Title
Phenomenology of creation of antihydrogen and measurement of antihydrogen properties

Author
Evans, Len Takahashi

Publication Date
2016

Peer reviewed|Thesis/dissertation

Phenomenology of creation of antihydrogen and measurement of antihydrogen properties

by

Len Takahashi Evans

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

and the Designated Emphasis

in

Computational and Data Science and Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jonathan S. Wurtele, Chair
Professor Joel Fajans
Professor John Strain

Fall 2016

Phenomenology of creation of antihydrogen and measurement of antihydrogen properties

Abstract

Phenomenology of creation of antihydrogen and measurement of antihydrogen properties

by

Len Takahashi Evans

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Jonathan S. Wurtele, Chair

This dissertation focuses on three ideas useful to the nonneutral plasma experiment at UC Berkeley and the ALPHA experiment at CERN. While these may seem like disparate ideas, in all cases, a careful mathematical treatment of the problems yield useful insights.

First, we present improvements to the analysis for diagnostics of temperature and density of plasmas in Penning-Malmberg trap experiments. Our new methods are faster and more accurate than previously used methods of analysis. This allows us to conduct these diagnostics in real-time without any input from a human.

We then theoretically consider the problem of enhanced cooling of an electron plasma from a coupling of the plasma to cavity modes. We make more rigorous the previous analysis that was done on the topic, and then extend these results to a longitudinally dynamic plasma. We compare our theoretical results with experimental observations from the Berkeley plasma trap and with simulations and find good agreement.

Finally, we consider analyses of ALPHA data to measure antihydrogen properties. We place a statistically rigorous bound on the antihydrogen charge from ALPHA data. This improves the previous bound on antihydrogen charge by a factor of 20, and assuming superposition, improves the bound on the positron charge anomaly by a factor of 25. Finally, we consider analyses for a future measurement of the gravitational mass of antihydrogen. We find that with reasonable constraints, a measurement of the gravitational mass of antihydrogen with a precision of 1% should be feasible.

To Aaron, Albert, Eric, Joanna, and Maribel

Contents

List of Figures

List of Tables

Acknowledgments

While I am not optimistic, I hope that in the future the climate in graduate schools for leaving academia looks better. The numbers simply do not work out for all (or even half) of graduate students to end up in an academic profession, so it is ridiculous that this is the expectation of all students. It is doubtful that future prospects will look any better; my research group already does not have funding and it is hard to believe the this will look better in the future. I hope that it can become more common in the future for students to take on industry internships over summers. I wish departments (and the professors in them) would make more efforts to highlight perspective industry careers instead of relying on student-led efforts. Students should feel free to take classes outside of their field without worrying that their adviser will look down upon them. Qualifying exams could be an excellent way to get interview practice if they were run optimally.

Regardless, I'm glad to have worked with excellent people during my time at Berkeley. Eric Hunter and Alex Povilus are extremely bright, motivated physicists and I believe both will be extremely succesful in the world of physics. Often the theory and experimental sides of our group were disjoint, but I feel Eric and I did a great job in bridging that gap. Andrey Zhmoginov was a great mentor and was always willing to talk about my projects even when he was extremely busy. I am happy that he has found a research position at Google where his full potential can be utilized. Andrew Charman is one of the brightest people I have ever met, and I am glad to have developed my statistical understanding through interactions with him.

One of the most rewarding experiences I had while at Berkeley was to be the head GSI for physics 7B for six semesters. I appreciate the work Cathy Bordel, Eric Corsini, Achilles Speliotopoulos, and Jonathan Wurtele all put into teaching their course and being open to the various ideas I had. Ultimately, I felt we were able to run the course more efficiently while having the students learn more. I would like to thank all the GSIs I worked with for being motivated and responsible and making the course run smoothly.

I would like to thank everyone who helped me have a smooth transition from physics into data science. Four years ago, I had never heard of data science, but through the career development initiative for the physical sciences (CDIPS) created by Gabriel Dunn, Caroline Soffiati and Joseph Thurakal, I realized that data science was the right place for me to be. It was a rewarding experience for me to work with CDIPS and bring data scientists to give talks on campus as well as organize data science workshops to help train the next generation of data scientists.

I would like to thank Kat Glaeser for being my mentor during the CDIPS data science workshop and sparking my initial interest in data science. Without her I am not sure I would have realized that data science was a feasible profession for me. I would like to thank Andreas Weigend, who despite his fame was willing to give a talk at our workshop for no cost, and even going so far as to cancel previous obligations to be at the workshop. I have enjoyed learning about data from a master and am glad that Andreas is now a personal friend. Finally, I'd like to thank the fraud team at Uber, Andre Bach, Ting Chen, Andrew

Duberstein, Vicente Silveira, and in particular, Karthik Ramasamy, for having me as a part of their team for a summer and helping me learn useful industry skills. I am glad they took the risk of bringing a physicist on board despite not the traditional training for data science. I look forward to working with them in the future.

Finally, I'd like to thank all my friends that I met at Berkeley. I am so glad Joanna Szornel found me in my office hours as she has become a close friend. The last four and a half years have had their ups and downs and I'm glad we were there for each other to support and celebrate these occasions. I will miss the lunchtime discussions about politics, video games, and grad school with Aaron Szasz and Eric Dodds, but I hope we will continue to get together every weekend to enjoy a great meal and boardgames. I hope that they will be following shortly in writing their theses. It was a great experience sharing an office with Albert Yuen. Albert's optimism for physics and academia as well as our conversations about data and life helped me enjoyed grad school more. Finally, I would like to thank Maribel Sierra. I am so glad that things were somehow able to work out between us. Thank you for listening to all my quibbles and celebrating my accomplishments. I relish all the adventures we have and hope for many more in the future.

Chapter 1

Introduction

Believe nothing from this book.
Except what you know to be true.
Test the knowledge.
Find your truth.

Liber Primus

The standard model of particle physics, based on quantum field theory, has been extremely successful in describing interactions of particles. With the recent discovery of the Higgs boson, all of the fundamental[1] theoretical predictions of the standard model have been discovered [8, 9]. Further, almost all measurement of the properties of these particles show excellent agreement with the standard model[2,3].

However, the standard model is not without issues. It suffers from a fine-tuning problem (the hierarchy problem) in that if there are energy scales relevant to particle physics larger than the weak scale (such as the gravitational scale), then it is unnatural for the Higgs to have the observed value of mass. Further, the theory has many free parameters that must be empirically determined, and some quantities, such as the CP violating phase, are extremely small without a known mechanism to make the values small [13]. Various cosmological observations are well-explained by dark energy and dark matter, yet the standard model does not include dark matter and has a scalar vacuum expectation value much too small to account for dark energy [14]. In addition, the observed abundance of matter over antimatter is not predicted by the standard model and accepted ideas of cosmology. Despite numerous efforts, quantum field theory has also been difficult to incorporate with general relativity [13].

[1] We say fundamental as there are composite mesons and baryons predicted that have not been discovered.

[2] Two measurements that show discrepancy at $> 3\sigma$ level are the muon magnetic moment measurements [10] and the branching ratio of the $\bar{B} \to D^{(*)}\tau\bar{\nu}_\tau$ decay [11].

[3] Neutrino masses generated through a Higgs mechanism are a straightforward addition to the standard model, so here they are considered part of the standard model. However, because of their small values, there are proposals that the masses are generated through other mechanisms such as the seesaw mechanism, which would be physics beyond the standard model [12].

Although there are many theories that attempt to solve these issues, their veracity can only be verified by careful experimental measurements. Collider experiments explore the high-energy regime. These experiments test the standard model through measurement of the properties of known particles and search for new particles. In the low-energy regime, properties of antimatter systems are carefully measured and compared to the theoretical expectations. These measurements constrain extensions of the standard model and any observed deviations would necessarily have to be incorporated into the more general theory.

Antihydrogen is an ideal test of quantum field theories as it is the simplest purely antimatter system, making it amenable to highly accurate theoretical predictions [15]. Further, hydrogen properties have been measured to great precision, and thus any deviation measured in antihydrogen will be readily apparent [7]. Antihydrogen is expected to have the same charge as hydrogen, which has charge $< 10^{-21}e$. However, unlike other uncharged antimatter such as antineutrons, antihydrogen has a significant magnetic moment due to its positron. This means that it can be trapped with magnetic field gradients achievable with current technology. The expected lack of charge makes antihydrogen apt for gravity studies, as stray electric and magnetic force will overpower the gravitational force for any charged particle [16]. An advantage of studies with antimatter in general are that they create a readily measured signal when they interact with normal matter (annihilate). As a result, while it may be hard to determine whether a antihydrogen atom *is* trapped, an experiment can determine with great certainty whether an antihydrogen atom *was* trapped. The only limitation to this method is the background from cosmic rays, which must be distinguished from an antihydrogen signal.

Thus, experimental tests of antihydrogen properties test fundamentals of quantum field theory and general relativity. Observed differences may offer key insights to guide more complete theories of particle interactions and ultimately quantum gravity [17, 18].

1.1 The CPT Theorem

The CPT theorem states that under the combined operation of charge conjugation (C), parity transformation (P), and time inversion (T), the standard model lagrangian is invariant. This feature is actually more general and any Lorentz invariant, local quantum field theory with a hermitian hamiltonian is CPT invariant [19].

Under CPT, four-vectors are reversed and particles are replaced with antiparticles of the same spin orientation [20]. Thus, we can interpret the CPT operation as replacing the particle equations of motion with the antiparticle equations of motion. Hence, if CPT is a good symmetry, the properties of hydrogen and antihydrogen (and in general matter systems and their corresponding anti-matter systems) should be equivalent [21].

To see this more concretely, consider the lagrangian of a charged particle interacting with an electromagnetic field [15]. This could be, for example, an electron interacting with the electromagnetic field produced by a proton in a hydrogen atom.

$$\mathcal{L}_{QED} = \bar{\psi}\gamma^{\mu}(i\partial_{\mu} - eA_{\mu} - eB_{\mu})\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.1}$$

Here, ψ denotes a fermion field and $\bar{\psi}$ its Dirac adjoint [15]. A^{μ} is the electromagnetic potential of the electron itself, whereas B^{μ} is the external electromagnetic potential. $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ is the electromagnetic field tensor. The Dirac matrices γ^{μ}, are a set of 4×4 matrices that are a representation of a Clifford algebra with three space dimensions and one time dimension. The fermion bilinear $\bar{\psi}\gamma^{\mu}\psi$ is the probability current of the fermion. Just as a regular current, this is odd under CPT, as it is a four-vector. The partial derivative and electromagnetic potentials are also odd under PT, and unchanged by C, so the first term of the lagrangian is invariant under CPT. Both of the tensors in the second term are CPT even, so that term is invariant under CPT. Hence, we expect properties of antihydrogen determined by electromagnetic forces (such as its spectrum) to be identical because of the CPT theorem.

The idea behind the connection between Lorentz invariance and CPT is given by the following argument [22]. If a hamiltonian of a theory is Lorentz invariant, then all tensors should have their indices contracted with other indices. As a result, even if four-vectors (or odd-rank tensors) change sign under CPT, the net effect is that any Lorentz invariant term will not change. This is the PT theorem in classical field theories [23]. While charge conjugation may appear fundamentally different from the space-time symmetries, the Feynman-Stückelberg interpretation, equating anti-particles as negative energy modes of the antiparticle with time reversed, offers the connection [15]. A *quantum* field theory necessarily has particles that under PT are transformed into their antiparticles under this interpretation. Thus, charge conjugation is necessary to fully leave the expressions invariant. The natural extension of the classical PT theorem to a quantum theory is the CPT theorem [24].

Experimental verification of the CPT theorem tests foundational axioms of quantum field theories and physics in general. Any violation found challenges quantum field theory at a fundamental level and is useful in motivating more general theories that solve some of the issues mentioned in the previous section.

1.2 Antihydrogen

CERN's antiproton decelerator (AD) facility houses numerous experiments to measure properties of antimatter and test the CPT theorem and other predictions of quantum field theory [25]. ALPHA[4] is one of these experiments that studies antihydrogen. ALPHA has made measurements of antihydrogen's spectrum [26], mass [18], and charge [27, 28].

Antihydrogen is dominantly produced in ALPHA through the recombination process $e^+ + e^+ + \bar{p} \rightarrow e^+ + \bar{H}^*$ (where $\bar{H}^*$ denotes antihydrogen in an excited state). Antihydrogen could also be produced in the process $e^+ + \bar{p} \rightarrow \bar{H}^* + \gamma$ [29] . While initially it may seem

[4]ALPHA formerly stood for antihydrogen laser physics apparatus, but has since abandoned this acronym.

like this would be the dominant process since it is a two body collision, the time to radiate a photon is slow compared to a collision time and hence the process happens quite slowly.

The mixing is achieved by trapping positrons and antiprotons in neighboring traps potential wells. The antiprotons are then provided with energy so that they spill over into the positron well. The antiprotons cool as they interact, via two-body collisions, with the positrons, and eventually recombine into antihydrogen. This is a non-equilibrium process in which simulations are necessary to model and optimize the procedure. For positron temperatures that ALPHA achieves, the reaction rate of the three body process is fast enough that all the antiprotons will be converted to antihydrogen during the mixing process.

The magnetic moment of ground state antihydrogen is mostly determined by the positron spin [16]. With trapping fields of order 1 T, only antihydrogen with temperatures $\lesssim 1$ K can be trapped. The antihydrogen energy will come mostly from the antiproton energy. As the two body collision rate of antiprotons and positrons is faster than the three body antihydrogen recombination, the antiprotons will thermalize with the positrons before forming antihydrogen [30]. Thus, cold positron plasmas will lead to cold antihydrogen, which ultimately leads to more antihydrogen atoms trapped. Currently, the trapping rate is limited to ~ 20 per experimental cycle ($\sim$100s) because cold plasma temperatures are not achievable.

We present results in this thesis on the measurement of antihydrogen charge from 24 antiatoms. While these bounds are competitive with measurements made on other systems [7], going forward, more particles will be necessary for more precise measurements. In this thesis, we discuss methods to achieve colder lepton plasma temperatures toward this goal.

1.3 Description of Chapters

Chap. 2 provides an overview of Penning-Malmberg traps, and in particular will highlight features of ALPHA and the plasma trap at UC Berkeley. The ALPHA trap is used to create antihydrogen and measure antihydrogen properties, while the Berkeley plasma trap is used to study the cooling process of lepton plasmas.

Chap. 3 discusses fitting of curves to a nonlinear functional form and the Levenberg-Marquardt algorithm for achieving this. We discuss our fast implementation of the Levenberg-Marquardt algorithm on graphical processing units (GPUs). This algorithm is an integral part of the analysis of data for the plasma diagnostics of Chap. 4 and 5.

Chap. 4 uses the methods of Chap. 3 to analyze field line integrated images of plasmas extracted from a Penning-Malmberg trap. The images can be used to estimate the density and length of a plasma. Our GPU code to analyze the images runs in a few seconds, compared to the minutes that a previous, non-GPU implementation took. Our new method yields increased accuracy and precision in the fits.

Chap. 5 describes the parallel temperature diagnostic of plasmas. We prove how low a temperature can feasibly be measured using the diagnostic. We then discuss our implementation of two different approaches to the problem. These approaches are designed to run in

a few seconds and be fully automated such that plasma temperatures are available with no human input.

In Chap. 6 we theoretically consider the cooling process of plasmas in the Berkeley plasma trap. While some of the initial results from the experiment were puzzling [5], we find that a careful analysis analysis is able to elucidate many of the results. These new cooling techniques will eventually be incorporated into ALPHA to improve the overall trapping rate of antihydrogen.

Chap. 7 of this thesis describes an experiment to bound the charge of antihydrogen using only 24 atoms to be $< 7 \times 10^{-9} e$ [28]. Assuming superposition, this constrains the charge difference of the positron and electron than any other measurement [7].

While the current best bound on the antihydrogen mass is large, we find in Chap. 8 evidence that a 1% measurement should be feasible in the future with a new trap designed for gravity studies with antihydrogen.

Chapter 2

The Berkeley Plasma Trap and ALPHA

Penning-Malmberg traps are used at CERN and Berkeley to capture and manipulate charged particles [31]. We start by describing the Penning-Malmberg trap and present results that will be referred to in the rest of the thesis. We then describe the Berkeley plasma trap and ALPHA. In particular, we highlight how these apparatuses deviate from a standard Penning-Malmberg trap. Typical parameter ranges for lepton plasmas in the ALPHA experiment and the Berkeley plasma trap are shown in Table 2.1.

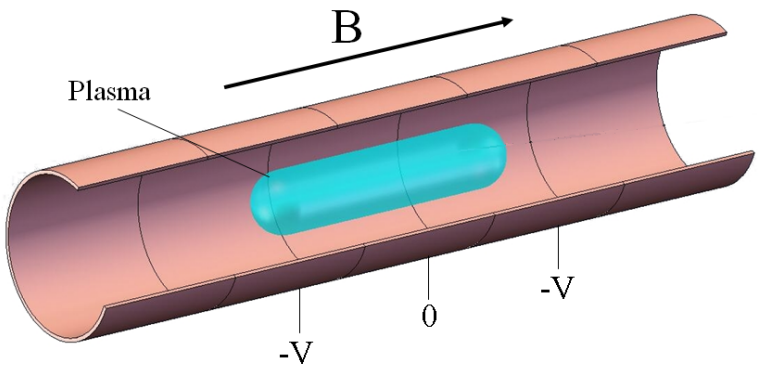

Figure 2.1: A schematic of a Penning Trap. The plasma is confined radially by a strong magnetic field. Electrodes are biased to trap the plasma axially. Image by A.P. Povilus [1].

Parameter	Typical Values (SI)	Typical Values (Energy units)
Number	$10^4 - 10^7$	$10^4 - 10^7$
Density (n_0)	$10^6 - 10^8$ cm^{-3}	$8 \times 10^{-9} - 8 \times 10^{-7}$ eV^3
Temperature (T)	$10 - 10^4$ K	$10^{-3} - 1$ eV
Length (l_p)	$1 - 10$ cm	$10^3 - 10^5$ eV^{-1}
Radius (R_p)	$1 - 5$ mm	$5 \times 10^3 - 2.5 \times 10^5$ eV^{-1}
Cyclotron frequency (ω_c)	$10 - 10^3$ GHz	$10^{-5} - 10^{-3}$ eV
Bounce frequency (ω_z)	$30 - 300$ MHz	$2 \times 10^{-8} - 10^{-7}$ eV

Table 2.1: Typical parameter ranges for lepton plasmas in ALPHA and the Berkeley plasma trap.

2.1 Penning-Malmberg Traps

The Penning-Malmberg trap is used to confine and manipulate plasmas. The basic configuration of the trap is a series of cylindrical electrodes that is surrounded by a solenoid magnet, as shown in Fig. 2.1.

The solenoid provides a strong magnetic field, and because charged particles will be tied to field lines (see Appendix A), the magnetic field provides confinement of the charged particles in the radial direction. To see this, consider the canonical angular momentum, which is conserved assuming azimuthal symmetry in the trap. Consider a plasma with $\mathcal{N}$ particles, and index each particle by a label i. Particle i has angular velocity $v_{\theta,i}$ and radial position r_i. In a strong B-field,

$$P_\theta = \sum_{i=1}^{\mathcal{N}} \left(m r_i v_{\theta,i} + \frac{qBr_i^2}{2} \right) \approx \frac{qB}{2} \sum_{i=1}^{\mathcal{N}} r_i^2. \tag{2.1}$$

Thus, the plasma radius is constant. Without collisions, a plasma particle undergoes a cyclotron motion at frequency $\omega_c = |q|B/m$ in a tight radius around its field line indefinitely. Plasma collisions will cause charged particles to jump between field lines, but this will not affect the overall plasma radius [32].

A Penning-Malmberg trap must be evacuated in order to have good radial confinement. Neutral particles in the trap are not affected by the electromagnetic fields in the trap. Thus, the neutral particles will not be azimuthally symmetric about the magnetic field axis. This means that collisions of plasma particles with neutral particles can lead to a net torque on the system, which will lead to the canonical momentum no longer being conserved. The argument given above is then invalid, and the plasma can expand. Other effects such as misalignment between the magnetic field and the electric potential will similarly lead to azimuthal symmetry breaking and hence lead to expansion of the plasma. These factors are controlled such that confinement of plasmas is possible for minutes in ALPHA and in the Berkeley plasma trap. This is long enough for the experiments conducted on the plasmas.

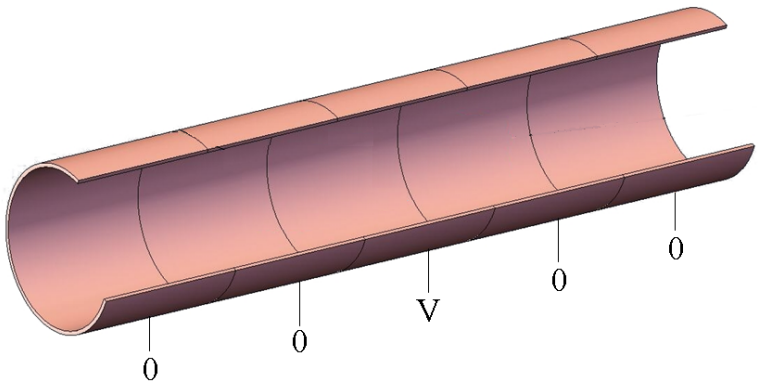

Figure 2.2: The potential everywhere inside the electrode configuration shown here is described by Eq. 2.2. For this expression, we assume there are additional grounded electrodes extending beyond the ones shown.

The electrodes in the trap can be biased to provide an electric potential well that traps the particles axially (that is, along the magnetic field lines). This will cause the particles to *bounce* between the turning points of the potential. The frequency of the bounce will be referred to as ω_z in this thesis. There is also a slow azimuthal $E \times B$ drift (see Appendix A for details) of each plasma particle from the confining field and the plasma self field. This will cause the plasma to rotate on axis, while keeping its shape radially. For parameters of interest to CERN and Berkeley, $\omega_c \gg \omega_z \gg \omega_{E\times B}$ (see Table 2.1).

Another advantage of the electrodes is that altering the electric potentials allows for control of the plasma. In particular, the diagnostics discussed later are destructive techniques where the plasma confining potential is lowered to let particles escape in a controlled manner.

Mathematical Description

Consider one electrode of the Penning-Malmberg trap with radius R_w and length L. As in Fig. 2.2, we bias the electrode to a potential V and ground all neighboring electrodes. We assume the neighboring electrodes extend indefinitely. The potential everywhere inside the trap is [3, 30]

$$V(r,z) = \frac{2V}{\pi}\int_0^\infty dk \frac{I_0(2kr/L)}{I_0(2kR_w/L)}\frac{\sin(k)}{k}\cos\left(\frac{2kz}{L}\right). \tag{2.2}$$

Now consider an electrode that is unbiased compared to its neighboring grounded electrodes in order to trap particles, as in Fig. 2.1. Near the center of this electrode, the potential is approximately harmonic, with

$$V(r,z) \approx V\left(S_{-1}\left(\frac{R_w}{L}\right) + \left(\frac{r^2}{L^2} - 2\frac{z^2}{L^2}\right) S_1\left(\frac{R_w}{L}\right)\right), \tag{2.3}$$

where

$$S_n(x) = \frac{2}{\pi}\int_0^\infty dk\ k^n \frac{\sin(k)}{I_0(2kx)}. \tag{2.4}$$

Thus, the potential is a trapping potential in either z or r, but not both. The magnetic field already offers radial confinement, and this confinement is not diminished by the radial part of the applied potential. This can be seen as electric potential not appearing in Eq. 2.1. Thus, the electric fields are used to trap the plasma in z.

2.2 Description of the Experiments

In ALPHA and the Berkeley plasma trap, the electrodes of the Penning-Malmberg trap are cooled to cryogenic temperatures so that the confined plasmas are expected to equilibrate to low temperatures. A large, superconducting magnet is used to generate the solenoidal field.

Berkeley Plasma Trap

The Berkeley plasma trap was designed to study electron plasmas. In particular, it has been used to study the cooling of plasmas in a cavity, and these results are discussed in Chap. 6. Though not discussed here, the trap has been used to study the rotating wall technique [33] and phase space islands [34].

The electrons in the Berkeley plasma trap are produced by a heated filament. This produces hot electrons with a temperature of $\sim$1 eV. To catch the electrons, an electrode near the center of the trap is raised to a potential such that electrons bounce off. Then, another electrode downstream (closer to the filament) is raised to a potential such that electrons are trapped. Once these electrons are trapped, the filament can be turned off and these captured plasmas can be manipulated by altering the electrode potentials.

The Berkeley plasma trap deviates from a standard Penning-Malmberg trap in that not all of the electrodes are perfectly cylindrical (that is, shaped like Fig. 2.1). Some electrodes have a radius significantly smaller than the neighboring electrodes. The radius of these electrodes also varies axially as shown in Fig. 2.2. This is called a *bulge* and the shape was designed to isolate specific modes of the cavity from the exterior. This is accomplished by designing the cavity to act as microwave cutoffs for the mode [2, 35]. This prevents leakage of microwave modes in the cavity and shields the cavity modes from external radiation at the cyclotron frequency, leading to reduced heating. Any leakage of the mode results in an undesired decrease in the Q of the cavity. In a manner similar to the Purcell effect [36], we expect that this cavity will allow for enhanced cyclotron emission by plasma particles in the bulge cavity. This is described in detail in Chap. 6.

