming language provided by Nvidia, and was the initial version of Aperture. The CUDA programming language has a few paradigms drastically different from ordinary programming languages, the most notable difference being that the computational power of GPUs depends critically on parallelization. A single instruction is applied to a set of 32 independent data elements by default, and to saturate the pipeline at least $\sim 1000$ data elements need to be processed in parallel. This calls for very different programming technique compared
to traditional procedural programming on CPU where most data is processed one element at a
time. We will discuss in detail how GPU architecture is different from traditional CPUs and the
challenges in designing software to take advantage of this architecture in appendix B.

![Flowchart](image)

Figure 2.9: Anatomy of one single timestep in Aperture code. Steps where communication between
nodes is required are colored red.

The structure of the Aperture code is modular by design. Each algorithm component is
encapsulated in a C++ class, and can be used mostly independently. The flow of the main loop
in the Aperture code is shown in figure 2.9: the code first read all the physical and numerical
parameters from a configuration file, then prepares the initial condition for the simulation. From
here it enters a loop over a specified number of timesteps; for each timestep it executes each
module in order, and writes out field and current distribution data at a set interval.

The GPU version of the code is designed to be run on a workstation with 1-4 GPUs and does not contain any multi-node parallelization. All particle and field data is stored and evolved on the GPU, and the total number of particles in the simulation is limited by the amount of graphical memory on the GPU board, which is typically 4-16GB depending on the model. The CPU version however, is completely parallelized using the MPI (Message Passing Interface) protocol, and can support arbitrary number of parallel CPU cores. The later version allows us to run much larger scale simulations, pushing the limit of parameters that we can access, closing the gaps between rescaled simulations and realistic physical parameters.

Parallelization of the CPU version over MPI is done using traditional domain decomposition. The overall domain is split into a number of patches matching the number of parallel computation “nodes”, each corresponds to a CPU core. Each core takes one patch of the domain and stores the field values and particles in that part of the domain. Therefore, when particles leave and enter the domain patches, communication is needed across the computational nodes.

Inter-node communication is buffered using guard cells. Guard cells are redundant cells that pad the boundary of a patch, which hold identical information as the corresponding cells in the neighboring node (figure 2.10). They are needed because in many situations such as evaluating finite difference operators (section 2.2.4) at the boundary of a patch, the field values in the cells of the neighboring patch are required. It is easiest to store them first in the guard cells, and update them every timestep after the field update.

There are two main types of guard cell communication for field values. For field solver (section 2.2.4), the guard cells need to be filled with the value of the corresponding cells in the neighboring node; for the current deposition (section 2.2.2), the charge \( \Delta \rho \) deposited into the guard cells need to be transferred to the neighboring node and added to the corresponding cells. For particles, all particles that move into guard cells at the end of each time step need to be moved to the corresponding cells in the neighbor, so that at the beginning of each time step, all guard cells are
empty. This also ensures that no current will ever flow through the outer boundary of the guard cells, and the current prefix sum described in section 2.2.2 can always begin with \( j = 0 \).

2. Add to deposit result

1. Copy field values

Figure 2.10: Guard cells between nodes. Two modes of inter-node communication. For field solver, the guard cells copies the values of the corresponding cells in the neighboring node (mode 1). For current deposition, the deposit results in the guard cells are added to the corresponding cells in the neighboring node (mode 2).

The CPU version of Aperture has been deployed on different platforms and shows excellent scaling properties. It has been run on a local workstation with 12 cores, the Habanero cluster at Columbia on \( \sim 250 \) cores, and the NASA Pleiades cluster on more than 10,000 cores. The code can reliably achieve a stable performance of 1.25 million particles per core per second per timestep.

### 2.4 Test Problems

In this section we present some well-known test cases to demonstrate the correctness of the code. Due to the coordinate flexibility of Aperture, we would like to test its correctness in multiple different coordinate systems. The following three test cases are in Cartesian, cylindrical, and spherical coordinates respectively, in the hope to represent the three main coordinate systems that we support in the code.
2.4.1 Relativistic Two-stream instability

The first test is relativistic two-stream instability in 2D. We consider a plasma beam with Lorentz factor of $\gamma_b$ and density $n_b$ passing through a background with density $n_p$. The maximum growth rate of the unstable mode is given by (see e.g. Bret et al., 2004):

$$\delta_{\text{TSI}} = \frac{\sqrt{3}}{2^{4/3}} \alpha^{1/3} / \gamma_b$$  \hspace{1cm} (2.100)

where $\alpha$ is the ratio of the number densities of the beam and background $\alpha = n_b/n_p$.

![Figure 2.11: Growth of two stream instability vs. theoretical value. Theoretical growth rate is $\delta_{\text{TSI}} \approx 0.1375$ from equation (2.100). The growth saturates at around $t\omega_p \sim 130$.](image)

We set up a neutral electron positron beam passing through a neutral electron positron plasma background, with density ratio $\alpha = 1$ and initial beam Lorentz factor $\gamma_b = 5.0$. We use a $512 \times 256$ box with 512 cells in the beam direction ($z$ direction in the $x$-$z$ plane), with roughly 5 cells per plasma scale. We assume periodic boundary conditions on all boundaries. Figure 2.11 shows the agreement of numerical growth of the electric field energy with the prediction, up until saturation.
2.4.2 Cylindrical Waveguide

Consider a cylindrical waveguide with radius $r_c$ and periodic boundary conditions at both ends, the system can be approximated as a long cylinder in the limit of $B_0 \rightarrow \infty$, except with discretized $k_z$. The side boundary of the cylinder is grounded, so $\Phi_e = 0$ at the boundary, which also implies that $E_{||} = 0$ on the boundary. With these assumptions, the set of Maxwell equations can be solved analytically (see e.g. Swanson, 2003). This calculation was also done in the context of pulsar flux tubes by Arons & Barnard (1986), where they also discussed the $B \rightarrow \infty$ limit. We repeat the calculation below and compare it with the numerical result obtained in the simulation.

The radial solutions are lowest Bessel functions $J_0$ and $J_1$, while the longitudinal solutions are simply plane waves:

\begin{align}
E_z &= \sum_{n=0}^{\infty} C_n J_0(k_{\perp} r) e^{i k_z z - i \omega t} \quad (2.101a) \\
E_r &= \frac{i k_z}{\epsilon - k_z^2} \partial_r E_z = \sum_{n=0}^{\infty} \frac{i k_z k_{\perp} e^2}{\omega^2 - c^2 k_z^2} C_n J_1(k_{\perp} r) e^{i k_z z - i \omega t} \quad (2.101b) \\
B_\phi &= \frac{i c}{\omega} (\partial_z E_z - \partial_r E_r) = \sum_{n=0}^{\infty} \frac{i \omega c k_{\perp}}{\omega^2 - c^2 k_z^2} C_n J_1(k_{\perp} r) e^{i k_z z - i \omega t} \quad (2.101c) \\
j_z &= \frac{i \omega^2_p}{4 \pi \omega} E_z \quad (2.101d)
\end{align}

where $k_{\perp}$ is determined by the radial boundary condition that $J_0(k_{\perp} r_c) = 0$, and $n$ sums over the zeros of the Bessel function. The fact that $E_z$ and $j_z$ are proportional to $J_0$ while $B_\phi$ and $E_r$ are proportional to $J_1$ is determined by our boundary condition that parallel electric field is zero at the side of the cylinder. We have well-defined waves propagating along the $z$ direction and the dispersion relation between $\omega$ and $k_z$ is:

\[
\left( \frac{\omega}{\omega_p} \right)^2 = \frac{1}{2} \left[ \left( 1 + k_{\perp}^2 \lambda_p^2 + k_z^2 \lambda_p^2 \right) \pm \sqrt{ \left( 1 + k_{\perp}^2 \lambda_p^2 + k_z^2 \lambda_p^2 \right)^2 - 4 k_z^2 \lambda_p^2} \right] \quad (2.102)
\]

This solution has two branches, as shown in figure 2.12. For small $k$, the lower branch describes
Alfvén waves, which correspond to low-frequency oscillations of the magnetic field. The polarization vector is dominated by $B_\phi$ and $E_r$. The upper branch represents Langmuir oscillations whose polarization vector is dominated by $E_z$. For larger $k$, the two branches change roles. The upper branch at higher $k$ corresponds to ordinary EM waves with polarization vector along $E_r$ and phase speed close to $c$, while the lower branch corresponds to plasma oscillations with frequency approaching $\omega \sim \omega_p$.

Since $k_\perp$ determines the shape and positions of the two branches of the dispersion relation, we need to select a particular $k_\perp$ in order to see a well-defined dispersion curve. In order to accomplish this, we start with a linear combination of normal modes with a wide range of $k_z$ that lie on both branches and let them evolve in time. The initial state is composed from normal mode exact solutions, so they should remain as normal modes after time evolution. Periodic boundary condition is assumed at both lower and upper boundaries. We used the 20th zero of the Bessel function $J_0$ to better separate the two branches. The numerical dispersion relation is obtained by taking the above initial condition and let it evolve for a light-crossing time of the box. Then we
take the real part of the 2D Fourier transform of the electric field and magnetic field amplitude along a fixed $r$ as a function of $z$ and $t$. The resolution is 1024 by 1024 with 10 cells in a single $\lambda_p$ in the $z$ direction, and the time step is taken to be $0.02\omega_p^{-1}$. Result is as shown in Figure 2.13.

Figure 2.13: Numerical dispersion relation for a cylindrical waveguide. 2D plot shows the real part of the discrete Fourier transform of $E_z(z, t)$ at $r = 1/4r_c$. White dotted curve shows the theoretical dispersion relation calculated from Equation (2.102).

The lower branch of the numerical dispersion relation curves down with respect to the theoretical line at high $k$. This is due to discretization on the grid which creates an effective dispersion relation that deviates from the theoretical one at small wavelengths. This can be seen simply from taking numerical derivatives of the simple solution $e^{ikx}$. The derivative, instead of $ike^{ikx}$, will look like

$$\frac{e^{ik(x+\Delta x)} - e^{ik(x-\Delta x)}}{2\Delta x} = i\frac{\sin(k\Delta x)}{\Delta x} e^{ikx} \quad (2.103)$$

This modifies the behavior of the dispersion relation when $k\Delta x \sim 1$, and reduces the phase velocity of the waves at small wavelengths. It leads to numerical Cherenkov radiation which is produced
when particles travel faster than the speed of the waves inside the medium. The wavelengths of
this kind of numerical radiation is always comparable to the grid scale, therefore can be damped
from the semi-implicit scheme we employed as described in section 2.2.4. In practice, we do not
observe strong sub-plasma scale oscillations which are characteristic of numerical Cherenkov
radiation.

2.4.3 Monopole magnetosphere

Michel (1973) found an analytic solution for the force-free magnetosphere of a rotating mag-
netic monopole. The solution is such that the toroidal field is proportional to the poloidal field:

\[ B_\phi = -\frac{\Omega r \sin \theta}{c} B_p \]  

(2.104)

The corotating electric field defined by \( E = -v_{\text{rot}} \times B / c \) only has a \( \theta \) component and it is equal to
\( B_\phi \) in magnitude.

We attempt to replicate this solution using PIC simulation. We found that field configuration
is identical to the analytic solution given above, with a deviation near the stellar surface which
consists of a parallel electric field that accelerates particles to speed close to \( c \). The simulation is
run in log-spherical coordinates with resolution 512 × 512 in log \( r \) and \( \theta \). Figure 2.14 shows the
solution for \( B_\phi \).

However, the magnetosphere cannot be force-free everywhere, as plasma is injected at the
stellar surface with velocity much smaller than \( c \), and particles have finite inertia. An acceleration
process is needed to accelerate the injected particles to speeds close to \( c \). In a later paper by Michel
(1974), he derived the acceleration profile for the space-charge limited flow, with terminal Lorentz
factor \( \gamma_0 \rightarrow \sigma^{1/2} \), where \( \sigma \) is the magnetization defined as \( \omega_B R_*^2 / \Omega R_{\text{LC}}^2 \). The particle acceleration
CHAPTER 2. PARTICLE-IN-CELL METHOD

Figure 2.14: Michel monopole solution. Color shows the ratio of toroidal vs poloidal magnetic field $B_\phi/B_p$, which should be equal to $-\Omega r \sin \theta / c$. The ratio becomes -1 at the light cylinder, which is $r \sin \theta = 6R_\ast$ in this simulation, marked with a white dashed line. Purple lines are the poloidal field lines, which remain monopolar.

is given by the equation

$$\frac{d}{dr} \left( r^2 \frac{dy}{dr} \right) - 2(y - y_0) = \frac{2 \sigma (\beta_0 - \beta)}{\beta_0}$$

(2.105)

with an overall latitude dependence in the form of $\cos \theta$. Our simulation shows also that $y_0 \sim \sigma^{1/2}$.

In addition, the acceleration profile follows the Michel solution as well (figure 2.15).
Figure 2.15: Acceleration of charge-separated flow for a rotating magnetic monopole. The numerical curve matches the Michel solution fairly well, especially at $\theta = 45^\circ$.

2.5 Discussions and Remarks

We have explained the fundamentals of the Particle-in-Cell technique and the design and structure of the Aperture code. Its novel features are support of radiative transfer models and the built-in flexibility of cuvilinear coordinates. The GPU version also enables the user to run medium-scale problems relatively quickly on a small cost-efficient computer/cluster. In the following chapters we will explore what can be done with this tool to further our understanding of the magnetosphere of neutron stars.

Aperture is not the top performing code on the market at the moment of writing. The VPIC code for example, can push 10 million particles per second per core. Load balancing issue further exacerbates the performance problem, since pair creation naturally leads to some domain patches having many more particles than others, thus becoming the lowest denominator for speed. An
CHAPTER 2. PARTICLE-IN-CELL METHOD

extreme scenario which arose in one of our simulations was that a handful of patches had more than 80% of the particles in the whole simulation box, in which case the simulation slows down to a crawl while most of the nodes are almost idling except for the few most loaded cores.

There is no universal way around this load balancing issue, since it is inherent to all PIC codes. Some like TRISTAN-MP implements dynamic rescaling of domain patches to shrink the loaded patches and grow the idling patches in hope of balancing out, but it does not work well with extreme particle density contrast. An approach that we implemented in Aperture is to annihilate $e^\pm$ pairs, getting rid of excess electrons and positrons in pairs when the total number of particles in a cell exceeds some limit. It helps dramatically with pulsar simulations where the current sheet contains many more particles than the surrounding.

With the help of Rui Hu, another hybrid parallelization scheme is being implemented which reduce the total number of domain patches, but the particles residing in those patches can by dynamically offloaded to other cores to process the particle push and current deposition. Since particle number is the main bottleneck, this scheme in theory can achieve near perfect parallel scaling, and accelerate simulations like pulsar magnetosphere by a factor of 10 or more. A complete vectorization of the code is also being developed, potentially can increase the raw performance by a factor of 2 to 4.

With the treatment of radiative transfer, the Aperture code can be used not only for magnetospheric problems but also for reconnection problems where plasma interaction with the radiation field is important, e.g. in black hole corona.
Chapter 3

Polar-Cap Discharge

3.1 Introduction

Magnetic field lines that pass through the light cylinder of a rotating neutron star are twisted and carry electric currents \( j_B = (c/4\pi)\nabla \times B \). These currents are sustained by electric field \( E_\parallel \) induced along the magnetic field \( B \), and ohmic dissipation \( E_\parallel j \) feeds the observed pulsar activity. The value of \( E_\parallel \) controls the energies of accelerated particles, creation of secondary electron-positron pairs, and emission of radio waves. The accelerating voltage has been discussed in many works on pulsars beginning from early papers in the 1970s (Sturrock [1971], Ruderman & Sutherland [1975], Goldreich & Julian [1969]).

The key dimensionless parameter of the polar-cap accelerator is

\[
\alpha = \frac{j_B}{c\rho_{\text{GJ}}},
\]

where \( \rho_{\text{GJ}} = -\Omega \cdot B / 2\pi c \) is the local corotation charge density of the magnetosphere (Goldreich & Julian [1969]).

For a special value of \( \alpha = \alpha_0 \) (close to unity) a steady state was found for the polar-cap flow with significant particle acceleration (Arons & Scharlemann [1979], Muslimov & Tsygan [1992]).
However, $\alpha$ is not, in general, expected to take this special value (e.g. Kennel et al., 1979). Global solutions for approximately force-free pulsar magnetospheres give $\alpha$ that significantly varies across the polar cap (Timokhin, 2006). In general, $\alpha$ can take any value from $-\infty$ to $+\infty$, depending on the polar cap distance from the rotation axis and the location inside the polar-cap region.

The character of the polar-cap accelerator strongly depends on $\alpha$ (Mestel et al., 1985; Beloborodov, 2008, hereafter B08) The solution with $\alpha = \alpha_0 \approx 1$ is a separatrix between two opposite regimes of efficient and inefficient acceleration. In particular, if $0 < \alpha < 1$, $E_\parallel$ is quickly screened in the charge-separated plasma flowing from the polar-cap surface. The electric field satisfies Maxwell equations that read (in the co-rotating frame of the star, see e.g. Fawley et al., 1977; Levinson et al., 2005)

$$\nabla \cdot \mathbf{E} = 4\pi (\rho - \rho_{\text{GJ}}), \quad (3.2)$$

$$\frac{\partial \mathbf{E}}{\partial t} = 4\pi (\mathbf{j}_B - \mathbf{j}). \quad (3.3)$$

If $0 < \alpha < 1$, there exists a velocity $v = \alpha c$ that allows the charge-separated flow $j = \rho v$ to satisfy both conditions $\rho = \rho_{\text{GJ}}$ and $j = j_B$. If the flow started from the conducting boundary (which has $E = 0$) with $v = \alpha c$, no electric field would be generated (then $\nabla \cdot \mathbf{E} = 0$ and $\partial \mathbf{E}/\partial t = 0$). The actual boundary has $v \neq \alpha c$, as charges are lifted from the polar-cap surface with a small initial $v$, comparable to the thermal velocity in the surface material. The deviation of $v$ from $\alpha c$ implies $\rho \neq \rho_{\text{GJ}}$ or $j \neq j_B$, which generates electric field. B08 argued that equations (3.2) and (3.3) with $0 < \alpha < 1$ always drive the flow toward $v = \alpha c$, like a pendulum is driven by gravity toward its equilibrium position. The resulting oscillations occur in space or time, according to equations (3.2) or (3.3), respectively. For example, the steady-state solution for a cold flow exhibits oscillations in space (Mestel et al., 1985; B08) The oscillatory behavior of the flow with $0 < \alpha < 1$ is, in essence, Langmuir oscillations; they are generated near the boundary where the flow is initially accelerated toward $v = \alpha c$.

1 Hereafter we will refer to this separatrix as $\alpha = 1$, neglecting the deviation of $\alpha_0$ from unity. Precise $\alpha_0$ depends on the curvature of magnetic field lines and the general relativistic effects (Muslimov & Tsygan, 1992), its exact value is close to unity and is not essential for the rest of the paper.
In this chapter, we investigate the accelerator with $0 < \alpha < 1$ in more detail. In Section 3.2, we write down the steady-state solution for the charge-separated flow, generalized to non-zero temperature of the polar-cap. We argue that the flow is unstable to small perturbations and can develop into a complicated time-dependent state with a broad momentum distribution. To explore the behavior of the flow, we perform fully kinetic time-dependent simulations. The method of simulations is described in Section 3.3 and the results are presented in Section 3.4. In Section 3.5, we consider the flow with mixed species of ions extracted from the surface. Our simulations show the turbulent oscillatory behavior of the flow with $0 < \alpha < 1$; particle acceleration in the flow is insufficient to ignite pair creation. Implications of this “dead zone” for radio emission and outer gaps in pulsars are discussed in Section 3.6.

3.2 Steady-state solution for a charge-separated flow

3.2.1 Basic equations

It is natural first to attempt to construct a simple model, assuming that the polar-cap flow is steady in the (rotating) frame of the neutron star. Given the steady magnetic field in this frame, and the steady boundary conditions at the stellar surface — an excellent static conductor that can supply charges with a given temperature, — one could expect a steady state to be established unless the flow is prone to an instability.

Consider a charge-separated flow from the polar cap that carries electric current $j_B$ along magnetic field $B$ (because of a strong field, particles are kept in the ground Landau state). In a steady state $j = j_B$ (equation 3.3). For simplicity, let us assume that $B$ is approximately perpendicular to the polar cap and let $z$ measure the altitude above the stellar surface. A particle of mass $m$ and charge $e$ that starts with a Lorentz factor $\gamma_0 \approx 1$ at $z = 0$ will accelerate as it moves along the magnetic field line,

$$\gamma(z) = \gamma_0 + a(z), \quad a = -\frac{e(\Phi - \Phi_0)}{mc^2},$$

(3.4)
where $\Phi$ is the electric potential. Gravitational acceleration (and centrifugal acceleration in the rotating frame) is neglected compared to the electric acceleration.

The electric potential satisfies Poisson equation,

$$\frac{d^2 \Phi}{dz^2} = -4\pi (\rho - \rho_{GJ}),$$  \hspace{1cm} (3.5)

where we assumed that the potential varies along $z$ much faster than in the transverse directions, i.e. the acceleration length $l_\parallel$ is much smaller than the characteristic transverse scale of the problem $l_\perp$, which may be associated with the size of the polar cap. This condition is satisfied for the flows considered below.\footnote{The term $-\rho_{GJ}$ may be viewed as a fixed background charge density.} The term $-\rho_{GJ}$ may be viewed as a fixed background charge density.

The charge density of the flow itself is given by

$$\rho(z) = j_B \int_1^\infty \frac{w(\gamma_0)}{v(\gamma_0, z)} d\gamma_0. \hspace{1cm} (3.6)$$

Here $v(\gamma_0, z)$ is the velocity of particles that started at $z = 0$ with initial Lorentz factor $\gamma_0$; note that $v^2/c^2 = 1 - \gamma^{-2}$ where $\gamma(z)$ is given by equation (3.4). Function $w(\gamma_0)$ describes the probability distribution of $\gamma_0$. The width of this distribution is controlled by the temperature of the polar cap $T$. For example, $w = \delta(\gamma_0 - 1)$ describes a cold polar cap ($T = 0$) where all particles have $\gamma_0 = 1$.

We multiply both sides of equation (3.5) by $da/dz = -(e/mc^2)d\Phi/dz$ and find,

$$\frac{mc^2}{2e} \frac{d}{dz} \left( \frac{da}{dz} \right)^2 = 4\pi \left[ \frac{j_B}{c} \int_1^\infty dp \frac{d}{dz} (\gamma_0, z) w(\gamma_0) d\gamma_0 - \frac{da}{dz} \rho_{GJ} \right]. \hspace{1cm} (3.7)$$

On the right-hand side, we used $da/dz = -d\gamma/z$ (equation 3.4) and $d\gamma/v = dp/c$. Integration of equation (3.7) in $z$ gives

$$\frac{\lambda_p^2}{2} \frac{da}{dz}^2 \int_1^\infty [p(\gamma_0, z) - p_0] w(\gamma_0) d\gamma_0 - \frac{a(z)}{\alpha}, \hspace{1cm} (3.8)$$

\footnote{Alternatively, the additional term $\nabla^2 \Phi$ could be moved to the right-hand side of equation (3.5) and included in the effective $\rho_{GJ}$.}
where
\[ p^2(y_0, z) = y^2 - 1 = [y_0 + a(z)]^2 - 1. \] (3.9)

In equation (3.8) we used \( a(0) = 0 \) and the boundary condition \( da/dz(0) = 0 \) (the stellar surface is modeled as a perfect conductor that can freely emit charges with \( E_\parallel(0) = 0 \)). We also used \( j_B(z) \approx \text{const} \) and \( \rho_{GJ}(z) \approx \text{const} \), as \( j_B \) and \( \rho_{GJ} \) do not vary much on the characteristic acceleration length \( \lambda_p \), which is defined by
\[ \lambda_p^2 = \frac{mc^3}{4\pi e j_B}. \] (3.10)

This length may be thought of as the plasma skin depth; it is related to the plasma frequency \( \omega_p \),
\[ \lambda_p = \frac{c}{\omega_p}, \quad \omega_p^2 = \frac{4\pi ne^2}{m}, \] (3.11)

with the characteristic plasma density \( n = j_B/ec \).

