Antiproton and positron dynamics in antihydrogen production

by

Chukman So

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jonathan S. Wurtele, Chair
Professor Joel Fajans
Professor Phillip Colella

Fall 2014

Antiproton and positron dynamics in antihydrogen production

Chukman So

Abstract

Antiproton and positron dynamics in antihydrogen production

by

Chukman So

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Jonathan S. Wurtele, Chair

The asymmetry between matter and antimatter in the universe and the incompatibility between the Standard Model and general relativity are some of the greatest unsolved questions in physics. The answer to both may possibly lie with the physics beyond the Standard Model, and comparing the properties of hydrogen and antihydrogen atoms provides one of the possible ways to exploring it. In 2010, the ALPHA collaboration demonstrated the first trapping of antihydrogen atoms, in an apparatus made of a Penning–Malmberg trap superimposed on a magnetic minimum trap. Its ultimate goal is to precisely measure the spectrum, gravitational mass and charge neutrality of the anti-atoms, and compare them with the hydrogen atom. These comparisons provide novel, direct and model–independent tests of the Standard Model and the weak equivalence principle. Before they can be achieved, however, the trapping rate of antihydrogen atoms needs to be improved.

This dissertation first describes the ALPHA apparatus, the experimental control sequence and the plasma manipulation techniques that realised antihydrogen trapping in 2010, and modified and improved upon thereafter. Experimental software, techniques and control sequences to which this research work has contributed are particularly focused on. In the second part of this dissertation, methods for improving the trapping efficiency of the ALPHA experiment are investigated. The trapping efficiency is currently hampered by a lack of understanding of the precise plasma conditions and dynamics in the antihydrogen production process, especially in the presence of shot–to–shot fluctuations. This resulted in an empirical development for many of the plasma manipulation techniques, taking up precious antiproton beam time and resulting in suboptimal performance. To remedy these deficiencies, this work proposes that simulations should be used to better understand and predict plasma behaviour, optimise the performance of existing techniques, allow new techniques to be explored efficiently, and derive more information from diagnostics.

A collection of numerical models for Penning–Malmberg trap plasmas are introduced, which are designed to simulate a major subset of the plasma manipulation techniques used in ALPHA, targeted at the plasma conditions available therein, and with near–real–time experimental usability in mind. The first of these is a zero–temperature plasma solver, which exploits the water bag model to compute the density and potential of a cold, stationary plasma with a given radial profile and electrode excitations. It is suited to analysing slow (or stationary) processes, where the variations applied are on a much slower time scale than the typical time between collisions in the plasma. The density and electric potential output by the solver inform the programming of the electrode voltages, which is of particular value when plasma bunches need to be weakly confined in shallow wells.

The second numerical model developed for this work is a radially–coupled Vlasov–Poisson solver, which evolves the axial phase space distribution of a plasma under the influence of (time-dependent) electrode excitations, from a given initial state. It takes into account the plasma self–field and the radial variations in potential and density, and assumes that radial transport is negligible. This model simulates processes where the dynamic behaviour of the plasma is critical to their outcome. It allows for tests of plasma manipulation techniques over a wide range of tunable parameters and plasma conditions prior to an actual experiment, potentially reducing the need for empirical tuning.

The third numerical model is an azimuthally averaged, energy–conserving Fokker–Planck solver for a discrete, non-regular grid distribution. It simulates the effects of weakly magnetised collisions on the bulk parallel and perpendicular velocity distributions of a plasma, as the particles collide among themselves. The collision coefficients are analytically calculated by azimuthally averaging the derivatives of the Rosenbluth potentials. This model is applicable to plasmas where self–collisions of antiprotons have a non-negligible effect, possible examples of which include the antiproton–positron mixture which exists during antihydrogen formation, and the antiproton cloud captured from the Antiproton Decelerator, the source of ALPHA's antiprotons.

The fourth numerical model is an azimuthally averaged Fokker–Planck model for intermediately magnetised collisions. It generalises the preceding model to study Fokker–Planck–type collisions of electrons, positrons and antiprotons in magnetic fields of arbitrary strength. Unlike the previous model, analytic solutions for collisions in arbitrarily strong magnetic fields are not known. The collision coefficients are therefore computed numerically via an adaptive Monte Carlo averaging of the colliding particles' changes in parallel and perpendicular velocities, over their impact parameter and their velocity phase angles. The collision process itself is simulated via a variable–time–stepping Boris particle pusher. This model is applicable to a wide range of processes involving cooling and thermalisation, which are critical to the ALPHA experiment.

The water bag and Vlasov models are employed to simulate the excitation of antiprotons during the antiproton–positron mixing process, which produces antihydrogen atoms and determines whether they can be confined by the magnetic minimum trap. The agreement

between the simulation and experimental measurements, analytic predictions and other existing simulations is demonstrated. The simulation is then used to optimise the excitation under various plasma conditions, and novel excitation techniques are proposed and explored.

The models developed throughout this work lay the foundation for a systematic analysis of the plasma phenomena in the experiment. Future work includes extending the result of the mixing simulation to study collisional and recombination effects, as well as applying the models to other processes in the experiment. It is also of interest to apply the collisional formulations in this work to particle–in-cell (PIC) models and to explore three–dimensional plasma effects.

To my mum and dad,
for whom I had not been the easiest kid to raise

Contents

List of Figures

Acknowledgments

I would like to take this opportunity to express my sincere gratitude to the members of the Wurtele–Fajans group at Berkeley for their invaluable knowledge and instruction throughout the course of this work. I would in particular like to thank my advisers Prof. Joanthan Wurtele and Prof. Joel Fajans for their encouragement and accommodation of my work at ALPHA, without which I would not have explored the world of experimental physics. My numerical works has also been made possible by Prof. Wurtele's generous provision of computational resources, both in local hardware and computer time on the Lawrencium supercluster at Lawrence Berkeley National Laboratory. For my friends Marcelo Baquero-Ruiz, Alex Povilus and Dr. Andrey Zhmoginov in Berkeley, I offer my earnest appreciation of the advices and knowledge they offered.

I would also like to thank all of ALPHA's members for their patience and the skills they shared. The perseverance they displayed has been an inspiration. In particular, the experience and wisdom of Dr. Will Bertsche, Dr. Paul Bowe, Dr. Eoin Butler, Dr. Makoto Fujiwara, Prof. Jeff Hangst and Dr. Niels Madsen made my time at CERN infinitely more fruitful and worthwhile. The company and camaraderie of my friends and colleagues Gorm Bruun Andersen, Mohammad Ashkezari, Marcelo Baquero-Ruiz, Tim Frisen, Andrea Gutierrez, Chris Orum Rasmussen and Joseph McKenna in the AD also made the long overnight shifts and laborious work bearable, and at times even worth looking forward to.

Chapter 1

Introduction

The ALPHA experiment, along with other antihydrogen experiments, aims to produce antihydrogen atoms and measure their physical properties in an attempt to discover discrepancies between matter and antimatter. This is essential, as the universe is matter–dominated according to astronomical observations, which implies that a level of asymmetry between matter and antimatter must have contributed to the Big Bang's process of equilibrating matter–antimatter and energy. However, laboratory– and accelerator–based experiments have so far failed to observe the level of asymmetry required. Moreover, the theory which summarises these particle physics experiments, the Standard Model, is incompatible with Einstein's theory of gravity. It is therefore essential for experiments to probe for the physics beyond the Standard Model, and to discover new asymmetries between matter and antimatter, to further our understanding of fundamental physics.

The core of the ALPHA apparatus is made of a Penning–Malmberg trap, which confines charged particles with a uniform axial magnetic field and an electric field created by a stack of independently controlled, hollow cylindrical electrodes. Antiprotons and positrons are captured, manipulated and cooled in the Penning–Malmberg trap, and subsequently mixed in a region of the electrode stack surrounded on the outside by a pair of mirror coils and an octupole magnet. These magnets form an magnetic minimum trap, which confines neutral particles via their magnetic moment by creating a magnetic minimum. The antihydrogen atoms created from the mixing of cold antiprotons and positrons come into being inside the magnetic minimum trap. Those anti-atoms with sufficiently low kinetic energy are confined by it. The long lifetime of the trapped anti-atoms allows them to decay from their highly excited states at recombination to their ground states. This in turn should allow their physical properties, such as spectrum and mass, to be precisely measured.

While ALPHA has demonstrated the long–duration trapping of antihydrogen atoms, the numbers obtained to date is not conductive of precision measurements. A higher number of trapped antihydrogen atoms is desirable as it improves the signal–to–noise ratio of the measurements, and reduces the impact of machine fluctuation. To improve the trapping rate, the production method of antihydrogen must be made more efficient. In this chapter,

we first give a brief summary of the history of antihydrogen research, the basic principle of the Penning–Malmberg trap on which the ALPHA experiment (as well as many other antihydrogen and plasma experiments) is based, the obstacles we face in achieving precision measurements, and how we propose to overcome them through plasma simulation. In Ch. 2 we describe in depth the hardware and experimental control sequences of the experiment, to provide a full picture of the processes we can simulate and improve. Many of the hardware and sequence developments were also made with our contribution. In Ch. 3 to 6 we outline the numerical models that can simulate various important aspects of the plasma physics in the experiment. In Ch. 7 we apply these models to one of the most critical and sensitive parts of the experimental sequence — the mixing of antiprotons and positrons — to demonstrate the capability our simulation models offer.

1.1 A history of antihydrogen research

Discovery of antimatter

The earliest scientific mention of the concept of antimatter dates surprisingly from the 1880's, before the era of quantum mechanics and relativity. Theoretical physicists attempted to explain the action–at–a–distance of classical gravity and electromagnetism with hydrodynamical models of the ether, an all–permeating, tenuous substance that mediates these forces [1]. For instance, Carl Pearson proposed the "ether squirt" theory in the 1880's, which suggested that ether emerges from particles, and these gravitational and electromagnetic forces originate from the interaction between these outflows. To conserve the total amount of ether, sinks for the ether must also exist which would behave in an opposite manner to normal particles under gravity and electromagnetism. He called these sinks "antimatter". The physicist William Hick also raised the concept of antimatter in the same period in his ether vortex theory. Naturally their classical attempts were made obsolete by Max Planck, whose proposal of the quantisation of photons in an attempt to explain the black body radiation spectrum ushered in the era of quantum physics. The concept of antimatter had fallen into obscurity.

In the early 1930s, the concept of antimatter emerged as we know it today due to the English physicist Paul Dirac. In 1928 he proposed an extension to Schrödinger's quantum wave equation [2] to take into account special relativistic effects. His theory explained the magnetic moment of the 1/2–spin electron, and gave accurate prediction of the fine structure splitting of the hydrogen spectrum. However, in addition to the electron solution, his equation had a solution for a positively charged particle with the same mass. Robert Oppenheimer and Igor Tamm further proved [3] in 1930 that this positively charged particle and an electron can combine and annihilate, leaving behind two energetic photons. Initially reluctant, Dirac finally identified this positive particle to be a new type of particle [3] in 1931, which he called an "anti-electron". It was also realised later that the Dirac equation predicted the presence of an antimatter partner to every particle with a half–integer

spin (fermions). These anti-particles would have the same spin and mass as their matter counterparts, but with opposite charges. A particle and a corresponding anti-particle would annihilate each other when put in close contact, leaving behind highly energetic photons. Conversely, photons with sufficient energy can create a particle–antiparticle pair out of the vacuum, a process called pair creation.

Soon after Dirac's prediction of a new positive particle, experimentalist Carl Anderson discovered [4] Dirac's particle in 1932 in a cloud chamber experiment measuring cosmic rays, proving the existence of antimatter. Anderson named this particle the positron, a convention still used today. With the advent of modern accelerators in the 1950s and the ability to accelerate particles to ever higher energies, other more massive antiparticles have been artificially produced. Antiprotons were first produced [5] at the Bevatron at Lawrence Berkeley National Laboratory in 1955, by smashing a beam of 6.2 GeV protons into a copper target, and detecting the scattered products in a bubble chamber. Antineutrons were produced at the same accelerator [6] in 1957, by first colliding the proton beam into a beryllium target to produce antiprotons. These antiprotons were subsequently allowed to interact with protons inside another downstream target. Some of them underwent a charge exchange collision, which led to antineutrons.

Astronomical observations

Parallel to these developments in quantum and particle physics, breakthroughs were also made in astronomy in the 1920s [7]. In 1924 the American astronomer Edwin Hubble measured the distances of what was then known as "spiral nebulae" by observing the relative brightness of a type of variable star in these nebulae. These variable stars, the Cepheid variable, have a highly regular correlation between their absolute brightness and pulsation period, and acted as standard candles with which Hubble deduced their distances. He proved that these spirals were indeed distant galaxies just like our own, and by comparing the measured distance of these galaxies to their receding speeds (which were measured earlier by Vesto Slipher in 1912 through the Doppler shift of their spectra), he discovered that more distant galaxies receded more rapidly. The universe is expanding. There were several competing theories to account for this expansion, but in 1964 astronomers Arno Penzias and Robert Wilson discovered the cosmic microwave background radiation exactly as predicted by the Big Bang model, hence confirming its validity. In the Big Bang model, the primordial universe started as a rapidly expanding microscopic point with immense energy and temperature. Matter and radiation existed in equilibrium through pair creation and annihilation. The expansion caused the temperature of this equilibrium to decrease, and fundamental particles of various masses became "frozen out" at various points in time, when the energy of the radiation was no longer sufficient to create a pair of that species. Eventually the radiation decreased to below the energy scale of the lightest species — electrons and positrons — and radiation and matter became fixed. The universe continued to expand, and the matter collapsed around fluctuations in density due to gravitational instability and formed all the

astronomical bodies we see today. The radiation on the other hand was "stretched" by the expansion of the universe, ever decreasing in temperature, and became the 2.7 K microwave background radiation observed by Penzias and Wilson.

The exact composition of the matter resulting from the Big Bang depends on pair creation, annihilation and decay dynamics. This allows particle models to be examined against astronomical observations of the matter composition of the universe. For instance, a particle model that is symmetric around matter and antimatter requires them to be present in exactly equal numbers in our universe. However, attempts to observe the astronomical presence of antimatter though the detection of either the cosmic ray from antimatter stars and galaxies (c.f. the BESS experiment [8]) or the activity along the interface between matter and antimatter domains have yielded no evidence of the missing antimatter. This indicates that some particle physics processes must have created an asymmetry between matter and antimatter in the early universe, in a process called baryogenesis. A successful particle model must provide mechanisms that can adequately explain baryogenesis and the composition of the universe.

More recent observations concerning the spinning of spiral galaxies, gravitational lensing around galaxies, the detailed structure of cosmic microwave background and the accelerating expansion of the universe indicate that a significant amount of gravitational mass and energy is present in our universe, but not directly observed. The lack of direct observation of these dark matter and energy means they must either be extremely weakly interacting, or highly scarce. This adds to the challenge of constructing a particle model which can explain the composition of the universe. A model has to account for these so far unobserved matter and energy, in addition to the observed (matter) stars and galaxies. Numerous experiments are currently being carried out in an attempt to detect these dark matter and energy, or to set bounds on the possible contribution from various particle species.

The Standard Model

There are three basic symmetries in fundamental physics: parity (P), charge (C) and time (T). An interaction is considered parity–symmetric if the physical law governing it remains the same upon a reversal of all coordinate vectors. Similarly, a charge–symmetric interaction remains invariant upon flipping the signs of all charges, and a time–symmetric interaction stays unchanged upon the reversal of time. These symmetries are observed in almost all common interactions, but Andrei Sakharov showed [9] in 1967 that there are at least three necessary conditions for baryogensis: 1. That there exist interactions which violate baryon number conservation, causing an excess of one type of matter to emerge from a net balance between matter and antimatter; 2. That C and CP–symmetries are violated, meaning the interactions favouring matter creation are not completely cancelled by their C or CP conjugates, which favour antimatter creation; 3. That these interactions happen out of thermal equilibrium, so as to prevent those matter-favouring interactions from being cancelled by their CPT conjugates.

The first observation of an interaction which violates at least some of these symmetries was carried out in 1956 by C. S. Wu et al. [10], regarding the beta decay of a ^{60}Co atom. They cooled and aligned cobalt atoms in a strong magnetic field, and observed whether there is any preference between the two possible decay modes (Fig. 1.1), which are mirror images of each other. If parity is observed in the weak interaction, decays should happen in equal frequency in both modes, but Wu et al. only observed electrons whose spin is anti-parallel to the cobalt atom's (mode 1 in Fig. 1.1), thus proving parity can be violated. However the joint CP symmetry is still obeyed in this interaction.

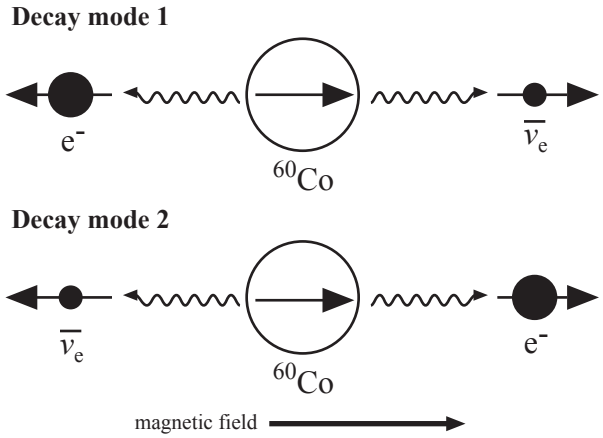

Figure 1.1: Two possible decay modes of a ^{60}Co atom.

Subsequently in 1964, Christenson et al. observed [11] the first CP–violating interaction in the decay of neutral K–mesons. The states of definite half–life for K–meson, K_L and K_S, are superpositions of the particle eignestates K^0 and $\bar{K}^0$. If CP symmetry is valid in K–meson's decay, the particle and antiparticle eigenstates of K–meson should be symmetric, and K_L and K_S should equal $(K^0 - \bar{K}^0)/\sqrt{2}$ and $(K^0 + \bar{K}^0)/\sqrt{2}$ respectively. The anti–symmetric superposition of K_L prevents it from decaying into two pions, which K_S is capable of, leading to the latter's much shorter half–life. However, Christenson et al. observed the rare two–pion decay of K_L, thus showing that K^0 and $\bar{K}^0$ are not exactly CP–symmetric.

The Standard Model is a quantum field theory which unifies electromagnetism, weak and strong interactions, and it is the state–of–the–art model in particle physics. It embeds the discovery of these (and many other) symmetry–breaking interactions and provides mechanisms to explain their existence. Every laboratory– and accelerator–based experiment so far is in agreement with the prediction of the Standard Model, which illustrates its extraordinary success.

However, the Standard Model is not a "theory of everything" — it has deficiencies both in its cosmological implications and incompatibility with gravity. The level of C and CP symmetry–breaking present in the Standard Model is well below that which is required to explain the predominance of matter over antimatter. Indeed according to the Standard

Model the amount of matter imbalance in the universe after Big Bang can at most only form "a single cluster of stars" [3]. The species and quantity of matter predicted by the Model also fail to explain the recent indirect observation of dark matter and dark energy. Moreover, the Standard Model is incompatible with Einstein's general theory of relativity, which on its own is highly successful in describing the fourth fundamental force of gravity. These are some of the most profound unsolved questions in physics, and their answers require the study of physics beyond the Standard Model. In particular, laboratory– and accelerator–based evidence of the break–down of the Standard Model or general relativity would help identify the way in which they are incomplete, and elucidate the more fundamental, unified law.

Tests of the Standard Model

One way to explore the physics beyond the Standard Model is to look for new CP or CPT violations, or to set the bounds thereof, in the four fundamental interactions. Numerous experiments have been or are currently being conducted to this end, each looking for violations in different particles and interactions. A few examples are as follow:

Mesons

Decays of accelerator–produced mesons are used to search for CP violations in weak interactions, since the energy scale in these machines means electromagnetic and gravitation interactions are relatively unimportant, and these particles are short–lived. For instance the KTeV experiment at Fermilab [12] and the NA48 experiments at CERN [13] discovered CP violations in rare K–meson decays in 1999; in 2001, the BaBar experiment at SLAC [14] and the Belle experiment at KEK [15] in Japan discovered CP violations in B–meson decays; and In 2011, the LHCb experiment at CERN discovered possible indications of CP violations in D–meson decays [16]. The violations discovered in these experiments so far fall within the predictions of the Standard Model.

Positron

A precision measurement of the positron magnetic moment can be compared to that of the electron, and deviation between the two would indicate a violation of CPT symmetry in the electromagnetic interaction of positrons. The Gabrielse group at Harvard is currently constructing a device for this purpose [17].

Antiproton

The TRAP experiment at CERN performed [18] a precision measurement of the charge–to–mass ratio of antiprotons in 1999 via a magnetic spectrometer, and compared it to that of proton, confirming their equality (and thus CPT compliance) within experimental limits. Subsequently the APEX experiment at Fermilab showed [19] in 2000 that the half–life of antiproton must be longer than 800,000 years with 90% confidence, again showing no CPT

violations up to this level. The ATRAP experiment at the AD at CERN measured [20] the magnetic moment of antiprotons in 2013, and showed it is compatible to that of proton within experimental error.

Positronium

Positronium is a short–lived bound system between an electron and a positron, which, much like the hydrogen atom, possesses various excited states and a spectrum. Since the electron and positron are structureless fundamental particles in the Standard Model, its spectrum is simple and can be predicted highly accurately. CPT violations would be revealed if its spectrum is measured to be different from the prediction of the Standard Model. M. Deutsch and S. Brown first measured [20] the Zeeman and hyperfine splitting of positronium in 1952, in agreement with the Standard Model prediction. However its short life time ($\sim$ 100 ns) prevented highly precise measurements.

Antihydrogen

The antihydrogen atom is a bound state between an antiproton and a positron. Intrinsically stable, its lifetime is only limited by vacuum and confinement quality. It is electrically neutral theoretically (see below for experimental measurement), and is made entirely of antimatter. These properties mean a precision measurement of its spectrum is feasible, since antihydrogen can be accumulated and cooled in a much longer time scale compared to positronium. Given that the spectrum of hydrogen has been measured [21] to extraordinary precision, and the Standard Model prediction of hydrogen (and therefore antihydrogen) spectrum is also highly precise, the antihydrogen spectrum offers a sensitive, model–independent test of CPT symmetry in a purely antimatter electromagnetic interaction. The ALPHA experiment has demonstrated [22] the first measurement of the hyperfine splitting of the antihydrogen atom in 2012, and the ALPHA, ASACUSA and ATRAP experiments at CERN are all attempting to perform precision spectral measurement on antihydrogen atoms.

The antihydrogen atom can also be used to test the gravitational weak equivalence principle. This principle states that the dynamics of a point mass in a gravitational field is solely determined by its mass, and independent of its composition and internal structures. In other words, under general relativity, matter and antimatter should behave in exactly the same manner. If antimatter is observed to behave differently under gravity compared to normal matter, general relativity must be incomplete. Non–neutral particles are not suitable for gravitational tests since any stray electromagnetic fields would overwhelm their gravitational responses, and the position or momentum of decaying systems cannot be sufficiently altered by gravity within their life–time compared to their initial spread. Antineutrons are difficult to manipulate due to their neutrality. This leaves the antihydrogen atom as the simplest candidate for antimatter gravity tests. The ALPHA experiment has established [23] a first bound on the gravitation mass of the antihydrogen atom in 2013, by measuring the gravitation bias of antihydrogen trajectories in a magnetic minimum trap during its shut–off.

The AEgIS experiment at CERN aims [24] to perform a measurement on the free fall of antihydrogen atoms in a moiré deflectometer in the near future. The GBAR experiment, also at CERN, aims [25] to directly measure the vertical free fall of ultra-cold ($\sim 20\mu$K) antihydrogen atoms down a vertical space a few tens of centimetres.

The electrical neutrality of the antihydrogen atom also serves to compare the charges of antiprotons and positrons, which, according to the Standard Model, should be exactly opposite. The presence of any fractional charge on an antihydrogen atom would open up physics beyond the Standard Model. This can be tested by measuring the response, if any, of antihydrogen atoms to a strong electric field, which is relatively simple and highly sensitive given the ease of creating a strong electric field and the stability of the anti-atom. The ALPHA experiment has set [26] a bound on the fractional charge of antihydrogen atom in 2014, consistent with the Standard Model's prediction of neutrality.

Producing antihydrogen

Relativistic

Antihydrogen was first produced in the LEAR antiproton decelerator in 1995 by Oelert et al. [27]. In that experiment, a beam of antiprotons was targeted at a cluster of Xenon atoms. Some of the antiprotons would collide with the Xenon nuclei, and part of that energy is consumed by the pair creation of electrons and positrons (among other possibilities). Some of these positrons would in turn be captured by antiprotons and form a highly–excited bound state. These anti-atoms were then detected in a magnetic spectrometer where the neutral species would show no deflection. These antihydrogen atoms were travelling at relativistic speeds and lasted for as long as the time taken to cover their beam path to the spectrometer, which is less than 100 ns.

Non–relativistic

In order to precisely measure the properties of antihydrogen, they must be at a sufficiently low temperature, which cannot be achieved with antihydrogen atoms moving at relativistic speeds. To overcome this, the ATHENA and ATRAP experiments used separate sources of antiprotons and positrons, cooled them individually to cryogenic temperatures in a Penning–Malmberg trap, and combined them to form low temperature antihydrogen. This methodology ensures the antiprotons, relatively easy to manipulate due to its charge, are as cold as possible before recombination with a positron. ATHENA first produced slow–moving antihydrogen in 2002 [28], followed by ATRAP in the same year [29]. In ATHENA, antiprotons from the AD were degraded using a thin foil to increase their energy spread, the slowest fraction of which were captured in a Penning–Malmberg trap using high voltage gates. They were subsequently cooled sympathetically with preloaded electrons. On the other end, positrons emitted from a radioactive ^{22}Na source were captured using a buffer gas, and accumulated in a Surko–type accumulator [30]. These positrons were then transferred to the

Penning–Malmberg trap and combined with the antiprotons to form slow–moving antihydrogen atoms. The anti-atoms were no longer confined due to their neutrality, and drifted to the Penning trap walls and annihilated. The annihilation products were detected in a silicon vertex detector and identified as the signature of an antihydrogen atom.

1.2 The basics of the Penning–Malmberg trap

Penning–Malmberg traps are used by ALPHA, ATRAP, AEgIS, and many plasma and ion trapping experiments to manipulate charged particles. These traps are compatible with the ultra–high vacuum (UHV) environment necessary for long–term antimatter storage, and the cryogenic temperatures necessary for particle cooling. A Penning–Malmberg trap is composed of a uniform solenoidal magnetic field and a stack of hollow cylindrical electrodes forming a long tube, aligned along the direction of the magnetic field (Fig. 1.2 a). The voltage of each electrode is individually controlled through external circuitry.

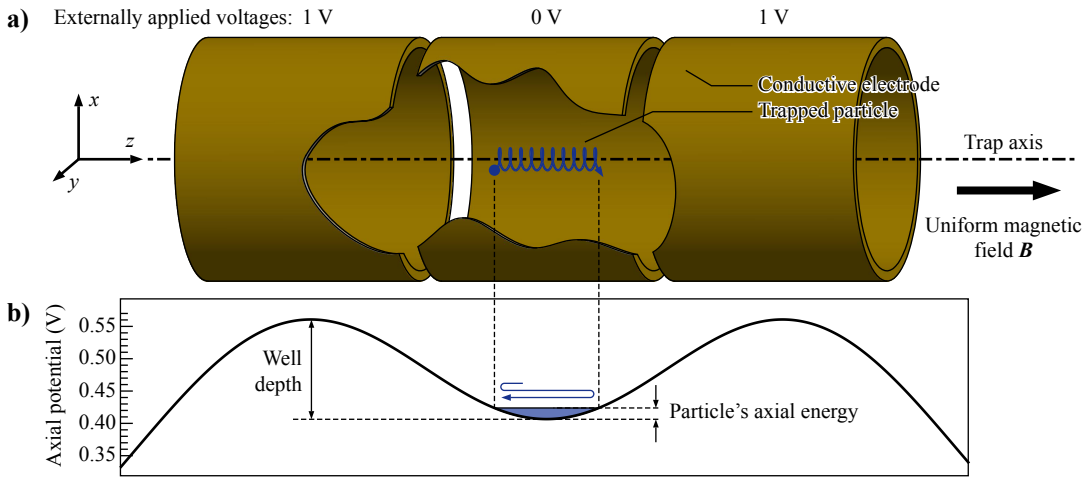

Figure 1.2: A schematic view of a basic three–electrode Penning–Malmberg trap, showing the a) physical geometry and b) the potential along the trap axis. A single particle is shown trapped.

Single particle dynamics

When voltages are applied to the electrodes in a Penning–Malmberg trap, they create an electrostatic field within the bore volume. Particles are usually confined near the axis of the electrodes, and in that region the field has a negligible radial component compared to the axial component. Ignoring collective effects, particle motion near the axis is simply given by the Lorentz force law and Newton's second law:

$$\begin{cases} \dot{\boldsymbol{v}}_\perp &= \omega_C \boldsymbol{v}_\perp \times \hat{\boldsymbol{z}} \\ \ddot{z} &= \dfrac{q}{m} E_z(z) \end{cases}$$

where $\boldsymbol{v}_\perp$ is the perpendicular velocity in the x-y plane as labelled in Fig. 1.2. The uniform magnetic field is given by $\boldsymbol{B} = B_0\hat{\boldsymbol{z}}$, and $\omega_C \equiv qB_0/m$. The parallel and perpendicular degrees of freedoms are completely decoupled, with the perpendicular motion a circular one at the cyclotron angular frequency ω_C, and the axial motion that of a free particle moving according to the axial electric field. The confinement of particles in a Penning–Malmberg trap therefore works in two parts: 1. Radially, particles are prevented from venturing far outward since trajectories are deflected into circles by the magnetic field; 2. Axially, the electrodes are biased to create an electrostatic well which deflects the axial movement of the particles towards the potential minima (Fig. 1.2 b). As long as the axial energies of the particles do not exceed the maximum deflection strength of the potential well (the well depth), they remain confined.

If the particle is not situated exactly on the trap axis, it experiences a radial component in the electric field. This radial component must point outward when positive particles are confined, since the Laplace equation requires the axial minimum of the potential to also be a radial maximum. The radially outward electric field causes the centre of the cyclotron motion (the gyrocentre) to drift in the $\hat{\boldsymbol{\theta}}$ direction at a velocity of $\boldsymbol{v}_D = E_r\hat{\boldsymbol{r}} \times B_0\hat{\boldsymbol{z}}/B_0^2$, but the gyrocentre remains approximately confined at a fixed radius. The accuracy of this approximation relies on the cyclotron radius being much smaller than the radius of the electrodes, such that the electric field across one cyclotron orbit is mostly uniform. This is usually the case in a Penning–Malmberg trap, given the strong magnetic field and low–temperature particles, both decreasing the size of the cyclotron orbit. In the ALPHA apparatus, the cyclotron radius of a typical positron and antiproton are $\sim 2 \times 10^{-7}$ m and $\sim 2 \times 10^{-5}$ m respectively, which are much smaller than the electrode radius $\sim 2 \times 10^{-2}$ m, the scale of variation for the electric field.

The cyclotron motion of the particle also leads to cyclotron radiation, causing the perpendicular energy of the particle to decrease. This radiative cooling continues until the power emitted via radiation is balanced by the power absorbed from the radiation in the environment. The cooling power is given by the Larmor formula for radiation [31]

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\frac{\mu_0 q^2 a^2}{6\pi c} = -\frac{\mu_0 q^4}{3\pi c}\frac{B^2}{m^3}E.$$

Here E is the kinetic energy contained in the perpendicular gyromotion, which decays with a time scale proportional to B^2/m^3. Therefore, the radiative cooling process is more rapid for lighter particles in stronger magnetic fields. Note that in an azimuthally symmetric trap, there is no single–particle mechanism for the cyclotron radiation to effect cooling to the axial movement, since these two degrees of freedom are decoupled (up to the uniform E-field approximation above). Collisions are required to help transfer the cooling effect to the axial degree of freedom.

Space charge and Debye shielding

The leading effect of having more than one particle in the trap is the mutual electrostatic repulsion of the particles, which tends to counter the confinement and eject the particles. The Penning–Malmberg trap, however, is able to confine multiple particles even in the collective regime. The outward acceleration due to the axial self–field, as well as the axial energies of the particles, is balanced by the compressive force due to the external field applied through the electrodes. The radial self–field, on the other hand, enhances the net radially electric field, and results in a more rapid $\boldsymbol{E} \times \boldsymbol{B}$ drift in the $\hat{\boldsymbol{\theta}}$ direction (compared to the single–particle case). However, as long as the scale of variation of the net radial electric field is much bigger than the size of the cyclotron orbits of the particles, the $\boldsymbol{E} \times \boldsymbol{B}$ drift still guarantees the gyrocentres do not move radially. The overall result is an ellipsoidal "cloud" of particles (plasma) confined along the trap axis. Each particle undergoes cyclotron motion while bouncing axially within the electrostatic trap formed by the combination of the external and self–field, and the gyrocentre drifts in the $\hat{\boldsymbol{\theta}}$ direction.

In steady–state and sufficiently deep inside the plasma, the net axial electric field (the sum of the self–field and external field) must be zero since the net axial particle flux is, by the definition of "steady state", zero, and particles move freely in $\hat{\boldsymbol{z}}$. This means the total potential $\phi(r, z)$ is only a function of r within the plasma and far away from the surface. Near the surface, and along a line of fixed radius, ϕ increases as one moves away from the bulk of the plasma. This increasing potential provides the axial deflecting force that keeps particles trapped. The more energetic particles can move further outward against this potential before eventually being turned back. This means the density of a plasma gradually decreases to zero near the surface as the total potential rises. Since the potential on the plasma surface varies like $\sim qn_0z^2/\epsilon_0$ (n_0 being the bulk density of the plasma), a typical particle with charge q travelling with an axial energy of k_BT (T being the plasma temperature) would stop within a distance of $\sqrt{\epsilon_0 k_B T/(q^2 n_0)}$ into the sheath. This distance, the characteristic thickness of the transitional sheath, is known as the Debye length λ_D. The net axial electric field vanishes within a few Debye lengths into the plasma.

The idea described above is quantitatively modelled by the Poisson–Boltzmann equation

$$\begin{cases} \nabla^2\phi(r,z) &= -\dfrac{qn(r,z)}{\epsilon_0} \\ n(r,z) &= \mathcal{N}(r)\exp\left(-\dfrac{q\phi(r,z)}{k_BT}\right), \end{cases} \tag{1.1}$$

where $n(r, z)$ is the number density of the particles, $\phi(r, z)$ is the net electric potential, and $\mathcal{N}(r)$ is the normalisation factor which sets the amount of particles at each radius. The Poisson equation uses the charge distribution to deduce the potential, while the Boltzmann factor describes how a group of thermal particles would arrange themselves according to a potential — locations of higher potential are exponentially less populated since particles are exponentially unlikely to have the energy to reach these locations. The Poisson–Boltzmann

equation is a non-linear differential equation due to the exponential Boltzmann factor, and has no general analytic solution. We will study the numerical solution to the equation in the limit of $T \to 0$ in Ch. 3.

Confinement

The Penning–Malmberg trap is well–suited to the long–term confinement of particles; antimatter particles are routinely confined for hours or even days [32]. This long lifetime is due to the conservation of the total canonical angular momentum of the plasma [33]

$$P_\theta = \sum_i^N \left(m r_i v_{\theta i} + \frac{q B_0 r_i^2}{2} \right) \approx \frac{q B_0}{2} \sum_i^N r_i^2.$$

The approximation on the last step is under the condition that the angular momentum due to the magnetic vector potential is much greater than the kinetic part, in the strongly magnetised limit. The angular momentum must remain conserved as long as the system is rotationally symmetric (i.e. no source of external torque), regardless of collisions. This requires $\sum r_i^2$ to remain constant, which means the plasma cannot diffuse radially outward.

1.3 Future experimental objectives and obstacles

The ALPHA apparatus has demonstrated an average trapping rate of one antihydrogen atom per attempt, each taking approximately 15 minutes. Measurements of some physical properties have been obtained from such a low trapping rate because the silicon vertex detector provided spatially and temporally resolved antihydrogen detection down to a single–atom sensitivity. However, we do not expect the precision achieved so far in these measurements to be able to discern the minute asymmetry between matter and antimatter, should any exist. The precision of the three types of measurement currently being pursued — spectral, gravitational and charge — depends critically on the number and temperature of the trapped anti-atoms:

- For a spectral measurement, a low antihydrogen temperature reduces the thermal motion of the anti-atoms, and decreases the Doppler spread of the spectrum (for the transitions susceptible to Doppler effects), making the lines more sharply defined.
- For a deflectometer–type gravitational measurement (e.g. the AEgIS experiment [24]), a low antihydrogen temperature helps reduce the thermal spread of the antihydrogen beam and allows a sharper moiré pattern to be resolved using finer gratings.
- For a trap escape–based gravitational or charge neutrality measurement, antihydrogen atoms with a lower temperature are more sensitive to small external forces. This allows

any external influence on their escape trajectory to be more clearly detected. For instance, in a gravitational measurement, a lower–temperature antihydrogen population would result in fewer anti-atoms being ejected from the trap in the "wrong" upward direction by simply having sufficient inertia to overcoming gravity. Any biasing of the mean vertex height would therefore be more prominent.

- Having a higher number of the anti-atoms in the trap increases the signal in all the above experiments, which helps to overcome background noise, reduce the relative impact of systematic error, and decrease the number of false signals from the cosmic ray background. These false signals, which originate mainly from cosmic muons scattering off the trap structure while passing through the experiment, occur at a constant rate and cannot be distinguished from antihydrogen annihilations by the detector.

- A stronger signal means data accumulation can be performed across fewer cycles of the experiment, which reduces the chance of machine fluctuations impacting the results.

Improving the trapping efficiency is a challenging task, given the difficulty in trapping antihydrogen. A number of factors have impeded improvement in the trapping rate:

- The greatest percentage loss of useful antiprotons occurs during the mixing of antiprotons and positrons, in which only one anti-atom is trapped, on average, out of the $\sim 10^4$ antiprotons used in the mixing process. Improving this efficiency requires reducing the temperature of the antiprotons and positrons, as well as improving the mixing scheme to minimise the velocity in which antiprotons are introduced into the positron plasma (which is currently achieved by an autoresonant [34] excitation of the axial oscillation of the antiprotons). The former is limited by cryogenic technology, electrical noise and other factors that are not yet fully understood, while the latter is difficult due to the fact that the self–potential of the plasmas is much greater than the 0.5 K magnetic minimum trap depth. Any small, shot–to–shot fluctuation in the particle numbers in either plasma can create a potential misalignment of $O(10)$ mV or higher. If a mixing scheme fails to account for this fluctuation, the energy gained by an antiproton traversing from the antiproton to the positron plasma can be of $O(100)$ K or higher, which eliminates any chance of producing a trappable anti-atom.

- Diagnostic access to a Penning–Malmberg trap plasma is limited by the trap geometry. For instance, the multi-channel plate imaging device, which measures the radial distribution of the particles by capturing the charges ejected axially from the trap, cannot provide any axial information. The temperature diagnostic, which deduces the axial temperatures by ejecting particles through a slow shallowing of the confining electrostatic well and correlating their escape timing to the well depth, does not yield the radial temperature. The self–potential of plasmas, knowledge of which is essential during the mixing of antiprotons and positrons, also cannot be measured directly.

- The special techniques used in the experimental sequence, like the rotating wall or evaporative cooling (Sec. 2.9), involve numerous tunable parameters, and are in most cases developed empirically. This means their development is time–consuming and the tuning is often incomplete due to the large parameter space. When new plasma conditions arise (due to machine fluctuations or improvement in prior manipulations), redevelopment is often required. This results in a slow development cycle that adds complexity to the experiment.

- Not all antiprotons available from the AD are captured for antihydrogen production. The AD delivers an antiproton bunch every ~ 100 s, while a production cycle of the apparatus typically takes 15 minutes. That means only one in about nine shots is captured, with the remainder rejected.

- Improving trapping yield by using more intense antiproton and positron bunches is difficult due to shot–to–shot fluctuations. The fluctuation of their self–potential grows with the number of particles, making an accurate, minimal energy mixing more difficult.

In order to make the most use of the antiproton available from the AD, ALPHA is commissioning a new apparatus with significant modifications to its structure and modes of operation (Fig. 1.3). Antiproton capture and storage has been moved to a separate Penning trap — the catching trap — which enables more long-term stacking and storage of antiprotons, independent of the positron and antihydrogen production cycle. A fraction of the stored antiprotons are to be ejected on demand, and delivered to the mixing trap for antihydrogen production. This uncouples the productions cycle from the AD, allowing more antiprotons to be used in each cycle, and potentially permits production while the AD is unavailable.

These innovative features necessitate new modes of operation for the apparatus and present new challenges. A longer–distance transfer of antiprotons between the catching trap and the mixing trap is necessary. Loading and removing cooling electrons are potentially obstructed in the catching trap due to the presence of a previously stored plasma. Diagnostic access is likewise limited. Experimentally it was observed that antiproton accumulation saturates at ~ 1 M antiprotons. Keeping the stored antiprotons at a low temperature and high density is a potential obstacle. The more intense antiproton bunches available to the mixing trap will require different manipulation techniques in light of its enhanced space charge.

To adapt the established antihydrogen trapping sequence to, and exploit the capabilities of the new hardware design, the detailed plasma dynamics for various techniques used in the Penning-Malmberg trap must be better understood. This requires a degree of predictive power for determining how the optimum parameters for a given technique vary with the plasma conditions (such as density or temperature). Such predictive capability reduces the parameter space that needs to be explored by experimental tuning, identifies the limits of

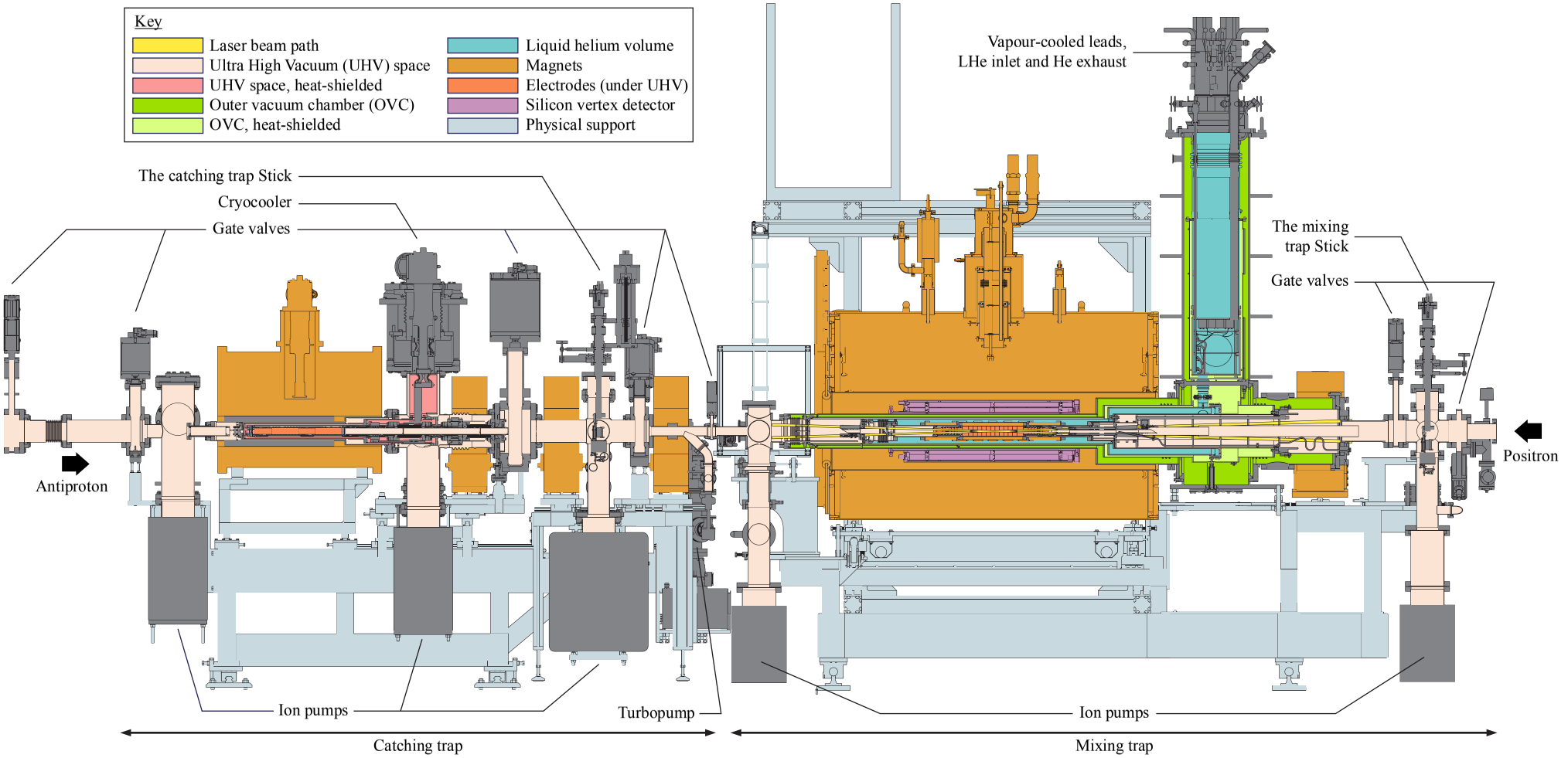

Figure 1.3: The structure of the ALPHA–2 apparatus. The Positron Accumulator is located on the right of the apparatus and is not shown.

a process, and helps determine when new techniques must be sought. Diagnostics like the multi–channel plate imaging and temperature measurement can be better analysed and may provide more information on the state of plasmas.

1.4 Overview of plasma simulations

The dynamics of a plasma is dictated by the interaction between the distribution of particles and the electromagnetic field, which is created by their charges and any externally controlled electrodes and magnets. At the most fundamental level, the system is described by the Lorentz force law and the Maxwell equations. The latter is usually simplified into the Poisson equation and a radiative cooling Larmor formula, assuming the particles are non-relativistic, which is mostly the case in antihydrogen experiments. Solving these equations for an N–particle plasma is, however, highly impractical due to the number of degrees of freedom and widely diverging spatial and temporal scales of the motion. It is therefore essential to separate the physics by their spatial and time scales, and develop separate models for them. We can then select and combine the models relevant to a process of interest, and construct a simulation that yields physically interesting results and is computationally manageable.

The first such separation concerns the force felt by each particle. This force can be divided into two components: the strong, sudden electrostatic force when another point charge passes nearby, and the much weaker bulk force due to other more distant particles in the plasma, the electrodes and the magnetic field. In this separation, we assume the chance of three (or more) particles all being in close proximity at the same time is negligible compared to the that of binary collisions, and that the collisions are uncorrelated to each other. Since the bulk force originates from the bulk plasma, it varies on a bigger spatial scale compared to the microscopic collisions. This means that nearby particles in the plasma all behave similarly under its influence. This allows the us to model the plasma as a *smooth distribution* instead of individual, singular particles. The evolution of the distribution is driven by the bulk forces, calculated self–consistently as a smooth field, as well as the statistical average of the collisional influences.

Further separation of the bulk motion is possible. Under the plasma conditions available in ALPHA, the following are the most important of the separated regimes of motion:

- **Cyclotron motion**
  The cyclotron motion in the r and θ directions is typically the fastest motion in the trap. For positrons this typically has a period of $\sim 4 \times 10^{-11}$ s, and for antiprotons $\sim 7 \times 10^{-8}$ s. Due to the high frequency of the cyclotron motion, very few perturbations address this degree of freedom. For perturbations at a much lower frequency than the cyclotron motion, the magnetic moment of the particles' cyclotron motion is conserved.

- **Axial bounce**
  In the z–direction particles move freely under the influence of the smoothed macro-

scopic electric field, which is created by the smoothed charge distribution and the boundary condition imposed by the trap electrodes. For positive (negative) particles the electrodes are biased such that a minimum (maximum) potential exists along the z axis. Positrons in a typical trap have an axial bounce period of $\sim 3 \times 10^{-7}$ s, and for antiprotons it is $\sim 3 \times 10^{-6}$ s. If an external perturbation is applied close to this time scale, the particles would respond dynamically and be accelerated in the z direction. This behaviour can be described by the Vlasov model, which resolves the detailed time evolution of the particles' phase space distribution subject to the action of time–varying forces (see Ch. 4). In contrast, if the perturbation is much slower than the axial bounce frequency ("much slower" being on the order of, or much longer than, the mean free time between collisions), these particles have enough time to equilibrate with the slow perturbation and behave quasi-statically. This is described by the Poisson–Boltzmann model (see Ch. 3).

- **E×B drift**
  The gyrocentres of these particles moves in the θ direction around the trap axis due to the E×B drift. The azimuthal motion only depends on the radial electric field and is uncoupled from the axial motion. A typical rotation period around the trap axis due to the E×B drift is $\sim 1 \times 10^{-5}$ s for positrons, and $\sim 7 \times 10^{-5}$ s for antiprotons. This motion is solely determined by the net perpendicular electric field.

- **Self–collision**
  Collisions lead to a gradual relaxation of the parallel (z) and perpendicular (r and θ) velocity distributions into thermal equilibrium, both within and between them. The relaxation process has different rates between the parallel and perpendicular directions, since the axial magnetic field introduces directionality to the microscopic collision process. The mean free time between collisions is $\sim 8 \times 10^{-7}$ s for positrons, and $\sim 7 \times 10^{-4}$ s for antiprotons. For phenomena with a time scale much shorter than the mean free time, collisional effects can be ignored. For phenomena with a time scale similar to the mean free time, the effect of collisions on the particles must be taken into consideration. Depending on the strength of the magnetic field, the effect of these collisions are described by either the weakly (Ch. 5), intermediately (Ch. 6) or strongly magnetised collision operator. If the phenomenon happens on a time scale much longer than the mean free time, one can assume there is sufficient time for the plasma to equilibrate through collision between each infinitesimal change, and the plasma is approximately always in equilibrium.

- **Radial transport**
  The bulk motion in the r direction is slow compared with all the other motions described above, driven by collisions with background gas, contaminating ions or broken azimuthal symmetry in the trap.

- **Radiative cooling**
  The electromagnetic interaction of each point charge to its own field, which is relativistic in nature, is not negligible in the very longest time scale. It leads to a radiative loss of energy in the perpendicular directions in the form of cyclotron radiation, taking energy away from the cyclotron motion. The cooling time scale for positrons is ~ 3 s, and $\sim 2 \times 10^{10}$ s for antiprotons [35].

In the case of a multi-species plasma, e.g. an antiproton–electron mixture or a antiproton–positron mixture, there are extra physics that result from the interaction between the two species:

- **Inter–species collision**
  Collisions between the two species can transfer momentum from one species to another, and can lead to a equilibration between their temperatures. The effect of this type of collision is also dependent on the magnetic field strength. The intermediately magnetised collision operator (Ch. 6) is designed for inter–species collision as well as self–collisions.

- **Recombination**
  When oppositely charged species collide, they sometimes combine into a bound–state. The initial antihydrogen momentum is determined by the antiproton momentum at the instant of recombination. Instead of being under the influence of the Penning–Malmberg trap, the new neutral anti–atom is now under the influence of the magnetic minimum trap. Recombination cannot happen in free–space due to the conservation of momentum; it has to either lose some of that momentum by emitting a photon (a radiative capture) or to a nearby third–party, e.g. another positron (a three–body recombination).

In this work we have developed four models: an axial equilibrium "water bag" model which solves for plasma shapes, an axial Vlasov–Poisson model which dynamically simulates the axial motion of a plasma distribution, and two Fokker-Planck models which give the statistical influence of collisions on distributions under two magnetisation regimes. They aim to cover the axial bounce and collisional regimes in the list above. While this is by no means exhaustive, these models are capable of simulating a diverse range of plasma manipulation techniques in ALPHA when suitably combined. Figure 1.4 shows the time scales of the motion regimes above, and the range of applicability of the models we have developed. For instance, the autoresonant excitation of antiprotons during the mixing of antiprotons and positrons has a frequency of ~ 250 kHz (the frequency of the antiproton axial oscillation) and lasts for ~ 1 ms, which puts it at a time scale between the two dashed red lines in Fig. 1.4. This means the positrons are best modelled with the Poisson–Boltzmann equation, as the self–collision of positrons is sufficient to ensure the positrons stay in quasi-equilibrium. The antiprotons are best modelled with the Vlasov equation, and the self–collision of the

antiprotons is mostly negligible. The two species interact through collision and recombination once the antiprotons and positrons overlap. The application of the models to the mixing process is further developed and tested in Ch. 7. Note that while the self–collisional effect of antiprotons seems negligible from Fig. 1.4, we have still developed its model, for two reasons. First, we are most interested in the slowest–moving antiprotons from which the trappable antihydrogen atoms form, and collisional effects are more significant for these slow–moving antiprotons. Second, the mean free time of antiprotons is inversely proportional to their density, which means an increase of antiproton density (for which the ALPHA-2 apparatus was designed) would also increase collisional effects.

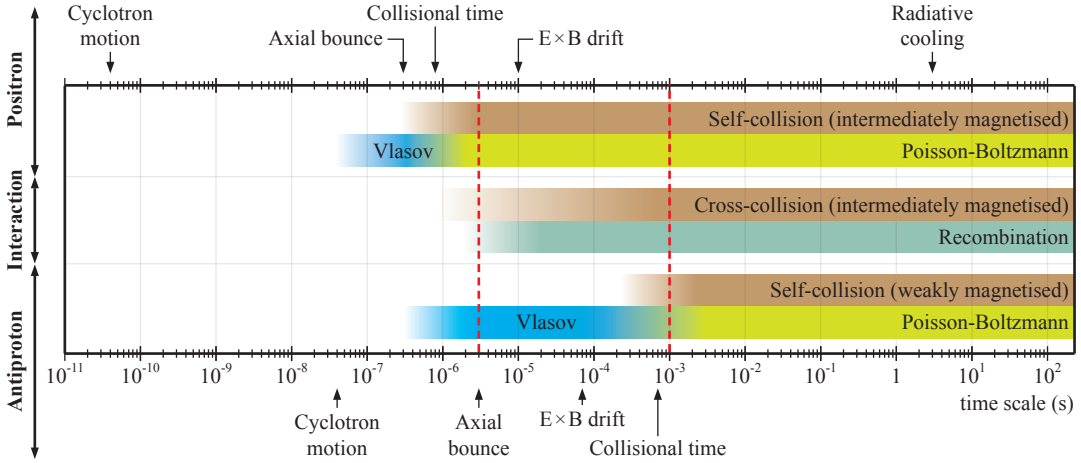

Figure 1.4: A schematic diagram of the time scales of various plasma physics dynamics in the ALPHA experiment, and the applicable range of time scales of various physics models. The periods of the cyclotron motion, axial bounce and E×B drift, as well as the mean free time between collisions and radiative cooling time scales for both species are marked on the axis.

Chapter 2

The ALPHA experiment

The main core of the ALPHA apparatus is functionally a multi–electrode Penning–Malmberg trap, with an magnetic minimum trap superimposed around the central region (see Fig. 2.1). Physically, an outer superconducting solenoid generates a uniform magnetic field required for the Penning–Malmberg trap. The bore of the solenoid is occupied by three layers of cylindrical structures which perform the majority of the experiment's functions: the outer silicon vertex detector, the middle cryostat, and the central ultra–high vacuum (UHV) chamber. The silicone vertex detector identifies the annihilation of antihydrogen atoms. The cryostat contains and cools the two mirror coils and the octupole coil which create the magnetic minimum trap. The UHV chamber contains the electrodes for the Penning–Malmberg trap, and is where the antiparticles are manipulated. The UHV chamber is open on both ends. A positron accumulator is situated to the right to provide positrons, while the Antiproton Decelerator (AD) beamline injects antiprotons from the left. A vertically articulated assembly of small devices (the Stick) can be placed onto the beamline between the main trap and the positron accumulator to perform various diagnostic functions, as well as to introduce electrons or radiation into the trap. The operation of these components are controlled by a sequencing system.