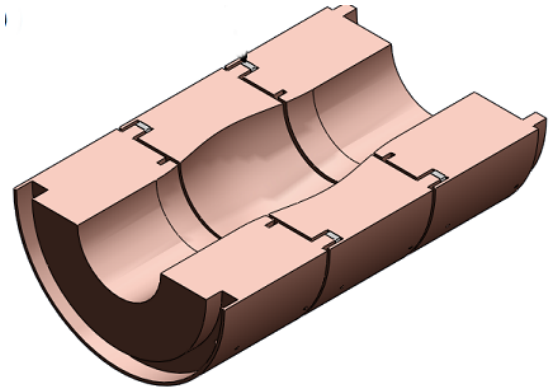

Figure 2.3: Some electrodes in the Berkeley plasma trap are designed to bulge to isolate the cavity modes from external radiation. Image by N. Evetts [2].

In theory, isolating part of the Penning-Malmberg trap is possible with physical doors that stop any microwave radiation. However, doors will also stop plasma from entering the trap and thus must open and close. Getting doors to operate in a magnetic, cryogenic environment is non-trivial and thus the bulge cavity is advantageous [1]. Experiments with doors and the bulge cavity is a planned future experiment for the Berkeley plasma trap.

ALPHA

The goal of the ALPHA experiment is to produce antihydrogen and measure antihydrogen properties. Antiprotons are produced for ALPHA by CERN's antiproton decelerator [25]. These antiprotons are captured in a Penning-Malmberg trap. Positrons are produced by a Na-22 source on the other side of the experiment. Once the antiprotons and positrons are appropriately cooled, they are brought together using an autoresonant process [37]. The antiprotons and positrons then combine to make antihydrogen, as described in the previous chapter.

Unfortunately, the Penning-Malmberg trap is unable to trap neutral particles. Thus, ALPHA superimposes an Ioffe trap (discussed below) on the Penning-Malmberg trap in order to trap the charged particles.

Ioffe Trap

In ALPHA, antihydrogen is trapped in an Ioffe trap [38], which is a $2n$-pole magnet for radial confinement with two mirror magnets for axial confinement. These coils are placed in the same region as the Penning-Malmberg trap such that charged and neutral particle trapping can be accomplished in the same space. Since the antihydrogen has a magnetic

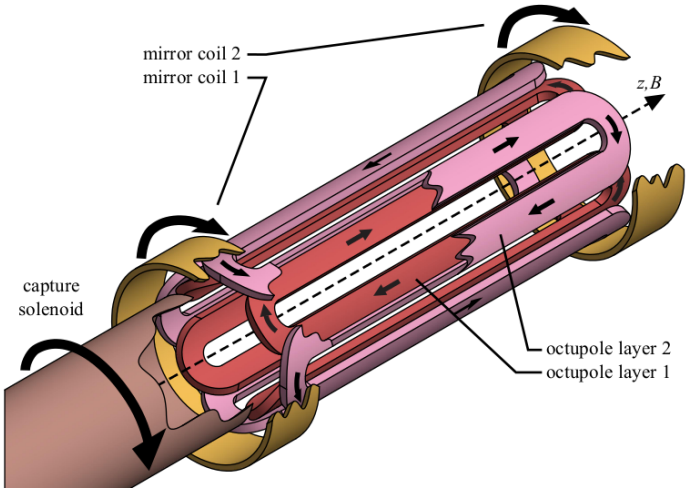

Figure 2.4: Neutral particles with a magnetic moment can be trapped with octupole and mirror magnets. The arrows in the figure show the direction of current. The electrodes of the Penning-Malmberg trap sit inside the capture solenoid. Image by C. So [3].

moment $|\vec{\mu}| = 0.56$ K/T [16], the energy of the antihydrogen due to these fields is

$$U = -\vec{\mu} \cdot \vec{B}. \tag{2.5}$$

Thus, antihydrogen can be trapped if there is a well in the magnetic field profile. However, if antihydrogen has kinetic energy that is much greater than the trap energy, the antihydrogen will not be trapped. Since it is difficult to construct trap depths larger than 1 T, this means antihydrogen needs to be colder than 0.56 K in order to be trapped. Thus, creating cold antihydrogen is necessary.

A radial magnetic gradient can be created with a multipole magnet. An ideal $2n$-pole magnet consists of $2n$ infinite, parallel wires evenly-spaced around a cylinder. Each wire carries the same current, but the direction of the current reverses between neighboring wires. This arrangement is only possible for even numbers, hence the name $2n$-pole. The field is proportional to r^{n-1}, with the minimum occurring at the center [39].

A physical multipole magnet is not infinite in length and consists of wires with finite thickness. Thus, the ideal r^{n-1} scaling is only approximately the field produced. However, as long as the particles are near both the radial and longitudinal center of the magnet, the approximation is good to within a few percent. It is also important to make sure the addition of the mirror coil and multipole fields do not cause the local minima of the magnetic field to shift off axis, as this could allow a mechanism for particles to escape [3].

The octupole and mirror fields obviously perturb the nearly uniform solenoidal field of the Penning trap. Thus, there is worry that these perturbations would cause losses in the plasma. Higher order multipoles are better to prevent this loss as they have a flatter profile

radially and thus better approximate the ideal Penning-Malmberg trap [40,41]. ALPHA uses an octupole as designing a higher order multipole magnet becomes an engineering challenge.

Chapter 3

Nonlinear Curve Fitting

In this chapter we provide the theoretical basis for fitting nonlinear curves to data using the Levenberg-Marquardt algorithm [42, 43]. We then discuss our implementation of the algorithm on graphical processing units (GPUs). The Levenberg-Marquardt algorithm is an integral part of the analysis of data for the plasma diagnostics described in Chap. 4 and 5.

3.1 Problem Formulation

Consider a function $f_{\boldsymbol{\beta}} : \mathbb{R}^d \to \mathbb{R}$ that is characterized by n fit parameters $\boldsymbol{\beta} \in \mathbb{R}^n$. Define $\mathcal{D} = \{(\vec{x}_i, y_i) | i \in \{1, ..., N\}\}$, a collection of N inputs $\vec{x}_i \in \mathbb{R}^d$ with N corresponding outputs $y_i \in \mathbb{R}$. We wish to find $\boldsymbol{\beta}$ such that $f_{\boldsymbol{\beta}}$ represents $\mathcal{D}$ "the best." Here, bold denotes a vector in $\mathbb{R}^n$ whereas the vector symbol denotes a vector in $\mathbb{R}^d$. We will assume that the residuals between the optimal fit and the data are Gaussian distributed with mean 0 and an unknown variance σ^2 [44]. Because of the law of large numbers, this is a justifiable and often used assumption. Further, we assume that each of the N inputs and outputs in $\mathcal{D}$ are independent of one another. We can then write the likelihood function of the data given the fit parameters as

$$p(\mathcal{D}|\boldsymbol{\beta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(f_{\boldsymbol{\beta}}(\vec{x}_i) - y_i\right)^2}{2\sigma^2}}. \tag{3.1}$$

Thus, the likelihood function is maximized when

$$\chi^2(\boldsymbol{\beta}) = \sum_{i=1}^{N} (f_{\boldsymbol{\beta}}(\vec{x}_i) - y_i)^2 \tag{3.2}$$

is minimized with respect to its argument $\boldsymbol{\beta}$. Note that we call this quantity χ^2 as by assumption the residual is a Gaussian random variable and thus this quantity is a sum of the squares of Gaussian random variables. For convenience, we define $\chi(\boldsymbol{\beta})$ as the vector of residuals,

$$\chi_i(\boldsymbol{\beta}) = f_{\boldsymbol{\beta}}(\vec{x}_i) - y_i, \tag{3.3}$$

which is consistent with the previous definition of $\chi^2(\boldsymbol{\beta})$. When the function $f_{\boldsymbol{\beta}}$ is linear in the parameters $\boldsymbol{\beta}$, standard linear regression (least-squares) techniques can be used [45, 46]. However, when $f_{\boldsymbol{\beta}}$ is nonlinear in the parameters the minimization must be handled more carefully.

3.2 Levenberg-Marquardt Algorithm

Define J as the $N \times n$ matrix corresponding to the gradient of $f_{\boldsymbol{\beta}}$ (or equivalently, $\chi(\boldsymbol{\beta})$) with respect to $\boldsymbol{\beta}$

$$J_{ij}(\boldsymbol{\beta}) = \left.\frac{\partial f_{\boldsymbol{\beta}}(\vec{x}_i)}{\partial \boldsymbol{\beta}_j}\right|_{\boldsymbol{\beta}}. \tag{3.4}$$

Consider a perturbation $\boldsymbol{\delta}$ away from $\boldsymbol{\beta}$. Assuming $f_{\boldsymbol{\beta}}$ is differentiable in $\boldsymbol{\beta}$ and the perturbation small, we see the change in the value of $f_{\boldsymbol{\beta}}$ is

$$\chi(\boldsymbol{\beta} + \boldsymbol{\delta}) \approx \chi(\boldsymbol{\beta}) + J(\boldsymbol{\beta})\boldsymbol{\delta}. \tag{3.5}$$

We wish to extremize $\chi^2(\boldsymbol{\beta})$, and thus we would like to find an update $\boldsymbol{\delta}$ such that it sets the gradient of $\chi^2(\boldsymbol{\beta})$ to zero. The gradient is

$$0 = \frac{\partial \chi^2(\boldsymbol{\beta} + \boldsymbol{\delta})}{\partial \boldsymbol{\beta}_j} = 2J^T(\boldsymbol{\beta} + \boldsymbol{\delta})\chi(\boldsymbol{\beta} + \boldsymbol{\delta}) \approx 2J^T(\boldsymbol{\beta} + \boldsymbol{\delta})(\chi(\boldsymbol{\beta}) + J(\boldsymbol{\beta})\boldsymbol{\delta}). \tag{3.6}$$

If J is invertible, this is satisfied when $\boldsymbol{\delta}$ is

$$\boldsymbol{\delta}_k = -J^{-1}(\boldsymbol{\beta})\chi. \tag{3.7}$$

However, it is unlikely that J will be invertible, as typically $N > n$, or the number of data points is usually larger than the number of fit parameters. Thus, we replace the inverse by the pseudo-inverse and get

$$\boldsymbol{\delta} = -J^{+}\chi = (J^T J)^{-1} J^T \chi. \tag{3.8}$$

$J^T J$ is a small matrix since often the number of fit parameters, n, is small. Because of this, the inverse is easy to compute. In our implementation, we quickly compute the inverse through a Cholesky decomposition, as the term multiplying $\boldsymbol{\delta}$ is positive-definite [46]. Using this solution as an update is the *Gauss-Newton* method for nonlinear fits. We will prove in the next section that the method is not guaranteed to converge locally or globally. The Levenberg-Marquardt algorithm introduces an unspecified regularization parameter λ, such that global convergence can be guaranteed for a sufficiently large choice of λ, when local convergence can be achieved [47].

Gauss-Newton algorithm

Assuming J has full column rank, J^TJ is invertible, and a Gauss-Newton update to the current guess $\boldsymbol{\beta}^{(k)}$ is

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + (J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}))^{-1}J^T(\boldsymbol{\beta}^{(k)})\chi. \tag{3.9}$$

We will investigate the convergence properties of this method. Consider the error at step $k+1$, $e^{(k+1)}$. Calling the optimum value $\boldsymbol{\beta}^*$ and the residual at the optimum value χ^*, this is [47]

$$\begin{aligned} e^{(k+1)} &= \boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^* \\ &= \boldsymbol{\beta}^{(k)} + (J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}))^{-1}J^T(\boldsymbol{\beta}^{(k)})\chi - \boldsymbol{\beta}^* \\ &= (J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}))^{-1}J^T(\boldsymbol{\beta}^{(k)})(J(\boldsymbol{\beta}^{(k)})e^{(k)} - \chi) \\ &= (J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}))^{-1}J^T(\boldsymbol{\beta}^{(k)})(-\chi^* + J(\boldsymbol{\beta}^{(k)})e^{(k)} - \chi + \chi^*) \end{aligned} \tag{3.10}$$

By the multidimensional mean value theorem [48], we have

$$\begin{aligned} J(\boldsymbol{\beta}^{(k)})e^{(k)} + \chi^* - \chi &= J(\boldsymbol{\beta}^{(k)})e^{(k)} - \int_0^1 dt\ J(\boldsymbol{\beta}^* + te^{(k)})e^{(k)} \\ &= \int_0^1 dt\ \left[J(\boldsymbol{\beta}^{(k)}) - J(\boldsymbol{\beta}^* + te^{(k)})\right] e^{(k)}. \end{aligned} \tag{3.11}$$

Assume that $J(\boldsymbol{\beta})$ is Lipschitz in the vicinity of $\boldsymbol{\beta}^*$, with Lipschitz constant γ. Since the fits in this thesis are fits to analytic functions, this assumption should be valid for the use cases. Then, we have

$$\begin{aligned} \left\| \int_0^1 dt\ \left[J(\boldsymbol{\beta}^{(k)}) - J(\boldsymbol{\beta}^* + te^{(k)})\right] e^{(k)} \right\| &\leq \int_0^1 dt \left\| J(\boldsymbol{\beta}^{(k)}) - J(\boldsymbol{\beta}^* + te^{(k)}) \right\| \|e^{(k)}\| \\ &\leq \gamma \|e^{(k)}\|^2 \end{aligned} \tag{3.12}$$

As we are optimizing $\|\chi\|_2$ we expect that at the optimal point, the gradient of $\|\chi\|_2$, which is proportional to $J^T(\boldsymbol{\beta}^*)\chi^*$, is zero. Plugging these results into Eq. 3.10 gives

$$\begin{aligned} \|e^{(k+1)}\| &\leq \|(J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}))^{-1}\| \left(\|J^T(\boldsymbol{\beta}^{(k)})\chi^*\| + \|J^T(\boldsymbol{\beta}^{(k)})\| \|J(\boldsymbol{\beta}^{(k)})e^{(k)} - \chi + \chi^*\| \right) \\ &\leq \|(J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}))^{-1}\| \left(\|J^T(\boldsymbol{\beta}^{(k)}) - J^T(\boldsymbol{\beta}^*)\| \|\chi^*\| + \gamma \|J^T(\boldsymbol{\beta}^{(k)})\| \|e^{(k)}\|^2 \right) \\ &\leq \gamma \|(J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}))^{-1}\| \left(\|\chi^*\| \|e^{(k)}\| + \|J^T(\boldsymbol{\beta}^{(k)})\| \|e^{(k)}\|^2 \right) \end{aligned} \tag{3.13}$$

There are two noteworthy aspects of the above equation. Global convergence is not guaranteed as the $\|e^{(k)}\|^2$ is large when the guess is far away from the optimum value. Furthermore, local convergence is not guaranteed. When $\gamma\|\chi^*\|$ is smaller than the smallest eigenvalue of J^TJ in the vicinity of $\boldsymbol{\beta}^*$, then the term in front of $\|e^{(k)}\|$ is less than 1 and the method converges. When this is not the case, local convergence is not guaranteed. In general, convergence is linear, but if $\|\chi^*\|$ is small (which could be reasonable for a fit with very little noise) then the convergence can be quadratic.

Regularizing the Gauss-Newton Algorithm

Similar to Tikhonov regularization, the Levenberg-Marquardt algorithm makes a simple modification of the update to be [49]

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + (J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}) + \lambda\mathbb{I})^{-1}J^T(\boldsymbol{\beta}^{(k)})\chi, \tag{3.14}$$

where $\mathbb{I}$ is the $n \times n$ identity matrix and $\lambda \in (0, \infty)$ is a parameter that is adjusted such that the residual decreases at each step of the algorithm. Note that now the restriction that J have full column rank is no longer strictly necessary, but we will show that in this case local convergence cannot be guaranteed. For the analyses in this thesis, the problems are well-formed and J has full column rank.

Note that for small λ, the Levenberg-Marquardt algorithm reduces to the Gauss-Newton algorithm. For large λ, the inversion becomes trivial and the descent is along $J(\boldsymbol{\beta})$, or the gradient descent direction. Thus, the Levenberg-Marquardt algorithm interpolates smoothly between Gauss-Newton and gradient descent.

The analysis of the change in error at each step can proceed exactly as before, and this gives

$$\|e^{(k+1)}\| \leq \left\|(J^T(\boldsymbol{\beta}^{(k)})J(\boldsymbol{\beta}^{(k)}) + \lambda\mathbb{I})^{-1}\right\| \left((\gamma\|\chi^*\| + \lambda)\|e^{(k)}\| + \gamma\|e^{(k)}\|^2 \left\|J^T(\boldsymbol{\beta}^{(k)})\right\|\right). \tag{3.15}$$

λ has the effect of increasing the eigenvalues of J^TJ. Thus, the eigenvalues of $(J^TJ + \lambda\mathbb{I})^{-1}$ are small as λ gets large. This means that there is a choice of λ that will make $\|(J^TJ + \lambda\mathbb{I})^{-1}\|$ smaller than $\gamma\|J^T(\boldsymbol{\beta}^{(k)}\|$, and hence global convergence is now guaranteed with this modification, provided the method is locally convergent. Note that this does not mean the method will converge to the global minimum, it simply means that the method will converge to *a* minimum given any initial condition.

The local convergence properties are shared with the Gauss-Newton method. In particular, the constraint that $\gamma\|\chi^*\|$ is less than the smallest eigenvalue of J^TJ in the vicinity of $\boldsymbol{\beta}^*$ is still necessary [50]. This is why full column rank in J is necessary.

For large λ, $J^TJ + \lambda\mathbb{I}$ has large eigenvalues and the update of the guess is small. Multiple steps with large λ thus lead to slow convergence. The algorithm starts with a small value of λ to try to take the largest step it can towards the optimum value of $\boldsymbol{\beta}$. In our code we start with $\lambda = 1$. The algorithm increases λ (by multiplying it by 2) until it finds a $\boldsymbol{\delta}$ that

reduces $\chi^2(\boldsymbol{\beta}+\boldsymbol{\delta})$, meaning it tries to take the largest step it can toward reducing the total residual. Once found, $\boldsymbol{\delta}$ is added to $\boldsymbol{\beta}$, and λ is set to a smaller value to try to take a big step in the next iteration. In our implementation, we divide λ by 3 when a successful update step is found. This process is repeated until the change, $\boldsymbol{\delta}$, is below a certain threshold chosen by the user, or a maximum number of steps is reached. Our termination condition is $\|\boldsymbol{\delta}\|_2 < t\|\boldsymbol{\beta}\|_2$, where t is a user-defined tolerance.

Good initial guesses for $\boldsymbol{\beta}$ are still necessary. A bad guess can converge to an unexpected minimum. In addition, convergence is faster if the initial guess is decent, and, as the Levenberg-Marquardt algorithm is the bottleneck in the analyses for plasma diagnostics discussed in this thesis, it is important to speed it up. We discuss choices of initial parameters when discussing the particular applications of the Levenberg-Marquardt algorithm.

Implementation on a GPU

We implemented the Levenberg-Marquardt algorithm on a GPU using the CUDA language [51]. The code is available in Sec. D.1. We chose to use a GPU as it is a relatively inexpensive way to speed up code and because there are well-developed libraries for linear algebra on GPUs.

We use the `-O3` and `-use_fast_math` compiler flags to make use of compiler optimizations. All the numbers are represented as floats and the float version of mathematical functions (such as expf) is used, as this is slightly faster than using double precision. We have not written the code with double precision so we do not know how much of a speed advantage there is from using floats.

Most of the parallelizations described here are from Ref. [52]. The Levenberg-Marquardt algorithm itself is not very parallelizable, as we have to find steps to sequentially improve the current best guess for all of the fit parameters. However, many of the intermediate steps can be split up into tiny problems that can be run in parallel. This attribute makes this problem ideal to being sped up by GPUs.

We wrote the code as a sequence of vector/matrix operations as much as possible. This was so that vector/matrix operations could be done using cuBLAS, as this is far more optimized than anything we would be able to write [53]. An interesting aspect of this problem is that we have many problems that are the size of the number of fit parameters, n, which is typically small. An example of this is the matrix inversion in Eq. 3.14. Using a GPU in this instance is overkill and may end up being slower due to the slow communication between the CPU and GPU. In contrast, there are also computations that depend on the number of data points, N, which is quite large ($\sim 10^6$) in our applications. An example of this is the computation of J^TJ in Eq. 3.14. These benefit greatly from the parallelism that GPUs offer. With this in mind, in executing vector operations for small vectors we used the CPU whereas for other tasks we use the GPU. As communication between the GPU and CPU is slow, there may be a slight speedup from doing all the computations (even the small ones) on the GPU, but we have not tested this.

Another positive aspect of this problem is that the gradient of $f_{\boldsymbol{\beta}}$ at each value of $\vec{x}_i$ can each be computed independently. Since these are each small computations, they are perfect to be executed on separate GPU nodes at the same time. Since we would like to fit to a particular, known, function, we analytically compute the derivative of $f_{\boldsymbol{\beta}}$ with respect to each of its parameters. These components of the gradient can all be found using only a few numerical operations. General purpose Levenberg-Marquardt fitters such as the one in LabView compute the gradient numerically with finite differences, and thus requires more computations and are less accurate [54]. Thus, by specializing to the specific cases we care about, we gain some speed and accuracy in our method.

Multiple GPUs and multiple CPU cores can also be exploited for a speedup [52]. In the Levenberg-Marquardt algorithm, often many steps (values of λ) are tried before an optimal one is found. While it depends on the problem, we found for our typical fits that this is true for about 30% of steps. Interestingly, this is much lower than for the applications considered in Ref. [52]. The optimal step is found more quickly by checking multiple potential steps (different values of λ) at the same time on different cores. Even if an optimal step would have been found with just one value of λ, it could be that the next trial value of λ would further reduce the residual and, hence, this could speed up the process. As we do not have a machine with multiple GPUs we were not able to implement this parallelization.

We find that the main bottlenecks in our code are matrix-matrix multiplications and matrix-vector multiplications. The next largest time sinks are when norms and updates are computed on large vectors. This makes sense since the vectors and matrices are large, the size of the the number of pixels. However, these operations are probably close to optimal because they are implemented in cuBLAS. As discussed previously, the best way to improve this is to get more GPU cores. While we have not made any definitive measurements, we noticed no change in runtime when memcpy[1] of a small vector was taken out (when it was deemed to be unnecessary). Thus, the bottleneck really seems to be computational rather than communicational.

[1] memcpy is a transfer of data from the CPU to the GPU.

Chapter 4

Radial Profile Diagnostic

In this chapter, we will discuss analysis of data from a plasma radial profile. The electric potentials confining a plasma in the Penning-Malmberg trap can be lowered to let the plasma escape. The plasma signal can be converted into a light signal to give a field line integrated image of the plasma. These images can be used to determine the shape and extent of the plasma [55]. Together with the knowledge of the initial confining potentials, the images yield the density of the plasma [3].

The analysis of plasma images was previously implemented with a multi-core code written in LabView with a runtime on the order of minutes. In order to get faster results, images were often downsampled at the cost of the accuracy. In addition, the optimization problem did not include the center coordinates of the plasma as fit parameters. Using the GPU Levenberg-Marquardt implementation discussed in Chap. 3, the runtime is improved to a few seconds and is able to fit the center location at the same time as the other fit parameters, yielding increased accuracy and precision.

This method only analyzes images with isolated plasma distributions with radial symmetry. We develop an algorithm to fit overlapping plasma distributions and plasma distributions that do not have radial symmetry.

4.1 Description

As discussed in Chapter 2, the plasmas at CERN and Berkeley are confined in a series of electrodes with adjustable voltages. The electrodes are biased to a high voltage (that is, $eV \gg T_{plasma}$) in order to trap the plasma particles. To gain knowledge of the plasma properties, the electrode voltages are lowered such that the charged particles can escape toward a micro-channel plate (MCP)/phosphor screen array [56]. This is shown in Fig. 4.1. The MCP amplifies the charge incident on it and this amplified charge is then converted to a light signal by a phosphor screen. This light can be collected on a CCD camera for a view of the field line integrated radial profile of the plasma. Examples of images are shown if Fig. 4.2. The light can also collected using a photodiode circuit that can read out a signal

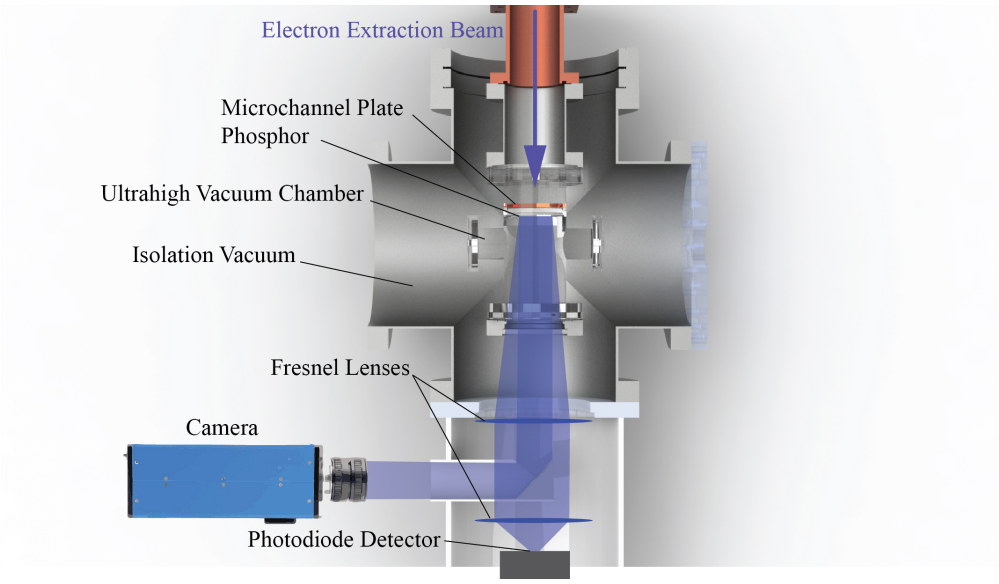

Figure 4.1: Extracted electron signal is converted to a light signal by the MCP, which can then be detected by a camera or photodiode detector. Image by A.P. Povilus [1]. The camera image is used in the diagnostic described in this chapter while the photodiode signal is used in the diagnostic described in Chap. 5.