A quick estimate for \( j_B \) and \( \lambda_p \) in pulsars may be obtained from the following consideration. The magnetic flux through the polar cap \( \Psi \) equals the flux through the light cylinder \( R_{LC} = c/\Omega \). The bundle of open field lines is strongly twisted at the light cylinder (toroidal component comparable to poloidal), and hence it carries electric current \( I \sim c\Psi/2\pi R_{LC} \), according to Stokes theorem. The current density near the star satisfies \( j_B/B \approx I/\Psi \) (which follows from the fact that \( j \) flows along \( B \)), which yields
\[ j_B \sim \frac{\Omega B}{2\pi}. \] (3.12)

This gives the plasma skin depth in the polar-cap accelerator,
\[ \lambda_p \sim \frac{c}{(\Omega\omega_B)^{1/2}}, \quad \omega_B = \frac{eB}{mc}. \] (3.13)

The scale \( \lambda_p \) is much smaller than the typical size of the polar cap \( r_{pc} \sim (R_{NS}^3\Omega/c)^{1/2} \), where \( R_{NS} \sim 10^6 \) cm is the radius of the neutron star.
CHAPTER 3. POLAR-CAP DISCHARGE

Figure 3.1: Steady-state solution for the charge-separated polar-cap flow with $\alpha = 0.8$. Two cases are shown: cold polar cap $T = 0$ (solid curve) and warm polar cap $kT/mc^2 = 0.03$ (dashed curve), which corresponds to average injection momentum $0.22mc$. Dotted curve shows the solution for a cold flow where all particles are injected with the same $p_0 = 0.22$.

3.2.2 Cold and warm solutions

Once the injection distribution $w(\gamma_0)$ is specified, it is straightforward to numerically integrate equation (3.8) and find $a(z)$. In our sample models we chose $w(\gamma_0) = (kT)^{-1} \exp[-(\gamma_0 - 1)/kT]$ with $kT/mc^2 = 0$ (cold) and 0.03 (warm); the average injection momentum $p_0$ in the warm model equals $0.22mc$. Figure 3.1 shows $\Phi(z)$ for the cold and warm solutions.

Figure 3.1 also shows a third model where all particles injected at the polar cap have $p_0 = 0.22$,
i.e. \( w(y_0) \) is a delta-function. In this model, equation (3.8) simplifies to

\[
\frac{\lambda_p^2}{2} \left( \frac{da}{dz} \right)^2 = p(y_0, z) - p_0 + \frac{a(z)}{\alpha},
\]  

(3.14)

[same as equation (3) in B08]. This flow is cold everywhere, i.e. its momentum distribution is described by \( f(p') = n \delta[p' - p(z)] \). As one can see in Figure [3.1] the cold model with \( p_0 \neq 0 \) provides an excellent approximation to the exact warm model that has the same average value of \( p_0 \).

The cold flow solution was discussed in earlier works (Mestel et al., 1985, B08) For \( 1 - \alpha \ll 1 \), the oscillation period is approximately given by (B08)

\[
z_0 \approx \frac{2^{3/2} \lambda_p}{1 - \alpha^2}, \quad 1 - \alpha \ll 1.
\]  

(3.15)

The precise period is obtained by numerical integration; e.g. \( z_0 = 11.0 \lambda_p \) for \( \alpha = 0.8 \). The momentum of the steady cold flow \( p(z) \) oscillates between the injection momentum \( p_0 \ll 1 \) and a maximum value \( p_{\text{max}} \). The minima and maxima are where \( da/dz = 0 \), and from equation (3.14) one finds

\[
p_{\text{max}} = \frac{2\alpha y_0 - (1 + \alpha^2)p_0}{1 - \alpha^2}.
\]  

(3.16)

### 3.2.3 Stability of the flow

Although the cold solution with \( p_0 = 0.22 \) reproduces very well the electric potential \( \Phi(z) \) of the exact warm solution with the same average \( p_0 \), the warm and cold flows are qualitatively different. Their different momentum distribution functions \( f(p, z) \) imply a qualitatively different response to small perturbations.

Consider first the cold-flow solution shown by the blue dotted curve in Figure [3.1] Since all particles are injected with the same momentum \( p_0 = 0.22 \), all of them follow a single trajectory in the phase space \( (z, p) \). They periodically reach the minimum momentum equal to \( p_0 \) at \( z_k = k z_0 \)
\( k = 0, 1, \ldots \) where potential \( \Phi \) reaches maximum. There are no particles with momenta \( p \approx 0 \), so a small perturbation cannot force any particles to reverse their direction of motion, and hence the perturbation will be advected along the flow. This flow may be expected to be stable.

In contrast, the warm flow (dashed curve in Figure 3.1) has a broad distribution of \( p_0 \) that extends from \( p_0 = 0 \). At each peak of the electric potential \( z_k \) there is a population of particles with nearly zero velocities. Consider a perturbation at \( z \approx z_k \). For example, suppose a small bunch \( \mathcal{A} \) of particles with momenta in a range \((p_1, p_1 + \Delta p)\) are slightly pushed forward while the rest of particles are unperturbed. This perturbation implies a local increase in electric current \( j > j_B \) and hence \( \partial E_\parallel / \partial t < 0 \) (equation 3.3), generating negative electric field \( \delta E_\parallel \) at \( z \approx z_k \) that tends to restore the condition \( j = j_B \). In contrast to the initial perturbation, the induced \( \delta E_\parallel \) affects all local particles, regardless of their momentum, not just bunch \( \mathcal{A} \). This has two implications: (1) The induced \( E_\parallel < 0 \) will easily and quickly reduce \( j \) back to \( j_B \) but will be unable to decelerate bunch \( \mathcal{A} \) to the momentum it would have in the steady state flow — bunch \( \mathcal{A} \) will continue to move to \( z > z_k \) with a larger momentum. (2) \( \delta E_\parallel < 0 \) will give very slow particles \( p \approx 0 \) negative velocities, creating a new bunch \( \mathcal{B} \) that slides backward down the potential hill. Bunch \( \mathcal{B} \) creates \( j < j_B \) at \( z < z_k \), and the system reacts there by inducing a small \( \delta E_\parallel > 0 \), which accelerates all local particles, regardless their momenta, not just bunch \( \mathcal{B} \). As a result, \( j \) quickly recovers to \( j_B \), however, bunch \( \mathcal{B} \) is not stopped from moving backward and away from \( z = z_k \).

Thus one can see that the perturbation creates permanent damage to the steady state that broadens the momentum distribution by creating backflowing particles. This perturbation is not advected away along the flow, and can develop further. The backflowing particles turn out to be trapped between two peaks of the electrostatic potential. Further development can be studied with kinetic time-dependent simulations; it eventually completely destroys the steady state solution.
3.3 Numerical setup

Our numerical method is similar to that used by Beloborodov & Thompson (2007, hereafter BT07) The plasma is modeled as a large number $N \sim 10^6$ of individual particles that flow along the magnetic field lines. We assume that the magnetic field is fixed in the co-rotating frame of the star; thus $j_B$ and $\rho_{GJ}$ are fixed. Then the problem becomes essentially one-dimensional, as discussed in detail in BT07. In the present chapter, we consider only charge-separated flows, with no pair creation. Three other differences from the magnetar simulation in BT07 are as follows: (1) The magnetar problem had $\alpha \gg 1$ ($\rho_{GJ}$ was negligible compared with $j_B/c$); in contrast, $\rho_{GJ}$ is crucial for polar-cap flows considered here. (2) The presence of gravity was essential for the closed-field circuit considered in BT07, where the global plasma flow was simulated (on a scale comparable to the radius of the star); in the problem considered here the electric fields are screened on a much smaller scale $\sim \lambda_p$ and the gravitational acceleration plays no role. (3) The flow behavior on the small scales $z \ll R_{NS}$ may be studied using a small computational box $H \ll R_{NS}$ with an open outer boundary (see below).

In the absence of pair creation, the flow is composed of particles lifted from the surface; we assume that all of them have the same mass $m$ and charge $e$. The particle motion is described by the equation,

$$\frac{dp_i}{dt} = \frac{eE_{\parallel}(z_i)}{mc}, \quad i = 1, \ldots, N,$$

(3.17)

where $p_i$ is the momentum of the $i$-th particle in units of $mc$, and $E_{\parallel}(z_i)$ is the self-consistent electric field at the particle location $z_i$. The field is found by integrating Gauss law (equation 3.2) along the magnetic field line,

$$E_{\parallel}(z_i) = 4\pi \left[ eN(z_i) - \rho_{GJ}z_i \right].$$

(3.18)

Here $N(z_i)$ is the column density of particles between $z = 0$ and $z = z_i$, and we used the boundary condition $E_{\parallel}(0) = 0$, as the material below the stellar surface is assumed to be a
very good conductor that can emit free charges. Divergence of the perpendicular component of electric field $E_\perp$ is neglected in equation (3.18) (see BT07 for discussion of this approximation). The approximation $|\nabla_\perp \cdot E_\perp| \ll |dE_\parallel/dz|$ is valid if the characteristic scale of the flow acceleration $z_0$ is smaller than the transverse scale $l_\perp$, which is limited by the polar-cap size $r_{pc}$; the condition $z_0 \ll r_{pc}$ is satisfied in the dead-zone models presented below. We also assume that $\rho_{GJ}$ is approximately constant on scale $z_0$. Equations (3.17) and (3.18) in essence describe a relativistic, time-dependent diode problem with an additional fixed background charge density $-\rho_{GJ}$.

As we track the motion of all particles individually, the continuity equation is automatically satisfied; for a charge-separated flow it is equivalent to charge conservation,

$$\frac{\partial \rho}{\partial t} + \frac{\partial j}{\partial z} = 0.$$  \hspace{1cm} (3.19)

equation (3.3) follows from equations (3.2) and (3.19), so we will not need equation (3.3). Instead, the parameter $j_B$ enters the problem as a boundary condition. The magnetic field lines are frozen in the excellent conductor below the stellar surface, which sustains $j(0) = j_B$. This condition is enforced in the simulation by injecting the charges in the computational box at $z = 0$ with the fixed rate $j_B$ (BT07).

The electric current $j_B$ is enforced at one boundary $z = 0$. Since the computational box has a finite size $H$, we also have to choose the boundary condition at $z = H$ and the value of $H$. In all sample models shown in this chapter we use the simplest boundary condition: particles moving out of the box are lost and no particles enter the box at $z = H$. This condition may be refined by allowing a small inflow of returning particles at the outer boundary. We ran test simulations that show that the refinements are not important as long as the boundary is sufficiently far, so that $H$ is much greater than the characteristic scale of the flow acceleration.

In the one-dimensional model, the transverse gradients are neglected, and the flow effectively has a slab geometry. Then it is sufficient to follow particles flowing through a small area $A$ of the slab. This allows one to chose a reasonable number of particles in the computational box,
$N \sim A H n$, e.g. $N \sim 10^6$, so that their dynamics can be followed in a reasonable computer time. On the other hand, $N$ should be large enough so that the plasma scale $\lambda_p$ contains many particles $N_p = A \lambda_p n$.

In summary, we choose $N$ and $H$ so that

$$\frac{H}{\lambda_p} \gg 1, \quad N_p = \frac{\lambda_p}{H} N \gg 1.$$  \hfill (3.20)

In this limit, the results are expected to be independent of the choice of $N$ and $H$ (we verify this by varying the two parameters in Section 3.4). For most of our simulations $H = 100 \lambda_p$ and $N \sim 10^6$. Another requirement is a small time step of the simulation, $\Delta t \ll \omega_p^{-1}$, so that plasma oscillations are well resolved.

### 3.4 Results

#### 3.4.1 Steady state tests

In our simulations and in reality the plasma above pulsar polar caps is collisionless. In the absence of pair creation it must satisfy the Vlasov equation,

$$\frac{\partial F}{\partial t} + \mathbf{v} \cdot \nabla F + \frac{d\mathbf{p}}{dt} \cdot \nabla_p F = 0,$$  \hfill (3.21)

where $F(t, z, p)$ is the particle distribution function in phase space. The electric current is $j(t, z) = \rho \bar{v}$ where $\bar{v}(t, z)$ is the average velocity of the particles. As a first simple test, consider a uniform flow with $\rho(z) = \rho_{GJ}$, $\bar{v}(z) = \alpha c$, and $E_\parallel(z) = 0$. It is easy to see from equations (3.17), (3.18) and (3.21) that the flow must remain in this state. This behavior is reproduced by our simulations. The steady uniform flow can have any momentum distribution $F(p)$ as long as $\bar{v} = \alpha c$. Note that it requires a continual injection of particles at $z = 0$ with the average velocity $\bar{v} = \alpha c$ (which also requires $0 < \alpha < 1$).
As a second test, consider a “cold” flow where all particles move with momentum $p(z)$, with zero momentum dispersion. Suppose the flow is injected at $z = 0$ with velocity $v_0 < \alpha c$. Then $E_\parallel$ must be generated, accelerating the flow. In a steady state, the solution for the cold flow must have the form, $F(z, p') = n(z) \delta[p' - p(z)]$, where $p(z)$ and $n(z)$ can be described analytically. We first test the special case $\alpha = 1$ (Michel, 1974). The flow is accelerated by the self-consistent $E_\parallel(z)$, and $p$ exceeds unity at $z \sim \lambda_p$. At heights $z \gg \lambda_p$, velocity approaches $c$, charge density of the flow $\rho = j_B/v$ approaches $\rho_{GJ}$, and electric field $E_\parallel$ asymptotes to a constant value,

$$E_\parallel = \left[ \frac{8\pi mcj_B}{e} (\gamma_0 - p_0) \right]^{1/2} \left[ 1 + O(p^{-1}) \right], \quad (3.22)$$

where $\gamma_0 = (1 - v_0^2/c^2)^{-1/2}$ and $p_0 = \gamma_0 \beta_0$. Then the flow momentum keeps growing linearly with $z$,

$$p(z) = [2(\gamma_0 - p_0)]^{1/2} \frac{z}{\lambda_p}, \quad z \gg \lambda_p. \quad (3.23)$$

This solution is reproduced by our simulations with “cold injection” — all injected particles at $z = 0$ have velocity $v_0$. After an initial relaxation period (comparable to the light crossing time of the computational box) the system forgot initial conditions and relaxed to the steady state shown in Figure [3.2] (in this example, $v_0 = 1/6$). The charge density of the flow is large near the polar cap surface and asymptotes to $\rho_{GJ}$ at $z \gg \lambda_p$, as expected.

Then we studied cold flows with $0 < \alpha < 1$ and fixed injection velocity $\beta_0$. We chose in our sample numerical model $\alpha = 0.8$ and $\beta_0 = 0.2$. The computational box was initially empty; the plasma injected at $z = 0$ filled the box on the dynamical timescale $\sim H/c$ and established a steady state shown in Figure [3.3]. The steady state is in perfect agreement with the analytical model of Section 2.2. The charge density $\rho(z)$ has spikes at $z = k\pi / \alpha \beta_0$ (where the flow has the minimum velocity $\beta_0$; the height of each spike is $\rho_{\text{max}} = j_B/\beta_0 = (\alpha/\beta_0)\rho_{\text{GJ}}$. The charge spikes are associated with maxima of the electric potential (Figure [3.3c]). The oscillating momentum has maxima $p_{\text{max}} = 3.6$, in excellent agreement with equation (3.16). The period of oscillation is
Figure 3.2: Test run for a cold-flow model with $\alpha = 1$ and $v_0 = c/6$. The flow relaxed to a steady state in the entire box $H = 10^2\lambda_p$ on the light-crossing timescale, $H/c$; the state of the system is shown at $t = 10H/c$. (a) Flow momentum per particle $\rho(z)$ in units of $mc$. (b) Charge density $\rho(z)$. 
Figure 3.3: Cold flow with $\alpha = 0.8$ and $\beta_0 = 0.2$ at time $t = 1.45 H/c$ (a) Momentum $p$ (in units of $mc$). (b) Charge density. (c) Electrostatic potential.
As anticipated in Section 2.3, we find that the steady state becomes unstable if we reduce $\beta_0$ to zero. Then any small perturbation (e.g. due to numerical error) completely destroys the steady state; instead, a time-dependent state forms, with a broadened momentum distribution function. A steady flow with a finite $\beta_0 \neq 0$ can also be destroyed, although in this case a finite, sufficiently large perturbation is required. In fact, this case provides a better setup for a numerical analysis of the instability, as we can control the form of the initial perturbation and then observe how it destroys the flow that was stable before the perturbation was applied. We made such an experiment with the flow with $\alpha = 0.8$ and $\beta_0 = 0.2$. We applied a perturbation that was localized in space and time — a small “kick” $\delta p$ was given to all particles located in a small region $\delta z = \lambda_p / 2$; in this experiment $\delta p$ had a Gaussian distribution with the mean value and dispersion equal to 0.02. We observed the following evolution. As the localized perturbation moved along with the background flow, it was greatly amplified when it reached the potential maximum (which corresponds to the minimum $p_0 \approx 0.2$ of the steady-state solution, see Figure 3.3), and some particles acquired a negative momentum, i.e. reversed their direction of motion. The reversed particles became mostly trapped between two potential maxima, but some of them were able to penetrate even further back, beyond the preceding potential peak. The perturbation further spread in the phase space and the damage to the initial steady-state solution was further amplified with time, in particular near the potential maxima. Eventually, the entire flow became strongly time-dependent and the regular periodic structure of potential peaks disappeared.

The amplification of small (linear) perturbations at the potential maximum can be understood as follows. Consider a particle whose Lorentz factor differs from that of the background cold flow by a small $\delta \gamma$. As the particle moves along with the flow, its deviation $\delta \gamma$ remains constant, as it travels in the same electrostatic potential as the background flow (cf. equation 3.4). Using the
relation \( \frac{d\gamma}{dp} = \beta \), we find the perturbation of momentum \( \delta p \) that corresponds to \( \delta \gamma \),

\[
\delta p = \frac{\delta \gamma}{\beta} \propto \beta^{-1}. \tag{3.24}
\]

It grows as the particle (and the background flow) decelerates near the potential maximum; the corresponding amplification factor \( \beta_0^{-1} \) is particularly large if \( \beta_0 \) is small.

Generation of backflowing particles at the potential peaks \( z_k \) plays the key role in disrupting the steady state. In a flow with a finite minimum velocity \( \beta_0 > 0 \), the external perturbation would need to rob particles energy \( \gamma_0 - 1 \approx \beta_0^2/2 \) before they could be reflected by the potential hill. Thus the gap \( \gamma_0 - 1 \) stabilizes the flow against infinitesimal perturbations, and only a sufficiently strong kick may disrupt the flow.

The trapped or backflowing particles have a deteriorating effect on the steady state because they are not advected away with the flow and instead repeatedly approach the same potential peaks, amplifying the perturbations. In addition, one can view the trapped particles as extra charge that distorts the electric field. Let \( N_{\text{trap}} \) be the number of particles trapped between two potential peaks \( z_{k-1} \) and \( z_k \); they create electric field \( E' = 4\pi e N_{\text{trap}} \) at \( z > z_k \). The corresponding distortion of the electrostatic potential \( \Phi' = -E'z \) grows linearly with \( z \) and becomes significant at sufficiently large \( z \) even if \( N_{\text{trap}} \) is small. The distance \( z \) required to produce \( e\Phi' \sim mc^2 \) is \( z \sim (N_p/N_{\text{trap}})\lambda_p \). This behavior is qualitatively confirmed by our numerical experiments with larger simulation boxes \( H \) — the flow was found to become more unstable with increasing \( H \).

### 3.4.2 Time-dependent state with warm particle injection

In a more realistic model, particles are lifted from the polar cap with a thermal velocity dispersion \( \Delta v_0 \sim v_0 \). The flow still starts with a small velocity \( \bar{v} \ll c \) and hence with a large charge density \( \rho \gg \rho_{\text{GJ}} \), which self-consistently generates the accelerating electric field. The basic acceleration mechanism is the same as for the cold flow shown in Figures 3.2 and 3.3. However, there is a new feature: particles with different initial velocities behave differently in the collective
electric potential, and the charge density $\rho(z)$ is changed from the cold-flow solution, even though $\Delta v_0 \ll c$. Some particles have $v \approx 0$ and can reverse their motion in the regions of growing potential ($E_{\parallel} < 0$), which greatly complicates the behavior of the distribution function $F(z, p)$.

In our simulations, we modeled the warm injection by a one-dimensional Maxwell distribution, which is a simple Gaussian with dispersion $\Delta v_0$ equal to the mean value $\bar{v}_0$; we chose $\bar{v}_0 = 0.2c$. As initial conditions we took the steady-state solution (Section 2). The main parameter of the flow is $\alpha$, and we calculated the evolution of the system for several values of $\alpha$ in the range $0 < \alpha < 1$. As expected, we found that the steady state was quickly destroyed and the flow kept oscillating in both space and time. The basic parameters of the flow remained, however, similar to the steady cold model. The average charge density (averaged over oscillations) is nearly equal to $\rho_{GJ}$ and the average velocity $\bar{v}$ is nearly equal to $\alpha c$, so that the condition $\bar{j} = j_B$ is satisfied. Figure 3.4 shows the evolution of the hydrodynamic velocity $\bar{v}(t)$ measured at a fixed location $z_1$ (we chose $z_1 = 50\lambda_p$, in the middle of the computational box; $\bar{v}$ was calculated by averaging over particles inside a small bin around $z_1$, of width $2\lambda_p$). The hydrodynamic velocity $\bar{v}(t)$ oscillates around $\alpha c$; these oscillations have a relatively small amplitude $\delta v \ll \bar{v}$.

The moderate value of the hydrodynamic velocity does not, in principle, exclude acceleration of a fraction of particles to much higher energies. We therefore also studied the momentum distribution of particles in the flow. Figure 3.5a shows a random snapshot of the particle distribution in the phase space for the flow with $\alpha = 0.8$. We randomly chose 1000 particles between $z = 0$ and $z = 100\lambda_p$ and the figure shows their locations in the two-dimensional phase space $(z, p)$. The simulation demonstrates the following:

(1) There is no high-energy tail in the momentum distribution.

(2) At each $z$, the momentum distribution has a pronounced narrow peak at $p_{\text{peak}}$. Thus, a large fraction of particles form a cold flow. The momentum of cold particles $p_{\text{peak}}$ is slightly above the average (hydrodynamical) $\bar{p}(z)$, and both are above (but comparable to) the value of $p_{\text{max}}$ predicted by the steady-state model.
Figure 3.4: Evolution of the hydrodynamical velocity $\bar{v}$ of the flow measured in the middle of the computational box. Three models are shown: $\alpha = 0.95$ (purple), 0.8 (blue) and 0.6 (dark green). In all three cases, the time-average value of $\bar{v}$ equals $\alpha$.

(3) There is a low-energy wing in the momentum distribution which extends to negative momenta (up to 10% of all particles have $p < 0$ in our sample model). This broad component of the particle distribution has a hydrodynamic velocity close to zero and does not contribute much to the current density; however it makes a significant contribution to charge density. In our sample model, about 20% of particles reside in the broad component, and this fact has a simple explanation. From the point of view of the cold stream dynamics, the broad component provides a background that offsets the effect of vacuum charge density $\rho_{GJ}$ by the fraction of 20%. This fraction equals $1 - \alpha$, so that the cold stream may move with $\nu \approx c$ and carry $j_B$ without the mismatch in charge density that would generate strong $E_\parallel$. In essence, the broad component with backflowing particles allows the plasma to self-organize so that the cold stream can keep $\nu \approx c$. This is in contrast to the steady-state solution in Section 2, where all particles form a stream with
Figure 3.5: Snapshot of 1000 randomly chosen particles in phase space for the flow with $\alpha = 0.8$ and $\beta_0 = 0.2$. Red dashed line shows the maximum momentum $p_{\text{max}}$ for the steady cold solution with the same $\alpha = 0.8$ and $\beta_0 = 0.2$ (Section 3.1). (a) Random snapshot for the simulation with box size $H = 100\lambda_p$. (b) Another random snapshot of a similar simulation with a larger computational box $H = 200\lambda_p$. 
CHAPTER 3. POLAR-CAP DISCHARGE

90

a positive velocity \( v \neq \alpha \), which result in the periodic deceleration and acceleration of the stream.