In the following we give a description of each of these components of the experiment, and outline its operation in a typical antihydrogen production cycle.

2.1 The outer solenoid

A warm bore superconducting solenoid manufactured by Kurchatov Institute [36] forms the outermost structure of the ALPHA apparatus, which generates a uniform magnetic field (maximum field strength of 1 T) within the bore and along its axis. The solenoid is kept at a superconducting temperature of 4.2 K by externally filled liquid helium and shielded by a liquid nitrogen jacket and a vacuum layer. The magnet is capable of persistent operation where the current in the solenoid windings forms an enclosed loop internally and does not

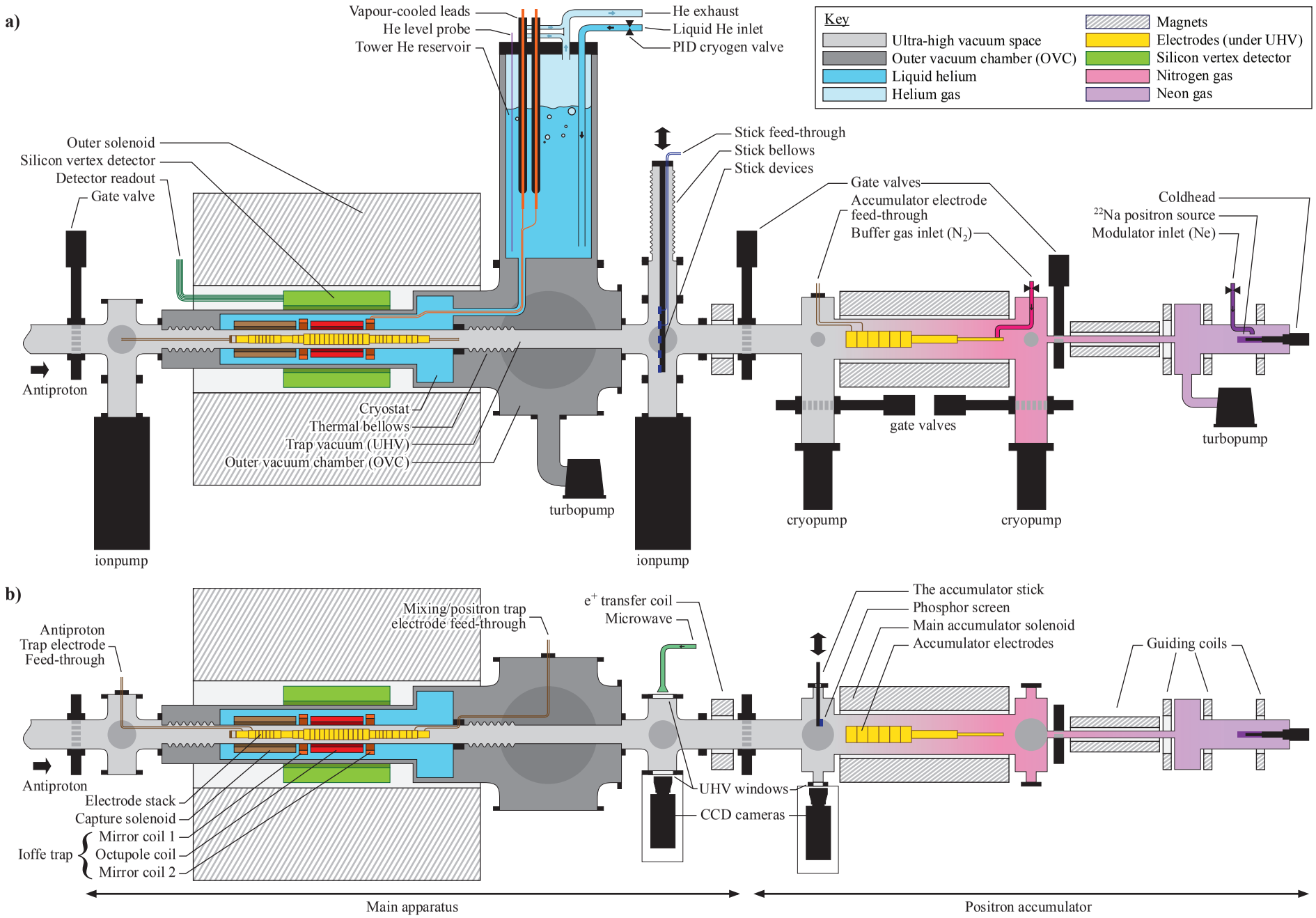

Figure 2.1: A schematic diagram of the a) side and b) top cross–section of the main ALPHA apparatus and the positron accumulator. The components are not drawn to scale, some physical support structure are omitted, and only one pair of wires is show for components sharing a similar electric feed–through pattern.

require an external power supply. A small heater is attached to a short segment of the windings. The persistent mode of the magnet can be switched off by activating the heater, which terminates the superconductivity of the segment. The internal current would instead go through a path parallel to the heated, non-superconducting segment, on which a power supply can control the current through the windings. A Lakeshore power supply unit is used to control the persistence operation of the solenoid, and provide the current to (extract the current from) the solenoid when the magnetic field of the solenoid is being increased (decreased). Note that the power supply needs to be set to the same current level as the magnet before the persistence switch can be turned off to prevent an inductive spike across the power supply.

2.2 The silicon detector

The outermost silicon vertex detector is made of 60 rectangular silicon wafers, arranged into three cylindrically concentric layers and two axial sets (see Fig. 2.2). Each pixel on these wafers registers the energy deposited by energetic particles passing through them, returning the position of such passage when read by an external readout system. When a particle pierces all three layers, the three coordinates registered in close temporal proximity are identified and used to reconstruct the direction and curvature (due to the solenoidal magnetic field) of the trajectory.

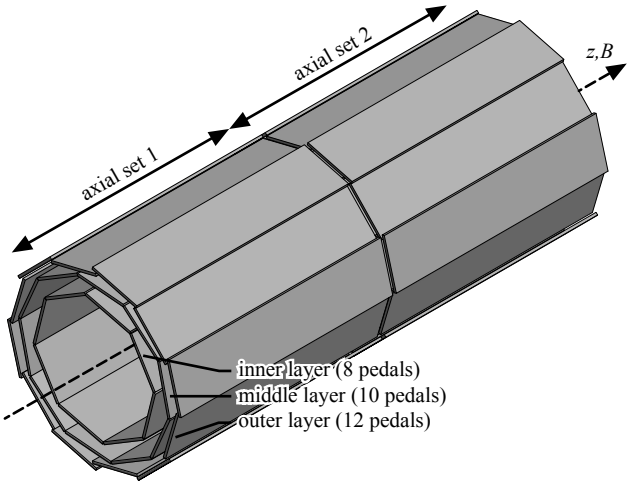

Figure 2.2: The silicon vertex detector, with the wafer pedals arranged into three radial layers and two axial sets.

To detect a trapped antihydrogen atom, the magnetic minimum trap is turned off and the anti-atom is allowed to annihilate. By measuring and extrapolating the trajectories of pi–mesons emanating from the annihilation between antiproton and proton, the exact position and time of the annihilation can be identified. The annihilation of the positron is

not detected, and at least two pi–meson tracks must be reconstructed to form an intersecting vertex. This annihilation vertex is then recorded as an antihydrogen atom, subject to various rejection criteria ("cuts") to eliminate mis-identifications [37]:

- The tracks being extrapolated must not be too close to co-linear, which makes them very likely to be the result of cosmic muons passing through the experiment. The silicon vertex detector does not distinguish muons from pi–mesons.
- The position of the vertex must be on the inner electrode wall (relaxed to take account of reconstruction inaccuracies) upon which antiprotons annihilate. Vertices which are not on the wall are likely to be from cosmic muons scattered by the structural material of the apparatus between their two passes through the detector.
- The axial position of the vertex must be within the two mirror coils of the magnetic minimum trap, between which the anti-atoms are originally trapped. While it is possible that antihydrogen atoms have drifted outside the mirror coils before annihilating, this should not impact the detection rate greatly since the silicon pedals do not extend too far beyond the coils. The reduced solid angle coverage near the ends means antihydrogen detection in that region is inefficient anyway. Vertices from far beyond the mirror coils are more likely to be cosmic muons than antihydrogen atoms.
- The timing of the annihilation must be within ~ 30 ms after the magnetic minimum trap shutdown, which is long enough for the current in the magnetic minimum trap coils to fully decay (see the next section). This small temporal window of detection minimises the chance of cosmic muon mis-identification since cosmic ray arrives at a constant rate, while antihydrogen escape falls off rapidly as the trap current is drained.

2.3 The cryostat

The middle cryostat is a "wet" volume of liquid helium, insulated on the outside face by the outer vacuum chamber (OVC) and a heat shield such that the silicon detector remains at room temperature. Immersed in the cryogenic volume are four magnet coils of various geometries [38], fabricated at Brookhaven National Laboratory. The coils are wound with an Niobium–Titanium wire (see Fig. 2.3), which become superconducting at temperatures below 10 K. A solenoidal coil on the left compliments the magnetic field of the outer solenoid, and increases the field strength inside its bore by 2 T when energised. This solenoid is used during antiproton capture and cooling (see Sec. 2.9). A race–track octupole coil surrounds the centre of the trap, and two mirror coils are positioned on the octupole's two axial limits. Together, the three latter coils form the magnetic minimum trap which, when energised, produces a region of minimum magnetic field strength at the centre. The octupole, with a maximum field strength of ~ 4 T on its inner surface, creates the radial field gradient, while the mirror coils, with a maximum axial field strength of ~ 1.2 T, create the axial gradient.

Neutral species can be trapped around the minimum by their magnetic moment, due to the magnetic force $\boldsymbol{F} = -\nabla(\boldsymbol{\mu} \cdot \boldsymbol{B})$ felt by a dipole [39]. The ALPHA magnetic minimum trap can confine ground state antihydrogen atoms with a maximum of ~ 0.5 K kinetic energy.

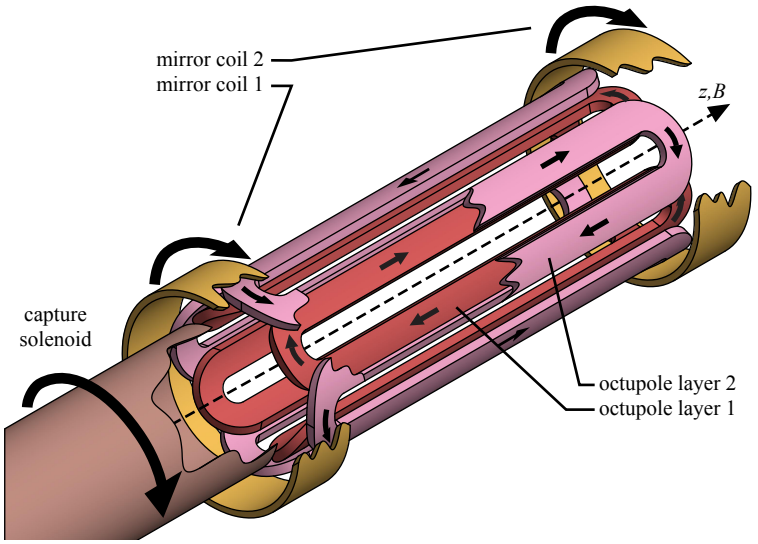

Figure 2.3: A schematic view of the cryostat magnet coils, showing the current directions in the four magnet coils. The octupole is made of layers of oppositely wound race–track coils such that the azimuthal currents on the two ends cancel out to the first order, and that the octupole would have no axial magnetic field contribution. Only two layers are shown in this schematic view, while the physical octupole contains eight counter–winding layers.

The cryostat coils do not have an internal persistent mode — i.e. their current is always driven through an external, non–superconducting circuit. This design is warranted since these coils require frequent ramp up and shut down during antihydrogen trapping operation. The current on these coils can be redirected from their original path through the power supplies to across a resistor array by high power switches made of IGBTs (insulated–gate bipolar transistors) and SCRs (silison–ctonrolled rectifiers), and dissipated as heat. This switching is either deliberately triggered, as in a rapid shutdown of the neutral trap (with a current time constant of ~ 10 ms), or automatically triggered due to the detection of a spike in voltage across the coils. This spike is indicative of some part of the coil having transitioned into a normally conductive state. If unchecked, the resistive heating in this segment would cause a run–away loss of superconductivity throughout the whole coil, and results in massive heating as all the magnetic energy stored in the field is converted into heat. This process, known as quenching, can potentially damage both the superconducting coils (from mechanical and thermal stress) and the cryostat (due to the loss of cryogen). A rapid shutdown of the magnet before the thermal runaway deposits the magnetic energy externally and protects the magnets against quench damage.

The liquid helium in the cryostat boils off due to the thermal load from imperfect insula-

tion, electrical connections from room temperature and infrared radiation through windows. To replenish the boil–off, liquid helium is fed into the cryostat from a vertical helium reservoir (the "tower"), which ensures the cryostat is fully immersed in liquid at all times. The liquid level in the tower is maintained by an inflow of liquid helium from an external 1000 litre storage dewer. The flow is regulated by a PID–controlled valve on the feed line, which responds to the liquid level inside the tower (as measured by a superconducting liquid helium level probe). The 1000 litre dewer is in turn manually refilled periodically with shipments of liquid helium arriving from the CERN liquefaction plant on standard 500 litre dewers. The cold gaseous helium exhaust, which is still at very low temperature, is used to pre–cool the current connections for the superconducting coils before they come into contact with liquid helium, thereby minimising the thermal load on the liquid — a configuration known as vapour–cooled leads. After cooling the connections, the now somewhat–warmer gaseous helium is returned to the CERN liquefaction plant on helium gas line for reuse.

2.4 The UHV chamber

The inner cylindrical UHV chamber is kept at a cryogenic temperature through direct thermal contact with the cryostat (as they are separated only by a non-insulated stainless steel wall). The volume is evacuated by a turbopump backed by a scroll pump during the initial pumping and baking after exposure to atmosphere, and kept at an ultra–high vacuum ($\ll 10^{-13}$ torr) by two ion pumps as well as the cryopumping action of the cold surfaces during steady state operation. The UHV chamber contains a stack of gold–plated copper electrodes which, together with the outer solenoid, form the Penning–Malmberg trap. To allow particle and diagnostic access, the cylindrical UHV volume is open on two ends.

The left UHV opening

The left opening of the electrode stack is capped by a 218 μm–thick aluminium foil, beyond which is a UHV–compatible gate valve and the AD beamline. The thin Al foil serves as a degrader foil which lets through antiprotons from the AD and increases their kinetic energy spread, which is essential in capturing antiprotons in the Penning–Malmberg trap (see Sec. 2.9). It separates the vacuum of the ALPHA apparatus and the AD beamline (which is also at UHV), when the gate valve is opened during normal operations. The foil also acts as a charge collector (a Faraday cup) that captures charged species being ejected from the Penning–Malmberg trap. These charges are then collected by an externally connected capacitor, the voltage across which indicates the absolute quantity of charges deposited on the foil.

The Faraday cup can also be used to analyse the axial temperature of a trapped plasma by slowly lowering the trap wall on one side (as opposed to a quick ejection when a simple charge count is desired). As the wall is lowered, the most energetic particles first escape, followed by slower–moving particles. By correlating the trap depth with the number of

particles escaped, the axial energy distribution of the plasma can be deduced, thus giving its temperature [40]. This simple picture is somewhat complicated by the self–field of the plasma, which decreases when particles escape, preventing a simple calculation of the trap depth purely based on the external voltages applied on the electrodes. However, if the plasma is assumed to be in thermal equilibrium, the temperature of the plasma can be deduced by the first few escaping particles. During the escape of these first particles, the self–field of the plasma remains essentially unchanged. This means that the net trap depth can be expressed as $d_0 - rt - \epsilon$, where the self–potential ϵ remains approximately constant in time, d_0 represents the initial vacuum trap depth, and r its rate of decrease . The rate of particle escape is then given by

$$r(t) = \mathcal{N} \exp\left(-\frac{d_0 - rt - \epsilon}{k_B T}\right)$$
$$\log(r(t)) = \log(\mathcal{N}) + \frac{\epsilon - d_0}{k_B T} + \frac{r}{k_B T} t.$$

Fitting the log of the rate of escape against time to a straight line during the escape of the first few particles, therefore, gives the temperature of the plasma. Note that this approximation relies on the fact that the amount of charge escaped is small compared to the overall charge in the plasma. Experimentally this is hampered by the limited sensitivity of the Faraday cup, due mainly to electrical noise. This forces the fitting to be done to the more lately escaping particles which rise above the noise background. Greater sensitivity of the Faraday cup therefore improves the accuracy of the temperature diagnostic, especially for bunches with a small number of particles.

The right UHV opening

The right end of the UHV chamber opens into a six–way cross, under the same vacuum (see Fig. 2.1). This cross is located outside the bore of the outer solenoid, but the latter's fringe field ensures that charged particles leaving or entering the Penning–Malmberg trap would follow the field lines. Particle sources and diagnostic tools placed in the cross can therefore inject and receive particles when placed on the correct field line which connects with the axis of the trap. A vertical assembly of small–sized devices called *The Stick* is therefore positioned in the cross. The assembly is actuated from the top by a step motor, and it places any one of the following devices along the axial field line:

1. A electron gun which injects electrons into the trap.
2. A micro–channel plate (MCP) which receives particles ejected by the trap. The plate is made of a microscopic honeycomb structure, with the "holes" facing the incoming particles, and the "comb walls" aligned at an oblique angle to the incoming particles. The "entrance" side of the plate is biased at a highly negative voltage compared to the "exit" side (of the order of 1 kV), creating a strong electric field. Particles ejected by the

trap travel along magnetic field lines determined by their in–trap transverse position. Upon ejection, all particles residing on the same line enter one channel on the MCP and strike the comb wall, knocking off electrons. These electrons are accelerated by the strong electric field and hit the comb wall again further downstream, giving rise to more electrons. This avalanche eventually exits on the exit side of the MCP, resulting in a strong electron pulse which number is proportional to the number and energy of the incoming particles received by the channel. These electrons are then converted into visible light upon impacting a phosphor screen at the back of the MCP. With each channel receiving all particles on its connected field line, the resultant image on the phosphor screen is therefore the axially integrated distribution $\sigma(r,\theta) = \int n(r,\theta,z)\,\mathrm{d}z$ of the plasma being ejected. A 45° mirror behind the phosphor screen allows a CCD camera positioned outside the front vacuum window on the cross to capture the image. In addition to the image, by measuring the net current drawn by the MCP (on the lines providing the MCP with high voltages), the amount of charge arriving on the plate can be sensitively determined due to the avalanche amplification in the honeycomb. The MCP therefore doubles as a Faraday cup with enhanced sensitivity, making it preferable to the "traditional" Faraday cup on the left opening for temperature diagnostics.

3. A microwave mirror which reflects microwave injected through the back vacuum window on the cross into the trap, used to stimulate some modes in plasmas and the hyperfine transition of antihydrogen atoms.

4. A microwave horn (antenna) which also injects microwave into the trap. Instead of passing the microwave through a window, a microwave waveguide directs the wave through a UHV feed–through into the horn, giving a better transmission efficiency.

5. A pass–through position, which is simply an empty space that allows the positron accumulator on the right of the cross to inject positrons into the trap.

2.5 The positron accumulator

The positron accumulator, first designed and developed by Surko et al. [30], is positioned to the right of the main apparatus, and contains a ^{22}Na source which decay provides a steady stream of positrons. These positrons are slowed through collision with neon atoms which are solidified on a 5 K cryocooler coldhead around the sodium source. These slowed positrons are then guided into a Penning–like trap with a converging axial magnetic field, generated with coils and solenoids. The trap, functioning as an accumulator for positrons, is made of a conventional water–cooled solenoid and five cylindrical electrodes, one of which is segmented azimuthally. The injection beam path and the trap volume is filled with a gradient of nitrogen buffer gas, with the highest pressure around the former and lowest around the latter. Incoming positrons collide with and excite these gas molecules, losing kinetic energy in the process and become trapped. Subsequent collisions causes the positron to lose energy

progressively, and migrate towards the bottom of the trap well. The pressure gradient is designed such that the injected beam and the more energetic positrons would venture into areas with higher nitrogen gas pressure, thus enhancing their slowing rate, while the slower positrons would remain at a low pressure region, enhancing their lifetime. Upon demand from the main apparatus, the nitrogen gas inflow is shut, the trap is evacuated to UHV by two cryopumps, the gate valve separating the positron accumulator and the main apparatus is opened, and the positrons are ejected. Positrons would stream through the cross into the main Penning–Malmberg trap. Additional pulsed magnet coils are energised during this transfer to ensure the field lines, along which the positrons follow, correctly connect the positron trap and the main trap without excessive divergence (which can result in a loss of positrons).

2.6 The central electrodes

There are four types of electrodes within the stack: high–voltage, thin–walled, segmented and normal (Fig. 2.4). High voltage electrodes have the smallest inner radius of 14.80 mm, and contain extra ceramic electrical insulation which allows them to reach a maximum voltage of ~ 2 kV without sparking. Thin–walled electrodes have the biggest inner radius of 22.28 mm, while normal and segmented electrodes both have an inner radius of 16.80 mm. All electrodes are rotationally continuous, except the segmented electrodes. Each of the segmented electrodes is azimuthally divided into sectors, and each sector is electrically isolated from one another. These four types of electrodes are assembled to form three distinct zones in the electrode stack: the antiproton catching trap on the left, the positron trap on the right, and the mixing trap at the centre (Fig. 2.4). The antiproton catching trap is tasked with cooling and tailoring antiprotons, and contains a pair of high voltage electrodes to act as gates for antiproton capture. It also contains one segmented electrode for rotating wall compression (see Sec. 2.9). Similarly, the positron trap receives and manipulates positrons, and it contains one high voltage electrode and one segmented electrode. Antiprotons and positrons are recombined to form antihydrogen in the central mixing trap, which is made exclusively of thin–walled electrode to maximise the neutral trap depth (since the octupole field strength is stronger the closer antihydrogen atoms can approach the windings without annihilating). Thin–walled electrodes are used only in the mixing trap but not the other two regions to leave room outside the latter for electrical connections, mechanical tensioning structures and thermal and magnetic sensors.

2.7 The sequencing system

The voltages on the electrodes dictates the trapped particles' motion inside the Penning–Malmberg trap. In the ALPHA apparatus electrode voltages is programmed through a sequencing system (Fig. 2.5). This system controls the slow ($\gtrapprox 10\mu s$) variation of the

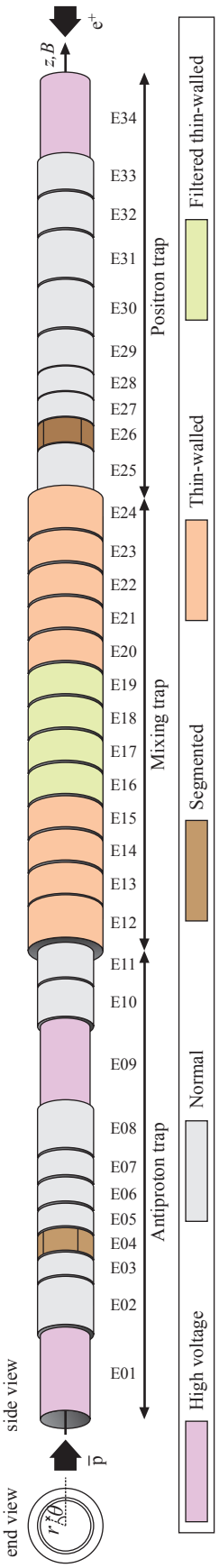

Figure 2.4: The electrode stack of the ALPHA Penning–Malmberg trap.

electrode voltages through five general purpose 16–bit analogue voltage output cards and a bank of amplifiers, while variations faster than this time scale is generated through other specialised components. These components include:

1. Rotating wall generator, which generates six phase–locked waves with sweepable frequency and amplitude. The generator is capable of storing three preset waves, triggered on two digital input lines. A wave output is either stopped according to the preset, or via a trigger on a third digital input line. It also returns its state (running or idle) on an output line. The presets are programmable through a COM port.

2. Pulser, which is an externally triggered fast solid state relay switch, with rise and fall time of ~ 10 ns, but cannot stay closed for longer than $\sim 100\mu s$.

3. Gate pulser, which is similar to a pulser except that it can stay closed indefinitely.

4. High–voltage power supply, which provides a set voltage up to ~ 3 kV upon trigger.

5. Signal generator, which generates simple frequency–sweepable waveforms. The waveforms are programmable through a COM port.

6. Arbitrary waveform generator, which outputs an arbitrary waveform according to a pre–stored, timed array of voltages upon trigger from a digital input line. It also has a digital output line which indicates its state (running or idle). The generator sits in a realtime PXI chassis, and is programmable through an network connection and the National Instruments DAQmx interface.

The sequencing system comprises of two halves, controlling the antiproton and the mixing / positron trap respectively. Each half contains a master sequencer software, which reads a pre–written sequence file and converts it into two arrays: an array of voltage states, and a timed array of digital trigger states. The latter indicates, in each timed state, whether the output trigger lines should be set high or low, and whether a sequence should wait for input trigger lines to become high or low. These two arrays are fed, via network connections, to the analogue and digital drivers on a real–time National Instruments PXI system. The drivers then convert and store these two arrays on the analogue output card and the digital input/output (I/O) card respectively. When the sequence is executed, the digital card manages the timing, advances through its array of timed states according to its internal clock and the state of input lines, and sends triggers to the various devices on its output lines. It also sends a trigger to the analogue output card each time the electrode voltages should be advanced to the next state, at which point the analogue card loads the next row on its voltage array and outputs new voltages on its channels ranging from -10 to 10 V. These voltages are then amplified by 14 times on specialised low–noise differential voltage amplifiers, except for four channels which are amplified by modified amplifiers with a stronger filtering, a slower response time and a gain of 7.2 instead of 14. These voltages from the amplifiers are then

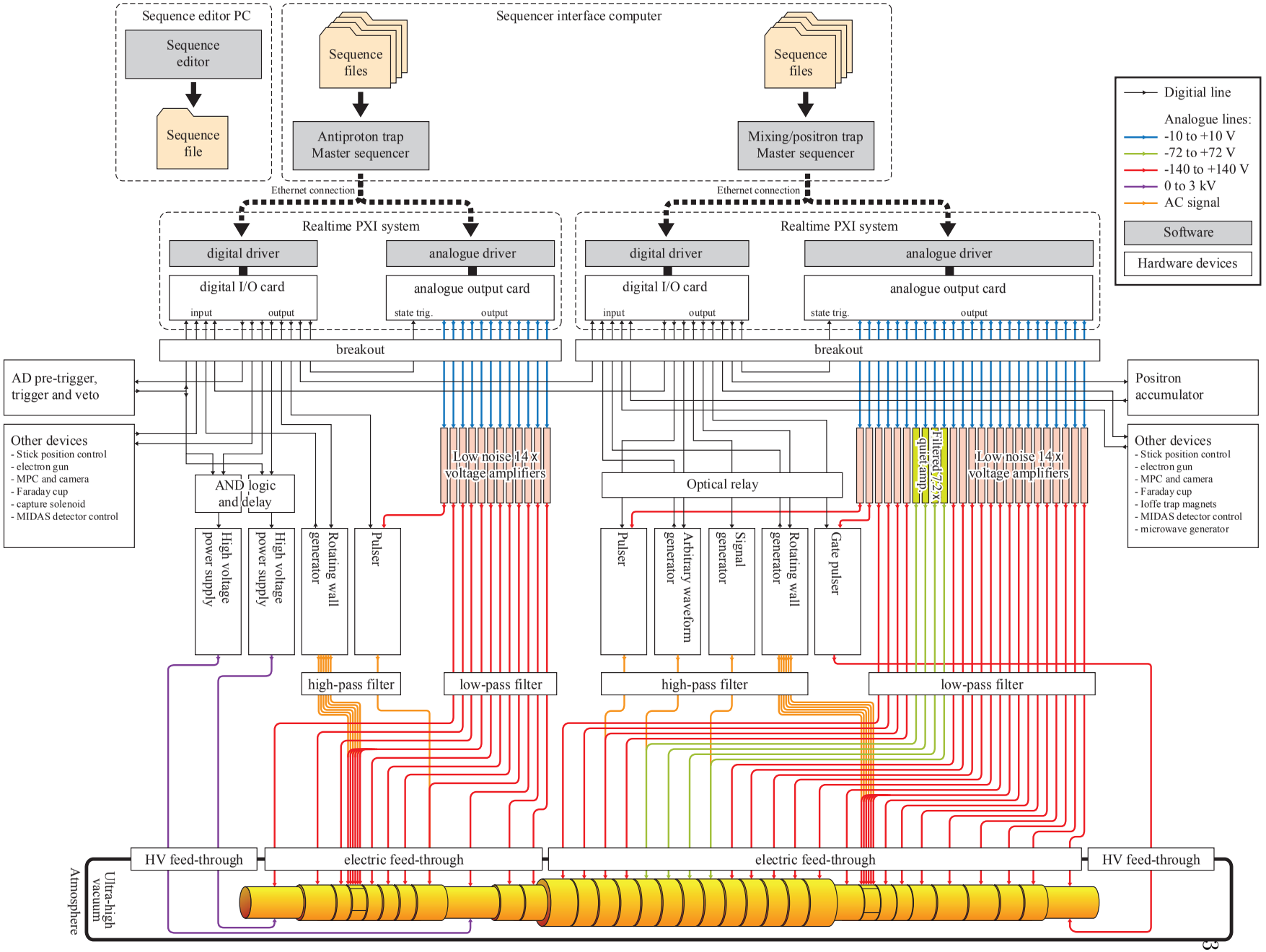

Figure 2.5: A schematic view of the sequencing system which controls the voltage of the main trap electrodes. Multiple digital lines are simplified into one where they start and end on the same devices.

passed through a low–pass ($\leq$ 50 kHz) filter to further isolate electrical noise. Meanwhile the output from the high–frequency specialised devices are passed through a high–pass filter and combined with the low–pass signal. These voltages are then conducted via D–sub style vacuum feed–throughs to the electrodes. The four lower gain channels are connected to E16 through E19, at where antiprotons and positrons are mixed during antihydrogen. The highly filtered amplifiers sacrifices the voltage dynamic range on these electrodes in exchange of minimising electrical noise in the the region, and ensures the two plasmas can be kept as cold as possible to maximise antihydrogen production.

The two halves of the sequencing system can operate independently, or synchronise with each other via two digital trigger channels, going in both directions, if required. The sequencing system also communicate with other particle sources, detectors and diagnostic devices via digital lines:

1. The AD: the AD sends trigger signals on two lines to the antiproton trap sequencer, one $\sim$ 100 s before the delivery of an antiproton beam, and one on beam delivery. The antiproton sequencer can also send a trigger to the AD, called a "veto", to prevent beam delivery by preventing a electrostatic kicker on the AD ring from firing and diverting the ring's beam into the extraction line towards ALPHA. This is useful when the experiment has already captured antiprotons and needs to prevent new energetic particles from entering the trap.

2. The positron accumulator: one trigger line from the accumulator to the mixing / positron sequencer indicates that the accumulator has finished one accumulation cycle and is ready to deliver its positron bunch. A trigger on a second line from the trap would then acknowledge it is ready to receive, and a third line carries a trigger to the trap indicating a positron bunch is being ejected. The mixing / positron sequencer also controls the firing of the magnet coil along the positron transfer beamline on a fourth trigger line.

3. The Stick and its devices: these devices are shared between the antiproton and mixing / positron sequencers, and a handshaking system is necessary to prevent both sequencers from accessing these devices at the same time. When a sequencer needs to take control of the stick, it would set a trigger line to the stick controller to the *high* state, and wait for an acknowledgement from the stick controller on a second trigger line before proceeding. For its part, the stick controller would only send out an acknowledgement when it is no longer occupied. To relinquish the control of the stick, the sequencer would set the first line to *low*, and wait for another acknowledgement. Lines three through eight are all trigger lines from the sequencer to the stick controller, which pass a bit pattern to select the state of the stick. A trigger on a ninth line from the sequencer then executes that selection, with a acknowledgement returned from the stick controller. Only commands from the controlling sequencer are executed and acknowledged. The state of the stick specifies the device to place onto the trap axis,

as well as the status of the aligned device. This includes whether the electron gun is on or off, and the voltage setting for the MCP–phosphor screen (to accommodate the imaging of different particle species, number and ejection energy).

4. MCP camera: it is also shared between the two sequencers, and triggered by one digital line from each sequencer to the camera. No collision avoidance handshaking is implemented for the camera since the shutter time is very short ($\sim$ 10 ms), and collisions are unlikely under normal operation.

5. Degrader Faraday cup: similar to the MCP camera, it is shared between the two sequencers, and triggered by one line from each. A second line is used to send a simultaneous trigger to the Faraday cup controller if a temperature diagnostic is being carried out instead of a simple count of charge.

6. Microwave: the microwave generator is capable of storing multiple preset waveform programmes, programmable through a network interface. One trigger line from the mixing / positron sequencer signals the generator to output the first pre–programmed microwave waveform, and additional triggers on the same line causes the generator to step to subsequent pre–programmed waveforms. Another trigger line controls a microwave switch on the waveguide to the microwave horn on the Stick, which is opened only while microwave is being injected through the horn. It helps to prevent unwanted radio frequency noise from entering the trap and interfering with the stored particles.

7. MIDAS: it is the detector data–logging software for the silicon vertex detector, and numerous other detectors, probes and gauges (e.g. scintillaor counts, NaI gamma ray detector levels, temperatures, pressures, voltages, cryogen levels, valve states, flow rates, etc). To be able to synchronise the timing of sequence execution with the data logged, the sequencers can send triggers to MIDAS to indicate when particular processes are being carried out, e.g. while ejecting particles or energising magnets, such that the detector data can be analysed later on.

8. Capture solenoid: the inner capture solenoid, wrapped around the antiproton trap, is controlled by the antiproton sequencer. The sequencer would set a trigger line to the magnet power controller to *high*, at which point the power supply would ramp up the solenoid current at a preset rate to a preset final value. When the trigger is returned to *low*, the solenoid is ramped back to zero.

9. magnetic minimum trap magnets: the three coils of the magnetic minimum trap are controlled separately through three trigger lines by the mixing / positron sequencer, in a similar fashion to the capture solenoid. Another three trigger lines initiate the rapid shutdown for these magnets.

2.8 Antihydrogen synthesis sequence

The following is a typical sequence involved in producing antihydrogen atoms in the ALPHA apparatus. This process is used, with minor variations, in the results published by ALPHA from 2009 to 2014 [22, 23, 26, 41, 42].

First the stick is moved to the electron gun position, and the capture solenoid is energised. The antiproton sequencer waits for the AD pre-trigger, which signals the arrival of an antiproton beam in 100 seconds. The electron gun is then activated to load electrons into the antiproton region. Typically ~ 20 million electrons are captured. The electron plasma is then split in half axially by lowering the voltage of the central electrode in a three–electrode trap. Half of the electron plasma is ejected, leaving ~ 10 million electrons behind. The electron plasma is further processed. The sequencer then waits for the AD trigger.

Upon receiving the AD trigger, which indicates the arrival of the antiproton beam, the high voltage power supplies for E1 and E9 are triggered through delay units while the antiprotons pass through the degrader foil. E9 is first energised in time to reflect the incoming antiprotons with $KE_{\parallel} \leq 3.4$ keV, and E1 is energised with a small delay to trap the reflected particles. Out of the $\sim 3 \times 10^7$ antiprotons provided by the AD, ~ 50 thousand are trapped by the HV electrodes. The antiprotons trapped by the HV electrodes collide with the pre–loaded electrons and lose energy. This sympathetic cooling process is allowed to take place over 80 seconds. The high voltage on E1 is then turned off, allowing antiprotons which has not lost enough energy to escape to the degrader and annihilate. Annihilations on the detectors are recorded on the scintillating detectors. ~ 20 thousand antiprotons remains at this point, at around 800 to 1000 K. The high voltage on E9 is turned off thereafter, and the sequencer turns on the AD veto to prevent further antiproton beam arrival. The rotating wall is triggered to compress the antiproton–electron mixture. The antiproton sequencer then waits for a synchronisation signal from the positron / mixing sequencer.

At the same time the antiproton region is processing the AD delivery, the stick control is transferred to the positron / mixing sequencer, which instruct the stick to move to the pass–through position to allow positrons passage. The voltage of E13 is raised, and the positron accumulator is triggered to deliver positrons, which are reflected by E13. The voltage of E34 is then raised in time through the gate pulser to trap the reflected positrons. ~ 10 million positrons are typically captured between the two electrodes. The captured positrons are then compressed axially into a smaller well in the positron region. Any contaminating ions introduced by the accumulator during the transfer is expelled, and the positron number is cut in half via a similar procedure as above. Other manipulations are done on the positron plasma, after which typically ~ 5 million positrons remain, at around 100 K. The rotating wall is then triggered to compress the positron plasma, and allowed to run indefinitely. A synchronisation signal is then sent to the antiproton sequencer, and the positron / mixing sequencer itself awaits the next synchronisation signal.

Upon receiving the signal, the antiproton sequencer proceeds with removing the cooling

electrons from the antiproton–electron mixture, by applying a fast electric pulse to a electrode nearby. The fast–moving electrons respond to the pulse and are ejected from the trap, while the heavy antiprotons are too slow to escape the well during the duration of the pulse. After this, the capture solenoid is powered down, and a synchronisation signal is sent to the positron / mixing sequencer. The signal is immediately followed by the ejection of the antiproton bunch towards the mixing region.

The positron / mixing sequencer, upon receiving this signal, stops the indefinite rotating wall on the positron plasma, and receives the antiproton bunch in the mixing region. The positron bunch is also transferred there. The two bunches are further processed, and the antiprotons are cooled via evaporative cooling. This is achieved by lowering one side of the antiproton confinement barrier, allowing the most energetic particles to escape. The remaining particles re-thermalise to a lower temperature. Typically $\sim$ 16 thousand antiprotons remain at this point, at a temperature of $\sim$ 250 K and a mean radius of $\sim$ 0.4 mm. The antiprotons and positrons are then moved immediately next to each other. A nested well is formed to contain the two species, and allow the injection of antiprotons into positrons later (see Fig. 2.6). The magnetic minimum trap is energised, and the positrons are evaporatively cooled. $\sim$ 3 million positron remains at this point, at $\sim$ 40 K and a mean radius of $\sim$ 0.5 mm. A frequency–chirped, sinusoidal perturbation is then applied on E16 through the arbitrary waveform generator to autoresonantly excite the axial oscillation of the antiprotons. The antiprotons gains axial energy and oscillate in increasing amplitude until they cross into the positron plasma. Most of the injected antiprotons recombine with positrons to become antihydrogen atoms. A small fraction of them are slow enough that they are confined by the magnetic minimum trap, while the majority escape and annihilate on the electrode wall. These annihilations are recorded by the scintillation detectors.

After the antihydrogen formation, the electrodes E1–17 are biased to eject negative particles (such as residual antiprotons or re-ionised antihydrogen atoms) to the left. Annihilations are recorded on the scintillating detectors. The electrodes E19–34 are then biased to eject negative particles to the right. Annihilations are also separately recorded. At this point the only remaining particles, beside the trapped antihydrogen atoms, are the positive unused positrons, which are ejected to the right onto the MCP and detected there. Four cycles of strong alternating electric field are then applied along the entire trap to clear any remaining charged particles.

Finally, the magnetic minimum trap is triggered to undergo a rapid shut–down. Pi–mesons are recorded on the silicon vertex detectors to reconstruct the annihilation vertex. Vertices that pass the cuts are identified as antihydrogen atoms. On average one anti-atom is detected per run.

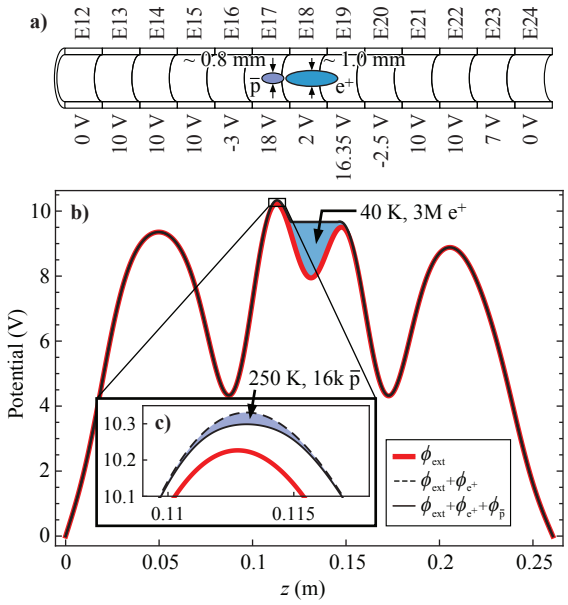

Figure 2.6: A view of the a) antiproton and positron positions in the trap and b) on–axis electrostatic potential during antiproton–positron mixing (step 18–22 in the mixing sequence), showing the contribution of the vacuum potential due solely to the electrodes (ϕ_{ext}), the positron self-potential ($\phi_{\text{e+}}$) and the antiproton self potential ($\phi_{\bar{\text{p}}}$). Insert c) shows the detail of the potential around the antiproton well.

2.9 Special techniques

Antiproton cooling

The difficulties in synthesising and trapping antihydrogen motivates the hardware design and the trapping sequence development of the apparatus. One of the biggest challenges is the difference in energy scale between the antiprotons delivered by the AD, and the shallow depth of the magnetic minimum trap. The AD delivers a beam of mostly monoenergetic antiprotons at an energy of 5.3 MeV, or 6×10^{10} K. These antiprotons have to be aggressively cooled since their velocities dominate the velocities of any antihydrogen atoms produced (positrons being much lighter than antiprotons), and the octupole field, which provides the radial confinement in the 0.5 K–deep magnetic minimum trap, cannot be significantly strengthened due to magnet technology limits and the restricted space. (The mirror coils are not as straightly restrained by space or winding design, making the octupole the limiting factor.) This means the antiprotons have to be cooled by a factor of 10^{11} within the apparatus.

Given the initial and target antiproton energies of 5.3 MeV and 0.5 K, a number of cooling techniques are used in the ALPHA apparatus to bring the antiprotons from one extremum to another while attempting to preserve the maximum number of anti-particles. These includes

the degrader foil, sympathetic cooling, evaporative cooling and collisional equilibration with positrons. Firstly, the degrader foil convert the mono-energetic beam from the AD into one with a wide spread of energy, when the antiprotons collide with the Al nuclei in the foil. Those antiprotons with energy below ~ 3.4 keV are then trapped by a pair of timed high voltage electrodes. This affords a $\sim$ 1600–fold decrease in energy while retaining $\sim 0.17\%$ of the antiprotons.

Subsequently, the antiprotons are cooled further via sympathetic cooling [43], in which the heavy antiprotons collide with a pre-loaded electron plasma and lose energy. The electrons in turn lose energy via cyclotron radiation. The electrons' cyclotron radiation provides a more rapid cooling mechanism compared with the antiprotons' own radiation, since the power of cyclotron emission is proportional to m^{-3}. The electrons therefore radiate energy 6×10^9 times more rapidly than the antiprotons. This process cools the antiprotons by a factor of ~ 50000 while retaining $\sim 40\%$ of them. This sympathetic cooling technique, while efficient, creates other issues. The cooling electrons need to be eventually discarded, which is accomplished via a fast electric pulse applied on a trap electrode. The fast electrons respond to the pulse and are ejected, while the slow antiprotons are too massive to be accelerated sufficiently to escape. This, however, does mean the antiprotons is heated by the pulse. It is therefore important to balance the cooling power of the electrons with the heating resulting from their removal. The pulse also needs be carefully programmed and applied appropriately to minimise heating (while still ensuring a full removal of electrons). If left standing with the electrons for too long, the antiprotons would also radially separate from the electrons in a manner similar to a centrifuge [44]. This can lead to an unstable plasma when the central electrons are removed. The hollow tube of antiprotons retained after the electron removal becomes azimuthally unstable due to the shear instability, leading to expansion and particle loss. We have therefore applied the sympathetic cooling technique in conjunction with the rotating wall (see below) to suppress the expansion and improve the antiproton density.

Further decrease in antiproton energy is accomplished via evaporative cooling [45]. This is achieved by lowering one side of the electrostatic barrier, such that the most energetic antiprotons can escape. These particles carries away a disproportionately large amount of the total energy in the antiproton bunch, and the remaining particles re-thermalise to a lower temperature. The shallowing of the well as a result of this barrier lowering also affords cooling via adiabatic expansion, as the particles on the two axial ends of the plasma climb up a stronger electrostatic well, and are accelerated back down by a weaker one. The temperature is decreased by a factor of ~ 2 while 80% of the antiprotons are retained. The resultant antiproton bunch is at ~ 250 K.

The final cooling happens when the antiprotons are injected into the positron plasma, which is accomplished by an autoresonant excitation of the antiprotons' axial oscillation (see below). As the antiprotons cross into the positron plasma, they equilibrate in temperature with the cold positrons through collisions. This process competes with recombination, in which an antiproton form a bound state (antihydrogen) with the positrons, taking itself

out of the collisional equilibration process. The momentum of the antihydrogen atom is determined by the momentum of the antiproton at the moment of recombination. The majority of antihydrogen atoms formed escape the magnetic minimum trap, and only those at the low–energy end of the distribution (< 0.5 K) result in trapped antihydrogen. On average one anti-atom is trapped, which gives a retention ratio of 0.006%.

Machine fluctuation tolerance

Another challenge to trapping antihydrogen is machine fluctuations. The antiproton bunches delivered by the AD numbers at $\sim 3 \times 10^7$ on average, but can fluctuate by $\pm 30\%$ or above from shot to shot, depending on its maintenance schedule and the behaviour of the upstream accelerators (from which the AD obtains the protons to create antiprotons with). The positron accumulator also has a shot–to–shot fluctuation of $\pm 10\%$ over a base of ~ 10 M positrons, depending on the condition of the Ne modulator and N_2 buffer gas, among many variables. These fluctuations causes the space–charge level of the plasmas in each run to be different from the last, and the processes in the apparatus must be designed to be robust against these fluctuations for the experiment to work. This consideration is especially important during the mixing of antiprotons and positrons: a 10% fluctuation in the positron space–charge corresponds to a ~ 0.2 V change in potential. If the mixing technique fails to take account of the fluctuation, the voltage of the wells can become misaligned, and the charges being ejected from one well into the other can be accelerated by a potential difference of that magnitude. This corresponds to a heating of 2000 K, which would effectively suppress all antihydrogen production and/or trapping.

The mixing of antiprotons and positrons is currently achieved by an autoresonant excitation of the antiprotons' axial oscillation [46]. A sinusoidal voltage perturbation is applied on E16 to create a small oscillating axial force on the antiproton plasma (see Fig. 2.6). The oscillation begins at a frequency above the antiproton axial bounce frequency in their E17 well (the linear resonance), and gradually decreases to a frequency below resonance. The response of the antiproton plasma is minimal at the beginning due to the off-resonance perturbation. As the perturbation frequency passes resonance, the bounce motion becomes phase–locked to the perturbation and grows in magnitude. Due to the non–linear nature of the E17 well, the bigger the antiproton bounce amplitude, the lower its frequency becomes. The phase–locking means the bounce amplitude stays in sync with the perturbation such that their frequencies match. This phase–locking happens regardless of the detailed shape of the well or the exact frequency of linear resonance, as long as the rate of change of the perturbation frequency is below a certain threshold, and the perturbation frequency range envelopes the resonance. Eventually the antiprotons gain enough energy to cross the separatrix and enter the positron plasma. Phase–lock is lost at this point due to the sudden change of bounce frequency, and antiprotons stop gaining energy. This mechanism allows for an accurate excitation of antiprotons up to the potential of the positron plasma, regardless of the shot–to–shot fluctuation of either species. It minimises the energy at which the antiprotons

transverse the positron plasma, and maximises the number of trappable antihydrogen atoms produced.

Radial control

A third challenge to trapping antihydrogen is the radial expansion of the plasmas in the Penning–Malmberg trap. As shown in Sec. 1.2, a single species plasma cannot expand radially in an axi-symmetric manner due to the conservation of the canonical angular momentum. This ensures a good confinement for plasmas. However, experimentally there are mechanisms which promote radial expansion, like scattering with background gas or contaminating ions, radial separation of antiproton–electron mixture, diocotron instability, imperfect azimuthal symmetry of the electrodes, or resistive energy loss. This expansion must be countered to prevent particle loss and to decrease the radius of the antiproton and positron bunches. The latter is important during the last stages of the particle preparation process, when the magnetic minimum trap is energised. The octupole component of the magnetic minimum trap destroys the azimuthal symmetry of the Penning–Malmberg trap, and particles at higher radius sees more of this asymmetry (as the octupole field increases like r^3). These higher–radius particles heat and escape the Penning–Malmberg trap rapidly [47, 48], and it is therefore essential that plasma be compressed radially to minimise this effect. Another motivation for radial compression is the fact that the magnetic minimum trap produces the minimum on the trap axis. Antihydrogen atoms produced off–axis due to radially dispersed antiproton and positron bunches come into being higher in the "dipole potential" $\boldsymbol{\mu} \cdot \boldsymbol{B}$. They therefore see less of a barrier from the magnetic minimum trap, and this reduces the likelihood that the anti-atoms can be trapped. To maximise the trapping rate, antihydrogen atoms should therefore be produced as close to the minimum of the magnetic minimum trap as possible.

The rotating wall technique [49] is used to compress the positron plasm. A rotating, approximately uniform electric field is applied in the x–y plane, through the segmented electrodes E26 (see Fig. 2.7). The rate of rotation of this electric field is chosen to couple with the frequencies of various collective modes of perpendicular motion of the plasma, such that the perpendicular energy of the plasma increases. This energy goes into both the cyclotron motion and the azimuthal momentum of the plasma. The former leads to an increase in perpendicular temperature of the plasma, which is slowly radiated away through cyclotron radiation; the latter results in the radial contraction of the plasma. This is because the azimuthal momentum is caused by the E×B drift of the particles. For this motion to speed up, the $\boldsymbol{E}$ component must increase, which is achieved by a decrease in plasma radius and an increase in the radial self–field.

The rotating wall is also used to compress the mixture of antiprotons and electrons. The frequency of the rotating wall is chosen to compress the electron portion of the mixture, and the antiprotons are observed to follow the electrons and decrease in radius as well [50].

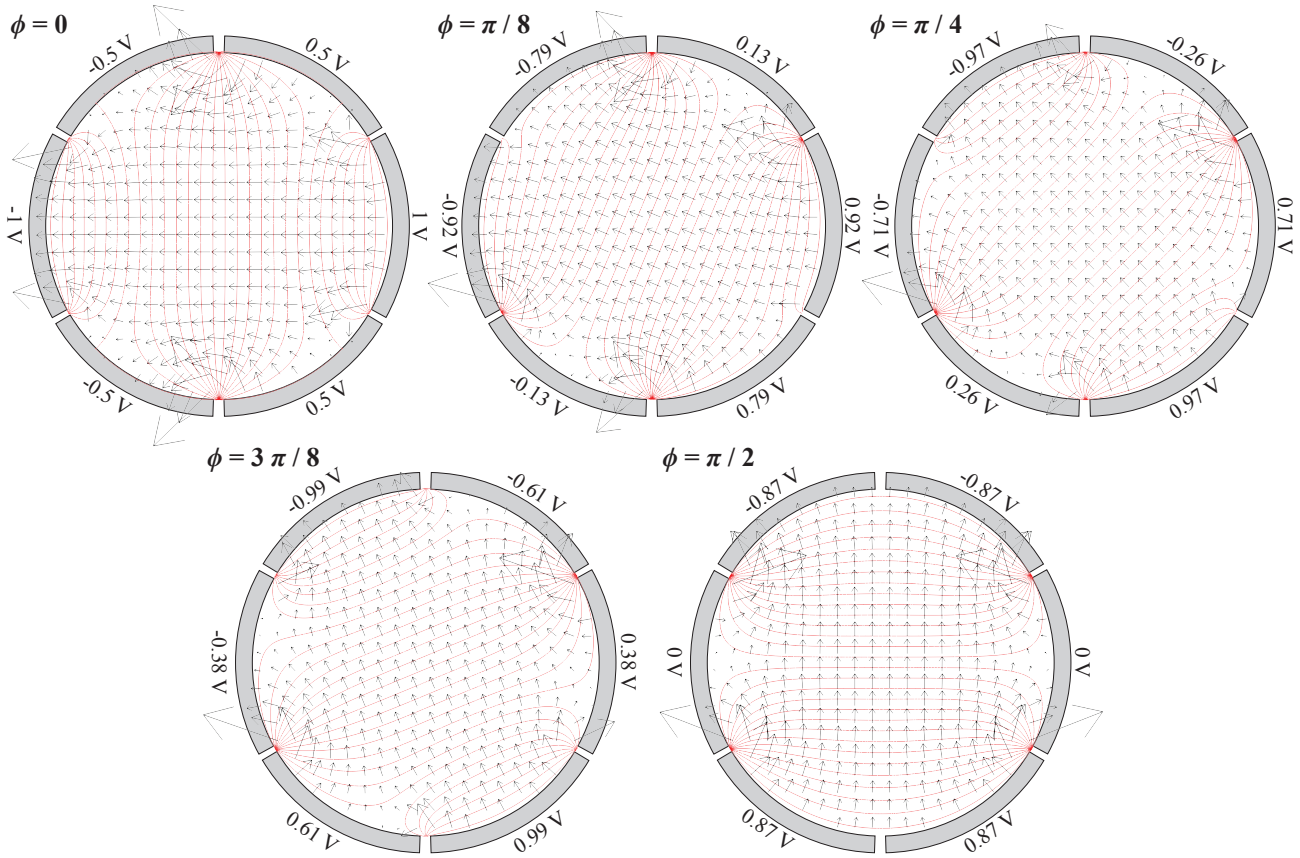

Figure 2.7: A cross-sectional view of the vacuum electric field (black arrows) and potential (red contour lines) of the segmented electrode contributed by the rotating wall generator. The five snapshots are taken at five different times covering a quarter of a period of the rotating wall. The voltages applied on the electrodes are as labelled.

2.10 Recent ALPHA results

Using the basic sequence described above (and its variations), the ALPHA experiment has demonstrated the long–time trapping of antihydrogen atoms and made the first measurements of the anti-atoms' spectrum, established a rough bound on their gravitational mass, and made a precision measurement of their charge neutrality:

Trapped antihydrogen

The ALPHA experiment first demonstrated [41] the trapping of antihydrogen atoms in 2010. In a total of 335 runs, 38 reconstructed annihilation vertices were consistent with that from trapped antihydrogen atoms. As a control, 246 runs were also conducted where extra positron heating was introduced. A strong perturbation was applied on the positrons immediately before they were mixed with antiprotons, which heated the former to $\sim$ 1100 K. This strongly

suppressed recombination and made the resultant antihydrogen atoms unlikely to be trappable. One reconstructed vertex was observed to be consistent with trapped antihydrogen, which is consistent with the expected signal–to–noise ratio of the antihydrogen detection method.

1000 second trapping and temperature measurements

Subsequently, in 2011, the experiment demonstrated [42] the confinement of antihydrogen atoms for a maximum of 2000 s, and more reliably for 1000 s, by holding the anti-atoms in the magnetic minimum trap before its shut down. The 1000 s confinement time was sufficient to ensure that the anti-atoms have enough time to reach their ground states through spontaneous emission (antihydrogen atoms exist in an excited state after recombination). The presence of ground state antihydrogen is essential for future spectroscopic measurements.