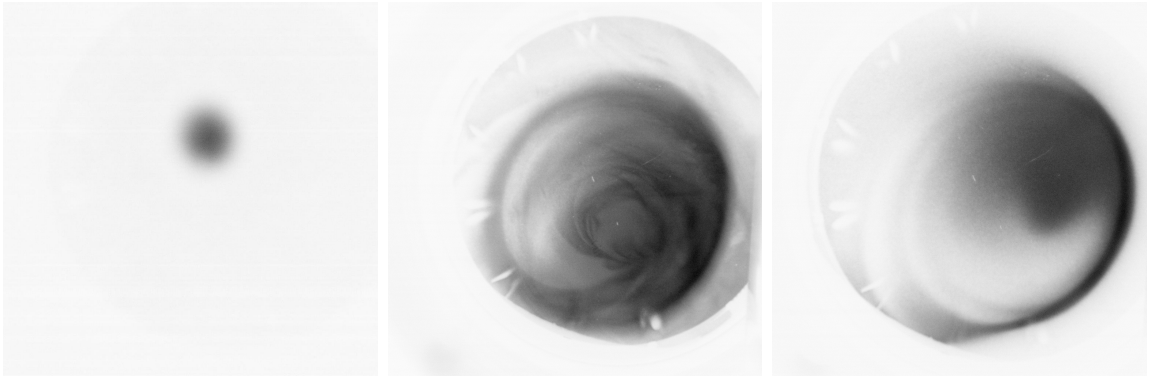

Figure 4.2: Plasma images from the light collected on the camera in Fig. 4.1. This is an accurate representation of the field line integrated radial profile of the plasma. The circular envelope in the image is the boundary of the Penning-Malmberg trap, and the white streaks near the edge of the circle are regions where the phosphor has become inactive. The first plot shows a typical plasma, whereas the latter two show plasmas in which instabilities have significantly affected the shape during extraction.

proportional to the number of charged particles escaping the trap as a function of time. This data is useful for measuring the temperatures of plasmas, described in Chap. 5.

In these experiments, a solenoid provides a strong magnetic field. As discussed in Sec. A.2, the charged particles stick to field lines as they approach the MCP. The plasma expands by a factor $\sqrt{B_{trap}/B_{MCP}}$ without its shape being distorted much by the extraction process. The $\vec{E} \times \vec{B}$ drifts also have a negligible contribution as its effect is azimuthally symmetric. Thus, this diagnostic provides an accurate field line integrated radial profile of the plasma that can be analyzed to infer the plasma's properties. Note that these measurements can only be obtained by destroying the plasma.

4.2 Applications of the Diagnostic

The image diagnostic can be useful for determining qualitative aspects of the trapped plasma. In particular, instabilities such as the diocotron are apparent as shearing on the images [57]. The image yields insights into the shape of the plasma, and can verify if the plasma will be able to fit through apertures. Based on where the plasma is centered on the image can reveal field misalignment. As discussed in Chap. 2, this is interesting as misalignments can lead to heating of the plasma.

The diagnostic yields quantitative information about the plasma, namely the density and length [3]. Up to normalization, the image gives $\mathcal{N}(r)$, the radial profile of the plasma. The assumption of thermal equilibrium in the axial direction yields the Poisson-Boltzmann equation

$$\begin{aligned} \Delta\phi(r,z) &= -qn(r,z) \\ n(r,z) &= \mathcal{N}(r)\exp\left(-\frac{q\phi(r,z)}{T}\right). \end{aligned} \tag{4.1}$$

Initially, an ellipsoid consistent with $\mathcal{N}(r)$ is chosen for $n(r,z)$. Iteratively, one solves for the self field using Poisson's equation (which is a time-consuming step) and then uses this self-field with the vacuum field to determine a new guess for the $n(r,z)$. In simple geometries (in particular an ideal Penning-Malmberg trap) analytic expressions for the potentials are available and Poisson's equation need not be solved. The new guess is used as an update for the old guess. This is then repeated until convergence. As described in Sec. B.2, the length scale of electric field fluctuations in a plasma is determined by the Debye length $\lambda_D = \sqrt{T/ne^2}$, where n is the density of the plasma. Thus, if Poisson's equation is solved numerically, it must be done on a grid with spacing of order λ_D. For cold ($\lesssim 300$ K) plasmas, λ_D is small enough that the calculation becomes computationally prohibitive.

In the cold plasma regime, or when $\lambda_D \ll R_p$, the plasma radius, we expect the potential at a given radius to be constant (see Sec. B.2). Thus, at a given radius, we expect that the axial profile will be a step function. The ends of this step function are unknown, and must be solved for. This method then proceeds as above, but now only the boundary needs

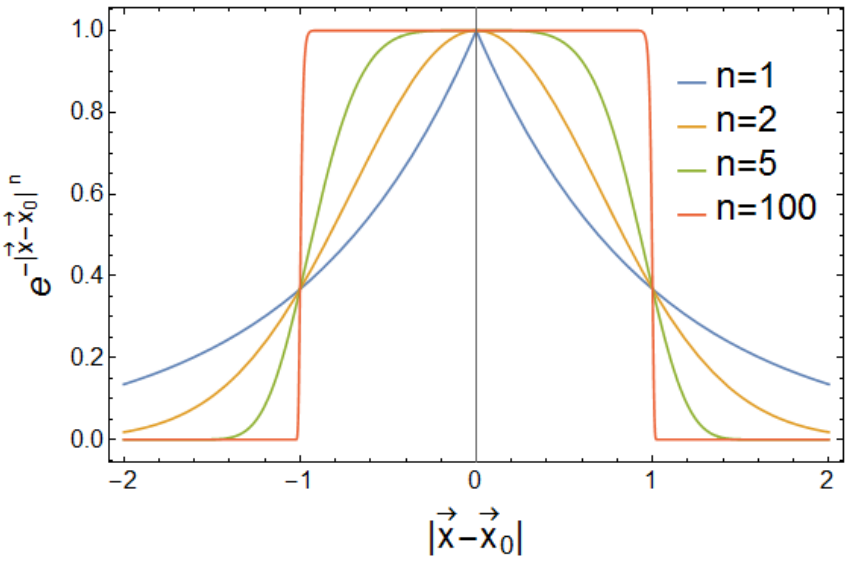

Figure 4.3: The modified Gaussian, Eq. 4.2, as a function of the shape parameter n. We have set $A = 1, B = 0$ and $R = 1$ for ease of plotting.

to be solved for, which is computationally advantageous [3]. The potential is calculated for the current guess, and is used to determine the force on a test particle at the plasma boundary. These forces are used to push the particles to obtain a new guess for the density function. Ultimately, there will be little force on test particles on the boundary. Thus, this waterbag method yields the full plasma profile $n(\vec{x})$, up to normalization, though in the zero temperature limit.

Though, in theory, these solvers works with direct MCP images, in practice, the images contain noise that could lead the algorithms astray. Thus, it is beneficial to be able to define the radial plasma distribution by a smooth function, and hence we develop a method to perform fits to the images. These smooth fits can also be advantageous to easily compare experimental plasmas with theory and simulation predictions.

4.3 Implementation of Image Fitter

We interpret the image as pixel intensities (y) as a function of pixel positions in two dimensions ($\vec{x}$). We use a mask to remove from the fit any pixels that are outside of the cavity or correspond to regions of inactive phosphor. We chose the function of the form

$$f(\vec{x}) = Ae^{-\left(\frac{|\vec{x}-\vec{x}_0|}{R}\right)^n} + B \tag{4.2}$$

to fit the radial profile of the plasma. This model has $d = 6$ fitting parameters: the center location of the plasma ($\vec{x}_0$) , the radius (R), the peak density (A), the offset (B), and a

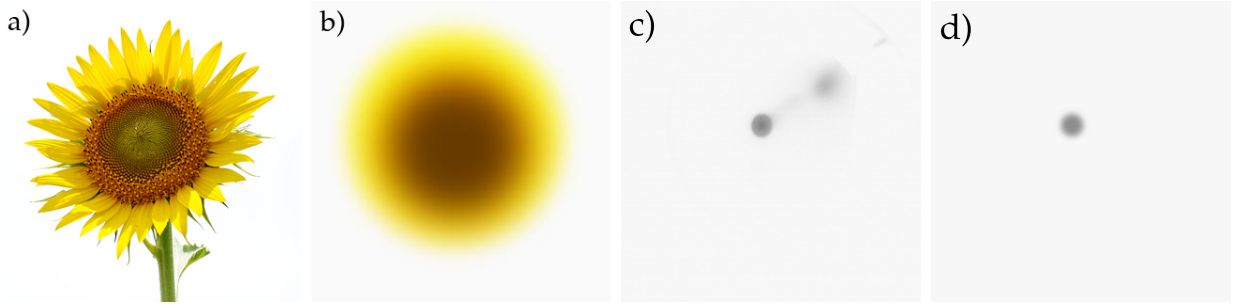

Figure 4.4: a) An image of a sunflower b) The RGB color intensity matrix for a) is passed into our modified Gaussian fitting algorithm and the result of the image fitter with the colors combined is shown c) A typical camera image of a plasma d) The result of the modified Gaussian fitting algorithm applied to the image in c). We see the fitter is able to find and fit the areas of high radial symmetry and is not influenced by the presence of noise in either of these cases. In the plasma image, the fitter is able to fit the brighter spot without trying to fit the fainter spot to the upper right of the bright spot in c).

shape parameter (n). The form in Eq. 4.2 is not theoretically justified, but is empirically justified by the following. As described in Sec. B.2, a zero temperature plasma has a flat radial distribution that cuts off sharply at the edges. Thus, if a plasma is infinite and well-approximated as a cylinder, the radial profile will have a sharp cutoff, corresponding to large n in Eq. 4.3. In contrast, for high temperature plasmas the distribution looks like a Gaussian, which obviously corresponds to $n = 2$ in Eq. 4.3. This modified Gaussian distribution gives a smooth way to go from one extreme to the other, as shown in Fig. 4.3. This form also has the property that it is analytically integrable over all of $\vec{x}$, which will be used later.

We use the Levenberg-Marquardt algorithm described in Chap. 3 to perform the fit. In the algorithm, we force the constraints $n, R > 0$ by increasing λ if there is ever a step $\boldsymbol{\delta}$ that yields $n < 0$ or $R < 0$, even if this results in a decrease of the residual. Distributions with $1 \geq n > 0$ have a kink at the origin and do not provide a good fit to our plasma images. But, the nonlinear fitter will sometimes need to search values of n in this regime to converge to its optimal value. Thus, while we allow for n to take values between 0 and 1, it is rare for the fitter to find these values optimal. No constraints are applied to the other parameters. However, while testing, we find the other parameters seem to avoid extreme values, provided that we start with decent initial values.

We introduce a hard cut-off of 500 steps of the Levenberg-Marquardt algorithm, but this condition is never reached during regular operation. Typically, convergence is achieved in about 50 steps and the fit can be performed in a few seconds with the GPU implementation of Levenberg-Marquardt. LabView's Levenberg-Marquardt routine takes over a minute to fit Eq. 4.2 with the center $\vec{x}_0$ fixed (that is, fitting fewer parameters), so the advantage of the GPU is clear. Fig. 4.4 shows examples of successful fits obtained from the algorithm.

In Fig. 4.5, we consider the strong scaling by looking at the runtime as a function of the

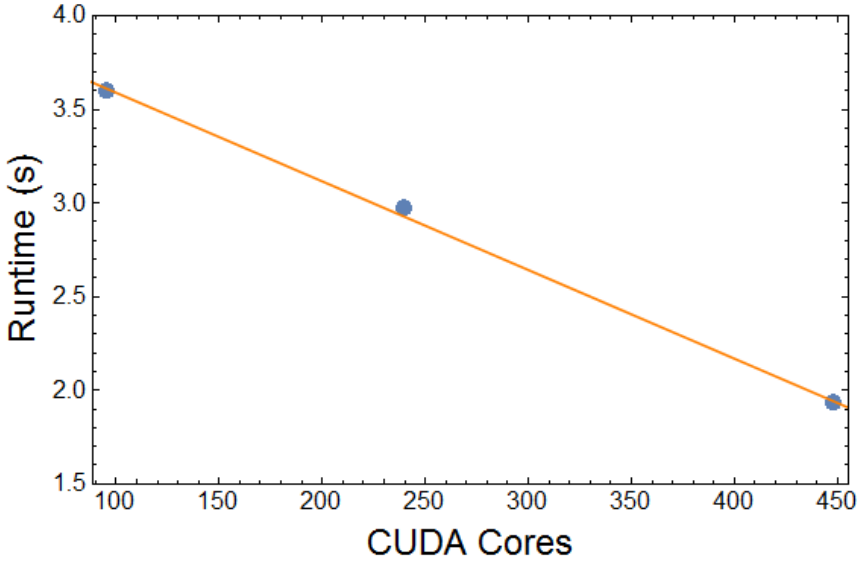

Figure 4.5: Runtime as a function of number of CUDA cores for analyzing a 1024x1024 pixel plasma image. The line shown is the linear fit. As expected, more processing power yields faster runtimes. The original LabView code required over a minute to analyze the image and fit fewer parameters.

number of CUDA cores available on the GPUs for a plasma image (1024x1024 pixels) [58]. The three nVidia GPUs used were a C2050 Tesla (with 3GB of memory and 448 cores), a C1060 Tesla (with 4GB of memory and 240 cores) and a GeForce GT 630M (with 2GB of memory and 96 cores). While this may not be a fair comparison as clock speeds may differ between the GPUs, we see a general trend that more cores gives us a shorter runtime. While the memory of the three GPUs were all different, we do not see a trend based on this. This indicates that the code is processor bound and not memory bound. We would expect this runtime to continue decreasing with more cores, but the trend probably would not continue past 1000 cores. With this many cores, the runtime could start to level off as the center-finding algorithm (which is serial) becomes increasingly important and there may be issues with load balancing on the GPUs when there are so many nodes.

Depending on the complexity of an image, the algorithm can take different times even if the size of the image is kept constant. Because of this, it is not easy to study the weak scaling, where the problem difficulty is scaled in tandem with the number of cores [58].

Initial Guesses

As mentioned in Chap. 3, in order to ensure fast convergence and to avoid convergence to false minima, we are careful in the selection of initial parameters for the algorithm. The following describes the selection process for initial values of each of the parameters in Eq. 4.2.

Estimation of the background intensity (B)

The initial estimate of the offset is taken to be the intensity of the pixel with the lowest intensity.

Estimation of the peak density (A)

The initial estimate of the peak density is taken to be the intensity of the pixel with the largest intensity, with the estimated offset, B, subtracted off.

Estimation of the shape parameter (n)

The shape parameter is difficult to estimate. This is the only parameter for which we do not alter the initial estimate of based on the available image. We set $n = 3$, as through testing on plasma images we found this was a reasonably fast and robust choice.

Estimation of the radius (R)

The integral, $\mathcal{I}$, of the distribution turns out to be useful in estimating the radius, R. Taking units where the size of a pixel is 1, $\mathcal{I}$ is the sum of the intensities of all the pixels in the image (with the estimated offset B subtracted off). If the plasma radius is reasonably small, this sum will be equal to the integral of the fit function function over $\mathbb{R}^2$. Even if the radius is larger (or the whole plasma is not contained in the image), the sum will only differ from the $\mathbb{R}^2$ integral of the fit function by an $O(1)$ factor. For an initial estimate of R for the Levenberg-Marquardt algorithm, this estimate is empirically found to be close enough to provide fast and reliable convergence.

To relate $\mathcal{I}$ to R, we note that

$$\int_{\mathbb{R}^2} d^2\vec{x}\ \ e^{-\left(\frac{|\vec{x}-\vec{x}_0|}{R}\right)^n} = \pi R^2 \Gamma\left(\frac{2+n}{n}\right). \tag{4.3}$$

Conveniently, for $n \in [1, \infty]$, which is the range of interest for most of our plasmas, the value of the gamma function is within a factor of 2 of 1. So, for the same reasons as above, for an initial estimate of R, it is sufficient to assume $\Gamma((2+n)/n) \approx 1$. This gives an initial estimate

$$R \approx \sqrt{\frac{\mathcal{I}}{A\pi}}. \tag{4.4}$$

Later, we will discuss using the singular value decomposition (SVD) of the coordinates of pixels that make up the Gaussian to estimate R. For circular distributions of plasmas, this is slower than the simple estimate given above and hence is not used here. However, for elliptical distributions of plasmas, the above integral is not as useful and hence an SVD is necessary.

Estimation of the center of the plasma ($\vec{x}_0$)

The algorithm used to estimate the center is an extension of that discussed in Ref. [59]. The main idea behind the algorithm is that, if a distribution is Gaussian (or more generally, radially symmetric about a center), the gradient at each point should point away from the center of the distribution. Thus, if we draw lines along the intensity gradient at every point, the point that minimizes the weighted distance to all of these lines should correspond to the center.

Following [59], we take the components of the gradient in the Robert's cross configuration, meaning the axes are tilted 45 degrees from the standard axes [60]. Call these new axes $\hat{u}$ and $\hat{v}$. Consider $\vec{x}_k$, the midpoint of the pixels at $\{(i,j),(i+1,j),(i,j+1),(i+1,j+1)\}$. The first-order approximation of the gradient there is

$$\vec{\nabla} y_k = (y_{i+1,j+1} - y_{i,j})\hat{u} + (y_{i,j+1} - y_{i+1,j})\hat{v}. \tag{4.5}$$

We wish to set the two components of this gradient equal to the "velocity" for $\vec{x}_k$ to obtain the desired line as a function of a parameter t. Rewriting $\hat{u}$ and $\hat{v}$ in terms of the standard directions and rearranging gives that the slope of the line is

$$m_k = \frac{(y_{i+1,j+1} - y_{i,j}) + (y_{i,j+1} - y_{i+1,j})}{(y_{i+1,j+1} - y_{i,j}) - (y_{i,j+1} - y_{i+1,j})}. \tag{4.6}$$

Then, the equation of the line is $\vec{x}^{(2)} = m_k(\vec{x}^{(1)} - \vec{x}_k^{(1)}) + \vec{x}_k^{(2)}$, where superscripts denote the components of $\vec{x}$. The minimum distance between a point $\vec{x}_c$ and this line is

$$d_k^2 = \frac{\left((\vec{x}_k^{(2)} - \vec{x}_c^{(2)}) - m_k(\vec{x}_k^{(1)} - \vec{x}_c^{(1)})\right)^2}{m_k^2 + 1}. \tag{4.7}$$

Allowing for weights w_k, by minimizing $\sum_k w_k d_k^2$, we find the optimal $\vec{x}_c$ that is our best guess for the plasma center. The result of this is

$$\begin{aligned} \vec{x}_c^{(1)} &= \frac{\langle m(\vec{x}^{(2)} - m\vec{x}^{(1)})\rangle - \langle m\rangle\langle \vec{x}^{(2)} - m\vec{x}^{(1)}\rangle}{\langle m\rangle^2 - \langle m\rangle} \\ \vec{x}_c^{(2)} &= \frac{\langle m(\vec{x}^{(2)} - m\vec{x}^{(1)})\rangle\langle m\rangle - \langle m^2\rangle\langle \vec{x}^{(2)} - m\vec{x}^{(1)}\rangle}{\langle m\rangle^2 - \langle m\rangle}, \end{aligned} \tag{4.8}$$

where for any function g,

$$\langle g\rangle = \frac{\sum_k \frac{g_k w_k}{m_k^2 + 1}}{\sum_k \frac{w_k}{m_k^2 + 1}}. \tag{4.9}$$

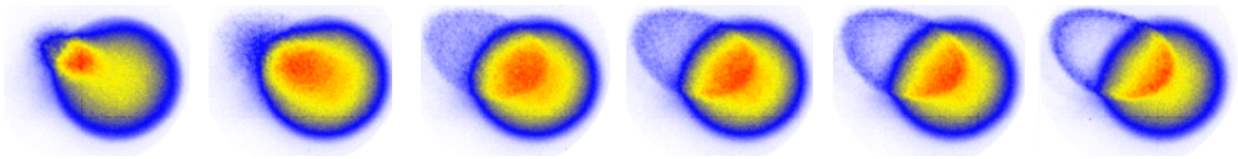

Figure 4.6: A series of images with varying degrees of overlap between the positron plasma and antiproton plasma. Image is modified from [4].

Initially, a guess for the center is established with weights w_k equal to the magnitude of the gradient vector squared at the the kth position. This is done as higher gradient vector norms tend to be more reliable in that they carry less error, and are likely to be close to the center of the distribution. However, this still suffers from the problem that if there are multiple valid centers, the algorithm will choose some point in between the multiple possible centers. Thus, we extended the algorithm in [59] by iterating the process with a weight of the magnitude of the gradient vector squared over the distance to the current best center guess squared. This process ignores large intensity values that are far away the current center guess, and hence allow the method to converge to a true center of a distribution rather than an average of all possible centers. We tested various powers of the distance between 1 and 5 and found that squaring produced the results that matched best with the center determined visually. We found that higher powers of distance tended to converge to a center more quickly, but the center it converged to was not good. These results may change if a higher order estimate of the gradient is used, but this was not tested.

When there is little noise and only one center, this method is slower than simpler methods such as the moment method [61], but produces similar results. However, when there are multiple possible centers this algorithm produces far better results. Further, the moment method will not find a center that is outside of the image, which is a somewhat common occurrence in the experimental plasma images.

We found the iteration usually takes 20 steps or fewer, independent of what type of picture is being analyzed. We place a hard cutoff for the iteration at 100 steps, but there were no test cases when this limit was reached. Though this code could be run quickly on a GPU, the center-finding algorithm runs on a CPU in about the time of one step of the Levenberg-Marquardt algorithm on a GPU, so we find that it is not the main bottleneck. Thus, there would be little performance gain in the overall algorithm from implementing the center-finding method on the GPU.

4.4 Fitting Multiple Plasma Profiles

As shown in Fig. 4.6, in ALPHA, it is sometimes desirable to image both the positron and antiproton plasma in the same image. This can be useful for studying mixing and

compression [4, 62]. The antiproton clouds are elliptical in shape, and thus instead of the functional form given by Eq. 4.2, we take the form

$$f(\vec{x}) = Ae^{-\left((\vec{x}-\vec{x}_0)^T\mathcal{R}^{-2}(\vec{x}-\vec{x}_0)\right)^{n/2}} + B, \tag{4.10}$$

where $\mathcal{R}^{-2}$ is a positive definite matrix with three free parameters, and the other variables are the same as Eq. 4.2. We call this a matrix squared to suggest that the elements of $\mathcal{R}$ have dimensions of length and as a reminder that $\mathcal{R}^2$ is positive definite. One way to characterize these free parameters would be the length of the semi-major axis, the length of the semi-minor axis, and the tilt of the ellipse. Thus, this model has $d = 8$ fit parameters.

If the antiproton and positrons clouds are not overlapping (or the overlap is small), the fitting is straightforward. We apply the center finding algorithm described in the previous section to find the center of one distribution. Then we fit the modified Gaussian to it, and subtract that fitted distribution from the image and proceed to fit the other distribution. However, in practice, the antiproton and positron clouds are often overlapping, and thus this naive idea is not sufficient.

Our solution is to determine systematically which pixels are representative of the positron cloud and which are representative of the antiproton cloud. We first use the center finding algorithm of the previous section to find the center of the positron distribution. We then fit the whole image to the form of Eq. 4.2. Call the obtained fit $f_0(\vec{x})$. The positron distribution is preferable to the antiproton distribution as it is usually bigger and it is a fit with fewer parameters. This means that the positron fit is better constrained and should, in general, yield more reliable results.