(4) Both the fluid momentum \( \bar{p}(z) \) and the peak momentum \( p_{\text{peak}} \) fluctuate in time (the corresponding curves in Figure 3.5 move in time). However, the qualitative form of the phase-space distribution remains similar to that in Figure 3.5.

Note that the average momentum \( \bar{p} \) does not correspond to the average velocity \( \bar{v} \) shown in Figure 3.4, in the sense that \( \bar{p} \neq \bar{\beta}(1 - \bar{\beta}^2)^{-1/2} \), because of the broad low-energy tail of the distribution function. Compared with velocity average, the averaging of momentum gives a higher weight to particles with large \( \beta \) because of the addition factor \( \gamma \) in \( p = \gamma \beta \). The average velocity remains close to \( \alpha \bar{c} \), and the averaged momentum is larger than \( \bar{\beta}(1 - \bar{\beta}^2)^{-1/2} \).

It should also be noted that the pronounced narrow stream and the broad low-energy stream together form a two-stream system. Conventionally one would expect the configuration to suffer from standard micro-instabilities. However, we do not observe such instabilities here. One important reason is that the flow remains turbulent even after establishing this quasi-steady state. The momentum of particles in the colder stream oscillates both in time and space, thus any particle would not be able to interact with a single wave for an extended period of time. In this case, exponential growth of particular waves cannot occur, which kills the instability.

As seen in Figure 3.5a, the flow momentum \( p_{\text{peak}} \) decreases near the outer boundary of the computational box \( z = H \). This is an artifact of the boundary condition (free escape with no backflow), which suppresses backflow density near the boundary. As a result, a modest negative electric field is induced near the boundary, decreasing \( p_{\text{peak}} \) so that the flow carries the required electric current \( j_B \). For comparison, Figure 3.5b shows a random snapshot of a similar model (in the same interval \( 0 < z < 100\lambda_p \)) that has twice as large computational box, \( H = 200\lambda_p \). As we increase \( H \), the boundary effect moves away to larger \( z \), affecting the flow properties only at \( z \approx H \). The fraction of backflowing particles measured inside the box remained unchanged at \( \sim 10\% \).

We also checked whether the flow momentum depends on the size of the computational box. We ran several simulations with the same \( \alpha = 0.8 \) and different box sizes \( H \). In each simulation,
we measured the fluid momentum $\bar{p}$ in the center of the box (using a bin $\Delta z = 2\lambda_p$) at time $t = 100\omega_p^{-1}$. The results are shown in Figure [3.6]. There is no systematic variation in $\bar{p}$ with the box size; the small variations ($\lesssim 10\%$) are consistent with the fluctuations of $\bar{p}$ in time for each model.

The other numerical parameter of the simulations is $N_p$ (Section 3). We checked that the results do not depend on $N_p$ as long as it is large; variations in $N_p$ around $\sim 10^4$ did not change the measured parameters of the flow.

### 3.5 Mixed ion flow and two stream instability

If $j_B > 0$ (which is equivalent to $\rho_G J > 0$ for $\alpha > 0$), the charge-separated flow pulled out from the polar cap is made of ions. Different species of ions may end up in such a flow, and they will be accelerated to different velocities.
The mixed ion flow shares many features with the identical-particle model studied in the previous sections. Steady state solutions exist for such flows, which can be obtained using the method described in Section 3.2.2. Ions with different masses and charges move with different hydrodynamical momenta and co-exist in a common, periodic electrostatic potential. This steady solution is prone to kinetic instability similar to that described in Sections 2 and 4. There is, however, an important new feature: the ion streams with different hydrodynamical momenta are prone to two-stream instability.

To study the behavior of the mixed ion flow we use a simple modification of our numerical simulation. Consider a mixture of protons and helium nuclei (alpha-particles). The particle injection at $z = 0$ now consists of two ions species; they have charges $e_1$ and $e_2 = 2e_1$, and masses $m_1$ and $m_2 = 4m_1$. The two species are injected with equal rates $\dot{N}_1 = \dot{N}_2$. Then alpha-particles carry electric current $j_2 = e_2 \dot{N}_2$ that is two times larger than the proton current $j_1 = e_1 \dot{N}_1$. Thus, $j_2 = (2/3) j_B$ and $j_1 = (1/3) j_B$ are maintained at the boundary.

To define a characteristic plasma skin depth $\lambda_p$ we use equation (3.10) where we replace $e, m, j_B$ by $(e_1, m_1, j_1)$ or, equivalently, by $(e_2, m_2, j_2)$ (note that $e_2 j_2 / m_2 = e_1 j_1 / m_1$). The characteristic plasma frequency is defined by $\omega_p = c / \lambda_p$.

Figure 3.7 shows a snapshot of the phase-space distribution of ions long after the beginning of the simulation. In this sample model $\alpha = 0.4$; the modest value of $\alpha$ (not close to unity) implies modest Lorentz factors of particles and fast development of instabilities. The flow exhibits the following features:

1. One period of the steady state solution is reproduced near the injection boundary $z = 0$. The period $z_0 \approx 3\lambda_p$ agrees with the result from numerical integration of the corresponding steady state model. (This feature is stable in our sample model because we chose a relatively large injection velocity $v_0 = 0.4c$.)

2. At larger $z$ the periodic flow becomes unstable and develops into a configuration similar to that in Figure 5, except that now we have two cold variable streams. Besides the cold streams, there is
a broad distribution of ions with smaller momenta and a negligible hydrodynamic velocity. In our sample model $\alpha = 0.4$ and correspondingly the broad component is self-organized to contain up to $\sim 60\%$ of particles, so that the streams may move with a relativistic speed without mismatch in charge density that would generate strong electric fields.

(3) Further from the boundary (at $z > 20\lambda_p$), the two streams develop a two-stream instability. The growth rate of the instability may be estimated using an idealized model of two streams with densities $n_1, n_2$ and velocities $v_1, v_2$. It is straightforward to derive the dispersion relation for...
Langmuir modes with frequency $\omega$ and wave-vector $k$ (e.g. Melrose, 1986); it gives,

$$1 - \frac{\omega_1^2}{\gamma_1^3(\omega - kv_1)^2} - \frac{\omega_2^2}{\gamma_2^3(\omega - kv_2)^2} = 0,$$  \hspace{1cm} (3.25)

where $\omega_1^2 = 4\pi n_1 e_1^2 / m_1$ and $\omega_2^2 = 4\pi n_2 e_2^2 / m_2$. Using $\omega_1 \approx \omega_2 \approx \omega_p$ and the characteristic values of velocities $v_1, v_2$ from our simulation, we find from equation (3.25) that the most unstable modes have $\omega$ comparable to $\omega_p$ and their growth rate is $\Gamma \sim 0.1\omega_p$. The distance over which Langmuir waves are amplified is roughly $10\lambda_p$. As a result of the instability, the two streams are smeared out at large $z$, in particular the stream of lighter ions. No significant particle acceleration is seen in the simulation.

3.6 Discussion

We have presented detailed one-dimensional time-dependent simulations of the plasma flow extracted from the polar caps of neutron stars. The simulations provide a fully kinetic description of the flow, with self-consistent electric field and particle distribution function. In this chapter, we focused on the regime $0 < \alpha < 1$, where $\alpha$ is the main parameter of the flow defined by equation (3.1). In agreement with the estimates of B08, we find that the particles are accelerated to Lorentz factors,

$$\gamma \approx \frac{1 + \alpha^2}{1 - \alpha^2},$$  \hspace{1cm} (3.26)

and are not capable of igniting pair creation. In this sense, flows with $0 < \alpha < 1$ are “dead.” They are sustained by a modest voltage, oscillating in space and time.

The simulations show how a kinetic instability develops and disrupts the ideal periodic structure found in analytical models of the dead zone (Section 2). We find that the momentum distribution function has two distinct parts — a variable “cold stream” and a broad wing at low momenta, which includes particles flowing backward to the polar cap. Even though the flow is turbulent, it shows no signs of particle acceleration to energies higher than that of the cold stream.
The value of parameter $\alpha$ depends on the location and geometry of the polar cap. The simplest magnetospheric configuration is that of a centered dipole. Then the parameter $\alpha$ depends on the angle between the magnetic and spin axes, $\xi$; besides, it varies across the polar cap. For nearly aligned rotators ($\xi \approx 0$), $0 < \alpha < 1$ in the central part of the polar cap and $\alpha < 0$ in a ring-shaped zone near the edge of the polar cap (Timokhin, 2006; Parfrey et al., 2012). In this case, the dead zone occupies the central part of the polar cap, and $e^{\pm}$ discharge must be confined to the ring, matching the phenomenological “hollow cone” model of pulsar emission. In contrast, the polar cap of an orthogonal rotator ($\xi \approx \pi/2$) has $|\alpha| \gg 1$, which enables $e^{\pm}$ discharge for the entire polar cap. At arbitrary misalignment $0 < \xi < \pi/2$, the values of $\alpha$ can be provided by global three-dimensional simulations of the magnetospheric structure (e.g. Spitkovsky, 2006) and should play a key role for the geometry of the radio beam.

We presented our results using plasma skin depth $\lambda_p$ as a unit of length and particle rest-mass $m_e c^2$ as a unit of energy. In this form, the results do not depend on the charge or mass of the particles extracted from the polar cap, as long as our assumption — that the flow is made of identical particles — is satisfied. In particular, equation (3.26) is valid for both electron flow ($\rho_{GJ} < 0$) and ion flow ($\rho_{GJ} > 0$), and the phase-space distribution shown in Figure 5 describes both cases. Note that the accelerating voltage is proportional to the particle mass; voltage implied by equation (3.26) is different for ions and electrons by the factor of $m_i/m_e \sim 2 \times 10^3$. The relatively high voltage in the ion flow, $e\Phi \approx \gamma m_i c^2(1 + \alpha^2)/(1 - \alpha^2)$, is still hardly sufficient to ignite $e^{\pm}$ pair discharge by a seed electron or positron.

The identical-particle model may not hold for an ion flow; in this case, new effects may enter the problem. Firstly, heavy ions pulled out from the polar cap may not be completely ionized and begin to lose electrons as they are accelerated and interact with the X-rays above the stellar surface; this process effectively creates new charges, reminiscent of pair creation (e.g. Jones, 2002). Secondly, the ion flow may be a mixture of different nuclei which will be accelerated to different Lorentz factors. The mixed ion flow is prone to two-stream instability, possibly leading to the
formation of plasma clumps and generation of coherent radio emission. In our simulations, we observe the expected two-stream instability, however we do not observe significant structure (clumps) in the turbulent flow. This may change in three-dimensional simulations. The frequency of excited waves \(\omega \approx 2\gamma_1^{1/2}/\omega_1\) is in the radio band, and coherent emission from clumps could create bright coherent emission. It remains to be seen whether this mechanism can contribute to the pulsar emission. If it does, it would create an additional component of the radio pulse. In the case of aligned rotator, the additional component would be generated in the central region of the polar cap, leading to a "hollow cone + core" structure of the radio pulse.

The charge-separated model of the dead zone can be modified to include possible backflooding particles from distant parts of the open field-line bundle (e.g. from a pair-producing outer gap). These particles can contribute to the current density and also serve as an additional background charge density, which may be modeled as a contribution to the effective "vacuum" charge density \(-\rho_{GJ}\). This would change the effective \(\alpha\) \cite{Lyubarskii1992, B08}, most likely reducing it.

An outer gap is expected to form in a charge-separated flow near the null surface \(\mathbf{B} \cdot \mathbf{\Omega} = 0\) \cite{Cheng1986}. On a given field line, the outer gap will be screened if it is loaded by multiple \(e^\pm\) pairs produced by discharge at the polar cap. Thus, suppression of \(e^\pm\) discharge near the field line footpoint is an essential condition for the existence of an outer-gap accelerator. Therefore, one can expect an outer gap to form on field lines with footpoints in the dead zone.

We did not simulate in this chapter flows with \(\alpha > 1\) or \(\alpha < 0\); in these cases particles must be strongly accelerated. This regime leads to an \(e^\pm\) discharge that must be unsteady, with a significant intermittent backflow \cite{B08}. A model for oscillating discharge can be studied in hydrodynamical approximation \cite{Levinson2005}, however a fully kinetic description is essential. We defer the kinetic time-dependent simulations with pair creation to a future paper.
Chapter 4

Global Pulsar Magnetosphere

4.1 Introduction

The standard picture of a pulsar magnetosphere assumes that it is filled with plasma and corotates with the neutron star with angular velocity $\Omega$ (Goldreich & Julian[1969] hereafter GJ). GJ considered the aligned rotator (magnetic dipole moment $\mu$ parallel to $\Omega$); then it was generalized to inclined rotators. The plasma sustains the “corotational” electric field $\mathbf{E} \approx -\mathbf{v}_{\text{rot}} \times \mathbf{B}/c$ (with $\mathbf{v}_{\text{rot}} = \Omega \times \mathbf{r}$), which implies the local charge density $4\pi \rho_{\text{GJ}} = \nabla \cdot \mathbf{E} \approx -2\Omega \cdot \mathbf{B}/c$. A key feature of the GJ model is the electric current $I_{\text{GJ}}$ flowing out of and into the star along the open magnetic field lines that extend to the light cylinder $R_{\text{LC}} = c/\Omega$. GJ showed that the open field lines are twisted and exert a spindown torque on the rotator. The circulating current is $I_{\text{GJ}} \approx \mu \Omega^2/c$, and the corresponding spindown power is $\dot{E} \approx \Omega^4 \mu^2/c^3$.

This picture was, however, never verified by a first-principle calculation and was questioned (Michel, 2004; Gruzinov, 2013). It was shown that charges lifted from the star by the rotation-induced electric field form the “electrosphere” — a corotating dome+torus structure, with a huge gap between them and no electric current (Jackson, 1976; Krause-Polstorff & Michel, 1985a). Although the electrosphere is prone to diocotron instability (Pétri et al., 2002; Spitkovsky & Arons, 2002), it was unclear if it could relax to the GJ state.
In addition to lifted charges, $e^\pm$ pairs are created around pulsars (Sturrock, 1971). This provides plasma capable of screening the electric field component parallel to the magnetic field, $E_\parallel$. The negligible plasma inertia and $E_\parallel = 0$ provide the “force-free” (FF) conditions, which imply GJ corotation. The global solution for FF magnetospheres was obtained using various numerical techniques (Contopoulos et al., 1999; Timokhin, 2006; Spitkovsky, 2006; Kalapotharakos & Contopoulos, 2009; Parfrey et al., 2012).

Its characteristic feature is a thin current sheet supporting a discontinuity of $\mathbf{B}$ and the Y-point near the light cylinder. It was verified with particle-in-cell (PIC) simulations that sprinkling pairs with a high rate everywhere around the neutron star would drive the magnetosphere to the FF configuration (Philippov & Spitkovsky, 2014).

A self-consistent model must, however, demonstrate how and where the plasma is created and to identify the regions of $E_\parallel \neq 0$ (called “gaps”) where particles are accelerated to high energies. Besides testing the FF approximation, the self-consistent model would show how the pulsar radiation is produced, how the plasma flows in the magnetosphere and gets ejected. This problem was posed soon after the discovery of pulsars and proved to be difficult. Three types of gaps were proposed: polar-cap gap (Sturrock, 1971; Ruderman & Sutherland, 1975), slot gap (Arons, 1983; Muslimov & Harding, 2004), and outer gap (Cheng et al., 1986).

The only reliable way to solve the problem is a first-principle calculation of the self-consistent dynamics of the electromagnetic field and pair discharge in the magnetosphere. Below we present such a direct numerical experiment. Our simulations are performed with a new 2.5D PIC code, developed from scratch and designed for neutron-star magnetospheres. The code calculates the fully relativistic dynamics of particles and fields on a curvilinear grid, traces the emission of gamma-rays and their conversion to pairs. The fields obey Maxwell equations, $\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E}$ and $\frac{\partial \mathbf{E}}{\partial t} = c \nabla \times \mathbf{B} - 4\pi \mathbf{J}$, and exert force on particles $e(\mathbf{E} + \mathbf{v} \times \mathbf{B}/c)$. We use Esirkepov (2001) charge-conserving scheme for calculating $\mathbf{J}$ and a semi-implicit algorithm for the field evolution. A detailed description of the code and tests was given in chapter 2.
Pair creation by accelerated particles occurs in two steps: production of gamma-rays and their conversion to \(e^{\pm}\). In many pulsars, the conversion is only efficient at \(r \ll R_{\text{LC}}\) where the magnetic field is strong. In young fast pulsars pairs can be created through photon-photon collisions inside and around the light cylinder \(\text{Cheng et al., 1986}\). For brevity, we call such rotators "type I." Pulsars with pair creation confined to \(r \ll R_{\text{LC}}\) will be called type II.

### 4.2 Problem formulation and simulation setup

The axisymmetric pulsar is described by its radius \(R_\star \approx 10\ \text{km}\), angular velocity \(\Omega\), and magnetic dipole moment \(\mu\) (aligned or anti-aligned with \(\Omega\)). These parameters set the energy scale of the problem. The neutron star is a nearly ideal conductor, and its rotation induces voltage \(\Phi_0 \approx \mu \Omega^2 / c^2\) across the footprint of the open field line bundle; it corresponds to possible particle acceleration up to Lorentz factors \(\gamma_0 = e\Phi_0 / m_e c^2\). We start our simulations with \(\Omega = 0\) and the vacuum dipole field. Then we gradually spin up the star: \(\Omega\) grows linearly until it reaches its final value at \(t_0 = 10R_\star / c\); \(\Omega = \text{const}\) afterwards.

Corotational charge density \(\rho_{\text{GJ}} \approx -\Omega \cdot \mathbf{B} / 2\pi c\) defines the characteristic particle density \(n = |\rho_{\text{GJ}}| / e\), plasma frequency \(\omega_p = (4\pi n e^2 / m_e)^{1/2}\), and skin-depth \(\lambda_p = c / \omega_p\). The magnetic field also determines the gyro-frequency of \(e^{\pm}\), \(\omega_B = eB / m_e c\), and ions, \(\omega_{B,i} = eB / m_i c\). In a dipole magnetic field the characteristic frequencies are related by \(\omega_p^2 = 2\omega_B^2\Omega\) and satisfy \(\Omega \ll \omega_p \ll \omega_B\).

The particle Larmor radius satisfies \(r_L \ll r\) at \(r \ll R_{\text{LC}}\), so particles move nearly along \(\mathbf{B}\). At the light cylinder, \(r_L \sim (\gamma / \gamma_0)R_{\text{LC}}\) may become comparable to \(R_{\text{LC}}\).

The characteristic \(\lambda_p\) at the polar cap is related to particle acceleration, as \(\gamma_0 \approx (1/4)(R_\star / R_{\text{LC}})(R_\star / \lambda_p)^2\).

Typical pulsars have \(R_\star / \lambda_p \approx 10^6 \gg 1\) and \(R_{\text{LC}} / R_\star \approx 10^2 - 10^3\). We scale down the big numbers, preserving the hierarchy of scales. The scale \(\lambda_p\) must be well resolved in the simulation, and the number of particles per grid cell must be large. This can only be achieved by increasing \(\lambda_p / R_\star\).

The simulations presented below have \(R_\star / \lambda_p \approx 100 - 130\), \(R_{\text{LC}} / R_\star = 6 - 10\), and \(\gamma_0 = 425\). We also reduced the ion mass, \(m_i = 5m_e\), and assumed ion charge number \(Z = 1\).
We use spherical coordinates $r, \theta, \phi$, and our grid is uniformly spaced in $\log r$ and $\theta$ to allow better resolution near the star, where it is most needed. The grid size is $512 \times 512$.

Three field components are continuous at the star surface which defines the boundary conditions: $B_r = 2\mu \cos \theta / R^3_\star$, $E_\theta = -\mu \Omega \sin 2\theta / R^2_\star$, and $E_\phi = 0$. The dipole configuration of the magnetosphere is set by the surface $B_r$, and its rotation is communicated from the star through the surface $E_\theta$. The star is also a source of electrons and ions for the magnetosphere. We make both particle species available below the surface and their extraction is self-consistently controlled by the local electric field. We neglect the work function, so that particles are easily lifted from the star.

The outer boundary is set at $R_{\text{out}} = 30R_\star \gg R_{\text{LC}}$. Here we place a “damping layer” of thickness $\Delta R = R_{\text{out}}/15$. The layer has resistivity and damps electromagnetic fields on a timescale $\sim \Delta R/c$. Particles escape freely; they are decoupled from the fields once they enter the damping layer. With this implementation, the boundary effectively absorbs waves and particles, which is equivalent to their free escape.

Once an electron or positron reaches the threshold energy $\gamma_{\text{thr}} m_e c^2$ it begins to emit curvature photons capable of pair creation. It depends on the curvature radius of the particle trajectory $R_c$ as $\gamma_{\text{thr}} = K(R_c/R_\star)^{1/3}$. In real pulsars $K \gtrsim 10^6$; we scale it down to $K = 20$ to allow copious pair creation in our numerical experiment. The photon emission rate is $\dot{N} = 0.25c(\gamma/R_c)$, where $\gamma$ is the particle Lorentz factor. The photon emission, propagation, and conversion are traced using Monte-Carlo technique. The free paths of photons $l$ have a distribution $P(l)$ with mean value $\bar{l}$ and dispersion $\Delta l$. The extreme case of $\bar{l} = 0$ is only relevant for discharge near magnetars [Beloborodov & Thompson 2007]; for ordinary pulsars, the delay $l/c$ should be included. In our simulations $\bar{l} = \Delta l = 2R_\star$ for photon-photon collisions (operating in rotators of type I) and $\bar{l} = \Delta l = 0.2R_\star$ for magnetic conversion (enabled at $r \lesssim 3R_\star$). The emitted photons have energies $E_{\text{ph}} \ll \gamma_{\text{thr}} m_e c^2$, and hence the secondary pairs are created with Lorentz factors $\gamma_s \ll \gamma_{\text{thr}}$. The condition $\gamma_s \ll \gamma_{\text{thr}} \ll \gamma_0$ ensures sufficient pair supply in the magnetosphere, and is satisfied in
CHAPTER 4. GLOBAL PULSAR MAGNETOSPHERE

Figure 4.1: Magnetosphere of type I aligned rotator (poloidal cross section) at \( t = 100 \). Vertical dashed line shows the light cylinder. Green curves show the magnetic flux surfaces. (a) Radial component of electric current density \( J_r \). (b) Net charge density \( \rho \). (c) Toroidal component of the magnetic field \( B_\phi \).

our simulations. Radiation reaction (energy loss due to gamma-ray emission) is explicitly included in the particle dynamics.

Hereafter distance is measured in \( R_\ast \), time in \( R_\ast/c \), energy in \( m_e c^2 \), magnetic and electric fields in \( m_e c^2/eR_\ast \), and charge density in \( m_e c^2/4\pi eR_\ast^2 \).

4.3 Rotators of type I

Figures 4.1–4.3 show the magnetosphere of the aligned rotator with \( R_{LC} = 6 \) and \( \mu = 1.5 \times 10^4 \) after 2.6 rotation periods. The energy density is almost everywhere dominated by the electromagnetic field, and the discharge finds a way to adjust and supply the charge density and electric currents demanded by the field. As a result, the magnetosphere shares several key features with the FF solution. Electric currents and Poynting flux flow through the light cylinder along the open magnetic field lines while the interior of the closed field-line zone has \( J = 0 \) and \( B_\phi = 0 \). The Y-point is observed near \( R_{LC} \).