In the same paper, the timing of antihydrogen escapes from the magnetic minimum trap during its rapid shutdown was correlated to its velocity distribution, using the fact that anti-atoms with higher kinetic energy escapes the shallowing neutral trap barrier earlier than lower–energy ones. By simulating the timing of annihilation [42, 51] with various initial antihydrogen velocity distributions, it was shown that the trapped anti-atoms' energy distribution is consistent with the tail-end of a Gaussian of a much higher temperature, i.e. $f(E) \propto \sqrt{E}e^{-E/k_BT} \sim \sqrt{E}$ for $T \gg 0.5$ K. This means the antihydrogen atoms were most likely created at a much higher temperature, and that only those with energy of 0.5 K or below were confined by the magnetic minimum trap. This measurement is consistent with the fact that ~ 1 anti-atoms are trapped from an initial population of $\sim 10^4$ created.

Hyperfine resonant interaction

In 2012 the experiment performed [22] the first–ever spectroscopic measurement on antihydrogen atoms, by subjecting them to microwaves that inverted the positron spin state. The anti-atoms were initially created in either the low–magnetic field seeking or high–magnetic field seeking states. The magnetic minimum trap only confined the low–field seeking anti-atoms. The high–field seeking anti-atoms were ejected upon formation. With microwave radiation of a suitable frequency (which is determined by the field experienced by the anti-atom), a low–field seeking anti-atom undergoes a stimulated transition and the positron spin state is inverted. This converts the low–field seeker into a high–field seeker, and causes the anti-atom to be immediately ejected from the magnetic minimum trap and annihilate.

In the experiment, three combinations of magnetic fields and microwave frequencies were attempted. In the first combination, the magnetic minimum trap was energised to its usual setting, and microwave resonant with the positron spin–flipping transition in that field (from low–field seeking to high–field seeking state) was injected after the charged particle clearing. In the 79 runs attempted, 1 antihydrogen atom was detected on trap shut–down (a rate of 0.01 anti-atoms per run), indicating that most antihydrogen atoms had been flipped and

ejected earlier. In the second configuration, the microwave remained at the same frequency, but the magnetic field was slightly increased in strength by 3.5 mT, which meant the microwave was no longer resonant with the (modified) spin–flipping frequency. In the 110 runs attempted, 23 antihydrogen atoms were detected (0.21 anti-atoms per run). In the third configuration, the microwave frequency was raised to match the increase magnetic field. In 24 attempts, 1 anti-atom was detected (0.04 anti-atoms per run). As a control, 52 runs were attempted without any microwave at the usual magnetic minimum trap settings, and 17 anti-atoms were detected (0.33 anti-atoms per run). Also, 48 runs were attempted without microwave at the higher octupole current, yielding 23 trapped anti-atoms (0.48 anti-atoms per run). Comparing these rates, it is obvious the on–resonance microwave has ejected nearly all anti-atoms. The off–resonance runs yields a lower trapping rate compared to the no–microwave control runs, since the off-resonance microwave can still flip the positron spin state of some of the trapped anti-atoms situated in regions of higher magnetic field strength inside the magnetic minimum trap.

Gravity

The trajectory of antihydrogen atoms during the magnetic minimum trap shutdown is influenced by gravitational force. If the weak equivalence principle is valid, antihydrogen atoms fall under gravity in the same manner as hydrogen atoms. For the horizontal trap orientation of ALPHA, this means the depth of the magnetic minimum trap is fractionally reduced on the bottom side compared to the top, and the anti-atoms preferentially escape downwards during the shutdown. This preferential escape is the most prominent for particles that escape late in the shutdown process since these particles have the lowest energy and gravity has the longest time to act on them. However, late–escaping particles are rare due to the distribution $f(E) \sim \sqrt{E}$, which reduces their statistical power. In 2013 ALPHA proposed [23] a statistical method to compare experimental vertices with Monte Carlo simulations of annihilations under various antihydrogen gravity scenarios. It compares the time–binned graphs of the vertices' average vertical position between the experiment and Monte Carlo simulations, and makes the best statistical use of both early– and late–escaping particles to set a bound on the gravitational mass of the antihydrogen atom. At the 95% confidence level, a gravitational mass for antihydrogen above 75 times of m_H, based on statistical effects alone, or 110 times of m_H, including worst–case systematic effects, can be ruled out for gravity, where m_H is the gravitational mass of the hydrogen atom. Similarly, a gravitation mass above 65 times of m_H can be ruled out for anti-gravity, where combined systematic and statistical effects are accounted for.

Charge neutrality

The overall charge neutrality of antihydrogen atoms can be tested by applying an electrostatic field on the anti–atoms and measuring their deflection. In 2014 ALPHA established [26] the fractional charge of the antihydrogen atom as $(-1.3 \pm 1.1 \pm 0.4) \times 10^{-8} e$, where e is

the elementary charge and the two bounds represent statistical and systematic errors. This was achieved by biasing the electrode stack during the magnetic minimum trap shutdown to produce a strong axial electric field. Similar to the effect of gravity, this would reduce the depth of the magnetic minimum trap on one end if the antihydrogen atoms possessed fractional charge, and cause the anti-atoms to preferentially escape in that direction. This would create an offset in the distribution of annihilation vertices along the axis. By simulating this process assuming various fractional charges, and comparing the resultant vertex distributions with the experimental one, the fractional charge of the antihydrogen atom was established.

Chapter 3

The water bag model for equilibrium plasma

The Poisson–Boltzmann equation (Eq. 1.1) describes a plasma in axial thermal equilibrium with its self-electric field and external fields in a Penning–Malmberg trap. It is applicable when the system is perfectly stationary, or when the time scale of the perturbations applied to the plasma is on a much longer time scale than the plasma's axial relaxation time. Equation 1.1 is formulated in such a way that only an axial equilibrium is assumed, and the radial profile (the amount of material in each cylindrical shell in the plasma) can be arbitrarily specified. This is because radial diffusion and advection is strongly suppressed in a Penning–Malmberg trap due to the strong magnetic field, and the time scale for radial–spatial equilibration is much longer than most manipulations in the trap. (Thermal equilibration in the perpendicular directions, on the other hand, involves no bulk spatial movement and occur much more rapidly.) It is therefore appropriate to treat the radial profile as fixed, and only solve for an axial equilibrium. This radial profile can be directly measured in ALPHA using the MCP imaging diagnostic.

A solver for the Poisson–Boltzmann equation requires a numerical solution to both the Poisson equation (the electrostatic field) and the the Boltzmann equation (the spatial distribution). The numerical solution to these two equations are described in the following sections. These solutions are designed for maximum efficiency and minimal hardware requirement, such that they can give near–real time results. This allows the resultant solver to be adopted in experimental operations to help design the electrode voltages necessary to confine various plasma bunches or to let escape particles from a electrostatic well.

3.1 Electrostatic field solver

The electrostatic potential $\phi(\boldsymbol{r})$ in the Penning–Malmberg trap is the solution to the Poisson equation $\nabla^2\phi(\boldsymbol{r}) = -\rho(\boldsymbol{r})/\epsilon_0$, where $\rho(\boldsymbol{r})$ is the space charge density due to the trapped

species. The boundary condition of $\phi(\boldsymbol{r})$ is specified by the voltages applied on the inner electrode surfaces. There are well–established finite element and finite volume methods for solving the Poisson equation numerically in cylindrical coordinates with arbitrary and irregular boundary conditions, which is advantageous since the electrodes have different radii. However these methods require re-computation on every grid point when the space charge distribution or the electrode voltages are altered. In the following we develop an approximate analytic solution to the Poisson equation which can be evaluated more efficiently to give quicker simulations.

For simplicity, the electrode stack is assumed to have a uniform inner radius, ignoring the different radii of the mixing trap and catching trap electrode. The eelctrode stack is also assumed to extend axially to $\pm\infty$. The cylindrical walls before E01 and after E34 are considered to be grounded. These simplifications give good approximation for processes near the trap axis and not in the immediate vicinity of the radial steps at E11–12 and E24–25 or the two ends of the electrode stack. In this case ϕ is the solution to the system

$$\begin{cases} \nabla^2\phi = -\dfrac{\rho}{\epsilon_0} \\ \phi(r = r_w, z) = V_n \quad \text{for } z \in \left[z_{n-1/2}, z_{n+1/2}\right], \end{cases} \tag{3.1}$$

where $z_{n+1/2}$ is the z–position of the gap between the n-th and $(n+1)$-th electrodes, V_n is the voltage applied on the n-th electrode, and r_w is the inner radius of the electrodes.

While straight forward, Eq. 3.1 can be further simplified by decomposing ϕ into two parts:

1. The vacuum potential, ϕ_{vac}, is the potential due to the voltages on the electrodes alone without any space charges. This is the homogeneous solution to Eq. 3.1 given by

$$\begin{cases} \nabla^2\phi_{\text{vac}} = 0 \\ \phi_{\text{vac}}(r = r_w, z) = V_n \quad \text{for } z \in [z_{n-1/2}, z_{n+1/2},]. \end{cases}$$

2. The self-potential, ϕ_{ch}, is the potential due to the space charges alone with a grounded boundary (i.e. all the electrodes are at 0 V). This corresponds to the particular solution to Eq. 3.1 given by

$$\begin{cases} \nabla^2\phi_{\text{ch}} = -\dfrac{\rho}{\epsilon_0} \\ \phi_{\text{ch}}(r = r_w, z) = 0. \end{cases}$$

Adding the two contributions, it is obvious that $\phi_{\text{vac}} + \phi_{\text{ch}}$ is a solution to Eq. 3.1. By the uniqueness theorem, it therefore must be the sole solution.

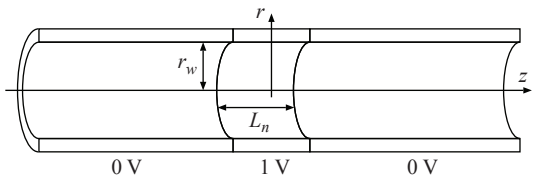

Figure 3.1: Boundary condition used to calculate the potential contribution from the n-th electrode in the Penning–Malmberg trap. The grounded cylinders on the left and right extend to infinity.

Vacuum potential

The vacuum potential can be obtained by solving the Laplace equation with the step-wise boundary condition on the electrodes. However, this requires a complete recalculation of the potential every time the voltages are altered. A more efficient method is to calculate the potential created by each electrode raised to a voltage of 1 V with all other electrodes grounded, i.e.

$$\begin{cases} \nabla^2 \phi_{v,n} = 0 \\ \phi_{v,n}(r = r_w, z) = \begin{cases} 1 \text{ for } z \in [-L_n/2, L_n/2] \\ 0 \text{ otherwise} \end{cases} \end{cases}, \tag{3.2}$$

where $L_n \equiv z_{n+1/2} - z_{n-1/2}$ is the length of the n-th electrode. Figure 3.1 shows the geometry for calculating $\phi_{v,n}$. The potential of a particular set of electrode voltage configuration $\{V_n\}$ is then given by

$$\phi_{\text{vac}}(r, z) = \sum_{n=1}^{34} V_n \, \phi_{v,n} \left(r, z - \frac{z_{n-1/2} + z_{n+1/2}}{2} \right).$$

This allows the individual $\phi_{v,n}$ to be computed only once to generate all possible ϕ_{vac} under any electrode voltage configuration.

We solve Eq. 3.2 using separation of variables to decompose $\phi_{v,n}$ into Fourier–Bessel components [52]. The solution is

$$\phi_{v,n}(r, z) = \frac{2}{\pi} \int_0^\infty \mathrm{d}k \, \frac{I_0(2kr/L_n)}{I_0(2kr_w/L_n)} \text{sinc}(k) \cos\left(\frac{2kz}{L_n} \right), \tag{3.3}$$

where $I_0(x)$ is the zeroth order modified Bessel function of the first kind, and $\text{sinc}(x) = \sin(x)/x$. This integral has no closed–form analytic solution, and must be numerically integrated. We have elected to numerically evaluate the integral using the simple trapezoidal rule, for each coordinate (r, z). This numerical integration converts the integral into a summation of the integrand evaluated at close intervals in k, across the entire domain $k \in [0, \infty)$. The spacing between evaluations δk needs to be sufficiently small such that the function does not change drastically between them. This requires a close examination of the behaviour of

the integrand against k. As k increases from zero, the sinc function oscillates with a period of $\sim 2\pi$, while the cosine oscillates with a period of $\pi L_n/z$. The ratio of the two Bessel functions decays exponentially at large k, forming an envelope around the two oscillations that eventually decay to zero. This asymptotic decay has a scale of $L_n/(r_w - r)/2$, which means for points close to the wall ($r \lessapprox r_w$), the integrand remains large at big k. Conversely, for points close to the axis, the decay is fast, and the integrand approaches zero for relatively small k. Knowing these scales, δk is chosen as

$$\delta k = \min\left(0.01 \times 2\pi, 0.01 \times \frac{\pi L_n}{z}, 0.04 \times \frac{L_n}{2(r_w - r)}\right)$$

to ensure a good resolution on the integrand. The numerical evaluation starts with the evaluation of the integrand at $k = \delta k/2$ as the first term in the summation, and more terms are added at each increment in k of δk. The summation continues until the fractional change to the final answer due to new terms being added become smaller than a desired precision, at which point the summation is truncated. Due to the oscillating nature of the integrand, the fractional change to the final answer is monitored every Δk to ensure the fast oscillation is averaged over between each comparison, where

$$\Delta k = \min\left(\max\left(0.25 \times 2\pi, 0.25 \times \frac{\pi L_n}{z}\right), 10 \times \frac{L_n}{2(r_w - r)}\right).$$

This prevents a premature truncation of the summation.

To evaluate the integrand, $I_0(x)$ is approximated using the algorithm described by Press et. al [53]. The $\mathrm{sinc}(x)$ function is evaluated simply as $\sin(x)/x$, except for $x < 10^{-3}$. In the latter case the sinc function is evaluated using the series expansion $\sum_{i=0} a_i$, where $a_0 = 1$ and $a_i = -a_{i-1}x^2/(4(i+1)(i+1.5))$.

Potential due to space charge

Using the same logic behind the vacuum potential computation, the space charge potential is also decomposed into contributions which are linearly superposed to yield ϕ_{ch}. The charge distribution, assumed to be azimuthally symmetric, is specified on a (r, z) grid of regular spacing Δr and Δz in each direction. Each cell on the grid physically represents a charged ring with thickness Δz and width Δr, located inside a grounded cylinder (see Fig. 3.2). The potential contribution created by this cell is constructed in steps. First the potential of an infinitesimal ring is calculated. It is then integrated radially to yield the potential of a circular annulus. The result is subsequently integrated axially to finally yield the potential of the cell. Explicitly, using eigenfunction expansion [52], the potential of an infinitesimally thin ring carrying a unit charge with radius r'' at $z = 0$ inside a grounded cylinder of radius r_w and infinite length is given by

$$\phi_{\mathrm{ring}}(r''; r, z) = \frac{1}{4\pi\epsilon_0}\frac{4}{r_w}\sum_{n=1}^{\infty}\frac{\exp(-\chi_{0n}z/r_w)}{\chi_{0n}}\frac{J_0(\chi_{0n}r/r_w)J_0(\chi_{0n}r''/r_w)}{J_1^2(\chi_{0n})}, \tag{3.4}$$

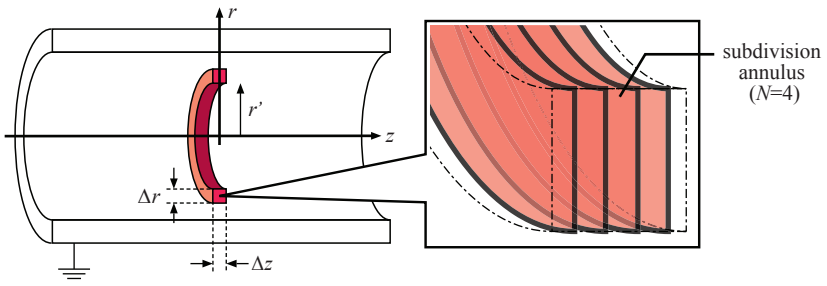

Figure 3.2: The boundary condition and charge distribution used in calculating the potential contribution from one pixel of charge density on the (r, z) grid. The grounded cylindrical boundary extends to infinity at both ends.

where $J_i(x)$ is the i-th order Bessel function of the first kind, and χ_{mn} is the n–th zero of the m–th order Bessel function of the first kind. Integrating this expression from $r'' = r'$ to $r' + \Delta r$, one obtains the potential due to a flat annulus with finite radial width. Expanding that integral to first order in Δr, and using the identity $J_0'(u) = -J_1(u)$, we arrive at the potential of an annulus with unit charge and radius from r' to $r' + \Delta r$ at $z = 0$:

$$\phi_{\text{annu}}(r'; r, z) \approx \frac{1}{4\pi\epsilon_0}\frac{4}{r_w}\sum_{n=1}^{\infty}\frac{\exp(-\chi_{0n}z/r_w)}{\chi_{0n}}\frac{J_0(\chi_{0n}r/r_w)}{J_1^2(\chi_{0n})}\times \left(J_0(\chi_{0n}r'/r_w) - \frac{r'\Delta r\chi_{0n}}{r_w(2r' + \Delta r)}J_1(\chi_{0n}r'/r_w)\right). \tag{3.5}$$

Finally, the potential of one cell is given by a sum of the contributions from the subdivision annuli, as shown in Fig. 3.2. Subdivision is used instead of analytic integration of Eq.(3.5) because of the better numerical convergence of the former. The potential due a cell with unit charge, radial extent from r' to $r' + \Delta r$, and axial extent from $-\Delta z/2$ to $\Delta z/2$ is given by

$$\phi_{\text{pixel}}(r'; r, z) = \frac{1}{N}\sum_{s=0}^{N-1}\phi_{\text{annu}}\left(r'; r, z - \frac{s - (N-1)/2}{N}\Delta z\right), \tag{3.6}$$

where N is the number of subdivisions (which should be even). Using this, the potential of a discrete charge distribution ρ_{ij} can be expressed as

$$\phi_{\text{ch}}(r, z) = \sum_{i,j}\rho_{ij}2\pi r_i\Delta r\Delta z\,\phi_{\text{pixel}}\left(r_i - \frac{\Delta r}{2}; r, z - z_j\right), \tag{3.7}$$

where the indices i and j runs through the grid in r and z dimensions respectively, and (r_i, z_j) is the centre of the (i, j) pixel.

Computationally, Eq. 3.5 is evaluated by first computing the summand at $n = 1$. Terms are added incrementally thereafter, and is truncated when the fractional change to the sum

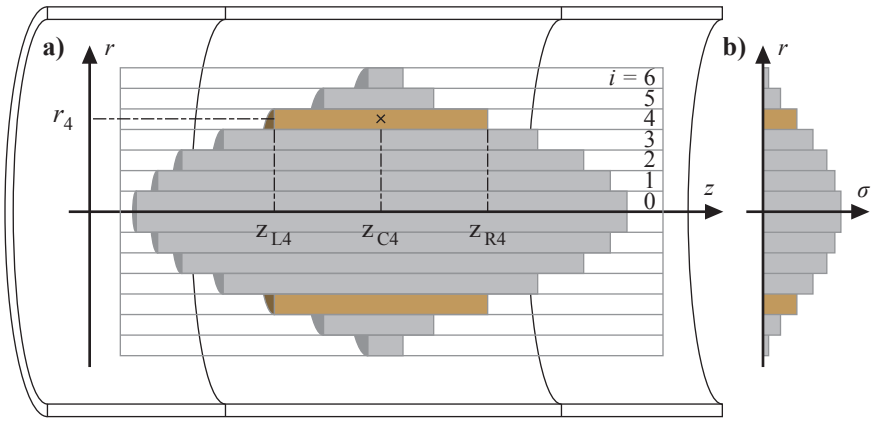

Figure 3.3: a) The geometry of a discretised water bag plasma model, where a plasma is divided into cylindrical shells indexed by i. The leftmost and rightmost boundary of each shell are denoted by z_{Li} and z_{Ri} respectively, and z_{Ci} is the middle point between the two. b) The radial profile of the plasma, given by the axial integral of the number density $n(r, z)$. The radial profile determines the amount of charge in each cylindrical shell, and is experimentally obtained from an MCP image.

due to additional terms decreases below a desired precision. This convergence is usually quite rapid due to the exponential decay factor $\chi_{0n}z/r_w$, which increases with n. When evaluating the terms in the summation, the Bessel functions of the first kind $J_0(x)$ and $J_1(x)$ are evaluated using the algorithm described by Press et. al [53], and χ_{mn} is precomputed and stored as a table.

3.2 Equilibrium distribution solver

Nonlinear solvers for Eq. 1.1 are well–established. However, a typical positron plasma in the ALPHA apparatus has a length of ~ 20 mm, and a Debye length of ~ 0.05 mm. These solvers offers impractical performance in these plasma conditions since they require a grid resolution of a fraction of the plasma Debye length. An alternative to solving the Poisson-Boltzmann equation directly is to exploit the water bag model, which applies to plasmas in the zero temperature limit. In this limit, plasmas have perfect Debye shielding and a zero Debye length, which means the thickness of the transition from the "bulk" of the plasma to the "outside" vacuum is vanishingly small, and the axial electric field E_z within the bulk must be zero. Particles lying on a line of constant radius r_i inside the plasma bulk, therefore, would see local translational symmetry along the line (since the net force is zero everywhere). This local symmetry requires the number density of the plasma to be a constant along this line, and the density must drops abruptly to zero at the leftmost and rightmost boundaries of the plasma, z_{Li} and z_{Ri}. (Here and henceforth, the index i is used to specify the radial grid point; see Fig. 3.3.)

Since the total charge at each radius is measurable experimentally using the MCP imaging

diagnostic, and since the density at each radius must be a boxcar function (i.e. the density is zero except inside the plasma, where the density is a constant), a plasma is fully specified by the boundaries z_{Li} and z_{Ri} for all i. The "correct" solution for the boundary should result in a net potential (the sum of the vacuum potential ϕ_{vac} due to the electrodes, and the self–potential ϕ_{ch}) that is constant between z_{Li} and z_{Ri} for each i. An algorithm which iteratively evolves the boundary to efficiently converge to this solution is developed. The algorithm first makes an initial guess for the plasma boundary (a simple ellipsoid centred around the minima of the vacuum potential, for instance), then solves for the corresponding total potential using the method described in Sec. 3.1. At each r_i, a test particle put at z_{Li} or z_{Ri} would move according to the z component of the total electric field at its position. Given that the plasma is itself a collection of charged particles, the boundary is expected to move in the same manner if a plasma is somehow initiated in this non-equilibrium shape. Based on this idea, the algorithm pushes the boundary at each i according to the axial electric field at the boundary; the proportionality between the field strength and the distance moved is chosen to maximise the "flatness" of the total potential inside the plasma achieved by that step. Practically, the algorithm proceeds thus:

1. For a given boundary $\{z_{Li}, z_{Ri}\}$ for $i \in [0, N_r - 1]$, the corresponding space charge density is given by

$$\rho(r_i, z) = \begin{cases} \dfrac{\sigma(r_i)}{z_{Ri} - z_{Li}} & \text{for } z \in [z_{Li}, z_{Ri}] \\ 0 & \text{otherwise.} \end{cases}$$

   Here $\sigma(r_i)$ is the radial charge density at r_i given by the MCP image. The self–potential $\phi_{\text{ch}}(r, z)$ can then be computed from the space charge, and the vacuum potential $\phi_{\text{vac}}(r, z)$ computed using the trap voltages applied on the electrodes, using the method described in Sec. 3.1. This gives the total potential $\phi = \phi_{\text{vac}} + \phi_{\text{ch}}$.

2. Using the total potential, the average axial electric fields in the left and right halves of the plasma are defined as

$$\bar{E}_{Li} = -\frac{\phi(r_i, z_{Ci}) - \phi(r_i, z_{Li})}{z_{Ci} - z_{Li}},$$
$$\bar{E}_{Ri} = -\frac{\phi(r_i, z_{Ri}) - \phi(r_i, z_{Ci})}{z_{Ri} - z_{Ci}},$$

   where $z_{Ci} \equiv (z_{Li} + z_{Ri})/2$ is the mid–point of the plasma.

3. The test boundary, which are slight modifications of the original boundary based on the average electric field, are then given by

$$z'_{Li} = z_{Li} + \Lambda \bar{E}_{Li},$$
$$z'_{Ri} = z_{Ri} + \Lambda \bar{E}_{Ri},$$
$$z'_{Ci} = (z'_{Li} + z'_{Ri})/2,$$

where Λ is an arbitrary number with the same sign as the plasma's charge.

4. Using the test boundaries, the test space charge density is given by

$$\rho'(r_i, z) = \begin{cases} \dfrac{\sigma(r_i)}{z'_{Ri} - z'_{Li}} & \text{for } z \in [z'_{Li}, z'_{Ri}] \\ 0 & \text{otherwise.} \end{cases}$$

Using this, the test self–potential ϕ'_{ch} and the test total potential $\phi' = \phi_{\text{vac}} + \phi'_{\text{ch}}$ are computed using the method described in Sec. 3.1. (Note that ϕ_{vac} is unchanged.)

5. From this test total potential, the following are defined:

$$\begin{aligned} \bar{E}'_{Li} &= -\frac{\phi'(r_i, z'_{Ci}) - \phi'(r_i, z'_{Li})}{z'_{Ci} - z'_{Li}} \\ \bar{E}'_{Ri} &= -\frac{\phi'(r_i, z'_{Ri}) - \phi'(r_i, z'_{Ci})}{z'_{Ri} - z'_{Ci}} \\ \hat{E}' &= \sum_i \left(\left| \bar{E}'_{Li} \right| + \left| \bar{E}'_{Ri} \right| \right). \end{aligned}$$

The value $\hat{E}'$ is a measure of the "flatness" of the net potential — a figure of merit for the choice of Λ.

6. Return to step 3 with another choice of Λ and obtain another $\hat{E}'$, until the Λ corresponding to the minimum $\hat{E}'$ is found. The search algorithm we use is a simple geometric search, with Λ divided or multiplied by 1.5 each time until a minimum is observed. An interpolation is then used to determine the optimal Λ.

7. Return to step 1, with $z_{Li} := z'_{Li}$ and $z_{Ri} := z'_{Ri}$ for all i, until $\hat{E}'$ becomes smaller than a desired tolerance.

The efficiency of this algorithm is maximised when used in conjunction with the electrostatic solver in Sec. 3.1. Unlike the finite volume or finite element methods for computing the potential from the space charge density, which must solve the entire domain, the solver described in Sec. 3.1 can evaluate the potential at individual points. This means that in steps 2 and 4, the potential at only $3 \times N_r$ points has to be evaluated (at $\phi(r_i, z_{Li})$, $\phi(r_i, z_{Ci})$ and $\phi(r_i, z_{Ri})$ for all i), thus cutting the time required for the algorithm significantly.

Note that the total potential so obtained remains a function of r within the plasma boundary. In using the experimental radial charge density $\sigma(r)$ from the MCP imaging diagnostic as the input of this algorithm, no assumption on whether the plasma has reached radial thermal equilibrium is made. Only an equilibrium in the axial direction is assumed. The actual potential difference across r inside the plasma is determined by the state of the physical plasma. Figure 3.4 shows the convergence of the algorithm when applied on the positron plasma in Fig. 2.6, just before the antiprotons and positrons are mixed during antihydrogen production.

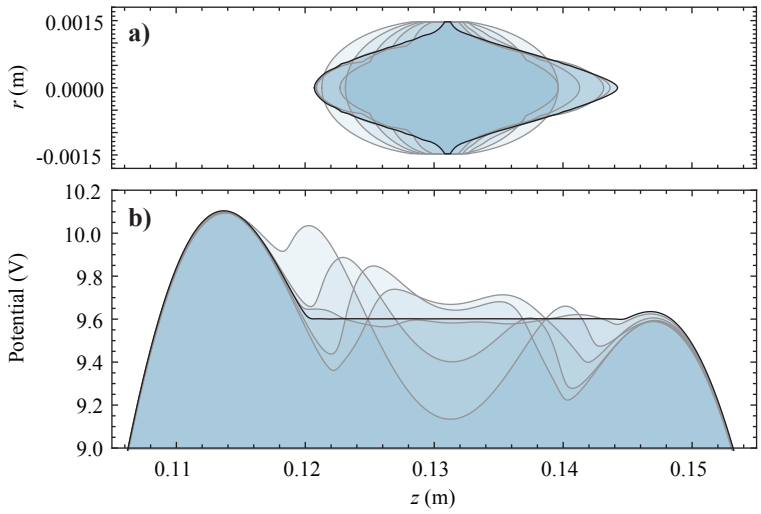

Figure 3.4: The convergence of the water bag solver, computing the positron plasma boundary and potential in the mixing trap just before mixing. a) The algorithm evolving the plasma boundaries z_{Li} and z_{Ri} from the initial ellipse in grey, to the final solution in black. The kinks in the boundary at the top and bottom of the figure are artefacts created by small errors in the outer tail of the radial charge density used as input for this solution. They contain little charge and have little effect. b) The corresponding total potential at $r = 0$ at each step of the convergence process, leading to the expected perfect Debye shielding of a zero-temperature plasma.

Chapter 4

The radially–coupled Vlasov solver for dynamic plasmas

While the Poisson–Boltzmann equation is valid when the external forces applied are sufficiently slow that a plasma responds quasi-statically, many manipulations in the experiment occur on a shorter time scale, e.g. the autoresonant excitation of antiprotons during mixing, or the axial ejection of positrons during a temperature diagnostic. In this case, a model has to follow the time–dependent particle motion in order to reproduce the behaviour of the physical system. As described in Sec. 1.4, the particle motion is separated into the local binary interaction and the bulk collective motion. The modelling of the binary interaction (collision) is deferred to Chs. 5 and 6, while the modelling of the bulk motion is discussed here. This bulk dynamics is dominated by axial motion; radial motion is negligible due to the strong magnetic field, and the azimuthal motion is decoupled. We ignore the radial diffusion in coordinate space due to collisions, and assume the system is azimuthally symmetric. This means diocotron modes are ignored, and non-axi-symmetric electric fields due to misaligned electrodes or the rotating wall are not modelled.

It is worth noting that, while radial motion is negligible, the axial motion does have radial dependence: particles at higher radii see a more shallow (axial) vacuum trap compared the on-axis ones due to the nature of the Laplace equation, and the former also see a weaker outward axial field due to the plasma's space charge compared to the latter. In order to capture this radial dependence of the axial motion, the plasma is modelled as a series of concentric cylindrical shells, with particles in each confined to move in the z direction. The bulk "flow" of material in the cylindrical shell at r and at time t is described by the distribution function $f(r, z, v_z; t)$, where z is the axial position and v_z is the axial velocity. The distribution function is defined such that $\sigma(r) f(r, z, v_z; t) 2\pi r \delta r \, \delta z \, \delta v_z$ gives the number of particles at radial position between r and $r + \delta r$, axial position between z and $z + \delta z$, and axial velocity between v_z and $v_z + \delta v_z$. Here $\sigma(r)$ is the radial profile of the plasma, which is measured experimentally through the MCP imaging diagnostic.

The evolution of the phase–space distributions is described by the Vlasov–Poisson–

Fokker–Planck equation

$$0 = \frac{\partial f}{\partial t} + \overbrace{v_z \frac{\partial f}{\partial z}}^{\text{inertial}} + \overbrace{a(r,z;t)\frac{\partial f}{\partial v_z}}^{\text{acceleration}} - \overbrace{\frac{\partial}{\partial v_z}(\nu(r,z,v_z;t)f)}^{\text{collisional drift}} - \overbrace{\frac{\partial^2}{\partial v_z^2}(D(r,z,v_z;t)f)}^{\text{collisional diffusion}}, \tag{4.1}$$

Equation 4.1 contains no term which leads to an exchange of material between radial shells. Within this equation, the particle motion at one radius has no influence on that at another radius. There is, however, radial coupling through the coefficients a, ν and D, which are computed separately. The four terms in the equation influence the evolution of the distribution in the following manner:

1. The inertial operator shifts the distribution in the z–direction at a rate of v_z. This corresponds to particles with axial velocity v_z free–streaming in the z–direction according to their inherent velocity.

2. The acceleration operator shifts the distribution in the v_z–direction at a rate of $a(r,z;t) \equiv -q/m\ \partial_z\phi(r,z;t)$. This represents the change in axial velocity afforded to the particles at (r,z) by the axial electric field at that position. The net potential $\phi(r,z;t)$ is obtained through the Poisson equation, using the space charge density $\rho(r,z;t) = q\sigma(r)\int f(r,z,v_z;t)\,\mathrm{d}v_z$, and the boundary condition given by the electrode voltages. The acceleration due to electric field is the primary mechanism through which the particles at different radii couple with each other, as the charges at each radius has a global influence on the electric field at all points.

3. The collisional drift term comes from the Fokker–Planck formulation of collisional effects (see Chs. 5 and 6). The drift term shifts the distribution at each (r,z,v_z) in the v_z–direction, and reflects the momentum transferred to the particles at (r,z,v_z) due to the collisions with other particles at the same (r,z) but different v_z.

4. Similar to the collisional drift term, the collisional diffusion term reflects the broadening of the distribution at (r,z,v_z) in the v_z–direction due to the collisions with other particles at the same (r,z) but different v_z.

There are existing solvers for Eq. 4.1, e.g. a spectral solver by Barth et al. [54] or a real space solver by Filbet et al. [55]. These implementations, however, have limitations in the way the coefficient a is evaluated (i.e. the method in which the Poisson equation is solved), and they cannot simulate the radial dependence of the plasma behaviour, being limited to the dynamics at $r = 0$. The error of these solvers and their computational requirements are also too high for our intended application. In order to develop a more suitable solver for Eq. 4.1, we opt to apply operator splitting, which separates the four operations into independent ones that act on the distribution consecutively in each time step. The distribution $f(r,z,v_z;t)$ is discretised on a 3–D phase space grid $\{r_i, z_j, v_{z\,k}\}$, and by discrete time steps $\{t_l\}$. Here

$i \in [0, N_r-1]$, $j \in [0, N_z-1]$ and $k \in [0, N_{v_z}-1]$ (see Fig. 4.2). The positions $r_i = (i+0.5)\Delta r$, $z_j = z_{\min} + (j+0.5)\Delta z$ and $v_{z\,k} = v_{z\,\min} + (k+0.5)\Delta v_z$ denote the centre of the grid point (i, j, k). The value of the distribution at time t_l in cell (i, j, k) is labelled by $f^l_{i,j,k}$.

4.1 The flux balanced method

The three advection operators in Eq. 4.1 take the general form of the advection operator in the PDE

$$\frac{\partial \psi(x,t)}{\partial t} + \frac{\partial}{\partial x}(u(x,t)\psi(x,t)) = 0, \quad \psi(x,0) = \psi_0(x), \tag{4.2}$$

where $\psi_0(x)$ is a given initial condition for ψ. This equation can be solved formally using the method of characteristics. Consider a starting point of (x_0, t_0). The 'flow' of the advection carries it to a new position at time t, labelled as $\chi(t; x_0, t_0)$. These χ are called the curves of characteristics, and are the solutions to the differential equation

$$\frac{\partial \chi(t; x_0, t_0)}{\partial t} = u(\chi, t), \quad \chi(t_0; x_0, t_0) = x_0. \tag{4.3}$$

Qualitatively this states that the movement of the point χ follows the velocity field $u(\chi, t)$. Using the resultant characteristics curve, the solution of the advection Eq. 4.2 can be expressed as

$$\psi(x_0, t_0) = \psi(\chi(t; x_0, t_0), t)\, \frac{\partial \chi(t; y, t_0)}{\partial y}\bigg|_{y=x_0}. \tag{4.4}$$

This solution encapsulates the Lagrangian picture of fluid motion: a parcel of fluid initially at point x_0 and time t_0 retains its material content as it is carried by the flow along the curve of characteristics $\chi(t; x_0, t_0)$, for all time t. The density in the parcel would, however, change as the parcel increase or decrease in size. The partial derivative factor in Eq. 4.4 accounts for this, as parcels cannot penetrate each other, and converging characteristics means the parcels are compressed, causing their density to increase.

The flux balanced method [56] makes use of the formal solution Eq. 4.4 to solve the advection Eq. 4.2 on a discrete grid, where $\psi^j_i = 1/\Delta x \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x\, \psi(x, t_j)$. The spatial index i increases in the x direction, such that $x_i = x_{\min} + (i+0.5)\Delta x$. Integer i denotes the centre of a cell, while $i-1/2$ and $i+1/2$ are the left and right boundaries of that cell. The temporal index j increases at every time step. Integrating Eq. 4.2 from $t = t_j$ to t_{j+1} and from $x = x_{i-1/2}$ to $x_{i+1/2}$ gives

$$\begin{aligned} 0 =& \Delta x \left(\psi^{j+1}_i - \psi^j_i\right) + \int_{t_j}^{t_{j+1}} \mathrm{d}t'\, u(x_{i+1/2}, t')\psi(x_{i+1/2}, t') \\ & - \int_{t_j}^{t_{j+1}} \mathrm{d}t'\, u(x_{i-1/2}, t')\psi(x_{i-1/2}, t'). \end{aligned} \tag{4.5}$$

Substituting $x_0 \to x_{i+1/2}$, $t_0 \to t'$ and $t \to t_j$ in Eqs. 4.3 and 4.4, using them to replace $dt_m\, u(x_{i+1/2}, t')$ and $\psi(x_{i+1/2}, t')$ respectively in the first integral of Eq. 4.5, and repeating a similar procedure for the second integral, one arrives at the flux balance equation

$$\Delta x\, \psi_i^{j+1} = \Delta x\, \psi_i^j - Q_{i+1/2}^j + Q_{i-1/2}^j, \tag{4.6}$$

where the flux across the cell boundary $x_{i+1/2}$ is given by

$$Q_{i+1/2}^j = \int_{\chi(t_j; x_{i+1/2}, t_{j+1})}^{x_{i+1/2}} \mathrm{d}x\, \psi(x, t_j). \tag{4.7}$$

This puts into mathematical form the idea that fluid exiting a cell must enter the next one, and the quantity exchanged is given by 'retracing' the characteristic curve from $(x_{i+1/2}, t_{j+1})$ to (χ, t_j). Since fluid elements cannot overtake each other, everything between χ and $x_{i+1/2}$ must have exited cell x_i and entered x_{i+1}. Also note that no approximation has been made up to this point. Eq. 4.6 is the exact solution of the time-stepping. However, it is incomplete as a numerical method, since in Eq. 4.7 $\psi(x, t_j)$ is not known for all x, but only as the cell averages ψ_i^j. Moreover $\chi(t_j; x_{i+1/2}, t_{j+1})$ needs to be solved using Eq. 4.3.

4.2 Reconstruction methods

As its name implies, a reconstruction method interpolates a function discretised on a grid and recreates a continuous function. This is used to reconstruct $\psi(x, t_j)$ from the discretised values of ψ_i^j to allow the evaluation of the flux integral Eq. 4.7. A reconstruction scheme is centrally important to the behaviour of a numerical advection operator, and there exist numerous schemes with distinct advantages and disadvantages. In the following we summerise some of the more well–known reconstruction schemes and compare their behaviour when used in an advection operator. We suppress the time index j in this section for simplicity; all quantities refer to the time step t_j.

Continuous linear reconstruction

The simplest first order interpolation scheme is

$$\psi(x) = \psi_i \frac{x_{i+1} - x}{\Delta x} + \psi_{i+1} \frac{x - x_i}{\Delta x} \qquad \text{for } x \in [x_i, x_{i+1}), \tag{4.8}$$

which is an interpolation using the cell centres as pivots. Figure 4.1 b shows the reconstruction of a distribution using this scheme. The reconstruction within cell i requires the data at $i-1$, i and $i+1$, called the "stencil" of this interpolation scheme. In Fig. 4.1 a a 1–D distribution is advected in a uniform velocity field using this reconstruction scheme, and several moving window snapshots of the distribution is shown. The window follows the distribution at the field's velocity, and a perfect reconstruction scheme is expected to

yield perfectly identical waveforms on the plot at all times. As evident from Fig. 4.1 a, however, this reconstruction method results in significant numerical defects during the advection process. In particular, these common problems are observed:

- Positivity: Areas in the distribution become negative as the pulse is advected, which is unphysical for a distribution function.
- Oscillation: The "downstream" tail of the distribution develops an oscillation which is not present in the initial distribution. This is due to the reconstruction scheme consistently underestimating (overestimating) the flux leaving convex (concave) regions of the distribution.
- Phase error: The peak of the distribution shifts to the left in Fig. 4.1 a, which indicates that the speed at which the distribution is shifted to the left is less than the actual velocity of the advection field.
- Numerical diffusion: The pulse is seen to grow in width and decrease in height upon advection in Fig. 4.1 a. Narrow details of the initial pulse is lost upon advection. These behaviours are indicative of the numerical diffusion introduced to the advection operator by this reconstruction scheme.

Many of the subsequent reconstruction methods are designed to counter these defects using various numerical techniques.

Piecewise linear reconstruction

Used by Fijalkow [56] in the discretisation of the Vlasov equation, this simple linear reconstruction scheme uses a piecewise linear discontinuous interpolant:

$$\psi(x) = \psi_i + (\psi_{i+1} - \psi_{i-1})\frac{x - x_i}{2\Delta x} \qquad x \in [x_{i-1/2}, x_{i+1/2}). \tag{4.9}$$

This scheme also has a stencil of $\{i-1, i, i+1\}$. Figures 4.1 c demonstrates the advection effect of this scheme and Fig. 4.1 d shows a reconstructed distribution. This scheme improves on the continuous linear reconstruction in most aspects, but the positivity of the distribution is not preserved. The numerical diffusion quickly erases the minor peak in the distribution in Fig. 4.1 c.

Uniformly non-oscillatory reconstruction (linear)

First described by Harten and Osher [57], this scheme uses a piecewise linear interpolant, but with its slope obtained through a more sophisticated process than the one used in the

piecewise linear reconstruction. First introduce the "modified minimum" function

$$\text{minmod}(a,b) \equiv \begin{cases} \text{sign}(a)\min(|a|,|b|) & \text{if sign}(a) = \text{sign}(b) \\ 0 & \text{otherwise.} \end{cases}$$

Next define the convexity at cell centre i and cell boundary $i+1/2$, respectively, as

$$D_i \equiv \psi_{i+1} + \psi_{i-1} - 2\psi_i$$
$$D_{i+1/2} \equiv \text{minmod}(D_i, D_{i+1}).$$

The slope of the interpolant in cell i is then

$$S_i \equiv \text{minmod}\left(\frac{\psi_{i+1}-\psi_i}{\Delta x} - \frac{D_{i+1/2}}{2\Delta x^2},\ \frac{\psi_i - \psi_{i-1}}{\Delta x} + \frac{D_{i-1/2}}{2\Delta x^2}\right),$$

which in effect equates the slope in cell i with either that at $i-1/2$ or $i+1/2$, whichever is smaller. In case the slope on the two sides differ in sign (e.g. at a local extremum), the interpolant is set to have a zero slope. The reconstructed function is then

$$\psi(x) = \psi_i + S_i\frac{x-x_i}{\Delta x} \qquad x \in [x_{i-1/2}, x_{i+1/2}). \tag{4.10}$$

This scheme has a stencil of $\{i-2,\ i-1,\ i,\ i+1,\ i+2\}$, and its use is demonstrated in Figs. 4.1 e and 4.1 f. As its name suggests, the uniformly non-oscillatory scheme prevents spurious oscillations, thus preserving the positivity of the distribution. There is still some level of phase error and numerical diffusion, however, as the small feature on the left is completely smoothed out by $t = 9000$ in Fig. 4.1 e.

Piecewise parabolic reconstruction

The Piecewise Parabolic Method (PPM) was developed by Colella and Woodward [58] for advection operators in hydrodynamic and Vlasov simulations. First define the slope S_i in cell i as a modification over the "usual" centre-difference slope $(\psi_{i+1} - \psi_{i-1})/2$:

$$\begin{aligned} S_i \equiv & \text{sign}\left(\psi_{i+1}-\psi_{i-1}\right) \times \\ & \min\left(\left|\frac{\psi_{i+1}-\psi_{i-1}}{2}\right|,\ 2\left|\psi_{i+1}-\psi_i\right|,\ 2\left|\psi_i - \psi_{i-1}\right|\right) \quad \text{if } \left(\psi_{i+1}-\psi_i\right)\left(\psi_i-\psi_{i-1}\right) > 0, \\ \equiv & 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned}$$

This modification enhances the representation of a steep slope; in the case cell i is a local extremum, the slope S_i is set to zero. Using this slope, the half–point is defined as

$$\psi_{i+1/2} \equiv \frac{\psi_i + \psi_{i+1}}{2} - \frac{S_{i+1} - S_i}{6}.$$

Now that we have the centre, the left–boundary and the right–boundary values of cell i (namely ψ_i, $\psi_{i-1/2}$ and $\psi_{i+1/2}$), there is sufficient information to fit a piecewise parabola in cell i. This is indeed the procedure in most cases; however, in some particular cases this would cause the interpolant in cell i to create extra local extremums that does not exist in the original discrete function (see Fig. 4.1 f) due to Runge's phenomenon, as higher order interpolations can oscillate between data points. This would lead to instability and oscillation in the advection operator. To prevent this, the interpolant in cell i is fitted to the "corrected" centre, left–boundary and right–boundary values ψ_i, $\psi_{L,i}$ and $\psi_{R,i}$, which are obtained through the following procedure:

1. $\psi_{L,i} := \psi_{i-1/2}$
2. $\psi_{R,i} := \psi_{i+1/2}$
3. if $(\psi_{R,i} - \psi_i)(\psi_i - \psi_{L,i}) \leq 0$:
   $\psi_{L,i} := \psi_i$
   $\psi_{R,i} := \psi_i$
4. if $(\psi_{R,i} - \psi_{L,i})(6\psi_i - 3\psi_{L,i} - 3\psi_{R,i}) > (\psi_{R,i} - \psi_{L,i})^2$:
   $\psi_{L,i} := 3\psi_i - 2\psi_{R,i}$
5. if $(\psi_{R,i} - \psi_{L,i})(6\psi_i - 3\psi_{L,i} - 3\psi_{R,i}) < -(\psi_{R,i} - \psi_{L,i})^2$:
   $\psi_{R,i} := 3\psi_i - 2\psi_{L,i}$.

The corrected boundary values ensure that the parabola inside cell i does not contain any internal extremum. In the case cell i itself is an extremum in the discrete grid, the parabola is set to horizontal. The PPM algorithm then interpolates between the three points with a parabolic function, from which $\psi(x)$ takes its value in cell i. This gives

$$\begin{aligned}\psi(x) =& \psi_{L,i} + (\psi_{R,i} - \psi_{L,i})\,\delta \\ &+ (6\psi_i - 3\psi_{R,i} - 3\psi_{L,i})\,(1-\delta)\delta \qquad x \in [x_{i-1/2}, x_{i+1/2}),\end{aligned} \tag{4.11}$$

where $\delta = (x - x_{i-1/2})/\Delta x$. The piecewise parabolic scheme has a stencil of $\{i-2,\ i-1,\ i,\ i+1,\ i+2\}$, and its effect is shown in Figs. 4.1 g and 4.1 h. This reconstruction scheme preserves positivity of the distribution, does not introduce spurious oscillations, and has small phase error and numerical diffusion compared to the other schemes tested.

Positive flux conserving reconstruction (linear)

Introduced by Filbet et al. [55] for discretising the Vlasov equation, this method uses a piecewise linear interpolant. The slope of the interpolant is obtained using centre–differencing, similar to the piecewise linear reconstruction, but includes an extra slope corrector ϵ_i to

preserve positivity of the distribution. It is given by

$$\epsilon_i = \begin{cases} \min\left(1, 2\dfrac{\psi_i}{\psi_{i+1} - \psi_i}\right) & \text{if } \psi_{i+1} > \psi_i \\ \min\left(1, -2\dfrac{\psi_{\max} - \psi_i}{\psi_{i+1} - \psi_i}\right) & \text{otherwise,} \end{cases}$$

where $\psi_{\max} = \max(\{\psi_i\})$ is the global maximum of the distribution. Using this correction factor the reconstructed function is

$$\psi(x) = \psi_i + \epsilon_i(\psi_{i+1} - \psi_i)\frac{x - x_i}{\Delta x} \qquad x \in [x_{i-1/2}, x_{i+1/2}). \tag{4.12}$$

The linear positive flux conserving scheme has a stencil of $\{i,\, i+1\}$, and its effect is demonstrated in Figs. 4.1 i and 4.1 j. The scheme, as its name implies, preserves positivity of the distribution, but it leads to a significant phase error and numerical diffusion for the smaller peak on the left of the distribution in Fig. 4.1 i.

Positive flux conserving reconstruction (parabolic)

An extension of the previous scheme, this method is also introduced by Filbet et al. [55]. It uses a piecewise parabolic interpolant instead of a linear one, and instead of one correction factor, this scheme employs two:

$$\epsilon_i^+ = \begin{cases} \min\left(1, 2\dfrac{\psi_i}{\psi_{i+1} - \psi_i}\right) & \text{if } \psi_{i+1} > f_i \\ \min\left(1, -2\dfrac{\psi_{\max} - \psi_i}{\psi_{i+1} - \psi_i}\right) & \text{otherwise} \end{cases}$$

$$\epsilon_i^- = \begin{cases} \min\left(1, 2\dfrac{\psi_{\max} - \psi_i}{\psi_i - \psi_{i-1}}\right) & \text{if } f_i > f_{i-1} \\ \min\left(1, -2\dfrac{\psi_i}{\psi_i - \psi_{i-1}}\right) & \text{otherwise,} \end{cases}$$

where $\delta = (x - x_{i-1/2})/\Delta x$. Using these two correction factors, the reconstructed function is given by

$$\begin{aligned} \psi(x) = \psi_i &+ \epsilon_i^+(\psi_{i+1} - \psi_i)\frac{2(x - x_i)(x - x_{i-3/2}) + (x - x_{i-1/2})(x - x_{i+1/2})}{6\Delta x^2} \\ &+ \epsilon_i^-(\psi_i - \psi_{i-1})\frac{2(x - x_i)(x - x_{i+3/2}) + (x - x_{i-1/2})(x - x_{i+1/2})}{6\Delta x^2} \qquad x \in [x_{i-1/2}, x_{i+1/2}). \end{aligned} \tag{4.13}$$

The parabolic positive flux conserving scheme has a stencil of $\{i-1,\, i,\, i+1\}$, and its effect is demonstrated in Figs. 4.1 k and 4.1 l. This scheme displays strong numerical diffusion and is not suitable for our use with the Vlasov equation.

Essentially non-oscillatory reconstruction (arbitrary order)

The essentially non-oscillatory reconstruction introduced by Filbet et al. [55] is slightly different from the previous schemes, in that it does not attempt to reconstruct $\psi(x)$ from the grid points and integrate it according to Eq. 4.7. Instead, this scheme first integrates ψ discretely. The discrete distribution ψ_i is specified at the cell centres x_i, which means the primitive is specified on the cell boundaries, i.e.

$$\Psi_{i+1/2} = \sum_{k=0}^{i} \psi_k \, \Delta x.$$

The discrete primitive is then interpolated to give the reconstructed $\Psi(x) \equiv \int \psi(x) \, \mathrm{d}x$, which in turn gives the flux as

$$Q_{i+1/2} = \Psi(x_{i+1/2}) - \Psi(\chi(t_j; x_{i+1/2}, t_{j+1})).$$

The essentially non-oscillatory scheme reconstructs $\Psi(x)$ from the grid averages $\{\Psi_{i+1/2}\}$ using a piecewise Newton polynomial of arbitrary order, which is fitted over a flexible stencil to achieve the smoothest curve. The lowest order reconstruction scheme ($n = 2$) has an interpolation stencil fixed at $\{i - 1/2,\, i + 1/2\}$, which gives

$$\Psi(x) = [\Psi_{i-1/2}] + [\Psi_{i-1/2}, \Psi_{i+1/2}](x - x_{i-1/2}) \qquad x \in [x_{i-1/2}, x_{i+1/2}), \tag{4.14}$$

where the divided differences, denoted by the square brackets, are defined iteratively by

$$\begin{aligned} &[F_1] = F_1, \\ &[F_1, \cdots, F_n] = \frac{[F_2, \cdots, F_n] - [F_1, \cdots, F_{n-1}]}{x_n - x_1}. \end{aligned}$$

For the next order of interpolation, the stencil of the piecewise interpolant can either be extended to the left or to the right. Labelling the first two points $i - 1/2$ and $i + 1/2$ of the stencil as p_1 and p_2 respective, the choice of the third point is determined by

$$p_3 = \begin{cases} i + 3/2 & \text{if } |[\Psi_{p1}, \Psi_{p2}, \Psi_{i+3/2}]| < |[\Psi_{p1}, \Psi_{p2}, \Psi_{i-3/2}]| \\ i - 3/2 & \text{otherwise.} \end{cases}$$

Using this choice of extension, the reconstructed primitive to the next order ($n = 3$) is expressed as

$$\begin{aligned} \Psi(x) =& [\Psi_{p1}] + [\Psi_{p1}, \Psi_{p2}](x - x_{p1}) \\ &+ [\Psi_{p1}, \Psi_{p2}, \Psi_{p3}](x - x_{p1})(x - x_{p2}) \qquad x \in [x_{i-1/2}, x_{i+1/2}). \end{aligned}$$

This process can be repeated iteratively, choosing to extend the stencil to either the left or right at each increase of the interpolation order. The resultant reconstructed primitive, up to n-th order, is

$$\Psi(x) = \sum_{k=1}^{n} \left([\Psi_{p1}, \cdots, \Psi_{pk}] \prod_{l=1}^{k-1} (x - x_{pl}) \right) \qquad x \in [x_{i-1/2}, x_{i+1/2}). \tag{4.15}$$

The effect of this reconstruction scheme for $n = 2$, 3 and 4 is shown in Figs. 4.1 m to 4.1 r. There is obvious improvement to the fidelity of the advection scheme with increasing interpolation order, and oscillation is suppressed. Numerical diffusion on prominent features is small at $n = 4$; however smaller details are still smoothed out quickly.

Barycentric reconstruction (arbitrary order)

In contrast to the local reconstruction methods introduced above, all of which which have a finite stencil, the Barycentric interpolation method [59] is a global method — all the points in the domain contribute to the reconstruction in each cell. The reconstructed distribution is given by

$$\psi(x) = \left(\prod_{i=0}^{N-1} (x - x_i) \right) \left(\sum_{i=0}^{N-1} \frac{w_i}{x - x_i} \psi_i \right) \tag{4.16}$$

where the weight w_i is given by

$$w_i = \frac{1}{\prod_{j=0,\, j\neq i}^{N-1} (x_i - x_j)}.$$

The result of the interpolation is shown in Fig. 4.1 t. This method suffers from Runge's phenomenon and is unsuitable for use with the Vlasov equation. Computationally this method requires $\mathrm{O}(N^2)$ operations per time step, which is expensive compared to the other schemes.