In the region where there is a significant contribution to the image due to the antiproton cloud, the image pixel values will be consistently larger than the expected value from just the positron cloud. We take our null hypothesis to be that $f_0(\vec{x})$ describes the positron cloud and for each pixel position $\vec{x}$ we define a p-value of observing a larger pixel intensity than that in the actual image.

$$\wp(\vec{x}) = \int_{y(\vec{x})}^{\infty} dy\ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(f_0(\vec{x})-y)^2}{2\sigma^2}} = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{f_0(\vec{x}) - y(\vec{x})}{\sqrt{2}\sigma}\right)\right) \tag{4.11}$$

σ is taken to be the standard deviation of the fit residuals. This is actually an overestimate considering all of the pixels that do not satisfy our hypothesis. However, it still sets a useful scaling for $\wp(\vec{x})$. We compute $\wp(\vec{x})$ for each pixel. We mark all pixels with $\wp(\vec{x}) < 1/\sqrt{|\mathcal{D}|}$ that are adjacent to another pixel with $\wp(\vec{x}) < 1/\sqrt{|\mathcal{D}|}$ to be ignored, where $|\mathcal{D}|$ is the total number of pixels used in the fit for $f_0(\vec{x})$. We choose this threshold so that the expectation of the number of pixels ignored that actually are consistent with the hypothesis is only two pixels, since p-values are approximately uniformly distributed between 0 and 1 when the null hypothesis is true [63]. The coincidence with neighboring pixels comes from the fact that we expect the antiproton cloud that we wish to exclude to be contiguous. Given this choice of threshold, we would expect $\sqrt{|\mathcal{D}|}$ pixels that satisfy the hypothesis to have $\wp(\vec{x}) < 1/\sqrt{|\mathcal{D}|}$ so without the coincidence the method would be too restrictive.

Then, Eq. 4.2 is fit to the distribution again, but this time excluding from the fit the points that were marked to be ignored. The initial parameters for the fit should be the parameters from $f_0(\vec{x})$ in order to ensure fast convergence. This process is repeated until the fit parameters do not change appreciably when the procedure is applied.

Once a good positron distribution has been determined through this method, we subtract off the fit positron distribution from the image. We then fit the remaining distribution to the form in Eq. 4.10.

The only issue left is how to initialize the fit parameters given the new form. Many of the fit parameters can be estimated the same as before, but the three free parameters in $\mathcal{R}^2$ requires attention.

We recall that singular value decompositions (SVD) will find the best fitting ellipse to a set of points [64]. When the value of the argument of the exponential in Eq. 4.10 is -1, the vector $\vec{x}-\vec{x}_0$ sits on the boundary of the ellipse defined by $\mathcal{R}$. Thus, we consider the coordinates of all points with pixel intensity greater than e^{-1} times the maximum pixel intensity. We then find the center of these points and subtract the center off of each point. These re-centered coordinates are concatenated into a matrix A. Note that the center should be quite close to the center found by the center finding algorithm. We use the SVD to decompose $A = U\Sigma V^\dagger$. Then, the product $V^\dagger \Sigma V$ will be our initial guess for $\mathcal{R}^2$.

If the antiproton image is cut off, the best fitting ellipse found from the SVD may not be representative of the actual ellipse boundary. As before, we expect this to only affect the elements of $\mathcal{R}^2$ by an $O(1)$ factor, but there are three free parameters being estimated by this method so the result is less robust than in previous cases. However, for an initialization of the Levenberg-Marquardt algorithm we believe this should be sufficient.

4.5 Future Work

We are still implementing the method for multiple plasma profiles. Of particular interest will be a careful study of how much the plasmas can be overlapping and still yield useful information of each underlying profile.

Acknowledgements

I would like to thank Alexander P. Povilus for writing the serial code that I optimized, and all of the LabView wrappers of the code to create an easy interface for experimentalists to use. Nathan Belmore is working on implementing the fitter for overlapping distributions.

Chapter 5

Temperature Diagnostic

As described in Chap. 1, cold positron plasmas are essential for production of trappable antihydrogen. To understand how to produce these cold positron plasmas, a method to measure plasma temperatures is necessary.

In a Penning-Malmberg trap, the electrode potentials can be lowered to let plasma particles escape in a controlled manner. ALPHA and the Berkeley plasma trap are able to obtain the charge escaping from the Penning-Malmberg trap as a function of time. We will show that a nonlinear fit of this data yields the plasma temperature. This method was originally formulated by Ref. [65]. Theoretical corrections to the diagnostic, such as the effect of space charge, were studied in Refs. [66, 67]. These effects complicate the fitting of data, and only the earliest arriving plasma particles can be used in the nonlinear fits. Selecting the region corresponding to the earliest arriving particles was previously done by hand. Not only was this process slow, but it was error-prone as different people could analyze the same dataset and get different temperatures. We fully automate the fitting to remove this bias. Further, our GPU implementation of the analysis runs in ~ 10 s. This speed is desirable when mapping the cooling rate of plasmas as a function of various parameters, such as the experiment described in Chap. 6.

5.1 Experiment

To determine the temperature of the plasmas, the potential confining the plasma is slowly lowered such that the hottest electrons escape (see Fig. 5.1). As described in Sec. 4.1, the electron signal is converted to a light signal which corresponds to the charge arriving at the MCP as a function of time. An example of the signal is shown in Fig. 5.2.

The rate at which the potential is lowered must be slow enough such that the time scale of the bounce motion of the plasma is irrelevant compared to the typical time that the potential changes. This is so the bounce motion does not "hide" the effect of the lowering potential. As we will assume in our analysis that the plasma is well-described by a Maxwellian distribution, we might expect that the collision frequency would be a relevant timescale for

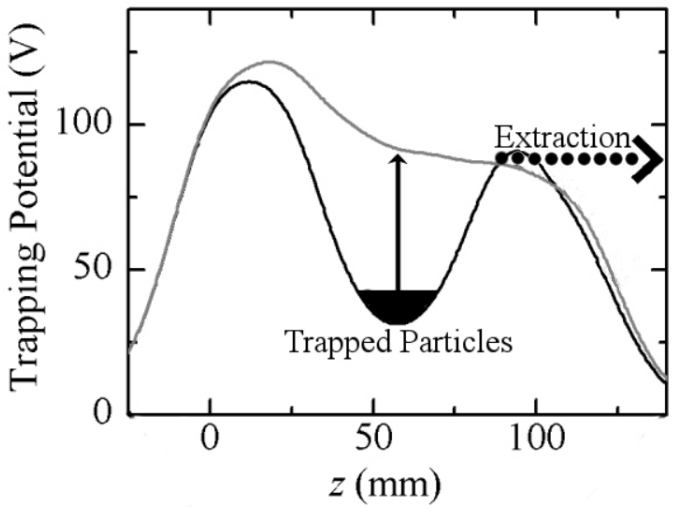

Figure 5.1: Particles are extracted from the trap by slowly (over milliseconds) lowering the potential under the trapped particles, allowing them to escape over the lower barrier of the well. Image is modified from an original created by A.P. Povilus [1].

how fast the potential can be ramped. In particular, collisions can repopulate the tail of a distribution and thus a previously trapped particle can gain enough energy to escape the well through collisions [66]. Experimentally, this is found not to be a limitation [68]. The true limitation comes from instabilities that develop as the plasma is extracted, for example the diocotron [69]. The ramp rate must be fast enough that the diagnostic terminates before these instabilities affect the distribution. This imposes a hierarchy

$$\omega_z^{-1} \ll t_{ramp} \ll t_{instab.}. \tag{5.1}$$

5.2 Theoretical Basis

Consider a plasma with temperature T with a net rotation ω. Let H denote its energy and p_θ its angular momentum. As described in Sec. B.2, the distribution function of the plasma provided equilibrium is reached is [1, 65, 67]

$$\begin{aligned} f(\vec{x},\vec{v}) &\propto \exp\left(-\frac{H-\omega p_\theta}{T}\right) \\ &\propto \exp\left(-\frac{m}{2T}\left(v_r^2+v_z^2+(v_\theta-r\omega)^2-2\frac{e}{m}(V_b(\vec{x})-\phi(\vec{x}))-r^2\omega^2+\omega r^2\omega_c\right)\right) \end{aligned} \tag{5.2}$$

Here, $V_b(\vec{x})$ is the external potential from biased electrodes while $\phi(\vec{x})$ is the plasma self-potential.

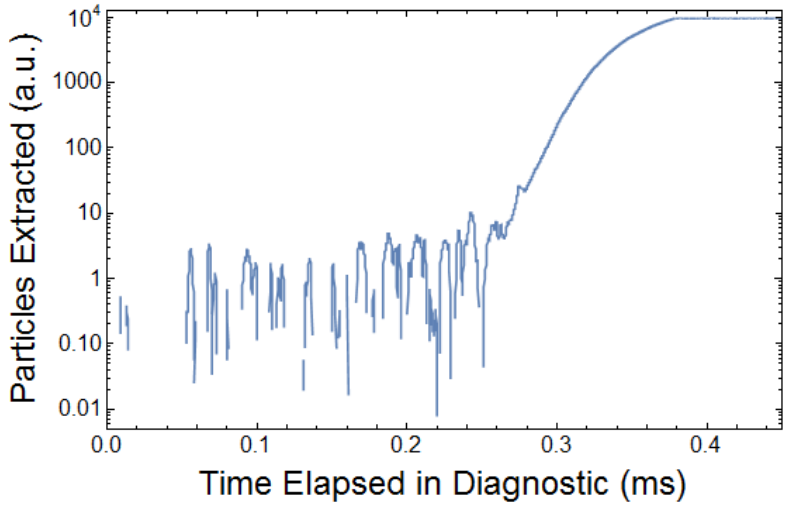

Figure 5.2: A typical signal from the extraction of an electron plasma at UC Berkeley. Temperatures are obtained by fitting the early signal to an exponential. Before 0.28 ms, the signal is dominated by background noise and fluctuations due to the discrete nature of the signal. As a pure exponential should appear linear on a log plot, we see that by 0.35 ms, the curve sub-exponential. The detector saturates near 0.38 ms.

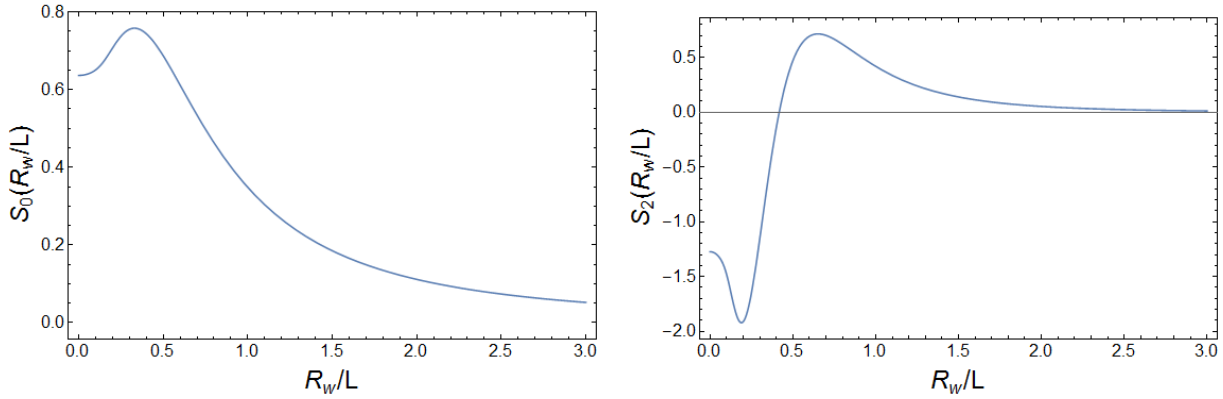

Figure 5.3: Plots of $S_0(R_w/L)$ (left) and $S_2(R_w/L)$ (right), which govern the amplitude and radial variation of the applied potential, see Eq. 5.3.

Suppose the plasma is trapped by two biased electrodes, as in Fig. 2.1. An electron escapes if its parallel kinetic energy is larger than than the potential barrier, $-e(V_b(r,t) - \phi(\vec{x}))$ evaluated at the z position of the center of one of the biased electrodes, where the potential is largest. This obviously ignores tunneling effects, but this should be a negligible contribution for plasmas considered here. From Sec. 2.1 we know that near the center of an electrode, V_b has form (assuming the other blocking electrode is far away)

$$V_b(r,t) \approx V_0(t)\left(S_0\left(\frac{R_w}{L}\right) + \frac{r^2}{L^2}S_2\left(\frac{R_w}{L}\right)\right), \tag{5.3}$$

if a voltage V_0 is applied to the electrode. The S functions are defined in Chap. 2, and plotted in Fig. 5.3. The self field does, in principle, depend on time as well, but we will assume that the diagnostic terminates before the self field changes appreciably. We will consider these effects in Sec. 5.3. The number that have escaped is thus all electrons with parallel kinetic energy greater than this potential barrier, or the integral of Eq. 5.2 over all space and velocities, with the constraint $v_z > \sqrt{-2e(V_b(r,t) - \phi(\vec{x}))/m}$. Assume that the potential barrier starts high enough that effectively all particles are trapped. Note that the v_r and v_θ integrals are independent of the v_z integral, so these integrals are simply a proportionality constant. Ignoring issues of statistical fluctuations and the distribution changing due to the changing potential (discussed later), the number of particles to escape for a given $V_0(t)$ is

$$\begin{aligned} N(V_0) &= \sqrt{\frac{2m}{\pi T}}\int d^3\vec{x}\; n(\vec{x}) \int_{\sqrt{\frac{-2e(V_b(r,t)-\phi(\vec{x}))}{m}}}^{\infty} dv_z\; \exp\left(-\frac{mv_z^2}{2T}\right) \\ &= \int d^3\vec{x}\; n(\vec{x})\left(1 - \mathrm{erf}\left(\sqrt{\frac{-e(V_b(r,t)-\phi(\vec{x}))}{T}}\right)\right), \end{aligned} \tag{5.4}$$

where $n(\vec{x}) = \int d^3\vec{v}\; f(\vec{x},\vec{v})$. Assuming we extract the temperature from the particles that arrive earliest to the detector, the space charge will not appreciably change and the dependence of $\phi(\vec{x})$ on $V_0(t)$ will be weak. Further, assume that the plasma dimensions are smaller than the electrode dimensions so that only the S_0 term of Eq. 5.3 contributes to the integral above. Thus

$$\begin{aligned} \frac{dN}{dV_0} &= S_0\left(\frac{R_w}{L}\right)\frac{e}{T}\int d^3\vec{x}\; n(\vec{x})\sqrt{\frac{T}{-\pi e(V_b(r,t)-\phi(\vec{x}))}}\exp\left(\frac{e}{T}(V_b(r,t)-\phi(\vec{x}))\right) \\ &\approx S_0\left(\frac{R_w}{L}\right)\frac{eN}{T}, \end{aligned} \tag{5.5}$$

where the second line is true when $-e(V_b(r,t) - \phi(\vec{x}))/T \gg 1$, or equivalently when the first order expansion of the error function in Eq. 5.4 describes the function well [70]. Thus,

we expect the signal to increase as an exponential, with the growth rate proportional to temperature. With knowledge of the voltages applied to electrodes, $V_0(t)$ is specified, which can then be used with the number of charged particles that reach the MCP as a function of time to extract the temperature [65, 66].

5.3 Limitations of the diagnostic

Minimum Measurable Temperature

We will write the plasma self potential as $\phi(\vec{x}) = \phi_0(\vec{x}) + \delta\phi(\vec{x})$, where $\phi_0(\vec{x})$ is the plasma self potential before any charge has been extracted, or when $-eV_b \gg T$. We expect $\delta\phi(\vec{x})$ to be a perturbative correction when particles initially start to escape in the diagnostic [71]. Near the center of the plasma, Gauss's law gives [72]

$$\phi_0(\vec{x}) = \phi_0(0) + \left(\frac{r}{2\lambda_D}\right)^2 \frac{T}{e} + ... \tag{5.6}$$

Assuming $n(r)$ is constant for $r \sim \lambda_D$, the contributions from higher order terms should be negligible for $r \sim \lambda_D$. This should be satisfied when $\lambda_D \ll R_p$, the plasma radius. While this may not be true for a hot plasma, we will show that the derivation in Sec. 5.2 only breaks down for low temperatures and hence ignore this issue. We use the above in Eq. 5.4 to find the number of particles escaped for a given applied blocking potential. Assuming a long plasma, only the radial integral is non-trivial. Assuming that $\lambda_D \ll L$, the length of the electrode (see Eq. 5.3),

$$N(V_b) = 2\pi l_p n(0) \int dr\ r \left(1 - \text{erf}\left(\sqrt{\mathcal{V} + \frac{e\delta\phi(\vec{x})}{T} + \frac{r^2}{4\lambda_D^2}}\right)\right), \tag{5.7}$$

where l_p is the length of the plasma, and for convenience we define $\mathcal{V} = -e(V_b - \phi_0(0))/T$. As we are evaluating the expression for a fixed V_b, the radial variation in the error function is determined by the parameter r^2/λ_D^2, as we have assumed that the contribution from $\delta\phi(\vec{x})$ is small. Recalling that for large arguments the error function is exponentially close to 1, this means that nearly all the escaped charges are coming from a few λ_D of the center [66]. This justifies the need for the earlier constraint that $n(r)$ is constant for $r \sim \lambda_D$. Note that Eq. 5.7 scales as the number of particles in a Debye cylinder. Notably, this number is independent of plasma density as λ_D^2 scales as $n(0)^{-1}$. Since $\lambda_D^2 \propto T$, low temperature plasmas have a limited number of particles obeying the form of Eq. 5.5. We expect a limitation to the validity of the diagnostic for low temperature plasmas.

Initially, we expect $\delta\phi$ to have a negligible contribution to the extracted number of particles. This assumption allows us to use Gauss's law on the extracted charge density (the integrand of Eq. 5.7) to find the electric field due to the extracted charge. In the limit $\mathcal{V} \gg 1$,

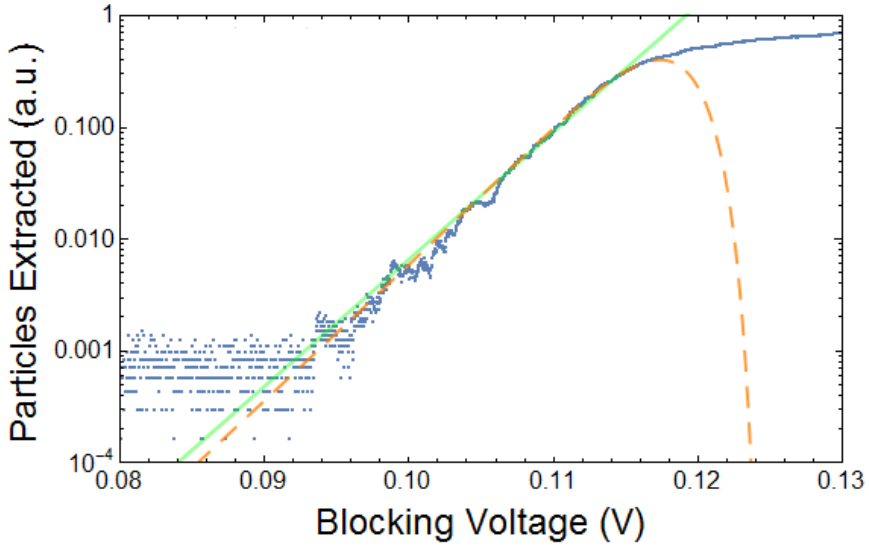

Figure 5.4: Signal from MCP/phosphor array during diagnostic extraction (shown in blue) of an electron plasma at UC Berkeley as a function of blocking voltage. The green line is an exponential fit to the curve, while the orange dotted line is a fit of the form of Eq. 5.10. We see that Eq. 5.10 extends the range of validity of the theory past the exponential region.

$$E(r) \approx \frac{2en(0)\lambda_D^2}{r}\frac{\exp(-\mathcal{V})}{\sqrt{\pi\mathcal{V}}}\left(1-e^{-\frac{r^2}{4\lambda_D^2}}\right). \tag{5.8}$$

We then integrate this from the wall (where the potential should vanish) to a point within r of the center. This yields

$$\delta\phi(r) = \frac{T\exp(-\mathcal{V})}{e\sqrt{\pi\mathcal{V}}}\left[\left(\gamma+\Gamma\left(0,\frac{R_w^2}{4\lambda_D^2}\right)+\ln\left(\frac{R_w^2}{4\lambda_D^2}\right)\right)+\left(\frac{r}{2\lambda_D}\right)^2\right]+..., \tag{5.9}$$

where γ is the Euler-Mascheroni constant. Note that for large arguments, the incomplete gamma function is zero. We plug this form in as $\delta\phi(\vec{x})$ in Eq. 5.7. Performing the integral, we find

$$N(V_b) \approx 4\pi l_p\lambda_D^2 n(0)\frac{\exp\left(-\mathcal{V}-\frac{\exp(-\mathcal{V})}{\sqrt{\pi\mathcal{V}}}\left(\gamma+\ln\left(\frac{R_w^2}{4\lambda_D^2}\right)\right)\right)}{\sqrt{\pi}\left(1+\frac{\exp(-\mathcal{V})}{\sqrt{\pi\mathcal{V}}}\right)\sqrt{\mathcal{V}+\frac{\exp(-\mathcal{V})}{\sqrt{\pi\mathcal{V}}}\left(\gamma+\ln\left(\frac{R_w^2}{4\lambda_D^2}\right)\right)}}. \tag{5.10}$$

While this expression is only valid for $\mathcal{V} \gg 1$ (in practice, it breaks down around $\mathcal{V} \approx 2$), it is the most accurate closed form solution to the problem. We show the deviation from the extraction signal in Fig. 5.4. More complete solutions require complicated numerical integrals, and hence this form is desirable for generating Monte Carlo samples across a

broad range of plasma parameters. Fitting to this form instead of the form of Eq. 5.5 may seem desirable, but we do not recommend it. This form may seem to reveal information about the density, but this dependence is weak, and so no more information about plasma parameters is available from this form. While this form may allow for more data for more accurate fits, other effects such as a poor choice of extraction time can make the signal deviate from exponential before the deviation in the form of Eq. 5.10 takes over. By fitting to a simpler function, we allow for the flexibility of still obtaining temperatures when effects such as these are pertinent. We will show, however, that fitting to a more complicated form becomes necessary for cold plasmas.

Suppose we would like to measure temperatures to a precision of α. In an ideal analysis (that is, analyzing a perfect signal with no noise), two factors affect the precision: the statistical uncertainty from the limited number of observations and the breaking down of our assumed exponential functional form in Eq. 5.5. By the law of large numbers, we would expect the error in T from an analysis to scale as $N(V_b)^{-1/2}$. Thus, the number of particles necessary for a precision of α is α^{-2}.

We design the temperature fitting algorithm to distinguish an exponential region from a non-exponential one. This algorithm will have some deviation from exponential that it will tolerate before it declares the curve no longer exponential. We estimate this deviation by determining when $dN/dV_b = 0$, using the form of Eq. 5.10. For typical plasma parameters, we find that $\gamma + \ln(R_w^2/4\lambda_D^2)$ takes values between 10 and 20. For these values, $dN/dV_b \approx$ zero when $\mathcal{V} \approx 2$, which corresponds to 2% of the available $4\pi l_p \lambda_D^2 n(0)$ being extracted. We let $\epsilon = 0.02$ represent this fraction. Hence, from Eq. 5.10, we find the limit

$$T_{min} = \frac{e^2}{4\pi l_p \epsilon \alpha^2} = 8 \text{ K} \left(\frac{1 \text{ cm}}{l_p}\right) \left(\frac{0.1}{\alpha}\right)^2. \tag{5.11}$$

Of course, in a real measurement, background noise and non ideal detector sensitivity will make measurement of such a temperature difficult. An experimental measurement can approach this theoretical limit by carefully reducing these effects. A circuit with a silicon photomultiplier has been designed for this purpose [73].

For a cold plasma, where this argument is relevant, we expect the potential in the plasma to be approximately constant (see Sec. B.2). Thus, in Eq. 5.4, we expect $n(\vec{x})$ should be largely independent of z, and the form of Eq. 5.7 is justified. Even though we assumed an infinite plasma in the derivation, we expect the results to hold for finite plasmas [66].

For plasmas colder than the limit in Eq. 5.11, a fit will have to be performed for a larger part of the extraction curve (that is, including the non-exponential region) to have sufficient statistics to constrain the temperature. Fig. 5.4 shows that Eq. 5.10 only extends the range of validity slightly, so a more accurate representation of the extraction curve is necessary. To solve for the plasma extraction curve generally, Eq. 5.7 must be solved self-consistently for every V_b. This can be done by iteratively including higher order perturbations in Eq. 5.7, or by using Poisson's equation as in Ref. [66]. These solutions are computationally intensive, but must be calculated for every extraction to fit a low temperature plasma extraction.

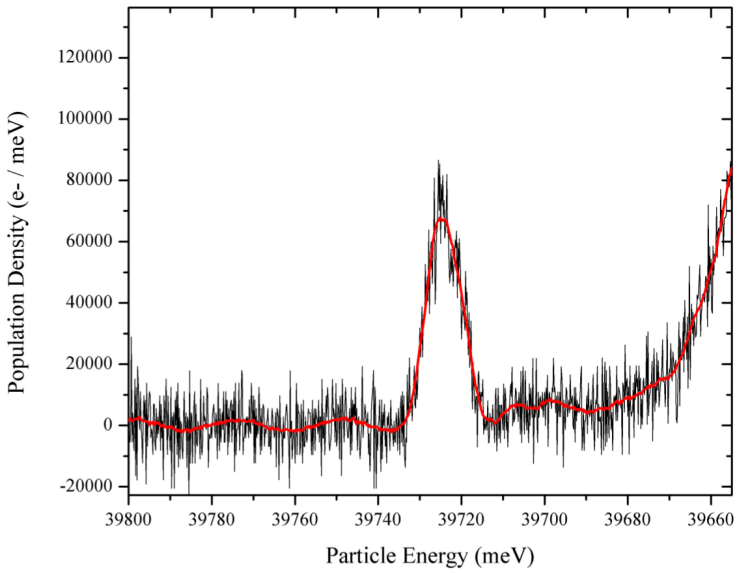

Figure 5.5: Signal from MCP/phosphor array during diagnostic extraction as a function of blocking voltage for a plasma with a phase space island. We see that part of the plasma is completely ejected before the bulk of the plasma follows. This is an effect not accounted for in the derivation of the equations pertaining to the temperature diagnostic. Figure taken from Ref. [1].