There are two distinct regions of negative and positive radial current density \( J_r \). The negative
current flows in the polar region around the magnetic axis. The positive current is concentrated in a current sheet supporting the jump of $B_\phi$ between the closed and open zones. Outside the light cylinder, the current sheet extends along the equatorial plane to support the flip of $B_\phi$ and $B_r$ across the equatorial plane.

Charge density $\rho = \nabla \cdot \mathbf{E}/4\pi$ also conforms to the expectations from the FF model (cf. Figure 16 in Parfrey et al., 2012). In particular, the current sheet is positively charged outside the Y-point and negatively charged inside the Y-point (see Timokhin 2006 for discussion). $\rho$ significantly deviates from the FF model in the neutral black region with $J_r = 0$; if the rotator approaches the “death line” for pair creation, $\gamma_{\text{thr}} \sim \gamma_0$, this region grows and occupies most of the magnetosphere. A similar neutral region was described by Yuki & Shibata (2012).

The two opposite currents are sustained by different mechanisms. The negative current in the polar region is carried by electrons lifted from the polar cap. There is no significant activity in this region; the particle acceleration is weak and pair creation does not occur. The absence of polar-cap activity is explained by the low positive value of $\alpha \equiv J_\parallel/c\rho GJ \sim 0.7 < 1$. It leads to easy screening of $E_\parallel$ by the charge-separated flow extracted from the star and the flow Lorentz factor comparable to $2\alpha/(1 - \alpha^2)$ (Beloborodov, 2008; Chen & Beloborodov, 2013). We observed
Figure 4.3: (a) Average ion energy in units of $m_e c^2$. (b) Ratio of total matter energy density $U_m$ to magnetic energy density $U_B = B^2/8\pi$.

The same behavior in the simulation of anti-aligned rotator where currents switch sign and the polar current is carried by ions extracted from the star.

The opposite current (the current sheet) is sustained by $e^\pm$ discharge at $r < R_{LC}$. It cannot be conducted by particles lifted from the star as its sign is opposite to that of the charge density demanded by the magnetosphere. Note also that $|\rho| \gg |\rho_{GJ}|$ in the current sheet, so $\rho_{GJ}$ is not important. The accelerating potential drop is $\Phi_\parallel \sim 2\pi \rho_\delta^2 \sim -(\delta/r)\Phi_0$ where $\delta$ is the sheet thickness and we used $2\pi r_\delta|\rho|c \sim I_{GJ} = c\Phi_0$. Pair creation is biased to the outer side of the sheet (a result of its curvature and the finite free path of photons), therefore the unscreened $\Phi_\parallel$ is largest on the inner side. The sheet thickness $\delta$ is set by the Larmor radius of particles near the Y-point.

Plasma outflows along the equatorial plane outside $R_{LC}$ and the Y-point resembles a nozzle formed by the open magnetic fluxes of opposite polarity. Two plasma streams come to the Y-point along the boundary of the closed zone and exchange their opposite $\theta$-momenta. Their collimation is achieved through gyration in the (predominantly toroidal) magnetic field, which communicates the $\theta$-momentum from one stream to the other. As a result the streams flow out in the direction
of their net momentum, which is radial (see also Shibata, 1985).

About 10% of current is carried by the ions extracted from the star at the footpoints of the current sheet. Ions experience no radiative losses and tap the full $\Phi_\parallel$ (Figure 4.3a). They have the largest Larmor radius $r_L$, so the ion streams show large oscillations around the equatorial plane (Figure 4.2c). The streams with smaller oscillations are formed by accelerated positrons with $\gamma$ limited by radiation reaction. Secondary particles have even smaller energies; they outflow almost exactly in the equatorial plane. The streams with different $r_L$ contribute to the thickening of the equatorial current sheet as seen in Figures 4.1-4.3.

Since ions do not create pairs, the discharge in the current sheet relies on the accelerated $e^\pm$. This requires continual recycling of created particles as seeds for new rounds of pair creation, which leads to voltage oscillations. The oscillations occur on the timescale $\sim R_{\text{LC}}/c = \Omega^{-1}$ and make the magnetosphere “breath” around $R_{\text{LC}}$.

There is a steep potential drop across the outer closed field lines toward the Y-point. The strong $E_r$ helps eject particles into the equatorial current sheet. In this region $E \approx B$ and the particle ejection across $B$ is assisted by the drop of $B$ near the Y-point on a scale $\sim r_L$.

The magnetic field dominates energy density everywhere except the Y-point region and the matter-dominated equatorial outflow (Figure 4.3b). This behavior is also visible in the angular distributions of the Poynting luminosity $L_P$ and matter kinetic power $L_m$ (Figure 4.4a). The integrated luminosities at $r = 2R_{\text{LC}}$ are $L_P \approx 0.7L_0$ and $L_m \approx 0.15L_0$ where $L_0 = \mu^2\Omega^4/c^3$. Both contribute to the energy flux from the rotator. This should be compared with the spindown power extracted from the star, $L_{\text{sd}} = L_P(R_*) \approx 0.88L_0$. The difference between $L_{\text{sd}}$ and $L_P + L_m$ is carried by particles bombarding the star (the backflow power is $\sim 0.1L_m$) and the gamma-rays.

At $r \gg R_{\text{LC}}$ we observe magnetic reconnection which strongly heats particles in the equatorial outflow, forms large plasmoids, kinks and wiggles. They are advected outward and do not affect the current structure near the Y-point.

The current sheet is the only source of gamma-rays (Figure 4.4b). The emission is strongly
anisotropic, peaking at $\pm \sim 12^\circ$ around the equatorial plane. In addition, there is a strong peak at the equator from the high-energy particles gyrating in the equatorial outflow.

The simulation for the anti-aligned rotator shows a similar magnetosphere but somewhat different discharge. The current sheet extracts and accelerates electrons from the star (instead of ions), which helps produce pairs. This leads to a more stable $\Phi_\parallel$ and reduces the ”breathing” of the magnetosphere, so the Y-point is nearly static.

We also performed runs for the aligned and anti-aligned rotators with different rotation rates, magnetic fields, and $\gamma_{\text{thr}}$. The exact position of the Y-point $R_Y$ depends on these parameters. With extremely low $\gamma_{\text{thr}}$ and copious supply of plasma $R_Y$ decreases. If $\gamma_{\text{thr}}/\gamma_0$ is increased to $\sim 1$, the magnetosphere transitions to the electrosphere state.

### 4.4 Rotators of type II

The simulation of type II rotators is the same as for type I except for three changes: we suppressed pair creation where $B < 400$, which roughly corresponds to $r \gtrsim 4$, increased $R_{\text{LC}}$ to 10 to better separate it from the pair creation zone, and increased $\mu$ to $2.5 \times 10^4$ to keep $\Phi_0$ unchanged.

As we spin up the star, there is an initial burst of pair creation due to the vacuum initial...
condition and the induced $E_\parallel$ accelerating charges extracted from the surface. The system is able to form an almost FF configuration for a short time, then $E_\parallel$ gets screened inside the pair-creation zone, and the magnetosphere relaxes to a different state. After about 2 rotations of the star, $\rho \approx \rho_{GJ}$ is sustained only at $r \lesssim 4$, and outside this zone the magnetosphere is close to the electrosphere solution (Figure 4.5). It stays in this state till the end of the simulation (about 8 rotations), with no pair creation.

In the final state, the magnetic field is everywhere similar to the dipole. The plasma forms the negative “dome” and positive “torus” with a vacuum gap between them at $r \gtrsim 4$. The gap is outside the pair-creation zone and finds no way for plasma supply. The unscreened $E_\parallel$ in the gap creates a large potential drop along the magnetic field lines, which leads to faster rotation of the magnetosphere in the equatorial region at $r \sim 5 - 8$ (cf. Wada & Shibata [2011]).

The magnetosphere is not completely dead at the end of the simulation. There is a weak negative current flowing out in the polar cap region, and a weak positive current leaking out from the tip of the “torus” region due to a strong $E_r \approx B$. At later times we expect the magnetosphere to evolve even closer to the electrosphere solution. The gap will tend to expand toward the null
point \( \rho_{\text{GJ}} = 0 \) on the star surface, into the zone where pair creation is possible. Then the discharge should reignite and prevent the inward expansion of the gap. The duration of our simulation is too short to see this late evolution; however, it is sufficiently long to show that type II rotators have a low level of activity most of the time.

Breaking the axisymmetry should initiate the diocotron instability \cite{PhilippovSpitkovsky2014}; however it is unlikely to prevent the inner zone from filling with plasma and shutting down pair creation. Our simulations suggest that type II rotators cannot relax to the GJ state — it is unstable even with suppressed azimuthal perturbations; allowing the diocotron instability cannot make the GJ state an attractor for the system.

### 4.5 Discussion

Our first conclusion is that significant activity of the axisymmetric pulsar requires pair creation enabled at \( r \sim R_{\text{LC}} \) (called type I in this chapter); this requires a sufficient optical depth to photon-photon collisions. If pair creation is limited to \( r \ll R_{\text{LC}} \) (type II), the return current is choked and the axisymmetric magnetosphere relaxes to the dome+torus state with suppressed electric currents and pair creation. Many observed pulsars are only capable of pair creation at \( r \ll R_{\text{LC}} \); we conclude that their activity and spindown should be a result of the misalignment of \( \Omega \) and \( \mu \). This conclusion supports the arguments of \cite{Michel2004} and disagrees with the models of \cite{GoldreichJulian1969}, \cite{RudermanSutherland1975}, and \cite{Gruzinov2013}.

Type I axisymmetric rotators are active, as long as discharge voltage \( \Phi_{\text{thr}} < \Phi_0 = \mu \Omega^2 / c^2 \). Their spindown power is \( L_{\text{sd}} \approx \mu^2 \Omega^4 / c^3 \) and their magnetic configuration is similar to the FF solution. Our numerical experiment shows, for the first time, how particle acceleration and \( e^\pm \) discharge self-organize to maintain this configuration. The result is quite different from the previously discussed “trio” of gaps: polar-cap gap, slot gap, and outer gap. Neither aligned nor anti-aligned rotators sustain pair creation in the polar cap outflow. Strong particle acceleration and pair creation occur in (and around) the return current sheet stretched along the boundary of...
the closed zone. The acceleration mechanism is different from the slot-gap models, which were
developed for the opposite, polar-cap current. It is also different from the outer gap model where
the null surface $\rho_{GJ} = 0$ plays a key role. We find that the current sheet has $|\rho| \gg |\rho_{GJ}|$, and its $E_\parallel$
is not controlled by the local value of $\rho_{GJ}$.

Our numerical experiment confirms the phenomenological description of the gamma-ray
source as an accelerator stretched along the boundary of the closed zone, which explains the
observed pulse profiles of GeV emission (Dyks & Rudak 2003). The angular distribution of
gamma-ray luminosity is determined by how $E_\parallel$ and $e^\pm$ discharge self-organize to sustain the
magnetospheric configuration; the geometry by itself does not determine this distribution.

Besides producing copious $e^\pm$ pairs, type I rotators eject a significant flux of ions, $\dot{N}_i$. The
anti-aligned rotator ejects $\dot{N}_i \approx I_{GJ}/e$ from the polar cap, with low energies. The aligned rotator in
our simulation ejects $\dot{N}_i \sim 0.1I_{GJ}/e$ along the current sheet, with much higher energies, which
carries $\sim 5\%$ of the spindown power.

It remains to be seen which features of the axisymmetric magnetosphere will hold for inclined
rotators. FF models provide guidance as they show where the current should flow. In contrast to
the axisymmetric case, inclined rotators have $\alpha < 0$ and $\alpha > 1$ in the central region of the polar
cap, which is required to activate $e^\pm$ discharge (Beloborodov 2008; Chen & Beloborodov 2013;
Timokhin & Arons 2013). A current sheet with $|\alpha| \gg 1$ is expected to form along the boundary
of the closed zone and produce gamma-rays similar to the mechanism seen in our simulations.

Our results show that key puzzles in pulsar physics can be solved using first-principle calcu-
lations, opening exciting opportunities for future modeling. This includes the global magnetic
configuration, particle acceleration, pair multiplicity, and broad-band radiation, from curvature
gamma-rays to coherent radio waves.
Chapter 5

Twisted Magnetar Magnetosphere

5.1 Introduction

Magnetars are neutron stars with ultrastrong magnetic fields ($B \gtrsim 10^{14}$ G) that display strong activity fed by dissipation of magnetic energy (see e.g. Mereghetti [2008], Turolla et al. [2015] for reviews). They produce strong outbursts and flares as well as bright persistent emission with a prominent hard X-ray component extending above 100 keV. These activities are associated with strong deformations of the external magnetosphere of the neutron star, resembling the activity of the solar corona (e.g. Thompson & Duncan [1995]). The magnetosphere is anchored in the solid crust of the star and its deformation is caused by crustal shear motions driven by ultrastrong internal magnetic stresses.

The speed of the surface motions is poorly known. Recent work suggests that the crust yields to internal stresses through an instability launching a thermoplastic wave (Beloborodov & Levin [2014]) or a Hall-mediated avalanche (Li et al. [2016]). In both cases the motion is plastic and should occur on a timescale much longer than the Alfvén crossing timescale (10 – 100 ms). It is expected to be fast enough to efficiently twist the external magnetosphere.

The surface shear motion launches Alfvén waves along the magnetic field lines and generates magnetospheric twist $\nabla \times \mathbf{B} \neq 0$ (Thompson et al. [2002], Parfrey et al. [2013] hereafter PBH13).
CHAPTER 5. TWISTED MAGNETAR MAGNETOSPHERE

Plasma is required to supply the current $j = (c/4\pi)\nabla \times B$. Beloborodov & Thompson (2007, hereafter BT07) found that plasma must be mainly supplied through $e^\pm$ discharge in the magnetosphere rather than through extraction of charges from the star. They performed simplified one-dimensional (1D) simulations of the discharge. In the simulations, the magnetosphere was replaced by a fixed, uniform field $B(x)$ connecting anode and cathode — metallic plates at $x_A$ and $x_C$. The fixed $\nabla \times B$ in this setup turns out to be equivalent to imposing an electric current through the plates into the computational box. When the pair creation was not allowed, the system quickly relaxed to a global “double layer” configuration, with surface charges of the opposite sign induced on the plates. The electric field between them gave a huge voltage $\Phi_e$ accelerating particles to ultra-high energies. When pair creation process was included in the simulation, the voltage dropped to a much lower value, just sufficient to sustain pair creation, and the current was supported through continual $e^\pm$ discharge. BT07 concluded that pair creation must be responsible for screening electric fields and regulating the magnetospheric activity of magnetars.

The simplified 1D model cannot, however, give a compete picture of the magnetospheric activity, for a few reasons. It does not show how $\nabla \times B$ is imparted in the first place, as the 1D model does not support Alfvén waves. The exclusion of this important degree of freedom may also put in question the double layer formation in the absence of pair creation, the necessity of the onset of pair creation, and the self-regulation of the discharge voltage seen in the 1D model. Note also that the electric field in the 1D (slab) geometry does not decrease with distance from the charge, and hence one cannot see a realistic distribution of the accelerating electric field along the magnetospheric field lines. Finally, the 1D model offers no way to follow the gradual resistive “untwisting” of the magnetosphere — its global evolution as a result of ohmic dissipation of the twist energy. The expected evolution must occur on the resistive timescale of months to years (regulated by voltage $\Phi_e$) and can be tested against observations.

An axisymmetric electrodynamic model of a resistively untwisting magnetosphere was developed by Beloborodov (2009, hereafter B09). This model assumed that a given fixed voltage $\Phi_e$ is
sustained on current-carrying field lines, without calculating the discharge that regulates $\Phi_e$. A surprising result was the formation of two distinct regions in the untwisting magnetosphere, with a sharp boundary between them — a “cavity” ($j = 0$) and a “j-bundle.” In essence, the untwisting process was found to be the growth of the cavity, erasing the currents in the j-bundle. A curious immediate implication was the prediction of shrinking hot spots on magnetars — the footprints of the shrinking j-bundle, where the stellar surface is heated by bombardment of accelerated particles. Shrinking hot spots have been observed in seven objects by now (see data compilation in Beloborodov & Li, 2016). All of these objects belong to the class of “transient magnetars” that show a sudden outburst and then gradually decay back to the quiescent state of low luminosity. A key parameter governing the j-bundle evolution is its poorly known voltage $\Phi_e$, which depends on how the $e^\pm$ discharge is self-organized and may be different on different magnetospheric field lines.

The goal of the present chapter is to overcome the limitations of the 1D discharge model and perform the first self-consistent calculation of the $e^\pm$ discharge in an axisymmetric twisted magnetosphere. The process can be simulated from first-principles using a full kinetic description of the magnetospheric plasma as a large number of charged particles moving in the self-consistent collective electromagnetic field. Such a direct numerical experiment will show how the twist and the electric current are created in the magnetosphere in response to crustal shear, and will follow the ensuing dissipative evolution of the twist.

The self-organization of the $e^\pm$ discharge should determine where the particles are created and accelerated. Should this occur near the footpoints of the magnetospheric field line or near its apex? Will the acceleration region be steady or move around? Answers to these questions may have important implications for nonthermal emission from magnetars. The voltage drop along the twisted field lines will control the dissipated power which feeds the observed emission. We expect to see how particles are accelerated in the current-carrying magnetic loop and rain down on the stellar surface to create hotspots. Finally, the established discharge voltage will determine
the life-time of the magnetic twist and the pattern of its evolution.

A suitable technique for such direct numerical experiments is the particle-in-cell (PIC) method, with pair creation implemented. This method has been successfully applied to the old problem of rotation-powered pulsars (Chen & Beloborodov 2014 hereafter CB14; Philippov et al. 2015; Belyaev 2014; Cerutti et al. 2015). The magnetar problem is different in important ways and in some ways easier to study using a global PIC simulation, as will be described below.

The chapter is organized as follows. In Section 5.2 we describe the theory of twisted magnetospheres in axisymmetric geometry, revisit the double-layer configuration (in the absence of pair creation), describe the mechanism of pair creation and basic electrodynamics of untwisting. This will be useful for understanding the simulation results and also introduces notation used in the chapter. Section 5.3 presents the setup of our numerical experiments. Section 5.4 describes the results and their implications. Finally in Section 5.5 we summarize our conclusions and provide an outlook for future studies.

5.2 Sustaining Currents in the Twisted Magnetosphere

Let us consider a dipole magnetic field around the star, and assume that its footpoints on the star are sheared in the azimuthal direction about the magnetic axis. In this case, the implanted twist is axisymmetric and its amplitude \( \psi \) is simply given by the azimuthal angle between the two footpoints of the magnetic field line. It is convenient to use spherical coordinates \( r, \theta, \phi \) with the polar axis being the axis of symmetry. The magnetospheric twist implies a toroidal component of the magnetic field \( B_\phi \neq 0 \), and the twist amplitude on a given magnetic field line is related to \( B_\phi \) by

\[
\psi = \int_{p}^{q} \frac{B_\phi}{Br \sin \theta} d\ell, \tag{5.1}
\]

where the integral is taken along the field line, and \( p, q \) are the two footpoints where the field line is anchored to the surface. As long as the implanted twist \( \psi \) is smaller than unity, the poloidal
magnetic field remains close to dipolar, and the deformation can be thought of as simply adding a toroidal component $B_\phi$ without changing the poloidal dipole component (B09). This induces $\nabla \times \mathbf{B}$ in the dipolar configuration that was originally curl-free. It must be sustained by an electric current in the magnetosphere, $\mathbf{j}$. The magnetic energy strongly dominates over the plasma energy, and hence the currents must be nearly force-free, $\mathbf{j} \times \mathbf{B} = 0$, i.e. flowing along the magnetic field lines.

The origin of the plasma that could carry the current is a non-trivial issue. The star can have a gaseous atmosphere, however for the typical surface temperature $kT < 1 \text{ keV}$ the atmosphere scale-height is tiny (centimeters), because of the strong gravity of the neutron star. The atmosphere does not provide enough plasma to conduct currents at large altitudes $r \sim R_\star$, where $R_\star = 10^{-13} \text{ km}$ is the neutron star radius.

Spinning of the neutron star and its magnetosphere with velocity $\mathbf{v}_{\text{rot}} = \mathbf{\Omega} \times \mathbf{r}$ implies a “co-rotation” electric field $\mathbf{E} = -\mathbf{v}_{\text{rot}} \times \mathbf{B}/c$ and requires charge density $\rho_{\text{GJ}} = \nabla \cdot \mathbf{E}/4\pi = -\mathbf{\Omega} \cdot \mathbf{B}/2\pi c$ (Goldreich & Julian [1969]). Magnetars are slow rotators, $\Omega \sim 1 \text{ Hz}$, and their $\rho_{\text{GJ}}$ is small. The currents demanded by the twisted magnetosphere are typically much stronger than $c\rho_{\text{GJ}}$.

The magnetosphere must make a special effort to avoid charge starvation and create sufficiently dense plasma to conduct the current $\mathbf{j}$ demanded by the twist. It achieves this by inducing an electric field $E_||$ (parallel to the magnetic field lines) that can accelerate particles and trigger pair creation. This implies a finite voltage in the magnetospheric electric circuit and a finite rate of ohmic dissipation.

### 5.2.1 Voltage without pair discharge

In the absence of pair creation, the star is the only available source of magnetospheric plasma. The lack of charges leads to induction of an electric field with a component parallel to the magnetic field, which can pull out charges from the star and accelerate them. Then the electric circuit is expected to relax to a static configuration similar to the relativistic double layer derived by
Carlqvist (1982) and observed in the 1D plasma simulations of BT07. It sustains the opposite surface charges at the two footpoints of the magnetic loop where the lifted particles still move slowly, \( v \ll c \), and create a large charge density \( \rho \sim j/\nu \).

The high charge density near the footpoints generates \( E_\parallel \) according to the Gauss law, and \( E_\parallel \) accelerates the flow on the plasma timescale \( \omega_p^{-1} = (m_e/4\pi e\rho)^{1/2} \). The flow density \( \rho \) is reduced to its minimum where its velocity approaches \( c \). As a result, the characteristic thickness of the surface charge layer is the plasma skin depth \( \lambda_p = c/\omega_p \) evaluated for the plasma density \( \rho \sim j/c \).

The surface charge \( \Sigma \sim (j/c)\lambda_p \) generates the self-consistent electric field that lifts and accelerates particles from the footpoint,

\[
E_\parallel \sim 4\pi \rho \lambda_p \sim \frac{4\pi j}{\omega_p},
\]

(5.2)

where \( \omega_p \) is the plasma frequency defined by

\[
\omega_p^2 = \frac{4\pi e\rho}{m_e}, \quad \rho = \frac{j}{c}.
\]

(5.3)

In other words, the surface charge near the anode and cathode is organized so that particles extracted from the star are accelerated to \( v \sim c \) over a length comparable to the plasma skin depth.

For simplicity, consider a symmetric double layer where the positive and negative charges have the same mass. In the 1D model, the electric field is almost constant between the two surface charges of the double layer, giving a voltage drop,

\[
\frac{e\Phi_e}{m_e c^2} = \frac{4\pi j e L}{\omega_p m_e c^2} = \frac{\omega_p L}{c} = \frac{L}{\lambda_p},
\]

(5.4)

where \( L \) is the size of the layer (the distance between the footpoints). Using \( j \sim \psi B/L \), one finds
for the typical parameters of a magnetar,

$$\frac{L}{\lambda_p} \sim 10^8 L^{3/2} \psi^{-1/2} B_{15}^{-1/2}, \quad (5.5)$$

which implies a huge voltage $\Phi_e$.

The estimate in Equation (6.3) is not valid, however, for a realistic magnetosphere, which is not one-dimensional. The current flows along the curved magnetic field lines and their dipolar geometry significantly changes the distribution of the net voltage sustained between the two footpoints.