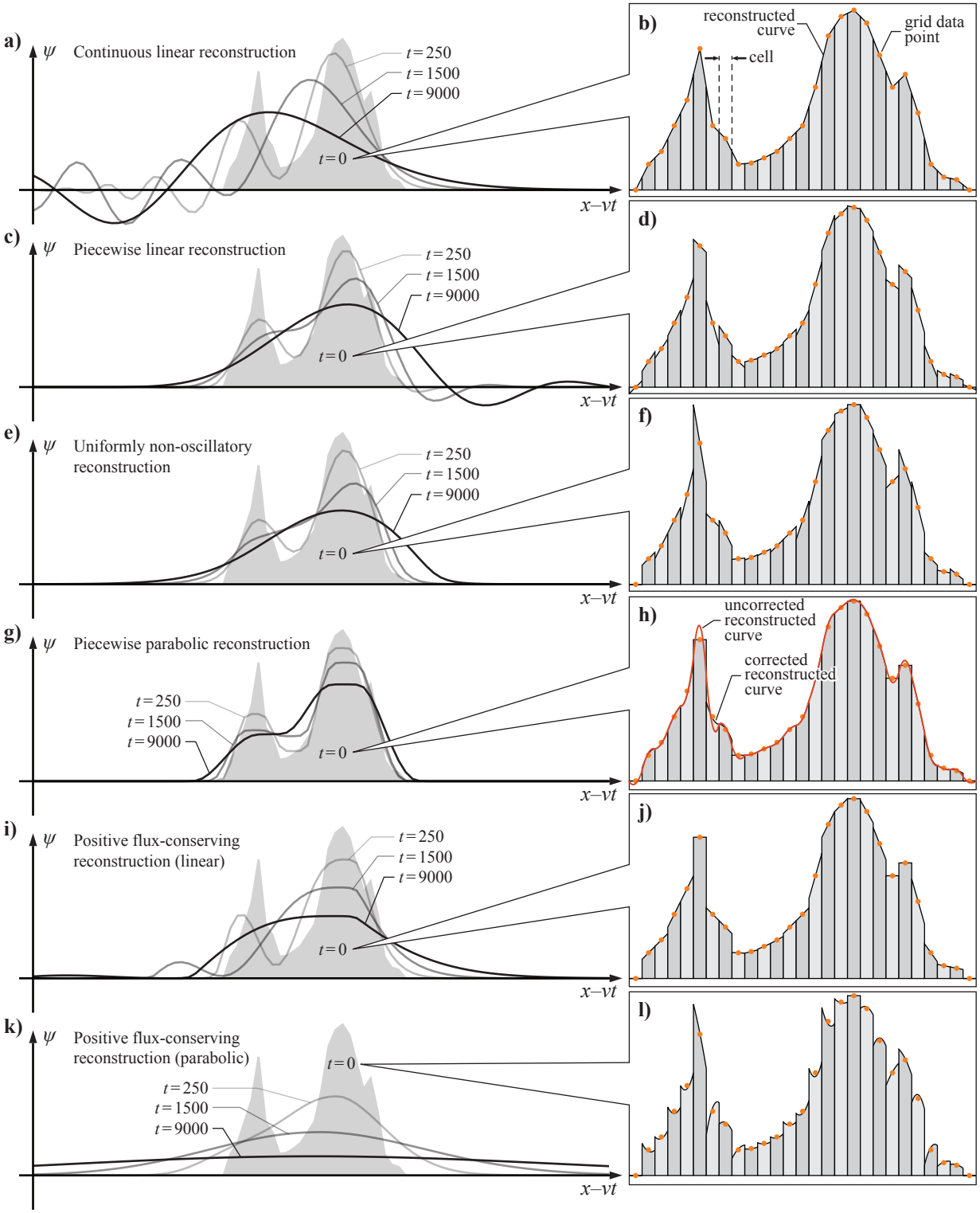

Figure 4.1: (Continued on next page)

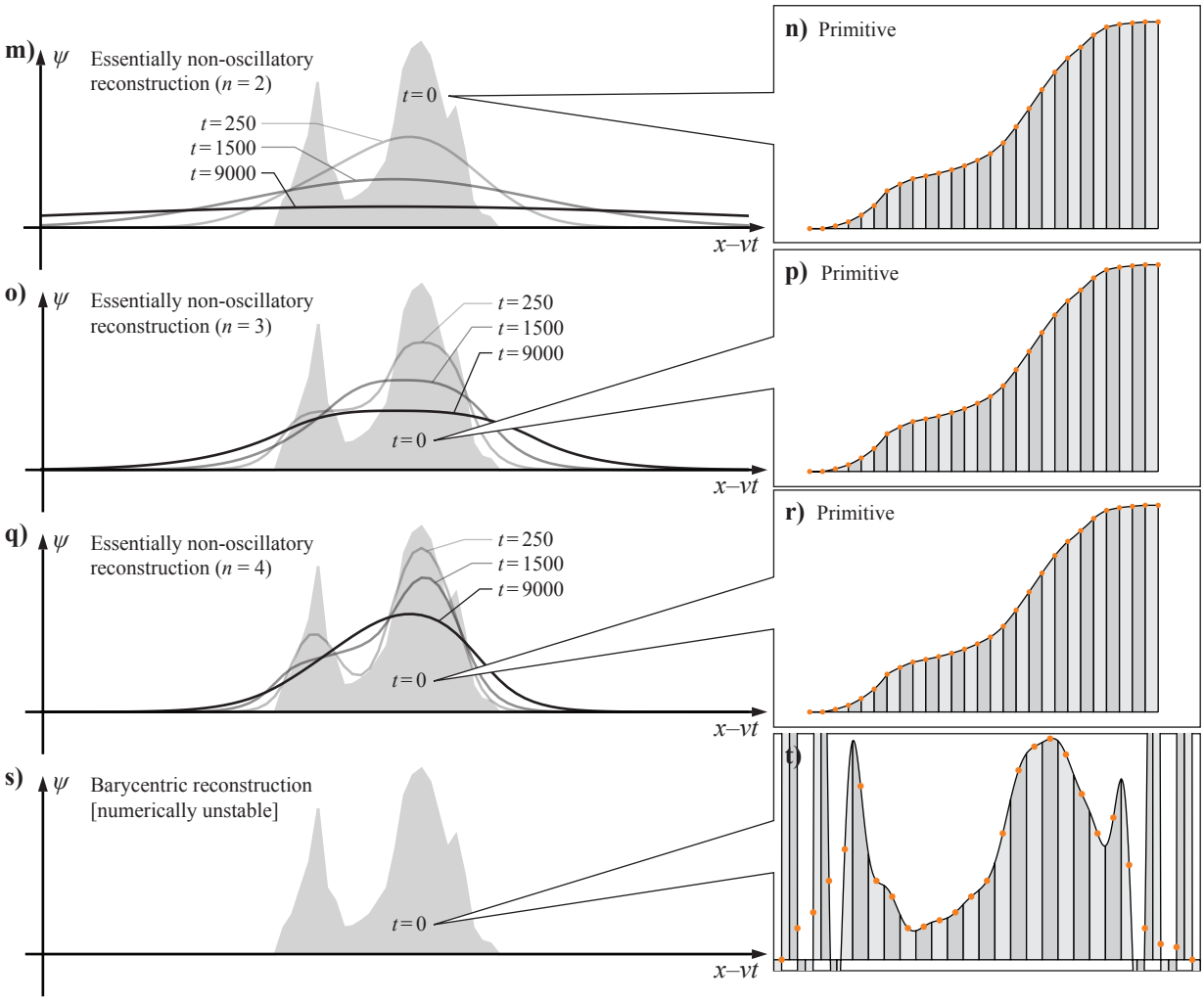

Figure 4.1: Snapshots of a distribution advected by a uniform velocity field, using various reconstruction schemes. A distribution with an initial shape shape shown in solid grey is advected by a uniform velocity field in the positive x–direction. In each plot several windowed snapshots of the distribution are displayed, with the window moving together with the pulse at the uniform velocity. The extent of the full simulation domain is much larger than shown, such that the distribution is not affected by boundary effects. An ideal advection scheme is expected to perfectly preserve the shape of the distribution as it is advected; the deterioration in its shape reflects the reconstruction error of the algorithm (the flux balanced method itself is analytically exact). The grid spacing and time steps of the advection equation simulated are both 1, in dimensionless units, and the velocity of the field is 0.2. The inset figures on the right show the reconstructed distribution at $t = 0$. In each plot, each grey band corresponds to one cell in the grid, and the orange points correspond to the centre of the cells.

4.3 Advection operator

Among the various reconstruction schemes tested, we selected the piecewise parabolic method (Eq. 4.11) to evaluate the flux integral Eq. 4.7, due to its numerical and computational performance. Using this flux, the distribution can be advanced to the next time step as per Eq. 4.6, which gives a numerical solution to the advection operator Eq. 4.2. Note that the choice of PPM reconstruction means advancing ψ_i^j to ψ_i^{j+1} requires the values of ψ_{i-2}^j, ψ_{i-1}^j, ψ_i^j, ψ_{i+1}^j and ψ_{i+2}^j — i.e. a stencil of five cells, symmetrically stretching along the axis of advection x_i. In the case of advection in a higher–dimensional distribution, each column of pixels along the direction of advection can be considered as a 1–D distribution, and the operator shifts the distribution along the column according to the flow field. The various columns spanning the directions orthogonal to the direction of advection are independent of each other as far as the advection operator is concerned, since the stencils of the operator do not span across columns.

On a practical computational grid, each of these columns of pixels is of finite length, which means boundary conditions are need for the two ends of the column. For a boundary where the flow carries material into the simulation domain, an insulating boundary condition is applied, i.e. $\psi_0^{j+1} = \max(0, \psi_0^j - Q_{1/2}^j)$. For a boundary where the flow carries material out of the domain, an absorbing boundary condition is applied, i.e. $\psi_{N-1}^{j+1} = \psi_{N-1}^j - (\psi_{N-1}^j - \psi_{N-2}^j)u_{N-1/2}^j(t_{j+1}-t_j)/(x_{N-1}-x_{N-2})$. The simulation domain is chosen to cover the majority of the distribution, such that the flow into or out of the boundary is negligible.

To apply this general formulation to each specific advection operator in the Vlasov Eq. 4.1, the variables in Eq. 4.2 are replaced by appropriate quantities:

1. Inertial operator: The advection column lies in the z–direction, and the different columns at various r and v_z are independent of each other. The characteristic curve ending at $z_{j+1/2}$ at t_{l+1} starts from $z_{j+1/2} - v_{z\,k}(t_{l+1} - t_l)$ at t_l. i.e.

$$
\begin{aligned}
&x := z \\
&u := v_z \\
&\chi(t_j; (r_i, z_{j+1/2}, v_{z\,k}), t_{j+1}) = \left(r_i, z_{j+1/2} - v_{z\,k}(t_{l+1} - t_l), v_{z\,k}\right).
\end{aligned}
$$

2. Acceleration operator: The advection column lies in the v_z–direction, and the columns at various r and z are independent. The characteristic curve ending at $v_{z\,k+1/2}$ at t_{l+1} approximately starts from $v_{z\,k+1/2} - a_{i,j}^l(t_{l+1}-t_l)$ at t_l. The acceleration $a_{i,j}^l$ is evaluated at the "start" time t_l, which is equivalent to using an explicit time–stepping scheme. This simplification is necessary since $a = -q/m\partial_z(\phi_{\text{vac}} + \phi_{\text{ch}})$, and for a time–stepping

from t_l to t_{l+1}, $\phi_{\rm ch}$ is only available explicitly at t_l.

$$x := v_z$$
$$u := a$$
$$\chi(t_j; (r_i, z_j, v_{z\,k+1/2}), t_{j+1}) = \left(r_i, z_j, v_{z\,k+1/2} - a^l_{i,j}(t_{l+1} - t_l)\right).$$

3. Drift operator: The advection column lies in the v_z–direction, and the columns at various r and z are independent. The characteristic curve ending at $v_{z\,k+1/2}$ at t_{l+1} starts from $v_{z\,k+1/2} + \nu^l_{i,j,k+1/2}(t_{l+1} - t_l)$ at t_l. This is also an explicit stepping scheme as ν is evaluated at the "old" time.

$$x := v_z$$
$$u := -\nu$$
$$\chi(t_j; (r_i, z_j, v_{z\,k+1/2}), t_{j+1}) = \left(r_i, z_j, v_{z\,k+1/2} + \nu^l_{i,j,k+1/2}(t_{l+1} - t_l)\right).$$

4.4 Diffusion operator

Given the diffusion term is usually small compared with other terms in the Vlasov Eq. 4.1, a simple, explicit forward–time–centred–space (FTCS) scheme [53] is used to discretise the diffusion operator after using the chain rule to expand the product $D\,f$ inside the second derivative:

$$\begin{aligned} f^{l+1}_{i,j,k} = f^l_{i,j,k} + \frac{t_{l+1} - t_l}{2\Delta x^2} \quad \Big(\quad & (2D^l_{i,j,k+1} + 2D^l_{i,j,k-1} - 8D^l_{i,j,k}) f^l_{i,j,k} \\ & + (2D^l_{i,j,k} + D^l_{i,j,k+1} - D^l_{i,j,k-1}) f^l_{i,j,k+1} \\ & + (2D^l_{i,j,k} - D^l_{i,j,k+1} + D^l_{i,j,k-1}) f^l_{i,j,k-1} \quad \Big), \end{aligned} \tag{4.17}$$

where $D^l_{i,j,k} \equiv D(r_i, z_j, v_{z\,k}; t_l)$. Note that the diffusion operator has a stencil stretching three cells in the v_z–direction, i.e. the time–stepping of one cell only requires information from its v_z neighbours. The various columns at different (r, z) are independent of each other.

4.5 Implementation

The discretised phase space distribution $f^l_{i,j,k}$ is a three dimensional table of numbers, spanning the r, z and v_z directions. To advance this table in time from t_l to t_{l+1}, the four operators in the Vlasov equation Eq. 4.1 are applied sequentially using operator splitting in the following procedure:

1. Solving for potential: The distribution is flattened in v_z to obtain the charge distribution in (r, z) at time t_l (see Fig. 4.2 b), and the corresponding total potential ϕ is obtained using the method outlined in Sec. 3.1.

2. Acceleration operator: Using this potential, the axial acceleration at time t_l at each point (r, z) is known. This acceleration modifies the velocity of each particle according to their (r, z) position. In the phase space picture, the acceleration operator shifts the distribution in each (r, z) column in the v_z–direction independently of each other.

3. Collision drift operator: The drift operator is next applied, which advects the distribution in v_z. The operation, like the acceleration operator, works on each v_z column individually, and the columns at various (r, z) are independent of each other.

4. Collision diffusion operator: The diffusion operator is next applied, which causes a spread of the distribution in v_z. The operation is independent between columns at different (r, z). After this operator, the velocity–aspect of the distribution is now updated to t_{l+1}, but the spatial aspect is still at t_l since all the advection and diffusion in v_z causes no movement of material in (r, z).

5. Inertial operator: Lastly the inertial operator is applied to advect the distribution along z in each (r, v_z) column independently, according to that column's axial velocity v_z. After this final operator, the distribution is now fully updated to t_{l+1}, and the time–stepping cycle can start over again.

4.6 Parallelisation

As summarised above, the four operators in the Vlasov equation work on columns either in the v_z– or z–direction, and each operator can be applied on orthogonal columns at the same time without requiring sophisticated synchronisation. No operator requires access to cells at multiple r positions (since there are no ∂_r terms in Eq. 4.1), which means that in terms of the Vlasov equation, the (z, v_z) planes at various r are decoupled from each other. This presents a natural two–level parallelisation scheme to maximise the computational resource addressable by the Vlasov simulation. Here we consider a common configuration of high performance computer clusters, where multiple machines are linked through high–speed interconnects and each machine contains multiple Central Processing Unit (CPU) cores. These cores have symmetric access to the memory within that machine, but data stored on different machines must be passed through the interconnects.

In our parallelisation scheme, each of these machines store one "slice" of the phase space distribution at a single r in its memory (Fig. 4.2 c). In steps 2 through 5 each machine applies the four operators along either z or v_z columns on its slice of the phase space, without having to retrieve information from other machines. Within one machine, each operator is further parallelised to work on multiple columns simultaneously through the OpenMP application programming interface (API), with each CPU core assigned to work on a fraction of the columns within that slice. It should be noted that there are operators along both z and v_z directions, which means the distribution within one slice is fully coupled and cannot be

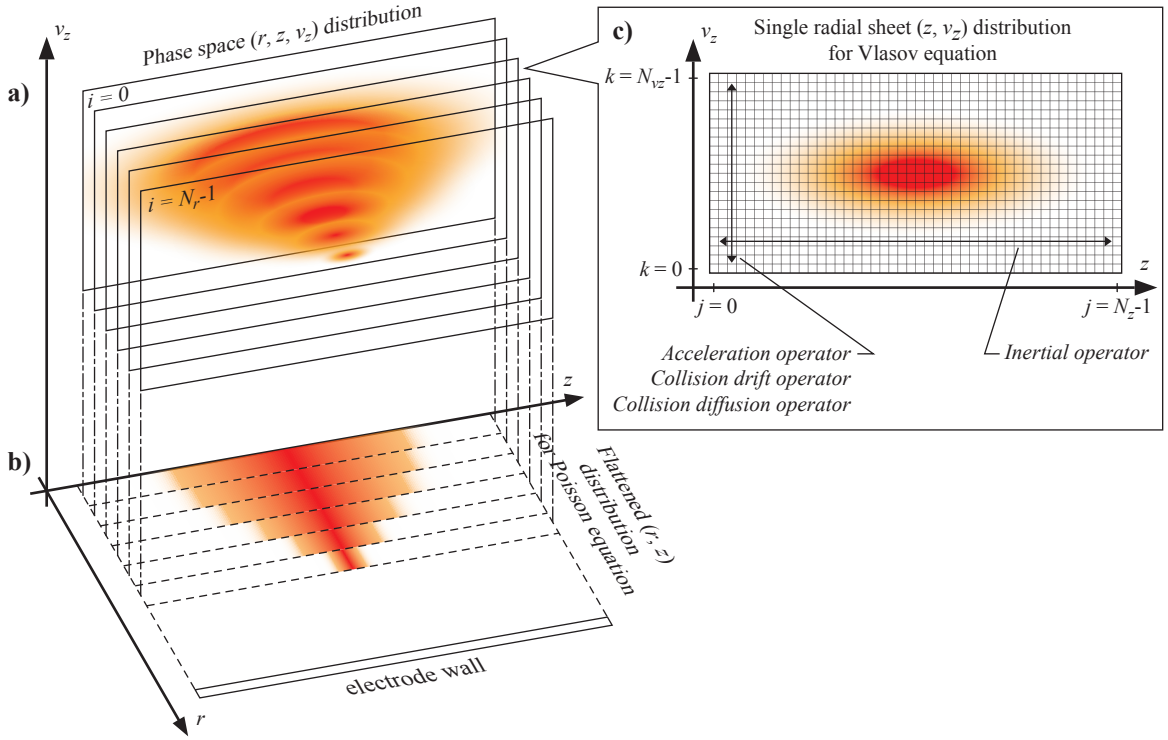

Figure 4.2: a) The geometry of the discretised phase space distribution. b) This distribution is flattened in v_z to produce the (r, z) distribution, from which the electric potential ϕ can be calculated from the Poisson equation. c) Each radial "slice" of the distribution at a fixed r is acted on by the four operators in the Vlasov equation to advance it in time. The direction of action of the operators is indicated in the figure, and each operator only requires the values of the distribution along the column of pixels it is acting on to advance the column in time.

further separated in the same manner as the radial dimension. But since the CPU cores have symmetric access to the machine's memory (and thus the full slice), this inter-machine parallelisation still offers full performance scaling. The implementation of step 1, in contrast, is more complex. In order for each machine to calculate the electric field acting on its radial slice, the full $\rho(r, z)$ is required. We choose to flatten the (z, v_z) distribution in each machine into a 1–D z–distribution (at a fixed r), and use the Open MPI library to circulate each machine's z–distribution to everyone else. Each machine then comes to possess the full (r, z) distribution, and can solve for the total potential using the method described in Sec. 3.1. While not computationally optimal, this approach is simple to implement compared to more sophisticated solutions.

This parallelisation scheme is equally applicable to a single asymmetric many–core machine, e.g. AMD's Bulldozer or Origin's Scalable Shared Memory Multi–Processor architecture. In these machines, a group of cores (a node) have access to its dedicated memory bank, and communication between the nodes is handled by a high–bandwidth inter-nodal

bus. There is no direct access by one node to another node's memory. The code assigns one radial slice to each node, and that node's cores will handle the four operators. The bus, on the other hand, synchronises the copy of the space charge distribution on each node before solving the Poisson equation.

4.7 Reduced domain

Computation efficiency is further increased by applying the operators only in areas of phase space with significant density. A convex profile on each phase space slice demarcates the area of computation (see Fig. 4.3), reducing the simulation domain. This does not alter the applications of the operators during time–stepping, except that each column of pixels can have different length. The shape of the boundary is dynamically adjusted to respond to the changing shape of the distribution. Upon the application of each operator, the amount of flux near the two boundaries of each column is monitored, and the boundaries are moved outward if the flux towards them exceed some threshold in order to keep the distribution within the domain. Conversely, if the flux depletes the cells near the boundaries, the boundaries would be moved inward to save computational resource. This ensures that the profile envelopes the area with significant density as closely as possible, and the flux hitting the computational boundary is minimal. The reduced computational domain has the added advantage that the extent of the phase space slices in z and v_z can be made arbitrarily big, and the algorithm can dynamically determine the area that requires actual computation.

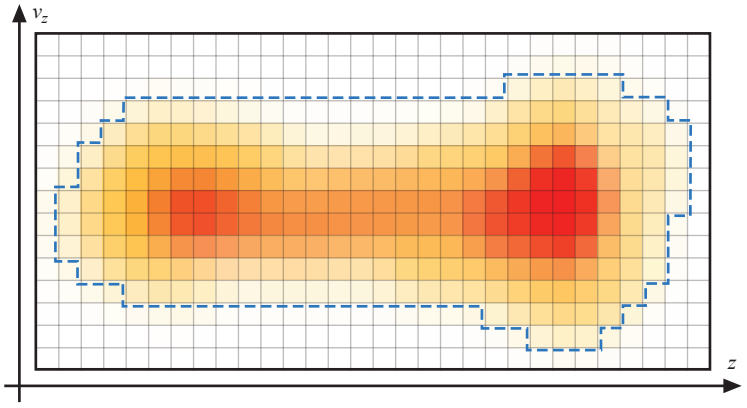

Figure 4.3: The convex profile, plotted in blue dotted line, restricts the simulation domain for each radial slice and conserves computational resource. Note that the profile has to be convex so that there is only one z column for each v_z, and vice versa.

4.8 Annealing initialisation

The initial phase space distribution $f^0_{i,j,k}$ needs to be set before the time–stepping begins. Under most situations, this initial condition is a plasma in an axial thermal equilibrium,

before perturbations are applied inside the simulation time. For low temperatures this equilibrium can be solve for using the waterbag solver described in Ch. 3. However, a more general and convenient method is available through the use of the collisional terms in the Vlasov equation Eq. 4.1. By setting the collisional drift and diffusion coefficients

$$D(r, z, v_z; t) = \text{constant} \qquad \nu(r, z, v_z; t) = D\frac{m}{k_B T} v_z,$$

the plasma will gradually settle from any arbitrary initial distribution into an equilibrium with axial temperature T. The speed of equilibration depends on the choice of the constant D. This process is known as numerical annealing, and can be used to generate a self-consistent equilibrium initial condition before the "proper" time–dependent simulation. This can be important as any error in the initial condition can introduce extra energy (and thus temperature) into the distribution. Numerical annealing can also be used as a standalone equilibrium solver for plasma of arbitrary temperature.

4.9 Comparison with numerical and analytic models

To benchmark our model, we simulated the autoresonant axial excitation of an antiproton plasma, and compared the result to other existing numerical Vlasov solvers' results, as well as first–order, single–particle analytic predictions which ignore the collective effects of a plasma. This is deferred to Sec. 7.2, where we observed a good agreement between our model and others.

Chapter 5

The weakly magnetised collisional operator

In Sec. 1.4 we qualitatively described the strategy of separating the microscopic collisions from the bulk interaction in plasma simulations. Under the binary collision approximation, the forces felt by a particle in a plasma is split into the averaged bulk force due to distant particles, and the short–range force felt during a binary collision. In this approximation, the bulk influence during the short duration of a collision is negligible as the collisional force acting on the two particles is much stronger than the bulk forces — with the possible exception of the magnetic force. In a Penning–Malmberg trap the uniform magnetic field acts on the pair of colliding particles during the collision process. The level of influence this magnetic force has on the outcome of the collision depends on the comparison of a number of scale lengths:

- the cyclotron radius of the incoming particle $\bar{v}_{\perp 1}/\omega_{C\,1}$
- the cyclotron radius of target particle $\bar{v}_{\perp 2}/\omega_{C\,2}$
- the mean distance of closest approach $q_1 q_2/(2\pi\epsilon_0\mu\bar{v}^2)$, where μ is the reduced mass $1/(1/m_1 + 1/m_2)$

If the cyclotron radii of both particles are much bigger than the distance of closest approach, the trajectories of the colliding particles would closely resemble those in free space, as the magnetic force does not significantly alter the trajectories within the scale of the collision. This is known as a weakly magnetised collision. On the other hand, if any of the two cyclotron radii is comparable to the distance of closest approach, special treatment is necessary. In this chapter we present an analytical proof to the qualitative picture above for handling weakly magnetised collisions. We also develop an efficient, energy–conserving numerical scheme to simulate the cumulative effect of collisions on the macroscopic distribution in the weakly magnetised regime. The simulation of intermediately magnetised collisions is deferred to the next chapter.

5.1 Rutherford scattering

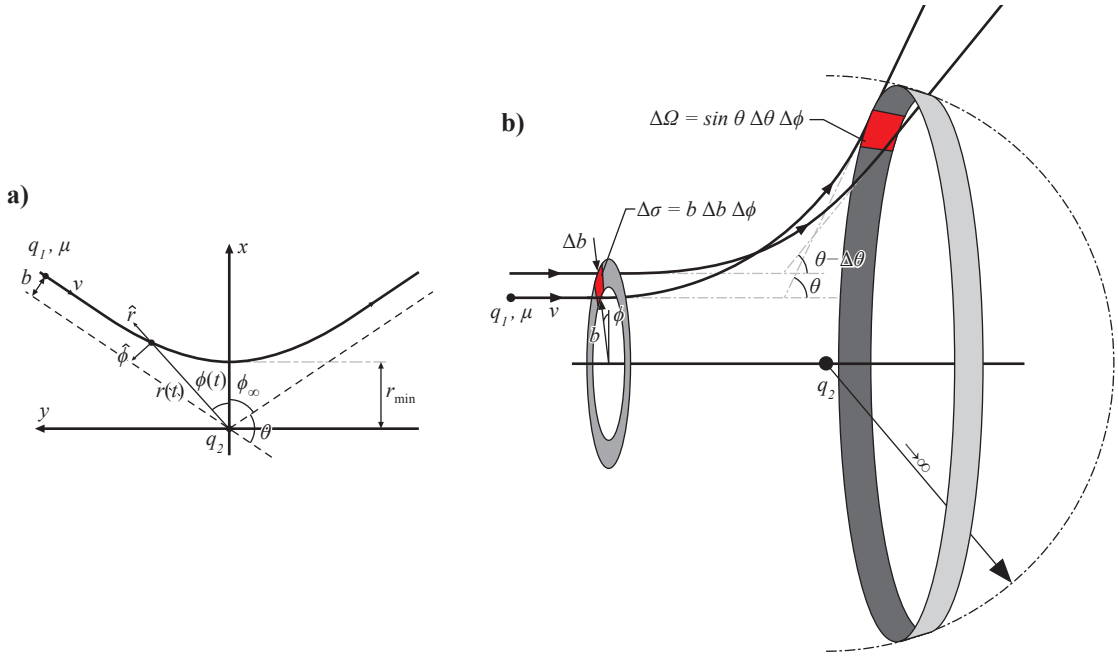

Figure 5.1: a) The coordinates and variables used to describe a Rutherford scattering, in the reduced–mass frame. The particle with charge and (reduced) mass q_1 and μ collides with a fixed collision centre with charge q_2, at an incoming relative speed of v and impact parameter b. The polar coordinates (r, ϕ) tracks the reduced mass during the collision process, and ϕ_∞ is the polar angle as $r \to \infty$. b) The definition of the collisional cross–section. Particles that are injected through the red patch with area $\Delta\sigma$ on the left must exit through the red patch on the right, which sustains a solid angle of $\Delta\Omega$.

To study the collective effect of collisions we first need to calculate the effect of individual collisions. Consider two particles of mass and charge (m_1, q_1) and (m_2, q_2) respectively, travelling towards each other at a relative speed v and impact parameter b. In the centre–of–mass frame, this problem can be described by $\mu\ddot{\boldsymbol{r}} = k/r^2\hat{\boldsymbol{r}}$, where $\boldsymbol{r} \equiv \boldsymbol{r}_2 - \boldsymbol{r}_1$ is the relative distance vector between the two particles, $\mu = (1/m_1 + 1/m_2)^{-1}$ is the reduced mass, and $k = q_1 q_2/(4\pi\epsilon_0)$ is the coefficient for the electrostatic interaction between the two particles. In the polar coordinates defined in Fig. 5.1 a, $\ddot{\boldsymbol{r}} = (\ddot{r} - r\dot{\phi}^2)\hat{\boldsymbol{r}} + (2\dot{r}\dot{\phi} + r\ddot{\phi})\hat{\boldsymbol{\phi}}$, and the equation of motion can be written as

$$\ddot{r} - r\dot{\phi}^2 = \frac{k}{\mu r^2} \qquad \frac{\mathrm{d}}{\mathrm{d}t}(r^2\dot{\phi}) = 0.$$

The second of these two equations can be expressed as $\mu r^2\dot{\phi} \equiv L =$constant, which is simply a statement of the conservation of angular momentum. Using the incoming particle's initial velocity, the angular momentum can also be expressed as $L = \mu b v$. This allows us to

eliminate $\dot{\phi}$ in the first equation to yield

$$\ddot{r} - \frac{L^2}{\mu^2 r^3} = \frac{k}{\mu r^2}. \tag{5.1}$$

The conservation of angular momentum also allows time derivatives to be replaced as $\partial_t = \dot{\phi}\partial_\phi = L/(\mu r^2)\partial_\phi$. Using this to replace the time derivatives in Eq. 5.1, and defining the variable $u \equiv 1/r$, we have

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2} + \left(u + \frac{k\mu}{L^2}\right) = 0. \tag{5.2}$$

Equation 5.2 has the solution $u(\phi) + k\mu/L^2 = \alpha\cos\phi + \beta\sin\phi$. By the symmetry of the coordinate system defined in Fig. 5.1 a, $\beta = 0$ since u should be symmetric around $\phi = 0$. Another boundary condition is available at the point of closest approach, where, by conservation of energy,

$$\frac{1}{2}\mu v^2 = \frac{k}{r_{\min}} + \frac{\mu}{2}\left(\frac{L}{\mu r_{\min}}\right)^2 \quad \Rightarrow \quad \frac{1}{r_{\min}} = \frac{\mu}{L^2}\left(\sqrt{k^2 + v^2 L^2} - k\right)$$

We thus have the second boundary condition ($u = 1/r_{\min}, \phi = 0$). Putting this into Eq. 5.2, the resultant trajectory curve is

$$\frac{1}{r(\phi)} = \frac{\mu k}{L^2}\left(\sqrt{1 + \left(\frac{vL}{k}\right)^2}\cos\phi - 1\right). \tag{5.3}$$

As defined in Fig. 5.1 a, the deflection angle θ is equal to $\pi - 2\phi_\infty$, where ϕ_∞ corresponds to the angle where $r \to \infty$. Setting the right side of Eq. 5.3 to zero, we can solve for these two angles as

$$\cos\phi_\infty = \frac{k}{\sqrt{k^2 + (vL)^2}} \quad \Rightarrow \quad \tan\frac{\theta}{2} = \frac{k}{vL} = \frac{q_1 q_2}{4\pi\epsilon_0 b\mu v^2}. \tag{5.4}$$

To obtain the differential cross–section of Rutherford scattering, we first take derivative of Eq. 5.4 against b to obtain $\mathrm{d}\theta/\,\mathrm{d}b = -\sin\theta/b$. Using the definition of the differential area $\mathrm{d}\sigma$ and differential solid angle $\mathrm{d}\Omega$ in Fig. 5.1 b, we can then write down the well–known Rutherford scattering cross–section

$$\left|\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right| = \left|\frac{b}{\sin\theta}\frac{\mathrm{d}b}{\mathrm{d}\theta}\right| = \left(\frac{q_1 q_2}{8\pi\epsilon_0\mu v^2}\right)^2 \frac{1}{\sin^4(\theta/2)}. \tag{5.5}$$

The absolute value sign is necessary since b and θ go in opposite sense: high impact parameter leads to small deflection angle and vice versa. However, as far as the cross–section area is concerned, the sign is not important.

5.2 Liouville's equation and BBGKY hierarchy

In this section we develop the general statistical description of collisions, following the approach by Moore [60]. At the most fundamental level, the N particles in a plasma can be described by the $6N$ dimension phase space distribution function $g(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N)$, where the vector $\boldsymbol{z}_1 \equiv \{\boldsymbol{r}_1, \boldsymbol{v}_1\}$ corresponds to the phase space position of one particle. Note that the particles in the distribution are considered indistinguishable, which means exchanging any $\boldsymbol{z}_i$ in the argument of g yields the same value. The distribution is the product of $6N$ Dirac delta functions, meaning only one point in the $6N$ phase space is non-zero, that being the exact state of the plasma. The evolution of the distribution g is given by Liouville's equation

$$\frac{\partial g}{\partial t} + \sum_{i=1}^{N} \left(\boldsymbol{v}_i \cdot \nabla_{\boldsymbol{r}_i} g + \frac{\boldsymbol{F}_i}{m} \cdot \nabla_{\boldsymbol{v}_i} g + \sum_{j=1,\, j\neq i}^{N} \frac{\boldsymbol{K}_{ij}}{m} \cdot \nabla_{\boldsymbol{v}_i} g \right) = 0,$$

where $\boldsymbol{F}_i = -\nabla_{\boldsymbol{r}_i}\phi$ is the bulk electrostatic force on particle i, and $\boldsymbol{K}_{ij} = -\nabla_{\boldsymbol{r}_i}\phi_{ij}$ is the pairwise force by particle j on particle i. The magnetic force on the particles are ignored since we are concerned with weakly magnetised collisions, and the bulk motion of the particles can be constrained to $\hat{\boldsymbol{z}}$ *a posteriori*. The Liouville's equation can be rewritten as

$$\left(\frac{\partial}{\partial t} + h(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \right) g = 0, \tag{5.6}$$

where the operator h is defined as

$$h(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) = \sum_{i=1}^{N} \left(\boldsymbol{v}_i \cdot \nabla_{\boldsymbol{r}_i} + \frac{\boldsymbol{F}_i}{m} \cdot \nabla_{\boldsymbol{v}_i} \right) + \frac{1}{2} \sum_{i,j=1}^{N} \frac{\boldsymbol{K}_{ij}}{m} \cdot (\nabla_{\boldsymbol{v}_i} - \nabla_{\boldsymbol{v}_j}). \tag{5.7}$$

The binary interaction term involving $\boldsymbol{K}_{ij}$ has been symmetrised by swapping the indices i and j.

Next we define the n-body distribution function

$$f_n(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n) = \frac{N!}{(N-n)!} \int g(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n, \boldsymbol{z}_{n+1}, \cdots, \boldsymbol{z}_N)\, \mathrm{d}^6 z_{n+1} \cdots \mathrm{d}^6 z_N,$$

which averages over the position of the last $N-n$ particles. Given that the particle labels in g are exchangeable, the one–body distribution $f_1(\boldsymbol{z}_1)$ gives the phase space particle density at $\boldsymbol{z}_1$, and the two–body distribution $f_2(\boldsymbol{z}_1, \boldsymbol{z}_2)$ gives the correlated probability of finding any two particles at $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ simultaneously. To derive a equation of motion for f_n, we integrate the Liouville's equation (Eq. 5.6) over the last $N-n$ variables and multiply both

sides by $N!/(N-n)!$ to obtain

$$\begin{aligned}
-\frac{\partial f_n}{\partial t} = \quad & \frac{N!}{(N-n)!}\int h(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n) g\,\mathrm{d}^6 z_{n+1}\cdots\mathrm{d}^6 z_N \\
& + \frac{N!}{(N-n)!}\int h(\boldsymbol{z}_{n+1},\ldots,\boldsymbol{z}_N) g\,\mathrm{d}^6 z_{n+1}\cdots\mathrm{d}^6 z_N \\
& + \frac{N!}{(N-n)!}\int \sum_{i=1}^{n}\sum_{j=n+1}^{N} \frac{\boldsymbol{K}_{ij}}{m}\cdot(\nabla_{\boldsymbol{v}_i}-\nabla_{\boldsymbol{v}_j}) g\,\mathrm{d}^6 z_{n+1}\cdots\mathrm{d}^6 z_N. \qquad (5.8)
\end{aligned}$$

The first integral in Eq. 5.8 is over variables not involved in the operator h, which means the operator can be commuted with the integral. The second integral is zero, which can be proven using the fact that g is zero outside of some bounded region in phase space. This means that $\int \boldsymbol{v}_i\cdot\nabla_{\boldsymbol{r}_i} g\,\mathrm{d}^3 r_i = \boldsymbol{v}_i\cdot\int\nabla_{\boldsymbol{r}_i} g\,\mathrm{d}^3 r_i$ has to be zero as g is evaluated at $g(\boldsymbol{r}_i\to\infty)$ using the gradient theorem. Similarly, $\int \boldsymbol{F}_i\cdot\nabla_{\boldsymbol{v}_i} g\,\mathrm{d}^3 v_i = 0$, and $\int \boldsymbol{K}_{ij}\cdot\nabla_{\boldsymbol{v}_i} g\,\mathrm{d}^3 v_i = 0$. The third integral can be simplified by dropping the $\nabla_{\boldsymbol{v}_j}$ term, as the j index runs over the integration variables, which means it is zero by the gradient theorem. The third integral is further simplified as follows:

$$\begin{aligned}
& \frac{N!}{(N-n)!}\int \sum_{i=1}^{n}\sum_{j=n+1}^{N} \frac{\boldsymbol{K}_{ij}}{m}\cdot\nabla_{\boldsymbol{v}_i} g\,\mathrm{d}^6 z_{n+1}\cdots\mathrm{d}^6 z_N \\
=& \frac{N!}{(N-n)!}\sum_{i=1}^{n}(N-n)\int \frac{\boldsymbol{K}_{i,n+1}}{m}\cdot\nabla_{\boldsymbol{v}_i} g\,\mathrm{d}^6 z_{n+1}\cdots\mathrm{d}^6 z_N \\
=& \sum_{i=1}^{n}\int \frac{\boldsymbol{K}_{i,n+1}}{m}\cdot\nabla_{\boldsymbol{v}_i}\left(\frac{N!}{(N-n-1)!}\int g\,\mathrm{d}^6 z_{n+2}\cdots\mathrm{d}^6 z_N\right)\mathrm{d}^6 z_{n+1} \\
=& \sum_{i=1}^{n}\int \frac{\boldsymbol{K}_{i,n+1}}{m}\cdot\nabla_{\boldsymbol{v}_i} f_{n+1}(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_{n+1})\,\mathrm{d}^6 z_{n+1}.
\end{aligned}$$

The first step is accomplished by swapping the indices j and $n+1$ in each term of the summation over j. This does not alter the value of g since $g(\cdots,\boldsymbol{z}_{n+1},\cdots,\boldsymbol{z}_j,\cdots) = g(\cdots,\boldsymbol{z}_j,\cdots,\boldsymbol{z}_{n+1},\cdots)$. The second step is done by commuting $\boldsymbol{K}_{i,n+1}$ and $\nabla_{\boldsymbol{v}_i}$ with the integral over the variables $\boldsymbol{z}_{n+2}$ to $\boldsymbol{z}_N$, given they have no overlap. The last step uses the definition of f_{n+1}. Putting these three integrals together, we have converted the Liouville's equation into

$$\begin{aligned}
& \left(\frac{\partial}{\partial t} + h(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_n)\right) f_n(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_n) \\
&= -\sum_{i=1}^{n}\int \frac{\boldsymbol{K}_{i,n+1}}{m}\cdot\nabla_{\boldsymbol{v}_i} f_{n+1}(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_{n+1})\,\mathrm{d}^6 z_{n+1}. \qquad (5.9)
\end{aligned}$$

This is the Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy, which gives the evolution of the n-particle distribution in terms of the next order $n+1$-particle distribution.

5.3 Boltzmann collision integral

Usually the first two equations in the BBGKY hierarchy are of the greatest physical interest, which we will analyse in this section using the approach described by Huang [61]. Explicitly, the first equation in the hierarchy, which is given by Eq. 5.9 with $n = 1$, is

$$\left(\frac{\partial}{\partial t} + \boldsymbol{v}_1 \cdot \nabla_{\boldsymbol{r}_1} + \frac{\boldsymbol{F}_1}{m} \cdot \nabla_{\boldsymbol{v}_1}\right) f_1(\boldsymbol{r}_1, \boldsymbol{v}_1)$$
$$= -\int \frac{\boldsymbol{K}_{12}}{m} \cdot \nabla_{\boldsymbol{v}_1} f_2(\boldsymbol{r}_1, \boldsymbol{v}_1, \boldsymbol{r}_2, \boldsymbol{v}_2)\, \mathrm{d}^3 r_2\, \mathrm{d}^3 v_2 \equiv \left(\frac{\partial f_1}{\partial t}\right)_{\mathrm{col}}. \tag{5.10}$$

This is the most important equation in the hierarchy. The left hand side of this equation contain the "streaming" terms which describe the bulk evolution of the one-particle distribution / phase space density according to the bulk force. This is the Vlasov equation, which we have in effect derived. The right hand side, on the other hand, represent the effects two–body correlations have on the evolution of the phase space density. The right hand side is therefore defined as the collisional operator $(\partial_t f_1)_{\mathrm{col}}$.

The second equation, given by substituting $n = 2$ into Eq. 5.9, is

$$\left(\frac{\partial}{\partial t} + \boldsymbol{v}_1 \cdot \nabla_{\boldsymbol{r}_1} + \frac{\boldsymbol{F}_1}{m} \cdot \nabla_{\boldsymbol{v}_1} + \boldsymbol{v}_2 \cdot \nabla_{\boldsymbol{r}_2} + \frac{\boldsymbol{F}_2}{m} \cdot \nabla_{\boldsymbol{v}_2} + \frac{\boldsymbol{K}_{12}}{m} \cdot \nabla_{\boldsymbol{v}_1}\right) f_2(\boldsymbol{r}_1, \boldsymbol{v}_1, \boldsymbol{r}_2, \boldsymbol{v}_2)$$
$$= -\int \left(\frac{\boldsymbol{K}_{13}}{m} \cdot \nabla_{\boldsymbol{v}_1} + \frac{\boldsymbol{K}_{23}}{m} \cdot \nabla_{\boldsymbol{v}_2}\right) f_3(\boldsymbol{r}_1, \boldsymbol{v}_1, \boldsymbol{r}_2, \boldsymbol{v}_2, \boldsymbol{r}_3, \boldsymbol{v}_3)\, \mathrm{d}^3 r_3\, \mathrm{d}^3 v_3. \tag{5.11}$$

This equation describes the evolution of the the two–body correlation in terms of the three–body correlation. This is qualitatively different from Eq. 5.10 in that there is no bulk streaming term; both sides describe collisional effects. The evolution of the two–body distribution on the left is expressed in terms of the three–body distribution on the right. Since the possibility of three particles being in close proximity simultaneously is generally much lower than the possibility of two particles being in close proximity, the magnitude of the right hand side of Eq. 5.11 involving f_3 is much smaller than the magnitude of the left hand side. The right hand side of Eq. 5.11 is therefore set to zero, as is customarily done in collision analysis. This terminates the BBGYK hierarchy and restricts our model to including only binary collisional effects.

Our objective is to eliminate f_2 altogether from Eqs. 5.10 and 5.11, and obtain a closed equation purely in terms of f_1. For this to be possible, simplifications have to be made. We first assume that particles only interact when their distance is within one Debye length, within which the two particles' trajectories are correlated. When the particles leave the interaction range, they become uncorrelated again. Mathematically, this means the two–body correlation f_2 is separable outside of the interaction range, i.e. $f_2(\boldsymbol{r}_1, \boldsymbol{v}_1, \boldsymbol{r}_2, \boldsymbol{v}_2) = f_1(\boldsymbol{r}_1, \boldsymbol{v}_1) f_1(\boldsymbol{r}_2, \boldsymbol{v}_2)$ when $|\boldsymbol{r}_2 - \boldsymbol{r}_1| > \lambda_D$. Using this, the integral on the right hand side of Eq. 5.10 is non-zero only when $|\boldsymbol{r}_2 - \boldsymbol{r}_1| < \lambda_D$. Secondly, the effect of collisions on

the macroscopic distribution is assumed to be slow in comparison to the time scale of the collisions themselves. For each combination of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ for a collision, there is a constant inflow of particles at $\boldsymbol{z}_1$ and a constant supply of targets at $\boldsymbol{z}_2$ into the collision region, as these flows are provided by the "outside" macroscopic distribution, which is slow–varying. This allows the ∂_t term on the left hand side of Eq. 5.11 to be dropped, as the effect of the collisions is quasi-steady state, compared to the magnitude of the other terms in Eq. 5.11. Thirdly, we assume that the bulk forces are negligible during the course of a collision, as the close–range interaction is much stronger than the bulk force. This means terms involving the bulk force $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ on the left hand side of Eq. 5.11 can be dropped when $|\boldsymbol{r}_2 - \boldsymbol{r}_1| < \lambda_D$. Putting all these simplifications together, Eq. 5.11 becomes

$$\left(\boldsymbol{v}_1 \cdot \nabla_{\boldsymbol{r}_1} + \boldsymbol{v}_2 \cdot \nabla_{\boldsymbol{r}_2} + \frac{\boldsymbol{K}_{12}}{m} \cdot \nabla_{\boldsymbol{v}_1}\right) f_2(\boldsymbol{r}_1, \boldsymbol{v}_1, \boldsymbol{r}_2, \boldsymbol{v}_2) \approx 0 \qquad \text{for } |\boldsymbol{r}_2 - \boldsymbol{r}_1| < \lambda_D.$$

Inserting this back into Eq. 5.10 to replace $\boldsymbol{K}_{12}$, while noting the integral is non-zero only over the domain $|\boldsymbol{r}_2 - \boldsymbol{r}_1| < \lambda_D$, we can rewrite the collision operator as

$$\left(\frac{\partial f_1}{\partial t}\right)_{\text{col}} = \int_{|\boldsymbol{r}_2 - \boldsymbol{r}_1| < \lambda_D} (\boldsymbol{v}_1 \cdot \nabla_{\boldsymbol{r}_1} + \boldsymbol{v}_2 \cdot \nabla_{\boldsymbol{r}_2})\, f_2(\boldsymbol{r}_1, \boldsymbol{v}_1, \boldsymbol{r}_2, \boldsymbol{v}_2)\, \mathrm{d}^3 r_2\, \mathrm{d}^3 v_2.$$

Introducing the centre of mass coordinates $\boldsymbol{R} = (\boldsymbol{r}_1 + \boldsymbol{r}_2)/2$ and $\boldsymbol{r} = \boldsymbol{r}_2 - \boldsymbol{r}_1$, and the corresponding $\boldsymbol{V} = \dot{\boldsymbol{R}}$ and $\boldsymbol{v} = \dot{\boldsymbol{r}}$, we can rewrite the gradients and the integral over $\boldsymbol{r}_2$ in the collisional integral into

$$\left(\frac{\partial f_1}{\partial t}\right)_{\text{col}} = \int \mathrm{d}^3 v_2 \boldsymbol{v} \cdot \int_{r < \lambda_D} \mathrm{d}^3 r \nabla_{\boldsymbol{r}} f_2.$$

The $\nabla_{\boldsymbol{R}}$ term is omitted assuming the background distribution is uniform over the collisional scale. Now define a cylindrical coordinate for the integral over $\boldsymbol{r}$, where the z axis is aligned in the direction of $\boldsymbol{v}$, and the cylindrical radius and azimuthal angle are labelled b and ϕ respectively. Since $\boldsymbol{v}$ is the relative velocity of the incoming particle, this coordinate system is also the one used in Fig. 5.1. In this coordinate, $\boldsymbol{v} \cdot \nabla_{\boldsymbol{r}}$ becomes $|\boldsymbol{v}|\partial_z$. The collision operator then becomes

$$\left(\frac{\partial f_1}{\partial t}\right)_{\text{col}} = \int \mathrm{d}^3 v_2\, |\boldsymbol{v}| \int_{b < \lambda_D} b\, \mathrm{d}b\, \mathrm{d}\phi\, (f_{2,\text{exit}} - f_{2,\text{entry}})\,,$$

where $f_{2,\text{entry}}$ refers to the distribution at the point where the collision trajectory enters the interaction range λ_D, and $f_{2,\text{exit}}$ refers to the exit point. Since these points are on the boundary of the interaction range, $f_{2,\text{exit}}$ and $f_{2,\text{entry}}$ should equal to $n_0^2 f_1(\boldsymbol{v}_1) f_1(\boldsymbol{v}_2)$ and $n_0^2 f_1(\boldsymbol{v}_1') f_1(\boldsymbol{v}_2')$ respectively to preserve continuity with the outside uncorrelated region. Here $\boldsymbol{v}_1'$ and $\boldsymbol{v}_2'$ are the final velocities of two particles after they collide with initial velocities $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ and impact parameter b. The 6–D distribution $f_1(\boldsymbol{r}, \boldsymbol{v})$ is replaced with $n_0 f_1(\boldsymbol{v})$, where n_0 is the spatial density of the particles, since the distribution is approximately uniform in

the collisional scale. Using the Rutherford scattering cross–section Eq. 5.5 to rewrite the integral over area into one over solid angle, we have

$$\left(\frac{\partial f_1}{\partial t}\right)_{\text{col}} = \int \mathrm{d}^3 v_2 n_0 \left|\boldsymbol{v}_2 - \boldsymbol{v}_1\right| \int_{\theta>\theta_{\min}} \mathrm{d}\Omega \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \left(f_1(\boldsymbol{v}_1') f_1(\boldsymbol{v}_2') - f_1(\boldsymbol{v}_1) f_1(\boldsymbol{v}_2)\right), \tag{5.12}$$

which is the Boltzmann collision integral. The integral over the scattering angle θ has a lower limit of $\theta_{\min}$, which corresponds to an *upper* limit of the impact parameter b (recall that the scattering angle increases as the impact parameter decreases). We have thus converted Eqs. 5.10 and 5.11 into a closed equation which involves only f_1.

5.4 Collisional Fokker–Planck equation

While the Coulomb collision integral forms a closed system that describes the evolution of f_1, it is not completely suitable for numerical computation. In this section we use the approach outlined by Montgomery and Tidman [62] to convert Eq. 5.12 into a differential instead of an integral–differential equation.

In the course of collision, the initial velocities of the two colliding particles $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ become $\boldsymbol{v}_1'$ and $\boldsymbol{v}_2'$ respectively when they leave the interaction range. Defining the change in velocity $\Delta\boldsymbol{v}_1 \equiv \boldsymbol{v}_1' - \boldsymbol{v}_1$ and $\Delta\boldsymbol{v}_2 \equiv \boldsymbol{v}_2' - \boldsymbol{v}_2$, conservation of momentum requires that $\Delta\boldsymbol{v}_1 + \Delta\boldsymbol{v}_2 = 0$. Furthermore, conservation of energy requires that the magnitude of the relative velocity before collision $\boldsymbol{v} \equiv \boldsymbol{v}_2 - \boldsymbol{v}_1$ and after collision $\boldsymbol{v}' \equiv \boldsymbol{v}_2' - \boldsymbol{v}_1'$ have to be equal. This means in the coordinates system of Fig. 5.1 b,

$$\Delta\boldsymbol{v}_2 = \frac{\Delta\boldsymbol{v}}{2} \qquad \Delta\boldsymbol{v}_1 = -\frac{\Delta\boldsymbol{v}}{2}$$
$$\Delta\boldsymbol{v} = |\boldsymbol{v}| \left((\cos\theta - 1)\hat{\boldsymbol{z}} + \sin\theta\cos\phi\ \hat{\boldsymbol{x}} + \sin\theta\sin\phi\ \hat{\boldsymbol{y}}\right).$$

In the last section we have shown that the collision integral is bounded below by a minimum scattering angle, due to Debye shielding. It can be shown that the integral can also be bounded above by $\theta_{\max} \ll \pi/2$, since the cumulative effect of large angle scattering (ones which changes $\boldsymbol{v}$ by an angle comparable to π) is much less than that of small angle scattering. This means we can consider $\Delta\boldsymbol{v}_1$ and $\Delta\boldsymbol{v}_2$ as small parameters without significantly changing the value of the collision integral (given appropriate lower and upper limits for θ). Expanding Eq. 5.12 to second order in $\Delta\boldsymbol{v}_1$ and $\Delta\boldsymbol{v}_2$, we have

$$\begin{aligned}\left(\frac{\partial f_1}{\partial t}\right)_{\text{col}} = \int \mathrm{d}^3 v_2 n_0 \left|\boldsymbol{v}_2 - \boldsymbol{v}_1\right| \int_{\theta_{\min}<\theta<\theta_{\max}} \mathrm{d}\Omega \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \times \\ \left\{ f_1(\boldsymbol{v}_1)\Delta\boldsymbol{v}_2 \cdot \frac{\partial f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2} + f_2(\boldsymbol{v}_2)\Delta\boldsymbol{v}_1 \cdot \frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1} \right. \\ \left. + \frac{1}{2} f_1(\boldsymbol{v}_1)\Delta\boldsymbol{v}_2\Delta\boldsymbol{v}_2 : \frac{\partial^2 f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2 \partial \boldsymbol{v}_2} + \frac{1}{2} f_1(\boldsymbol{v}_2)\Delta\boldsymbol{v}_1\Delta\boldsymbol{v}_1 : \frac{\partial^2 f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1}\end{aligned}$$

$$+ \Delta\boldsymbol{v}_1\Delta\boldsymbol{v}_2 : \left(\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\frac{\partial f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2}\right)\Bigg\}. \tag{5.13}$$

As only $\Delta\boldsymbol{v}_1$ and $\Delta\boldsymbol{v}_2$ depend on the scattering angles θ and ϕ, f_1 and its gradients can be extracted outside the integral over the solid angle Ω. Using the differential cross–section Eq. 5.5, and assuming the ordering $\pi/2 \gg \theta_{\max} \gg \theta_{\min} > 0$, we have

$$\int_{\theta_{\min}<\theta<\theta_{\max}} \mathrm{d}\Omega \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\Delta\boldsymbol{v} = -\frac{q^4}{4\pi\epsilon_0^2\mu^2|\boldsymbol{v}|^3}\log\left(\frac{\theta_{\max}}{\theta_{\min}}\right)\hat{\boldsymbol{z}}$$
$$\int_{\theta_{\min}<\theta<\theta_{\max}} \mathrm{d}\Omega \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\Delta\boldsymbol{v}\Delta\boldsymbol{v} = \frac{q^4}{4\pi\epsilon_0^2\mu^2|\boldsymbol{v}|^2}\log\left(\frac{\theta_{\max}}{\theta_{\min}}\right)(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}),$$

where $\mu = m/2$ is the reduced mass. All small terms of $O(\theta_{\min}^2)$ or $O(\theta_{\max}^2)$ and higher have been dropped, leaving only the logarithmic terms. Rewriting the unit vectors in a coordinate–independent form, Eq. 5.13 becomes

$$\begin{aligned}\left(\frac{\partial f_1}{\partial t}\right)_{\mathrm{col}} =& -\frac{n_0 q^4\log(\theta_{\max}/\theta_{\min})}{4\pi\epsilon_0^2\mu^2}\frac{1}{2}\int \mathrm{d}^3v_2 \frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}\cdot\left(f_1(\boldsymbol{v}_1)\frac{\partial f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2} - f_1(\boldsymbol{v}_2)\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\right)\\ &+ \frac{n_0 q^4\log(\theta_{\max}/\theta_{\min})}{4\pi\epsilon_0^2\mu^2}\frac{1}{8}\int \mathrm{d}^3v_2\left(\frac{\hat{\boldsymbol{I}}}{|\boldsymbol{v}|} - \frac{\boldsymbol{v}\boldsymbol{v}}{|\boldsymbol{v}|^3}\right):\\ &\left(f_1(\boldsymbol{v}_1)\frac{\partial^2 f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2\partial \boldsymbol{v}_2} + f_1(\boldsymbol{v}_2)\frac{\partial^2 f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1\partial \boldsymbol{v}_1} - 2\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\frac{\partial f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2}\right).\end{aligned} \tag{5.14}$$

The first of the two integrals in Eq. 5.14 can be simplified as follows:

$$\begin{aligned}&\int \mathrm{d}^3v_2 \frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}\cdot\left(f_1(\boldsymbol{v}_1)\frac{\partial f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2} - f_1(\boldsymbol{v}_2)\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\right)\\ &= \int \mathrm{d}^3v_2 f_1(\boldsymbol{v}_2)\left(-f_1(\boldsymbol{v}_1)\frac{\partial}{\partial \boldsymbol{v}_2} - \frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\right)\cdot\frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}\\ &= \int \mathrm{d}^3v_2 f_1(\boldsymbol{v}_2)\left(f_1(\boldsymbol{v}_1)\frac{\partial}{\partial \boldsymbol{v}_1}\cdot\frac{\boldsymbol{v}}{|\boldsymbol{v}|^3} - \frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\cdot\frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}\right)\\ &= \int \mathrm{d}^3v_2 f_1(\boldsymbol{v}_2)\left(\frac{\partial}{\partial \boldsymbol{v}_1}\cdot\left(f_1(\boldsymbol{v}_1)\frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}\right) - 2\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\cdot\frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}\right)\\ &= \frac{\partial}{\partial \boldsymbol{v}_1}\cdot\left(f_1(\boldsymbol{v}_1)\frac{\partial}{\partial \boldsymbol{v}_1}\int \mathrm{d}^3v_2 f_1(\boldsymbol{v}_2)\frac{1}{|\boldsymbol{v}|}\right) - 2\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\cdot\int \mathrm{d}^3v_2 f_1(\boldsymbol{v}_2)\frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}.\end{aligned}$$

In going from the first line to the second, we have used integration by parts to convert the $\partial_{\boldsymbol{v}_2}$ acting on $f_1(\boldsymbol{v}_2)$ into one acting on $\boldsymbol{v}/|\boldsymbol{v}|^3$. The boundary term is ignored since $f_1(\boldsymbol{v}_2)$ is bounded. In the third line we used the fact that $\partial_{\boldsymbol{v}_2} = -\partial_{\boldsymbol{v}_1}$ when it is acting on a function purely of $\boldsymbol{v} = \boldsymbol{v}_2 - \boldsymbol{v}_1$. The fourth line exploits the chain rule. The fifth uses the vector equality $\partial_{\boldsymbol{v}_1}(1/|\boldsymbol{v}|) = \boldsymbol{v}/|\boldsymbol{v}|^3$.