Effects not Included in the Analysis

There are additional effects that complicate the temperatures from the diagnostic. Recall that the potential is ramped slowly compared to a typical bounce time so that bounce effects are not apparent in a dump. Thus, the potential is adiabatically changed and will lead to some cooling in the parallel direction. The second adiabatic invariant (see Sec. A.1) will remain constant through this process. This invariant scales as $J \propto \sqrt{T_\parallel} l_p$, where l_p is the length of the plasma. Thus, in this case $T_{\parallel,f} = T_{\parallel,0} l_{p,0}^2 / l_{p,f}^2$. Waterbag simulations show that this could be a 50-300% effect on temperature. Cooling may be happening in more than one dimension, so simulations would be necessary to fully characterize this effect.

Finally, we assume that the equilibrium is simple enough to be described by the form in Eq. 5.2. When there are driving sources (which could be due to noise), phase space islands can form [34]. As shown if Fig. 5.3, part of the plasma is extracted in a clearly non-exponential manner, making temperatures extracted from it meaningless. If the number of trapped particles is small, the thermodynamic assumptions that justify Eq. 5.2 may not be valid. Finally, the hole left in the center by the plasma extraction can drive a diocotron instability that further complicates the analysis of extracted particles at late times [69].

5.4 Continuous Fitter

One method employed to extract the temperature is to fit Eq. 5.5 to the raw signal off of a charge collector (assuming it is continuous). A naive way to perform this fit would be to take the logarithm of the signal and perform a linear fit. However, because of electronic noise, the signal could become negative and its logarithm become ill-defined. Further, a linear fit to the logarithm fits an objective function that is the L_2 norm of the residual of the log of the data minus the log of the fit, which is a perverse choice for an objective function. Instead, as in Chap. 3, we assume that the residuals between the data and fit are independent and Gaussian distributed with mean 0 and unknown variance. We then proceed to use the Levenberg-Marquardt algorithm from Chap. 3 to determine the fit parameters.

Suppose $V_b(t)$ is linear in t. From Eq. 5.5, we then expect the signal to be exponential in time. Allowing for background noise that could contribute to a constant offset, we fit a 3 parameter function

$$N(t) = A + e^{\frac{t-t_0}{T}}. \tag{5.12}$$

Here, A is the aforementioned offset, T, is proportional to the temperature, and t_0 is a parameter relating to when the exponential starts to rise. Note that a prefactor in front of the exponential is equivalent to a redefinition of t_0 and thus would be redundant.

As argued Sec. 5.3, there are various mechanisms that will cause the expression in Eq. 5.5 to be invalid. The exponential form will only hold at early times. However, what constitutes an early time is not obvious and differs from plasma to plasma. Thus, a proper window of the data to fit Eq. 5.12 on must be chosen to obtain valid results. In prior analyses of this data, the selection was done by hand, which may lead to inconsistencies among people performing the fits. In addition, the Berkeley plasma trap sometimes operates in a state where it does a plasma dump every few seconds, and it is not feasible to process all this data manually.

Our algorithm to find the window is motivated by the assumption that sources of noise in the system should be equivalent before and during the plasma extraction period. Thus, we calculate the noise level before any plasma is extracted and compare this value to the noise level in fits of multiple possible choices of fit window. If the noise in the fit is smaller than the noise in the baseline, the fit is performing better than expected, and hence is overfitting. If the noise in the fit is larger than the noise in the baseline, the exponential assumption is invalid.

Following [1] and as shown in Fig. 5.4, we define three times to assist in finding the optimal window: t_-, a time sufficiently before any charge has reached the MCP, t_0, a guess at the time when enough charges have arrived at the MCP to obtain a good fit, and t_+, a time when the signal is clearly in the nonlinear regime.

To establish these times, we first fit the initial 0.1 ms of the data to a constant to establish an estimate of the baseline (A) and the noise level (σ). We then define t_- to be the point where the signal value is $\xi\sigma$ larger than the baseline value, where ξ is an adjustable parameter.

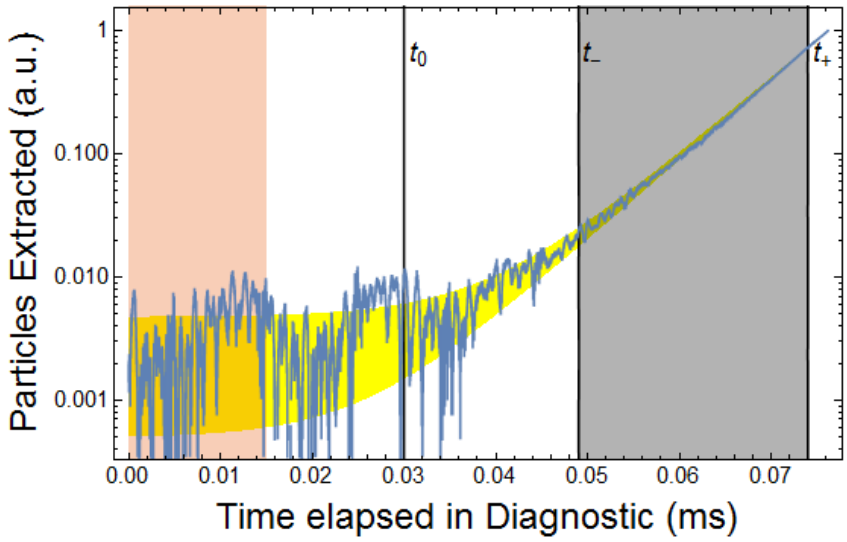

Figure 5.6: The blue line is the charge at the MCP as a function of time during a plasma temperature measurement. Our automated fitter calculated a temperature value of $T = 79 \pm 2$ K, and this fit is shown as the yellow band. The orange shaded region is fit to establish the baseline and noise level. The black vertical lines are, in order from left to right, t_0, t_- and t_+. The black shaded region shows the region from which $\mathcal{N}$ intervals are evenly selected.

If known, ξ would be chosen such that $\xi\sigma$ is greater than the amplitude of low frequency noise that could mimic the signal. In our final implementations, we used $\xi = 10$. This further ensures that there is a sufficiently long fit region to allow a well-specified temperature to be obtained from the fit. We then choose t_0 to be 0.4 ms before t_-. We do not choose t_0 to be the first datapoint, as there can be low frequency noise that will make the variance appear larger in the region from the first datapoint to t_-. The time t_+ is chosen to be a fraction (usually 0.85) of the $t \to \infty$ asymptote of the curve, or the detector saturation.

We then define $\mathcal{N}$ times, where the (zero-indexed) ith time is

$$t_i = t_- + \frac{i}{\mathcal{N}-1}(t_+ - t_-). \tag{5.13}$$

We then fit Eq. 5.12 for each of the $\mathcal{N}$ intervals $[t_0, t_i]$. Define N_i as the number of observations in $[t_0, t_i]$. To compare these fits, we consider the log-likelihood (this is just the log of Eq. 3.1), of each of the fits,

$$l_i = -N_i \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{j=1}^{N_i}(f_{\boldsymbol{\beta}}(\vec{x}_j) - y_j)^2. \tag{5.14}$$

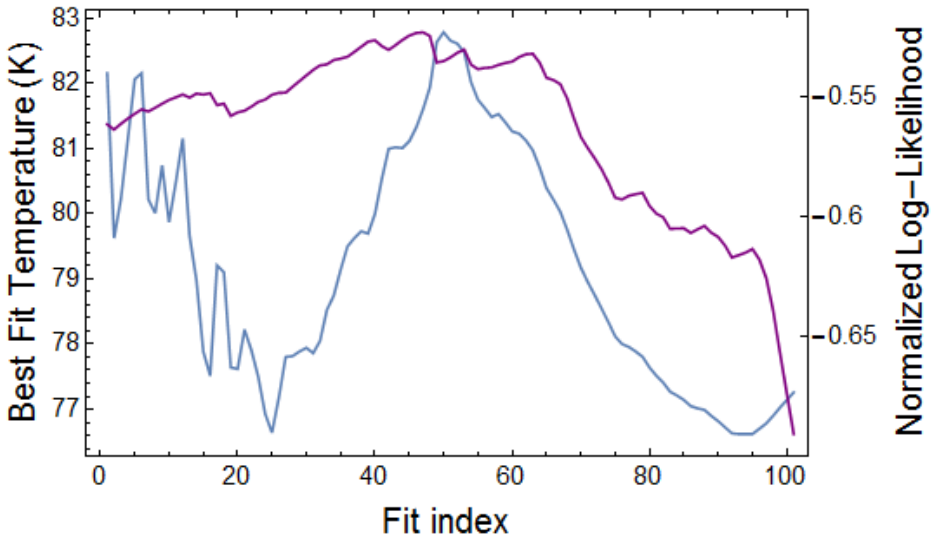

Figure 5.7: The temperature (blue line) and normalized log-likelihood (purple line) as a function of the $\mathcal{N} = 100$ intervals considered in choosing the optimal fit, for the same plasma dump as Fig. 5.4. The temperature and error evaluated from these values gives 79 ± 2 K. Note that below index 20, the variance in best fit temperatures is large due to the relatively small number of points used to obtained the fits. Above index 70, the curve is poorly modeled by an exponential and the normalized log-likelihood becomes small.

However, this suffers from the fact that it scales linearly with N_i, the number of points in the fit, and this is increasing for each of the $\mathcal{N}$ intervals (note that the second term is a χ^2 distribution and thus it also scales as N_i). Thus, to keep comparisons fair, we consider the normalized log-likelihood, which, dropping the now constant first term becomes

$$\tilde{l}_i = -\frac{1}{2N_i\sigma^2}\sum_{j=1}^{N_i}(f_{\boldsymbol{\beta}}(\vec{x}_j) - y_j)^2. \tag{5.15}$$

The value of σ is the standard deviation on the constant fit to the first 0.1 ms of data. To obtain the best value of temperature with an uncertainty, we perform weighted averages of T and T^2, where the weights are given by $e^{|\tilde{l}-\frac{1}{2}|}$ of the fit. The $\frac{1}{2}$ is, as discussed earlier, to ensure that the fit is not overfitting. As shown in Fig. 5.4, we find that this penalization has the desired behavior of fits getting worse as $t_i \to t_-$, and these overfit temperatures are weighted less.

We find that for all $\mathcal{N} \gtrsim 50$, the temperature and error values are similar. In our implementation, we set the default value to 100, but this is an adjustable parameter. We compared the results of this implementation with manual window selection and other ideas for automation and found that the results were comparable within error.

For this implementation, we set a hard cut-off of 1000 steps for the Levenberg-Marquardt algorithm. Typically optimal values are found in about 20 steps.

Using our GPU implementation of Levenberg-Marquardt, the ALPHA temperature fits are completed in a few seconds with a nVidia GeForce 780 GTX GPU with 2304 cores. We also have a version of our code that can run on any computer using LabView's standard implementation of Levenberg-Marquardt. This takes a few minutes to run, so the advantage of the GPU is clear. With the GPU, temperature values are available in real-time with no human input. This code has also been adopted by the UC Berkeley plasma experiment.

As the GPU version runs quickly, we have not tried to optimize the selection of initial parameters. In particular, a good estimate for T may help convergence of the algorithm.

5.5 Discrete (Poisson) Fitter

The methods in the previous section ignore statistical fluctuations due to the discrete nature of electron arrivals. We have also assumed that the noise is a random variable with constant variance, which may be a restrictive assumption. Further, while the previous statistical methods have been shown to be robust, they are rather *ad hoc* in that they are not derived from basic statistical principles. For these reasons, it was necessary to formulate a temperature fitter that incorporates these effects.

Ref. [73] has developed a circuit with a silicon photomultiplier (SiPM) for increased timing resolution on electron arrivals. In particular, the circuit can distinguish the $\sim$100 earliest arriving electrons. Using this data, one could imagine fitting temperatures to the arrival times of electrons to the MCP. However, because of the variability in the response of the entire system of MCP, phosphor screen, and the photodetector to a single electron, it becomes difficult to construct an appropriate discrimination function to distinguish electrons from noise. In addition, as the timing resolution is not perfect, pileup complicates distinguishing late arriving electrons.

We bin the data as this has two desirable effects: the variability the response of a single electron is averaged over by binning and thus a simpler discrimination function can be used and high-frequency noise is smoothed over. The method can thus make use of single particle arrival times when it is available, without being plagued by pileup when particle response signals begin to overlap. Effectively, this is a rigorous way to interpolate between fitting electron arrival times and the continuous fitter of Sec. 5.4 depending on the available data.

Much like the analysis in Sec. 4.4, the way we will determine temperatures from the data is to make a "bad" fit, and use the parameters from that fit as a null hypothesis to throw out the data that is not consistent with the exponential assumption.

With the form of Eq. 5.5, we know that the expected number of counts per bin will be (assuming for now no noise or background)

$$\Lambda_i = A\left(\exp\left(-\frac{V_{i+1}}{T}\right) - \exp\left(-\frac{V_i}{T}\right)\right), \tag{5.16}$$

where V_i and V_{i+1} are the lower and upper ends of the voltage bin. Each of the $\mathcal{B}$ bins is an independent Poisson process with rate Λ_i. For convenience, define the vector $\mathcal{T} = (T, A)$. Thus, the likelihood function is [74]

$$p(\mathcal{D}|\mathcal{T}) = \prod_{i=1}^{\mathcal{B}} \frac{\Lambda_i^{n_i}}{n_i!}\exp(-\Lambda_i). \tag{5.17}$$

Using Bayes' rule, and choosing scale invariant priors $\pi(A) = 1/A$ and $\pi(T) = 1/T$, we have

$$\begin{aligned}\ln(p(\mathcal{T}|\mathcal{D})) =&(\mathcal{N}-1)\ln(A) - \ln(T) - A\left(\exp\left(-\frac{V_{B+1}}{T}\right) - \exp\left(-\frac{V_1}{T}\right)\right)\\ &+\sum_{i=1}^{\mathcal{B}} n_i \ln\left(\exp\left(-\frac{V_{i+1}}{T}\right) - \exp\left(-\frac{V_i}{T}\right)\right) + const.\end{aligned} \tag{5.18}$$

We can find the *maximum a posteriori* (MAP) estimate for T and A using the Levenberg-Marquardt algorithm on this expression. Call the MAP estimate $\mathcal{T}^*$. We expand Eq. 5.18 about the MAP estimate to write

$$\ln(p(\mathcal{T}|\mathcal{D})) \approx \ln(p(\mathcal{T}^*|\mathcal{D})) + \frac{1}{2}(\mathcal{T}-\mathcal{T}^*)^\mu(\mathcal{T}-\mathcal{T}^*)^\nu \partial_\mu\partial_\nu \ln(p(\mathcal{T}^*|\mathcal{D})), \tag{5.19}$$

where the derivatives are with respect to $\mathcal{T}$. Hence, we see the the likelihood function is approximately Gaussian about its MAP estimate and a Laplace approximation is valid provided the entries of the matrix $\partial_\mu\partial_\nu \ln(p(\mathcal{T}^*|\mathcal{D}))$ are much less than the corresponding entries in $\mathcal{T}^*$.

Now we consider the choice of the optimal fit interval. We use the fact that if the form in Eq. 5.16 is invalid, this is quantifiable. For reasons described in Sec. 5.3, we expect the signal to predictably become sub-exponential when the assumptions are invalid. Thus, we consider the one-sided p-value of counts for the ith bin, or the events smaller than the number, n_i observed in that bin

$$\wp_i = \sum_{j=0}^{n_i} \frac{\Lambda_i^j}{j!}\exp(-\Lambda_i) = \frac{\Gamma(n_i+1,\Lambda_i)}{n_i!}. \tag{5.20}$$

As the Λ_i depend on T and A, we average this over the posterior for these quantities to calculate

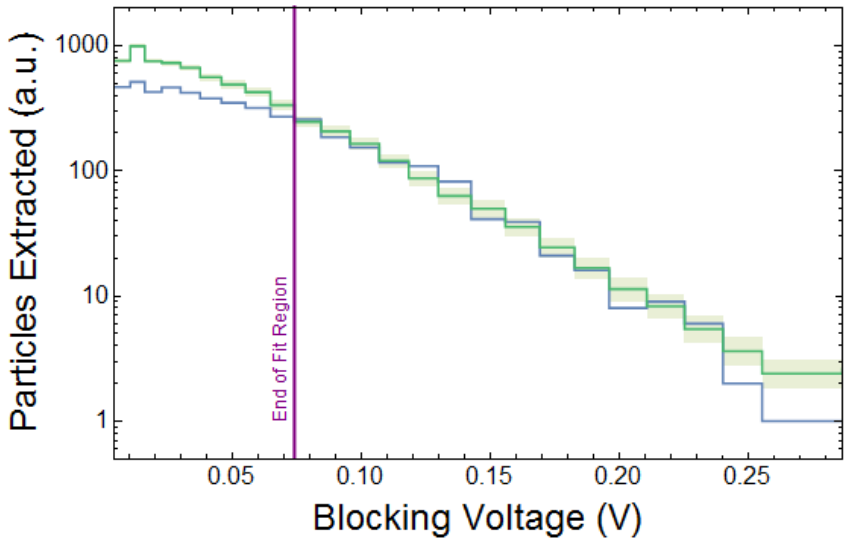

Figure 5.8: (Top) The discrete fitter applied to antiproton extraction data from ALPHA. The blue curve is the detected number as a function of blocking voltage. The green curve is the best fit curve using all bins past the purple line, with $T = 400 \pm 10$ K. The 1σ variation in fit parameters is shown as the partially transparent region around the solid green curve. The optimal fit region shown here is when the fit region is the same as the region consistent with the fit.

$$\langle \wp_i \rangle = \int_0^\infty dA \int_0^\infty dT \; p(\mathcal{T}|\mathcal{D}) \wp_i. \tag{5.21}$$

This integral could be computed numerically, but we find it simpler to compute it using the standard expansion of Gaussian integrals [15] using the form of Eq. 5.19 for $p(\mathcal{T}|\mathcal{D})$. This gives

$$\langle \wp_i \rangle \approx p(\mathcal{T}^*|\mathcal{D}) \frac{2\pi\Gamma(n_i + 1, \Lambda_i(\mathcal{T}^*))}{n_i!\sqrt{\det(-\partial_\mu \partial_\nu p(\mathcal{T}^*|\mathcal{D}))}} \sim \frac{\Gamma(n_i + 1, \Lambda_i(\mathcal{T}^*))}{n_i!}. \tag{5.22}$$

We have extracted the Gamma function and factorial as this is the only part that depends on the bin index, i.

$\langle \wp_i \rangle$ should have values ranging between 0 and 1 when the fit model describes the data in the bin well (i.e. Eq. 5.5 still describes the number of counts in the bin), but will be extremely small when the assumption breaks down. To choose the optimal interval, an initial fit is done with part of the data, comprising of $\mathcal{B}'$ bins. We then take our new fit region to be all data before (but not including) two neighboring bins with $\langle \wp_i \rangle \leq 1/\mathcal{B}'$ with the $\mathcal{T}^*$ from the fit of $\mathcal{B}'$ bins. This new fit region is then used to obtain a new fit. This is repeated until the fit interval chosen by the method no longer changes, or a cycle is reached in the algorithm. As

p-values are approximately uniformly distributed when the null hypothesis is true, we expect the false positive rate by requiring two neighboring bins to satisfy the rejection condition to be $1/\mathcal{B}'^2$, which is small for a reasonable number of bins [63]. A fit with this method is shown in Fig. 5.8. An implementation of the fitter is available in Sec. D.2.

There is a worry that when the chosen fit region becomes too large, the method will converge to a poor fit value. A poor fit will have a large estimated variance for the residuals and hence will allow many more poor points into the fit that are consistent with the null hypothesis. This scenario is avoided by placing a cutoff on the log-likelihood per points in fit, as with the continuous fitter. If the log-likelihood per points in fit is twice the original value, the new fit curve should not be kept and the size of the fit region should be halved.

We need initial guesses for A and T to initialize the Levenberg-Marquardt algorithm. We do this by considering the $\mathcal{N} \to \infty$ limit of Eq. 5.18. For simplicity, assume that the voltage bins are spaced evenly by an amount δV. In this limit, Eq. 5.18 becomes

$$\ln(p(T, A|\mathcal{D})) \approx \mathcal{N} \ln(A) - A \exp\left(-\frac{V_{\mathcal{B}+1}}{T}\right) - \mathcal{N}\frac{\bar{V}}{T} + \mathcal{N} \ln\left(1 - \exp\left(\frac{\delta V}{T}\right)\right) + const. \tag{5.23}$$

where

$$\bar{V} = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{B}} n_i V_{i+1}. \tag{5.24}$$

We take partial derivatives of the expression above and set it equal to zero to analytically get estimates of the optimal T and A in this limit. From this, we get that

$$\begin{aligned} T^* &= \frac{\delta V}{\ln\left(\frac{\bar{V} - V_{\mathcal{B}+1} + \delta V}{\bar{V} - V_{\mathcal{B}+1}}\right)} \overset{\delta V \to 0}{\approx} \bar{V} - V_{\mathcal{B}+1} \\ A^* &= \mathcal{N} \exp\left(\frac{V_{\mathcal{B}+1}}{T^*}\right) \end{aligned} \tag{5.25}$$

The uncertainty in these quantities can be estimated by the second derivative of the log-likelihood evaluated at these maximal points. For A, this gives $\sigma_A = A^*/\sqrt{\mathcal{N}}$, confirming the expected $1/\sqrt{N}$ scaling for a Poisson rate. T appears in an exponential and hence has a more complicated expression. In the $\delta V \to 0$ limit, $\sigma_T = T^*/\sqrt{\mathcal{N}(1 + V_{\mathcal{B}+1}^2/T^{*2})}$. This confirms the $1/\sqrt{\mathcal{N}}$ scaling given earlier for the uncertainty in temperature.

If accuracy is not a primary concern, and $\mathcal{N}$ is sufficiently big (larger than 100), these expressions are sufficient to quickly estimate temperature and uncertainty. This skips the bulk of the time as a nonlinear optimization is not necessary.

Background rate

We modify Eq. 5.16 with a new term b accounting for the background rate

$$\Lambda_i = A\left(\exp\left(-\frac{V_{i+1}}{T}\right) - \exp\left(-\frac{V_i}{T}\right)\right) + b. \tag{5.26}$$

The prior, $\pi(b)$, is taken to contain any known information regarding the distribution of background signal. If the prior knowledge is complicated, such as in the case of Gaussian noise passing through a discriminator, $\pi(b)$ should be estimated by Monte Carlo methods.

We might expect low frequency noise to cause b to change as a function of time (or equivalently voltage). If this is the case, the b in the above expression can be replaced with a bin-specific b_i. Fitting a background rate for each bin will result in too many free parameters, so some model relating each of the bins is ultimately necessary.

5.6 Future Work

We are still developing a robust, fully-automated implementation of the discrete fitter. Once completed, it will be interesting to compare how it performs relative to the continuous fitter. We would expect that when single particle resolution is not available, the two would perform similarly, but the discrete fitter should be able to utilize the single particle information better when it is available. In addition, the size of bins to use is an open question. With these issues sorted out, the method can be used at ALPHA and UC Berkeley.

When methods such as evaporative cooling are employed, it is possible for plasmas to have temperatures below the limit of Eq. 5.11. In these cases, the extraction signal must be fit to a form that includes plasma self-field effects. A fully automated scheme to perform these fits will be a useful future development.

The analysis given here is derived under the assumption that a sudden approximation for particles leaving the confinement well is valid. In doing so, we assume that the plasma distribution does not rethermalize during extraction, and hence we expect the collision frequency to be a relevant parameter. However, experiments show that the collision frequency is an irrelevant timescale for the problem. Formalizing the analysis in such a way that the collision frequency does not appear to be a relevant parameter would be a useful theoretical development.

An open question is to rigorously quantify the effect of adiabatic cooling and to characterize finite length effects. An analytic analysis of the method, as shown in Sec. 5.2, is difficult and thus this question can only be studied with simulations. The plasmas studied here are in a regime ill-suited for particle-in-cell (PIC) methods, as the plasma temperatures are cold enough that the corresponding Debye lengths will make the simulations computationally prohibitive [75–77]. As the Green's function of Poisson's equation can be written analytically in an ideal Penning-Malmberg trap [78], an idealized system could be studied using an N-body simulation with a Barnes-Hut [79] or fast multipole method [80]. However,

we are not aware of attempts to implement or to even derive theoretical results for a Barnes-Hut or fast multipole method in the a cylindrical geometry. Deviations from the cylindrical symmetry, and accounting for differing electrode sizes (as is the case in ALPHA) would also be difficult to incorporate.

Experimentally, it is of interest whether the diagnostic can be used as a non-destructive technique. One could lower the potential up to a point where ~ 100 particles are extracted, and then raise the potential. From these extracted particles, we could potentially estimate the plasma temperature with the plasma still in the trap. This would also be a study of whether the tail of a thermodynamic distribution thermalizes at the same rate as other parts of the distribution, as well as allow for more control in studies of cooling of plasmas.