The corrected voltage may be estimated as follows. Since $\lambda_p$ is small compared with the thickness of the j-bundle, the surface charge remains thin and its structure is not changed from the 1D model. The self-consistent electric field extracting charges from the footpoint is still described by Equation (5.2). However, with increasing altitude the electric field must be reduced on a scale comparable to the horizontal size of surface charge $W$ (thickness of the j-bundle). The resulting potential drop saturates at $\Phi_e \sim E_\parallel W$, which gives

$$\gamma_{DL} = \frac{e\Phi_e}{m_e c^2} \sim \frac{W}{\lambda_p}. \quad (5.6)$$

It is smaller than the 1D estimate by the factor of $W/L$. For instance a j-bundle of thickness $W \sim 0.1 R_\ast$ at the stellar surface and length $L \sim 10 R_\ast$ would sustain a voltage $\sim 10^{-2}$ smaller than predicted by the 1D model. This is still a huge voltage and particles that tap the full potential drop will be able to induce pair discharge, making the double layer model inconsistent.

One should also note that $E_\parallel = \mathbf{E} \cdot \mathbf{B}/B$, and hence the voltage,

$$\Phi_e = \int_p^q E_\parallel d\ell, \quad (5.7)$$

has a pure inductive origin. One should think of $E_\parallel$ as $c^{-1} \partial A_\parallel / \partial t$, the result of the slow decay
of the ultrastrong twisted magnetic field (BT07). $e\Phi_e$ measures the energy gain of charge $e$ completing the electric circuit, and this released energy is extracted from the magnetic twist energy. A potential electric field would be unable to support any significant voltage between the footpoints, as they are connected by an excellent conductor — the crust.

The induction electric field $\mathbf{E}$ still satisfies the Gauss law $\nabla \cdot \mathbf{E} = 4\pi \rho$; as long as the untwisting process occurs much slower than the light crossing of the system, one can think of the dissipation as a quasi-steady process. The inductive double layer is similar to a normal electrostatic double layer except that the integral of $\mathbf{E}$ along the full closed circuit (including the part closing through the crust, where $\mathbf{E} = 0$) does not vanish and instead equals $\Phi_e$. There is no external emf applied to the circuit below the stellar surface; the only emf sustaining the current is the induction emf due to the twist decay in the magnetosphere itself.

### 5.2.2 Voltage with pair discharge

The mechanism of secondary $e^\pm$ creation by relativistic particles in the magnetar magnetosphere involves an intermediate step of gamma-ray production. It occurs through resonant Compton scattering of photons flowing from the star by particles accelerated in the magnetosphere. A target photon with energy $E_t \sim 1$ keV can be resonantly scattered by an electron with Lorentz factor $\gamma$ if the photon energy measured in the electron rest frame matches $\hbar \omega_B$, where $\omega_B = eB/m_e c$.[1] The resonance condition reads

$$\gamma(1 - \beta \cos \theta_X)E_t = \hbar \omega_B, \quad (5.8)$$

where $\theta_X$ is the angle of the target X-ray with respect to local magnetic field line (the electron moves along the field line). The energy $\hbar \omega_B$ equals $m_e c^2$ for the characteristic magnetic field $B_Q = m_e^2 c^3 / e\hbar \approx 4.4 \times 10^{13}$ G and scales linearly with $B$.

---

[1] This simple resonance condition remains valid in ultrastrong fields $B \gg B_Q$ when one takes into account the electron recoil in scattering and the fact that the target photon is propagating almost parallel to $\mathbf{B}$ when viewed in the electron rest frame, because of the relativistic aberration effect (BT07).
Magnetars supply plenty of keV photons, and the electron Lorentz factor required for resonant scattering at $B \sim B_Q$ is moderate, $\gamma \sim 10^3$. It is far below the electron Lorentz factors that would be reached in the double layer discussed in the previous section.

After the scattering, the photon energy is boosted by a factor comparable to $\gamma^2$, putting the originally keV photon into the GeV range, $E_\gamma \sim 1$ GeV. Such energetic gamma-rays can easily convert to $e^\pm$ pairs in the strong magnetic field, as soon as the gamma-ray pitch angle with respect to the magnetic field, $\theta_\gamma$, is large enough to satisfy the threshold condition,

$$E_\gamma \sin \theta_\gamma > 2m_e c^2. \quad (5.9)$$

In the region near the star where $B > 10^{13}$ G the conversion occurs practically immediately following resonance scattering (Beloborodov, 2013).

The efficiency of pair creation implies a quick development of electric discharge until the number of created particles becomes sufficient to screen the accelerating electric field. The process develops in a runaway (exponential) manner and hence the accelerating voltage is unlikely to grow beyond a characteristic value that makes particles capable of resonant scattering. This condition defines a “threshold” for discharge, which corresponds to a characteristic electron Lorentz factor $\gamma_{\text{thr}}$.

### 5.2.3 Characteristic timescales and energy scales

The shortest timescale of interest is the plasma scale $\omega_p^{-1}$. It describes the growth rate of the local accelerating electric field in response to charge starvation (BT07). It also determines the thickness of the surface charge $c/\omega_p$ in the double-layer configuration.

The characteristic dynamic timescale of the electric circuit is the light crossing time or the Alfvén crossing time of the system,

$$t_A = \frac{L}{c} \sim 0.3 L_7 \text{ ms}, \quad (5.10)$$
where $L$ is the length of the magnetospheric field line. The group speed of Alfvén waves is always directed along the magnetic field lines and its value is close to $c$ in the magnetically dominated corona.

The longest timescale in the problem is the lifetime of the magnetic twist. The finite voltage sustaining the magnetospheric current implies a finite ohmic dissipation rate, so the magnetic twist energy $E_{\text{twist}}$ must dissipate with time,

$$\frac{dE_{\text{twist}}}{dt} \approx \frac{d}{dt} \int \frac{B^2}{8\pi} dV \sim I \Phi_e,$$

(5.11)

where $I$ is electric current flowing through the magnetosphere. The voltage $\Phi_e$ controls the timescale of this evolution,

$$t_{\text{ohm}} \sim \frac{E_{\text{twist}}}{I \Phi_e}.$$  (5.12)

Using the characteristic $I \lesssim \psi (c/4\pi) BR_*$ and $\gamma_{\text{thr}} \sim 10^3$ one can estimate that $t_{\text{ohm}}$ is comparable to one year. This theoretical timescale for untwisting is comparable to the observed decay timescale in transient magnetars following an outburst of activity.

Because of the vast separation of timescales, $t_{\text{ohm}} \gg t_A$, the ohmic dissipation of the magnetospheric twist can be viewed as a quasi-steady process slowly draining the twist energy. Unsteadiness of the discharge may lead to strong variability in the electric circuit, however it occurs on very short timescales, which would be hard to resolve observationally.

The characteristic scales for energy (or electron Lorentz factor $\gamma$) also have an important hierarchy. The highest energy corresponds to $\gamma_{\text{DL}}$, which would only be achieved in the absence of pair creation. It is given by Equation (5.6) and can exceed $10^6$. The next characteristic $\gamma$ is determined by the threshold for $e^\pm$ discharge $\gamma_{\text{thr}}$, which is comparable to $10^3$. Both $\gamma_{\text{DL}}$ and $\gamma_{\text{thr}}$ are much greater than unity.
CHAPTER 5. TWISTED MAGNETAR MAGNETOSPHERE

5.2.4 Mechanism of untwisting

An integral form of the Faraday’s induction law $\partial \mathbf{B}/\partial t = -c \nabla \times \mathbf{E}$ leads to a simple equation describing resistive evolution of the axisymmetric twist (Beloborodov 2011),

$$\dot{\psi} = 2\pi c \frac{\partial \Phi_e}{\partial f}.$$ (5.13)

Here $f(r, \theta)$ is the poloidal magnetic flux function (constant along a magnetic flux surface), which serves to label the magnetic field lines. For any given point $(r, \theta)$, $f$ is defined as the magnetic flux through the circle about the axis of symmetry passing through the point; $f = 0$ on the axis of symmetry. In particular, for a dipole poloidal field with a dipole moment $\mu$ the flux function is given by

$$f = \frac{2\pi \mu \sin^2 \theta}{r}, \quad 0 \leq f \leq f_{\text{max}} = \frac{2\pi \mu}{R_*}.$$ (5.14)

Note that $\sin^2 \theta/r = \text{const}$ along a dipole field line. It is convenient to use the dimensionless flux function

$$u \equiv \frac{f}{f_{\text{max}}} = \sin^2 \theta_*,$$ (5.15)

where $\theta_*$ is the polar angle of the magnetic field line footprint on the stellar surface.

Equation (5.13) shows that the twist must decrease where $\partial \Phi_e/\partial f < 0$ and increase where $\partial \Phi_e/\partial f > 0$. The fact that $\Phi_e(f_{\text{max}}) = 0$ (the field line $f_{\text{max}}$ is confined to the star, which we approximate as an ideal conductor) implies $\partial \Phi_e/\partial f < 0$ at some $f < f_{\text{max}}$. This region with large $f$, comparable to $f_{\text{max}}$, corresponds to the inner magnetosphere near the equator, with short field lines. B09 showed that this fact leads to immediate formation of a “cavity” with $j = 0$ in the equatorial region near the star, and the cavity expands on the timescale $t_{\text{ohm}}$, erasing the magnetospheric currents. The currents are “sucked” into the star, so that they close inside the conductor.

From the untwisting equation it is evident that the profile of $\Phi_e(f)$ plays the key role for the twist evolution. Voltage regulated by pair discharge is expected to satisfy the condition
\( e \Phi_e \sim \gamma_{\text{thr}} m_e c^2 \). Its variation with \( f \) over a region \( \Delta f = f_{\text{max}} \Delta u \) gives the characteristic twist evolution timescale,

\[
t_{\text{ohm}} = \frac{\dot{\psi}}{\dot{\psi}} \sim \frac{\mu}{c \Phi_{\text{thr}} R_*} \psi \Delta u.
\]

(5.16)

The dimensionless quantities \( \Delta u \) and \( \psi \) are comparable to unity, and the characteristic timescale is set by the ratio \( \mu/\Phi_{\text{thr}} \). Note however that \( t_{\text{ohm}} \) can strongly differ for different magnetic field lines. In particular, if there is a region with a flat dependence of \( \Phi_e(f) \), \( \partial \Phi_e/\partial f = 0 \), then the local \( t_{\text{ohm}} = \infty \) and the twist angle \( \psi \) is “frozen”, waiting for the cavity expansion to reach the region (B09).

Another interesting implication of Equation (5.13) is that on some field lines the twist may grow as the magnetosphere untwists. In particular, a decrease of \( \Phi_e \) toward the magnetic axis, \( \partial \Phi_e/\partial f > 0 \), leads to \( \dot{\psi} > 0 \). This effect will be observed in the simulations below. Together with the cavity expansion, this means that the twist relocates toward the axis with a decreasing energy \( E_{\text{twist}} \) but possibly with increasing amplitude \( \psi \) in some regions before being completely dissipated.

### 5.3 Setup of the simulation

#### 5.3.1 Implanting the twist

Our simulation starts with a pure dipole magnetosphere, with a magnetic moment \( \mu \) and no magnetic twist, \( B_\phi = 0 \). The twist is gradually implanted by shearing the stellar surface with a latitude-dependent angular velocity \( \omega(\theta) \parallel \mu \). The profile of \( \omega(\theta) \) determines the profile of the implanted twist; we choose a profile similar to previous magnetohydrodynamic (MHD) and force-free electrodynamic (FFE) simulations of twisted magnetospheres [Mikic & Linker 1994; PBH13],

\[
\omega(\theta, t) = \omega_0(t) \frac{\Theta}{\sin \theta} \exp \left[ (1 - \Theta^4)/4 \right],
\]

(5.17)
where $\Theta = (\theta - \pi/2)/\Delta \theta_m$ and $\Delta \theta_m = \pi/4$ is a measure of the width of the sheared region. This profile gives a smooth twist that is centered at $\theta = \pi/4$ and decreases to zero at the equator. The prefactor $\omega_0(t)$ describes the rate of implanting the twist. It is smoothly increased from zero at $t = 0$ to a chosen maximum value, kept at this value for some time, and then smoothly switched off back to zero.

As long as the duration $t_{\text{shear}}$ of the surface shear $\omega \neq 0$ is shorter than the resistive timescale of the magnetosphere, $t_{\text{shear}} \ll t_{\text{ohm}}$, ohmic dissipation may be neglected during time $t_{\text{shear}}$. Then the implanted twist profile is given by

$$\psi(\theta) = \int_0^{t_{\text{shear}}} \omega(\theta, t) \, dt.$$  \hspace{1cm} (5.18)

We choose $t_{\text{shear}} = 40R_*/c$. Then the shearing stage is sufficiently short compared with the total duration of our simulation $t_{\text{sim}} = 350R_*/c$ but longer than or comparable to the Alfvén crossing time $t_A$ of the sheared region, so that twist implanting is a relatively gentle process. The maximum shear angle (near $\theta = \pi/4$) is $\psi_{\text{max}} \approx 1.6$ radian in the simulations presented below.

After the twist implantation is finished, $\omega$ is kept at zero and the boundary condition at the stellar surface becomes simply a perfect static conductor. Magnetars are slow rotators, and their light cylinders $R_{\text{LC}} \gtrsim 10^4 R_*$ are well beyond the twisted, dissipative region. The slow spinning of the star is neglected in the present chapter, which corresponds to $R_{\text{LC}} = \infty$.

The implanted twist $\psi \sim 1$ is moderate and expected to result in moderate inflation of the poloidal magnetic field lines. The main effect of surface shearing is creating a strong $B_\phi$ in the magnetosphere. Analytical arguments (e.g. [Uzdensky, 2002]) and FFE simulations (PBH13) show that a stronger $\psi \gtrsim 3$ will result in a global instability of the magnetosphere, which we do not intend to study in this chapter and defer to future work.
5.3.2 Surface atmospheric layer

We start the simulation with a complete vacuum around the star and create a dense neutral atmospheric layer at the stellar surface by injecting warm electron-ion plasma at \( R_\star \). The atmosphere scale-height \( h \) is determined by the particle injection temperature and gravity of the star. We choose a Maxwellian injection velocity with the mean value \( v_0 \approx 0.1c \) and the gravitational acceleration \( g = g_0 / r^2 \) with \( g_0 = 0.5R_\star c^2 \). This gives the hydrostatic scale-height

\[
h \approx \frac{v_0^2}{2g_0} \approx 0.01R_\star.
\]

This is a much thicker atmospheric layer than the magnetar would have at a surface temperature \( kT \lesssim 1 \text{ keV} \). However, it is sufficiently thin and still resolved by our numerical grid (see below). The characteristic time it takes to form the atmosphere is short, \( t_{\text{atm}} \sim h/v_0 = 0.1R_\star/c \). Throughout the simulation particles are continually injected and absorbed by the star, sustaining a steady atmosphere at \( t \gg t_{\text{atm}} \).

The injection rate is chosen high enough to ensure a high density at the base of the atmosphere,

\[
n_{\text{atm}} \gg \frac{j}{ev_0}.
\]

The density is exponentially reduced with altitude on the scale \( h \), and steeply drops to a low value below \( j/ec \). Therefore, in the absence of \( E_\parallel \) the hydrostatic plasma is not capable of conducting the electric current \( j \) required in the twisted magnetosphere.

Where the atmospheric density \( n(r) \) falls below \( j/ec \), electric field \( E_\parallel \) is expected to develop in response to charge starvation and lift particles from the atmosphere. The thin and dense atmospheric layer merely makes plasma available, with no special injection assumptions at the stellar surface. The numerical experiment must show how the system responds to the surface shear described in Section 5.3.1 and whether the induced \( E_\parallel \) will self-organize to conduct the magnetospheric currents that allow the twist to be implanted.
5.3.3 Creation of $e^\pm$ pairs

If $E_\parallel$ accelerates the lifted electrons to high Lorentz factors $\gamma > \gamma_{\text{thr}}$, pair creation will be ignited. In this chapter, we use the simplest implementation of this process: we choose a fixed value for $\gamma_{\text{thr}}$ and let a new $e^\pm$ pair be instantaneously created every time an electron (or positron) reaches $\gamma_{\text{thr}}$. This may be a reasonable approximation for the $e^\pm$ discharge near the star where $B \gg 10^{13}$ G (Beloborodov, 2013). However, it becomes poor at larger distances where the magnetic field is weak and resonantly scattered photons have lower energies.

An additional simplification in our implementation is the prescription for the energy of the created pair. We will assume that the pair takes a fixed energy $\Delta E$ from the primary particle, and shares it equally, i.e. the new $e^+$ and $e^-$ each receives $\Delta E/2$ (including the rest mass). Total energy and momentum parallel to $B$ is conserved in the pair creation process.

Thus, we do not track the propagation of any high-energy photons, which is significantly simpler than the discharge model of CB14 developed for pulsars. The simplified version appears adequate for the first axisymmetric PIC model of magnetars. It should be sufficient to demonstrate some basic features of plasma self-organization in response to shearing of the magnetospheric footpoints, followed by ohmic dissipation of the twist. The results may be used as a benchmark for future more advanced simulations. Future simulations will have explicitly implemented resonant scattering process, so that $\Delta E$ will be the energy of the resonantly scattered photon, which may convert to $e^\pm$ with a delay. Both $\gamma_{\text{thr}}$ and $\Delta E$ will vary with the local magnetic field, see Beloborodov (2013) for a detailed discussion.

5.3.4 Rescaling of large numbers in the problem

Any PIC simulation must resolve the plasma skin depth $\lambda_p = c/\omega_p$, which is a demanding condition on the computational grid, as $\lambda_p$ is a microscopic scale and the ratio $R_*/\lambda_p$ is huge (comparable to $10^8$ in magnetars). Similar to the PIC simulations of rotation-powered pulsars, this issue is resolved by rescaling the parameters of the problem so that $\lambda_p$ remains much smaller than
the stellar radius, $\lambda_p \sim 10^{-2} R_*$, but becoming sufficiently large to be well resolved. This rescaling has two main implications:

(1) Similar to the pulsar problem, the increased $\lambda_p$ implies a reduction of the energy scales (cf. CB14). In particular, the maximum voltage that can be induced in a magnetar magnetosphere is given by $\gamma_{DL}$ (Equation 5.6), which now becomes moderate, $\gamma_{DL} \sim 10^2$. To respect the hierarchy of the energy scales $1 \ll \gamma_{thr} \ll \gamma_{DL}$, a good choice for the discharge threshold in the numerical experiment is $\gamma_{thr} \sim 10$. Secondary pairs receive the energy $\Delta E$, which must be a fraction of $\gamma_{thr} m_e c^2$. We will fix $\Delta E = 3.5 m_e c^2$ for all simulations presented below.

(2) The rescaling of $\lambda_p$ changes the lifetime of the implanted twist, as seen from the following estimate. The value of $\lambda_p = c / \omega_p$ is related to the electric current density $j$ by Equation (5.3), and the characteristic value of $j$ scales with the magnetic dipole moment of the star $\mu$: $j \sim \psi (c/4\pi) (\mu / R^4_*)$. This gives,

$$\left( \frac{\lambda_p}{R_*} \right)^2 \sim \frac{m_e c^2 R_*^2}{e \mu \psi}.$$  \hspace{1cm} (5.21)

Combining this relation with Equation (5.16) for the resistive evolution timescale, one obtains

$$t_{ohm} \sim \gamma_{thr}^{-1} \left( \frac{R_*}{\lambda_p} \right)^2 \frac{R_*}{c}. \hspace{1cm} (5.22)$$

One can see that the rescaling of $\lambda_p$ to $\sim 10^{-2} R_*$ reduces the resistive timescale to $t_{ohm} \sim 10^3 (R_*/c)$ when $\gamma_{thr} \sim 10$. This is fortunate, as the untwisting evolution can now be observed during a reasonably long simulation. With the realistic $\lambda_p / R_* \sim 10^{-8}$ and $\gamma_{thr} \sim 10^3$ one would have $t_{ohm} \sim 10^{13} R_*/c$.

Another large number that should be rescaled in the simulation is the ion-to-electron mass ratio $m_i / m_e \approx 2 \times 10^3$. We use $m_i / m_e = 10$. This rescaling is useful for two reasons: (1) The characteristic ion plasma frequency $\omega_{p,i} = (4 \pi n_i e^2 / m_i)^{1/2}$ is not very much smaller than $\omega_p$, so that $\omega_{p,i} < r/c$ is well satisfied, and (2) $m_i c^2$ becomes comparable to $\gamma_{thr} m_e c^2$. The latter coincidence is also expected for the real magnetar discharge.
It is also useful to evaluate the surface magnetic field $B_\ast \sim \mu/R_\ast^3$, which can be expressed from Equation (5.21), and then estimate the characteristic gyro-frequency,

$$\omega_B = \frac{e B_\ast}{m_e c} \sim \frac{c}{R_\ast} \left( \frac{R_\ast}{\lambda_p} \right)^2,$$

(5.23)

where $\lambda_p$ corresponds to the current density supporting a twist $\psi \sim 1$. One can see that the particles are very strongly magnetized, $\omega_B \sim 10^4 c/R_\ast$, and hence expected to move along the magnetic field lines, similar to real magnetars. The characteristic gyro-frequency is also related to another important parameter of the magnetosphere — the ratio of magnetic and plasma energy densities,

$$q = \frac{B^2}{4\pi \gamma n m_e c^2} = \frac{\omega_B^2}{\gamma \omega_p^2}.$$  (5.24)

For real parameters of magnetars this ratio is $q \sim 10^{17}$. The characteristic parameters chosen in our simulations give $q \sim 10^3$. This is still very much above unity, so the magnetosphere is nearly force-free as it should be.

The parameter $q$ also determines the Lorentz factor of Alfvén waves, $\gamma_A \approx q^{1/2}$. For a real magnetar, this gives $\gamma_A \gg \gamma \sim \gamma_{\text{thr}}$. This condition is satisfied in our rescaled numerical experiment as long as $\gamma_{\text{thr}} \ll 30$.

### 5.3.5 Evolving the fields and the plasma: Aperture

The particle-in-cell (PIC) method provides an efficient technique to simulate plasma from first principles. The electromagnetic fields are evolved on a grid according to Maxwell equations with the source (electric current and charge density) provided by the plasma that is self-consistently evolved in the electromagnetic field. The plasma is represented directly as a large number of individual particles. The simulation follows the motion of each particle by calculating the applied forces. The motion of the plasma particles creates electric current which is interpolated onto the grid and then used as the source term in the Maxwell equations to update the electromagnetic
field. The method well describes the plasma behavior at the microscopic kinetic level as long as the plasma skin depth is well resolved by the grid and the number of particles per grid cell is much larger than one.

Our simulations are performed using the PIC code Aperture\footnote{Aperture is a recursive acronym: Aperture is a code for Particles, Electromagnetic fields, and Radiative Transfer at Ultra-Relativistic Energies.} The code was originally developed for the PIC simulations of rotationally powered pulsars (CB14). The code can follow pair creation with or without explicit tracking of high-energy photons. In the present work we use the simplified implementation of pair creation (Section 5.3.3) and do not use the radiative transfer module. The code is fully relativistic and designed to work on curvilinear grids. This is particularly important for problems with natural spherical geometry, such as the plasma dynamics around a spherical star in a region extending far beyond the stellar radius.

The simulations presented below are done in 2.5D, which means that our grid is 2D (in the poloidal plane) but all vector quantities are fully 3D, and we solve the full Maxwell equations assuming axisymmetry. Particles in the simulation may be thought of as rings with poloidal and toroidal velocity components. We use a spherical $r, \theta$ grid with logarithmic spacing in $r$ and uniform spacing in $\theta$. For all of the simulations shown in this chapter, the grid size is $384 \times 384$ and the timestep $\Delta t = 10^{-3} R_\star / c$.