The second integral in Eq. 5.14, on the other hand, can be simplified as

$$\begin{aligned}
&\int \mathrm{d}^3 v_2 \left(\frac{\hat{\boldsymbol{I}}}{|\boldsymbol{v}|} - \frac{\boldsymbol{v}\boldsymbol{v}}{|\boldsymbol{v}|^3}\right) : \left(f_1(\boldsymbol{v}_1)\frac{\partial^2 f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2 \partial \boldsymbol{v}_2} + f_1(\boldsymbol{v}_2)\frac{\partial^2 f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1} - 2\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\frac{\partial f_1(\boldsymbol{v}_2)}{\partial \boldsymbol{v}_2}\right) \\
&= \int \mathrm{d}^3 v_2 f_1(\boldsymbol{v}_2) \left(f_1(\boldsymbol{v}_1)\frac{\partial^2}{\partial \boldsymbol{v}_2 \partial \boldsymbol{v}_2} + \frac{\partial^2 f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1} + 2\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\frac{\partial}{\partial \boldsymbol{v}_2}\right) : \left(\frac{\hat{\boldsymbol{I}}}{|\boldsymbol{v}|} - \frac{\boldsymbol{v}\boldsymbol{v}}{|\boldsymbol{v}|^3}\right) \\
&= \int \mathrm{d}^3 v_2 f_1(\boldsymbol{v}_2) \left(f_1(\boldsymbol{v}_1)\frac{\partial^2}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1} + \frac{\partial^2 f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1} - 2\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\frac{\partial}{\partial \boldsymbol{v}_1}\right) : \left(\frac{\hat{\boldsymbol{I}}}{|\boldsymbol{v}|} - \frac{\boldsymbol{v}\boldsymbol{v}}{|\boldsymbol{v}|^3}\right) \\
&= \int \mathrm{d}^3 v_2 f_1(\boldsymbol{v}_2) \left\{\frac{\partial^2}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1} : \left(f_1(\boldsymbol{v}_1)\frac{\partial^2 |\boldsymbol{v}|}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1}\right) - 4\left(\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\frac{\partial}{\partial \boldsymbol{v}_1}\right) : \left(\frac{\hat{\boldsymbol{I}}}{|\boldsymbol{v}|} - \frac{\boldsymbol{v}\boldsymbol{v}}{|\boldsymbol{v}|^3}\right)\right\} \\
&= \frac{\partial^2}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1} : \left(f_1(\boldsymbol{v}_1)\frac{\partial^2}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1}\int \mathrm{d}^3 v_2 f_1(\boldsymbol{v}_2)|\boldsymbol{v}|\right) - 8\frac{\partial f_1(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1} \cdot \int \mathrm{d}^3 v_2 f_1(\boldsymbol{v}_2)\frac{\boldsymbol{v}}{|\boldsymbol{v}|^3}.
\end{aligned}$$

In going from the first line to the second, integration by parts is used on derivatives against $\boldsymbol{v}_2$. The third line uses $\partial_{\boldsymbol{v}_2} = -\partial_{\boldsymbol{v}_1}$ when the derivative acts on a function purely of $\boldsymbol{v}$. The fourth makes use of the chain rule, and the identity $\partial_{\boldsymbol{v}_1}\partial_{\boldsymbol{v}_1}|\boldsymbol{v}| = \hat{\boldsymbol{I}}/|\boldsymbol{v}| - \boldsymbol{v}\boldsymbol{v}/|\boldsymbol{v}|^3$. The fifth exploits the identity $\partial_{\boldsymbol{v}_1} \cdot (\hat{\boldsymbol{I}}/|\boldsymbol{v}| - \boldsymbol{v}\boldsymbol{v}/|\boldsymbol{v}|^3) = 2\boldsymbol{v}/|\boldsymbol{v}|^3$.

Putting these two simplifications back into Eq. 5.14, we notice that the second term coming from the simplification of the two integrals cancel each other. The resulting expression can be written as

$$\left(\frac{\partial f_1}{\partial t}\right)_{\mathrm{col}} = \frac{\partial}{\partial \boldsymbol{v}_1} \cdot \left(f_1(\boldsymbol{v}_1)\frac{\partial h(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1}\right) + \frac{\partial^2}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1} : \left(f_1(\boldsymbol{v}_1)\frac{\partial^2 g(\boldsymbol{v}_1)}{\partial \boldsymbol{v}_1 \partial \boldsymbol{v}_1}\right) \tag{5.15}$$

where the Rosenbluth potentials [63] $h(\boldsymbol{v}_1)$ and $g(\boldsymbol{v}_1)$ are defined by

$$h(\boldsymbol{v}_1) = -2\Gamma \int \mathrm{d}^3 v_2 f_1(\boldsymbol{v}_2)\frac{1}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|} \tag{5.16}$$

$$g(\boldsymbol{v}_1) = \frac{\Gamma}{2}\int \mathrm{d}^3 v_2 f_1(\boldsymbol{v}_2)|\boldsymbol{v}_2 - \boldsymbol{v}_1|, \tag{5.17}$$

and $\Gamma = \log(\theta_{\max}/\theta_{\min})n_0 q^4/(4\pi\epsilon_0^2 m^2)$. This is the collision operator expressed as a Fokker–Planck equation, which describes the collective velocity space drift and diffusion caused by collisions. The derivatives of the Rosenbluth potentials give the drift and diffusion coefficients.

5.5 Velocity–cylindrical coordinates

In our weakly magnetised system, the particles undergo gyro-rotation in between collisions. Assuming the gyromotion is uncorrelated, the polar angle of the velocities of particles are

randomised. This means we can express the phase space density f_1 as a function of v_z and $v_\perp \equiv \sqrt{v_x^2 + v_y^2}$, and not of $\theta \equiv \tan^{-1}(v_y/v_x)$. (Here the z axis is along the magnetic field.) In the cylindrical coordinates, the vector and tensor ∇ operators are generally expressed as

$$\frac{\partial F(r,\theta,z)}{\partial \boldsymbol{r}} = \hat{\boldsymbol{r}}\frac{\partial F}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial F}{\partial \theta} + \hat{\boldsymbol{z}}\frac{\partial F}{\partial z} \qquad\qquad \frac{\partial}{\partial \boldsymbol{r}} \cdot \boldsymbol{V} = \frac{1}{r}\frac{\partial (rV_r)}{\partial r} + \frac{1}{r}\frac{\partial V_\theta}{\partial \theta} + \frac{\partial V_z}{\partial z}$$

$$\begin{aligned}\frac{\partial^2 F(r,\theta,z)}{\partial \boldsymbol{r}\partial \boldsymbol{r}} = {} & \hat{\boldsymbol{r}}\hat{\boldsymbol{r}}\left(\frac{\partial^2 F}{\partial r^2}\right) + \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}\left(\frac{1}{r^2}\frac{\partial^2 F}{\partial \theta^2} + \frac{1}{r}\frac{\partial F}{\partial r}\right) + \hat{\boldsymbol{z}}\hat{\boldsymbol{z}}\left(\frac{\partial^2 F}{\partial z^2}\right) \\ & + 2\hat{\boldsymbol{r}}\hat{\boldsymbol{\theta}}\left(\frac{\partial^2}{\partial r \partial \theta}\frac{F}{r}\right) + 2\hat{\boldsymbol{r}}\hat{\boldsymbol{z}}\left(\frac{\partial^2 F}{\partial r \partial z}\right) + 2\hat{\boldsymbol{\theta}}\hat{\boldsymbol{z}}\left(\frac{1}{r}\frac{\partial^2 F}{\partial \theta \partial z}\right)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial \boldsymbol{r}\partial \boldsymbol{r}} : \overleftrightarrow{\boldsymbol{T}} = {} & \frac{1}{r}\frac{\partial^2 (rT^{rr})}{\partial r^2} + \left(\frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} - \frac{1}{r}\frac{\partial}{\partial r}\right) T^{\theta\theta} + \frac{\partial^2 T^{zz}}{\partial z^2} \\ & + \frac{1}{r^2}\frac{\partial^2 (rT^{r\theta})}{\partial r \partial \theta} + \frac{1}{r}\frac{\partial^2 (rT^{rz})}{\partial r \partial z} + \frac{1}{r}\frac{\partial^2 T^{\theta z}}{\partial \theta \partial z}.\end{aligned}$$

Applying these to rewrite the Fokker–Planck equation (Eq. 5.15), we first discard the ∂_θ terms since f_1 is considered to have no θ–dependence. Labelling $\boldsymbol{v}_1 \equiv (\tau\cos\theta,\, \tau\sin\theta,\, \xi)$ and $\boldsymbol{v}_2 \equiv (\tau^*\cos\theta^*,\, \tau^*\sin\theta^*,\, \xi^*)$, we have

$$\begin{aligned}\left(\frac{\partial f_1}{\partial t}\right)_{\text{col}} = {} & \frac{1}{\tau}\frac{\partial}{\partial \tau}\left(\tau f_1 \frac{\partial h}{\partial \tau} - \frac{f_1}{\tau}\frac{\partial g}{\partial \tau}\right) + \frac{\partial}{\partial \xi}\left(f_1 \frac{\partial h}{\partial \xi}\right) \\ & + \frac{1}{\tau}\frac{\partial^2}{\partial \tau^2}\left(\tau f_1 \frac{\partial^2 g}{\partial \tau^2}\right) + \frac{\partial^2}{\partial \xi^2}\left(f_1 \frac{\partial^2 g}{\partial \xi^2}\right) + \frac{2}{\tau}\frac{\partial^2}{\partial \tau \partial \xi}\left(\tau f_1 \frac{\partial^2 g}{\partial \tau \partial \xi}\right).\end{aligned}$$

Averaging both sides by $1/(2\pi)\int_0^{2\pi} \mathrm{d}\theta$, it can be shown that

$$\left(\frac{\partial f_1}{\partial t}\right)_{\text{col}} = -\left(\frac{1}{\tau}\frac{\partial}{\partial \tau}\tau\right) J^\tau - \frac{\partial}{\partial \xi} J^\xi \tag{5.18}$$

where J^τ and J^ξ are the phase space fluxes in the radial and axial directions. These fluxes are given by

$$J^\tau = f_1 \nu^\tau + \left(\frac{1}{\tau}\frac{\partial}{\partial \tau}\tau\right)(f_1 D^{\tau\tau}) + \left(\frac{\partial}{\partial \xi}\right)(f_1 D^{\tau\xi}) \tag{5.19}$$

$$J^\xi = f_1 \nu^\xi + \left(\frac{\partial}{\partial \xi}\right)(f_1 D^{\xi\xi}) + \left(\frac{1}{\tau}\frac{\partial}{\partial \tau}\tau\right)(f_1 D^{\tau\xi}). \tag{5.20}$$

The variables ν and D are essentially the advection and diffusion coefficients, which determines the flow and diffusion of the distribution in the (τ, ξ) phase space. These coefficients are defined by

$$\nu^\tau(\tau,\xi) \quad = \int 2\pi\tau^* \,\mathrm{d}\tau^* \,\mathrm{d}\xi^* f_1(\tau^*,\xi^*)\, \mathcal{C}^\tau(\tau,\tau^*,\xi^*-\xi) \tag{5.21}$$

$$\nu^{\xi}(\tau,\xi) \quad = \int 2\pi\tau^* \,\mathrm{d}\tau^* \,\mathrm{d}\xi^* f_1(\tau^*,\xi^*)\, \mathcal{C}^{\xi}(\tau,\tau^*,\xi^*-\xi) \tag{5.22}$$

$$D^{\tau\tau}(\tau,\xi) = \int 2\pi\tau^* \,\mathrm{d}\tau^* \,\mathrm{d}\xi^* f_1(\tau^*,\xi^*)\, \mathcal{C}^{\tau\tau}(\tau,\tau^*,\xi^*-\xi) \tag{5.23}$$

$$D^{\xi\xi}(\tau,\xi) = \int 2\pi\tau^* \,\mathrm{d}\tau^* \,\mathrm{d}\xi^* f_1(\tau^*,\xi^*)\, \mathcal{C}^{\xi\xi}(\tau,\tau^*,\xi^*-\xi) \tag{5.24}$$

$$D^{\tau\xi}(\tau,\xi) = \int 2\pi\tau^* \,\mathrm{d}\tau^* \,\mathrm{d}\xi^* f_1(\tau^*,\xi^*)\, \mathcal{C}^{\tau\xi}(\tau,\tau^*,\xi^*-\xi). \tag{5.25}$$

Here the "interaction" coefficients $\mathcal{C}^{(*)}$ specify the advective and diffusional influence felt by the distribution at (τ,ξ), when the particles in that region of the phase space collide with particles in another region at (τ^*,ξ^*). These coefficients are given by

$$\mathcal{C}^{\tau}(\tau,\tau^*,\xi^*-\xi) \;\equiv \frac{\Gamma}{4\pi^2}\iint \mathrm{d}\theta\,\mathrm{d}\theta^* \left(2\frac{\partial}{\partial\tau}\frac{1}{|\boldsymbol{v}_2-\boldsymbol{v}_1|}+\frac{1}{2\tau^2}\frac{\partial}{\partial\tau}|\boldsymbol{v}_2-\boldsymbol{v}_1|\right) \tag{5.26}$$

$$\mathcal{C}^{\xi}(\tau,\tau^*,\xi^*-\xi) \;\equiv \frac{\Gamma}{4\pi^2}\iint \mathrm{d}\theta\,\mathrm{d}\theta^* \left(2\frac{\partial}{\partial\xi}\frac{1}{|\boldsymbol{v}_2-\boldsymbol{v}_1|}\right) \tag{5.27}$$

$$\mathcal{C}^{\tau\tau}(\tau,\tau^*,\xi^*-\xi) \equiv -\frac{\Gamma}{4\pi^2}\iint \mathrm{d}\theta\,\mathrm{d}\theta^* \left(\frac{1}{2}\frac{\partial^2}{\partial\tau^2}|\boldsymbol{v}_2-\boldsymbol{v}_1|\right) \tag{5.28}$$

$$\mathcal{C}^{\xi\xi}(\tau,\tau^*,\xi^*-\xi) \equiv -\frac{\Gamma}{4\pi^2}\iint \mathrm{d}\theta\,\mathrm{d}\theta^* \left(\frac{1}{2}\frac{\partial^2}{\partial\xi^2}|\boldsymbol{v}_2-\boldsymbol{v}_1|\right) \tag{5.29}$$

$$\mathcal{C}^{\tau\xi}(\tau,\tau^*,\xi^*-\xi) \equiv -\frac{\Gamma}{4\pi^2}\iint \mathrm{d}\theta\,\mathrm{d}\theta^* \left(\frac{1}{2}\frac{\partial^2}{\partial\tau\partial\xi}|\boldsymbol{v}_2-\boldsymbol{v}_1|\right). \tag{5.30}$$

These can be explicitly evaluated using the relation $|\boldsymbol{v}_2-\boldsymbol{v}_1|^2 = \tau^{*2}+\tau^2+(\xi^*-\xi)^2 - 2\tau^*\tau\cos(\theta^*-\theta)$, which yields

$$\mathcal{C}^{\tau}(\tau,\tau^*,\xi^*-\xi) \;= \frac{\Gamma}{\pi}\left\{\frac{\partial}{\partial\tau}\left(\frac{2}{\sqrt{2\tau\tau^*}}R_2(k)\right)+\frac{1}{4\tau^2}\frac{\partial}{\partial\tau}\left(2\sqrt{2\tau\tau^*}R_1(k)\right)\right\} \tag{5.31}$$

$$\mathcal{C}^{\xi}(\tau,\tau^*,\xi^*-\xi) \;= \frac{\Gamma}{\pi}\frac{\partial}{\partial\xi}\left(\frac{2}{\sqrt{2\tau\tau^*}}R_2(k)\right) \tag{5.32}$$

$$\mathcal{C}^{\tau\tau}(\tau,\tau^*,\xi^*-\xi) = -\frac{\Gamma}{4\pi}\frac{\partial^2}{\partial\tau^2}\left(2\sqrt{2\tau\tau^*}R_1(k)\right) \tag{5.33}$$

$$\mathcal{C}^{\xi\xi}(\tau,\tau^*,\xi^*-\xi) = -\frac{\Gamma}{4\pi}\frac{\partial^2}{\partial\xi^2}\left(2\sqrt{2\tau\tau^*}R_1(k)\right) \tag{5.34}$$

$$\mathcal{C}^{\tau\xi}(\tau,\tau^*,\xi^*-\xi) = -\frac{\Gamma}{4\pi}\frac{\partial^2}{\partial\tau\partial\xi}\left(2\sqrt{2\tau\tau^*}R_1(k)\right), \tag{5.35}$$

where $k = (\tau^{*2} + \tau^2 + (\xi^* - \xi)^2)/(2\tau\tau^*)$, and the complete elliptic integrals R_1 and R_2 are defined by

$$\begin{aligned} R_1(k) &\equiv \int_{-1}^{1} \frac{\mathrm{d}\beta}{\sqrt{1-\beta^2}} \sqrt{k-\beta} \\ &= -\frac{2(k^2-1)}{3}(R_D(2k, 2(k+1), k+1) + R_D(2k, 2(k-1), k-1)) + 4\sqrt{(}k) \\ R_2(k) &\equiv \int_{-1}^{1} \frac{\mathrm{d}\beta}{\sqrt{1-\beta^2}\sqrt{k-\beta}} \\ &= 2R_F(0, k+1, k-1). \end{aligned}$$

In the second equality for both R_1 and R_2, we used the table of elliptic integrals by Carlson [64] to express R_1 and R_2 in terms of the symmetric elliptic integrals R_D and R_F. This is particularly useful since these functions can be numerically evaluated using existing algorithms [53]. It can also be shown that $R_1'(k) = 1/2R_2(k)$ and $R_2'(k) = -R_1(k)/(2(k^2-1))$. Using these relations, we can explicitly evaluate the derivatives against τ and ξ in $\mathcal{C}^{(*)}$ (Eqs. 5.31 to 5.35), and rewrite these coefficients into forms which can be readily computed:

$$\mathcal{C}^{\tau} = -\frac{\Gamma}{2\pi\tau^{3/2}\sqrt{2\tau^*}} \left\{ \left(2\frac{\tau/\tau^* - k}{k^2-1} - \frac{\tau^*}{\tau} \right) R_1(k) + \left(1 + \frac{\tau^*}{\tau} k \right) R_2(k) \right\} \tag{5.36}$$

$$\mathcal{C}^{\xi} = \frac{\Gamma}{\pi\sqrt{2}(\tau\tau^*)^{3/2}} \frac{\xi^* - \xi}{k^2-1} R_1(k) \tag{5.37}$$

$$\mathcal{C}^{\tau\tau} = -\frac{\Gamma\sqrt{2\tau^*}}{4\pi\tau^{3/2}} \left(-\frac{1}{2} \left(1 + \frac{(\tau/\tau^* - k)^2}{k^2-1} \right) R_1(k) + kR_2(k) \right) \tag{5.38}$$

$$\mathcal{C}^{\xi\xi} = -\frac{\Gamma}{2\pi\sqrt{2\tau\tau^*}} \left(\frac{(\xi^* - \xi)^2}{2(k^2-1)\tau\tau^*} R_1(k) + R_2(k) \right) \tag{5.39}$$

$$\mathcal{C}^{\tau\xi} = -\frac{\Gamma(\xi^* - \xi)}{4\pi\tau^{3/2}\sqrt{2\tau^*}} \left(\frac{\tau/\tau^* - k}{k^2-1} R_1(k) + R_2(k) \right), \tag{5.40}$$

Equations 5.18 to 5.25 and 5.36 to 5.40 form the collisional Fokker–Planck equation in azimuthally–averaged cylindrical coordinates. The relatively cumbersome explicit expressions allow for direct discretisation of the derivatives and integrals, which we develop in the following section.

5.6 Discretisation scheme

Discretisation schemes for the advection and diffusion operators have previously been developed in Ch. 4 for the Vlasov equation. However, the magnitude of the advection and diffusion operators in the collisional Fokker–Planck equation (Eqs. 5.18 to 5.20) are comparable, while the Vlasov equation is dominated by the advection terms. The conservation of

momentum and energy is much more important for a numerical solution to the collisional Fokker–Planck equation, while the solution to the Vlasov equation has a greater need to preserve the phase space structures of the distribution while it is distorted by the flow field. The difference in the numerical requirements necessitates a new discretisation scheme for the collisional Fokker–Planck equation. Here, we modify the approach by Chacón et al. [65] to develop an energy–conserving scheme for the collisional Fokker–Planck equation.

Consider a rectangular grid spanning the τ and ξ phase space, indexed $i \in [0, N_\tau - 1]$ and $j \in [-N_\xi, N_\xi - 1]$ respectively. The grid is specified as follows, and also illustrated in Fig. 5.2:

1. First define $\Delta\tau_{i+1/2}$ and $\Delta\xi_{j+1/2}$ which are the distance in τ and ξ directions between cell centres. These spacings do not need to be uniform, but the list of lengths for ξ should be symmetric around $\xi = 0$, i.e. $\Delta\xi_{j+1/2} = \Delta\xi_{-j-3/2}$.

2. The cell centres are then defined as $\tau_{i+1} \equiv \tau_i + \Delta\tau_{i+1/2}$ and $\xi_{j+1} \equiv \xi_j + \Delta\xi_{j+1/2}$. The first radial cell centre is defined to be $\tau_0 \equiv 1/2\Delta\tau_{-1/2}$, and the first axial cell centre is similarly defined as $\xi_0 \equiv 1/2\Delta\xi_{-1/2}$.

3. The cell boundaries are defined as the midpoint between cell centres, i.e. $\tau_{i+1/2} \equiv (\tau_{i+1} + \tau_i)/2$ and $\xi_{j+1/2} \equiv (\xi_{j+1} + \xi_j)/2$. The first boundary in the axial and radial directions are $\tau_{-1/2} \equiv 0$ and $\xi_{-1/2} \equiv 0$ by definition.

4. The cell widths $\Delta\tau_i$ and $\Delta\xi_j$ are defined as the distance between the boundaries, i.e. $\Delta\tau_i \equiv \tau_{i+1/2} - \tau_{i-1/2}$ and $\Delta\xi_j \equiv \xi_{j+1/2} - \xi_{j-1/2}$.

5. The "alternative" cell centre is defined as the midpoint between cell boundaries, i.e. $\bar{\tau}_i \equiv (\tau_{i+1/2} + \tau_{i-1/2})/2$ and $\bar{\xi}_j \equiv (\xi_{j+1/2} + \xi_{j-1/2})/2$. Note that $\bar{\tau}_i = \tau_i$ and $\bar{\xi}_j = \xi_j$ if the cell widths are constant.

6. Noting that a cell in (τ, ξ) is actually a torus in velocity–space, the cell (i, j) has a volume of $\Delta\Omega_{ij} = 2\pi\bar{\tau}_i\Delta\tau_i\Delta\xi_j$. The area of the annulus separating cell (i, j) and $(i, j + 1)$ is $\Delta A_{i,j+1/2} = 2\pi\bar{\tau}_i\Delta\tau_i$, and the area of the cylindrical wall between cell $(i - 1, j)$ and (i, j) is $\Delta W_{i-1/2,j} = 2\pi\tau_{i-1/2}\Delta\xi_j$.

The discrete distribution takes on values of $f_{ij} \equiv f_1(\tau_i.\xi_j)$ at the centre of each cell. The distribution is also assumed to be symmetric in ξ, i.e. the same number of particles travelling at a positive axial velocity is the same as the number travelling at the opposite negative velocity. This means $f_{ij} = f_{i,-j-1}$, and the axial flux J^ξ should be zero at $\xi = 0$ by symmetry. Similarly J^τ is also zero at $\tau = 0$ since there can be no flux through a zero–area line.

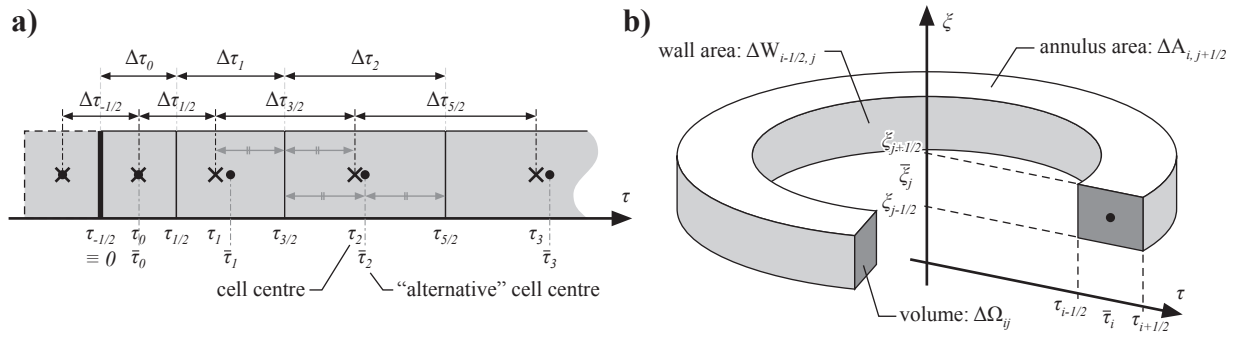

Figure 5.2: a) The definition of the radial τ grid spacing and position variables. The axial ξ grid is similarly defined, except that it extends to negative indices, and is symmetric around $\xi = 0$. b) The definition of the volume and area elements associated with cell (i, j).

Using this grid, we can discretise Eq. 5.18 by considering the total flux that passes into or out of a cell through the annulus or cylindrical walls. This means

$$f_{ij}(t+\Delta t) = f_{ij}(t) + \frac{\Delta t}{\Delta\Omega_{ij}}(J^{\tau}_{i-1/2,j}\Delta W_{i-1/2,j} - J^{\tau}_{i+1/2,j}\Delta W_{i+1/2,j} + J^{\xi}_{i,j-1/2}\Delta A_{i,j-1/2} - J^{\xi}_{i,j+1/2}\Delta A_{i,j+1/2}). \tag{5.41}$$

The first (second) term in the brackets refers to the radial flux flowing into (out of) cell (i, j) through the inner (outer) cylindrical wall, and the third (fourth) term refers to the axial flux flowing into (out of) cell (i, j) through the lower (upper) annulus. These fluxes come from the discretisation of Eqs. 5.19 and 5.20, which we choose to write as

$$J^{\tau}_{i+1/2,j} = (f\nu^{\tau})_{i+1/2,j} + \frac{1}{\tau_{i+1/2}\Delta\tau_{i+1/2}}\Big((\bar{\tau}fD^{\tau\tau})_{i+1,j} - (\bar{\tau}fD^{\tau\tau})_{i,j}\Big) + \frac{1}{\Delta\xi_j}\Big((fD^{\tau\xi})_{i+1/2,j+1/2} - (fD^{\tau\xi})_{i+1/2,j-1/2}\Big) \tag{5.42}$$

$$J^{\xi}_{i,j+1/2} = (f\nu^{\xi})_{i,j+1/2} + \frac{1}{\Delta\xi_{j+1/2}}\Big((fD^{\xi\xi})_{i,j+1} - (fD^{\xi\xi})_{i,j}\Big) + \frac{1}{\bar{\tau}_i\Delta\tau_i}\Big(\tau_{i+1/2}(fD^{\tau\xi})_{i+1/2,j+1/2} - \tau_{i-1/2}(fD^{\tau\xi})_{i-1/2,j+1/2}\Big), \tag{5.43}$$

where the non-cell centre points for f are given by the interpolations

$$(f\nu^{\tau})_{i+1/2,j} \equiv \frac{(f\nu^{\tau}\tau\Delta\tau)_{i,j} + (f\nu^{\tau}\tau\Delta\tau)_{i+1,j}}{2\tau_{i+1/2}\Delta\tau_{i+1/2}} \tag{5.44}$$

$$(f\nu^{\xi})_{i,j+1/2} \equiv \frac{(f\nu^{\xi}\xi\Delta\xi)_{i,j+1} + (f\nu^{\xi}\xi\Delta\xi)_{i,j}}{2\xi_{j+1/2}\Delta\xi_{j+1/2}} \tag{5.45}$$

$$(fD^{\tau\xi})_{i+1/2,j+1/2} \equiv \left(\sum_{i'=i}^{i+1}\sum_{j'=j}^{j+1}(fD^{\tau\xi}\Delta\Omega)_{i',j'}\right) \Big/ \left(\sum_{i'=i}^{i+1}\sum_{j'=j}^{j+1}\Delta\Omega_{i',j'}\right) \tag{5.46}$$

and the advection and diffusion coefficients are discretised as

$$\nu^{\tau}_{i,j} = \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \mathcal{C}^{\tau}(\tau_i, \tau_{i^*}, \xi_{j^*} - \xi_j) \tag{5.47}$$

$$\nu^{\xi}_{i,j} = \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \mathcal{C}^{\xi}(\tau_i, \tau_{i^*}, \xi_{j^*} - \xi_j) \tag{5.48}$$

$$D^{\tau\tau}_{i,j} = \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \mathcal{C}^{\tau\tau}(\tau_i, \tau_{i^*}, \xi_{j^*} - \xi_j) \tag{5.49}$$

$$D^{\xi\xi}_{i,j} = \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \mathcal{C}^{\xi\xi}(\tau_i, \tau_{i^*}, \xi_{j^*} - \xi_j) \tag{5.50}$$

$$D^{\tau\xi}_{i,j} = \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \mathcal{C}^{\tau\xi}(\tau_i, \tau_{i^*}, \xi_{j^*} - \xi_j). \tag{5.51}$$

This choice of discretisation and interpolation is chosen so that the total energy of the distribution, $E \equiv \sum \Delta\Omega_{ij} f_{ij} (\tau_i^2 + \xi_j^2)/2$, remains unchanged between time steps. To see that is the case, we use Eq. 5.41 to write the change in total energy in one time step as

$$\begin{aligned}
\frac{\Delta E}{\Delta t} &= \sum_{i,j} \frac{\tau_i^2}{2} \left(J^{\tau}_{i-1/2,j} \Delta W_{i-1/2,j} - J^{\tau}_{i+1/2,j} \Delta W_{i+1/2,j} \right) \\
&\quad + \sum_{i,j} \frac{\xi_j^2}{2} \left(J^{\xi}_{i,j-1/2} \Delta A_{i,j-1/2} - J^{\xi}_{i,j+1/2} \Delta A_{i,j+1/2} \right) \\
&= \sum_{i,j} \left(\frac{\tau_{i+1}^2 - \tau_i^2}{2} J^{\tau}_{i+1/2,j} \Delta W_{i+1/2,j} + \frac{\xi_{j+1}^2 - \xi_j^2}{2} J^{\xi}_{i,j+1/2} \Delta A_{i,j+1/2} \right) \\
&= 2\pi \sum_{i,j} \left(J^{\tau}_{i+1/2,j} \tau^2_{i+1/2} \Delta\tau_{i+1/2} \Delta\xi_j + J^{\xi}_{i,j+1/2} \bar{\tau}_i \xi_{j+1/2} \Delta\tau_i \Delta\xi_{j+1/2} \right).
\end{aligned} \tag{5.52}$$

In the first summation in the first equality, we eliminated the ξ^2 term since the flux J^{τ} cannot change the "flattened" distribution in ξ, and the axial energy is therefore unchanged. Similarly the τ^2 term in the second summation is dropped. In the second equality we shifted the indices, and used the fact that $J^{\tau}_{-1/2,j} = 0$ by symmetry. $J^{\tau}_{N_\tau - 1/2,j}$, $J^{\xi}_{i,-N_\xi - 1/2}$ and $J^{\xi}_{i,N_\xi - 1/2}$ are zero, assuming the distribution is entirely contained inside the simulation domain. In the third equality we used the grid definitions. Inserting Eqs. 5.42 and 5.43 into Eq. 5.52, we obtain

$$\frac{\Delta E}{2\pi\Delta t} = \epsilon^{\tau} + \epsilon^{\tau\tau} + \epsilon^{\tau\xi} + \epsilon^{\xi} + \epsilon^{\xi\xi} + \epsilon^{\xi\tau}$$

where the first three ϵ terms are contributed by J^{τ} and the last three by J^{ξ}. These are given explicitly by

$$\epsilon^{\tau} = \sum_{i,j} (f\nu^{\tau})_{i+1/2,j} \tau^2_{i+1/2} \Delta\tau_{i+1/2} \Delta\xi_j$$

$$
\begin{aligned}
&= \sum_j \Delta\xi_j \frac{\sum_i \tau_{i+1/2}(f\nu^\tau\tau\Delta\tau)_{i+1,j} + \sum_i \tau_{i+1/2}(f\nu^\tau\tau\Delta\tau)_{i,j}}{2} \\
&= \sum_{i,j} (f\nu^\tau\tau\bar{\tau}\Delta\tau\Delta\xi)_{i,j} \\
\epsilon^{\tau\tau} &= \sum_{i,j} \Big((\bar{\tau} f D^{\tau\tau})_{i+1,j} - (\bar{\tau} f D^{\tau\tau})_{i,j} \Big) \tau_{i+1/2} \Delta\xi_j \\
&= -\sum_{i,j} f_{i,j} D^{\tau\tau}_{i,j} (\tau_{i+1/2} - \tau_{i-1/2}) \bar{\tau}_i \Delta\xi_j \\
&= -\sum_{i,j} (f D^{\tau\tau} \bar{\tau} \Delta\tau\Delta\xi)_{i,j} \\
\epsilon^{\tau\xi} &= \sum_{i,j} \Big((f D^{\tau\xi})_{i+1/2,j+1/2} - (f D^{\tau\xi})_{i+1/2,j-1/2} \Big) \tau^2_{i+1/2} \Delta\tau_{i+1/2} \\
&= \sum_i \tau^2_{i+1/2} \Delta\tau_{i+1/2} \Big(\sum_j (f D^{\tau\xi})_{i+1/2,j+1/2} - \sum_j (f D^{\tau\xi})_{i+1/2,j-1/2} \Big) = 0 \\
\epsilon^{\xi} &= \sum_{i,j} (f\nu^\xi)_{i,j+1/2} \bar{\tau}_i \xi_{j+1/2} \Delta\tau_i \Delta\xi_{j+1/2} \\
&= \sum_i \bar{\tau}_i \Delta\tau_i \frac{\sum_j (f\nu^\xi \xi \Delta\xi)_{i,j+1} + \sum_j (f\nu^\xi \xi \Delta\xi)_{i,j}}{2} \\
&= \sum_{i,j} (f\nu^\xi \bar{\tau} \xi \Delta\tau\Delta\xi)_{i,j} \\
\epsilon^{\xi\xi} &= \sum_{i,j} \Big((f D^{\xi\xi})_{i,j+1} - (f D^{\xi\xi})_{i,j} \Big) \bar{\tau}_i \xi_{j+1/2} \Delta\tau_i \\
&= -\sum_{i,j} f_{i,j} D^{\xi\xi}_{i,j} (\xi_{j+1/2} - \xi_{j-1/2}) \bar{\tau}_i \Delta\tau_i \\
&= -\sum_{i,j} (f D^{\xi\xi} \bar{\tau} \Delta\tau\Delta\xi)_{i,j} \\
\epsilon^{\xi\tau} &= \sum_{i,j} \Big((\tau f D^{\tau\xi})_{i+1/2,j+1/2} - (\tau f D^{\tau\xi})_{i-1/2,j+1/2} \Big) \xi_{j+1/2} \Delta\xi_{j+1/2} \\
&= \sum_j \xi_{j+1/2} \Delta\xi_{j+1/2} \Big(\sum_i (\tau f D^{\tau\xi})_{i+1/2,j+1/2} - \sum_i (\tau f D^{\tau\xi})_{i-1/2,j+1/2} \Big) = 0.
\end{aligned}
$$

The total energy change per time step is then

$$
\begin{aligned}
\frac{\Delta E}{\Delta t} &= \sum_{i,j} (2\pi\bar{\tau}\Delta\tau\Delta\xi f)_{i,j} (\tau\nu^\tau - D^{\tau\tau} + \xi\nu^\xi - D^{\xi\xi})_{i,j} \\
&= \sum_{i,j} (f\Delta\Omega)_{i,j} (\tau\nu^\tau - D^{\tau\tau} + \xi\nu^\xi - D^{\xi\xi})_{i,j}
\end{aligned}
$$

$$= \sum_{i,j} \sum_{i^*,j^*} (f\Delta\Omega)_{i,j} (f\Delta\Omega)_{i^*,j^*} (\tau_i \mathcal{C}^{\tau}_{i,j,i^*,j^*} - \mathcal{C}^{\tau\tau}_{i,j,i^*,j^*} + \xi_j \mathcal{C}^{\xi}_{i,j,i^*,j^*} - \mathcal{C}^{\xi\xi}_{i,j,i^*,j^*}).$$

Substituting Eqs. 5.26 to 5.30 for the coefficients $\mathcal{C}^{(*)}$, this becomes

$$\begin{aligned}
\frac{\Delta E}{\Delta t} &= \frac{\Gamma}{4\pi^2} \iint \mathrm{d}\theta \, \mathrm{d}\theta^* \sum_{i,j} \sum_{i^*,j^*} (f\Delta\Omega)_{i,j} (f\Delta\Omega)_{i^*,j^*} \left(2\tau \frac{\partial}{\partial \tau} \frac{1}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|} \right. \\
&\quad \left. + 2\xi \frac{\partial}{\partial \xi} \frac{1}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|} + \frac{1}{2} \frac{\partial^2}{\partial \tau^2} |\boldsymbol{v}_2 - \boldsymbol{v}_1| + \frac{1}{2\tau} \frac{\partial}{\partial \tau} |\boldsymbol{v}_2 - \boldsymbol{v}_1| + \frac{1}{2} \frac{\partial^2}{\partial \xi^2} |\boldsymbol{v}_2 - \boldsymbol{v}_1| \right)^{\dagger} \\
&= \frac{\Gamma}{4\pi^2} \iint \mathrm{d}\theta \, \mathrm{d}\theta^* \sum_{i,j} \sum_{i^*,j^*} (f\Delta\Omega)_{i,j} (f\Delta\Omega)_{i^*,j^*} \times \\
&\quad \left(2\boldsymbol{v}_1 \cdot \frac{\partial}{\partial \boldsymbol{v}_1} \frac{1}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|} + \frac{1}{2} \frac{\partial}{\partial \boldsymbol{v}_1} \cdot \frac{\partial}{\partial \boldsymbol{v}_1} |\boldsymbol{v}_2 - \boldsymbol{v}_1| \right)^{\dagger}.
\end{aligned}$$

Here the † superscript indicates that the expression in the braces is evaluated at $\boldsymbol{v}_1 = (\tau_i \cos\theta, \tau_i \sin\theta, \xi_j)$ and $\boldsymbol{v}_2 = (\tau_{i^*} \cos\theta^*, \tau_{i^*} \sin\theta^*, \xi_{j^*})$. In the second equality we used the fact that the integral over θ means we can freely re-introduce the ∂_θ terms in the gradient and Laplacian operator — these terms integrate to zero due to azimuthal continuity. Using the identities $\partial_{\boldsymbol{v}_1}(1/|\boldsymbol{v}_2 - \boldsymbol{v}_1|) = (\boldsymbol{v}_2 - \boldsymbol{v}_1)/|\boldsymbol{v}_2 - \boldsymbol{v}_1|^3$ and $\partial_{\boldsymbol{v}_1} \cdot \partial_{\boldsymbol{v}_1} |\boldsymbol{v}_2 - \boldsymbol{v}_1| = 2/|\boldsymbol{v}_2 - \boldsymbol{v}_1|$, we thus have

$$\begin{aligned}
\frac{\Delta E}{\Delta t} &= \frac{\Gamma}{4\pi^2} \iint \mathrm{d}\theta \, \mathrm{d}\theta^* \sum_{i,j} \sum_{i^*,j^*} (f\Delta\Omega)_{i,j} (f\Delta\Omega)_{i^*,j^*} \left(2 \frac{\boldsymbol{v}_1 \cdot (\boldsymbol{v}_2 - \boldsymbol{v}_1)}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|^3} + \frac{1}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|} \right)^{\dagger} \\
&= \frac{\Gamma}{4\pi^2} \iint \mathrm{d}\theta^* \, \mathrm{d}\theta \sum_{i^*,j^*} \sum_{i,j} (f\Delta\Omega)_{i^*,j^*} (f\Delta\Omega)_{i,j} \left(2 \frac{\boldsymbol{v}_2 \cdot (\boldsymbol{v}_1 - \boldsymbol{v}_2)}{|\boldsymbol{v}_1 - \boldsymbol{v}_2|^3} + \frac{1}{|\boldsymbol{v}_1 - \boldsymbol{v}_2|} \right)^{\dagger} \\
&= \frac{\Gamma}{4\pi^2} \iint \mathrm{d}\theta \, \mathrm{d}\theta^* \sum_{i,j} \sum_{i^*,j^*} (f\Delta\Omega)_{i,j} (f\Delta\Omega)_{i^*,j^*} \frac{1}{2} \left(-2 \frac{|\boldsymbol{v}_2 - \boldsymbol{v}_1|^2}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|^3} + \frac{2}{|\boldsymbol{v}_2 - \boldsymbol{v}_1|} \right)^{\dagger} \\
&= 0.
\end{aligned}$$

In the second equality we have simply swapped the starred and the un-starred indices, and in the third we take an average of the first two lines, which turns out to be zero. This discretisation scheme therefore ensures that the total energy of the distribution is conserved by the collisional Fokker–Planck operator.

5.7 Computational implementation

Equations 5.41 to 5.51 form the numerical scheme of the simulation. Upon every time step, Eq. 5.41 is evaluated for each cell, and each evaluation involves the summation of the coefficients $\mathcal{C}^{(*)}$ over the whole distribution. This means each time step requires $O((N_\tau \times N_\xi)^2)$

evaluations of the $\mathcal{C}^{(*)}$, which makes the collisional Fokker–Planck equation computationally expensive.

The first technique we used to reduce the computational requirement is to store and evolve the $j \geq 0$ half of the distribution only, given that we assumed the distribution is symmetric around ξ. This requires a modification to Eqs. 5.47 to 5.51 to include the "mirrored half" of the distribution that is omitted:

$$\begin{aligned}
\nu^{\tau}_{i,j} &= \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \big(\mathcal{C}^{\tau}(\tau_i, \tau_{i^*}, |\xi_{j^*} - \xi_j|) + \mathcal{C}^{\tau}(\tau_i, \tau_{i^*}, \xi_{j^*} + \xi_j)\big) \\
\nu^{\xi}_{i,j} &= \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \big(\mathrm{sgn}(j^* - j)\mathcal{C}^{\xi}(\tau_i, \tau_{i^*}, |\xi_{j^*} - \xi_j|) - \mathcal{C}^{\xi}(\tau_i, \tau_{i^*}, \xi_{j^*} + \xi_j)\big) \\
D^{\tau\tau}_{i,j} &= \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \big(\mathcal{C}^{\tau\tau}(\tau_i, \tau_{i^*}, |\xi_{j^*} - \xi_j|) + \mathcal{C}^{\tau\tau}(\tau_i, \tau_{i^*}, \xi_{j^*} + \xi_j)\big) \\
D^{\xi\xi}_{i,j} &= \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \big(\mathcal{C}^{\xi\xi}(\tau_i, \tau_{i^*}, |\xi_{j^*} - \xi_j|) + \mathcal{C}^{\xi\xi}(\tau_i, \tau_{i^*}, \xi_{j^*} + \xi_j)\big) \\
D^{\tau\xi}_{i,j} &= \sum_{i^*,j^*} \Delta\Omega_{i^*,j^*} f_{i^*,j^*} \big(\mathrm{sgn}(j^* - j)\mathcal{C}^{\tau\xi}(\tau_i, \tau_{i^*}, |\xi_{j^*} - \xi_j|) - \mathcal{C}^{\tau\xi}(\tau_i, \tau_{i^*}, \xi_{j^*} + \xi_j)\big).
\end{aligned}$$

where the sign function $\mathrm{sgn}(x) = x/|x|$ if $x \neq 0$ or 0 otherwise. We have used the fact that $\mathcal{C}^{\tau}$, $\mathcal{C}^{\tau\tau}$ and $\mathcal{C}^{\xi\xi}$ are symmetric in their third argument, while $\mathcal{C}^{\xi}$ and $\mathcal{C}^{\tau\xi}$ are anti-symmetric. These expressions for the advection and diffusion coefficients limit the evaluation of $\mathcal{C}^{(*)}$ to only positive third argument values.

The second speed–up technique is to pre-compute the coefficients $\mathcal{C}^{(*)}$ and store them as three–dimensional tables in memory. The three dimensions span the three arguments of the $\mathcal{C}^{(*)}$. The particular values at which the first two arguments are evaluated simply follow the radial grid, i.e. $\tau \in \{\tau_i\}$ and $\tau^* \in \{\tau_i\}$. The third argument $\xi^* - \xi$ is more complicated as it is the axial distance between any two cells (including the mirrored half). Given that the grid size is not necessarily uniform, there are potentially $N_\xi(2N_\xi - 1)$ possible values for the third argument. This exceeds the memory capacity of most systems. We therefore construct the axial grid such that the spacing at small ξ are uniform, and only progressively enlarge the spacing for the few most outward cells at large ξ. This reduces the number of possible distances between cells significantly. All of these possible axial distances are then calculated, sorted, and used as the values for the third argument of $\mathcal{C}^{(*)}$ to build the tables.

The third speed–up technique is to use a Crank–Nicolson style implicit stepping [53] instead of the explicit one as described by Eqs. 5.41 to 5.43. These three equations can be written into the general form of

$$\frac{f_{ij}(t + \Delta t) - f_{ij}(t)}{\Delta t} = \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} M_{(ij),(kl)} f_{kl}(t),$$

where the value of cell (i, j) after a time step depends on the closest 3×3 cells on the step prior, and M is a linear combination of the ν and D coefficients. The Crank-Nicolson stepping simply replaces $f_{kl}(t)$ on the right by the time–symmetric $(f_{kl}(t) + f_{kl}(t + \Delta t))/2$. This leads to the implicit scheme

$$\left(\overleftrightarrow{\boldsymbol{I}} - \frac{\Delta t}{2}\overleftrightarrow{\boldsymbol{M}}\right)\boldsymbol{f}(t + \Delta t) = \left(\overleftrightarrow{\boldsymbol{I}} + \frac{\Delta t}{2}\overleftrightarrow{\boldsymbol{M}}\right)\boldsymbol{f}(t).$$

Here the 2–D grid spanned by (i, j) is flattened into a 1–D list indexed by $a \equiv N_\xi i + j$, and $\boldsymbol{M}$ is a band–diagonal matrix with band width $2N_\xi+3$. For each time step, the matrix on the right is inverted using existing algorithms like LU decomposition [53], and the vector $\boldsymbol{f}(t+\Delta t)$ is solved. This scheme has the advantage of allowing much larger time steps to be taken without introducing numerical instability, and therefore reducing the overall computational requirement. Note that this scheme is not completely implicit, as the coefficients ν and D involved in $\boldsymbol{M}$ are still evaluated using the explicit $\boldsymbol{f}(t)$. It is not possible to construct a complete implicit scheme as the collisional Fokker–Planck operator is not a linear differential equation.

5.8 Comparison with analytic model

As a benchmark of our simulation scheme, we compare it with an analytic solution. While the simulation model is applicable to all distributions, analytic solution of the collisional Fokker–Planck equation exists only for a few special cases. One of the analytic solutions, first obtained by Ichimaru and Rosenbluth [66], is applicable when the distribution is thermal (Gaussian) in both the parallel and perpendicular directions, but at different temperatures. The distribution relaxes through weakly magnetised collisions, and the temperatures in the two directions equilibrate in an exponential decay. The analytic solution for the temperatures is given by

$$\frac{\mathrm{d}T_\perp(t)}{\mathrm{d}t} = \gamma(T_z - T_\perp), \qquad T_z(t) + 2T_\perp(t) = T_z(0) + 2T_\perp(0),$$

where the second equation comes from energy conservation. The decay factor γ is given by

$$\gamma = \frac{8\sqrt{\pi}}{15}\frac{n_0 q^4}{(4\pi\epsilon_0)^2\sqrt{m}T_{\mathrm{eff}}^{3/2}}\log\left(\frac{\lambda_D}{\bar{b}}\right).$$

The effective temperature is defined as

$$\frac{1}{T_{\mathrm{eff}}^{3/2}} = \frac{15}{4}\int_{-1}^{1}\mathrm{d}a\frac{a^2(1-a^2)}{((1-a^2)T_\perp(0) + a^2T_z(0))^{3/2}},$$

which lies between the initial parallel and perpendicular temperatures $T_z(0)$ and $T_\perp(0)$. The value $\bar{b}$ is twice the distance of closest approach, given by

$$\bar{b} = 2b_{\min}, \qquad b_{\min} = \frac{q^2}{4\pi\epsilon_0 k_B T_{\mathrm{eff}}}.$$

Comparing this analytic solution with the numerical model, we use the initial distribution

$$f_{ij}(0) = \mathcal{N} \exp\left(-\frac{m\tau_i^2}{2k_B T_\perp(0)} - \frac{m\xi_j^2}{2k_B T_z(0)}\right)$$
$$\mathcal{N} = 1/\left(\sum_{i,j} \Delta\Omega_{ij} \exp\left(-\frac{m\tau_i^2}{2k_B T_\perp(0)} - \frac{m\xi_j^2}{2k_B T_z(0)}\right)\right)$$

and compute the coefficients $\mathcal{C}^{(*)}$ coefficients using

$$\Gamma \equiv \frac{n_0 q^4}{4\pi\epsilon_0^2 m^2} \log\left(\frac{\theta_{\max}}{\theta_{\min}}\right) = \frac{n_0 q^4}{4\pi\epsilon_0^2 m^2} \log\left(\frac{\lambda_D}{b_{\min}}\right).$$

The logarithmic term is simplified using Eq. 5.4, together with the fact that $\theta_{\min}$ and $\theta_{\max}$ are both much smaller than $\pi/2$. This means that $\theta_{\min}$ and $\theta_{\max}$ are approximately $2q^2/(4\pi\epsilon_0\lambda_D\mu\bar{v}^2)$ and $2q^2/(4\pi\epsilon_0 b_{\min}\mu\bar{v}^2)$ respectively.

Figure 5.3 compares the evolution of the parallel and perpendicular temperatures during the equilibration process in an antiproton plasma, predicted by both the analytic model and our numerical model, where $T_z(0) = 800$ K and $T_\perp(0) = 250$ K. The density of the antiproton plasma is at $8 \times 10^{12}\text{m}^{-3}$, which is typical in the ALPHA apparatus. A good agreement is observed.

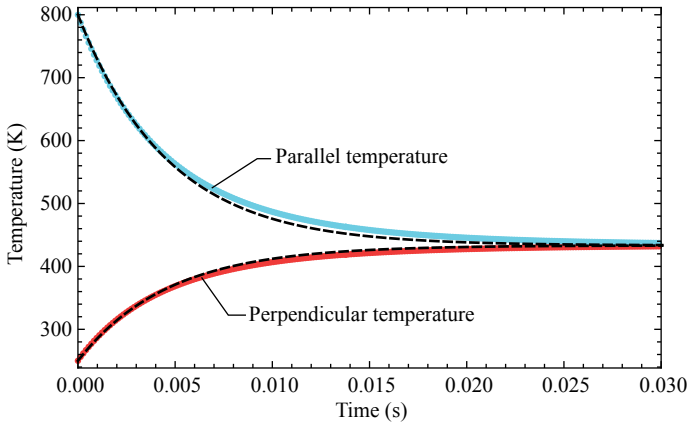

Figure 5.3: The equilibration of parallel and perpendicular temperatures of a plasma due to weakly magnetised collisions. The solid coloured lines show the result from the numerical model, and the dashed lines show the analytic solution.

Chapter 6

The intermediately magnetised collisional operator

As discussed in Ch. 5, collisions are considered weakly magnetised when the magnetic field is sufficiently weak that the curvature of the particles' trajectories are much bigger than their distance of closest approach. In this case the effect of the magnetic field is negligible. In the other extreme, the magnetic field can be so strong that that the particles can complete numerous gyro-cycles during the course of a collision. If this is the case, the magnetic moments of their gyromotion stay approximately conserved throughout the collision, and the motion of the gyrocentres can be obtained by averaging the forces acting on the particles over the cyclotron motion. This regime of strongly magnetised collision is treated by O'Neil [67]. Collisions involving positrons in the ALPHA apparatus are, however, neither weakly nor strongly magnetised, since the typical positron cyclotron radius is $\sim 2 \times 10^{-7}$ m, which is very close to the typical distance of closest approach $\sim 4 \times 10^{-7}$ m. In the length scale of a closest approach, a positron would have undergone ~ 0.3 cycles of gyromotion. This means for processes involving positron collisions (e.g. the collisional equilibration between antiprotons and positrons during antihydrogen production), a different numerical model is required to solve for their effects.

The intermediately magnetised regime of collisions is a difficult regime to study, as the outcome of collision events cannot be written down analytically in closed expressions. We have therefore chosen to evaluate the collision outcome numerically, and average their effects in a Fokker–Planck equation assuming weak collisions dominate the bulk collisional effects.