Acknowledgements

Andrew E. Charman contributed useful discussions regarding the statistics in this chapter. Alexander P. Povilus translated my nonlinear fitter for this particular problem and made a LabView wrapper to aid in the automated fitting. Eric Hunter provided useful guidance with using LabView and experimental techniques, as well as discussions about the statistical analysis. Nathan Belmore worked on the user-interface of the temperature fitter for ALPHA such that it would match the user experience of the old temperature fitting routine. Sabrina Shanman conducted useful simulations to better understand the continuous and discrete fitters. Sabrina Shanman and Cheyenne Nelson worked on the implementation and automation of the discrete fitter. Mike Zhong read this chapter and provided feedback for improvements.

Chapter 6

Cavity Enhanced Cooling

Cold lepton plasmas are imperative to produce trappable antihydrogen. In addition to trapping, the magnetic field of the Penning-Malmberg causes leptons to execute circular, cyclotron orbits, thereby radiating away transverse energy [72]. The methods of Chap. 5 allow us to measure the temperature of plasmas as they cool under this mechanism.

If the plasma is in a cavity with resonances near the cyclotron frequency of the plasma particles, we expect enhanced cooling akin to the Purcell mechanism [36]. The Berkeley plasma trap measured this effect experimentally [5]. In this chapter, we present a theoretical description of cavity enhanced cooling and compare with experimental and simulation results.

We first fill in the holes of the derivations of Ref. [81] and confirm its main results. We then provide a reinterpretation of the results that is useful to understand a longitudinally dynamic plasma.

We show that as the number of plasma particles, $\mathcal{N}$, in the cavity is increased, the plasma cooling rate transitions between two regimes. For small $\mathcal{N}$, the cooling rate is a Lorentzian function of magnetic field about the $\omega_c = \omega$ resonance. For large $\mathcal{N}$, the optimal cooling is not at $\omega_c = \omega$, but when $\omega_c = \omega \pm \omega_z$. The critical $\mathcal{N}$ is determined by the variation in cyclotron frequency across the cavity. We also predict the cooling rate as a function of axial position in the cavity.

6.1 The Experiment

In the measurement, we first trap a reservoir of $\sim 10^8$ electrons outside of the bulge cavity, and then periodically transfer $10^3 - 10^6$ electrons downstream into the bulge cavity. The electron transfer procedure reproducibly initializes the test electrons at a high temperature of ~ 1 eV. The electrons then cool via cyclotron radiation, potentially with cavity enhancements. The confining potentials are then lowered to measure the temperature of the plasma through the methods of Chap. 5. For a given plasma, this procedure takes a few seconds, limited by the cooling time of the plasmas. Thus, it is relatively quick to scan the cooling of plasmas as a function of magnetic field strength, number of plasma particles, the length of

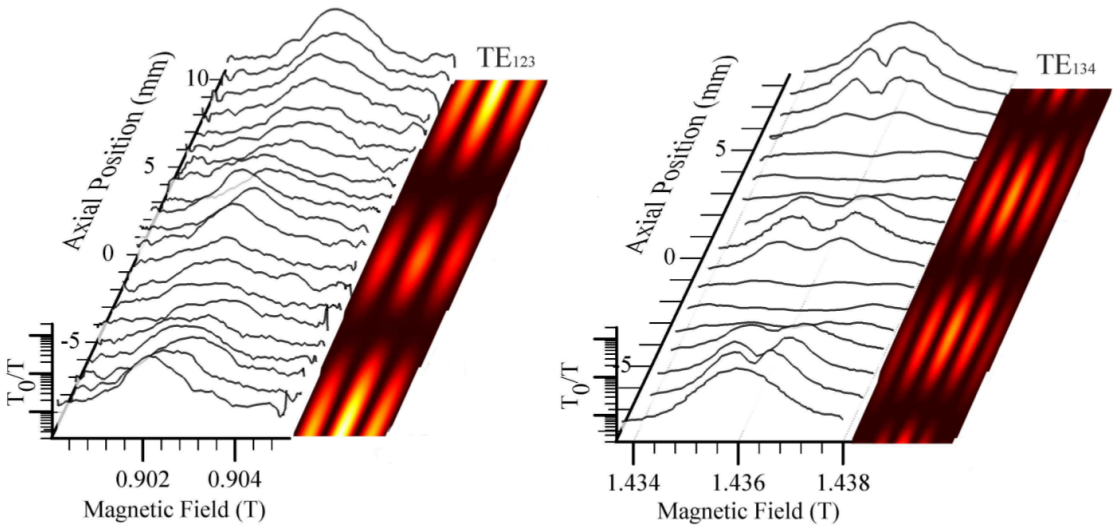

Figure 6.1: Cooling enhancements: (left) TE_{123} with 2×10^4 electrons, (right) TE_{134} with 3×10^5 electrons. For each mode, the left waterfall plot shows the ratio T_0/T as a function of the magnetic field and plasma position (T_0 is the typical off-resonant temperature for each dataset), the right color contour plot shows the mode structure. Figure from Ref. [5].

the plasma, and the center position of the plasma. Fig. 6.1 shows the result of such a scan for two cavity modes.

As will be shown later, in the single-particle picture, we expect that cooling should occur when the particles are in areas where the cavity mode electric field magnitude is large. It was initially surprising that this behavior did not continue when the number of particles in the plasma was increased, as shown in Fig. 6.1. This phenomenon will be explained in this chapter.

6.2 Problem Formulation

We follow Ref. [81] and [82] to derive equations relating the coupling of the plasma particles to the cavity mode. We use cylindrical coordinates (r, θ, z) to describe the geometry, with z pointing along the axis of the solenoidal magnetic field in the Penning-Malmberg trap. Denote the volume of the cavity V. We expand the vector potential in cavity modes $\vec{\psi}_j(\vec{x})$, with boundary conditions for a non-conducting cavity. Each cavity mode has energy $\omega_j |a_j|^2$.

$$\vec{A}_{rad}(\vec{x}, \vec{v}, t) = \sum_j \frac{1}{\sqrt{2V\omega_j}} (a_j \vec{\psi}_j(\vec{x}) + a_j^\dagger \vec{\psi}_j^*(\vec{x})) \tag{6.1}$$

The normalization condition for $\vec{\psi}_i(\vec{x})$ is given by

$$\int d^3\vec{x}\ \vec{\psi}_i(\vec{x}) \cdot \vec{\psi}_j^*(\vec{x}) = \delta_{ij} V, \tag{6.2}$$

making $\vec{\psi}_i(\vec{x})$ a dimensionless quantity. For a cylindrical cavity with TE modes, the modes are characterized by the azimuthal index l, the radial index n, and the axial index k. These modes are typically labeled TE_{lnk} [72]. The functional form is, for a cavity of radius R_w and length L,

$$\vec{\psi}(\vec{x}) = \sqrt{\frac{2}{{x'_{ln}}^2 (J_l^2(x'_{ln}) - J_{l+1}^2(x'_{ln}))}} \left(\frac{ilR_w}{r} J_l\left(\frac{x'_{ln} r}{R_w}\right) \hat{r} + x'_{ln} J'_l\left(\frac{x'_{ln} r}{R_w}\right) \hat{\theta} \right) e^{il\theta} \sin\left(\frac{\pi k z}{L}\right), \tag{6.3}$$

where x'_{ln} is the nth zero of J'_l. The radial extent of the plasmas in the Berkeley plasma trap are quite small with respect to the variation of the cavity modes in the radial direction. As a result, only the $l = 1$ mode will have a strong coupling to the plasma, as $J'_l(0) = 0$ for $l \neq 1$. For a non-cylindrical cavity such as the bulge cavity used in the measurement, $\vec{\psi}(\vec{x})$ will have a different form. These effects are discussed in Sec. 6.5.

Assume the cooling dynamics are dominated by the plasma particles coupling to just one cavity mode. We will thus drop the label j (we will subsequently use j as a particle index). Expand the total vector potential as $\vec{A}_0 + \vec{A}_{rad}$ where $\vec{A}_0$ describes the static, solenoidal $\vec{B}$ field and $\vec{A}_{rad}$ has the radiation effects. Let $\vec{x}^{(j)}$ and $\vec{p}^{(j)}$ denote the position and momentum of the jth particle with mass m and charge $-e$. Assume there are $\mathcal{N}$ such particles. The hamiltonian for the particles and radiation field in the trap is

$$H = \frac{1}{2}\omega(a^\dagger a + a a^\dagger) + \sum_{j=1}^{\mathcal{N}} \frac{(\vec{p}^{(j)} + e\vec{A}_0(\vec{x}^{(j)}) + e\vec{A}_{rad}(\vec{x}^{(j)}))^2}{2m} - e\Phi(\vec{x}^{(j)}), \tag{6.4}$$

where $\Phi(\vec{x})$ satisfies Poisson's equation with appropriate boundary conditions. Treating the particles classically, Hamilton's equations give

$$\begin{aligned} \vec{v}^{(j)} \equiv \dot{\vec{x}}^{(j)} &= \frac{\vec{p}^{(j)} + e\vec{A}_0(\vec{x}^{(j)}) + e\vec{A}_{rad}(\vec{x}^{(j)})}{m} \\ \dot{\vec{p}}^{(j)} &= -e\vec{v}^{(j)} \cdot \vec{\nabla}(\vec{A}_0(\vec{x}^{(j)}) + \vec{A}_{rad}(\vec{x}^{(j)})) + e\vec{\nabla}\Phi(\vec{x}^{(j)}) - e\vec{v}^{(j)} \times \vec{\nabla} \times (\vec{A}_0(\vec{x}^{(j)}) + \vec{A}_{rad}(\vec{x}^{(j)})) \end{aligned} \tag{6.5}$$

Write $\vec{A}_0(\vec{x}) = rB\hat{\theta}/2$ and eliminate $\dot{\vec{p}}$ for $\dot{\vec{v}}$. This gives

$$\dot{\vec{v}}^{(j)} = -\omega_c \vec{v}^{(j)} \times \hat{z} + \frac{e}{m}\vec{\nabla}\Phi(\vec{x}^{(j)}) + \frac{e}{m}\frac{\partial}{\partial t}\vec{A}_{rad}(\vec{x}^{(j)}) - \frac{e}{m}\vec{v}^{(j)} \times \vec{\nabla} \times \vec{A}_{rad}(\vec{x}^{(j)}) \tag{6.6}$$

As would be expected from Newton's law. The equation for the z motion of the particle is

$$\frac{dv_z^{(j)}}{dt} = \frac{e}{m}\frac{\partial\Phi(\vec{x}^{(j)})}{\partial z} - \frac{e}{m}(\vec{v}^{(j)} \times \vec{\nabla} \times \vec{A}_{rad}(\vec{x}^{(j)})) \cdot \hat{z}. \tag{6.7}$$

If we assume that $\Phi(\vec{x})$ is fully specified by its mean field expression and ignore the effect of the final, magnetic term, the motion in z is fully specified. This assumes that the collision rate is slow enough that for timescales relevant to energy loss of the plasma, the trajectories of the particles are not significantly altered.

In the perpendicular direction, we again ignore the magnetic effects due to the radiation field. In addition, we ignore the electric potential as, on average, over a gyroperiod, this term will average to zero. This gives

$$m\frac{d\vec{v}_\perp^{(j)}}{dt} = -\frac{i\omega e}{\sqrt{2V\omega}}a\vec{\psi}(\vec{x}^{(j)}) - m\omega_c\vec{v}_\perp^{(j)} \times \hat{z} \tag{6.8}$$

Define $u = (\hat{r} - i\hat{\theta}) \cdot \vec{v}_\perp$ and $\psi = (\hat{r} - i\hat{\theta}) \cdot \vec{\psi}$. The equation for u is

$$m\frac{du^{(j)}}{dt} = -\frac{i\omega e}{\sqrt{2V\omega}}a\psi(\vec{x}^{(j)}) - im\omega_c u^{(j)} \tag{6.9}$$

We can integrate to write this as

$$u^{(j)}(t) = \exp(-i\omega_c t)\left[u^{(j)}(0) - \frac{i\omega e}{m\sqrt{2V\omega}}\int_0^t dt' \exp(i\omega_c t')\, a(t')\psi(\vec{x}^{(j)}(t'))\right]. \tag{6.10}$$

Now consider the radiation field. The Heisenberg equation for a is

$$\frac{da}{dt} = i[H, a]. \tag{6.11}$$

This is analogous to Hamilton's equation, and thus will give identical results as the classical computation [81]. a commutes with the electrostatic potential, but not the other terms in the hamiltonian. Thus

$$\begin{aligned}\frac{da}{dt} &= -i\omega a - \frac{ie}{\sqrt{2V\omega}}\sum_{j=1}^{\mathcal{N}}\vec{\psi}^*(\vec{x}^{(j)}) \cdot \vec{v}^{(j)} \\ &= -i\omega a - \frac{ie}{2\sqrt{2V\omega}}\sum_{j=1}^{\mathcal{N}}\left(\psi^*(\vec{x}^{(j)})u^{(j)} + \psi(\vec{x}^{(j)})u^{(j)^*}\right)\end{aligned} \tag{6.12}$$

Having a conductive cavity is equivalent to taking $\omega \to \omega - \frac{\omega}{2Q} - i\frac{\omega}{2Q}$, with Q the quality factor of the cavity [72]. We redefine $\omega \to \omega(1 - 1/2Q)$, as the original ω is not a measurable quantity for the experiment. With this change, to order $1/Q$, the complex cavity frequency is $\omega - i\frac{\omega}{2Q}$. We ignore the u^* term in Eq. 6.12 as it counter-propagates with a. Thus,

$$\frac{da}{dt} = -\left(i\omega - \frac{\omega}{2Q}\right) a - \frac{ie}{2\sqrt{2V\omega}} \sum_{j=1}^{\mathcal{N}} \psi^*(\vec{x}^{(j)}) u^{(j)} \tag{6.13}$$

Together with Eq. 6.9, this defines a set of $\mathcal{N}+1$ oscillator equations. These equations can be simulated using a standard numerical integrator [2,83]. We take a different approach and consider analytic approaches to the problem. Plugging in Eq. 6.10 to Eq. 6.13 gives

$$\begin{aligned}\frac{da}{dt} = &- \frac{ie}{2\sqrt{2V\omega}} \sum_{j=1}^{\mathcal{N}} \psi^*(\vec{x}^{(j)}) u^{(j)}(0) e^{-i\omega_c t} - \left(i\omega + \frac{\omega}{2Q}\right) a \\ &- \frac{e^2}{4mV} \sum_{j=1}^{\mathcal{N}} \psi^*(\vec{x}^{(j)}) \int_0^t dt'\ e^{-i\omega_c (t-t')} a(t') \psi(\vec{x}^{(j)}(t')).\end{aligned} \tag{6.14}$$

The cavity modes will vary on a length scale comparable to the dimensions of the cavity. Thus, the variation of a particle about its guiding center coordinate should have a negligible contribution to the variation of ψ. Hence, we assume that the time variation in ψ is dominated by the z motion of the particles. As argued previously, for times less than a collision time, we expect the z motion to be fully specified by choosing a form for $\Phi(\vec{x})$. This reduces the $\mathcal{N}+1$ oscillator equations into just 1 equation, albeit a complicated integro-differential equation. We will consider various cases of Eq. 6.14.

6.3 A longitudinally static plasma

In this section we reconsider the analysis of Ref. [81], in which particles are assumed static in z. This will provide insights into the more general case of bouncing plasmas.

We allow all of the particles to have differing cyclotron frequencies. As long as the maximal variation in the cyclotron frequency across the plasma ($\Delta\omega_c$) is much smaller than the average cyclotron frequency, this can be done without significantly altering the derivation given in Sec. 6.2. We also assume that $\Delta\omega_c$ is not so large that it introduces coupling to multiple cavity modes, which would also make the derivation of Sec. 6.2 invalid. Eq. 6.14 becomes

$$\frac{da}{dt} = -\sqrt{\frac{m}{\omega}} \sum_{j=1}^{\mathcal{N}} \hat{\beta}_j u^{(j)}(0) e^{-i\omega_c^{(j)} t} - \left(i\omega + \frac{\omega}{2Q}\right) a - \sum_{j=1}^{\mathcal{N}} |\hat{\beta}_j|^2 \int_0^t dt'\ e^{-i\omega_c^{(j)}(t-t')} a(t'). \tag{6.15}$$

Where for clarity we have defined

$$\hat{\beta}^{(j)} = \frac{e}{2\sqrt{2mV}} \psi^*(\rho_{gc}^{(j)}, \theta_{gc}^{(j)}, z^{(j)}) \tag{6.16}$$

The hat above β will be used to distinguish this β from a related but different expression in the dynamic case. As the integral is a convolution, the equation can be solved by Laplace transform. Doing this yields

$$\tilde{a}(s) = \frac{a(0) - \sqrt{\frac{m}{\omega}} \sum_{j=1}^{\mathcal{N}} \frac{\hat{\beta}_j u^{(j)}(0)}{s + i\omega_c^{(j)}}}{s + i\omega + \frac{\omega}{2Q} - \sum_{j=1}^{\mathcal{N}} \frac{|\hat{\beta}_j|^2}{s + i\omega_c^{(j)}}} \tag{6.17}$$

By the Laplace inversion formula, solving Eq. 6.15 reduces to finding the poles of the above expression [84]. Through basic manipulations, this is equivalent to finding the roots of

$$\left(s + i\omega + \frac{\omega}{2Q}\right) \prod_{j=1}^{\mathcal{N}} (s + i\omega_c^{(j)}) - \sum_{j=1}^{\mathcal{N}} |\hat{\beta}_j|^2 \prod_{k \neq j} (s + i\omega_c^{(k)}) = 0. \tag{6.18}$$

The imaginary parts of the solutions to Eq. 6.18, s^*, correspond to the frequency of the phase evolution of each term that contributes to $a(t)$. The real parts of s^* correspond to the decay time constant of each term that contributes to $a(t)$. Note that Eq. 6.18 shows that initial conditions have no effect on the decay rates of $a(t)$, though they can affect which terms are important in the Laplace inversion.

The weakly coupled plasma

When $\mathcal{N}\langle|\beta|^2\rangle \ll \omega^2/4Q^2$ (where $\langle\rangle$ denotes a particle average), the first term in Eq. 6.18 dominates, and the $\mathcal{N}+1$ roots are approximately the $\mathcal{N}$ roots $s_j^* \approx -i\omega_c^{(j)}$ for each particle and $s^* \approx -i\omega - \omega/2Q$. Perturbatively including the effect of the second term in Eq. 6.18 shows that the purely imaginary roots, s_j^*, pick up a correction

$$s_j^* = -i\omega_c^{(j)} + \frac{i|\hat{\beta}_j|^2}{(\omega - \omega_c^{(j)}) - i\frac{\omega}{2Q}} \tag{6.19}$$

The real part of these are small under the assumption $\mathcal{N}\langle|\beta|^2\rangle \ll \omega^2/2Q$, and thus, these solutions can lead to cooling of the plasma for long times. The other solution picks up a correction

$$s^* = -i\omega - \frac{\omega}{2Q} - \sum_{j=1}^{\mathcal{N}} \frac{i|\hat{\beta}_j|^2}{(\omega - \omega_c^{(j)}) - i\frac{\omega}{2Q}} \tag{6.20}$$

The real part of this expression is not significantly altered by the correction. Since $2Q/\omega$ is a short time, this solution does not contribute to cooling for long times, and here we will consider it an ignorable transient. Applying the Laplace inversion formula to Eq. 6.17 shows that the leading order contribution to the cavity mode is thus (assuming $a(0) = 0$),

$$a(t) = \sum_{j=1}^{\mathcal{N}} \frac{\sqrt{\frac{m}{\omega}}\hat{\beta}_j u^{(j)}(0) e^{-i\omega_c^{(j)} t}}{\left(i(\omega - \omega_c^{(j)}) + \frac{\omega}{2Q}\right)}. \tag{6.21}$$

We find

$$|a(t)|^2 = \frac{m}{\omega}\left|\sum_{j=1}^{\mathcal{N}} \frac{\hat{\beta}_j u^{(j)}(0) e^{-i\omega_c^{(j)} t}}{\left(i(\omega - \omega_c^{(j)}) + \frac{\omega}{2Q}\right)}\right|^2 \approx \frac{m}{\omega}\sum_{j=1}^{\mathcal{N}} \frac{|\hat{\beta}_j|^2 |u^{(j)}(t)|^2}{(\omega - \omega_c^{(j)})^2 + \frac{\omega^2}{4Q^2}}. \tag{6.22}$$

We assume that over timescales relevant to cooling, the difference in cyclotron frequencies will make off-diagonal terms in the sum zero for the second approximate equality. This can also be interpreted as looking at the DC component of the cavity mode energy for the purposes of the following argument. In addition, we may expect that the off-diagonal terms will average to zero without a spread in cyclotron frequencies, as the random phases in $u^{(j)}(0)$ should make the average of the off-diagonal terms zero. We have also used that to this level of approximation, $u^{(j)}(0)e^{-i\omega_c^{(j)} t} = u^{(j)}(t)$. Replace the sum by a particle average. We assume the $\hat{\beta}_j$ (depending on particle positions) are independent of the velocities $u^{(j)}(t)$. We also define $T_\perp \equiv m\langle |u(t)|^2\rangle/2$ to get

$$U_{cav} \equiv \omega |a(t)|^2 = 2T_\perp \left\langle \frac{\mathcal{N}|\hat{\beta}|^2}{(\omega - \omega_c)^2 + \frac{\omega^2}{4Q^2}} \right\rangle \tag{6.23}$$

For a nonzero DC component of the energy to exist, there must be an energy balance between the energy lost to the walls and the energy provided to the radiation field. Energy is absorbed by the walls at a rate $\omega U_{cav}/Q$, so it must be that $\omega U_{cav}/Q$ of energy is being supplied from elsewhere. The only source for this is the plasma, so the plasma must be losing energy at this rate. Putting this together gives a cooling rate

$$\Gamma_\perp \equiv -\frac{1}{\mathcal{N}T_\perp}\frac{d\mathcal{N}T_\perp}{dt} = \frac{\omega U_{cav}}{\mathcal{N}QT_\perp} = \frac{2\omega}{Q}\left\langle \frac{|\hat{\beta}|^2}{(\omega - \omega_c)^2 + \frac{\omega^2}{4Q^2}} \right\rangle. \tag{6.24}$$

In this limit, optimal cooling is achieved on resonance, $\omega = \langle\omega_c\rangle$, and when the maximal spread in cyclotron frequencies, $\Delta\omega_c \ll \omega/2Q$. This cooling is similar to the single particle cooling seen in Ref. [85], and thus we refer to this as the single particle cooling regime. Note that this answer could have also been obtained by looking at the real part of Eq. 6.19, as this is the timescale that cavity radiation exists in the cavity, and cooling cannot happen when there is no radiation field inside the cavity.

In Fig. 6.2 we show this cooling rate as a function of magnetic field and plasma position (holding the length fixed) inside a cylindrical cavity. Comparing to Fig. 6.1, the features in the TE_{123} plot are similar, but the TE_{134} plot does not show splitting and peaks at different axial locations than the experimental observations.

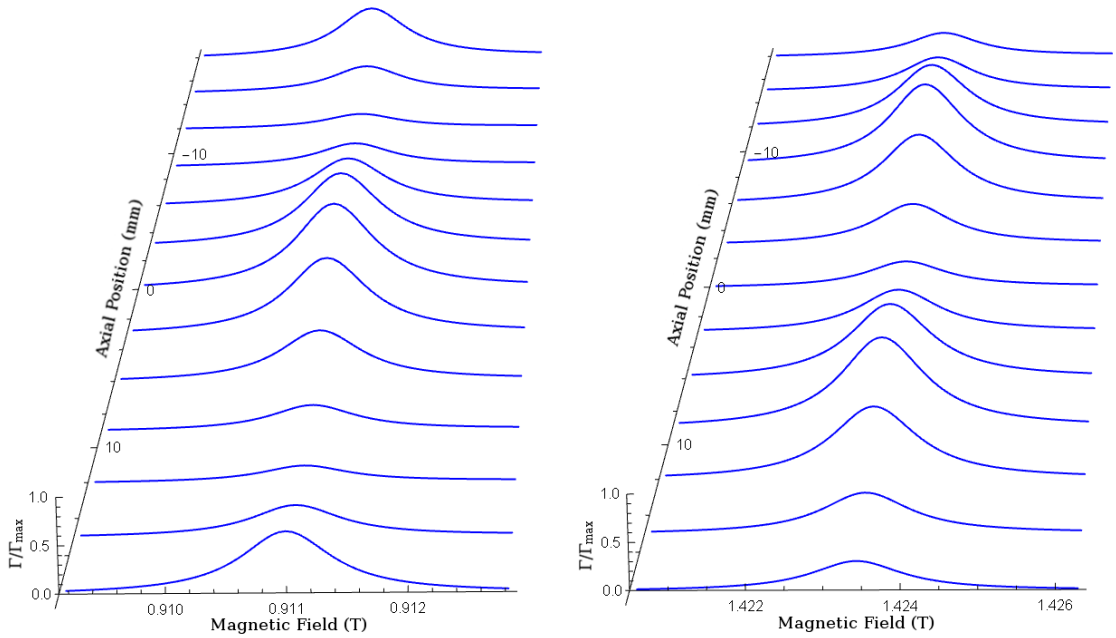

Figure 6.2: Cooling enhancements for (left) TE_{123} and (right) TE_{134} calculated from Eq. 6.24. We use the cavity dimensions of the Berkeley plasma trap with the cavity mode of form Eq. 6.3 to calculate the cooling rate. We assume that electrons are thermally distributed with $T = 10000$ K in z, under a potential $e\Phi(z) = m\omega_z^2 z^2/2$, and $\omega_z = 40\pi \times 10^6$ s^{-1}. We set the cavity $Q = 1000$. We assume the spread in cyclotron frequencies $\Delta\omega_c \ll \omega/2Q$, and that the plasma has negligible radius to calculate the cooling rate. $\Gamma_{max} = 49$ s^{-1} for the left plot and $\Gamma_{max} = 69$ s^{-1} for the right.