The outer boundary of the simulation box is set at $r_{\text{out}} = 30 R_\star$ and employs a damping condition that lets outgoing electromagnetic waves and particles escape the box, preventing reflection. We did not detect any appreciable reflection of waves from the outer boundary. Note also that most of the active (current carrying) field lines are closed well inside the box and do not cross the outer boundary.

The shear motion of the stellar surface during the twist implantation stage $t < t_{\text{shear}} = 40 R_\star / c$ is equivalent to imposing a tangential electric field at the boundary. The field corresponding to the surface motion with velocity $v$ in the lab frame is given by $E = -v \times B / c$. It corresponds to zero electric field in the comoving frame of the stellar crust, which is assumed to be an ideal conductor.
This gives the following boundary condition at \( r = R_\star \),

\[
E(t, \theta) = -\frac{(\omega(t, \theta) \times r) \times B}{c}.
\]  

(5.25)

The initial state is a dipole field and the normal component of the magnetic field at the surface remains unchanged during the simulation.

### 5.3.6 Units

A set of natural units can be defined as follows. All lengths are measured in units the stellar radius \( R_\star \) and time is measured in \( R_\star / c \). The corresponding velocity unit is the speed of light \( c \). We define the dimensionless electromagnetic field and current density as

\[
\tilde{E} = \frac{eR_\star E}{m_e c^2}, \quad \tilde{B} = \frac{eR_\star B}{m_e c^2}, \quad \tilde{j} = \frac{4\pi eR_\star^2 j}{m_e c^3}.
\]  

(5.26)

Hereafter we will use tilde to denote dimensionless quantities, e.g. \( \tilde{r} = r / R_\star \), \( \tilde{t} = ct / R_\star \), etc.

### 5.4 Results

In all simulations presented below the magnetic field strength at the pole of the star is \( \tilde{B}_{\text{pole}} = 4 \times 10^4 \). It corresponds to \( \tilde{\omega}_B = 4 \times 10^4 \). We focus on the simulation with \( \gamma_{\text{thr}} = 10 \), as it gives the best re-scaled model of real magnetars (Section 5.3.4). Simulations with different \( \gamma_{\text{thr}} \) are only discussed in Section 5.4.3.

#### 5.4.1 Initial relaxation

During the initial stage of the simulation \( \tilde{t} < \tilde{t}_{\text{shear}} = 40 \) the dipole magnetosphere is twisted by the surface shearing motion described in Section 5.3.1. The surface motion induces a parallel electric field \( E_\parallel \), which lifts charges from the atmospheric layer into the magnetosphere and
accelerates them. The electron Lorentz factors quickly reach $\gamma_{\text{thr}}$ and $e^\pm$ discharge is triggered within a single Alfvén time of the twisted field line bundle.

The $e^\pm$ plasma created by the discharge screens $E_\parallel$, and the voltage along the current loop temporarily drops, shutting down the discharge. As the created pairs are lost to the star on the light-crossing time, a charge-starved region with significant $E_\parallel$ develops again. This first happens near the equatorial plane. As a result, an equatorial gap with strong $E_\parallel$ emerges and begins to accelerate particles, sustaining the pair creation process. The gap structure and how the $e^\pm$ discharge is sustained will be described in more detail in Section 5.4.2.

It is clear from the simulation that a magnetospheric source of pair plasma is established in the twisted magnetosphere on a timescale not much longer than the light crossing time, before the surface shearing ends at $t_{\text{shear}}$. Pair creation becomes the dominant source of plasma; the extraction of particles from the atmospheric layer is only important at the initial stage igniting the $e^\pm$ discharge. After the pair discharge is activated, only a small fraction of the magnetospheric current is carried by the particles lifted from the surface. In particular, we observed that less than 1% of the current is carried by the ions.

We also observed that the twist implantation at $t < t_{\text{shear}}$ is accompanied by excitation of Alfvén waves, which bounce back and forth along the magnetospheric field lines. Similar waves were observed in FFE simulations (PBH13). The waves are damped in the magnetosphere at later times, and the initial relaxation period is followed by the gradual evolution on a much longer timescale $\tilde{t}_{\text{ohm}} \gg 100$.

After the surface shearing stopped at $t_{\text{shear}}$, the electric discharge persisted for the rest of the simulation. It continually supported the electric current in the slowly untwisting magnetosphere, and the created particles continually bombarded the star. The duration of the simulation $t_{\text{sim}} = 350$ was approximately 9 times longer than $t_{\text{shear}}$ and comparable to the expected resistive timescale $t_{\text{ohm}}$ estimated in Section 5.2. The observed gradual evolution of the magnetospheric twist and 3Alfvén waves are reflected from the rigid sphere and trapped in the magnetosphere. Our simulation neglects the fact that the crustal material has a finite strength, which can lead to plastic damping of Alfvén waves in the crust (Li & Beloborodov 2013).
currents on the timescale $\sim t_{\text{ohm}}$ will be described in Section 5.4.4.

### 5.4.2 The equatorial gap

A key aspect of the discharge self-organization is how and where particles are accelerated. The simulation clearly shows the formation of a quasi-steady “gap” with a strong $E_{\parallel}$ concentrated around the equatorial plane (Figure 5.1). The gap thickness $\ell_{\text{gap}}$ is smaller than radius, and its voltage is near the threshold for $e^{\pm}$ discharge,

$$\Phi_{\text{gap}} \approx \ell_{\text{gap}} E_{\text{gap}}, \quad e\Phi_{\text{gap}} \approx \gamma_{\text{thr}} m_e c^2. \quad (5.27)$$

Particles are accelerated in the gap and most of the pair creation events happen around this region.

As seen in Figure 5.1, the gap has a rather sharp boundary; $E_{\parallel}$ is screened outside it by the created $e^{\pm}$ plasma. The drop of $E_{\parallel}$ across the two boundaries of the gap is sustained by the layers of positive and negative charge ($\pm \Sigma$ above and below the equatorial plane, respectively), according to Gauss law $\nabla \cdot E = 4\pi \rho$. The charged layers are self-consistently sustained by the difference in velocities of positive and negative charges passing through them in the self-organized $E_{\parallel}$.

In essence, the gap is a double layer. It has been compressed toward the equatorial plane to a minimum thickness $\ell_{\text{gap}}$ that is still capable of sustaining particle acceleration to $\gamma_{\text{thr}}$. Similar to the double layer described in Section 5.2.1 the charge layers sandwiching the gap have the thickness comparable to the local plasma skin depth $\lambda_p$ (evaluated for charge density $\sim j/c$) (Figure 5.2). The electric field in the gap is $E_{\text{gap}} \sim 4\pi(j/c)\lambda_p$ and its voltage is

$$e\Phi_{\text{gap}} \sim \frac{\ell_{\text{gap}}}{\lambda_p} m_e c^2. \quad (5.28)$$

The self-regulation of the gap voltage to $\Phi_{\text{gap}} \approx \Phi_{\text{thr}}$ controls the gap thickness $\ell_{\text{gap}} \sim \gamma_{\text{thr}} \lambda_p$.

Unlike normal double layers, particles accelerated in the gap are not brought from outside; instead, the gap feeds itself with particles. The accelerated particles create secondary $e^{\pm}$ of lower
energies near the gap exit, and some of the secondary particles are reversed by $E_\parallel$ and accelerated toward the opposite boundary of the gap, where they create new pairs, etc.

The multiplicity of the pair plasma is defined by $M = (\rho_+ - \rho_-)c/j$, where $\rho_+$ and $\rho_-$ are the charge densities of the positrons and electrons, respectively. One can see in Figure 5.3 that $M$ in the gap is close to 1, i.e. the gap contains the minimum amount of plasma needed to conduct the electric current. This is consistent with no screening in the gap that allows the strong $E_\parallel$ to be sustained. Pair multiplicity in other parts of the j-bundle is close to 2, just sufficient to screen $E_\parallel$. Apparently, the discharge in the simulation is self-organized to carry the current with the minimum voltage $\Phi_e \approx \Phi_{\text{gap}} \approx \Phi_{\text{thr}}$ and the minimum rate of pair creation.
Figure 5.2: Charge density in the magnetosphere, averaged in the same way as in Figure 5.1. Note the thin charged layers bounding the equatorial gap across the magnetic field lines. The layers extend into the inner magnetosphere along the inner boundary of the j-bundle. The charged structure observed on the field lines extending to $\tilde{r} \sim 9$ approximately corresponds to the outer boundary of the j-bundle (see Figure 5.6).

Figure 5.4 shows the average hydrodynamic momenta of electrons and positrons. It is apparent that both species are accelerated across the equatorial gap to the threshold Lorentz factor $\gamma_{\text{thr}} = 10$. The move with almost speed of light in the opposite directions and make approximately equal contributions to the current density, consistent with $M \approx 1$. Outside the gap, $M \approx 2$ together with the charge neutrality condition $n_+ \approx n_-$ implies that the current is carried by one species while the other creates the neutralizing, nearly static, background. This is indeed observed in Figure 5.4.

The gap voltage is not exactly steady and shows quasi-periodic “breathing” with time. This must assist the gap in reversing some of the secondary particles so that they can cross the gap
and accelerate to $\gamma_{\text{thr}}$, sustaining the pair creation cycle. Most of the accelerated particles escape the gap and get absorbed by the star.

Since the magnetosphere was set up to be symmetric about the equatorial plane, the fact that the current is strongly dominated by created pairs implies symmetric bombardment of the two footprints of the $j$-bundle. Thus, our simulation shows two symmetric hot spots (or rather rings, due to the axial symmetry) in the northern and southern hemispheres of the star.

As discussed in BT07 and Section 5.2.1, the voltage $\Phi_e$ in the magnetospheric circuit is purely inductive. The parallel electric field $E = -c^{-1}\partial A/\partial t$ is associated with the slow dissipation of $B_\phi$ rather than an electrostatic potential. Note also that the dissipation rate $E \cdot j = E_\parallel j$ is localized in the gap while the untwisting of $B_\phi$ also occurs outside the gap. The re-distribution of the dissipated $B_\phi$ along the $j$-bundle into the screened region with $E_\parallel \approx 0$ occurs through the Alfvén mode, which can propagate without dissipation. The Alfvén timescale $t_A \sim r/c$ is much shorter than
the untwisting timescale $t_{\text{ohm}}$, and so the magnetosphere slowly evolves through the sequence of global twist equilibria of a decreasing energy $E_{\text{twist}}$, even though the magnetic energy is converted to heat only near the equator.

### 5.4.3 Dependence on the threshold voltage

While the simulation with $\gamma_{\text{thr}} = 10$ is the most adequate re-scaled version of the magnetar magnetosphere (Section 5.3.4), we also performed simulations with $\gamma_{\text{thr}} = 20, 100, \text{and } \infty \text{ (no pair creation). All other parameters of the four simulations were identical.}$

Figure 5.5 shows the evolution of the twist energy $E_{\text{twist}}$ in the simulations with the four different values of $\gamma_{\text{thr}}$. An obvious trend is observed: a higher threshold voltage for discharge, $e\Phi_{\text{thr}} = \gamma_{\text{thr}}m_ec^2$, leads to a higher dissipation rate and a shorter lifetime of the magnetic twist. When $\gamma_{\text{thr}} \gg 10$, the dissipation becomes so strong that it affects the initial stage of the twist implantation at $\tilde{t} < \tilde{t}_{\text{shear}} = 40$, so that a substantial part of the twist amplitude (and the corresponding energy $E_{\text{twist}}$) is lost before it could be implanted.
The extreme model with $\gamma_{\text{thr}} = \infty$ gives so strong dissipation that $E_{\text{twist}}$ does not reach even 10% of its target value. It is instructive to compare this simulation with the expected dissipation rate in the pair-free configuration described in Section 5.2.1. From equation (5.6), we can estimate the voltage drop of the double layer as $\gamma_{\text{DL}} = \tilde{\Phi} \sim \sqrt{\tilde{j} \tilde{W}}$. The initial width of the j-bundle near the star is $W \sim 1$. The target current density reaches $\tilde{j} \sim 3 \times 10^4$ if the twist is fully implanted. This estimate gives $\gamma_{\text{DL}}$ comparable to 200; the actual voltage in the simulation reaches somewhat higher values. The high voltage develops early during the shearing stage and results in strong dissipation, which does not allow $\tilde{j}$ to approach $3 \times 10^4$.

The simulation with $\gamma_{\text{thr}} = 100$ enables the pair discharge, which buffers the voltage growth in the j-bundle and allows a stronger twist to be implanted. The simulations with $\gamma_{\text{thr}} = 20$ and, in particular, $\gamma_{\text{thr}} = 10$, allow almost full implantation of the target twist with small ohmic losses. The subsequent slow resistive evolution is similar in the two models, as both have $\Phi_{\text{thr}}$ well below the double-layer voltage and sustain a long-lived discharge activity in the j-bundle. As expected, the untwisting timescale $t_{\text{ohm}}$ is reduced by a factor of 2 as $\gamma_{\text{thr}}$ is increased from 10 to 20 (see Equation 5.16).

These results unambiguously demonstrate that the energy dissipation timescale is controlled by the pair creation threshold, confirming the conclusion of BT07. In real magnetars, we expect $\gamma_{\text{thr}} \ll \gamma_{\text{DL}}$ (Section 5.2). Therefore, the most relevant model is the one with low $\gamma_{\text{thr}} = 10$, which is still high enough to accelerated particles to ultra-relativistic energies and produce relativistic secondary $e^{\pm}$.

### 5.4.4 Expanding cavity

Figure 5.6 shows the resistive evolution of the j-bundle. The untwisting of the magnetic field lines proceeds as anticipated in Section 5.2.4 through formation of a cavity $j = 0$ that expands from the inner magnetosphere near the equator (large flux function $u$). Figure 5.7 shows the evolution of the poloidal current $j_p$ until the end of the simulation at $\tilde{t}_{\text{sim}} = 350$. We chose to
Figure 5.5: Evolution of the twist magnetic energy $E_{\text{twist}}$. Four simulations are shown with discharge thresholds $\gamma_{\text{thr}} = 10, 20, 100, \infty$. We use the exact expression for $E_{\text{twist}} = \int \frac{(B^2 - B_0^2)}{8\pi} dV$, where $B_0$ is the initial dipole field. It takes into account that besides $B^2_\phi/8\pi$ part of the twist energy is stored in the inflated poloidal magnetic field, which becomes important when the twist amplitude $\psi$ exceeds unity.

show $j_p/B_p$ because this quantity is constant along the magnetic field lines (after averaging over short-timescale fluctuations), as expected in a nearly force-free magnetosphere — currents flow along the magnetic field lines. Therefore, $j_p/B_p$ is a function of the magnetic field line, which we label by the parameter $u = \sin^2 \theta_\star$ (see Equation (5.15)). Note the expansion of the region where $j_p = 0$ toward the magnetic axis, from $u \approx 0.75$ to $u \lesssim 0.55$.

Figure 5.8 shows the evolution of the integrated twist angle $\psi$ defined in Equation (5.1). The untwisting proceeds from near the equator, where the twist angle decreases over time, but the twist angle is not simply erased, but relocated from the inner magnetosphere to the outer parts, as expected from the untwisting Equation (5.13).

A curious feature is observed to develop on the magnetic field lines with $u$ around 0.22: the twist angle $\psi$ grows and approaches 3.5 toward the end of the simulation. This feature is also seen
Figure 5.6: Color plot showing the evolution of the poloidal current density $j_p$ in the simulation with $\gamma_{\text{thr}} = 10$. Four snapshots are shown: (a) $\tilde{t} = 30$, (b) $\tilde{t} = 120$, (b) $\tilde{t} = 230$, and (d) $\tilde{t} = 350$. Note that when $j_p = 0$ then also $\tilde{j} = 0$.

in the current structure shown in Figures 5.6 and 5.7. The strongly twisted, narrow bundle of field lines is inflating with time and eventually opens up, causing a magnetospheric instability (cf. PBH13). Our simulation stopped right at the onset of this development, since we would like to limit our study to the quasi-steady untwisting regime. An important difference from over-twisting studied in PBH13 is that here it is not driven by excessive surface shear. Instead, it results from resistive evolution of the implanted twist while the crust is static.
Figure 5.7: Evolution of the poloidal current distribution in the magnetosphere in the simulation with $\gamma_{\text{thr}} = 10$. The ratio $j_p/B_p$ (constant along the magnetic field lines) is shown versus the poloidal flux function defined in Equation (5.15); $\theta_*$ is the polar angle of magnetic field line footprint on the star. The different curves show snapshots at times $\tilde{t} = 50, 100, 150, 200, 250, 300$, and 350.

Figure 5.8: Evolution of the twist angle $\psi$ in the simulation with $\gamma_{\text{thr}} = 10$. 
5.5 Discussion

We have performed the first axisymmetric particle-in-cell simulations of the twisted magnetospheres of magnetars. The simulations demonstrate from first principles that electric $e^\pm$ discharge is self-organized in the magnetosphere to sustain the electric current $j$ demanded by the magnetospheric twist.

The results of our numerical experiment may be summarized as follows.

1. Shear motion of the stellar surface on a timescale $t_{\text{shear}} < t_{\text{ohm}}$ successfully implants a magnetic twist in the magnetosphere. The twist is supported by continual electric current due to self-organized $e^\pm$ discharge.

2. Particles are accelerated along the magnetic field lines to Lorentz factors $\gamma \approx \gamma_{\text{thr}}$, just sufficient to ignite pair creation. The voltage sustaining the electric circuit, the dissipation rate, and the lifetime of the twist are all regulated by $\gamma_{\text{thr}}$.

3. Particle acceleration is localized in a gap near the equatorial plane (Figure 5.1). The gap has the electric field $E_\parallel \sim 4\pi (j/c)\lambda_p$ and width $\ell_{\text{gap}} \sim \gamma_{\text{thr}}\lambda_p$, where $\lambda_p = (m_e c^3 / 4\pi e j)^{1/2}$ is the local plasma skin-depth. The plasma density in the gap is close to the minimum value $n = j / ec$ required to conduct the electric current. Continual $e^\pm$ creation occurs near the two exits from the gap.

4. The magnetospheric current is carried by electrons and positrons created in the magnetosphere rather than electrons and ions extracted from the atmospheric layer on the stellar surface. The created particles rain onto the footprints of the $j$-bundle, creating two hot spots.

5. Resistive untwisting of the magnetosphere occurs on the timescale $t_{\text{ohm}}$ estimated in Equation (5.16), in agreement with theoretical expectations. The evolution proceeds as predicted in B09: a cavity with $j = 0$ quickly forms in the inner magnetosphere and gradually expands, erasing the remaining electric currents.
6. A curious feature was observed in the untwisting process: while the twist energy was decreasing as expected from ohmic dissipation, the twist amplitude \( \psi \) grew in a narrow bundle of field lines at the outer boundary of the twisted region. This over-twisted bundle inflated so much that it eventually opened up.

Our results confirm that the untwisting magnetospheres naturally create shrinking hot spots (footprints of the shrinking \( j \)-bundle), which have been detected in 7 transient magnetars. The evolution timescale inferred from the simulations (Equation 5.16) is consistent with the decay timescale observed in transient magnetars (months to years).

One unknown in the setup of our numerical experiment is the profile of the surface shear. However, basic features observed in the simulation, in particular voltage regulation through \( e^\pm \) discharge and the cavity expansion, should be generic and independent of the details of the twist profile. It is less clear how generic is the formation of the narrow over-twisted bundle. This could be further explored with simulations of different shear profiles.

An important caveat in the simulation setup is the simplified “on the spot” prescription for pair creation, with the created \( e^\pm \) pair taking a significant energy fraction from the primary particle. As briefly discussed in Section 5.3.3, this prescription is reasonable if the twist is confined to the region of ultrastrong magnetic field near the star, \( B \gtrsim B_Q \). Pair creation in weaker fields tends to occur with high multiplicities, which can launch a dense \( e^\pm \) outflow and efficiently screen \( E_\parallel \) in the equatorial region (Beloborodov, 2013). Then the gap may have to split into two gaps and move away from the equator, closer to the star.

How the discharge will self-organize in this case can only be explored using a more detailed implementation of the pair creation process. The future simulation will directly track the high-energy photons produced by resonant scattering and their conversion to pairs, without prescribing any \( \gamma_{\text{thr}} \). This will be the focus of our future work, and we expect it to establish the gap location on magnetic field lines extending far from the star. This part of the magnetosphere is interesting for two reasons: (1) the \( j \)-bundle activity tends to concentrate on the extended field lines, and (2)
the nonthermal emission is able to escape the outer magnetosphere while almost all resonantly scattered photons in the region $B \gg 10^{13}$ G convert to pairs (Beloborodov, 2013). Gap location on the extended field lines influences the hard X-ray spectrum emitted by the twisted magnetosphere, and thus can be tested against observations. Phase-resolved hard X-ray spectra have been measured for several magnetars and fitted by the $e^\pm$ outflow model (e.g. Hascoët et al., 2014; An et al., 2015), which assumes an electric gap near the star. Direct PIC simulations of the $e^\pm$ discharge of high multiplicity can verify or disprove this assumption.

We did not study in this chapter what happens when the magnetosphere is over-twisted and becomes unstable. This phenomenon is associated with the observed giant flares of magnetars, an extreme analogy of solar flares. The over-twisted magnetosphere inflates and creates a thin current sheet separating magnetic fluxes of opposite polarities. The current sheet becomes unstable to the tearing mode, which leads to magnetic reconnection and ejection of plasmoids from the magnetosphere (Lyutikov, 2003; PBH13), resembling the mechanism of coronal mass ejections from the sun (e.g. Mikic & Linker, 1994). Our preliminary studies using Aperture show similar behavior. One difficulty encountered by such simulations is the huge pair creation rate in the dissipative current sheet, which must result in quick thermalization of the released magnetic energy. A scheme describing this transition needs to be developed and will be a topic for future work.
Chapter 6

Further Explorations in Extension of PIC

6.1 General Relativity

This is an exploratory effort of expanding a PIC code to support general relativistic simulations. To my knowledge at the time of this writing, no such code exists, and a successfully implemented code like this can open up the possibility of simulating e.g. the pair discharge process around a black hole. Here I outline a possible implementation of a full GR PIC code. In all of the following chapter, Einstein summation convention is assumed, where repeated indices are contracted and summed over. Greek indices $\mu, \nu$, etc. denote summation over 0, 1, 2, 3; Latin indices $i, j$, etc. only go over spatial indices 1, 2, 3.

6.1.1 Metric and Observer

This subsection is purely for reference. We record the results and move on. When referring to a definition of a quantity one should check back to this subsection. We start with the $3 + 1$ split following MacDonald & Thorne (1982), writing the metric as

$$ds^2 = (\beta^2 - \alpha^2) dt^2 + 2 \beta_i dx^i dt + \gamma_{ij} dx^i dx^j$$

(6.1)
where \( \alpha \) is called the “lapse function” and \( \beta \) is called the “shift vector”. We consider only metric that are time independent meaning \( \partial_t g_{\alpha\beta} = 0 \). We follow Komissarov (2004) in defining the fiducial observer (FIDO) 4-velocity:

\[
    n_\mu = (-\alpha, 0, 0, 0)
\]

The spatial metric tensor, which is used to raise and lower spatial indices, is defined using this FIDO velocity

\[
    \gamma_{\alpha\beta} = g_{\alpha\beta} + n_\alpha n_\beta
\]

Note since \( n_\alpha \) only has the temporal component \( \gamma_{ij} \) is the same as \( g_{ij} \). The other useful identities are:

\[
    n^\mu = \frac{1}{\alpha} (1, -\beta^i)
\]

\[
    g^{0\mu} = -\frac{1}{\alpha} n^\mu
\]

\[
    g^{ij} = \gamma^{ij} - \beta^i \beta^j / \alpha^2
\]

and

\[
    g = -\alpha^2 \gamma, \quad \beta^i = \gamma^{ij} \beta_j, \quad g = \det g_{\mu\nu}, \quad \gamma = \det \gamma_{ij}
\]

### 6.1.2 Field Equations

We define as usual the Maxwell tensor from the vector potential

\[
    F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu
\]

where the physical meaning of \( A_\mu \) will come clear later. In GR we are supposed to replace \( \partial_\mu \) with \( \nabla_\mu \), but since \( \Gamma^\mu_{\alpha\beta} \) is symmetric in the lower indices, it is okay to simply use partial derivative here.
The upper $F^\mu{}^\nu$ are defined simply by raising the indices

$$F^\mu{}^\nu = g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta}$$  \hfill (6.9)

The covariant Maxwell equations are the usual Maxwell equations with covariant derivative in place of ordinary partial derivative:

$$\nabla_\nu F^{\ast\mu\nu} = 0, \quad \nabla_\nu F^{\mu\nu} = I^\mu$$  \hfill (6.10)

where $I^\mu$ is the 4-current and $F^{\ast\mu\nu}$ is defined by equation (6.17). Again, to associate physical meaning to this current we need to relate it with particle motion, which we will do subsequently.