6.1 The phase–randomised collisional Fokker–Planck equation

In Ch. 5 we derived a collisional Fokker–Planck equation by exploiting the fact that the collective effect of the collisions is dominated by small–angle scattering, and in these cases

the change in velocity on each collision is small. The first difficulty in extending this model for intermediately magnetised collisions is the magnetic rotation of the perpendicular velocity through the course of the collision (see Fig. 6.1). The rotation is negligible in the weakly magnetised case, while in the strongly magnetised case the phase introduced to the perpendicular velocity is close to that of the zeroth order gyromotion. In the intermediately magnetised case, however, this phase is large but not easily predicable, even when considering small-angle scattering. It is therefore desirable to write down a equation in terms of the change in the *magnitude* of the perpendicular velocity, rather than in terms of the polar components of the change in the velocity vector.

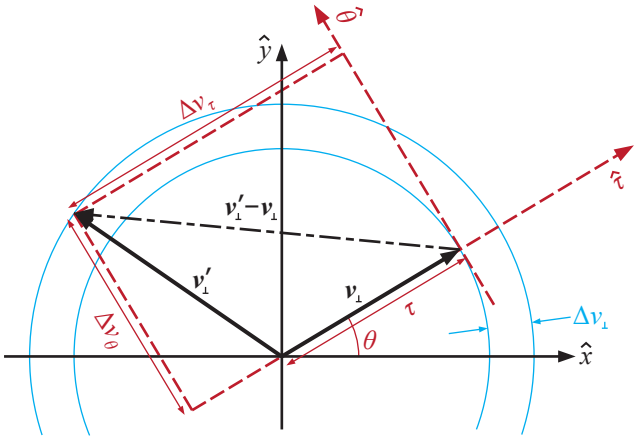

Figure 6.1: The coordinates and variables describing the change in perpendicular velocity of a particle in a collision. The perpendicular velocities before and after a collision are given by $\boldsymbol{v}_\perp$ and $\boldsymbol{v}_\perp'$ respectively. The change in perpendicular velocity, $\boldsymbol{v}_\perp' - \boldsymbol{v}_\perp$, can be written in polar coordinates in terms of Δv_τ and Δv_θ. However, due to the magnetic field–induced rotation of the velocity during the course of the collision, neither Δv_τ nor Δv_θ are small quantities. The true small quantity, $\Delta v_\perp$, is the different between the magnitude of the initial and final perpendicular velocities.

Firstly we define the change in the magnitude of $\boldsymbol{v}_\perp$ as

$$\Delta v_\perp \equiv \sqrt{(\tau + \Delta v_\tau)^2 + \Delta v_\theta^2} - \tau, \tag{6.1}$$

where τ is the magnitude of the initial perpendicular velocity, and Δv_τ and Δv_θ are the radial and azimuthal components of the (vector) change in perpendicular velocity (see Fig. 6.1). Expending $\Delta v_\perp$ in powers of Δv_τ and Δv_θ, it can be shown that

$$\frac{\Delta v_\perp}{v_\perp} = \sum_{i=1}^{\infty} \sum_{j=0}^{i} \sum_{k=0}^{j} \frac{(-1)^{i-1}(2i-2)!}{2^{i+j-1}(i-1)!(j-k)!k!(i-j)!} \left(\frac{\Delta v_\tau}{\tau}\right)^{i+j-2k} \left(\frac{\Delta v_\theta}{\tau}\right)^{2k}. \tag{6.2}$$

Restructuring the summations such that each distinct combination of powers for Δv_τ and Δv_θ appears only in one combination of the summation indices, and using the relations

$$\sum_{j=s}^{t-1} \frac{(-1)^j(4t-2j-4)!}{(2t-j-2)!(j-s)!(2t-2j-1)!} = \frac{2^{2t-2s-1}(-1)^s(2t-3)!}{(2t-2s-1)!(s-1)!} \qquad t \geq 2\ ,\ s \leq t \qquad (6.3)$$

$$\sum_{j=s}^{t} \frac{(-1)^{j+1}(4t-2j-2)!}{(2t-j-1)!(j-s)!(2t-2j)!} = \frac{2^{2t-2s}(-1)^{s-1}(2t-2)!}{(2t-2s)!(s-1)!} \qquad t \geq 1\ ,\ s \leq t, \qquad (6.4)$$

it can be shown that

$$\frac{\Delta v_\perp}{\tau} = \frac{\Delta v_\tau}{\tau} - \sum_{t=1}^{\infty}\sum_{s=1}^{t} \frac{(-1)^s}{2^{2s-1}s!(s-1)!}\Bigg[- \frac{(2t-1)!}{(2t-2s+1)!}\left(\frac{\Delta v_\tau}{\tau}\right)^{2t-2s+1}\left(\frac{\Delta v_\theta}{\tau}\right)^{2s} + \frac{(2t-2)!}{(2t-2s)!}\left(\frac{\Delta v_\tau}{\tau}\right)^{2t-2s}\left(\frac{\Delta v_\theta}{\tau}\right)^{2s}\Bigg]. \qquad (6.5)$$

The lone Δv_τ term on the right hand side of Eq. 6.5 originates from the $i = 1, j = 0, k = 0$ term of Eq. 6.2. This term has to be singled out since Eq. 6.3 is not valid for $t = 1$.

Next, defining the operators

$$\hat{\mathcal{O}}_\tau \equiv \frac{1}{\tau}\frac{\partial}{\partial\tau}\tau, \qquad \hat{\mathcal{O}}_\theta \equiv -\frac{1}{\tau}\frac{\partial}{\partial\tau} = -\hat{\mathcal{O}}_\tau\frac{1}{\tau}, \qquad \hat{\mathcal{O}}_\xi \equiv \frac{\partial}{\partial\xi},$$

we obtain several important relations which will become useful later. The most important is the permutation relation

$$\frac{1}{\tau^n}\hat{\mathcal{O}}_\tau = \hat{\mathcal{O}}_\tau\frac{1}{\tau^n} + \frac{n}{\tau^{n+1}}.$$

This permutation allows the operator $\hat{\mathcal{O}}_\tau$ to be gathered to the left hand side of expressions. Using this, it can be shown that

$$\hat{\mathcal{O}}_\theta^m\hat{\mathcal{O}}_\tau^n = (-1)^m \sum_{i=0}^{m+n-1} k_i^{m,n}\frac{n!}{i!}\hat{\mathcal{O}}_\tau^{i+1}\frac{1}{\tau^{2m+n-i-1}} \qquad (6.6)$$

where the coefficient $k_i^{m,n}$ is given by the iterative relation

$$k_i^{m+1,n} = \sum_{j=i-1}^{m+n-1}(j+1)k_j^{m,n} \qquad k_j^{1,n} = 1.$$

Equation 6.6 is another gathering operation which moves all the operators to the left hand side of expressions. Closer inspection of Eq. 6.6 reveals that the compound operator $\hat{\mathcal{O}}_\theta^m\hat{\mathcal{O}}_\tau^n$ gives rise to a series of derivatives against τ, with orders ranging from 1 to $m+n$. As argued

later, we are only interested in the $\hat{\mathcal{O}}_\tau$ and $\hat{\mathcal{O}}_\tau^2$ terms. Writing out these terms explicitly, we have from Eq. 6.6

$$\hat{\mathcal{O}}_\theta^m \hat{\mathcal{O}}_\tau^n = (-1)^m \frac{(2m+n-2)!}{2^{m-1}(m-1)!} \left(\hat{\mathcal{O}}_\tau \frac{1}{\tau^{2m+n-1}} + \hat{\mathcal{O}}_\tau^2 \frac{1}{\tau^{2m+n-2}} + \cdots \right), \tag{6.7}$$

for $m+n \geq 2$. For $m+n=1$, the right hand side of Eq. 6.7 is still definite, but does not give the correct expansion.

Now that we have the necessary relations, we can move onto the derivation of the collision model. In Ch. 5 we used the BBGKY hierarchy to derive the collisional Fokker–Planck equation; here we base our derivation on an equivalent but simpler method by Bellan [33]. Consider the collisional transfer function $F(\boldsymbol{v}, \Delta\boldsymbol{v})$, which gives the probability a particle with initial velocity $\boldsymbol{v}$ would come to possess the velocity $\boldsymbol{v}+\Delta\boldsymbol{v}$ due to collisions with other particles in a time of Δt. Since a particle must possess *some* velocity after Δt,

$$\int F(\boldsymbol{v}, \Delta\boldsymbol{v}) \, \mathrm{d}^3\Delta\boldsymbol{v} = 1. \tag{6.8}$$

Using the transfer function, the time evolution of a distribution can be written as

$$f(\boldsymbol{v}, t) = \int f(\boldsymbol{v}-\Delta\boldsymbol{v}, t-\Delta t) F(\boldsymbol{v}-\Delta\boldsymbol{v}, \Delta\boldsymbol{v}) \, \mathrm{d}^3\Delta\boldsymbol{v}. \tag{6.9}$$

Expanding the integrand in powers of $\Delta\boldsymbol{v}$ and Δt, we have

$$\begin{aligned}
&f(\boldsymbol{v}-\Delta\boldsymbol{v}, t-\Delta t) F(\boldsymbol{v}-\Delta\boldsymbol{v}, \Delta\boldsymbol{v}) \\
&= f(\boldsymbol{v}, t) F(\boldsymbol{v}, \Delta\boldsymbol{v}) + (-\Delta t) \frac{\partial}{\partial t} (f(\boldsymbol{v}, t) F(\boldsymbol{v}, \Delta\boldsymbol{v})) \\
&\quad + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{i!j!k!} (-\Delta v_x)^i (-\Delta v_y)^j (-\Delta v_z)^k \frac{\partial^i}{\partial v_x^i} \frac{\partial^j}{\partial v_y^j} \frac{\partial^k}{\partial v_z^k} (f(\boldsymbol{v}, t) F(\boldsymbol{v}, \Delta\boldsymbol{v})),
\end{aligned} \tag{6.10}$$

where we retain terms only up to the first order of Δt as Δt is arbitrarily small. In contrast, all orders of the velocity–space expansion are retained since the quantities Δv_x and Δv_y are not necessarily small in the intermediately magnetised regime. Putting Eq. 6.10 back into Eq. 6.9, and making use of Eq. 6.8, we have

$$\begin{aligned}
\frac{\partial f}{\partial t} &= \int \frac{\mathrm{d}^3\Delta\boldsymbol{v}}{\Delta t} \Bigg[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{i!j!k!} \\
&\qquad \frac{\partial^i}{\partial v_x^i} \frac{\partial^j}{\partial v_y^j} \frac{\partial^k}{\partial v_z^k} (-\Delta v_x)^i (-\Delta v_y)^j (-\Delta v_z)^k \Bigg] f(\boldsymbol{v}, t) F(\boldsymbol{v}, \Delta\boldsymbol{v}) \\
&\equiv \left\langle \frac{1}{\Delta t} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{i!j!k!} \frac{\partial^i}{\partial v_x^i} \frac{\partial^j}{\partial v_y^j} \frac{\partial^k}{\partial v_z^k} (-\Delta v_x)^i (-\Delta v_y)^j (-\Delta v_z)^k \right] f(\boldsymbol{v}, t) \right\rangle.
\end{aligned} \tag{6.11}$$

Here, the angled brackets denote the operation $\int \mathrm{d}^3\Delta\boldsymbol{v}F(\boldsymbol{v},\Delta\boldsymbol{v})$, keeping in mind that F is acted on by the derivatives.

Introducing the cylindrical coordinates $(v_x,v_y,v_z)=(\tau\cos\theta,\tau\sin\theta,\xi)$ as depicted in Fig. 6.1, we replace the rectilinear quantities in Eq. 6.11 using the relations

$$\frac{\partial}{\partial v_x}=\cos\theta\frac{\partial}{\partial\tau}-\frac{\sin\theta}{\tau}\frac{\partial}{\partial\theta},\qquad \frac{\partial}{\partial v_y}=\sin\theta\frac{\partial}{\partial\tau}+\frac{\cos\theta}{\tau}\frac{\partial}{\partial\theta},\qquad \frac{\partial}{\partial v_z}=\frac{\partial}{\partial\xi}$$
$$\Delta v_x=\cos\theta\Delta v_\tau-\sin\theta\Delta v_\theta,\quad \Delta v_y=\sin\theta\Delta v_\tau+\cos\theta\Delta v_\theta,\quad \Delta v_z=\Delta v_\xi.$$

Applying these in Eq. 6.11, and after some arithmetic, we arrive at

$$\begin{aligned}\frac{\partial f}{\partial t}=\Bigg\langle\!\!\Bigg\langle\frac{1}{\Delta t}\Bigg(&\hat{\mathcal{O}}_\tau(-\Delta v_\tau)f+\hat{\mathcal{O}}_\xi(-\Delta v_\xi)f\\
&+\frac{1}{2}\hat{\mathcal{O}}_\tau^2(-\Delta v_\tau)(-\Delta v_\tau)f+\frac{1}{2}\hat{\mathcal{O}}_\xi^2(-\Delta v_\xi)(-\Delta v_\xi)f+\hat{\mathcal{O}}_\tau\hat{\mathcal{O}}_\xi(-\Delta v_\tau)(-\Delta v_\xi)f\\
&+\sum_{t=1}^{\infty}\sum_{s=1}^{t}\Bigg[\frac{(2s-1)!!}{(2s)!(2t-2s)!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s}(-\Delta v_\tau)^{2t-2s}(-\Delta v_\theta)^{2s}\\
&\qquad+\frac{(2s-1)!!}{(2s)!(2t-2s+1)!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s+1}(-\Delta v_\tau)^{2t-2s+1}(-\Delta v_\theta)^{2s}\Bigg]f\\
&+\hat{\mathcal{O}}_\xi\sum_{t=1}^{\infty}\sum_{s=1}^{t}\Bigg[\frac{(2s-1)!!}{(2s)!(2t-2s)!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s}(-\Delta v_\tau)^{2t-2s}(-\Delta v_\theta)^{2s}\\
&\qquad+\frac{(2s-1)!!}{(2s)!(2t-2s+1)!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s+1}(-\Delta v_\tau)^{2t-2s+1}(-\Delta v_\theta)^{2s}\Bigg](-\Delta v_\xi)f\\
&+\cdots\Bigg)\Bigg\rangle\!\!\Bigg\rangle.\end{aligned}\tag{6.12}$$

A number of important assumptions and simplifications have been made here. Firstly, the double angle brackets indicate that both sides of Eq. 6.12 have been averaged by $1/2\pi\int_0^{2\pi}\mathrm{d}\theta$. Assuming the collisions are uncorrelated, the velocity phase angles introduced by collisions randomise the distribution f in the θ–direction. We have therefore assumed f to have no θ–dependence, in which case $\partial f/\partial t$ remains unchanged upon the phase angle averaging. For the right hand side, the phase angle averaging eliminates all terms involving ∂_θ due to the continuity condition in θ. Furthermore, we discarded terms which contain $\hat{\mathcal{O}}_\tau^n\hat{\mathcal{O}}_\xi^p$ where $n+p\geq 3$, as well as those which contain $\hat{\mathcal{O}}_\theta^m\hat{\mathcal{O}}_\tau^n\hat{\mathcal{O}}_\xi^p$ where $m+n\geq 1$ and $p\geq 2$, a choice which we will justify later. The two (identical) summations in Eq. 6.12 can be simplified as

$$\begin{aligned}\sum_{t=1}^{\infty}\sum_{s=1}^{t}\Bigg[&\frac{(2s-1)!!}{(2s)!(2t-2s)!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s}(\Delta v_\tau)^{2t-2s}(\Delta v_\theta)^{2s}\\
&-\frac{(2s-1)!!}{(2s)!(2t-2s+1)!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s+1}(\Delta v_\tau)^{2t-2s+1}(\Delta v_\theta)^{2s}\Bigg]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\hat{\mathcal{O}}_\theta(\Delta v_\theta)^2 - \frac{1}{2}\hat{\mathcal{O}}_\theta\hat{\mathcal{O}}_\tau(\Delta v_\tau)(\Delta v_\theta)^2 \\
&\quad + \sum_{t=2}^{\infty}\sum_{s=1}^{t}\left[\frac{1}{(2t-2s)!2^s s!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s}(\Delta v_\tau)^{2t-2s}(\Delta v_\theta)^{2s}\right. \\
&\qquad\qquad \left. - \frac{1}{(2t-2s+1)!2^s s!}\hat{\mathcal{O}}_\theta^s\hat{\mathcal{O}}_\tau^{2t-2s+1}(\Delta v_\tau)^{2t-2s+1}(\Delta v_\theta)^{2s}\right] \\
&= -\frac{1}{2}\hat{\mathcal{O}}_\tau\tau\left(\frac{\Delta v_\theta}{\tau}\right)^2 + \frac{1}{2}(\hat{\mathcal{O}}_\tau + \hat{\mathcal{O}}_\tau^2\tau)\tau\left(\frac{\Delta v_\tau}{\tau}\right)\left(\frac{\Delta v_\theta}{\tau}\right)^2 \\
&\quad + \sum_{t=2}^{\infty}\sum_{s=1}^{t}\frac{(-1)^s}{2^{2s-1}(s-1)!s!}\left[\frac{(2t-2)!}{(2t-2s)!}(\hat{\mathcal{O}}_\tau + \hat{\mathcal{O}}_\tau^2\tau)\tau\left(\frac{\Delta v_\tau}{\tau}\right)^{2t-2s}\left(\frac{\Delta v_\theta}{\tau}\right)^{2s}\right. \\
&\qquad\qquad \left. - \frac{(2t-1)!}{(2t-2s+1)!}(\hat{\mathcal{O}}_\tau + \hat{\mathcal{O}}_\tau^2\tau)\tau\left(\frac{\Delta v_\tau}{\tau}\right)^{2t-2s+1}\left(\frac{\Delta v_\theta}{\tau}\right)^{2s}\right] \\
&= \frac{1}{2}\hat{\mathcal{O}}_\tau^2\tau^2\left(\frac{\Delta v_\theta}{\tau}\right)^2 \\
&\quad + (\hat{\mathcal{O}}_\tau + \hat{\mathcal{O}}_\tau^2\tau)\tau\sum_{t=1}^{\infty}\sum_{s=1}^{t}\frac{(-1)^s}{2^{2s-1}(s-1)!s!}\left[\frac{(2t-2)!}{(2t-2s)!}\left(\frac{\Delta v_\tau}{\tau}\right)^{2t-2s}\left(\frac{\Delta v_\theta}{\tau}\right)^{2s}\right. \\
&\qquad\qquad \left. - \frac{(2t-1)!}{(2t-2s+1)!}\left(\frac{\Delta v_\tau}{\tau}\right)^{2t-2s+1}\left(\frac{\Delta v_\theta}{\tau}\right)^{2s}\right] \\
&= \frac{1}{2}\hat{\mathcal{O}}_\tau^2\tau^2\left(\frac{\Delta v_\theta}{\tau}\right)^2 + (\hat{\mathcal{O}}_\tau + \hat{\mathcal{O}}_\tau^2\tau)\tau\left(\frac{\Delta v_\tau}{\tau} - \frac{\Delta v_\perp}{\tau}\right) \\
&= \frac{1}{2}\hat{\mathcal{O}}_\tau^2\Delta v_\theta^2 + (\hat{\mathcal{O}}_\tau + \hat{\mathcal{O}}_\tau^2\tau)(\Delta v_\tau - \Delta v_\perp).
\end{aligned}
$$

In the first equality we have explicitly written out the $t = 1$ terms. In the second equality the compound operator $\hat{\mathcal{O}}_\theta^m\hat{\mathcal{O}}_\tau^n$ is expanded using Eq. 6.7, which is applicable since now $m + n \geq 2$. In applying Eq. 6.7 we choose to discard terms with $\hat{\mathcal{O}}_\tau^3$ and above, in line with our previous choice of retaining terms. In the third equality we re-introduce the explicitly written terms back into the summation as the $t = 1$ term, with the compensation of the $(\Delta v_\theta)^2$ term. In the fourth equality we use Eq. 6.5 to replace the summation. Putting this back into Eq. 6.12, and using the fact that $(\Delta v_\perp)^2 = (\Delta v_\tau)^2 + (\Delta v_\theta)^2 + 2\tau\Delta v_\tau - 2\tau\Delta v_\perp$, we have

$$
\begin{aligned}
\frac{\partial f}{\partial t} &= \left\langle\!\!\left\langle \frac{1}{\Delta t}\Big(- \hat{\mathcal{O}}_\tau\Delta v_\perp f - \hat{\mathcal{O}}_\xi\Delta v_\xi f \right.\right. \\
&\qquad \left.\left. + \frac{1}{2}\hat{\mathcal{O}}_\tau^2(\Delta v_\perp)^2 f + \frac{1}{2}\hat{\mathcal{O}}_\xi^2(\Delta v_\xi)^2 f + \hat{\mathcal{O}}_\tau\hat{\mathcal{O}}_\xi\Delta v_\perp\Delta v_\xi f + \cdots\Big)\right\rangle\!\!\right\rangle \\
&= -\hat{\mathcal{O}}_\tau\left\langle\!\!\left\langle\frac{\Delta v_\perp}{\Delta t}\right\rangle\!\!\right\rangle f - \hat{\mathcal{O}}_\xi\left\langle\!\!\left\langle\frac{\Delta v_\xi}{\Delta t}\right\rangle\!\!\right\rangle f
\end{aligned}
$$

$$- \hat{\mathcal{O}}_\tau^2 \left\langle\!\!\left\langle - \frac{1}{2}\frac{(\Delta v_\perp)^2}{\Delta t} \right\rangle\!\!\right\rangle f - \hat{\mathcal{O}}_\xi^2 \left\langle\!\!\left\langle - \frac{1}{2}\frac{(\Delta v_\xi)^2}{\Delta t} \right\rangle\!\!\right\rangle f$$
$$- 2\hat{\mathcal{O}}_\tau \hat{\mathcal{O}}_\xi \left\langle\!\!\left\langle - \frac{1}{2}\frac{\Delta v_\perp \Delta v_\xi}{\Delta t} \right\rangle\!\!\right\rangle f + \cdots . \tag{6.13}$$

In the second equality we used the facts that $\int \mathrm{d}^3\Delta\boldsymbol{v}$ and $1/2\pi \int_0^{2\pi} \mathrm{d}\theta$ commute with $\hat{\mathcal{O}}_\tau$ and $\hat{\mathcal{O}}_\xi$, and that the distribution f contains no dependence on $\Delta\boldsymbol{v}$ and θ, to move the double angle brackets. Equation 6.13 is formally the same as a Fokker–Planck equation in polar coordinates, but with the perpendicular evolution determined by the change in the *magnitude* of perpendicular velocity in collisions. Here our choice of discarding terms in Eqs. 6.7 and 6.12 is finally justified: these discarded terms can be gathered using a similar technique as demonstrated above, and be expressed in terms of $\hat{\mathcal{O}}_\tau^i \hat{\mathcal{O}}_\xi^j (\Delta v_\perp)^i (\Delta v_\xi)^j$ where $i + j \geq 3$. Under the assumption that the collective effects of collisions is dominated by small–angle scattering, $\Delta v_\perp$ and Δv_ξ are small quantities, and these third and higher order terms are discarded, as is customarily done in the derivation of Fokker–Planck equations.

Defining the advection coefficients

$$\nu^\tau(\tau,\xi) \equiv \left\langle\!\!\left\langle \frac{\Delta v_\perp}{\Delta t} \right\rangle\!\!\right\rangle = \frac{1}{\Delta t}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int \mathrm{d}^3\Delta\boldsymbol{v} F((\tau\cos\theta,\tau\sin\theta,\xi),\Delta\boldsymbol{v})\Delta v_\perp \tag{6.14}$$
$$\nu^\xi(\tau,\xi) \equiv \left\langle\!\!\left\langle \frac{\Delta v_\xi}{\Delta t} \right\rangle\!\!\right\rangle = \frac{1}{\Delta t}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int \mathrm{d}^3\Delta\boldsymbol{v} F((\tau\cos\theta,\tau\sin\theta,\xi),\Delta\boldsymbol{v})\Delta v_\xi \tag{6.15}$$

and the diffusion coefficients

$$\begin{aligned} D^{\tau\tau}(\tau,\xi) &\equiv -\frac{1}{2}\left\langle\!\!\left\langle \frac{(\Delta v_\perp)^2}{\Delta t} \right\rangle\!\!\right\rangle \\ &= -\frac{1}{2\Delta t}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int \mathrm{d}^3\Delta\boldsymbol{v} F((\tau\cos\theta,\tau\sin\theta,\xi),\Delta\boldsymbol{v})(\Delta v_\perp)^2 \end{aligned} \tag{6.16}$$
$$\begin{aligned} D^{\xi\xi}(\tau,\xi) &\equiv -\frac{1}{2}\left\langle\!\!\left\langle \frac{(\Delta v_\xi)^2}{\Delta t} \right\rangle\!\!\right\rangle \\ &= -\frac{1}{2\Delta t}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int \mathrm{d}^3\Delta\boldsymbol{v} F((\tau\cos\theta,\tau\sin\theta,\xi),\Delta\boldsymbol{v})(\Delta v_\xi)^2 \end{aligned} \tag{6.17}$$
$$\begin{aligned} D^{\tau\xi}(\tau,\xi) &\equiv -\frac{1}{2}\left\langle\!\!\left\langle \frac{\Delta v_\perp \Delta v_\xi}{\Delta t} \right\rangle\!\!\right\rangle \\ &= -\frac{1}{2\Delta t}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int \mathrm{d}^3\Delta\boldsymbol{v} F((\tau\cos\theta,\tau\sin\theta,\xi),\Delta\boldsymbol{v})\Delta v_\perp \Delta v_\xi, \end{aligned} \tag{6.18}$$

we can express Eq. 6.13 as

$$\frac{\partial f}{\partial t} = -\frac{1}{\tau}\frac{\partial}{\partial\tau}\tau J^\tau - \frac{\partial}{\partial\xi}J^\xi, \tag{6.19}$$

where the fluxes J^τ and J^ξ are given by

$$J^\tau = \nu^\tau f + \frac{1}{\tau}\frac{\partial}{\partial\tau}\tau(D^{\tau\tau}f) + \frac{\partial}{\partial\xi}(D^{\tau\xi}f), \tag{6.20}$$

$$J^\xi = \nu^\xi f + \frac{\partial}{\partial\xi}(D^{\xi\xi}f) + \frac{1}{\tau}\frac{\partial}{\partial\tau}\tau(D^{\tau\xi}f). \tag{6.21}$$

Comparing Eqs. 6.19 to 6.21 with the weakly magnetised collisional Fokker–Planck equation in Ch. 5 (Eqs. 5.18 to 5.20), we see that they share exactly the same numerical structure, and we can therefore reuse the discretisation scheme in Ch. 5 for the intermediately magnetised regime. The only major difference here is that the advection and diffusion coefficients in Ch. 5 are analytically calculated, while in the intermediately magnetised regime they are numerically evaluated, the method for which is described in the next section.

6.2 Evaluating the collision coefficients

From Eqs. 6.14 to 6.18, it is obvious that the advection and diffusion coefficients are various averages of the change of parallel and perpendicular velocities suffered by a particle travelling at an initial velocity of $(\tau\cos\theta, \tau\sin\theta, \xi)$. The average is over all possible changes in velocity, and weighted by its likelihood of occurrence. As depicted in Fig. 6.2, a unique collision between two particles is specified by eight parameters: the incoming particle's parallel and perpendicular velocities (ξ and τ) and its phase angle (θ), the target particle's parallel and perpendicular velocities (ξ^* and τ^*) and its phase angle (θ^*), and their relative impact position (b and ϕ). The change of parallel and perpendicular velocities, $\Delta v_\perp$ and Δv_ξ, are fully specified by these eight parameters:

$$\Delta v_\perp = \Delta v_\perp(\tau, \theta, \xi; \tau^*, \theta^*, \xi^*; b, \phi), \quad \Delta v_\xi = \Delta v_\xi(\tau, \theta, \xi; \tau^*, \theta^*, \xi^*; b, \phi).$$

To compute the probability of a certain change in velocity happening, it is therefore necessary to know the probability of the target being in the vicinity of some velocity $(\tau^*\cos\theta^*, \tau^*\sin\theta^*, \xi^*)$ first. This latter probability is given by $f^*(\tau^*, \xi^*)\tau^*\,\mathrm{d}\tau^*\,\mathrm{d}\theta^*\,\mathrm{d}\xi^*$, where f^* is the velocity distribution for the target particle (keeping in mind that it can be a different species). Here we assume the target species has a uniform θ^* distribution, just as the incoming species does. Using this, we can express the advection and diffusion coefficients as

$$\nu^\tau(\tau,\xi) = \int 2\pi\tau^*\,\mathrm{d}\tau^* \int \mathrm{d}\xi^* f^*(\tau^*,\xi^*)C^\tau(\tau,\xi;\tau^*,\xi^*) \tag{6.22}$$

$$\nu^\xi(\tau,\xi) = \int 2\pi\tau^*\,\mathrm{d}\tau^* \int \mathrm{d}\xi^* f^*(\tau^*,\xi^*)C^\xi(\tau,\xi;\tau^*,\xi^*) \tag{6.23}$$

$$D^{\tau\tau}(\tau,\xi) = \int 2\pi\tau^*\,\mathrm{d}\tau^* \int \mathrm{d}\xi^* f^*(\tau^*,\xi^*)C^{\tau\tau}(\tau,\xi;\tau^*,\xi^*) \tag{6.24}$$

$$D^{\xi\xi}(\tau,\xi) = \int 2\pi\tau^*\,\mathrm{d}\tau^* \int \mathrm{d}\xi^* f^*(\tau^*,\xi^*)C^{\xi\xi}(\tau,\xi;\tau^*,\xi^*) \tag{6.25}$$

$$D^{\tau\xi}(\tau,\xi) = \int 2\pi\tau^* \,\mathrm{d}\tau^* \int \mathrm{d}\xi^* f^*(\tau^*,\xi^*) C^{\tau\xi}(\tau,\xi;\tau^*,\xi^*) \tag{6.26}$$

where the $C^{(*)}$ coefficients are given by

$$C^{\tau} = \frac{1}{\Delta t}\int_0^{10\lambda_D}\int_0^{2\pi}\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\pi(10\lambda_D)^2}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta^*}{2\pi}\Delta v_\perp(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi) \tag{6.27}$$

$$C^{\xi} = \frac{1}{\Delta t}\int_0^{10\lambda_D}\int_0^{2\pi}\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\pi(10\lambda_D)^2}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta^*}{2\pi}\Delta v_\xi(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi) \tag{6.28}$$

$$C^{\tau\tau} = -\frac{1}{2\Delta t}\int_0^{10\lambda_D}\int_0^{2\pi}\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\pi(10\lambda_D)^2}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta^*}{2\pi}(\Delta v_\perp(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi))^2 \tag{6.29}$$

$$C^{\xi\xi} = -\frac{1}{2\Delta t}\int_0^{10\lambda_D}\int_0^{2\pi}\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\pi(10\lambda_D)^2}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta^*}{2\pi}(\Delta v_\perp(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi))^2 \tag{6.30}$$

$$C^{\tau\xi} = -\frac{1}{2\Delta t}\int_0^{10\lambda_D}\int_0^{2\pi}\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\pi(10\lambda_D)^2}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta^*}{2\pi}\Delta v_\perp(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi)\times \Delta v_\xi(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi). \tag{6.31}$$

Here the integral over the impact parameter b is bound above by $10\lambda_D$, since the electrostatic interaction between the particles is exponentially shielded on a length scale of λ_D by Debye shielding. At distances beyond a few multiples of λ_D (which we choose to be $10\lambda_D$) the interaction between the particles is negligible, and $\Delta v_\perp$ and Δv_ξ go to zero.

The expressions for $C^{(*)}$ contain the term $1/\Delta t$, which is necessary as the functions $\Delta v_\perp(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi)$ and $\Delta v_\xi(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi)$ give the change in velocity per *collision*, while the $C^{(*)}$ give the collisional effect per *time*. The quantity Δt is therefore the conversion factor between the two — the number of collisions per time. Using the geometry in Fig. 6.3, we have

$$1 = n^*\pi(10\lambda_D)^2|\xi - \xi^*|\Delta t,$$

where n^* is the number density of the target particles in real space. In the reverse situation where the incoming and the target particles' species are exchanged, we also define

$$1 = n\pi(10\lambda_D)^2|\xi - \xi^*|\Delta t^*.$$

The evaluation of Eqs. 6.27 to 6.31 cannot be done analytically, since there exists no closed analytic solution to the trajectory of intermediately magnetised collisions. The simplest way to evaluate Eqs. 6.27 to 6.31 numerically is to discretise the four–dimensional integral over b, ϕ, θ and θ^* into a four–dimensional sum, using the basic trapezoidal rule. This is numerically impractical though, since the number of evaluations of the integrand (i.e. the

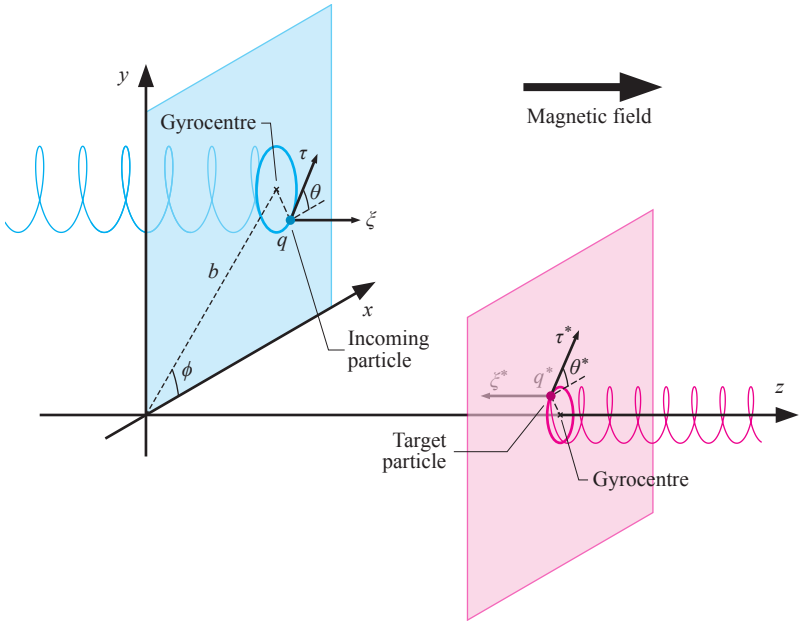

Figure 6.2: The geometry of the initial condition of two particles colliding in a magnetic field. The two particles have the initial velocities $(\tau\cos\theta, \tau\sin\theta, \xi)$ and $(\tau^*\cos\theta^*, \tau^*\sin\theta^*, \xi^*))$ respectively. The relative impact position is given by the impact parameter b and the angle ϕ, and the particles are initially positioned $20\lambda_D$ apart.

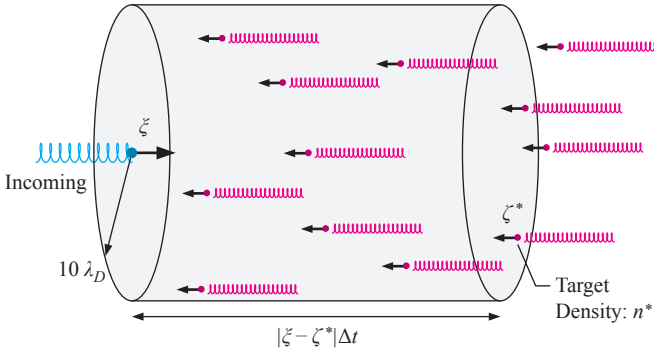

Figure 6.3: The geometry of magnetised collisions between two species of particles. The number of collisions within a time interval of Δt is given by the number of target particles contained within the shaded volume. Here $10\lambda_D$ is the maximum impact parameter within which two particles are considered to have interacted.

number of simulated collisions) required is of $O(N^4)$, where N is the number of divisions per dimension.

A well–known method of efficiently evaluating higher–dimensional integrals is the Monte Carlo method. Instead of regularly dividing and sampling the four dimensions, the integrand is evaluated at points drawn randomly from within the integration domain, and the average of these evaluations is multiplied by the volume of the integration domain to give the value of the integral. This method is generally suited to higher dimensional integrals, but it is still slow to converge in our case. This is mainly due to the collisions' strong and sharp dependence on the relative impact position (b, ϕ) and phase angles θ and θ^*. A small offset in any of them can change the helical gyro-orbit of the two particles, and heavily alter the relative distance of closest approach between them. In addition, the upper bound of the impact parameter b is chosen to be $10\lambda_D$, a large distance compared to the fall–off of the interaction strength, which means most evaluations of the integrand are essentially "inert". It is therefore inefficient to sample the integration domain at uniform density.

To further improve the efficiency of computation, we first recognise that rotational symmetry means ϕ is a cyclic variable when the impacting particles are evenly distributed across all possible phase angles (see Fig. 6.2). Instead of sampling it, we can set $\phi = 0$ for simplicity. For the remaining three variables b, θ and θ^*, we have chosen to apply an adaptive Monte Carlo method, which adjusts the sampling density of the integration space, such that more points are concentrated at regions where they would minimise the error of the integration. The prerequisite of such a scheme is a method of estimating statistical error. We have chosen the bootstrapping method for this purpose, due to its ease of implementation, and the fact that it makes no assumption about the shape of the underlying distribution. In our bootstrapping scheme, consider a set of samples a_i and their associated weights w_i, where $i \in [0, N-1]$. The weighted average of the set is given by

$$\langle a \rangle = \left(\sum_{i=0}^{N-1} a_i w_i \right) \Big/ \left(\sum_{i=0}^{N-1} w_i \right).$$

To estimate the statistical error of this average, we first generate a new set of samples a'_i and weights w'_i, where $i \in [0, N-1]$. This new set is generated from the old set via sampling with replacement, i.e. $a'_i = a_{g_i}$ and $w'_i = w_{g_i}$ where g_i is a uniform random number between 0 and $N-1$. The average corresponding to the new set can be computed as

$$\langle a \rangle' = \left(\sum_{i=0}^{N-1} a'_i w'_i \right) \Big/ \left(\sum_{i=0}^{N-1} w'_i \right).$$

This re-sampling is then repeated a large number of times (which we choose to be 1,000), yielding 1,000 distinct values for $\langle a \rangle'$. As the re-sampling imitates the different possibilities the a_i would have turned out had the process of generating them been redone, the distribution of the values for $\langle a \rangle'$ represents an estimate of the statistical spread of the average $\langle a \rangle$. The

1,000 values for $\langle a \rangle'$ are then sorted, and the 49th and 950th entries in the sorted list are identified as the lower and upper bounds for the average at the 90% confidence level. It should be noted that these bounds are independent of the number of samples N, as long as N is large enough that the distribution of a is well sampled.

During the computation of the collision coefficients, note should be taken concerning our assumption that $\Delta v_\perp$ and Δv_ξ are small values. This assumption is required in order to express the collisional effect as a Fokker–Planck–type operator, which means large–angle scatterings needs to be excluded from the averaging process. This exclusion is also present in Ch. 5 for weakly magnetised collisions in the form of the upper limit imposed on the scattering angle. In the case of intermediately magnetised collisions, these large–angle scatterings come from "reflecting" collisions, where the colliding particles reflect off each other in the axial direction instead of passing by. We exclude these cases from the averaging by rejecting collisions where $\Delta v_\xi/(\Xi - \xi) > 1$, where $\Xi \equiv (m\xi + m^*\xi^*)/(m + m^*)$ is the centre–of–mass axial velocity of the particles. This is equivalent to rejecting collisions where the change of axial velocity of one particle is over half of that in a perfectly reflecting collision. For a treatment of collisions that includes the large–angle scattering effect, see the work of Glinsky et al. [68].

We can now describe the adaptive Monte Carlo method used to compute C^τ, C^ξ, $C^{\tau\tau}$, $C^{\xi\xi}$ and $C^{\tau\xi}$. Here, the values for τ, ξ, τ^* and ξ^* are given as the arguments for these $C^{(*)}$ coefficients, while the values for b, θ and θ^* are sampled by the Monte Carlo method. The value of ϕ is fixed at zero due to symmetry.

1. Initialise the refinable 3–D parameter space by dividing it in a rectangular pattern, forming an array of boxes as shown in Fig. 6.4. The boxes are indexed by m, and each is associated with a weight

$$V_m \equiv (b_{m,\max} - b_{m,\min})(\theta_{m,\max} - \theta_{m,\min})(\theta^*_{m,\max} - \theta^*_{m,\min}).$$

2. For each box m, prepare a set of initial conditions

$$s_{m,n} = \{\tau, \mathrm{dran}[\theta_{m,\min}, \theta_{m,\max}], \xi, \tau^*, \mathrm{dran}[\theta^*_{m,\min}, \theta^*_{m,\max}], \theta^*, \mathrm{dran}[b_{m,\min}, b_{m,\max}], 0\},$$

   where $n \in [0, N_m - 1]$, and the dran function returns a random real number within the range specified at a uniform probability. The first three numbers in each entry specify the velocity of the incoming particle, the fourth to sixth numbers specify the velocity of the target particle, and the last two numbers specify the relative impact position. From each of these initial conditions, a particle pusher is used to simulate the collision process, which yields the change in perpendicular and parallel velocities of both the

incoming and target particles after the collision. Using these, we can compute

$$c_{m,n} = \Big\{ \Delta v_\perp, \Delta v_\xi, -\frac{1}{2}(\Delta v_\perp)^2, -\frac{1}{2}(\Delta v_\xi)^2, -\frac{1}{2}\Delta v_\perp \Delta v_\xi, \\ \Delta v_\perp^*, \Delta v_\xi^*, -\frac{1}{2}(\Delta v_\perp^*)^2, -\frac{1}{2}(\Delta v_\xi^*)^2, -\frac{1}{2}\Delta v_\perp^* \Delta v_\xi^* \Big\}.$$

3. Compute the averages within each box

$$\langle c \rangle_m [\![r]\!] = \sum_{\substack{n=0 \\ \text{passing}}}^{N_m - 1} \frac{c_{m,n}[\![r]\!] s_{m,n}[\![6]\!]}{N_m},$$

where $r \in [0, 9]$. The "passing" condition indicates that only terms where $c_{m,n}[\![1]\!]/(\Xi - \xi) \leq 1$ are included in the summations. The $s_{m,n}[\![6]\!]$ factor is included to account for the fact that the impact parameter b is a radial variable in polar coordinates. Using the bootstrapping method described above, the lower and upper bounds at 90% confidence level, $\langle c \rangle_m^{5\%}[\![r]\!]$ and $\langle c \rangle_m^{95\%}[\![r]\!]$ are also calculated. The average impact parameter

$$\langle b \rangle_m = \sum_{\substack{n=0 \\ \text{passing}}}^{N_m - 1} \frac{s_{m,n}[\![6]\!]}{N_m}$$

is also calculated.

4. Compute the solutions

$$\langle\!\langle c \rangle\!\rangle [\![r]\!] = \Big(\sum_m V_m \langle c \rangle_m [\![r]\!] \Big) \Big/ \Big(\sum_m V_m \langle b \rangle_m \Big).$$

5. Calculate, for each box, the error

$$\mathrm{err}_m [\![r]\!] = \frac{V_m \Big(\langle c \rangle_m^{95\%}[\![r]\!] - \langle c \rangle_m^{5\%}[\![r]\!] \Big) \Big/ \Big(\sum_m V_m \langle b \rangle_m \Big)}{|\langle\!\langle c \rangle\!\rangle [\![r]\!]|}.$$

This gives the percentage error contributed by each box to the ten solutions.

6. Scan through $\mathrm{err}_m[\![r]\!]$ for all m and r to locate several of the biggest entries. The number of biggest entries sough depends on the parallelisation scheme — see Sec. 6.4. If all entries originate from the same r, this r is recorded and skipped over on the next iteration while scanning for the biggest $\mathrm{err}_m[\![r]\!]$, to ensure all coefficients are examined across the iterations. The indices of the boxes from which these biggest entries originate are recorded.

7. A fixed number of extra collisions are launched from each of these boxes. The number of collisions launched per box depends on the parallisation scheme (Sec. 6.4). The outcome of the collisions are appended to the lists $s_{m,n}$, $c_{m,n}$ and $w_{m,n}$, and N_m is correspondingly incremented.

8. The $\langle c \rangle_m [\![r]\!]$, $\langle c \rangle_m^{5\%} [\![r]\!]$ and $\langle c \rangle_m^{95\%} [\![r]\!]$ for these boxes must be recomputed after the addition of extra collisions.

9. Scan through all boxes m, and identify those where N_m exceeds a chosen threshold. These boxes are split into eight by dividing them into two equal halves in each of the three dimensions, as shown in the red–highlighted box in Fig. 6.4 a. The events in the old, undivided box are sorted into the eight new boxes, according to the values in $s_{m,n}$. Note that the boundaries, the weight V_m and the number of events N_m for the newly created boxes have to be updated. The $\langle c \rangle_m [\![r]\!]$, $\langle c \rangle_m^{5\%} [\![r]\!]$ and $\langle c \rangle_m^{95\%} [\![r]\!]$ for these newly created boxes are also recomputed.

10. Go to step 4, until a desired accuracy is reached.

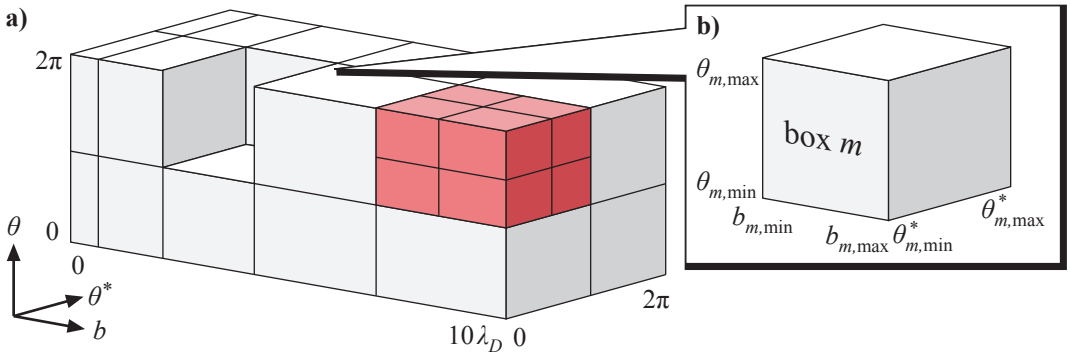

Figure 6.4: A schematic view of the subdivision structure of the adaptively sampled 3–D parameter space. The space is initially divided into a rectangular pattern, and boxes which contribute the most to the error in the final answer is subdivided, an example of which is shown in the red box. The extent of box m is labelled in b).

Qualitatively, this scheme launches collisions from the regions in the b–θ–θ^* parameter space where their bootstrapping error contribution to the answers is the greatest. As the number of collisions from these regions increase, they are subdivided to allow for more fine–grained resolution, such that the peak error region can be targeted more accurate by the sampling. One execution of this numerical scheme yields the solutions $\langle\!\langle c \rangle\!\rangle [\![r]\!]$, where $r \in [0, 9]$. The first five of these solutions ($r = 0$ to 4), when divided by Δt, correspond to the coefficients $C^{\tau}(\tau, \xi; \tau^*, \xi^*)$, $C^{\xi}(\tau, \xi; \tau^*, \xi^*)$, $C^{\tau\tau}(\tau, \xi; \tau^*, \xi^*)$, $C^{\xi\xi}(\tau, \xi; \tau^*, \xi^*)$ and $C^{\tau\xi}(\tau, \xi; \tau^*, \xi^*)$. The last five of the solutions ($r = 5$ to 9), divided by Δt^*, correspond to $C^{\tau *}(\tau, \xi; \tau^*, \xi^*)$,

$C^{\xi*}(\tau,\xi;\tau^*,\xi^*)$, $C^{\tau\tau*}(\tau,\xi;\tau^*,\xi^*)$, $C^{\xi\xi*}(\tau,\xi;\tau^*,\xi^*)$ and $C^{\tau\xi*}(\tau,\xi;\tau^*,\xi^*)$, which are the coefficients for the collisional Fokker–Planck equation describing the evolution of the "target" species due to collisions with the "incoming" species.

6.3 Boris particle pusher

The task of a particle pusher in the context of computing Eqs. 6.26 to 6.31 is to simulate the collision of two particles with the given parameters $(\tau,\theta,\xi;\tau^*,\theta^*,\xi^*;b,\phi)$, and obtain the resultant change in perpendicular and parallel velocities for both particles after the collision is complete. We consider two particles with charge and mass (q,m) and (q^*,m^*) being launched into each other in a uniform magnetic field $\boldsymbol{B}=B\hat{\boldsymbol{z}}$. The two particles' positions and velocities are labelled $(\boldsymbol{r},\boldsymbol{v})$ and $(\boldsymbol{r}^*,\boldsymbol{v}^*)$, and the simulation is executed in a frame which moves along z at a velocity of $\Xi=(m\xi+m^*\xi^*)/(m+m^*)$. This does not affect the outcome of the simulation, as the magnetic field lies in the z–direction. The initial z positions of the particles are chosen such that they would meet at $z=0$ (barring interaction). On the x-y plane the gyrocentre of the starred particle is fixed at the origin, while the gyrocentre of the un-starred is at $(b\cos\phi, b\sin\phi)$ (see Fig. 6.2). The *actual* x-y positions of the particles includes the radius and phase of the gyro-rotation. Quantitatively, the initial condition of the two particles are given in Cartesian coordinates by

$$
\begin{aligned}
\boldsymbol{r}_i &= \left(b\cos\phi-\frac{\omega_C}{\tau}\sin\theta, b\sin\phi+\frac{\omega_C}{\tau}\cos\theta, \mathrm{sgn}(\xi^*-\xi)20\lambda_D\frac{\Xi-\xi}{\xi^*-\xi}\right)\\
\boldsymbol{r}_i^* &= \left(-\frac{\omega_C^*}{\tau^*}\sin\theta^*, \frac{\omega_C^*}{\tau^*}\cos\theta^*, -\mathrm{sgn}(\xi^*-\xi)20\lambda_D\frac{\xi^*-\Xi}{\xi^*-\xi}\right)\\
\boldsymbol{v}_i &= (\tau\cos\theta,\tau\sin\theta,\xi-\Xi)\\
\boldsymbol{v}_i^* &= (\tau^*\cos\theta^*,\tau^*\sin\theta^*,\xi^*-\Xi),
\end{aligned}
$$

where the cyclotron radii $\omega_C=qB/m$ and $\omega_C^*=q^*B/m^*$.

To evolve the initial condition forward in time, the Boris particle pusher scheme [69] is used, which is a leapfrog algorithm where the position and velocity of the particles are offest from each other by half a time step, as shown in Fig. 6.5 a. The forwarding of the position from $n-1/2$ to $n+1/2$ makes use of the velocity at n, while the forwarding of the velocity from n to $n+1$ makes use of the electric and magnetic fields evaluated at the particle's position at $n+1/2$. To keep the half–step staggering consistent, Δt has to be constant in this scheme. A slight modification, shown in Fig. 6.5 b, splits the stepping of the position and allows both the particle position and velocity to be updated synchronously. This makes adjusting Δt more convenient. The synchronous scheme starts with an update of the positions from n to $n+1/2$:

$$
\boldsymbol{r}_{n+1/2}=\boldsymbol{r}_n+\frac{\Delta t}{2}\boldsymbol{v}_n,\qquad \boldsymbol{r}^*_{n+1/2}=\boldsymbol{r}^*_n+\frac{\Delta t}{2}\boldsymbol{v}^*_n.
$$

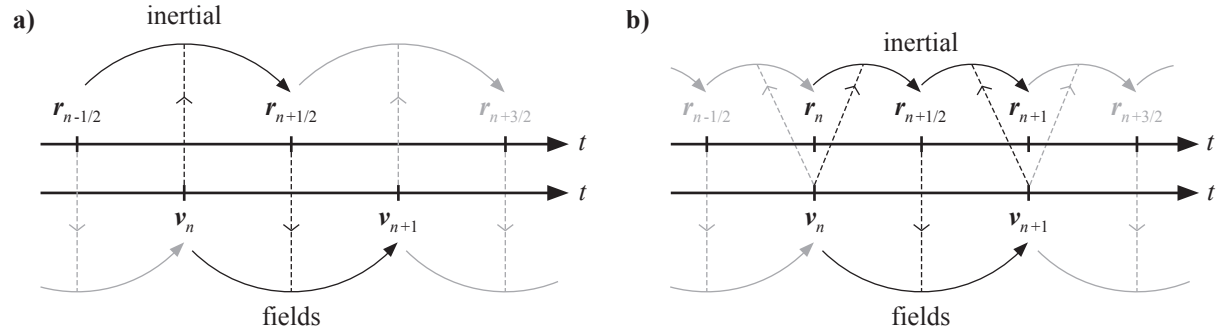

Figure 6.5: A schematic view of the Boris particle pusher scheme, showing the a) time–staggered leapfrog scheme suitable for uniform time–stepping, and the b) time–synchronous scheme suitable for non-uniform time–stepping. Only the position and velocity of one particle is shown.

Next we define

$$\boldsymbol{u} = \boldsymbol{v}_n + \frac{\Delta t}{2}\frac{\boldsymbol{F}}{m}, \qquad \boldsymbol{u}^* = \boldsymbol{v}_n^* - \frac{\Delta t}{2}\frac{\boldsymbol{F}}{m^*}.$$

Using them, the velocities are updated from n to $n+1$:

$$\boldsymbol{v}_{n+1} = \boldsymbol{u} + \frac{\omega_C \Delta t}{1+(\omega_C \Delta t/2)^2}\left(\boldsymbol{u}\times\hat{\boldsymbol{z}} + \frac{\omega_C \Delta t}{2}(\boldsymbol{u}\times\hat{\boldsymbol{z}})\times\hat{\boldsymbol{z}}\right) + \frac{\Delta t}{2}\frac{\boldsymbol{F}}{m}$$
$$\boldsymbol{v}_{n+1}^* = \boldsymbol{u}^* + \frac{\omega_C^* \Delta t}{1+(\omega_C^* \Delta t/2)^2}\left(\boldsymbol{u}^*\times\hat{\boldsymbol{z}} + \frac{\omega_C^* \Delta t}{2}(\boldsymbol{u}^*\times\hat{\boldsymbol{z}})\times\hat{\boldsymbol{z}}\right) - \frac{\Delta t}{2}\frac{\boldsymbol{F}}{m^*}.$$

Here $\boldsymbol{F}$ is the electrostatic force from the starred particle to the un-starred particle, evaluated using the positions at step $n+1/2$:

$$\boldsymbol{F} = \frac{qq^*}{4\pi\epsilon_0}\frac{e^{-\mathcal{R}/\lambda_D}}{\mathcal{R}^3}\left(1+\frac{\mathcal{R}}{\lambda_D}\right)(\boldsymbol{r}_{n+1/2} - \boldsymbol{r}_{n+1/2}^*), \qquad \mathcal{R} = |\boldsymbol{r}_{n+1/2} - \boldsymbol{r}_{n+1/2}^*|.$$

The exponential and $(1+\mathcal{R}/\lambda_D)$ factors originate from the Debye shielding of the electrostatic potential. Finally the position is updated again from $n+1/2$ to $n+1$:

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_{n+1/2} + \frac{\Delta t}{2}\boldsymbol{v}_{n+1}, \qquad \boldsymbol{r}_{n+1}^* = \boldsymbol{r}_{n+1/2}^* + \frac{\Delta t}{2}\boldsymbol{v}_{n+1}^*.$$

During the simulation, the two particles spend most of the time undergoing simple gyromotion without much interaction, as they are launched from well beyond the interaction cut–off λ_D. As the Boris particle pusher is a sympletic scheme which preserve magnetic moment where the electric field negligible, relatively large time steps can be taken in the outer region to conserve computational effort. However, within the region of interaction the time steps need to be resolved much more finely, especially when the particles come sufficiently close to each other. We choose to monitor the total energy

$$E_n = \frac{1}{2}m\boldsymbol{v}_n^2 + \frac{1}{2}m^*\boldsymbol{v}_n^{*2} + \frac{qq^*}{4\pi\epsilon_0}\frac{e^{-|\boldsymbol{r}_n^* - \boldsymbol{r}_n|/\lambda_D}}{|\boldsymbol{r}_n^* - \boldsymbol{r}_n|}$$

and look for its non-conservation between time steps. If the error $|E_n - E_{n-1}|$ exceeds a chosen threshold, the step size Δt is decreased, and if the error is below a (usually smaller) threshold the step size is increased. We impose an upper limit of $\min(0.11 \times 2\pi/\omega_C, 0.11 \times 2\pi/\omega_C^*, 0.0005 \times \lambda_D/|\xi^* - \xi|)$ on the step size, such the the cyclotron motion is adequately simulated at the maximum step size, and the time stepping do not become so broad that the closest approach between the two particles is entirely bypassed between steps.