The strongly coupled plasma

As $\mathcal{N}\langle|\beta|^2\rangle$ grows large, the contribution from the second term in Eq. 6.18 becomes sizable and can no longer be treated perturbatively. Consider the solutions to this second term,

$$\sum_{j=1}^{\mathcal{N}} |\hat{\beta}_j|^2 \prod_{k\neq j}(s + i\omega_c^{(k)}) = 0. \tag{6.25}$$

We will look at perturbations about the solutions to this polynomial. Note that the polynomial in Eq. 6.25 is an order $\mathcal{N}-1$ polynomial, and hence has $\mathcal{N}-1$ roots. Assume that all roots of Eq. 6.25 are purely imaginary. We then see that i factors out and the resulting polynomial has only positive coefficients. We notice that this form is the Lagrange interpolating polynomial for points $\{(-\omega_c^{(j)}, |\hat{\beta}_j|^2 \prod_{k\neq j}(\omega_c^{(j)} - \omega_c^{(k)}))|j \in \{1, ..., \mathcal{N}-1\}\}$ [86]. In this collection of points, if the $\omega_c^{(j)}$ are ordered from smallest to largest, the ordinate will

flip signs after each increment of j. By the intermediate value theorem, this means that the polynomial must have $\mathcal{N}-1$ real roots [48]. Thus, our initial assumption of purely imaginary roots was valid. This also indicates that the magnitude of each of the $\mathcal{N}-1$ roots is between successive $\omega_c^{(j)}$.

Let $-i\omega_*^{(k)}$ denote the $\mathcal{N}-1$ roots of Eq. 6.25. Take a perturbation of $-i\omega_*^{(k)}$ and insert into Eq. 6.18 to find

$$s_k^* = -i\omega_*^{(k)} + \frac{\left(i(\omega-\omega_*^{(k)})+\frac{\omega}{2Q}\right)\prod_{j=1}^{\mathcal{N}} i(\omega_c^{(j)}-\omega_*^{(k)})}{\sum_{j=1}^{\mathcal{N}}|\hat{\beta}_j|^2\sum_{l\neq j}\prod_{i\neq j,l} i(\omega_c^{(i)}-\omega_*^{(k)})}. \tag{6.26}$$

In particular, through the manipulation of the above equation, and defining $\Delta\omega_c$ as the maximal difference in cyclotron frequencies across the plasma, we can bound

$$-\mathrm{Re}(s_k^*) \leq \frac{\omega\Delta\omega_c^2}{2Q\mathcal{N}(\mathcal{N}-1)\langle|\hat{\beta}|^2\rangle}, \tag{6.27}$$

and thus in this limit, these solutions predict cooling for long times (that is, seconds), provided $\Delta\omega_c$ is smaller than $\sqrt{\mathcal{N}\langle|\hat{\beta}|^2\rangle}$. This bounds the plasma cooling rate as when there is no radiation field in the cavity, the plasma can no longer cool.

There are two roots remaining in Eq. 6.18, and for these we consider the other term in Eq. 6.18. Dividing through by our factorization of Eq. 6.25 yields

$$\left(s+i\omega+\frac{\omega}{2Q}\right)\frac{\prod_{j=1}^{\mathcal{N}}(s+i\omega_c^{(j)})}{\prod_{l=1}^{\mathcal{N}-1}(s+i\omega_*^{(l)})} - \mathcal{N}\langle|\beta|^2\rangle \approx \left(s+i\omega+\frac{\omega}{2Q}\right)(s+i\bar{\omega}_c) - \mathcal{N}\langle|\hat{\beta}|^2\rangle = 0. \tag{6.28}$$

Here, $s+i\bar{\omega}_c$ is the result of the polynomial long division of the ratio of two polynomials in the first term.

This can be solved to give

$$s_\pm = \frac{1}{2}\left(-i(\omega+\bar{\omega}_c) - \frac{\omega}{2Q} \pm \sqrt{\left(i(\omega-\bar{\omega}_c)+\frac{\omega}{2Q}\right)^2 + 4\mathcal{N}\langle|\hat{\beta}|^2\rangle}\right). \tag{6.29}$$

In the limit $N\langle|\hat{\beta}|^2\rangle \gg \omega^2/4Q^2$, we see that both of these solutions decay rapidly, and we will consider them as ignorable transients.

Using the poles of Eq. 6.25 and applying the Laplace inversion formula to Eq. 6.17 shows that the leading order contribution to the cavity mode is thus (assuming $a(0)=0$),

$$a(t) = \sqrt{\frac{m}{\omega}} \sum_{l=1}^{\mathcal{N}-1} \frac{\sum_{j=1}^{\mathcal{N}} \hat{\beta}_j u^{(j)}(0) \prod_{k\neq j} (\omega_c^{(k)} - \omega_*^{(l)}) e^{-i\omega_*^{(l)} t}}{\left(\left(i\frac{\omega}{2Q} - (\omega - \omega_*^{(l)})\right)(\bar{\omega}_c - \omega_*^{(l)}) - \mathcal{N}\langle|\hat{\beta}|^2\rangle\right) \prod_{k\neq l}(\omega_*^{(k)} - \omega_*^{(l)})}. \tag{6.30}$$

The dominant contribution to the inner sum (indexed by j) is when the product in the numerator does not include terms such that $\omega_c^{(k)} \approx \omega_*^{(l)}$. Hence, the term is maximal when $j \approx l$. We confirm this intuition by choosing polynomial and exponential variations of ω_c for $\mathcal{N} \sim 100$ particles and find that this is indeed the case. Using this to simplify gives

$$a(t) = \sqrt{\frac{m}{\omega}} \sum_{l=1}^{\mathcal{N}-1} \frac{\hat{\beta}_j u^{(j)}(0) \prod_{k\neq l} (\omega_c^{(k)} - \omega_*^{(l)}) e^{-i\omega_*^{(l)} t}}{\left(\left(i\frac{\omega}{2Q} - (\omega - \omega_*^{(l)})\right)(\bar{\omega}_c - \omega_*^{(l)}) - \mathcal{N}\langle|\hat{\beta}|^2\rangle\right) \prod_{k\neq l}(\omega_*^{(k)} - \omega_*^{(l)})} \tag{6.31}$$

Again, for $\mathcal{N} \sim 100$ particles and polynomial and exponential profiles of magnetic field across the cavity, we find this implies $\omega|a(t)|^2 \propto \Delta\omega_c^2$, where $\Delta\omega_c$ is the spread in cyclotron frequency across the plasma (see also Eq. 6.27). In the Berkeley plasma trap, we expect $\Delta\omega_c^2$ to be small, and hence suppressed cooling rates are observed as $\mathcal{N}$ is increased.

Connecting the Limits

We can estimate the transition region by asymptotically matching the strongly coupled cooling rate (we use the bound of Eq. 6.27) with the weakly coupled cooling rate, Eq. 6.24 [87]. By doing this, we (approximately) find the critical $\mathcal{N}$ such that we transition from one regime to the other. Assuming resonance such that $\langle\omega_c\rangle = \omega$, we find

$$\mathcal{N}_{critical} \sim \frac{\omega\Delta\omega_c}{2Q\langle|\beta|^2\rangle}. \tag{6.32}$$

We compare cooling rates for a varying number of particles for the TE_{111} mode in the Berkeley plasma trap in Fig. 6.3. The experiment finds, in this case, $\mathcal{N}_{critical} \approx 5 \times 10^5$ particles. Using our expression above, this implies $\Delta\omega_c/\omega_c \approx 3 \times 10^{-5}$. The value of $\Delta\omega_c$ cannot be measured, but this value is plausible.

Now consider the decay rate of total energy of the cavity, or $\mathcal{N}T_\perp\Gamma_\perp$. Notice from Eq. 6.27 that in the large $\mathcal{N}$ limit this decay rate decreases with increasing $\mathcal{N}$. Further Eq. 6.24 shows that in the small $\mathcal{N}$ limit, this decay rate increases with increasing $\mathcal{N}$. Hence, maximal cooling is achieved when $\mathcal{N} \approx \mathcal{N}_{critical}$. At this value, we find

$$\Gamma_{max} = \frac{8Q\langle|\hat{\beta}|^2\rangle}{\omega}. \tag{6.33}$$

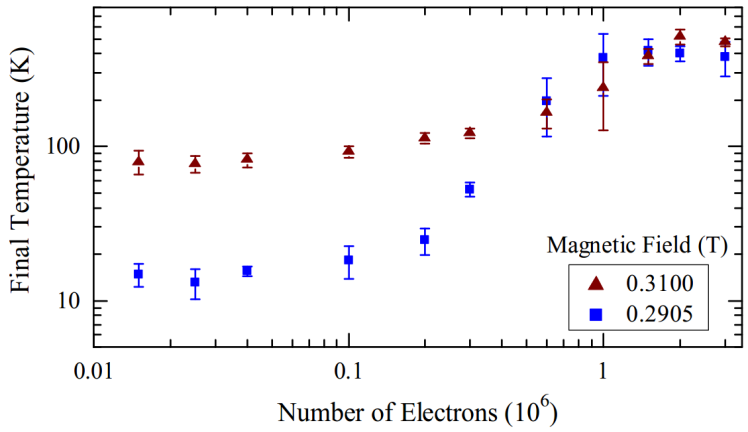

Figure 6.3: quilibrium temperature of plasmas with $10^4 - 10^6$ electrons. In the dark red dataset the field is detuned 19.5 mT (546 MHz) from the TE_{111} resonance. Figure from Ref. [5].

This matches the scaling of Ref. [81] when $\omega/Q \sim \sqrt{\mathcal{N}\langle|\hat{\beta}|^2\rangle}$. By Eq. 6.32, we also find this implies $\sqrt{\mathcal{N}\langle|\hat{\beta}|^2\rangle} \sim \Delta\omega_c$, again confirming the scaling of Ref. [81].

Superradiance

While the longitudinally static plasma problem was simple enough to be solved by Laplace transform, we will find that for the dynamic case a similar approach becomes intractable. We will provide a reinterpretation of the previous results with useful ideas to extend to the more general case.

We can multiply Eq. 6.13 by a^* to obtain the time rate of change of energy in the radiation field

$$\frac{dU_{cav}}{dt} = -\frac{\omega U_{cav}}{Q} - \frac{ie\sqrt{\omega}}{\sqrt{2V}} \sum_{j=1}^{\mathcal{N}} \left(\psi^*(\vec{x}^{(j)}) a^* u^{(j)} - \psi(\vec{x}^{(j)}) a u^{(j)*} \right). \tag{6.34}$$

The plasma can exchange energy with the cavity mode only when the last term is nonzero. We will refer to this term as the change in energy of the superradiant mode, or dU_{SR}/dt.

In particular, if $a(t) \propto \exp(i\omega t)$ and $u^{(j)}(t) = u^{(j)}(0)\exp(i\omega_c^{(j)} t)$, we find that

$$\frac{dU_{SR}}{dt} \propto \langle \mathrm{Im}\left(u(0)\psi(\vec{x})\exp(i(\omega_c - \omega)t)\right)\rangle. \tag{6.35}$$

In this static plasma approximation, $\vec{x}$ is unchanging. Thus once zero, the only way dU_{SR}/dt can become nonzero in a longitudinally static plasma is to have a spread in ω_c across the

plasma particles. This shows that a spread in cyclotron frequencies is necessary for cooling (but does not constrain its magnitude).

We will replace the sums in Eq. 6.17 by averaged quantities. We will refer to the cooling rates obtained from this as bulk cooling rates. This obviously assumes that there are enough particles that a thermodynamic average can be calculated. We further assume $\Delta\omega_c \ll \omega/2Q$ to simplify the expressions. The polynomial to solve for bulk cooling rates, the analog of Eq. 6.18, is

$$\left(s + i\omega + \frac{\omega}{2Q}\right)(s + i\langle\omega_c\rangle) - \mathcal{N}\langle|\hat{\beta}|^2\rangle = 0. \tag{6.36}$$

The solution to this quadratic equation is

$$s_\pm^* = -i\left(\frac{\omega + \langle\omega_c\rangle}{2}\right) - \frac{\omega}{4Q} \pm \sqrt{\left(i\left(\frac{\omega - \langle\omega_c\rangle}{2}\right) + \frac{\omega}{4Q}\right)^2 + \mathcal{N}\langle|\hat{\beta}|^2\rangle} \tag{6.37}$$

When $\mathcal{N}\langle|\hat{\beta}|^2\rangle \ll \omega^2/4Q^2$, we have

$$\mathrm{Re}(s_+^*) \approx \frac{\omega}{2Q}\frac{\mathcal{N}\langle|\hat{\beta}|^2\rangle}{(\omega - \omega_c)^2 + \frac{\omega^2}{4Q^2}} \tag{6.38}$$

Which is $\mathcal{N}/4$ times the single particle cooling rate, Eq. 6.24. This suggests, but by no means proves, that in this limit, cooling is constrained by the ability of the superradiant mode to release energy.

As $\mathcal{N}$ gets larger, the overall cooling rate is no longer limited by the bulk rates in Eq. 6.38, but by how much energy can be put into the superradiant mode. Eq. 6.35 shows that this is only achieved by a spread in cyclotron frequencies. The rate of energy entering the superradiant mode is equal to the energy out if $\Delta\omega_c \approx 2Q\mathcal{N}\langle|\hat{\beta}|^2\rangle/\omega$. This places a limit on the number of particles that can be cooled at

$$\mathcal{N}_{max} \sim \frac{\omega\Delta\omega_c}{2Q\langle|\hat{\beta}|^2\rangle}. \tag{6.39}$$

This is precisely the critical number we found in Eq. 6.32, giving credence to the preceding argument. For $\mathcal{N} > \mathcal{N}_{max}$, energy is taken out of the superradiant mode faster than it can be supplied, and explains why in this limit, the cooling rate becomes proportional to $\Delta\omega_c$, as in Eq. 6.31.

6.4 A longitudinally dynamic plasma

Now assume that the $z^{(j)}(t)$ motion is harmonic. This would occur if the potential in Eq. 6.7 is harmonic. We write

$$z^{(j)}(t) = L\left(\alpha + \zeta_0^{(j)}\sin(\omega_z t + \phi^{(j)})\right) \tag{6.40}$$

and plug this form into $\psi(\vec{x}^{(j)})$. Here, α denotes the normalized center coordinate of the plasma, $\zeta_0^{(j)}$ denotes the amplitude of the oscillation, related to the particle's parallel energy, ω_z is the bounce frequency, and $\phi^{(j)}$ is the initial bounce phase of the particle. Eq. 6.14 becomes

$$\begin{aligned}\frac{da}{dt} =& -i\sqrt{\frac{m}{\omega}}\sum_{j=1}^{\mathcal{N}}\beta^{(j)}\sin\left(\pi k\left(\alpha+\zeta_0^{(j)}\sin(\omega_z t+\phi^{(j)})\right)\right)u^{(j)}(0)e^{-i\omega_c t} - \left(i\omega+\frac{\omega}{2Q}\right)a \\ &-\sum_{j=1}^{\mathcal{N}}|\beta^{(j)}|^2\int_0^t dt'\; e^{-i\omega_c(t-t')}a(t')\sin\left(\pi k\left(\alpha+\zeta_0^{(j)}\sin(\omega_z t+\phi^{(j)})\right)\right) \\ &\qquad\times\sin\left(\pi k\left(\alpha+\zeta_0^{(j)}\sin(\omega_z t'+\phi^{(j)})\right)\right),\end{aligned} \tag{6.41}$$

where $\beta^{(j)}$ is defined as

$$\beta^{(j)} = \frac{e}{2\sqrt{2mV}}\psi^*(\rho_{gc}^{(j)}, \theta_{gc}^{(j)}, :), \tag{6.42}$$

where the colon in the argument on ψ denotes that the z dependent part of ψ (that is, $\sin(\pi k z/L)$) is divided out.

We will determine the bulk decay rates of the plasma by averaging over the initial bounce phase, assuming that they are initially uniform. As before, we expect this to be proportional to the plasma cooling rate when $\mathcal{N}\langle|\beta|\rangle^2$ is not large. Doing this gives [88]

$$\begin{aligned}\frac{da}{dt} =& -i\mathcal{N}\sqrt{\frac{m}{\omega}}\sin(\pi k\alpha)\langle\beta J_0(\pi k\zeta_0)u^{(j)}(0)\rangle e^{-i\omega_c t} - \left(i\omega+\frac{\omega}{2Q}\right)a \\ &-\frac{\mathcal{N}}{2}\int_0^t dt'\;\left\langle|\beta|^2 J_0\left(2\pi k\zeta_0\sin\left(\frac{\omega_z(t-t')}{2}\right)\right)\right\rangle e^{-i\omega_c(t-t')}a(t') \\ &+\frac{\mathcal{N}}{2}\cos(2\pi k\alpha)\int_0^t dt'\;\left\langle|\beta|^2 J_0\left(2\pi k\zeta_0\cos\left(\frac{\omega_z(t-t')}{2}\right)\right)\right\rangle e^{-i\omega_c(t-t')}a(t').\end{aligned} \tag{6.43}$$

We then perform the average over the bounce amplitudes $\zeta_0^{(j)}$. Assuming a Maxwellian distribution, the probability distribution for ζ_0 is

$$p(\zeta_0) = \frac{\zeta_0}{\zeta_{th}^2}\exp\left(-\frac{\zeta_0^2}{2\zeta_{th}^2}\right), \tag{6.44}$$

where $\zeta_{th}^2 = T/m\omega_z^2 L^2$. Note that for $T < 1$ eV, $\zeta_{th} \ll 1$. Using this in the average yields [70]

$$\begin{aligned}\frac{da}{dt} =& -i\mathcal{N}\sqrt{\frac{m}{\omega}}\sin(\pi k\alpha)\langle\beta J_0(\pi k\zeta_0)u(0)\rangle e^{-i\omega_c t} - \left(i\omega + \frac{\omega}{2Q}\right)a \\ &- \frac{\mathcal{N}\langle|\beta|^2\rangle}{2}\int_0^t dt'\ \exp\left(-2\pi^2k^2\zeta_{th}^2\sin^2\left(\frac{\omega_z(t-t')}{2}\right) - i\omega_c(t-t')\right)a(t') \\ &+ \frac{\mathcal{N}\langle|\beta|^2\rangle}{2}\cos(2\pi k\alpha)\int_0^t dt'\ \exp\left(-2\pi^2k^2\zeta_{th}^2\cos^2\left(\frac{\omega_z(t-t')}{2}\right) - i\omega_c(t-t')\right)a(t').\end{aligned} \tag{6.45}$$

This is still a convolution, and thus Laplace transforms can be used to determine the solution. However, the Laplace transform of the exponential of a trigonometric function squared function has no analytic solution. We use the fact that $\pi k\zeta_{th} \ll 1$ and expand the exponentials.

$$\begin{aligned}s\tilde{a}(s) - a(0) =& \frac{-i\mathcal{N}\sqrt{\frac{m}{\omega}}\sin(\pi k\alpha)\langle\beta J_0(\pi k\zeta_0)u(0)\rangle}{s+i\omega_c} - \left(i\omega + \frac{\omega}{2Q}\right)\tilde{a}(s) \\ &- i\frac{\mathcal{N}\langle|\beta|^2\rangle}{2}\frac{\tilde{a}(s)}{s+i\omega_c}{}_2F_2\left(\frac{1}{2},1;1-i\frac{s+i\omega_c}{\omega_z},1+i\frac{s+i\omega_c}{\omega_z};-2\pi^2k^2\zeta_{th}^2\right) \\ &+ i\frac{\mathcal{N}\langle|\beta|^2\rangle}{2}\cos(2\pi k\alpha)\tilde{a}(s)e^{\frac{\pi s}{\omega_z}}\sum_{\iota=0}^{\infty}\frac{B_{-1}(-n+i\frac{s+i\omega_c}{\omega_z},1+2n)\left(\frac{\pi k\zeta_{th}}{\sqrt{2}}\right)^{2\iota}}{\omega_z n!}\end{aligned} \tag{6.46}$$

Here, ${}_2F_2$ is a hypergeometric function and B_{-1} denotes the incomplete beta function [70]. We again look for the poles of $\tilde{a}(s)$, the real part of which gives us the cooling rate of the superradiant mode by the Laplace inversion formula. Expanding to $O(\pi^2k^2\zeta_{th}^2)$, this is equivalent to the solutions of

$$\begin{aligned}0 =& -\frac{\mathcal{N}\langle|\beta|^2\rangle}{2}\cos(2\pi k\alpha)\left(\omega_z^2\left(1-\pi^2k^2\zeta_{th}^2\right) + \left(1-2\pi^2k^2\zeta_{th}^2\right)(s+i\omega_c)^2\right) \\ &+ \frac{\mathcal{N}\langle|\beta|^2\rangle}{2}\left(\omega_z^2\left(1-\pi^2k^2\zeta_{th}^2\right) + (s+i\omega_c)^2\right) + \left(\frac{\omega}{2Q}+s+i\omega\right)\left(\omega_z^2(s+i\omega_c) + (s+i\omega_c)^3\right).\end{aligned} \tag{6.47}$$

We solve for the four roots numerically [70]. The dominant contribution peaks when $\omega_c = \omega$, but the $O(\pi^2k^2\zeta_{th}^2)$ contribution introduces poles with real components that peak at a detuning of $\pm\omega_z$ from the $\omega = \omega_c$ resonance. In general, the $O(\pi^{2n}k^{2n}\zeta_{th}^{2n})$ terms introduce poles with real part detuned $\pm n\omega_z$ from the $\omega_c = \omega$ resonance. This explains the splitting of the resonances in Fig. 6.1, and can be interpreted as the plasma seeing a Doppler shifted cavity frequency as it bounces in the trap. This has been observed previously in Ref. [85].

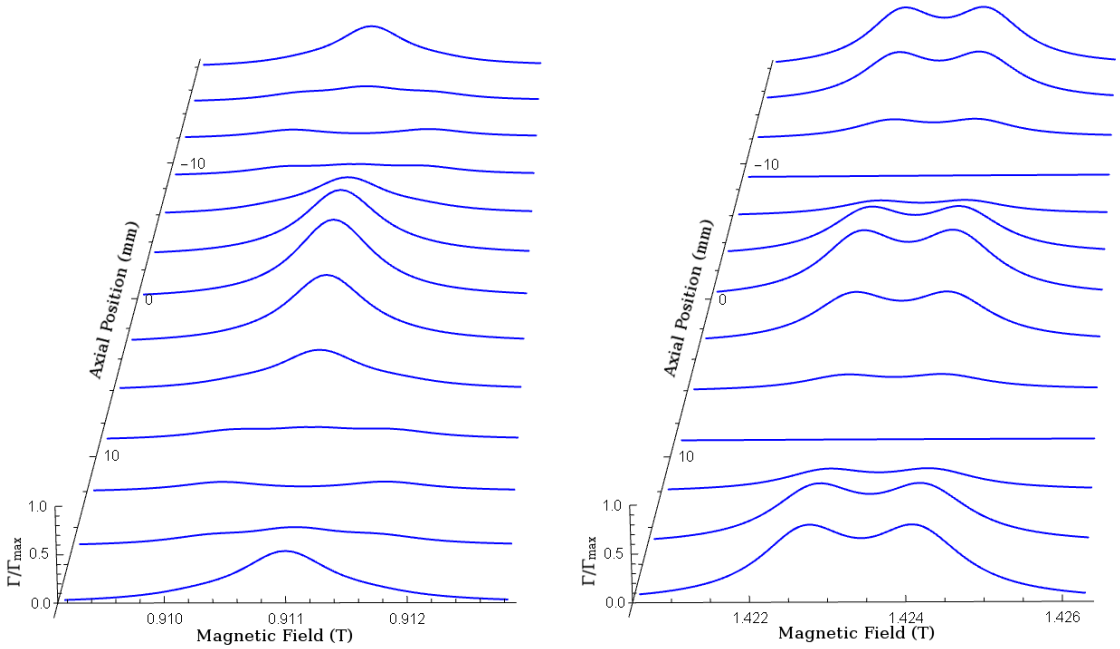

Figure 6.4: Cooling enhancements: (left) TE_{123} with 2×10^4 electrons, (right) TE_{134} with 3×10^5 electrons as calculated from Eq. 6.47. We use the cavity dimensions of the Berkeley plasma trap with the cavity mode of form Eq. 6.3 to calculate the cooling rate. We assume $\Delta\omega_c \ll \omega/2Q$, and that the plasma has negligible radius to calculate the cooling rate. These plots set $T = 10000$ K, $Q = 1000$, and $\omega_z = 40\pi \times 10^6$ s^{-1}. $\Gamma_{max} = 50$ s^{-1} for the left plot and $\Gamma_{max} = 16$ s^{-1} for the right. For the right plot, we assume that the $O(\pi^0 k^0 \zeta_{th}^0)$ term in the expansion of Eq. 6.46 does not contribute to the cooling rate. We find good qualitative agreement between the cooling rates here and the cooling rates in Fig. 6.1.