Taking the first equation, the temporal and spatial parts read separately:

$$\frac{1}{\sqrt{-g}} \partial_t(\sqrt{-g} F^{\ast 0i}) = 0$$  \hfill (6.11)

$$\frac{1}{\sqrt{-g}} \partial_t(\sqrt{-g} F^{\ast i0}) + \frac{1}{\sqrt{-g}} \partial_j(\sqrt{-g} F^{\ast ij}) = 0$$  \hfill (6.12)

since $\sqrt{-g} = \alpha \sqrt{\gamma}$, we can write the spatial equation as:

$$\frac{1}{\sqrt{\gamma}} \partial_t(\sqrt{\gamma} \alpha F^{\ast i0}) + \frac{1}{\sqrt{\gamma}} \partial_j(\alpha \sqrt{\gamma} F^{\ast ij}) = 0$$  \hfill (6.13)

This can be simplified to something we can identify with, by defining new fields $B^i = \alpha F^{\ast i0}$ and $E_i = \frac{1}{2} \alpha \varepsilon_{ijk} F^{ijk}$ (we also used the fact that $\sqrt{\gamma}$ is time-independent):

$$\partial_t B^i + \varepsilon_{ijk} \partial_j E_k = 0$$  \hfill (6.14)

where $\varepsilon_{ijk} = \sqrt{\gamma} \varepsilon_{ijk}$ is the Levi-Civita pseudo-tensor and $\varepsilon_{ijk}$ is the totally antisymmetric symbol which is either 1 or $-1$. Similarly $\varepsilon_{ijk} = \varepsilon_{ijk} / \sqrt{\gamma}$ where $\varepsilon_{ijk}$ is numerically the same as $\varepsilon_{ijk}$. This equation looks like the usual Maxwell equation where $\partial_t \mathbf{B} = -\nabla \times \mathbf{E}$, however we try not to rely
on the usual curl and bold-face vector notation, and make clear whenever we talk about a field component whether it is the upper or lower index version.

The spatial part of the second Maxwell equation reads:

\[
\frac{1}{\sqrt{\gamma}} \partial_t (\sqrt{\gamma} \alpha F^{0i}) + \frac{1}{\sqrt{\gamma}} \partial_j (\alpha \sqrt{\gamma} F^{ij}) = \alpha I^i \tag{6.15}
\]

Similarly we can identify new fields \( D^i = \alpha F^{0i} \) and \( H_i = \frac{1}{2} \alpha \epsilon_{ijk} F^{jk} \), such that this equation reads:

\[
- \partial_t D^i + \epsilon^{ijk} \partial_j H_k = \alpha I^i \tag{6.16}
\]

There are in total of 4 dynamic fields: \( B^i, E_i, D^i, \) and \( H_i \). They are not independent so we want to find their relations to cut down the number of fields. From the definition of the Maxwell field tensor we have in vacuum

\[
\star F_{\alpha\beta} = \frac{1}{2} \epsilon^{\alpha\beta\mu\nu} F_{\mu\nu}, \quad F_{\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\alpha\beta} \star F_{\alpha\beta} \tag{6.17}
\]

so we can find that

\[
H_i = \star F_{0i}, \quad E_i = F_{i0}, \quad B^i = \frac{1}{2} \epsilon^{ijk} F_{jk}, \quad D^i = \frac{1}{2} \epsilon^{ijk} \star F_{jk} \tag{6.18}
\]

To figure out the relations between them we can take:

\[
\alpha D^i = \alpha^2 F^{0i} = \alpha^2 g^{0\alpha} g^{i\beta} F_{\alpha\beta} \\
= \alpha^2 g^{00} g^{ij} F_{0j} + \alpha^2 g^{0j} g^{i0} F_{j0} + \alpha^2 g^{0i} g^{jk} F_{jk} \\
= \gamma^{ij} E_j \frac{\beta^i \beta^j}{\alpha^2} E_j + \frac{\beta^j \beta^j}{\alpha^2} E_j + \beta^i g^{ik} F_{jk} \\
= \gamma^{ij} E_j + \beta^j \gamma^{ik} e_{jkl} B^l - \frac{1}{\alpha} \beta^j \beta^k e_{jkl} B^l \tag{6.19}
\]
Since the last term on the left hand side is zero due to antisymmetry we have:

\[ E_i = \alpha \gamma_{ij} D^j + e_{ijk} \beta^j B^k \]  
(6.20)

Similarly we can work from \( \alpha B^i \) to get

\[ H_i = \alpha \gamma_{ij} B^j - e_{ijk} \beta^j D^k \]  
(6.21)

Since we can express \( E_i \) and \( H_i \) in terms of \( B^i \) and \( D^i \) and some metric coefficients, we will call \( B^i \) and \( D^i \) our dynamic fields. It is also convenient that equations (6.14) and (6.16) are already time evolution equations of these two dynamic fields. Therefore our general strategy of solving them will be to figure out the auxiliary fields \( E_i \) and \( H_i \) from equations (6.20) and (6.21), then use them to update the fields \( B \) and \( D \).

It is also worth writing down the integral version of the equations (6.14) and (6.16). If \( \Sigma \) is the cell face in direction \( i \), then we have

\[
\partial_t \int_{\Sigma} \frac{1}{2} e_{ijk} D^j dx^j \wedge dx^k = \oint_{\partial \Sigma} H_i dx^i - \int_{\Sigma} \frac{1}{2} \alpha e_{ijk} l^i dx^j \wedge dx^k \]  
(6.22)

\[
\partial_t \int_{\Sigma} \frac{1}{2} e_{ijk} B^j dx^j \wedge dx^k = -\oint_{\partial \Sigma} E_i dx^i \]  
(6.23)

### 6.1.3 Particle Equations

Now that we decided that \( B^i \) and \( D^i \) should be our dynamic fields, we need to find how they affect particles. To do that we use the Lagrangian formalism and write down the action as follows (Landau & Lifshitz):

\[
S = \sum \int m c d\tau - \sum \int \frac{e}{c} A_\mu dx^\mu - \frac{1}{16\pi c} \int F_{\mu\nu} F^{\mu\nu} \sqrt{-g} d\Omega \]  
(6.24)

where the sum is over all particles, and \( d\Omega \) is the spacetime volume element.
We try to follow Landau & Lifshitz to find the equation of motion for a single particle. The principle of least action states that (for the two terms of the action):

$$\delta S = \delta \int \left( mc \, d\tau - \frac{e}{c} A_\mu \, dx^\mu \right) = 0$$ (6.25)

We do the variation of the first part in the following way. Starting with $d\tau^2 = -ds^2 = -g_{\mu\nu} dx^\mu dx^\nu$, we have

$$\delta d\tau^2 = 2d\tau \delta d\tau = -\delta (g_{\mu\nu} dx^\mu dx^\nu) = -dx^\rho dx^\sigma \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \delta x^\rho - 2g_{\mu\nu} dx^\mu \delta x^\nu$$ (6.26)

we can then write the variation of the whole action

$$\delta S = -mc \int \left[ \frac{1}{2} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \partial_\sigma g_{\mu\nu} \delta x^\sigma + g_{\rho\nu} \frac{dx^\rho}{d\tau} \frac{dx^\sigma}{d\tau} \delta x^\sigma \right] d\tau - \frac{e}{c} \int (A_\mu \delta x^\mu + \delta A_\mu dx^\mu)

= -mc \int \left[ \frac{1}{2} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \partial_\sigma g_{\mu\nu} \delta x^\sigma - \frac{d}{d\tau} \left( g_{\rho\nu} \frac{dx^\rho}{d\tau} \right) \delta x^\rho \right] d\tau - \frac{e}{c} \int (\partial_\sigma A_\mu \delta x^\sigma - \partial_\sigma A_\mu dx^\sigma \delta x^\mu)

= -mc \int \left[ \frac{1}{2} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \partial_\sigma g_{\mu\nu} - \frac{d}{d\tau} \left( g_{\rho\sigma} \frac{dx^\rho}{d\tau} \right) \delta x^\rho \right] d\tau - \frac{e}{c} \int (\partial_\sigma A_\mu - \partial_\mu A_\sigma) \frac{dx^\mu}{d\tau} \delta x^\sigma d\tau$$ (6.27)

Since the variation should vanish for any $\delta x^\sigma$, the integrand should vanish therefore we get our dynamic equation (as expected):

$$g_{\mu\sigma} \frac{du^\mu}{d\tau} + \frac{1}{2} (\partial_\sigma g_{\mu\rho} + \partial_\rho g_{\mu\sigma} - \partial_\sigma g_{\rho\mu}) u^\rho u^\nu = \frac{e}{c} (\partial_\sigma A_\mu - \partial_\mu A_\sigma) u^\nu$$ (6.28)

where $u^\mu = dx^\mu / d\tau$ is the 4-velocity of the particle. We shall always use $u^\mu$ to denote the standard 4-velocity from now on. In other words we have

$$mc \frac{Du^\mu}{D\tau} = \frac{e}{c} F_{\mu\nu} u^\nu, \quad \text{or} \quad mc \frac{Du_\mu}{D\tau} = \frac{e}{c} F_{\mu\nu} u^\nu$$ (6.29)

These two are equivalent because the covariant derivative defined by Christoffel symbols are automatically metric-compatible, so $\nabla_\sigma g_{\mu\nu} = 0$. We can therefore freely raise and lower indices.
inside covariant derivatives.

If we take the spatial part of the second equation, and multiply it by $d\tau / dt$, we get (from now on we ignore factors of $c$)

$$m \frac{Du_i}{D\tau} \frac{d\tau}{dt} = eF_iu^\nu \frac{d\tau}{dt}, \quad m \frac{Du_i}{Dt} = eF_i \frac{dx^\nu}{dt} \quad (6.30)$$

For simplicity of notation lets use $Du/Dt$ to refer to the product $(Du_i/D\tau)(d\tau/dt)$ although it might not be a proper covariant derivative. We also define $\nu^\mu = dx^\mu / dt$. From here on $u^\mu$ will mean the ordinary 4-velocity and $\nu^\mu$ means the derivative with respect to coordinate time. Note $\nu^0 = 1$ by definition. We can now write

$$m \frac{Du_i}{Dt} = e(F_{i0}\nu^0 + F_{ij}\nu^j) = e(E_i + e_{ijk}B^k)$$

$$= e \left( \alpha\gamma_{ij}D^j + e_{ijk}\beta^jB^k + e_{ijk}\nu^j B^k \right) \quad (6.31)$$

The last line is because we want to express particle acceleration in terms of our dynamic fields $B^i$ and $D^i$. Due to the form of the equation, it is convenient to define a new velocity $\bar{\nu}$ such that

$$\alpha\bar{\nu}^i - \beta^i = \nu^i \quad (6.32)$$

Thus we will have

$$m \frac{Du_i}{Dt} = e\alpha(\gamma_{ij}D^j + e_{ijk}\bar{\nu}^jB^k) \quad (6.33)$$

which looks like the ordinary Lorentz force.

Now the remaining problem is that we only have an update equation for $u_i$, so we need a relation between $u_i$ and $\bar{\nu}^j$. This is relatively simple since we know $u^\mu u_\mu = -1$ for a time-like worldline. Let’s call $u^0 = dt/d\tau = \gamma_p$. Then

$$u_i = g_{iv}u^v = \beta_i\gamma_p + \gamma_{ij}\gamma_p\nu^j = \alpha\gamma_p\gamma_{ij}\bar{\nu}^j \quad (6.34)$$
And the particle Lorentz factor $\gamma_p$ is related to $\vec{v}^i$ by (from $u^\mu u_\mu = -1$)

$$\gamma_p^2 = \frac{1}{\alpha^2 (1 - \gamma_{ij} \vec{v}^i \vec{v}^j)} \quad (6.35)$$

So the Lorentz factor is similar to conventional one with $\vec{v}$, except for the extra factor of $\alpha$.

Because by definition we have $v^i = dx^i/dt$, the particle positions can be updated by

$$\frac{dx^i}{dt} = v^i = \alpha \vec{v}^i - \beta^i \quad (6.36)$$

The only remaining problem with particle dynamics is that we have a hard time evaluating the covariant derivative $Du_i/Dt$. An approximation might be simply replace it with $du_i/dt$. We will not go into possible different schemes here.

Additional note on implementation of the particle pusher. We need to find an efficient way to store the quantities that we need, which in this case is $\vec{v}^i$. However when the particles become very relativistic, $\vec{v}^i$ is close to 1 and therefore prone to numerical truncation error. It’s much better to store some form of momentum. We could store directly $u_i$ which we update, but then it is slightly difficult to find $\gamma_p$ from it, where a matrix inversion of $\gamma_{ij}$ is involved. As a compromise, we can define

$$\vec{p}^i = \alpha \gamma_p \vec{v}^i \quad (6.37)$$

so that we have the following relations:

$$u_i = \gamma_{ij} \vec{p}^j, \quad \gamma_p^2 = \frac{1}{\alpha^2} (1 + \gamma_{ij} \vec{p}^j \vec{p}^j) \quad (6.38)$$

The particle momentum update step now should look like the following: First we find $u_i$ using equation (6.38), and $\vec{v}^i$ using equation (6.37), then we use equation (6.33) and the modified Vay algorithm (6.1.4) to find $\Delta u_i$ over the time step. When we have the new $u_i$ we can invert back to find the $\vec{p}^i$ again by inverting the matrix $\gamma_{ij}$. This step will be simple in metrics where $\gamma_{ij}$ is diagonal. If
we store $\vec{p}^i$ as the momentum of the particles, then $\vec{v}^i$ can simply be calculated by dividing it by $\alpha \gamma p$. It might be useful to store $\alpha \gamma p$ of the particle (where $\alpha$ is evaluated at the particle position) in the particle storage as well.

### 6.1.4 Modified Vay Algorithm

Vay (2008) showed a stable algorithm for pushing particles in an electromagnetic field. It was designed to solve the discretized equation

$$
\frac{\gamma^{n+1} \mathbf{v}^{n+1} - \gamma^n \mathbf{v}^n}{\Delta t} = \frac{q}{m} \left( \mathbf{E}^{n+1/2} + \frac{\mathbf{v}^n + \mathbf{v}^{n+1}}{2} \times \mathbf{B}^{n+1/2} \right)
$$

(6.39)

where $\gamma^n = (1 - \mathbf{v}^n \cdot \mathbf{v}^n)^{-1/2}$. Notice that we only have one kind of vector in this equation, namely $\mathbf{v}$, or in our notation, the contravariant vector $\nabla^i$. If we discretize our equations in the same way, then we will find the following equation:

$$
\frac{\gamma_{ij}(\vec{p}^i)^{n+1} - \gamma_{ij}(\vec{p}^i)^n}{\Delta t} = \frac{\alpha q}{m} \left( \gamma_{ij} \mathbf{D}^j + \sqrt{\gamma} \epsilon_{ijk} \frac{(\vec{v}^i)^n + (\vec{v}^i)^{n+1}}{2} \mathbf{B}^k \right)
$$

(6.40)

where $\vec{p}^i = \alpha \gamma p \vec{v}^i$ is similar to how usual relativistic momentum is related to velocity. The crucial difference from (6.39) is the appearance of the metric matrix $\gamma_{ij}$ on the left hand side.

The usual Vay algorithm solves equation (6.39) using two steps. The first step is to split the equation into two parts and define (where $\mathbf{u} = \gamma \mathbf{v}$)

$$
\mathbf{u}' = \mathbf{u}^n + \frac{q \Delta t}{m} \left( \mathbf{E}^{n+1/2} + \frac{\mathbf{v}^n}{2} \times \mathbf{B}^{n+1/2} \right)
$$

(6.41)

then one solves the resulting implicit equation

$$
\mathbf{u}^{n+1} = \mathbf{u}' + \frac{q \Delta t}{m} \left( \frac{\mathbf{u}^{n+1}}{2 \gamma^{n+1}} \times \mathbf{B}^{n+1/2} \right)
$$

(6.42)

If one writes $\mathbf{b} = (q \Delta t/2m) \mathbf{B}^{n+1/2}$, then the solution to the above equation is simply (writing $\gamma$
instead of \( \gamma^{n+1} \) for simplicity)

\[
\mathbf{u}^{n+1} = \left[ \mathbf{u}' + (\mathbf{u}' \cdot \mathbf{b})\mathbf{b}/\gamma^2 + \mathbf{u}' \times \mathbf{b}/\gamma \right] / (1 + b^2/\gamma^2)
\]  
(6.43)

This equation is still incomplete because \( \gamma^{n+1} \) appears on the right hand side, so we need to carry out the second step, which is to solve for \( \gamma \). We can do that by dotting \( \mathbf{u}^{n+1} \) to equation (6.42) and make use of the definition of \( \gamma \), \( \gamma = \sqrt{1 + u^2} \):

\[
\gamma^2 = 1 + u^2 = 1 + \mathbf{u}' \cdot \mathbf{u} = 1 + \frac{u'^2 + (\mathbf{u}' \cdot \mathbf{b})^2/\gamma^2}{1 + b^2/\gamma^2}
\]  
(6.44)

This is a quadratic equation for \( \gamma^2 \) and we can solve it to find \( \gamma \). Finally one can plug this back into equation (6.43) to find the updated momentum \( \mathbf{u}^{n+1} \).

In our GR case things becomes more complicated. Due to the appearance of the \( \gamma_{ij} \) matrix in equation (6.40), the direction solution to the equation (6.43) does not work anymore. We still define the intermediate \( \mathbf{u}' \) as

\[
u_i' = u_i^n + \frac{\alpha q \Delta t}{m} \left( \gamma_{ij} D^j + \sqrt{\gamma} \epsilon_{ijk} \frac{\bar{\epsilon}^j}{2} B^k \right)
\]  
(6.45)

Now one needs to solve the following equation:

\[
\begin{pmatrix}
  \gamma_{ij} - \frac{\epsilon_{ijk} b^k}{\alpha \gamma_p} \\
\end{pmatrix}
\bar{\mathbf{p}}^j = u_i'
\]  
(6.46)

where \( b^k = (\alpha \sqrt{\gamma} q \Delta t/2m)B^k \). The difference is that \( \gamma_{ij} \) would have been \( \delta_{ij} \) in the case of flat space. However, this is still a 3 \( \times \) 3 matrix equation and one can invert it by brute force.

Assuming it is done, we can plug the equation into the relation of \( \gamma_p \) with \( \bar{\mathbf{p}}^i \) which is

\[
\alpha^2 \gamma_p^2 = 1 + \gamma_{ij} \bar{\mathbf{p}}^i \bar{\mathbf{p}}^j = 1 + \bar{\mathbf{p}}^i u_i' = 1 + \left( \gamma_{ij} - \frac{\epsilon_{ijk} b^k}{\alpha \gamma_p} \right)^{-1} u'_i u_j
\]  
(6.47)
At this point it is obvious we should call the combination $\alpha \gamma_p$ the convenient Lorentz factor $\bar{\gamma}_p$.

Again this turns out to be a quadratic equation for $\bar{\gamma}_p$, albeit more complicated. I plugged the matrix into Mathematica to find the full inverted result (which is surprisingly not that complicated):

$$\bar{\gamma}_p^2 = 1 + \frac{(u'_i b^i)^2 + \bar{\gamma}_p^2 \sigma}{\bar{\gamma}_p^2 \det \gamma + \gamma_{ij} b^i b^j}$$

(6.48)

where the symbol $\sigma$ is (assuming $\gamma_{ij}$ is symmetric, which is always true for a metric):

$$\sigma = u'_1 (\gamma_{22} \gamma_{33} - \gamma_{23}^2) + u'_2 (\gamma_{11} \gamma_{33} - \gamma_{13}^2) + u'_3 (\gamma_{11} \gamma_{22} - \gamma_{12}^2)$$

$$+ 2u'_1 u'_2 (\gamma_{13} \gamma_{23} - \gamma_{12} \gamma_{33}) + 2u'_1 u'_3 (\gamma_{12} \gamma_{32} - \gamma_{13} \gamma_{22}) + 2u'_2 u'_3 (\gamma_{21} \gamma_{31} - \gamma_{23} \gamma_{11})$$

(6.49)

It would be nice to simplify this $\sigma$ term further! However even at this form one should be able to code it into an algorithm. After one solves equation (6.48) for $\bar{\gamma}_p$ one can plug it into equation (6.46) to solve for $\bar{p}^i$. This is the modified Vay algorithm for curved space.

### 6.1.5 Effect of Gravity

Since we have written our particle push equation in complete form:

$$m \frac{Du_i}{Dt} = e \alpha (\gamma_{ij} D^j + e_{ijk} \bar{\gamma}^j B^k)$$

(6.50)

it would be nice to solve it completely accounting for GR effects as well! I think it can be done.

The covariant derivative on the left hand side can be expanded as

$$\frac{Du_i}{D\tau} = \frac{du_i}{d\tau} - \Gamma^\alpha_{\beta i} u_\alpha = \frac{du_i}{d\tau} - \frac{1}{2} \left( g_{i\alpha,\mu} + g_{\mu\alpha,i} - g_{i\mu,\alpha} \right) u^\alpha u^\mu$$

(6.51)

Two terms in the bracket are antisymmetric with respect to $\alpha$ and $\mu$ whereas $u^\alpha u^\mu$ is symmetric, so we only have

$$\frac{Du_i}{D\tau} = \frac{du_i}{d\tau} - \frac{1}{2} g_{i\alpha,i} u^\mu u^\alpha$$

(6.52)
and if we convert the time to coordinate time we have

$$\frac{Du_i}{Dt} = \frac{du_i}{dt} - \frac{1}{2} g_{\mu\alpha,i} v^\alpha g^{\mu\nu} u_\nu$$

(6.53)

The inverse metric $g^{\mu\nu}$ was inserted so that the right hand side also has lower indices $u_\nu$. One can totally pre-store the coefficients $g_{\mu\alpha,i} g^{\mu\nu}/2$ on the grid, then convolve with $v^\alpha$ to get the matrix in front of $u_\nu$.

Now to combine the Lorentz force with the gravitational correciton, one can do the so-called “drift-kick” trick. The idea is that, we want to separate a single timestep into two half-steps. In the first half-step, the particle simply drifts along the geodesic, and we have only

$$\frac{u_i^{n+\frac{1}{2}} - u_i^n}{\Delta t/2} = \frac{1}{2} g_{\mu\alpha,i} v^\alpha g^{\mu\nu} u_\nu$$

(6.54)

Since the right hand side is quadratic in $u$, we either use a fully explicit scheme, evaluating the right hand side at time step $n$, or we use a “semi-implicit” scheme and solve the following matrix equation

$$\frac{u_\mu^{n+\frac{1}{2}} - u_\mu^n}{\Delta t/2} = \frac{1}{2} g_{\alpha\beta,\mu} (v^\alpha)^n g^{\beta\nu} (u_\nu)^{n+\frac{1}{2}}$$

(6.55)

Note that to do this we should save $u_0$ as part of the dynamic variables now.