The collision is considered "complete" when the relative z distance between the two particles has decreased as they are launched towards each other, reached a minimum during collision, and increased to above $20\lambda_D$ again, as the two passes by each other. At this point the time–stepping is terminated, and the change of the velocities are returned:

$$\Delta v_\perp = \sqrt{(\boldsymbol{v}_f \cdot \hat{\boldsymbol{x}})^2 + (\boldsymbol{v}_f \cdot \hat{\boldsymbol{y}})^2} - \sqrt{(\boldsymbol{v}_i \cdot \hat{\boldsymbol{x}})^2 + (\boldsymbol{v}_i \cdot \hat{\boldsymbol{y}})^2}, \quad \Delta v_\xi = \boldsymbol{v}_f \cdot \hat{\boldsymbol{z}} - \boldsymbol{v}_i \cdot \hat{\boldsymbol{z}}$$
$$\Delta v_\perp^* = \sqrt{(\boldsymbol{v}_f^* \cdot \hat{\boldsymbol{x}})^2 + (\boldsymbol{v}_f^* \cdot \hat{\boldsymbol{y}})^2} - \sqrt{(\boldsymbol{v}_i^* \cdot \hat{\boldsymbol{x}})^2 + (\boldsymbol{v}_i^* \cdot \hat{\boldsymbol{y}})^2}, \quad \Delta v_\xi^* = \boldsymbol{v}_f^* \cdot \hat{\boldsymbol{z}} - \boldsymbol{v}_i^* \cdot \hat{\boldsymbol{z}}$$

where $\boldsymbol{v}_f$ and $\boldsymbol{v}_f^*$ here are the final velocities of the particles.

6.4 Parallelisation on Graphical Processing Units

Using the adaptive Monte Carlo sampling technique described in Sec. 6.2, each computation of the $C^{(*)}$ at a specific $(\tau, \xi; \tau^*, \xi^*)$ on average requires the simulation of $O(10^6)$ particles, each of which consists of $O(10^6)$ time steps. This level of computation effort requires special hardware and parallelisation to keep the time requirement at a practical level. In light of the recent advances in general purpose computing in Graphical Processing Units (GPUs), and the computational power available in these units, we have chosen to exploit GPU computing using the NVIDIA CUDA platform. The CUDA platform offers a C–like language environment, where ordinary codes are executed on the CPU and stored in the system memory as usual. In addition, it contains extensions which allow data and computation tasks to be transferred to the GPU upon request, on which a large number of parallel tasks can be computed simultaneously. Each of these tasks is executed on a CUDA core on the GPU, and has access to a local memory private to that task. These tasks are organised into groups called "blocks", the members of which have access to a block–wide shared memory, and the tasks in a block are guaranteed to be executed simultaneously on the hardware (which is otherwise not guaranteed, if there are more tasks than cores). Most default numerical C functions are supported on the GPU.

During step 2 and 7 in the adaptive sampling process in Sec. 6.2, collisions are launched with a variety of randomised initial conditions, and simulated using the Boris pusher. These are the most computationally intensive steps in the whole scheme, with minimal memory operations, which makes them ideal candidates for off–loading onto the GPU. The hardware structure and driver design of the GPU means it is necessary to provide a very high number of collisions to the GPU for simultaneous computation to fully exploit its computational power.

On an NVIDIA GeForce GTX 780 GPU, we have identified $2^{14} = 16,384$ collisions, separated into 64 blocks, as an optimal number. (As the collisions in our case are unrelated to each other, and do not require simultaneous comutation or shared memory, common sense suggests a block size of 1 would give the GPU maximum flexibility in assigning the tasks to the processors. However practical experiment indicates a block size below 16 would significantly degrade performance, possibly due to overheads in block allocation in the hardware.) For step 2, this number of simultaneous collisions computed is achieved by requesting 1024 collisions to be launched from each box, and processing 16 boxes at a time. For step 7, we have chosen to first identify 16 boxes that require further refinement in step 6, and request 1024 collisions to be launched from each of them.

As CUDA is a relatively new technology, its usage is not completely streamlined. Several observations of its behaviour should be mentioned as a precaution: 1. Double precision (64 bit) computation is a necessity in the Boris pusher to ensure its accuracy. Double precision is supported by CUDA 1.3 or above; however the current version of the compiler defaults to CUDA 1.0, which only supports single precision (32 bit) real numbers, and double precision variables declared on the GPU are automatically degraded to single precision without warning messages. 2. The gaming version of the NVIDIA GPUs handles the graphical rendering of the operating system's interface, even when it is executing a CUDA programme. If a batch of tasks transferred to the GPU for one simultaneous computation take longer than ~ 2 s (which the Boris pusher typically does), the operating system will identify the unresponsive card as having crashed, and reset its driver, hence terminating the computation on it. It is therefore advisable to disable the driver reset function in the operating system before using CUDA. 3. In Microsoft Windows the GPU driver is replaced by a software rendering driver when a user logs in through Remote Desktop Connection, which will disable CUDA functionality. It is necessary to have physical access to the computer to start a CUDA programme using the local display. One proven workaround for this issue is to use third party remote control programmes like VNC, which does not replace the GPU driver.

Apart from the Boris particle pusher, the other aspects of the scheme are executed on the CPU in normal C code. Most operations have negligible computation requirement, apart from the bootstrapping error estimate, which can be efficiently parallelised on the multi-core CPU by computing the 1,000 bootstrapped samples for the average on various CPU cores in parallel.

6.5 Sampling the collision coefficients

In Sec. 6.2 we established a method of evaluating $C^{(*)}$ at specific values of τ, ξ, τ^* and ξ^*. From the numerical scheme described in Ch. 5 (Eqs. 5.41 to 5.46), and the definitions for the advection and diffusion coefficients for intermediately magnetised collisions (Eqs. 6.22 to 6.26), it is evident that we require the evaluation of $C^{(*)}$ at all possible combinations of the grid points for the un-starred species, $(\tau, \xi) = (\tau_i, \xi_j)$, and the grid points in the starred

species, $(\tau^*, \xi^*) = (\tau_{i^*}, \xi_{j^*})$. However, computing $C^{(*)}$ for all these possible combinations is computationally prohibitive. We have chosen to sample the arguments for $C^{(*)}$ adaptively and interpolate between them. As stated in Sec. 6.3, the change of the parallel and perpendicular velocities for the un-starred and starred particles are only a function of the *relative* parallel velocity $v_z - v_z^*$ between them due to the translation symmetry in the z-direction, which allows us to express

$$C^{(*)}(\tau, \xi; \tau^*, \xi^*) = C^{(*)}(\tau, \xi - \xi^*; \tau^*, 0).$$

Furthermore, the mirror symmetry in the z-direction means

$$\begin{aligned}
C^{\tau}(\tau, -\xi; \tau^*, 0) &= C^{\tau}(\tau, \xi; \tau^*, 0) & \qquad C^{\tau *}(\tau, -\xi; \tau^*, 0) &= C^{\tau *}(\tau, \xi; \tau^*, 0) \\
C^{\xi}(\tau, -\xi; \tau^*, 0) &= -C^{\xi}(\tau, \xi; \tau^*, 0) & \qquad C^{\xi *}(\tau, -\xi; \tau^*, 0) &= -C^{\xi *}(\tau, \xi; \tau^*, 0) \\
C^{\tau\tau}(\tau, -\xi; \tau^*, 0) &= C^{\tau\tau}(\tau, \xi; \tau^*, 0) & \qquad C^{\tau\tau *}(\tau, -\xi; \tau^*, 0) &= C^{\tau\tau *}(\tau, \xi; \tau^*, 0) \\
C^{\xi\xi}(\tau, -\xi; \tau^*, 0) &= C^{\xi\xi}(\tau, \xi; \tau^*, 0) & \qquad C^{\xi\xi *}(\tau, -\xi; \tau^*, 0) &= C^{\xi\xi *}(\tau, \xi; \tau^*, 0) \\
C^{\tau\xi}(\tau, -\xi; \tau^*, 0) &= -C^{\tau\xi}(\tau, \xi; \tau^*, 0) & \qquad C^{\tau\xi *}(\tau, -\xi; \tau^*, 0) &= -C^{\tau\xi *}(\tau, \xi; \tau^*, 0).
\end{aligned}$$

These relations allow $C^{(*)}$ to be fully explored by limiting the evaluations to within $(\tau, \xi; \tau^*, 0)$, where all three arguments are restricted to non-negative values. The first and third arguments range from zero to the maximum of the un-starred and starred species' grids in the perpendicular direction respectively, while the second argument ranges from zero to the *sum* of the maximum of the un-starred and starred species' grids in the parallel direction. In order words,

$$\tau \in [0, \tau_{\max}], \qquad \tau^* \in [0, \tau^*_{\max}], \qquad \xi \in [0, \xi_{\max} + \xi^*_{\max}].$$

To sample these three dimensions, we discretise the parameter space firstly along ξ into $\{\xi_k, k \in [0, N-1]\}$. At each ξ_k, the two remaining dimensions τ and τ^* are discretised into a grid of $\{\tau_{k,l}, l \in [0, N_k - 1]\}$ and $\{\tau^*_{k,m}, m \in [0, N^*_k - 1]\}$ (see Fig. 6.6). The ten $C^{(*)}$ are evaluated at the grid points as

$$C^{(*)}_{k,l,m} \equiv C^{(*)}(\tau_{k,l}, \xi_k; \tau^*_{k,m}, 0),$$

using the method outlined in Sec. 6.2. To obtain $C^{(*)}$ in between these evaluation points, we first interpolate on the $\tau - \tau^*$ plane (Fig. 6.6 b) using two dimensional linear interpolation to compute the values for $C^{(*)}$ at the queried value of τ and τ^*, and then linearly interpolate between the planes for the queried value of ξ. Linear interpolation is used instead of higher order schemes to prevent Runge's phenomenon, since the evaluation grid has limited resolution due to computational limitations.

To improve the accuracy of this discrete representation of $C^{(*)}$, the grid can be refined by either introducing additional columns or rows in the $\tau - \tau^*$ plane at a particular k, or

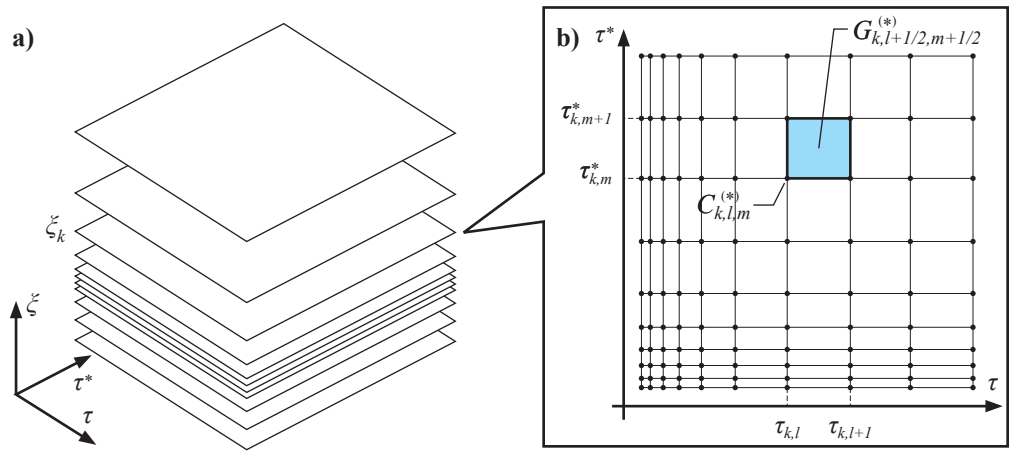

Figure 6.6: The coordinates for evaluating and interpolating the collision coefficients $C^{(*)}$. The three dimensional parameter space is first discretised in ξ into "horizontal" slices, as shown in a), and at each value for ξ a non-uniform grid specifies the points of evaluation, shown as black dots in b). The grid in $\tau - \tau^*$ space can be refined by considering the difference between the first and second order interpolation within each cell in b), which is shown in blue and has a value of $G^{(*)}_{k,l+1/2,m+1/2}$.

by inserting additional slices in ξ. For the first type of refinement, we require a way to determine the columns and rows which require refinement most urgently. This is achieve by first introducing the slopes in the τ and τ^* directions

$$s^{(*)}_{k,l+1/2,m} = \frac{C^{(*)}_{k,l+1,m} - C^{(*)}_{k,l,m}}{\tau_{k,l+1} - \tau_{k,l}}, \qquad s^{(*)}_{k,l,m+1/2} = \frac{C^{(*)}_{k,l,m+1} - C^{(*)}_{k,l,m}}{\tau^*_{k,m+1} - \tau^*_{k,l}},$$

from which we can define the averaged second derivatives

$$a^{(*)}_{k,l+1/2,m} = \frac{s^{(*)}_{k,l+3/2,m} - s^{(*)}_{k,l+1/2,m}}{\tau_{k,l+2} - \tau_{k,l}} + \frac{s^{(*)}_{k,l+1/2,m} - s^{(*)}_{k,l-1/2,m}}{\tau_{k,l+1} - \tau_{k,l-1}},$$
$$a^{(*)}_{k,l,m+1/2} = \frac{s^{(*)}_{k,l,m+3/2} - s^{(*)}_{k,l,m+1/2}}{\tau^*_{k,m+2} - \tau^*_{k,m}} + \frac{s^{(*)}_{k,l,m+1/2} - s^{(*)}_{k,l,m-1/2}}{\tau^*_{k,m+1} - \tau^*_{k,m-1}}.$$

This in turn gives the volume contained in the "gap" between a linear and quadratic interpolation on the $\tau - \tau^*$ plane, which we use to indicate the fidelity of the the discrete representation of $C^{(*)}$:

$$G^{(*)}_{k,l+1/2,m+1/2} = \frac{\Delta\tau_{k,l+1/2}\Delta\tau^*_{k,m+1/2}}{24}\Bigg| \left(a^{(*)}_{k,l+1/2,m} + a^{(*)}_{k,l+1/2,m+1}\right)(\Delta\tau_{k,l+1/2})^2 + \left(a^{(*)}_{k,l,m+1/2} + a^{(*)}_{k,l+1,m+1/2}\right)(\Delta\tau^*_{k,m+1/2})^2\Bigg|,$$

where $\Delta\tau_{k,l+1/2} = \tau_{k,l+1} - \tau_{k,l}$ and $\Delta\tau^*_{k,m+1/2} = \tau^*_{k,m+1} - \tau^*_{k,m}$. Note that here the $(*)$ superscript can be replaces by τ, ξ, $\tau\tau$, $\xi\xi$, $\tau\xi$, $\tau*$, $\xi*$, $\tau\tau*$, $\xi\xi*$ or $\tau\xi*$, for a total of ten values. If we were to introduce an extra column mid-point between $\tau_{k,l}$ and $\tau_{k,l+1}$ in the ξ_k slice, the improvement to the overall accuracy of the representation is proportional to

$$\frac{\xi_{k+1} - \xi_{k-1}}{2} \sum_{m=0}^{N^*_k - 2} G^{(*)}_{k,l+1/2,m+1/2}.$$

Similarly, if we were to introduce an extra row between $\tau^*_{k,m}$ and $\tau^*_{k,m+1}$ in the ξ_k slice, the improvement is proportional to

$$\frac{\xi_{k+1} - \xi_{k-1}}{2} \sum_{l=0}^{N_k - 2} G^{(*)}_{k,l+1/2,m+1/2}.$$

By comparing these numbers for all k, l and m, one can determine the priority in which additional columns and rows should be inserted to achieve the best improvement in fidelity.

For the second type of refinement — inserting a new slice at a new ξ — we first define the two–dimensional linear interpolation on the $\tau - \tau^*$ plane in the ξ_k slice, $C^{(*)}_k(\tau, \tau^*)$. From this, we can define the slope

$$s^{(*)}_{k+1/2}(\tau, \tau^*) = \frac{C^{(*)}_{k+1}(\tau, \tau^*) - C^{(*)}_k(\tau, \tau^*)}{\xi_{k+1} - \xi_k}$$

and the second derivative

$$a^{(*)}_{k+1/2}(\tau, \tau^*) = \frac{s^{(*)}_{k+3/2}(\tau, \tau^*) - s^{(*)}_{k+1/2}(\tau, \tau^*)}{\xi_{k+2} - \xi_k} + \frac{s^{(*)}_{k+1/2}(\tau, \tau^*) - s^{(*)}_{k-1/2}(\tau, \tau^*)}{\xi_{k+1} - \xi_{k-1}}.$$

The "area" contained between a linear and quadratic interpolation in the ξ direction is then given by

$$G^{(*)}_{k+1/2}(\tau, \tau^*) = \frac{(\Delta\xi_{k+1/2})^3}{12} \left| a^{(*)}_{k+1/2}(\tau, \tau^*) \right|,$$

where $\Delta\xi_{k+1/2} = \xi_{k+1} - \xi_k$. The improvement to the fidelity of the discrete $C^{(*)}$ representation when an extra slice is introduced at the mid–point between ξ_k and ξ_{k+1} is then proportional to

$$\int_0^{\tau_{\max}} \tau \, \mathrm{d}\tau \int_0^{\tau^*_{\max}} \tau^* \, \mathrm{d}\tau^* G^{(*)}_{k+1/2}(\tau, \tau^*).$$

By comparing this quantity for all k, we can determine at where additional slices should be inserted.

6.6 Energy conservation

Using a similar methodology to Sec. 5.6, it can be shown that the net energy change caused by the discrete Fokker–Planck operator per time step for the collision between two species is given by

$$\frac{\Delta E}{\Delta t} = \sum_{ij}\sum i^* j^* (f\Delta\Omega)_{i,j}(f^*\Delta\Omega^*)_{i^*,j^*}\Big[m\big(\tau_i C^{\tau}_{i,j,i^*,j^*} - C^{\tau\tau}_{i,j,i^*,j^*} + \xi_j C^{\xi}_{i,j,i^*,j^*} - C^{\xi\xi}_{i,j,i^*,j^*}\big) + m^*\big(\tau^*_{i^*} C^{\tau *}_{i,j,i^*,j^*} - C^{\tau\tau *}_{i,j,i^*,j^*} + \xi^*_{j^*} C^{\xi *}_{i,j,i^*,j^*} - C^{\xi\xi *}_{i,j,i^*,j^*}\big)\Big]. \tag{6.32}$$

The first four terms inside the square brackets arise from the energy change in the distribution f of the un-starred species, while the last four terms arise from the energy change in f^* of the starred species. Using the definitions of the $C^{(*)}$, Eqs. 6.27 to 6.31, it is easy to show that

$$\begin{aligned}
\Delta E =& \sum_{ij}\sum_{i^*j^*}(f\Delta\Omega)_{i,j}(f^*\Delta\Omega^*)_{i^*,j^*}\int_0^{10\lambda_D}\int_0^{2\pi}\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\pi(10\lambda_D)^2}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta^*}{2\pi} \\
&\times\Bigg[m\bigg(\tau\Delta v_\perp + \frac{1}{2}(\Delta v_\perp)^2 + \xi\Delta v_\xi + \frac{1}{2}(\Delta v_\xi)^2\bigg) \\
&+ m^*\bigg(\tau^*\Delta v^*_\perp + \frac{1}{2}(\Delta v^*_\perp)^2 + \xi^*\Delta v^*_\xi + \frac{1}{2}(\Delta v^*_\xi)^2\bigg)\Bigg]^\dagger \\
=& \sum_{ij}\sum_{i^*j^*}(f\Delta\Omega)_{i,j}(f^*\Delta\Omega^*)_{i^*,j^*}\int_0^{10\lambda_D}\int_0^{2\pi}\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\pi(10\lambda_D)^2}\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta^*}{2\pi} \\
&\times\frac{1}{2}\Bigg[m\bigg((\tau+\Delta v_\perp)^2 + (\xi+\Delta v_\xi)^2 - \tau - \xi\bigg) \\
&+ m^*\bigg((\tau^*+\Delta v^*_\perp)^2 + (\xi^*+\Delta v^*_\xi)^2 - \tau^* - \xi^*\bigg)\Bigg]^\dagger.
\end{aligned}$$

The dagger superscript indicates the changes to velocity are evaluated for a pair of particles with initial velocities (τ_i, θ, ξ_j) and $(\tau^*_{i^*}, \theta^*, \xi^*_{j^*})$, colliding at impact parameter b and impact angle ϕ. The expression in the square bracket is obviously the change to the total energy of the particle pair after the collision. As two–particle collisions in a uniform magnetic field must preserve the total energy of the particles, we have shown that our discretisation scheme is energy conserving.

Numerically, if we calculate all the collision coefficients $C^{(*)}_{i,j,i^*,j^*}$ for all combinations of $\{i, j, i^*, j^*\}$ through the adaptive Monte Carlo algorithm, the precision of the energy conservation of the collisional Fokker–Planck operator only depends on the Boris particle pusher's ability to conserve energy while simulating individual collisions. As we adaptively adjust the Boris pusher's time–step size using the total energy error as an indicator, we have direct control over the precision of energy conservation.

However, as we cannot directly compute all $C^{(*)}$ due to computational constrains, interpolation is necessary. Special attention is require during the interpolation to preserve energy conservation. From Eq. 6.32, and using the symmetry conditions for $C^{(*)}$, the Fokker–Planck equation is energy–conserving if the following are satisfied for all τ, τ^* and ξ:

$$\begin{aligned} 0 =& m(\tau C^{\tau}(\tau,\xi;\tau^*,0) - C^{\tau\tau}(\tau,\xi;\tau^*,0) + \xi C^{\xi}(\tau,\xi;\tau^*,0) - C^{\xi\xi}(\tau,\xi;\tau^*,0)) \\ &+ m^*(\tau^* C^{\tau *}(\tau,\xi;\tau^*,0) - C^{\tau\tau *}(\tau,\xi;\tau^*,0) - C^{\xi\xi *}(\tau,\xi;\tau^*,0)) \qquad (6.33) \\ 0 =& mC^{\xi}(\tau,\xi;\tau^*,0) + m^* C^{\xi *}(\tau,\xi;\tau^*,0). \qquad (6.34) \end{aligned}$$

The first condition corresponds to the conservation of energy, while the second corresponds to the conservation of momentum in the axial direction. These conditions are automatically satisfied at the τ, τ^* and ξ where the coefficients are directly evaluated, barring the small numerical error in the Boris pusher.

Multiplying Eq. 6.34 on both sides by ξ, we have

$$0 = m\xi C^{\xi}(\tau,\xi;\tau^*,0) + m^*\xi C^{\xi *}(\tau,\xi;\tau^*,0). \qquad (6.35)$$

Using the fact that the linear superposition of Eqs. 6.33 and 6.35 at different τ, τ^* and ξ still yields zero on the left hand side, we see that if we linearly interpolate between two directly evaluated points a and b using

$$\begin{aligned} C^{\tau}(\tau,\xi;\tau^*,0) &= \frac{w_a\tau_a C^{\tau}(\tau_a,\xi_a;\tau_a^*,0) + w_b\tau_b C^{\tau}(\tau_b,\xi_b;\tau_b^*,0)}{\tau} \\ C^{\xi}(\tau,\xi;\tau^*,0) &= \frac{w_a\xi_a C^{\xi}(\tau_a,\xi_a;\tau_a^*,0) + w_b\xi_b C^{\xi}(\tau_b,\xi_b;\tau_b^*,0)}{\xi} \\ C^{\tau\tau}(\tau,\xi;\tau^*,0) &= w_a C^{\tau\tau}(\tau_a,\xi_a;\tau_a^*,0) + w_b C^{\tau\tau}(\tau_b,\xi_b;\tau_b^*,0) \\ C^{\xi\xi}(\tau,\xi;\tau^*,0) &= w_a C^{\xi\xi}(\tau_a,\xi_a;\tau_a^*,0) + w_b C^{\xi\xi}(\tau_b,\xi_b;\tau_b^*,0) \\ C^{\tau\xi}(\tau,\xi;\tau^*,0) &= w_a C^{\tau\xi}(\tau_a,\xi_a;\tau_a^*,0) + w_b C^{\tau\xi}(\tau_b,\xi_b;\tau_b^*,0), \end{aligned}$$

and

$$\begin{aligned} C^{\tau *}(\tau,\xi;\tau^*,0) &= \frac{w_a\tau_a^* C^{\tau *}(\tau_a,\xi_a;\tau_a^*,0) + w_b\tau_b^* C^{\tau}(\tau_b,\xi_b;\tau_b^*,0)}{\tau^*} \\ C^{\xi *}(\tau,\xi;\tau^*,0) &= \frac{w_a\xi_a C^{\xi *}(\tau_a,\xi_a;\tau_a^*,0) + w_b\xi_b C^{\xi *}(\tau_b,\xi_b;\tau_b^*,0)}{\xi} \\ C^{\tau\tau *}(\tau,\xi;\tau^*,0) &= w_a C^{\tau\tau *}(\tau_a,\xi_a;\tau_a^*,0) + w_b C^{\tau\tau *}(\tau_b,\xi_b;\tau_b^*,0) \\ C^{\xi\xi *}(\tau,\xi;\tau^*,0) &= w_a C^{\xi\xi *}(\tau_a,\xi_a;\tau_a^*,0) + w_b C^{\xi\xi *}(\tau_b,\xi_b;\tau_b^*,0) \\ C^{\tau\xi *}(\tau,\xi;\tau^*,0) &= w_a C^{\tau\xi *}(\tau_a,\xi_a;\tau_a^*,0) + w_b C^{\tau\xi *}(\tau_b,\xi_b;\tau_b^*,0), \end{aligned}$$

where w_a and w_b are the interpolation weights, energy and momentum conservation for the interpolated point is preserved.

6.7 Comparison with analytic model

As is the case with the weakly magnetised collisional Fokker–Planck model in the previous chapter, analytic solutions and approximations to intermediately magnetised collisions only exist for a few special cases. Glinsky et al. [68] obtained the thermal equilibration rate of a plasma's parallel and perpendicular temperatures through intermediately magnetised collisions, assuming the velocity distribution is Gaussian in both degrees of freedom, and the temperature difference between them is small. This rate was obtained through a Monte Carlo simulation of the collisions, and the result converged to the weakly and strongly magnetised equilibration rates in the limiting cases. The equilibration of the parallel and perpendicular temperatures, T_z and $T_\perp$, of a plasma with density n, particle mass m and charge q is described by

$$\frac{\mathrm{d}T_\perp(t)}{\mathrm{d}t} = \gamma(T_z - T_\perp), \qquad T_z(t) + 2T_\perp(t) = T_z(0) + 2T_\perp(0),$$

where the second equation comes from energy conservation. The decay factor γ is given by

$$\gamma = n\bar{v}\bar{b}^2 I(\bar{\kappa}),$$

where $\bar{v} \equiv \sqrt{2k_B\bar{T}/m}$ is the average speed of the particles. The mean temperature is $\bar{T} \equiv (2T_\perp(0) + T_z(0))/3$, and is conserved through time. The quantity $\bar{b}$ is twice the distance of closest approach, given by

$$\bar{b} = 2b_{\min}, \qquad b_{\min} = \frac{q^2}{4\pi\epsilon_0 k_B \bar{T}}.$$

The value $\bar{\kappa}$ is a dimensionless number reflecting the degree of magnetisation of the plsama, and is defined as

$$\bar{\kappa} = \frac{\omega_C \bar{b}}{\bar{v}} = \frac{qB\bar{b}}{m\bar{v}}.$$

The function $I(\bar{\kappa})$ reflects the magnetisation effect on the equilibration rate, and its value is tabulated in Table 6.1, reproduced from Glinsky et al. [68]. For values of $\bar{\kappa}$ other than what is available in Table 6.1, we have chosen to interpolate the function I in the Log–Log scale.

We apply this result to the thermal equilibration of a positron plasma with $T_z(0) = 40$ K, $T_\perp(0) = 20$ K, and a density of $n = 7.0 \times 10^{13}\mathrm{m}^{-3}$, in a magnetic field of $B = 1$ T, which is a typical condition in the ALPHA apparatus. The magnetisation number for this plasma is $\bar{\kappa} = 11.9$, which is neither weakly nor strongly magnetised, and the corresponding interpolated value for $I(\bar{\kappa})$ is 6.57×10^{-3}. This should be compared to the near–unity magnitude for when the collisions are weakly magnetised. Equilibration is suppressed by the presence of the magnetic field.

Comparing this with our collisional Fokker–Planck model, we first sample the collision coefficients in the three–dimensional space spanned by τ, τ^* and ξ, in the range which is

$\bar{\kappa}$	$I(\bar{\kappa})$
1.00×10^{-4}	$1.753(63) \times 10^{0}$
1.00×10^{-3}	$1.335(44) \times 10^{0}$
1.00×10^{-2}	$9.26(45) \times 10^{-1}$
1.00×10^{-1}	$5.90(36) \times 10^{-1}$
3.33×10^{-1}	$3.81(18) \times 10^{-1}$
9.99×10^{-1}	$1.927(46) \times 10^{-1}$
1.25×10^{0}	$1.572(38) \times 10^{-1}$
2.50×10^{0}	$8.17(16) \times 10^{-2}$
5.00×10^{0}	$3.34(20) \times 10^{-2}$
1.25×10^{1}	$5.91(37) \times 10^{-3}$
2.50×10^{1}	$9.19(38) \times 10^{-4}$
5.00×10^{1}	$7.42(27) \times 10^{-5}$
1.00×10^{2}	$2.74(13) \times 10^{-6}$
2.00×10^{2}	$2.94(11) \times 10^{-8}$
5.00×10^{2}	$9.48(44) \times 10^{-12}$
1.00×10^{3}	$2.527(61) \times 10^{-15}$
2.00×10^{3}	$5.16(24) \times 10^{-20}$
5.00×10^{3}	$1.531(57) \times 10^{-28}$
1.00×10^{4}	$2.90(50) \times 10^{-37}$

Table 6.1: The numerical values of the function $I(\bar{\kappa})$, which reflects the magnetisation effect on collisional equilibration. The digits inside brackets indicates the statistical error in the last two digits of the mantissa. Reproduced from Glinsky et al. [68].

relevant to the $T_z(0) = 40$ K, $T_\perp(0) = 20$ K plasma. The sampled points are shown in Fig. 6.7. It should be noted that as we are studying the self–collisions of positrons, the un-starred and starred species refer to the same population. This symmetry means that the collision coefficients must be diagonally symmetric in the τ–τ^* space. By limiting the evaluations to the lower half triangle, and mirroring the results to the upper half, computational resources can be saved.

Interpolating these coefficients between the evaluated points using the energy–conserving scheme described above, and putting them into the discrete, energy–conserving Fokker–Planck equation, we can simulate the equilibration of a heterogeneous positron plasma with an initial distribution $f(0) \propto \exp(-mv_\perp^2/(2k_B T_\perp(0)) - mv_z^2/(2k_B T_z(0)))$. Preliminary comparison shows that the time scale of the behaviour between the two models agree with each other, although our model predicts an equilibration curve which is somewhat different in shape from the exponential decay predicted by Glinsky et al. This difference may be the result of an insufficient resolution of the collision coefficients. The true cause of the difference is being investigated. The application of the intermediately magnetised collision operator to the equilibration between positrons and antiprotons is also being pursued.

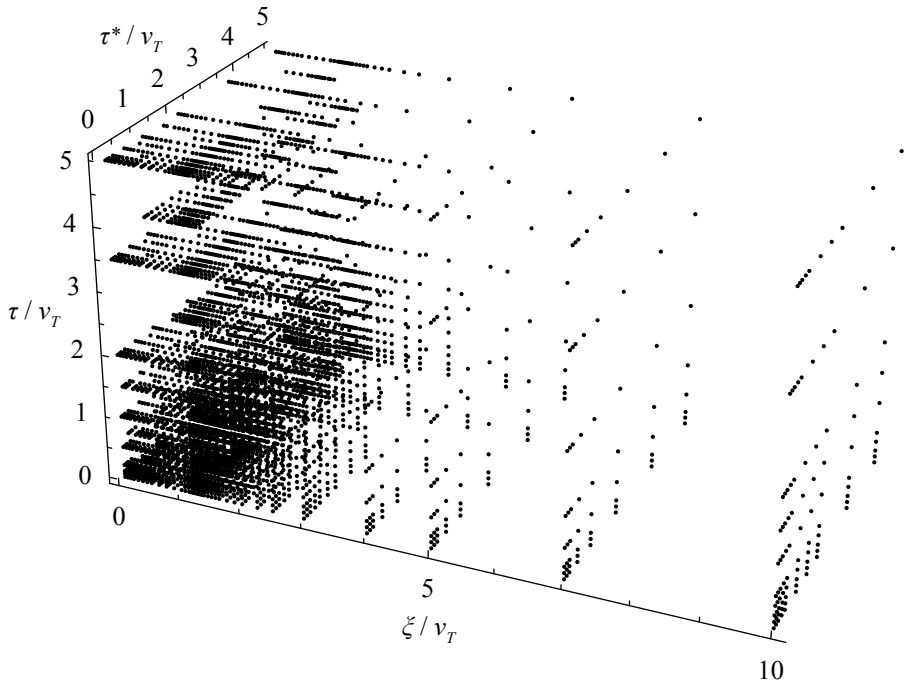

Figure 6.7: The adaptively sampled points in the collision coefficients' argument space, where they are directly evaluated. Here the self–collisions of a positron plasma at around $T = 40$ K is being studied, and $v_T = \sqrt{2k_BT/m}$.

Chapter 7

Mixing simulation

One of the possible — and most important — applications of the numerical models developed above is to simulate the mixing of antiprotons with positrons. As explained in Sec. 2.9, ALPHA achieved antihydrogen trapping using an autoresonant perturbation to excite the axial oscillation of antiprotons neighbouring a positron plasma. This method of mixing the two species is tolerant to machine fluctuations and minimises the energy of the resultant antihydrogen atoms. Despite this, the mixing and trapping process still result in the greatest percentage loss of particle in the entire experimental sequence. To improve antihydrogen yield, it is imperative to understand the details of the mixing and recombination process and how it can be improved. Qualitatively, the process starts with an autoresonant perturbation applied on the antiprotons, which excites their axial oscillation until they cross into the neighbouring positron plasma. The positrons, on the other hand, remain quasi-static since the perturbation is far below their bounce frequency. The self–field of the two species and the vacuum field from the electrodes (including the time–dependent perturbation) interact with each other during this excitation. Upon injection, the antiprotons usually travel at much higher speeds than the positrons (by virtue of their thermal motion), and the collision between them tend to cool the former and heat the latter. At the same time recombination is under progress, which depletes antiprotons and positrons and results in antihydrogen atoms. The cross–section of recombination is a function of the relative speed between the two particles, and the momentum of the antiproton at the instant of recombination determines the momentum of the resultant antihydrogen. This means there is a competition between collisional equilibration and recombination. If recombination is more rapid, antihydrogen atoms are formed soon after antiprotons are injected, and the energy spectrum of the antihydrogen is whatever the injected antiprotons' is. If the collisional equilibration happens more rapidly, on the other hand, there is stronger heating on the positrons and the antiprotons are better thermalised and cooled. Antihydrogen formation would be delayed due to the higher positron temperature, and their energy spectrum is closer to a thermal distribution (which is probably cooler than the first scenario, but it is unclear if the number of antihydrogen atoms with energy below 0.5 K has increased).

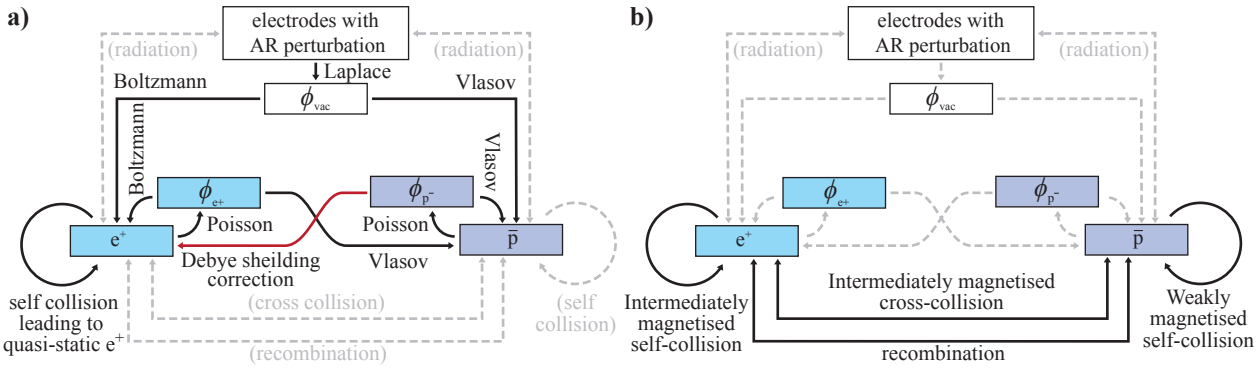

Figure 7.1: A schematic view of the interaction between various physics elements in the two phases of the injection simulation, with a) phase 1 simulating the excitation process, and b) phase 2 simulating the collisional equilibration and recombination between the two species. An arrow indicates the interaction between two objects, and its label indicates the numerical model used to simulate that interaction. Greyed–out arrows indicate interactions that are not significant in the time scale of that phase, and is thus ignored. The red arrow highlights the corrective algorithm used.

To qualitatively understand the detailed balance between these disparate processes, we have opted to simulate the mixing and recombination in a two–phase simulation separated by the collisional equilibration time scale of antiprotons, as explained below and also shown in Fig. 7.1.

Phase 1: the water bag – Vlasov hybrid model. This phase is focused primarily on solving for the spatial evolution of the plasmas during the antiproton excitation process. The antiprotons have an axial bounce period of ~ 3 μs, which is also the time scale of the perturbation applied, and the temporal resolution required of the simulation. The typical duration of this phase (the duration of the perturbation) is ~ 1 ms.

- The positron plasma has sufficient time to axially equilibrate with the external perturbation through collision. The positron plasma's low temperature of ~ 40 K, it is modelled quasi-statically using the waterbag equilibrium solver (Ch. 3), with the electrodes providing the background field ϕ_{vac} (including the perturbation). The influence from the antiprotons on the positrons is ignore for now, since the former is much fewer in number. This means the positron distribution is solved for as $f_{e^+}(r, z)$, with the distribution in v_z and $v_\perp$ assumed to remain Gaussian at 40 K, as a function of the perturbation (in this case the voltage of electrode E16). This decoupling of positrons from antiprotons allows the waterbag solution to be pre-computed, thus saving computer time.

- The antiprotons are simulated dynamically using the Vlasov model (Ch. 4), with the electric field $\phi \equiv \phi_{\text{vac}} + \phi_{e^+} + \phi_{\bar{p}}$ (the former two fields pre-computed by the waterbag

solver). The antiprotons' initial condition at the start of the simulation is obtained using the numerical annealing technique (Sec. 4.8), with $T = 250$ K. This means the antiproton distribution is solved for as $f_{\bar{p}}(r, z, v_z)$, with the distribution in $v_\perp$ assumed to remain Gaussian at 250 K.

- The time scale of the self–collision between antiprotons is much longer than the time scale in this phase of the simulation, and its effect is thus ignored (see Fig. 1.4). The cross–collision between antiprotons and positrons is also ignored since the two species only overlap each other momentarily towards the end of this phase. The self–collision between positrons is implicitly handled through the water bag solver. Since the external perturbation is slow compared to both the cyclotron frequency and the axial bounce, the energy in the positrons' axial and perpendicular motion remains unchanged, and the collisional operator takes no effect on the velocity distribution in either direction.

- The influence from antiprotons to positrons is small due to the former's comparative low number. However this does have a non-negligible effect in terms of the Debye shielding afforded on the antiprotons by the positrons when the two start to overlap. The positrons will rearrange themselves in response to the antiproton self-field to maintain the overall zero–field inside the bulk of the positron plasma. This can have an important impact on antiproton energy at the moment of injection. To simulate this small but important positron response, a correction algorithm is placed in the simulation that modifies the positron potential according to the antiproton self–field, and forces the total potential ϕ to be a constant against z within the modified positron boundary. For each radial slice at r_i, we first define $\Lambda_i \equiv \int_{L_i}^{R_i} \phi_{\bar{p}(r_i,z)}\, dz/(R_i - L_i)$. Also define the function

$$\phi_G(r_i, z) \equiv (\phi_{e^+} + \phi_{\text{vac}})(r_i, (L_i + R_i)/2) - (\phi_{\text{vac}} + \phi_{\bar{p}})(r_i, z) + \Lambda_i$$

  The modified boundaries $\tilde{L}_i$ and $\tilde{R}_i$ are defined as the roots of

$$\begin{aligned}\phi_G(r_i, \tilde{L}_i) =& \phi_{e^+}(r_i, L_i)\\ \phi_G(r_i, \tilde{R}_i) =& \phi_{e^+}(r_i, R_i),\end{aligned}$$

  that are closest to L_i and R_i respectively. Further introduce $\mathcal{L}_i$ and $\tilde{\mathcal{L}}_i$, which equal L_i and $\tilde{L}_i$ respectively if $\partial_z\phi_{e^+}(r_i, L_i) \geq \partial_z\phi_G(r_i, \tilde{L}_i)$. Otherwise they are the solutions of the simultaneous equations

$$\begin{cases} \phi_G(r_i, \tilde{\mathcal{L}}_i) = \phi_{e^+}(r_i, \mathcal{L}_i) \\ \partial_z\phi_G(r_i, \tilde{\mathcal{L}}_i) = \partial_z\phi_{e^+}(r_i, \mathcal{L}_i), \end{cases}$$

  which satisfy $\tilde{\mathcal{L}}_i \in [\Xi_i, \tilde{L}_i]$, where Ξ_i is the z value of the first maximum of $\phi_G(r_i, z)$ left of $\tilde{L}_i$. If multiple solutions exist, the one corresponding to the largest $\tilde{\mathcal{L}}_i$ is chosen.

If no such solution exists, $\tilde{\mathcal{L}}_i$ is set to Ξ_i, and $\mathcal{L}_i$ is the root of $\phi_G(r_i, \Xi_i) = \phi_{e^+}(r_i, \mathcal{L}_i)$ cloest to Ξ_i. A similar procedure is repeated for $\mathcal{R}_i$ and $\tilde{\mathcal{R}}_i$. The modified positron self–potential is then given by

$$\tilde{\phi}_{e^+}(r_i, z) = \begin{cases} \phi_{e^+}(r_i, z + (\mathcal{L}_i - \tilde{\mathcal{L}}_i)) & \text{for } z < \tilde{\mathcal{L}}_i \\ \phi_G(r_i, z) & \text{for } \tilde{\mathcal{L}}_i \leq z \leq \tilde{\mathcal{R}}_i \\ \phi_{e^+}(r_i, z + (\mathcal{R}_i - \tilde{\mathcal{R}}_i)) & \text{for } z > \tilde{\mathcal{R}}_i. \end{cases}$$

This $\tilde{\phi}_{e^+}$ is then used in place of ϕ_{e^+} in the Vlasov solver, and completes the red arrow in Fig. 7.1 a.

- Alternatively, and possibly more simply, the water bag solver can be executed upon every time-stepping in the Vlasov solver to derive the quasi-equilibrium state of the positrons, with the electrostatic influence from the antiprotons properly taken account of. This is possible with sufficient speed–up in the water bag solver, and indeed necessary when antiproton space charge start to become comparable to the positrons'.

Phase 2: the collision–recombination model. This phase is focused on the evolution of the velocity–space distribution of the two species due to collisional equilibration and recombination, which happen over a time scale of ~ 20 ms. After phase 1, the perturbation is terminated, and the system reaches a steady state where the antiprotons stream through the positron plasma at a constant rate while undergoing a wide axial bounce. The Debye shielding of the positrons means the positron density is also a constant against z within the plasma. These means the variation in z for both species' distribution can now be discarded as long as the overlapping region of the two species is concerned, and the phase 2 of the model will focus on this volume.

- Discarding the axial dependence, the antiproton distribution from the end of the phase 1 simulation is converted from $f_{\bar{p}}(r, z, v_z)$ to $f_{\bar{p}}(r, v_z, v_\perp)$ by averaging over z within the positron plasma. The $v_\perp$ degree of freedom is initially assumed to follow a Gaussian at 250 K. Similarly, for positrons, the distribution is converted from $f_{e^+}(r, z)$ to $f_{e^+}(r, v_z, v_\perp)$ by averaging it over z within the positron plasma, and the v_z and $v_\perp$ degrees of freedom are initially assumed to follow a Gaussian at 40 K.

- The weakly magnetised collision operator (Ch. 5) is used to model the effect of collisions between antiprotons on $f_{\bar{p}}$. The weakly magnetised version is used since the typical distance of closest approach between antiprotons is much smaller than their cyclotron orbit radius. This operator reflects how the self–collision of antiprotons transfers energy from the axial degree of freedom to the perpendicular one, and also thermalises the shape of the distribution within each degree of freedom as well.

- The more general intermediately magnetised collision operator is used to model the effect of collisions between positrons and antiprotons on both f_{e^+} and $f_{\bar{p}}$. Since the antiprotons are usually at a higher temperature than the positrons, and the energy equilibration is more efficient in the axial than the perpendicular direction, the collisions mainly lead to an axial cooling of the antiprotons and heating of the positrons.

- The intermediately magnetised collision operator is also used to model the effect of collisions within the positrons on f_{e^+}, which tends to transfer energy from the positron's axial degree of freedom (which is heated by the antiprotons) to the perpendicular.

- The recombination operator takes away particles from both distributions according to the recombination cross–section, and give the timing and energy spectrum of the antihydrogen atoms formed.

This separation of the collisional time scale has the distinction of restricting the number of degrees of freedom in the distribution to only three or less, which is essential in terms of computation. It is in principle possible to construct a numerical scheme combining the Vlasov equation and the collisional operators, but this would require handling a full 4–D distribution $f(r, z, v_z, v_\perp)$. Assuming each dimension apart from r is discretised to $O(N)$ grid points (r gridding can be sparse to $O(1)$), each time step would require $O(N^3)$ operations for the Vlasov equation, and $O(N^5)$ operations for the collision operators. Compared with the split scheme, in phase 1 the Vlasov equation requires $O(N^2)$ operations per time step, and in phase 2 the collision operators require $O(N^4)$ operations per time step. The size of the time step in phase 2 can be increased as well since only collisional and recombination are simulated, which are comparatively slow processes. The speed–up is significant. Note that this clear–cut separation of the two phases is only approximate: collision and recombination occurs during the excitation process when the two species start to overlap, and the spatial dynamics still plays a part as the space charge of the two species recombine and deplete.

In this chapter we first present a simple single–particle picture of the autoresonant excitation process, then go on to use the water bag – Vlasov model to simulate the autoresonant excitation of the antiproton plasma, and study the resultant energy spectrum of the injected antiprotons. The results presented in this chapter has previously been published in [70].

7.1 Basic principle of autoresonance

Autoresonant excitation has been applied to, and observed in, a wide variety of systems [34]. The principle of autoresonant excitation is most transparent in the case of a single particle in an anharmonic well. Autoresonant excitation only works when there is a monotonic relation between the amplitude and frequency of an oscillator; here we assume it is monotonically decreasing, which is the case for the antiproton well in Fig. 2.6. The oscillation frequency at vanishing amplitude is called the linear resonance, and is denoted by ω_0. A fixed frequency

perturbation at this ω_0 results in a limited excitation since, as the particle is excited, its oscillation frequency changes with amplitude and consequently loses phase lock with the drive. An autoresonant perturbation instead starts at a frequency above ω_0, and is chirped towards a lower frequency. As the frequency passes through ω_0, the particle becomes phase–locked to the perturbation, provided certain conditions are satisfied. The amplitude of the oscillator motion changes such that its frequency automatically matches that of the perturbation. Fajans and Friedland [71] gave an analytic treatment of the autoresonant excitation process for an oscillator with an equation of motion of

$$\ddot{z} + \omega_0^2 \left(1 - \frac{4}{3}\beta z^2\right) z = \epsilon \cos\left(\theta_D(t)\right) = \epsilon \mathrm{Re}\left(e^{i\theta_D(t)}\right), \tag{7.1}$$

and derived the conditions that must be met in order for phase–locking to occur. Here ϵ denotes the drive amplitude, $\theta_D(t) = \omega_0 t - \alpha t^2/2$ the drive phase angle, $-\alpha$ the rate of change of the drive frequency, or chirp rate, and β the nonlinearity of the oscillator. The motion starts at a large negative t with $z = \dot{z} = 0$, and the perturbation passes through the linear frequency at $t = 0$. The derivation starts by separating the fast and slow motion of the oscillator:

$$z(t) = \mathrm{Re}\left(a(t)e^{i\theta_P(t)}\right). \tag{7.2}$$

Here $a(t)$ is the slowly varying time–dependent oscillator amplitude and $\theta_P(t)$ is its phase. The amplitude $a(t)$ and phase difference $\delta(t) \equiv \theta_P - \theta_D$ vary on a timescale $\gg 1/\omega_0$.

For t near zero, and ignoring higher harmonics, substituting Eq. 7.2 into Eq. 7.1 gives

$$\dot{I} = -\frac{\epsilon}{\sqrt{2}\omega_0}\sqrt{I}\sin(\delta) \tag{7.3}$$

$$\dot{\delta} = \alpha t - \omega_0\beta I - \frac{\epsilon}{2\sqrt{2}\omega_0}\frac{1}{\sqrt{I}}\cos(\delta), \tag{7.4}$$

where $I(t) \equiv a^2(t)/2$. The phase difference $\delta(t)$ becomes locked near π. This is a stable phase, since if δ becomes smaller than π, the perturbation begins to do positive work on the oscillator, accelerating it back to $\delta = \pi$; when δ becomes larger than π, negative work is done, slowing it down.

Given δ is expected to stay around π and only change slowly, we define, from Eq. 7.4, the solution of $\dot{\delta} = 0$ as $I_0(t)$. Next, we define $S(t) \equiv \omega_0\beta + \epsilon/(2\sqrt{2}\omega_0 I_0^{3/2})$ and $\Delta(t) \equiv I_0(t) - I(t)$. The coupled Eqs. 7.3 and 7.4 are then simplified into the Hamiltonian

$$H(\delta, \Delta) = S\Delta^2/2 + V_{\mathrm{pseudo}}(\delta), \tag{7.5}$$

$$V_{\mathrm{pseudo}}(\delta) = -\frac{\alpha}{S}\delta + \frac{\epsilon}{\sqrt{2}\omega_0}\sqrt{I_0}\cos(\delta). \tag{7.6}$$

For the pseudo-particle to remain trapped in the pseudo-potential, V_{pseudo} must have parts with positive and negative slopes against δ for all times. The cosine part of Eq. 7.6 has a

slope varying between $-\epsilon\sqrt{I_0}/(\sqrt{2}\omega_0)$ and $\epsilon\sqrt{I_0}/(\sqrt{2}\omega_0)$, while the linear part has a slope of $-\alpha/S$. This means $|\epsilon/(\sqrt{2}\omega_0)\sqrt{I_0}| > |\alpha/S|$ is required at all times to maintain phase locking. At $I_0 = (\epsilon/(2\sqrt{2}\omega_0^2\beta))^{3/2}$ the inequality becomes most difficult to satisfy, at which point the requirement for phase locking becomes

$$\epsilon > \epsilon_{\rm cr} \equiv 2\sqrt{2}\sqrt{\frac{\omega_0}{\beta}}\left(\frac{\alpha}{3}\right)^{3/4}. \tag{7.7}$$

Generalizing the single particle dynamics above to the excitation of a antiproton bunch in a nested Penning–Malmberg trap is not trivial. If a test particle is placed in the potential formed by the external electrodes and the space charge, and subject to the autoresonant perturbation, it will not exhibit the behaviour above since the net electrostatic well does not have a monotonic relationship between amplitude and frequency. It is therefore important to analyse the collective behaviour of the plasma to understand its autoresonant excitation. Barth et al.[54] presented theoretical results showing that the self–field causes the plasma to remain coherent during an autoresonant perturbation. The first experimental observation of phase–locking and excitation in the collective regime was presented by Andresen et al.[46].

7.2 Comparisons with numerical and analytic models

In this section, the water bag – Vlasov model is compared with (1) the analytic model (Eqs. 7.5, 7.6 and 7.7), (2) a leap–frog single particle pusher that neglects the self–field of the antiproton bunch and treats it as a single particle, but evolves it under the same positron and vacuum fields as in the water bag – Vlasov model, and (3) a 1–D collisionless spectral Vlasov–Poisson model used by Barth et al.[54], which does not model any radial variation and solves the Poisson equation using an approximate radial cut–off in place of a true radial profile.

Time–resolved autoresonant excitation

The different models are applied to the autoresonant excitation of a 250 K, 10,000 antiproton bunch with a radius of ~ 0.7 mm (defined by the 90% material–inclusive equi–density contour). The particles are confined by an anharmonic electrostatic well with a linear bounce frequency of 412.7 kHz, created by the electrodes as shown in Fig. 7.2. The antiproton bunch is excited by an autoresonant perturbation applied to an electrode to the right of the bunch. The perturbation frequency changes linearly from 420 kHz to 200 kHz at a chirp rate of -200 MHz/s, and an amplitude of 0.14 V. A 10–period transition is present before the start of the chirp, where the perturbation amplitude is linearly increased from 0 to its full amplitude, at the starting frequency. Similarly, the perturbation amplitude is linearly decreased to 0 after the chirp at the stopping frequency in 10 periods. In the simulation, a further 20 periods are present (measured in terms of the stopping frequency), during which no perturbation

is applied, before the simulation is terminated. The results from the different models are displayed in Fig. 7.3, showing similar predicted behaviour between the models, in terms of both the energy and the phase of the antiproton bunch. The analytic model prediction shows a slightly higher excitation at late time and high amplitude, since it includes only the 4th order non–linearity of the confining well, while other models include all orders by using a physical confining potential. Also note that only the two Vlasov–based models can predict the spread in energy, since they take account of the finite–size effect of the antiproton bunch. The energy spread predicted by the spectral Vlasov model is higher than the water bag – Vlasov model, possibly due to the much lower resolution allowed in the former, leading to stronger numerical diffusion.

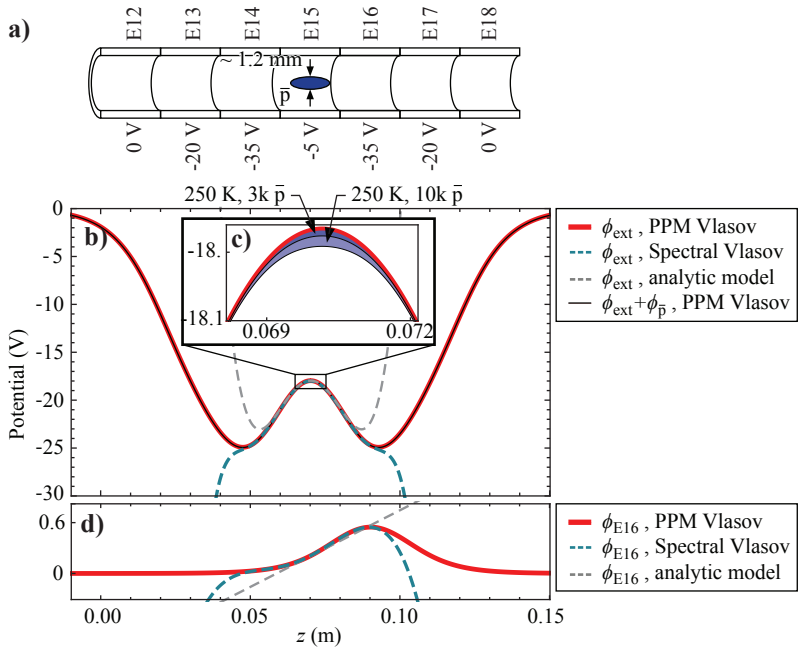

Figure 7.2: Potentials and geometry for measuring the autoresonant excitation of a antiproton bunch. a) The physical setup; the perturbation is applied on electrode E16. b) The external potential created by the electrodes at $r = 0$. c) A close–up of b, emphasizing the effect of various antiproton space charges. d) The perturbation created at $r = 0$ when 1 V is applied to E16. The potentials used in the water bag – Vlasov model are deduced by solving the 2D Poisson equation as outlined in Sec. 3.1. Those in the spectral Vlasov solver are analytic fits up to z^6. The background vacuum potential ϕ_{vac} in the analytic model is a fit up to z^4, while the perturbative force is assumed to be a constant.

Perturbation amplitude threshold for pickup

In Fig. 7.4, the analytic prediction for the critical drive amplitude ϵ_{cr} (Eq. 7.7) is compared with the single particle model and the water bag – Vlasov model. The set up is identical

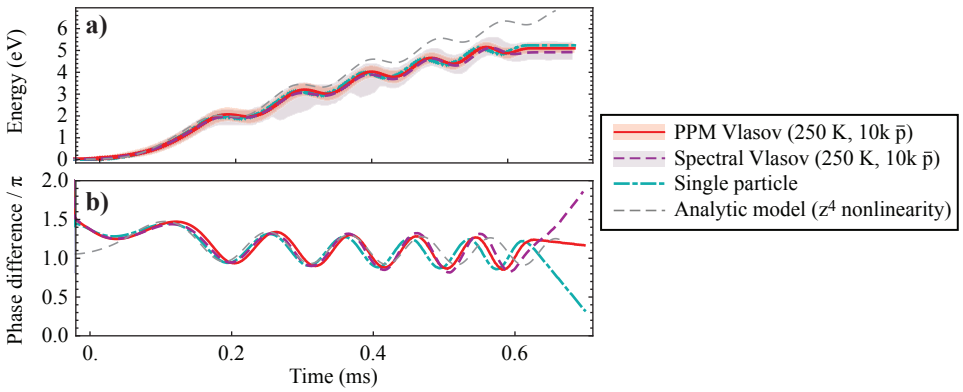

Figure 7.3: Time evolution of a) the energy and b) phase angle of the antiproton distribution, as predicted by different numerical and analytic models. The colour bands around the water bag – Vlasov and spectral Vlasov results indicate the lower 5% and upper 95% boundaries of the spread in energy of the phase space distribution. The phase difference is defined as $\theta_P - \theta_D$, where θ_P is the phase angle of the centre of charge of the distribution, and θ_D the phase angle of the autoresonant perturbation.

to that shown in Fig. 7.2, with a 250 K, 10,000 antiproton bunch subjected to different perturbations. These perturbations all start from 420 kHz with the 10–period ramp–up, and end at 360 kHz with the 10–period ramp–down, but with various amplitudes and the chirp rates. At each chirp rate on the horizontal axis of Fig. 7.4, multiple simulations with different drive amplitudes are executed, and a sudden jump in the final averaged antiproton energy is observed when the drive amplitude exceeds the critical value. This critical drive amplitude is plotted in the vertical axis of Fig. 7.4, together with the analytical prediction. Good agreement between the models is again observed.

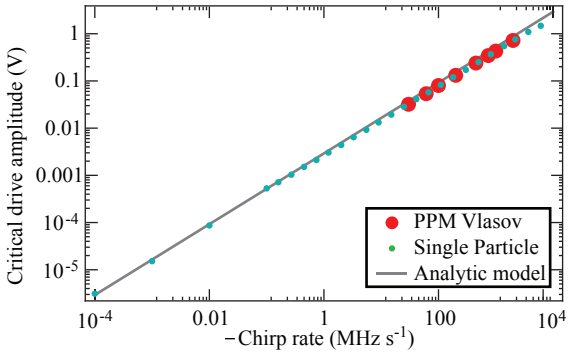

Figure 7.4: Critical autoresonant perturbation amplitude for various chirp rates. The prediction of the analytic model and the results from the single particle and water bag – Vlasov model are compared.

7.3 Comparisons with experiment

The results of experimental runs are compared with predictions of the water bag – Vlasov model and the single particle model. In the first comparison, a antiproton bunch is subjected to an autoresonant perturbations in an anharmonic well without any neighbouring positrons, and the resultant antiproton axial energy gain is analysed. In the second comparison, a antiproton bunch is subjected to perturbations next to an positron plasma. Some fraction of the antiprotons obtain sufficient energy to enter the positron plasma. The distribution of the kinetic energy of the injected antiprotons predicted by simulation is compared with the number of antihydrogen measured from experiment.

Final antiproton energy distribution versus drive amplitude and stopping frequency

A 250 K, 3000 ± 1000 antiproton bunch is prepared in the anharmonic well shown in Fig. 7.2, which has a linear frequency of 412.7 kHz. The particles are subjected to various autoresonant perturbations, all of which start from 420 kHz, have a chirp rate of -200 MHz/s and include the 10–period ramp–up and ramp–down. In the first series of runs, the stopping frequency is fixed at 360 kHz, and the drive amplitude varies between 0 V and 0.161 V. In the second series of runs, the drive amplitude is fixed at 0.15 V and the stopping frequency varies between 355 kHz and 390 kHz. From the simulations, the final energy of the antiprotons post–perturbation is derived from the final phase space distributions / single particle states, while experimentally, the final energy is measured by a temperature analysis ejection. The final energies obtained from the models and the experiment are compared in Fig. 7.5. Simulations with a lower initial antiproton temperature of 30 K are included to demonstrate temperature effects. The centres of charge of the bunch predicted by the models agree well with experimental measurements, but the water bag – Vlasov model predicts a broader energy distribution than observed in experiment when using a 250 K antiproton bunch.

Note that the experimental data in Fig. 7.5 have been fitted to correct for experimental systematics:

1. There is a time synchronization mismatch between the voltage controller for the electrodes and the silicon vertex detector, expected to be within 0.1 ms, introducing a possible offset between the escape time reported by the detector and the actual escape time with respect to the voltage changes being made on the electrodes during a dump. This is accounted for by a time shift of the detector signal such that the detector count from the 0 V drive amplitude experiment in Fig. 7.5 b corresponds to an average energy in well of 0 eV. This time offset is then fixed for all other experimental measurements of energy distributions.

2. The experimental drive amplitude quoted hitherto is the amplitude on the electrode, which is 0.54 times the amplitude at the waveform generator, due to the wiring impedance and high–pass filter between the generator and the electrodes (see Fig. 2.5). This conversion factor is derived by fitting the horizontal position of the jump in Fig. 7.5 a between the experiment and the simulation. This factor is then used in the analysis of all other experimental runs.

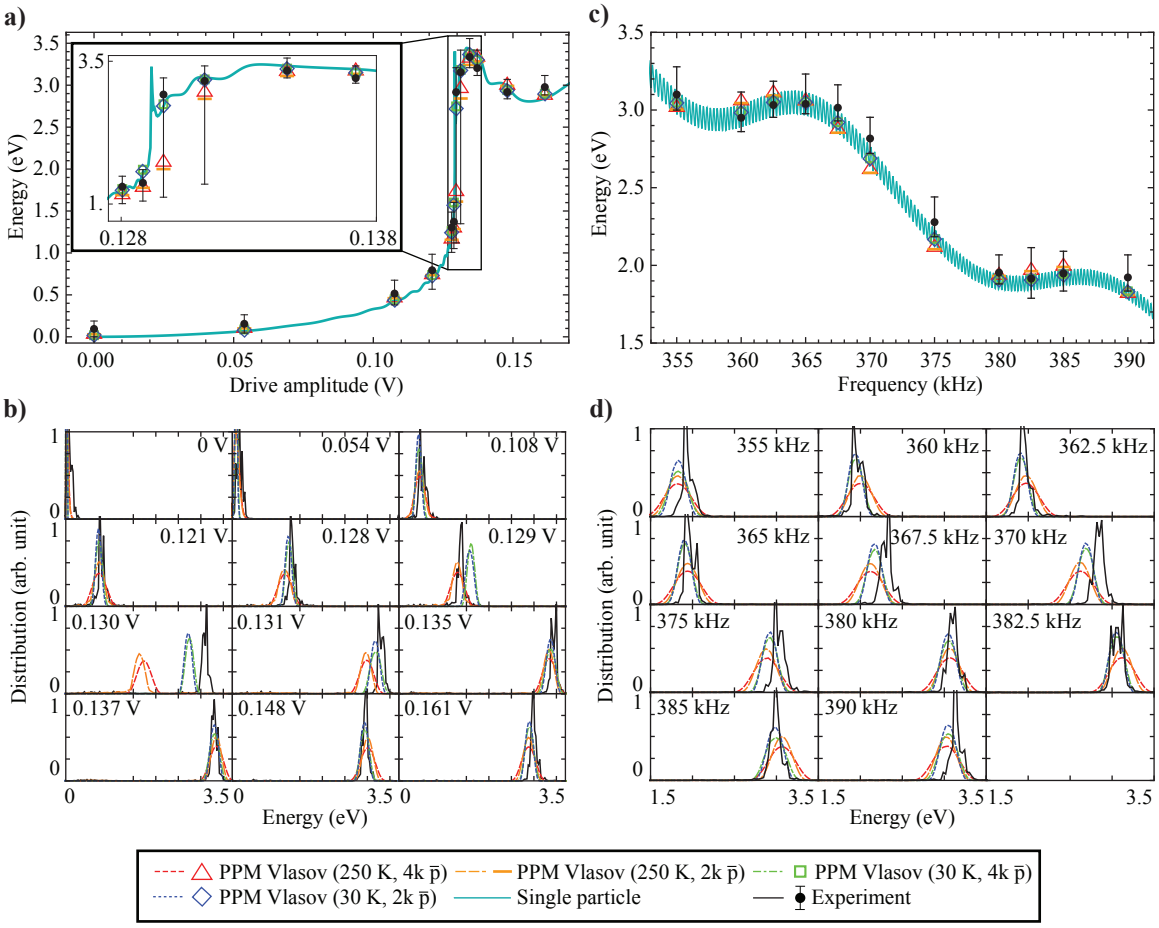

Figure 7.5: The final energy of an antiproton bunch after various autoresonant perturbations, as measured in the experiment and predicted by the single particle and the water bag — Vlasov models. a) The final antiproton energy after autoresonant perturbations of various amplitudes and a fixed stopping frequency of 360 kHz. The dotted lines indicate the lower 5% and upper 95% boundaries of the antiproton energy distribution. b) The energy distribution function for each of the data points in a. c) The final antiproton energy after autoresonant perturbations of various stopping frequencies and a fixed amplitude of 0.15 V. d) The energy distribution function for each of the data points in c. The experimental data have been altered to correct for experimental systematics — see main text.

Injection ratio versus stopping frequency

Figure 2.6 shows the experimental setup injecting antiprotons into a positron plasma. A 250 K, 16,000 antiproton bunch is placed in a nested well next to a 40 K, 3×10^6 positron plasma, and subjected to an autoresonant perturbation. The antiproton well has a linear frequency of 297.4 kHz, and the perturbation starts at 325 kHz with a 10–period ramp–up to an amplitude of 0.08 V. It is then chirped down to 250 kHz at -120 MHz/s, and ends with a 10–period ramp–down. A fraction of the antiprotons gain enough energy to enter the positron plasma. Due to Debye shielding, the total potential within the positron plasma is a constant (in z), and therefore each of the injected antiprotons moves across the positron plasma at a constant speed. The simulated distribution of the axial speed of the injected antiprotons as they travel through the positron plasma after the perturbation is plotted in Fig. 7.6 a, together with several snapshots of the antiproton distribution during the autoresonant perturbation in Fig. 7.6 b.

In the experiment, the speed distribution of the injected antiprotons cannot be measured directly. Instead they collide and recombine with positrons. Most of the resultant antihydrogen are not confined by the magnetic minimum trap. They drift and eventually annihilate on contact with the electrode wall. These annihilations are recorded by the silicon vertex detector, and the total number of annihilations within a 1 s window after the perturbation is plotted in black in Fig. 7.7 c, against the stopping frequencies of the perturbation. This number of annihilation indicates the fraction of antiprotons that enters the positron plasma and successfully forms antihydrogen. Figures 7.7 a and b show the simulated fraction of antiprotons which successfully injects and has a kinetic energy or radius below various values. For instance, the "KE $<$ 100 K" curve in Fig. 7.7 a plots the fraction of antiproton, out of the original 16,000, that successfully enters the positron plasma and travels across it with a kinetic energy below 100 K as a function of the perturbation's stopping frequency. Qualitatively, the simulation shows that a chirp stopping below $\sim$ 290 kHz is necessary for injection. The injected fraction increases as increasingly long chirps are used, but the fraction injected at lower energies ($<$ 100 K) slowly saturates when the stopping frequency is below $\sim$ 240 kHz. The simulation also shows that antiprotons at smaller radii are injected earlier, while those at the outer radii require a longer chirp to reach injection. It is also observed that stopping frequencies much lower than 200 kHz (not shown in Fig. 7.6) cause the speed distribution of the injected antiprotons to broaden, which is expected since the perturbation, having no frequency relation to the bounce orbits of the injected population, only acts as a heating signal. The number of antihydrogen formed, as measured in the experiment, increases with the length of the perturbation, until saturating at a stopping frequency of $\sim$ 250 kHz (see Fig. 7.6 c). This roughly agrees with the simulation.

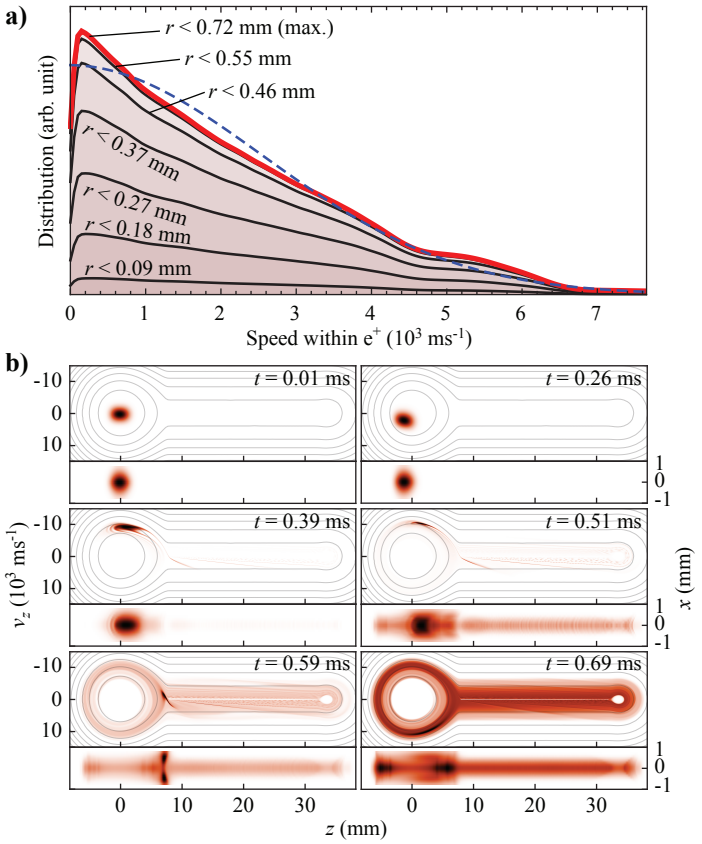

Figure 7.6: a) The simulated distribution in speed of injected antiprotons as they travel across the positron plasma, conditioned on the radius. The blue dotted curve shows a reference thermal distribution of antiprotons at 800 K, that has the same area under the curve as the all–radii curve. The total number of injected antiproton is 7,400 (out of the 16,000 initial antiprotons). b) Snapshots of the simulated antiproton distributions at various t during the perturbation, which starts at $t = 0$. The contours are lines of constant total energy, and increase by 0.25 eV (2900 K) per contour. At each time the (z, v_z) phase space at $r = 0$ is displayed, together with the (x, z) charge density.

7.4 Injection limits

The main adjustable parameters of an antihydrogen production and trapping run in ALPHA are the numbers, radial sizes and temperatures of the positron plasma and antiproton bunch, as well as the ending frequency, chirp rate and amplitude of the autoresonant perturbation. To maximize the production rate of trappable antihydrogen, it is instructive to know which of the parameters the rate is most sensitive to, and what limit these parameters pose to the rate. Within the confines of the water bag – Vlasov model, one can predict the kinetic energy distribution of injected antiprotons as a function of the initial antiproton parameters; the results are relatively insensitive to positron parameters since the positron plasma is assumed to evolve according to the quasi–static water bag model. Their impact is most keenly seen

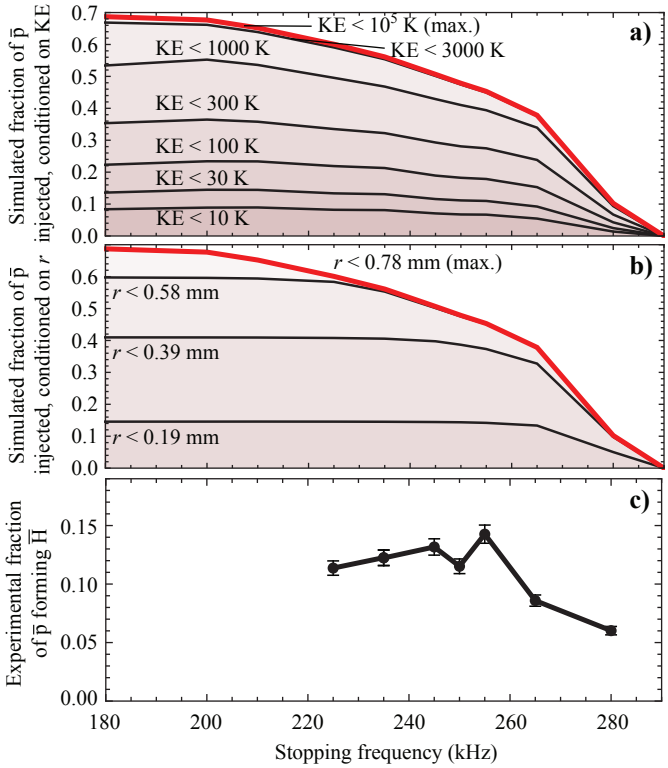

Figure 7.7: The performance of autoresonant injection against the sweep's stopping frequency, compared between simulation and experiment. a) The fraction of antiprotons injected into the positron plasma, conditioned on their kinetic energy in the positron plasma. b) Same as a, except the curves are conditioned on the radius. c) The number of antiprotons from experiment that successfully inject into the positron plasma and form antihydrogen, divided by the estimated initial number of antiprotons.

after the injection, in the second phase of the overall simulation. Still, some qualities of the injected antiprotons can be safely assumed to lead to better antihydrogen trapping. Here we assume the trappable antihydrogen come mainly from the low–energy portion of the injected antiprotons (defined as < 500 K; other definitions yield similar results), since the portion with significantly higher energy would have a small collision and recombination cross–section with positron. At best, these high–energy antiprotons have no impact on the number of trappable antihydrogen produced, and at worst they lead to the heating of the positron plasma and delay the recombination of the low–energy antiprotons, during which time they can heat up by equilibrating with positrons.

This motivates us to study the impact these antiproton and perturbation parameters have on the fraction of antiproton injected at below 500 K, which is shown in Fig. 7.8. Different antiproton bunches with various initial numbers and temperatures are fed into the model in the configuration shown in Fig. 2.6. At each initial number and temperature, various perturbations are applied. The optimal stopping frequency and strength, defined as that

which yields the highest fraction of antiprotons injected with energy < 500 K, is identified. The corresponding fraction of antiprotons injected below various energies is then plotted as a function of the initial antiproton number and temperature. (The starting frequency and chirp rate are kept constant at 325 kHz and -120 MHz/s respectively to keep the parameter space manageable.)

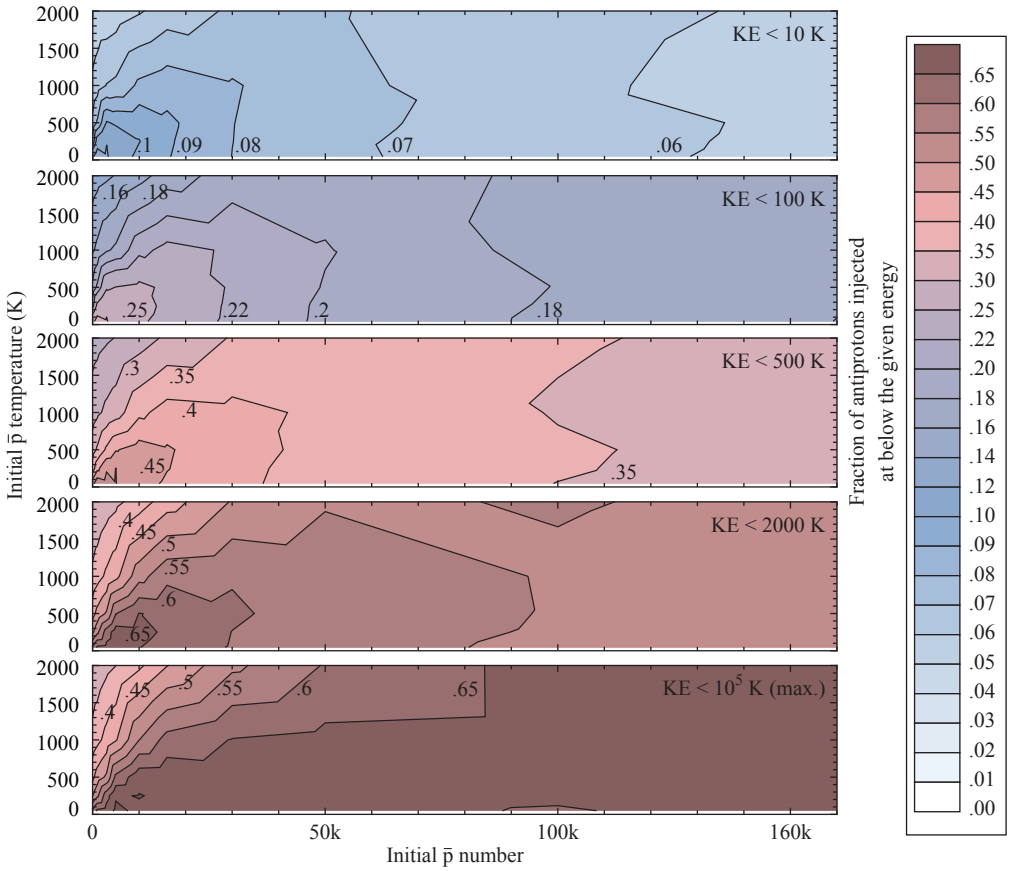

Figure 7.8: The performance of autoresonance injection against antiproton number and temperature, according to simulation. Each plot shows the fraction of antiprotons injected by a perturbation into the positron plasma with energy below the indicated value, for various initial antiproton number and temperature. Each antiproton bunch with a specific initial number and temperature is injected using the optimal perturbation that leads to the highest injection ratio at KE < 500 K — i.e., these contours reflect the best–case capability of a conventional autoresonant perturbation.

The self–field of the antiprotons tends to equalize the electric field felt by different parts of the bunch during the perturbation. The fraction of antiprotons that can be excited by the autoresonant perturbation is therefore a function of the density of the bunch [54], which is, in turn, determined by the initial number and temperature. This explains the drop in the injected fraction on the upper left corner in the plots in Fig. 7.8: the thermal spreads of these bunches are too great compared with their self–field to remain coherent during the

perturbation, and some parts of these bunches fail to be excited.

The low–energy injected fraction decreases as the antiproton number increases. This is caused by a mixing of the antiproton's self–field into the energy of the injected antiprotons, due to the electrostatic interaction between the main bulk of the positrons and antiprotons. The overall injected fraction does not decrease since the antiprotons are injected into the positron plasma at higher energies. This trend means the absolute number of antiprotons injected at low energies increases sub–linearly with an increasing initial number of antiprotons.

That the low–energy injected fraction does not significantly improve once the initial antiproton temperature reaches below ~ 250 K indicates the energy spread of the injected antiprotons is dominated by space charge effects rather than by their initial temperature, once the latter is below ~ 250 K. The autoresonant injection technique fails to make full use of the low temperatures of the initial bunch.

7.5 Incremental injection

Various schemes to overcome the limitations of the autoresonance injection have been investigated using the water bag – Vlasov model, and one of these ideas, the so–called incremental injection technique, offers some interesting injection characteristics. This type of scheme has been studied in ALPHA before [72, 73], but under different plasma conditions and with a somewhat different procedure. In this scheme, a antiproton bunch is positioned next to an positron plasma in the configuration shown in Fig. 2.6. An autoresonant perturbation, with an amplitude of 0.08V and starting at 325 kHz, is then applied on electrode E16 to excite the axial oscillation of the antiproton bunch, but is stopped before the whole bunch is injected. The voltage on E16 is subsequently decreased linearly to reduce the voltage separation between the positron and antiproton wells, until the rest of the antiproton bunch is entirely injected into the positron plasma. (The rate of the linear ramp is assumed to be slow enough that the positron plasma can redistribute radially through diffusion in case of evaporative escape.) The main tunable parameters in this scheme are the stopping frequency of the perturbation and the stopping voltage of the linear ramp. For a fixed stopping frequency, the optimal ramp depth — that which yields the highest injection ratio at KE < 10 K, a choice that will be justified later — is determined by running multiple simulations. This optimal ramp depth depends on the stopping frequency, but typically lies within -6 to -7 V. The resultant optimised injection statistics are shown in Fig. 7.9, for a stopping frequency between 250 and 325 kHz (with the upper limit corresponding to a zero–length perturbation). The sudden shift in injection behaviour at around 290 kHz is expected, since that is where the perturbation passes the linear resonance of the antiproton well (297.4 kHz) and starts to excite the antiproton bunch.

From Fig. 7.9 a, the total (unconditioned) injection fraction of antiproton is observed to increase as a longer perturbation is used. Figure 7.9 b further shows that the increase in

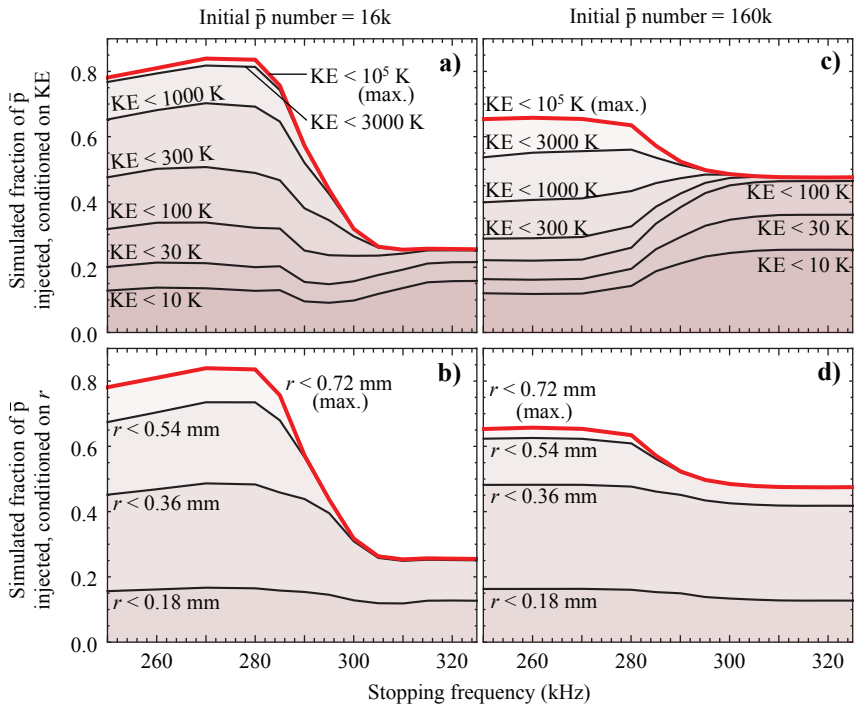

Figure 7.9: The performance of incremental injection against the sweep's stopping frequency, according to simulation. a) The fraction of antiprotons injected, out of an initial 16,000, conditioned on their kinetic energy in the positron plasma. b) Same as a, except that the ratios are conditioned on radius. c) Same as a, except that 160,000 antiprotons are used. (d) Same as c, except that the ratios are conditioned on radius.

the injected fraction comes from improved injection fractions at the outer radii. This can be explained as follows: when only a short chirp is used, the antiprotons reside, more or less, at the bottom of the antiproton well. The antiproton well becomes shallow as the electrode is ramped. However, this decreasing antiproton well depth is not constant across all radial shells. This is *not* due to the fall–off of the vacuum field (the antiproton bunch only has a radius of $\sim$ 0.8 mm, which is much smaller than the 22.3 mm radius of the electrode wall), but rather to the fall–off of the positron self–field. This fall–off causes the outer radial shells to have a higher antiproton well depth than the inner ones (see Fig. 7.10). When the antiproton well depth at the centre reaches zero (or, more accurately, reaches the level of the thermal spread of the antiprotons), the antiprotons on that shell start to inject, but the outer shells are still confined. Continuing the ramp further will not help inject the outer radii antiprotons. Instead, positrons start to evaporate since the left wall of the positrons (being also the antiproton well) now has a hole at small r. This self–adjusting process of the positrons causes the shape of the antiproton well to remain constant so long as there is still a significant remnant positron population. The autoresonant perturbation pre-excites the antiprotons at the outer radii, and helps them overcome the residual well by giving them more energy beforehand, thereby allowing their injection.

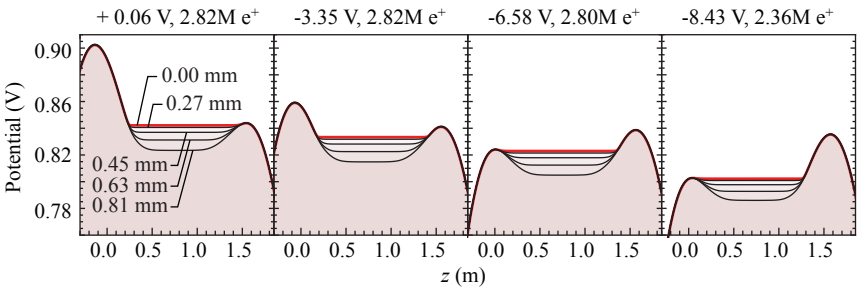

Figure 7.10: Four snapshots of the external potential seen by the antiprotons during incremental injection. The numbers displayed on top of each plot are the voltages applied on E16, and the number of remaining positrons (the rest being lost to evaporative escape). The curves in each plot display the potential at different radii, as indicated.

However, Fig. 7.9 a shows that this increase in the total injection fraction comes mostly from an increase in the fraction at high kinetic energy; the low energy injection fraction actually decreases, indicating a strong broadening in the energy distribution of the injected antiprotons due to the application of the autoresonant pre-excitation. This broadening becomes more pronounced with more antiprotons (see Fig. 7.9 c). A purely linear ramp without any pre–excitation (the rightmost limit in Fig. 7.9) is eventually going to produce more trappable antihydrogen than a ramp with pre–excitation, as the initial antiproton number increases. This is possible because the injected fraction at the outer radii increases with antiproton number. The enhanced space charge of the antiproton bunch fills the residual antiproton well at the outer radii, and causes the antiprotons on those radii to inject in higher numbers in a pure linear ramp. Using antiproton space charge to overcome the residual well rather than pre–excitation also prevents the broadening associated with it. Apart from these two techniques, one can also conceive of a scheme using a antiproton bunch with smaller radial size, so that the variation of the antiproton well depth among the radii of interest is reduced. A positron plasma with a larger radial size can also reduce the variation of the antiproton well depth for the radii with a significant antiproton population. Each of these possibilities has its experimental difficulty. In an antihydrogen production run the neutral trap is energised, the increased antiproton bunch size associated with higher particle numbers might lead to octupole–induced heating or particle loss [74]. An antiproton bunch with a smaller radius can be realised using a strong rotating wall compression, but there are technical and theoretical limits. A larger positron plasma is easily prepared experimentally, but the octupole–induced heating of the bunches is sensitive to their radial size. These are some of the technical issues that must be resolved to achieve a higher antihydrogen production rate.

Finally, Fig. 7.11 shows the performance of a pure linear ramp injection (without pre–excitation) when applied to antiprotons of various initial numbers and temperatures. The final ramp depth for each antiproton bunch is optimised to give the highest injection fraction at KE < 10 K. As argued above, the low–energy injection ratio improves with higher initial

antiproton number due to the space charge filling of the residual antiproton well. However, the converse is also true: low–energy injection performance deteriorates for a lower number of antiprotons. At below $\sim$ 20k antiprotons, the autoresonant injection technique offers a better performance than the pure linear ramp injection. One can also see from Fig. 7.11 that the energy distribution of the injected antiprotons is much "colder" than its autoresonance counterpart: there is hardly any difference between the fractions at KE $<$ 100 K and at KE $< 10^5$ K. Having a cold distribution of injected antiprotons should significantly improve the production rate of trappable antihydrogen, both in terms of more rapid recombination and minimal heating on the positrons. This is the motivation for the initial choice of looking at antiprotons injected at KE $<$ 10 K: we have demonstrated that it is possible to have both a higher total injected fraction for antiprotons, *and* a "colder" distribution of them.

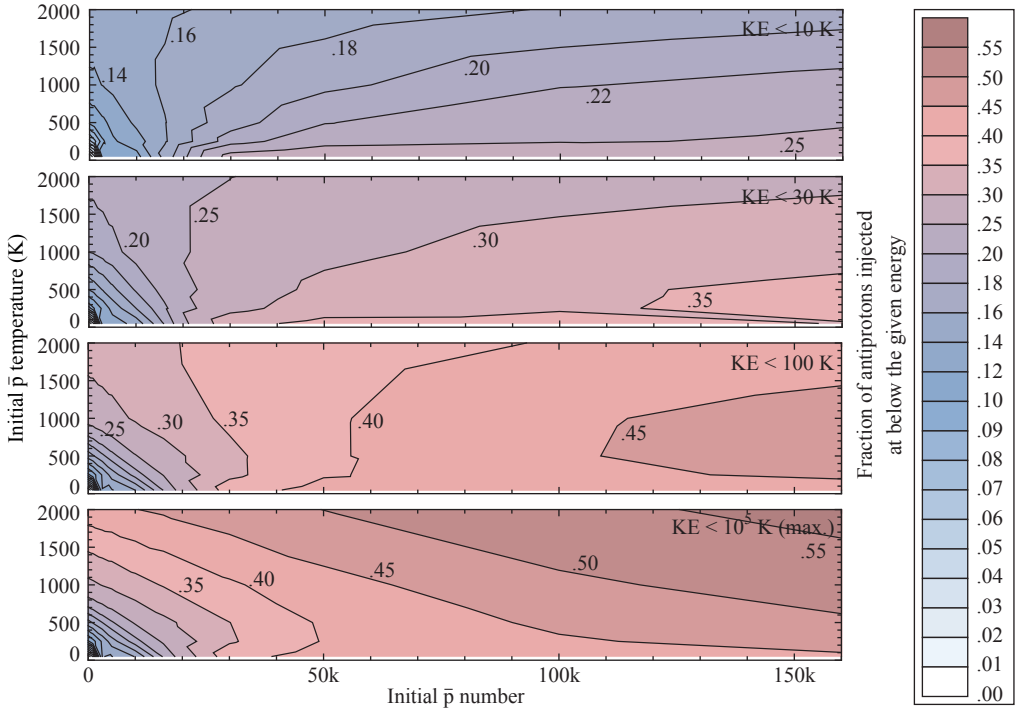

Figure 7.11: The performance of incremental injection against the number and temperature of the antiprotons, according to simulation. Each plot shows the fraction of antiprotons injected into the positron plasma with kinetic energy below the indicated value, for various initial antiproton number and temperature. Each antiproton bunch with a specific initial number and temperature is injected using the optimal ramp that leads to the highest injection ratio at KE $<$ 10 K.

Comparing the quantitative advantage, when using an autoresonant perturbation to inject a 250 K, 16k antiproton bunch into a positron plasma, $\sim$ 1.4k are injected at a KE $<$ 10 K, with an overall energy distribution fitted to a Maxwellian of 800 K. When using a pure linear ramp to inject the same bunch, $\sim$ 2.7k are injected at a KE $<$ 10 K, with an overall energy distribution fitted to 40 K. When using a pure linear ramp to inject a 250 K, 160k

antiproton bunch, $\sim$ 40k are injected at a KE $<$ 10 K, with an overall energy distribution fitted to 60 K.

Chapter 8

Conclusions

The ALPHA experiment has demonstrated the first trapping of antihydrogen atoms in 2010, and the first measurement of the hyperfine splitting of the antihydrogen spectrum in 2012, both major steps forward in the field of antihydrogen research. They provided the proof–of–concept that the design of the ALPHA apparatus, the manipulation techniques, the diagnostics and the detector system are suited to antihydrogen studies. We are, nonetheless, quite far from the ultimate goal of precision antihydrogen measurements, limited by the number and temperature of the trapped anti-atoms. If there exists any difference between the properties of hydrogen and antihydrogen, it is most likely present at a very small level. The spectral lines of the hydrogen atom has been measured to the 10^{-14} level of precision, and the measurement process requires access to numerous ultra-cold atoms. Replicating it for antihydrogen will be a highly challenging prospect. Experimentally, progress is being made with the commissioning of the ALPHA-2 apparatus, which enables more efficient catching and utilisation of antiprotons. The additional laser access and new neutral trap magnets are designed to allow more accurate laser and microwave spectral measurements. Beyond the hardware, however, progress is also needed in terms of the plasma manipulation techniques and the way they are developed, so that we can exploit the new apparatus as efficiently as possible. These developments are currently hampered by a lack of information on the state of plasmas, and the lack of tools to predict plasma behaviour. Currently, every time plasma conditions changes due to fluctuations in the particle sources (the AD and the positron accumulator) or improvement in the upstream processes in the apparatus, time–consuming and sub-optimal empirical tuning of downstream manipulation techniques are required.

Several numerical solvers have been developed to model a range of common plasma processes in the apparatus. These models lay the foundation for a more systematic analysis of the plasma phenomena in the apparatus, and their parallelised computational design allows some of them to have near–real–time usability in the experiment. The objective of their development is to improve our understanding of the plasma conditions in the trap (via enhanced diagnostic analysis), to predict the effects of manipulations on plasmas, and to speed up the development process of sequences. They can also help discover new techniques

which can qualitatively improve the synthesis process. The efficacy of these numerical tools will be tested during the commissioning and development of the ALPHA-2 apparatus.

The first model developed is a water bag solver, applicable to equilibrium plasmas at the zero–temperature limit. It iterates the boundary of a plasma to solve for a density distribution which yields the perfect Debye shielding expected of a zero–temperature plasma. This code has a numerical kernel Poisson solver, which allows the potential to be calculated for individual points, and maximises the scheme's performance. Given that a significant portion of the plasmas in the ALPHA experiment are at cyrogenic temperatures, this solver gives a good approximation of their equilibrium density and potential inside the Penning–Malmberg trap. A LabVIEW interface has been written for this solver which enables users to examine the confinement and geometry of plasmas given the plasma radial profile and the voltages of the electrodes. It provides an essential, and often used, functionality during sequence–programming when shallow wells for plasmas are being designed, for instance during the evaporative cooling of positrons.

The second model is a radially–coupled Vlasov–Poisson solver, applicable to dynamic plasmas with negligible radial transport. The solver decomposes the plasma into radially concentric shells, and evolves the z–v_z distributions in each shell using the Vlasov equation according to the electric field created by the plasma and the electrodes. The advection operators in the Vlasov equation are discretised using the flux balance method, in conjunction with the piecewise parabolic reconstruction method. The electric field is solved for using the numerical kernel method developed for the previous model. For optimal performance, a two–level parallelisation scheme using openMP and MPI is employed to give the code access to multiple cores in multiple server nodes. The simulation domain is dynamically adjusted in response to the phase space distribution to maximise computational efficiency. This model is suitable for simulating dynamic plasma processes with time scales near the axial bounce frequency, which includes a wide range of plasma manipulation techniques in ALPHA: species separation, mixing, temperature diagnostic, evaporative cooling, etc. In addition, the numerical annealing function embedded into the Vlasov–Poisson model allows it to solve for equilibrium plasma density at arbitrary temperatures. This generalises the zero–temperature water bag solver, and provides more accurate solutions to ill–confined or higher temperature plasmas. A LabVIEW interface has been developed for convenient access to this functionality during experimental operation.

The third model is an implicit, energy-conserving, azimuthally averaged Fokker–Planck solver for weakly magnetised collisions. It simulates the effect self–collisions have on the parallel and perpendicular velocity distributions of a plasma in which the magnetic effect during binary collisions is negligible. For the plasma parameters in the ALPHA apparatus, this corresponds to the self–collisions of antiprotons. The collision coefficients are computed by azimuthally averaging the derivatives of the Rosenbluth potentials. The solver is parallelised using openMP to address multiple cores on a single computer. This model is primarily devised to study the effects of antiproton–antiproton collision during the mixing and antihy-

drogen formation process. It is anticipated that the improved antiproton numbers available in ALPHA-2 may cause self–collisions to have a significant influence on the slow–travelling antiprotons in the mixture, from which trappable antihydrogen atoms form.

The fourth model generalises the previous weakly magnetised collision model. It simulates the effect of Fokker–Planck–type collisions between particles of different species, in magnetic fields of arbitrary strength. It reuses the discretisation scheme of the Fokker–Planck equation from the weakly magnetised model, but implements new collision coefficients appropriate for general collisions in arbitrary magnetic fields. Since analytic solutions to these general collisions are not known, the collision coefficients are calculated numerically using an adaptive Monte Carlo averaging algorithm. The algorithm evolves many pairs of colliding particles across all possible impact parameters and velocity phase angles. The sampling density in the impact parameter and phase angles space is dynamically adjusted, such that areas of higher bootstrapping error are sampled more densely. It then averages the change of the colliding particles' parallel and perpendicular velocities to yield the collision coefficients, as functions of the colliding particles' initial velocities. The collisions are simulated using an adaptive–time–stepping Boris pusher, which is parallelised on a GPU using the NVIDIA CUDA platform for maximum performance. This model is devised to simulate the effects of collisions in the antiproton–positron mixture that are not weakly magnetised — namely, those among positrons and between positrons and antiprotons. These collisions are expected to have an important effect on the velocity distribution of the antiprotons, and thus the antihydrogen atoms.

Two of these models are used in a detailed investigation of the antiproton–positron mixing process. The Vlasov model is used to simulate the autoresonant excitation of a standalone antiproton plasma, and the post–excitation energy distribution of the particles are compared to the results from existing solvers, analytic models and experiment measurements, yielding good agreement. The waterbag and the Vlasov–Poisson models are then combined to simulate the autoresonant axial excitation of the antiprotons adjacent to a positron plasma, a technique used to mix the two species and create antihydrogen. The excitation process directly influences the number of antihydrogen atoms trapped, and maximising the trapping efficiency requires minimising the axial energy of the antiprotons which cross into the positron plasma, among other dependencies that are not yet fully understood. The combined model is used to optimise the autoresonant perturbation for various plasma conditions, such that the antiprotons are injected at minimal energy. It is shown that the autoresonant technique becomes increasingly ineffective as more antiprotons are used in the mixing process. The model is then used to investigate novel mixing techniques, showing that in the high–antiproton intensity scenario, a slow and smooth merging of the antiproton and the positron well injects more antiprotons at low energies than the autoresonant excitation technique does.

Further work should extend the result of the antiproton excitation simulation using the collisional models to study the collisional and recombination interaction between the two

species. The ultimate goal is to develop a combination of models which provides a full prediction of the number of trapped antihydrogen given an excitation technique and a set of plasma conditions. Achieving this will serve two purposes: for a given plasma condition created in the experiment, the simulation can study and optimise excitation techniques; and for a given excitation technique that is known to be feasible experimentally, the simulation can predict the optimal plasma conditions for it. The latter informs the development of the plasma preparation sequence before mixing, which is important as many plasma manipulations trade off one desirable plasma characteristic for another. For instance, the rotating wall technique increases density while increasing the plasma temperature, and the evaporative cooling technique yields plasma of lower temperature, but with few numbers.

Future work should also include applying the models to simulating other processes, and improving the usability of these codes by developing graphical interfaces. Candidates processes suited to these models include the temperature diagnostic, axial species separation (by either electric pulse or resonant excitation), and sympathetic and evaporative cooling. For the temperature diagnostic, the Vlasov model can be used to simulate the slow ejection of particles and calculate the space charge effect on the inferred temperature. Detailed measurement and simulation may allow a more accurate analysis of the experimental particle escape timing, and yield a better estimate of the axial velocity distribution in the plasma. For the species separation and sympathetic and evaporative cooling, the simulation allows for numerical studies of different choices of particle numbers, timing and electrode voltages, identifying those that lead to improved separation or cooling.

Other numerical tools that may extend our ability to simulate important plasma processes in the ALPHA apparatus include three–dimensional particle–in–cell (PIC) codes and N–body codes. Their ability to simulate non-axisymmetric plasmas and radial transport is essential for processes like the rotating wall, diocotron instability and radial separation. Particle codes are more suited to these processes since distribution function–based algorithms (like the Vlasov equation) become computationally infeasible for higher–dimensional motion. The effects of non-Fokker–Planck collisions in strong magnetic fields, where v_z exchanging collisions are not negligible, should also be investigated.

Bibliography

[1] H. Kragh, *Quantum generations: a history of physics in the twentieth century* (Princeton University Press, 2002).

[2] P. A. M. Dirac, "The quantum theory of the electron", Proc. R. Soc. Lond. A **117**, 778 (1928).

[3] H. Quinn and Y. Nir, *The mystery of the missing antimatter* (Princeton University Press, 2008).

[4] C. D. Anderson, "The positive electron", Phys. Rev. **43**, 491 (1933).

[5] O. Chamberlain et al., "Observation of antiprotons", Phys. Rev. **100**, 947 (1955).

[6] J.Button et al., "Antineutron production by charge exchange", Phys. Rev. **108**, 1557 (1957).

[7] S. Dodelson, *Modern cosmology* (Academic Press, 2003).

[8] T. Yoshida et al., "BESS–polar experiment", Adv. Space Res. **33**, 1755 (2004).

[9] A. D. Sakharov, "Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe", J. Exp. Theor. Phys. **5**, 24–27 (1967).

[10] C. S. Wu et al., "Experimental test of parity conservation in beta decay", Phys. Rev. **105**, 1413 (1957).

[11] J. Christenson et al., "Evidence for the 2π decay of the K_2^0 meson", Phys. Rev. Lett. **13**, 138 (1964).

[12] A. Alavi-Harati et al., "Observation of direct CP violation in $K_{S,L} \to \pi\pi$ decays", Phys. Rev. Lett. **83**, 22 (1999).

[13] V. Fanti et al., "A new measurement of direct CP violation in two pion decays of the neutral kaon", Phys. Lett. B **465**, 335 (1999).

[14] B. Aubert et al., "Measurement of CP-violating asymmetries in B^0 decays to CP eigenstates", Phys. Rev. Lett. **86**, 2515 (2001).

[15] K. Abe et al., "Observation of large CP violation in the neutral B meson system", Phys. Rev. Lett. **87**, 091802 (2001).

[16] A. Carbone, "A search for time-integrated CP violation in $D^0 \to h^- h^+$ decays", (2012).

[17] S. Hoogerheide, "Trapped positrons for high precision magnetic moment measurements", PhD thesis (Havard University, 2013).

[18] G. Gabrielse et al., "Precision mass spectroscopy of the antiproton and proton using simultaneously trapped particles", Phys. Rev. Lett. **82**, 3198 (1999).

[19] S. H. Geer and D. C. Kennedy, "A new limit on the antiproton lifetime", Astrophys. J. **532**, 648 (2000).

[20] J. DiSciacca et al., "One-particle measurement of the antiproton magnetic moment", Phys. Rev. Lett. **110**, 130801 (2013).

[21] P. J. Mohr et al., "Codata recommended values of the fundamental physical constants: 2006", (2007).

[22] C. Amole et al., "Resonant quantum transitions in trapped antihydrogen atoms", Nature **483**, 439 (2012).

[23] C. Amole et al., "Autoresonant-spectrometric determination of the residual gas composition in the ALPHA experiment apparatus", Nat. Comm. **4**, 1785 (2013).

[24] S. Aghion et al., "A moiré deflectometer for antimatter", Nat. Comm. **5**, 4538 (2014).

[25] P. Debu, "Gbar", Hyperfine Interactions **212**, 51–59 (2012).

[26] C. Amole et al., "An experimental limit on the charge of antihydrogen", Nat. Comm. **5**, 3955 (2014).

[27] G. Baur et al., "Production of antihydrogen", Phys. Lett. B **368**, 251 (1996).

[28] M. Amoretti et al., "Production and detection of cold antihydrogen atoms", Nature **419**, 456 (2002).

[29] G. Gabrielse et al., "Driven production of cold antihydrogen and the first measured distribution of antihydrogen states", Phys. Rev. Lett. **89**, 233401 (2002).

[30] F. Wysocki et al., "Accumulation and storage of low energy positrons", Hyperfine Interactions **44**, 185 (1988).

[31] D. J. Griffiths, *Introduction to electrodynamics*, 3rd (Prentice Hall, 1999).

[32] B. M. Jelenkovic et al., "Sympathetically laser–cooled positrons", Nucl. Instr. Meth. in Phys. Res. B **192**, 117 (2002).

[33] P. M. Bellan, *Fundamentals of plasma physics* (Cambridge University Press, 2004).

[34] L. Friedland, "Autoresonance in nonlinear systems", Scholarpedia **4**, 5473 (2009).

[35] E. Butler, "Antihydrogen formation, dynamics and trapping", PhD thesis (Swansea University, 2011).

[36] A. V. Dudarev et al., "Superconducting magnet for non-neutral plasma research", in Magnet technology conference – 15, beijing (1997).

[37] G. B. Andersen et al., "Antihydrogen annihilation reconstruction with the ALPHA silicon detector", Nucl. Instrum. Meth. A **684**, 73–81 (2012).

[38] W. Bertsche et al., "A magnetic trap for antihydrogen confinement", Nucl. Instr. Meth. Phys. Res. A **556**, 746 (2006).

[39] D. E. Pritchard, "Cooling neutral atoms in a magnetic trap for precision spectroscopy", Phys. Rev. Lett. **51**, 1336 (1983).

[40] D. L. Eggleston et al., "Parallel energy analyzer for pure electron plasma devices", Phys. Fluids B **4**, 3432–3439 (1992).

[41] G. B. Andersen et al., "Trapped antihydrogen", Nature **468**, 673 (2010).

[42] G. B. Andersen et al., "Confinement of antihydrogen for 1,000 seconds", Nat. Phys. **7**, 558 (2011).

[43] D. S. Hall and G. Gabrielse, "Electron cooling of protons in a nested penning trap", Phys. Rev. Lett. **77**, 1962–1965 (1996).

[44] G. B. Andersen et al., "Centrifugal separation and equilibration dynamics in an electron-antiproton plasma", Phys. Rev. Lett. **106**, 145001 (2011).

[45] G. B. Andersen et al., "Evaporative cooling of antiprotons to cryogenic temperatures", Phys. Rev. Lett. **105**, 013003 (2010).

[46] G. B. Andersen et al., "Autoresonant excitation of antiproton plasmas", Phys. Rev. Lett. **106**, 025002 (2011).

[47] J. Fajans et al., "Critical loss radius in a penning trap subject to multipole fields", Phys. Plasmas **15**, 032108 (2008).

[48] G. B. Andersen et al., "Magnetic multipole induced zero-rotation frequency bounce-resonant loss in a Penning-Malmberg trap used for antihydrogen trapping", Phys. Plasmas **16**, 100702 (2009).

[49] X.-P. Huang and other, "Steady–state confinement of non-neutral plasma by rotating electric fields", Phys. Rev. Lett. **78**, 875 (1997).

[50] G. B. Andersen et al., "Compression of antiproton clouds for antihydrogen trapping", Phys. Rev. Lett. **100**, 203401 (2008).

[51] C. Amole et al., "Discriminating between antihydrogen and mirror-trapped antiprotons in a minimum-B trap", New J. Phys **14**, 105010 (2012).

[52] J. D. Jackson, *Classical electrodynamics*, 3rd (John Weiley and Sons, Inc., 1999).

[53] W. H. Press et al., *Numerical recipes*, 3rd (Cambridge University Press, 2007).

[54] I. Barth et al., "Autoresonant transition in the presence of noise and self-fields", **103**, 155001 (2009).

[55] F. Filbet et al., "Conservative numerical schemes for the Vlasov equation", J. Compu. Phys. **172**, 166 –187 (2001).

[56] E. Fijalkow, "A numerical solution to the Vlasov equation", Compu. Phys. Comm. **116**, 319 –328 (1999).

[57] A. M. Harten and S. Osher, "Uniformly high-order accurate nonoscillatory schemes. I", SIAM J. Numer. Anal. **24**, 279–309 (1987).

[58] P. Colella and P. R. Woodward, "The piecewise parabolic method (PPM) for gas-dynamical simulations", J. Compu. Phys. **54**, 174 –201 (1984).

[59] J.-P. Berrut and L. N. Trefethen, "Barycentric Lagrange interpolation", SIAM Rev. **46**, 501 –517 (2004).

[60] J. Moore, *Physics 212: statistical mechanics II lecture IV*, (2006) `http://socrates.berkeley.edu/~jemoore/Moore_group,_UC_Berkeley/Physics_212_files/phys212ln4.pdf`.

[61] K. Huang, *Statistical mechanics* (John Wiley and Sons, Inc., 1987).

[62] D. C. Montgomery and D. A. Tidman, *Plasma kinetic theory* (McGraw–Hill Book Company, 1964).

[63] M. N. Rosenbluth et al., "Fokker–Planck equation for an inverse–square force", Phys. Rev. Lett. **107**, 1 (1957).

[64] B. C. Carlson, "A table of elliptic integrals of the second kind", Math. Compu. **49**, 595–606 (1987).

[65] L. Chacón et al., "An implicit energy–conservative 2D Fokker–Planck algorithm I. difference scheme", J. Compu. Phys. **157**, 618–653 (2000).

[66] S. Ichimaru and M. N. Rosenbluth, "Relaxation process in plasmas with magnetic field. temperature relaxations", Phys. Fluids **13**, 11 (1970).

[67] T. M. O'Neil, "Collision operator for a strongly magnetized pure electron plasma", Phys. Fluids **26**, 8 (1983).

[68] M. E. Glinsky et al., "Collisional equipartition rate for a magnetized pure electron plasma", Phys. Fluids B **4**, 1156 (1992).

[69] J. P. Boris, "Relativistic plasma simulation — optimization of a hybrid code", in Fourth conference on the numerical simulation of plasma (Naval Research Laboratory, 1970), pp. 3–67.

[70] C. Amole et al., "Experimental and computational study of the injection of antiprotons into a positron plasma for antihydrogen production", Phys. Plasmas **20**, 043510 (2013).

[71] J. Fajans and L. Friedland, "Autoresonant (non–stationary) excitation of a pendulum, plutinos, plasmas and other nonlinear oscillators", **69**, 1096 (2001).

[72] N. Madsen, "Antihydrogen from merged plasmas — cold enough to trap?", AIP Conf. Proc. **862**, 164–175 (2006).

[73] G. B. Andresen et al., "Antihydrogen formation dynamics in a multipolar neutral anti-atom trap", Phys. Lett. B **685**, 141–145 (2010).

[74] G. B. Andresen et al., "Magnetic multiple induced zero-rotation frequency bounce-resonant loss in a penning-malmberg trap used for antihydrogen trapping", Phys. Plasmas **16**, 100702 (2009).