After we have “drifted” the particles for a half-step, we evaluate the “kick” on the particle due to electromagnetic force. This is computed using the modified Vay algorithm detailed in the above subsection. We solve the following equation

$$\frac{u_i' - u_i}{\Delta t} = \frac{\alpha q}{m} \left( Y_{ij} D^j + e_{ijk} \frac{\ddot{v}^j + \dot{v}^j}{2} B^k \right)$$

(6.56)

where the primed quantities are the momentum and velocity after the kick. One can then compute
CHAPTER 6. FURTHER EXPLORATIONS IN EXTENSION OF PIC

$u'_0$ from $u'_i$ using the constraint $u_\mu u^\mu = -1$. We have:

$$u_0 = \alpha \gamma p \left[ y_{ij} \beta^i \bar{\gamma}^j - \alpha \right] \quad (6.57)$$

Then we use these quantities to drift the particles for another half-step

$$\frac{u_\mu^{n+1} - u_\mu^{n+1/2}}{\Delta t/2} = \frac{1}{2} g_{\alpha \beta, \mu} (u^{\alpha})^{n+1/2} g^{\beta \nu} (u_\nu)^{n+1} \quad (6.58)$$

One question that may be asked is whether the energy of the particle is conserved during the update. This is on the list of tasks that we might go back to a bit later.

### 6.1.6 Current Deposition

One needs to define the current $I^\mu$ that occurs in the Maxwell equations (6.10) in terms of particle motion. Again following Landau & Lifshitz we do that by referring to the action (6.24). We introduce the charge density $\rho$ by summing the charge contained in a certain spatial volume element

$$de = \rho \sqrt{\gamma} dx^1 dx^2 dx^3 \quad (6.59)$$

Multiplying this by $dx^i$ we have

$$dedx^i = \rho dx^\mu \sqrt{\gamma} dx^2 dx^3 = \frac{\rho}{\alpha} \sqrt{-g} d\Omega \frac{dx^\mu}{dt} \quad (6.60)$$

where $d\Omega = dt dx^1 dx^2 dx^3$. We therefore define the current 4-vector as

$$I^\mu = \frac{\rho c}{\alpha} \frac{dx^\mu}{dt} \quad (6.61)$$

Charged particles are represented by delta function in the current and charge distributions.
Let’s define \( \int \delta(r) d^3x = 1 \), then the charge and current densities (Landau & Lifshitz) are

\[
\rho = \sum_p \frac{q_p}{\sqrt{\gamma}} \delta(r - r_p), \quad I^\mu = \sum_p \frac{q_p c}{\alpha \sqrt{\gamma}} \delta(r - r_p) \frac{dx^\mu}{dt} \quad (6.62)
\]

The continuity equation is \( \nabla_\mu I^\mu = 0 \) and using the divergence form (6.63) and the above definition we can see it is indeed satisfied. The form therefore suggests that in depositing the current the correct velocities we should use are \( \bar{v}^i \), definitely not \( \bar{v}^i \).

We can also find the integral form of the continuity equation. Since we have

\[
\frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} I^\mu) = 0 \quad (6.63)
\]

we can separate into spatial and temporal part, and integrate both over a given cell:

\[
\partial_t \int_C \alpha \sqrt{\gamma} I^0 d^3x + \int_C \partial_i (\alpha \sqrt{\gamma} I^i) d^3x = 0 \quad (6.64)
\]

We can use Gauss’s theorem to replace the second integral. Because \( \int_\Sigma d\omega = \oint_{\partial\Sigma} \omega \) we have

\[
\partial_t \int_C c \rho \sqrt{\gamma} d^3x + \oint_{\partial C} \frac{1}{2} e_{ijk} \alpha I^i dx^j \wedge dx^k = 0 \quad (6.65)
\]

Notice that the second integral is exactly the integral that occurs in the integral form of the Maxwell equations (6.22). Therefore if using integral formalism, we only need to separate the change of charge in a given cell into different directions, find the current flux easily and then directly use them to update the dynamic fields \( B^i \) and \( D^i \).

To find the current fluxes, we can again use the Esirkepov scheme described in section 2.2.2. The reason the algorithm is still applicable is that charge movement in one direction will only cause current on the cell surface in that direction, even when the cell is not rectangular.
6.1.7 Concluding remarks

The algorithms outlined above is a possible way to extend the PIC technique to a general relativistic regime. This could potentially lead to new findings about plasma dynamics very close to the black hole, which is very non-trivial and have only been studied in either FFE or GRMHD (e.g., Komissarov 2004). It is known that current sheets naturally form in the black hole ergosphere, and to study the reconnection and possible particle acceleration there, a full GR PIC code is required. This is a very promising work in progress that in the long run might open a variety of new research opportunities.
Bibliography


—. 2009, The Astrophysical Journal, 703, 1044

—. 2011, Astrophysics and Space Science Proceedings, 21, 299

Carlqvist, P. 1982, Astrophysics and Space Science, 87, 21


Erber, T. 1966, Reviews of Modern Physics, 38, 626

Esirkepov, T. Z. 2001, Computer Physics Communications, 135, 144


—. 2013, Aristotelian Electrodynamics solves the Pulsar: Lower Efficiency of Strong Pulsars, arXiv e-print 1303.4094


Hockney, R. W., & Eastwood, J. W. 1981, Computer Simulation Using Particles


—. 1985b, Astronomy and Astrophysics, 144, 72


Longair, M. S. 2011, High Energy Astrophysics


Lyubarskij, Y. E. 1992, Astronomy and Astrophysics, 261, 544


Mazets, E. P., Golentskij, S. V., & Guryan, Y. A. 1979a, Soviet Astronomy Letters, 5, 641


Melrose, D. B. 1986, Instabilities in Space and Laboratory Plasmas
Mereghetti, S. 2008, Astronomy and Astrophysics Review, 15, 225


—. 2004, Advances in Space Research, 33, 542


Pétri, J. 2007, Astronomy and Astrophysics, 469, 843


Pierce, J. R. 1954, Theory and Design of Electron Beams


Spitkovsky, A. 2004, in , 357


Swanson, D. G. 2003, Plasma Physics and Controlled Fusion, 45, 1069

Tajima, T. 1989, Computational plasma physics. With applications to fusion and astrophysics.


Turolla, R., Zane, S., & Watts, A. L. 2015, Reports on Progress in Physics, 78, 116901

Umeda, T., Omura, Y., & Matsumoto, H. 2001, Computer Physics Communications, 137, 286

Umeda, T., Omura, Y., Tominaga, T., & Matsumoto, H. 2003, Computer Physics Communications, 156, 73


Yee, K. 1966, Antennas and Propagation, IEEE Transactions on, 14, 302

Appendix A

Higher order finite-difference schemes

In this appendix we summarize the way to extend the finite difference scheme described in chapter 2 to higher orders. We assume symmetrically staggered Yee grid as described in section 2.2.1. In 1D the position where the derivative is evaluated is staggered by half a cell with respect to where the function value is evaluated (figure A.1).

\[
\begin{align*}
    f\left(x - \frac{3\Delta x}{2}\right) &= f(x) - \frac{3\Delta x}{2} f'(x) + \frac{1}{2!} \left(\frac{3\Delta x}{2}\right)^2 f''(x) - \frac{1}{3!} \left(\frac{3\Delta x}{2}\right)^3 f'''(x) + O(\Delta x^4) \quad (A.1) \\
    f\left(x - \frac{\Delta x}{2}\right) &= f(x) - \frac{\Delta x}{2} f'(x) + \frac{1}{2!} \left(\frac{\Delta x}{2}\right)^2 f''(x) - \frac{1}{3!} \left(\frac{\Delta x}{2}\right)^3 f'''(x) + O(\Delta x^4) \quad (A.2) \\
    f\left(x + \frac{\Delta x}{2}\right) &= f(x) + \frac{\Delta x}{2} f'(x) + \frac{1}{2!} \left(\frac{\Delta x}{2}\right)^2 f''(x) + \frac{1}{3!} \left(\frac{\Delta x}{2}\right)^3 f'''(x) + O(\Delta x^4) \quad (A.3) \\
    f\left(x + \frac{3\Delta x}{2}\right) &= f(x) + \frac{3\Delta x}{2} f'(x) + \frac{1}{2!} \left(\frac{3\Delta x}{2}\right)^2 f''(x) + \frac{1}{3!} \left(\frac{3\Delta x}{2}\right)^3 f'''(x) + O(\Delta x^4) \quad (A.4)
\end{align*}
\]

Figure A.1: Positions where the function is evaluated and its derivative is evaluated in 1D.

Now if we would like to find the 4th order accurate symmetric numerical derivative, then we need to solve the following set of equations for \(f'(x)\):
The solution is (note that the 4th order term automatically cancels by symmetry, the expression is accurate to 4th order in $\Delta x$):

$$f'(x) = \frac{9}{8} f(x + \Delta x/2) - f(x - \Delta x/2) - \frac{1}{24} f(x + 3\Delta x/2) - f(x - 3\Delta x/2) + O(\Delta x^5) \quad (A.5)$$

One can repeat this procedure for $f'(x)$ up to any order in $\Delta x$. Table A.1 summarizes the coefficients for the results up to 6th order accuracy.

<table>
<thead>
<tr>
<th>Order</th>
<th>$f\left(x - \frac{5\Delta x}{2}\right)$</th>
<th>$f\left(x - \frac{3\Delta x}{2}\right)$</th>
<th>$f\left(x - \frac{\Delta x}{2}\right)$</th>
<th>$f\left(x + \frac{\Delta x}{2}\right)$</th>
<th>$f\left(x + \frac{3\Delta x}{2}\right)$</th>
<th>$f\left(x + \frac{5\Delta x}{2}\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4th</td>
<td>0</td>
<td>0/24</td>
<td>-9/8</td>
<td>9/8</td>
<td>-1/24</td>
<td>0</td>
</tr>
<tr>
<td>6th</td>
<td>-3/640</td>
<td>25/384</td>
<td>-75/64</td>
<td>75/64</td>
<td>-25/384</td>
<td>3/640</td>
</tr>
</tbody>
</table>

Table A.1: Summary of 1D finite difference operator coefficients for evaluating $f'(x)$.

Note that to achieve accuracy in terms of numerical truncation error, we sacrifice locality: more and more adjacent grid points are used to evaluate the local derivative. This has implications for communication, as the number of guard cells for each boundary needs to be expanded to match the amount of information that is needed to evaluate the derivative: for 2nd order derivative we only need 1 guard cell, for 4th we need 2, and for 6th we need 3 layers of guard cell, etc.

The number of guard cells is not the only problem however. When solving the continuity equation (2.24) the finite difference divergence operator used must match the one used in updating the Maxwell equations. If we use the above 4th order solution for evaluating curls in the Maxwell equations, then we need to use the same 4th order operator for the continuity equation. To my knowledge, there is no way to accomplish this using the Buneman current deposition scheme (section 2.2.2), but it is possible using the Esirkepov scheme. After one splits the continuity equation into components, we now have:

$$\frac{9}{8} [j_x(x + \Delta x/2) - j_x(x - \Delta x/2)] - \frac{1}{24} [j_x(x + 3\Delta x/2) - j_x(x - 3\Delta x/2)] = \frac{\Delta x}{\Delta t} \Delta \rho_x(x) \quad (A.6)$$
Now it is not possible to solve these equations using a simple prefix sum, since the current at different grid positions are coupled. This becomes a set of $N$ linear equations where $N$ equals the number of grid points in that direction. It involves a band-diagonal matrix with alternating coefficients, which is known to have unstable properties when solving numerically. However, instead of solving for $j$ directly we can define $\Delta j_i = j_i - j_{i-1}$ where $i$ labels the grid points. Now the system of equations becomes

$$-\frac{1}{24} \Delta j_{i-1} + \frac{13}{12} \Delta j_i - \frac{1}{24} \Delta j_{i+1} = -\frac{\Delta x}{\Delta t} \Delta \rho_i$$

(A.7)

which is a symmetric tri-diagonal system that is diagonally dominant. This kind of systems has standard solvers and is generally numerically stable. After one solves $\Delta j$ for every grid point, a prefix sum can be done similar to the 2nd order case to obtain $j$ for each cell. For 6th order finite difference, the matrix becomes a band-diagonal system with 5 non-zero diagonals:

$$\frac{3}{640} \Delta j_{i-2} - \frac{29}{480} \Delta j_{i-1} + \frac{1067}{960} \Delta j_i - \frac{29}{480} \Delta j_{i+1} + \frac{3}{640} \Delta j_{i+2} = -\frac{\Delta x}{\Delta t} \Delta \rho_i$$

(A.8)

which is also relatively simple to solve numerically.

Boundary condition is another problem. The main difficulty is the lack of information across the boundary of the simulation domain, which means that the typical symmetric finite difference cannot be used. A one-sided finite difference scheme is required. Figure [A.2] shows two scenarios for evaluating the one sided derivatives. One can find the correct coefficients by writing down similar equations as (A.1) and the solutions are given in tables [A.2] and [A.3]

<table>
<thead>
<tr>
<th>Order</th>
<th>$f(x)$</th>
<th>$f(x + \Delta x)$</th>
<th>$f(x + 2\Delta x)$</th>
<th>$f(x + 3\Delta x)$</th>
<th>$f(x + 4\Delta x)$</th>
<th>$f(x + 5\Delta x)$</th>
<th>$f(x + 6\Delta x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-3/2</td>
<td>2</td>
<td>-1/2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-25/12</td>
<td>4</td>
<td>-3</td>
<td>4/3</td>
<td>-1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>-147/60</td>
<td>6</td>
<td>-15/2</td>
<td>20/3</td>
<td>-15/4</td>
<td>6/5</td>
<td>-1/6</td>
</tr>
</tbody>
</table>

Table A.2: One-sided derivatives for $f'(x)$ at the domain boundary. Scenario (a) in figure [A.2]
Figure A.2: Several scenarios for evaluating one-sided derivatives.

Table A.3: One-sided derivatives for \( f'(x) \) near the domain boundary. Scenario (b) in figure A.2
Appendix B

GPU architecture and CUDA

B.1 The GPU Architecture

In the recent decade we saw the rise of Graphics Processing Units (GPUs) as an alternative way to carry out general purpose computations. GPUs are originally designed specifically for intensive floating point calculations that are very common in video games and 3D rendering. For these applications, typically the goal is to compute the color of a pixel on the screen, where each pixel is relatively independent from others. In order to compute millions of pixels every frame and keep up with more than 60 frames per second, GPUs are designed to be massive parallel machines. A modern GPU typically has hundreds to thousands of cores, compared to up to less than 20 cores in a top-of-the line CPU. In terms of raw floating point operations per second, a single modern GPU can break 5 TetraFLOPs per second which not that many years ago was only achievable by supercomputers. In addition, the GPU on-board memory has usually 5-6 times more bandwidth than the ordinary DRAM, which further improves the computation throughput especially for memory bound codes.

However, the many cores of the modern GPU come at a price, as each is less competent than a core of an equivalent CPU. The GPU cores are organized into symmetric multiprocessors (SMPs), and all the cores in an SMP share one control unit while each core has only an ALU (Arithmetic
Logic Unit) which is used to perform arithmetic and logic operations (figure B.1). This means that the same statement is carried out not only by one core but by a group of cores, on a chunk of data at the same time. At the time of this writing, the popular Nvidia K40 processor is the most common one found in local and national clusters. It has 15 SMPs each with 192 CUDA cores, giving a total of 2880 cores.

This particular architecture is perfect for operations that naturally call for parallelism, but it poses a serious programming challenge for general tasks. For example, when an algorithm requires iterations where each step depends on the previous one (e.g. numerical integration of an ODE), there is no way to do it in parallel, and the GPU structure falls back to executing everything in series, which is very slow due to the inherently lower clock speed. Another problem lies deeper in the architecture itself. The many cores in an SMP are organized into “warps” of typically 32 cores each. An instruction is passed to a warp and executed by all 32 cores simultaneously regardless of the content of the instruction or any conditional statements. If an if statement is part of the instruction and segregates the warp into two branches, then the two branches will both be executed by all threads, with inappropriate results discarded. The worst case is when all 32 instances belong to different conditional branches, the same instructions will be executed 32 times, effectively erasing any parallelization.

The GPU version of Aperture is developed using the CUDA parallel computing platform, created by Nvidia Corporation, and designed to be executed on Nvidia GPUs. The CUDA platform
is a software layer that gives direct access to the GPU’s virtual instruction set and parallel computational elements, for the execution of compute kernels. CUDA is designed to work with programming languages such as C, C++, and Fortran. Since Aperture was developed in C++, it is natural to blend CUDA into the code base. In the remaining part of this appendix, we will describe the execution model of CUDA and outline how I designed Aperture to take advantage of this model.

B.2 The CUDA execution model

As mentioned in section B.1, GPU relies on massive parallelization for its speed advantage over ordinary CPUs. This is explicitly built into the execution model of CUDA. Functions allowed to run on the GPU are called compute kernels, and are annotated with the keyword \texttt{__global__} in front of the function definition. When a compute kernel is launched, a set of additional parameters is given specifying the number of threads that will execute the kernel in parallel. It is very typical to launch hundreds of thousands of threads at the same time.

Threads in CUDA are organized into thread blocks. Each thread block can have at most 1024 threads. Multiple blocks can be launched at the same time, and there is no upper limit on the number of blocks launched. At execution time, thread blocks are assigned to SMPs. On K40 up to 16 blocks can be sent to an SMP at a time, and each block is executed in warps of 32 threads. For example, if a compute kernel is launched with 128 blocks of 512 threads each, in reality up to 64 randomly chosen warps from 16 blocks can be executed at the same time for each SMP, and it is all up to the scheduler in the SMP to decide at runtime.

Apart from the warp execution model, each thread block also has a small pool called “shared memory” that is accessible by all threads inside the block. This piece of shared memory is similar to the L1 cache in ordinary CPUs, which has very low latency compared to the main memory. In addition, threads in a given block can write to the shared memory \textit{atomically}, avoiding race

\footnote{In fact, the user can choose how the 64KB is split between L1 cache and the shared memory.}
condition. The same operation is much more costly and difficult to do on the main memory. In K40 each block can have up to 64KB of shared memory. If all threads in a block access a common part of the main memory repeatedly, loading that part of main memory into shared memory first can accelerate the memory access speed by a factor of more than a hundred.

In general, memory speed is an important aspect of the GPU program design. The different levels of memory accessible to a program forms a hierarchy in terms of speed and latency. The system RAM is the slowest, since every access from the GPU need to go through the PCIe bus, which has a latency of ~ 1000 cycles and a bandwidth of only a fraction of the system RAM bandwidth. The on board GPU memory (known as global memory in CUDA) does not suffer from this issue, and due to design its bandwidth is much higher than the system RAM, however it still suffers from very high latency: an access to the GPU main memory requires ~ 200 clock cycles. A faster but very limited memory space is the L1 cache/shared memory as mentioned above. An access to the shared memory only has latency of a few clock cycles, and multiple threads reading the same address can be done in a single operation. Figure B.2 shows the GPU memory hierarchy and the perspective memory bandwidth/latency.

Therefore the optimal strategy is to either keep all the computation data on GPU main memory, or to overlap memory copy with computation as much as possible. When processing data on GPU main memory, try to manipulate a local block of data at a time by loading them to the shared memory of a thread block, do the calculation in parallel compute kernels, and then save them back to the GPU main memory.
Figure B.2: Memory hierarchy of Nvidia GPUs. Shared memory only have a latency of $\sim 1$ clock cycle. Global memory (GPU on-board memory) has a latency of $\sim 100$ cycles, and the system RAM has an access latency of more than 1000 cycles and relatively low total bandwidth.

### B.3 Parallelization and optimization in Aperture

PIC codes are naturally very suited for GPUs due to the algorithm being readily parallelizable. A PIC code typically deals with millions to billions of particles, each relatively independent of each other especially for collisionless plasma. On a grid scale, the Maxwell solver is also typically parallel, each cell only requires the information of a few adjacent cells and updates independently. Moreover, GPU excels at floating point operations which is the main kind of arithmetic operations used in a PIC code. The only potential problem for parallelization lies in current deposition, where by definition multiple particles need to be processed and then write the result to the same cell, which can only be done serially.
Particle pusher

Most of the particle pusher algorithm is completely parallelizable. Given the values of $E$ and $B$ fields at the particle location, every particle can be processed independently to update their momenta and positions. This lends itself well to the parallel structure of GPUs. The only nontrivial optimization is to interpolate the field values to particle position, which involves frequent random access to the field array which is not a strength of the GPU architecture.

The way we optimize this part of the code is to subdivide the whole computational domain in “tiles”, such that the 6 field components of each tile can fit into the shared memory of a thread block. In 2D simulations we found that $8 \times 8$ tiles works very nicely. In the compute kernel, a thread block is assigned to every tile in the domain, and the threads first load the $E$ and $B$ field values from global memory into the shared memory of the block. Then the threads in the block work through the particles inside the tile in parallel, interpolating the field values to the local particle position using the values in the shared memory, then update the particle momentum using Vay pusher (section 2.2.3).

A small sacrifice for this algorithm is that we need to have the particles sorted by tile all the time. Therefore at each time step we sort them immediately after particle move and production of new pairs, since both operations mess up the particle order: movement between tile boundary, and because all new particles are added to the end of the particle array. However we found this to be the best way since current deposition also benefits from having a sorted particle array (see B.3).

Current deposition

Current deposition poses a main problem for a massively parallelized architecture like the GPU, since by definition many threads (particles) need to write the same memory location (current values on the grid), which cannot be done in parallel. The only thing we can do is to use duplicates and minimize clashing memory access as much as possible.

The current deposition in Aperture is handled as follows. Similar to particle pusher, current
deposition delegates the work for each tile to a thread block. Every thread block maintains a temporary array in the shared memory to store the deposited current values. Every thread in the block processes one particle at a time, and uses atomic add to add the deposited result from each particle to the array. To minimize clashing, each thread \( i \) in the block will go through the \( M \) cells in the tile from \( i \mod M \) to \( (i + M) \mod M \). On top of this, since there is some headroom in shared memory, each thread block actually holds 4 different temporary arrays, and particles are evenly split to deposit to the 4 different arrays, only adding up the results in the end.

![Figure B.3: 4-color scheme for current deposition to avoid overlapping of guard cells. Each square represents a tile, and tiles of the same color are processed at the same time. Since no two neighboring tiles are of the same color, there is no issue of guard cell overlapping.](image)

Since during one time step, particles might move across tile boundaries, therefore current deposition needs to be done with a layer of guard cells around each tile. However this creates a problem where multiple thread blocks may want to update overlapping regions of the global memory at the same time. The way around this issue is to color all tiles in the domain in 4 different colors (figure): 4 different kernels are invoked in sequence, each only updating the tiles with one single color. This avoids any memory access issue between thread blocks and has minimal impact on performance as long as the domain is large enough, so that each kernel call saturates the compute pipeline.
**Pair creation**

Pair creation poses a similar problem as current deposition. Since new particles are appended to the end of the main particle array, every photon that is to convert to a pair will need to access the end of the array and race condition may occur unless the process is serialized.

To avoid race condition, we need to pre-calculate the final position of the created photon/pair in the particle array. This is done using a two pass scan of the particle array. During the first pass, the compute kernel examines every particle in parallel, maintaining a temporary array that marks whether a particle will emit a photon or a photon will convert to a pair during this timestep. For every photon emitting particle, the thread will atomically add 1 to a temporary variable marking the total number of photons emitted in this tile, then store this number in the corresponding position in the marker array, which will act as the position of the resulting photon in the main photon array. Then during the second pass, actual photons will be added to the main array in parallel, using the values in the marker array as offsets.

![Diagram](image)

**Figure B.4:** Thematic illustration for parallel pair creation. Each square represents a particle and the partition represents a tile. During the first pass the compute kernel checks for photon-emitting particles, and mark them with the number within the tile (first two rows). Then a prefix sum is carried out on the number of emitting particles in each tile. During the second pass, the index of each emitting particle is added by the number of emitting particles in the proceeding tiles, which now represents the absolute position of the photon in the final array.