By the arguments of Sec. 6.3, when $\mathcal{N}\langle|\beta|^2\rangle$ becomes large enough such that the dominant contribution (that is $O(\pi^0 k^0 \zeta_{th}^0)$) is suppressed, the cooling by the higher order contributions will be the only mechanism for cooling. This explains why for low $\mathcal{N}$, we observe cooling without a split resonance, but for high $\mathcal{N}$, we observe we observe two resonances split by $2\omega_z$ in Fig. 6.1. For even higher $\mathcal{N}$, we would expect to observe two resonances split by $4\omega_z$, coming from the $O(\pi^4 k^4 \zeta_{th}^4)$ term in Eq. 6.46. The higher order cooling rates are proportional to $\zeta_{th}^2 \propto T$, and hence as the plasma cools, the cooling will slow. For a sufficiently cold plasma, this means there is little benefit from the cavity and the cooling will be dominated by the free space form.

We show in Fig. 6.4 the theoretical predictions. We see that the puzzling fact cooling appeared at antinodes in Fig. 6.1 is naturally a part of this theory. To obtain the high $\mathcal{N}$ limit of the expression (such that the dominant term is suppressed), we take only the $O(\pi^2 k^2 \zeta_{th}^2)$ term in the expansion of Eq. 6.46, and do not include the $O(\pi^0 k^0 \zeta_{th}^0)$ term.

6.5 Miscellaneous effects

In this section we discuss the addition of effects not accounted for in the analysis of Sec. 6.4. Many of the effects can be included exactly, but complicate the derivations of Sec. 6.3 and 6.4.

Field Gradient

Assume that there is a linear field gradient $\omega_c(\zeta) = \omega_{c,0}(1 + \kappa\zeta)$ across the plasma. We consider the effect of a linear field gradient by modifying Eq. 6.9 to

$$\frac{du^{(j)}}{dt} = -\frac{i\omega e}{\sqrt{2V\omega}m} a\psi(\vec{x}^{(j)}) - i\omega_{c,0}(1 + \kappa\zeta_0^{(j)} \sin(\omega_z t + \phi^{(j)}))u^{(j)}. \tag{6.48}$$

The solution to this is

$$\begin{aligned} u^{(j)}(t) =& e^{-i\left(\omega_{c,0}t + \frac{\zeta_0^{(j)}\kappa\omega_{c,0}}{\omega_z}\left(\cos(\phi^{(j)}) - \cos(\omega_z t + \phi^{(j)})\right)\right)} \\ &\times \left[u^{(j)}(0) - \frac{i\omega e}{m\sqrt{2V\omega}} \int_0^t dt'\, e^{i\left(\omega_{c,0}t' + \frac{\zeta_0^{(j)}\kappa\omega_{c,0}}{\omega_z}\left(\cos(\phi^{(j)}) - \cos(\omega_z t' + \phi^{(j)})\right)\right)} a(t')\psi(\vec{x}^{(j)}(t')) \right] \end{aligned} \tag{6.49}$$

We proceed as before with the calculation of the bulk decay rates. We find that, for $\kappa\omega_{c,0} \ll \omega/2Q$, there is little change in the bulk decay rates due to the field gradient. When this assumption is not true, the magnetic field gradients can lead to complicated plasma trajectories, and hence are difficult to model here [89]. From the arguments of Sec. 6.3, we expect that cooling rates become proportional to $\Delta\omega_c$, as $\mathcal{N}$ grows large, and hence by introducing a field gradient there may be enhanced cooling. This is confirmed by some of the simulation results in Sec. 6.6.

Non-cylindrical cavities

The bulge cavity used in the Berkeley plasma trap alters the shape of the cavity modes. To model this effect, we use the numerical solutions provided by Ref. [2] and fit the z dependence at $r = 0$ as a Fourier series. While this complicates the analysis, the cooling rates can be calculated in exactly the same manner as Sec. 6.4.

Anharmonic Wells

The analysis of Sec. 6.4 can be interpreted as considering the dominant Fourier component of the bounce motion of the plasma particles. In this interpretation, if we wish to consider a

triangle wave instead of harmonic motion (velocity is constant but changes direction at the walls), we simply take $\zeta_0 \to 8\zeta_0/\pi^2$ to account for the Fourier expansion of the triangle wave. In general, this is valid for any Fourier decomposition provided the Fourier component at ω_z dominates.

Finite temperature walls and heating sources

Finite temperature walls can be included by introducing a vacuum expectation value for the radiation field. In Eq. 6.14, we could take $a(t) \to a(t) - a_0$ with $\omega|a_0|^2 = T_{wall}$ to include this effect. A heating source is modeled by adding a constant term to Eq. 6.9. We are not aware of any new effects from adding these corrections, but the two will ultimately affect the final temperature.

Free space emission

We assume that the plasma does not interact with the free space radiation of plasma particles. In this case, free space emission is included by taking, in Eq. 6.9, $\omega_c \to \omega_c - i\gamma$, where γ is the free space cooling rate. This is the Larmor cooling rate given by [72]

$$\gamma = \frac{2e^2\omega_c^2}{9\pi m}. \tag{6.50}$$

In general, this leads to cooling at a rate γ when all other cooling methods are slower, which is most readily seen by making the replacement in Eq. 6.21 and 6.31.

If the plasma does interact with the free space radiation of plasma particles, the effect must be included as a series of cavity modes.

6.6 Comparison with Simulations

In Fig. 6.5 we compare our analytic model, that is the bulk cooling rates calculated from Eq. 6.47, with simulations of the $\mathcal{N}+1$ oscillator equations given in Eq. 6.9 and 6.13 [83]. The simulations allow to more easily include some of the effects of Sec. 6.5. These simulations included an anharmonicity and a gradient in the magnetic field, and also use the true cavity mode profile of the true cavity rather than the cavity mode profile of a cylindrical cavity. As expected, the analytic model describes the early time behavior of the cavity mode well, but disagrees from the simulation as the magnetic gradient effects become important. As the magnetic field gradient is increased, the agreement becomes poorer, as expected from the analysis in Sec. 6.3.

In Fig. 6.6, we compare the simulation for different $\mathcal{N}$. We see that as $\mathcal{N}$ is increased, the effect of $\Delta\omega_c$ becomes more prominent in the DC value of the cavity mode energy. This is in agreement with the predictions from Sec. 6.3.

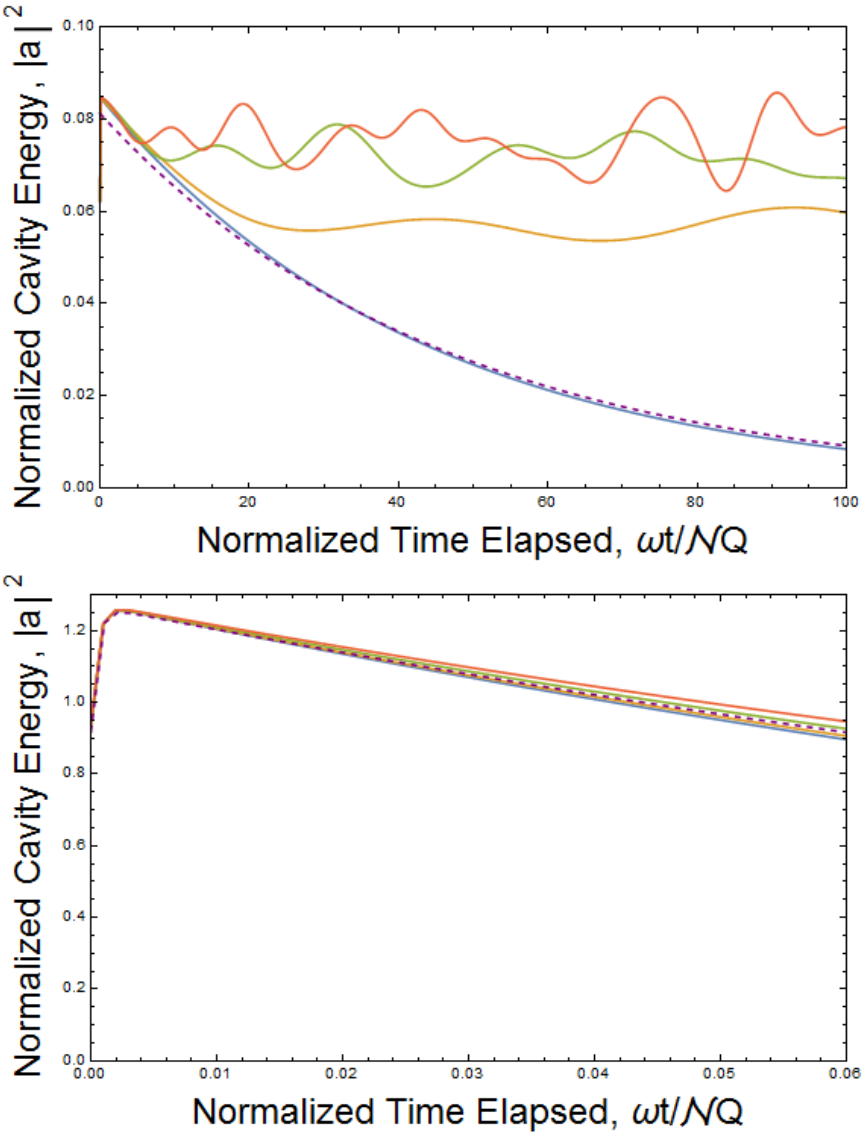

Figure 6.5: Energy in the TE_{121} cavity mode as a function of time for a cavity with $\mathcal{N} = 4096$ particles with initial temperature $T = 10000$ K from our analytic model (dashed, purple) and from simulation of Eq. 6.9 and 6.13 (solid lines). Both plots show the same data, with different timescales. We assumed that $\langle\omega_c\rangle = \omega$, and that the plasma was at the center of the cavity. For the simulation, the blue curve sets $\Delta\omega_c/\omega_c = 0$, the light orange curve sets $\Delta\omega_c/\omega_c = 4 \times 10^{-7}$, the green curve sets $\Delta\omega_c/\omega_c = 1.2 \times 10^{-6}$, and the dark orange curve sets $\Delta\omega_c/\omega_c = 2 \times 10^{-6}$. The gradient is assumed to vary linearly across the cavity.

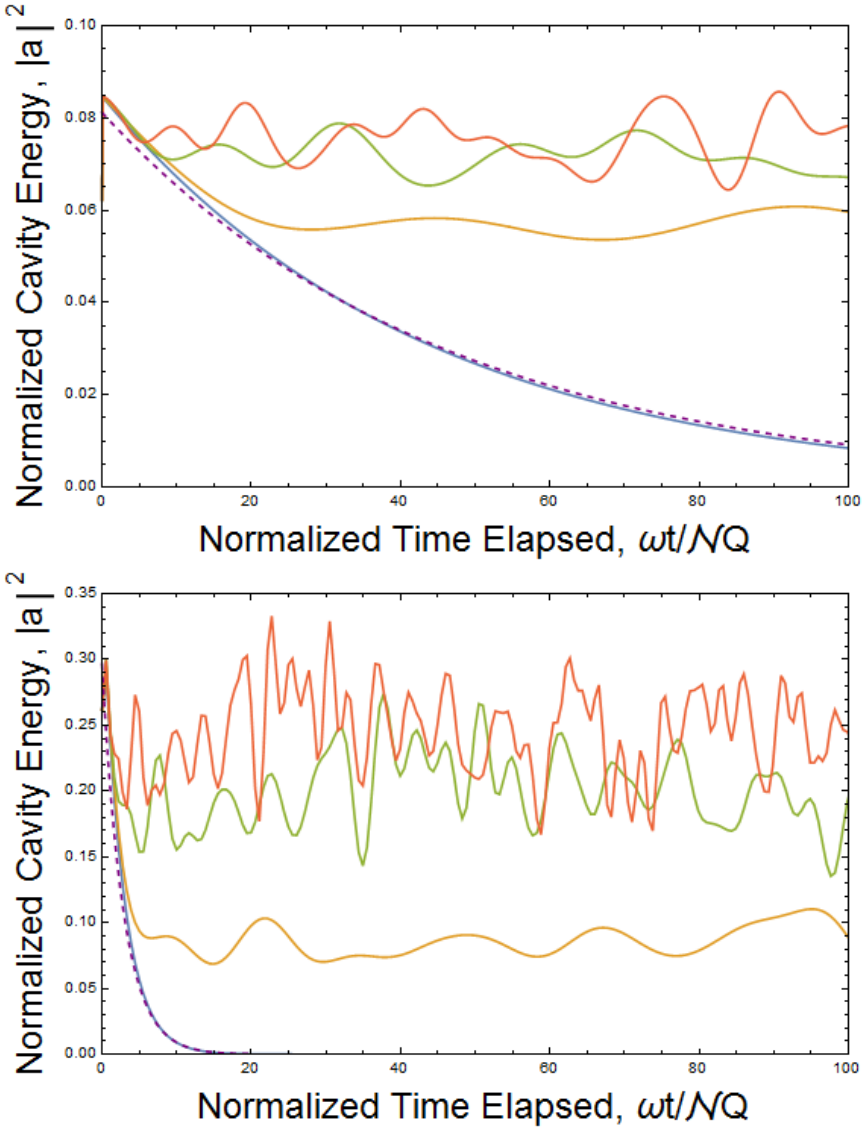

Figure 6.6: Energy in the TE_{121} cavity mode as a function of time for a cavity with particles with initial temperature $T = 10000$ K from our analytic model (dashed, purple) and from simulation of Eq. 6.9 and 6.13 (solid lines). The top plot sets $\mathcal{N} = 256$, while the bottom plot sets $\mathcal{N} = 1024$. We assumed that $\langle \omega_c \rangle = \omega$, and that the plasma was at the center of the cavity. For the simulation, the blue curve sets $\Delta\omega_c/\omega_c = 0$, the light orange curve sets $\Delta\omega_c/\omega_c = 4 \times 10^{-7}$, the green curve sets $\Delta\omega_c/\omega_c = 1.2 \times 10^{-6}$, and the dark orange curve sets $\Delta\omega_c/\omega_c = 2 \times 10^{-6}$. The gradient is assumed to vary linearly across the cavity.

6.7 Future Work

While we have shown good agreement between our theoretical results and the experiment, an open question is the effect of field gradients. A future upgrade to the Berkeley plasma trap will include a small solenoid for high control of magnetic gradients. We expect that the cooling $\propto \Delta\omega_c$ of Eq. 6.31 should hold generally. However, the field gradients could complicate the trajectory of plasma particles and lead to particle losses, which we have not considered here [90]. When the magnetic gradient becomes large, the plasma may be able to couple to multiple cavity modes, which has not been considered here.

Our analysis shows that a particular linear combination of electron oscillators (here called the superradiant) is the relevant mode through which the plasma ultimately cools. Formalizing and understanding the energy transfer between the electrons and the superradiant mode will be crucial to fully understand the cooling problem.

We have assumed throughout that collisions will thermalize the parallel and perpendicular temperatures of the plasmas (note that the only quantity that is measurable is the parallel temperature). We assume that there are no observable effects from this thermalization, but we have not proved this fact, and a careful analysis should be conducted. In particular, as the plasma cools, the collision rate becomes increasingly relevant. Collisions may introduce correlations between plasma particles as well as another mixing mechanism into the superradiant mode, and these effects may be relevant to the cooling of plasmas.

We have also not studied the effect of noise in the system. Because of the fluctuation dissipation theorem, if there is dissipation through the walls, we expect there to be fluctuations in the radiation field [91]. While we expect to be in the *hemi-classical* limit, there may be observable effects from the discrete nature of photons [92].

At colder temperatures, other effects become important. As $\zeta_{th}^2 \propto T$ becomes small, there is no efficient mechanism for a plasma where $\mathcal{N}\langle|\beta|^2\rangle \gg \omega^2/4Q^2$ to cool. In this limit, there is little benefit from the cavity and we expect most of the plasma cooling to happen from free space cooling. Further, at these temperatures, our harmonicity assumption, Eq. 6.40, becomes poor, and thus the results may not be generalizable to this regime. Careful experimental measurements can show that this is indeed the case and explore any additional effects. At extremely cold temperatures ($\sim$ 1 K) quantum effects become relevant and a fully quantum mechanical treatment of the problem would be necessary.

Acknowledgements

I thank many useful discussions with Jonathan Wurtele and Eric Hunter on this topic. Andrew E. Charman provided guidance with the initial formulation of the problem. Francis Robicheaux shared his code for the solving the oscillator equations and offered many insights into the problem.

Chapter 7

Bounding the antihydrogen charge

In this chapter we describe a stochastic acceleration measurement made by ALPHA to bound the charge, Qe, of antihydrogen, where e is the charge of the proton [28]. I present details of my contribution to this work, primarily performing simulations and the Bayesian analysis conducted to place the bound.

ALPHA placed, at 68.3% confidence, a bound of $|Q| < 7.1 \times 10^{-10}$ [28], which is over 20 times better than the previous bound obtained by the ALPHA experiment [27]. As the antiproton charge is within 7×10^{-10} (at 90% confidence) of $-e$, if we assume that superposition holds, this is also the best bound on the charge anomaly of the positron by a factor of 25 [7, 93, 94].

7.1 Measurement

Recall from Sec. 2.2 that in ALPHA, antihydrogen atoms are trapped in an Ioffe trap occupying the same physical space as a Penning-Malmberg trap. The electrodes of the Penning-Malmberg trap can be biased to create electric fields in the region where antihydrogen is trapped. If antihydrogen is charged, it would accelerate due to these fields and, if given enough energy, would no longer be trapped by the mirror and octupole fields of the Ioffe trap.

The charge measurement procedure is the following. Antihydrogen atoms are created in the trap and are allowed to relax with no controlled electromagnetic variations for 0.4 s. After this, we apply potentials on the Penning-Malmberg trap electrodes that vary between $\sim \pm 100$ V. The potentials are applied in repetitions consisting of 1698 transitions over 1.698 s. The transitions are spaced at pseudo-random intervals drawn from a a uniform distribution with average 1 ms and standard deviation of 0.2 ms such that the time of kicks are approximately independent [95, 96]. 50 identical repetitions are applied to the anti-atoms, with each repetition spaced by 2.298 s, resulting in $N = 84,900$ total potential transitions. Due to differences in amplifiers and filters driving the electrodes, the potential on the electrode as a function of time is not identical between electrodes. These differences

have been characterized and are taken into account exactly in the simulations.

If antihydrogen is charged, each of the potential transitions non-adiabatically perturbs the kinetic energy of the anti-atom. Let $\Delta\Phi$ be the potential change at each transition, ≈ 100 V. Note that $\Delta\Phi$ varies over the volume of the trap and hence has significant uncertainties. The potential transitions can either increase or decrease the energy of the anti-atom by $Qe|\Delta\Phi|$, and thus the process corresponds, approximately, to a random walk in energy if the times of kicks are independent. A simple random walk model predicts that an anti-atom with putative charge Qe would gain energy $\sim |Q|e\Delta\Phi\sqrt{N}$. Note that simulations suggest that the energy gain should scale as a slightly higher power of N [96]. Under these assumptions, an anti-atom would gain enough energy to escape the trap of depth E_{Well} if

$$|Q| \gtrsim \frac{E_{Well}}{e\Delta\Phi\sqrt{N}} \approx 1.6 \times 10^{-9}. \tag{7.1}$$

Note that such a relation relies on the assumption that the potential transitions applied to the antihydrogen atoms are independent. In particular, if the kicks become phase-locked with the antihydrogen bounce motion, it can damp the motion of the anti-atom making it harder for the anti-atom to escape. This is clearly not ideal given the low trapping rates at ALPHA. Independence of the potential transitions is a temporally local requirement, and hence there is no advantage to using a random set of repetitions over a pseudo-random one [96].

We define stochastic trials as those described above in which varying potentials are applied to electrodes. We define null trials as those in which the anti-atoms are held for the same amount of time, but no potentials are applied to the electrodes. We compare the number of anti-atoms that remain after the null and stochastic treatments. If a smaller number remain after the stochastic treatment, this is strong evidence that the anti-atoms are responding to the electric fields produced by the electrodes. In both stochastic and null trials, we observed 12 total events over 10 trials of each type. Since we observed no statistically obvious difference between the two protocols, this suggests that the charge, if any, of antihydrogen is below that required to escape the trap, namely $|Q| \lesssim 1.6 \times 10^{-9}$ from our analytic estimate, Eq. 7.1. The measurement relies only on whether anti-atoms survive or not, which is minimally affected by systematic issues such as the robustness of the annihilation detectors. Sufficient samples to reveal an assymetry in deflection are not necessary as in Ref. [27]. This is advantageous as only 0-3 anti-atoms are trapped per trial, and thus high statistics would require a large number of trials and time.

7.2 Simulations of Antihydrogen Orbits

The anti-atom equation of motion in the trap is

$$M\ddot{\vec{x}}(t) = \vec{\nabla}[\vec{\mu} \cdot \vec{B}(\vec{x})] + Qe[\vec{E}(\vec{x}, t) + \dot{\vec{x}}(t) \times \vec{B}(\vec{x})]. \tag{7.2}$$

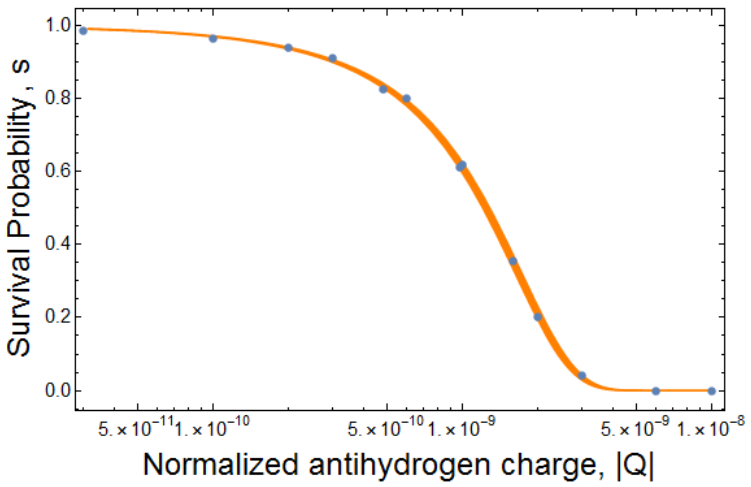

Figure 7.1: Simulated maximum likelihood estimate of the survival probability s as a function of the normalized antihydrogen charge $|Q|$ for the stochastic trials. The points are the number of anti-atoms surviving at the given $|Q|$ value divided by the number of anti-atoms surviving at $Q = 0$ (the null simulation). The orange band of varying thickness is the 68.3% confidence region from a Bayesian fit to this data.

Here, $\vec{x}(t)$ is the center-of-mass position of an anti-atom at time t, $\vec{E}(\vec{x}, t)$ and $\vec{B}(\vec{x})$ are the applied electric and magnetic fields, M is the antihydrogen mass, and $\vec{\mu}$ is the experimentally verified magnetic moment of antihydrogen [26]. A trapped anti-atom must be low-magnetic-field-seeking, so the direction of $\vec{\mu}$ is assumed to adiabatically track the direction anti-parallel to $\vec{\nabla}|\vec{B}|$.

We wish to estimate $s(|Q|)$, the survival probability of anti-atoms with a charge $\pm Q$ when subjected to the fields described in the previous section. Let τ be the time corresponding to the end of a full experimental cycle, and let $I(\vec{x})$ be an function that is 1 if $\vec{x}$ is in the trap and 0 if $\vec{x}$ is outside the trap. Formally, the survival probability can be written

$$s(|Q|) = \mathbb{E}(I(\vec{x})), \tag{7.3}$$

where the expectation is taken over $p(\vec{x}, \vec{v}, \tau)$, the probability of finding an anti-atom with charge Qe at position $\vec{x}$ and velocity $\vec{v}$ at time τ. The difficulty is that the distribution $p(\vec{x}, \vec{v}, \tau)$ is unknown. As shown in Sec. B.2, the time propagation of $p(\vec{x}, \vec{v}, t)$ is equivalent to the ensemble average of the time propagation of test particles. Thus, we can estimate the integral above by simulating the trajectories of particles and determining whether the particles escape the trap. If the initial distribution of simulated test particles matches the expected initial distribution, by the law of large numbers, the survival fraction will converge to the expectation as more test particles are simulated.

The simulations are similar to those in [18, 27, 96–99], in which numerous tests of robustness of the dynamical model and numerical approximations have been performed. A symplectic leap-frog propagator is used for numerical integration of Eq. 7.2. The electric potentials in the simulation are determined from a precise model of the trap geometry, using the COMSOL finite element code to solve the Laplace equation, and the measured time-history of the potentials on each electrode [100]. The magnetic field is calculated from high-accuracy analytic expansions derived from numerical Biot-Savart modeling [99].

The simulations follow the experimental procedures closely. Simulated anti-atoms are launched in a volume corresponding to the volume of the positrons in the experiment with an energy distribution obeying a truncated 3D Maxwellian distribution. The truncation energy, 0.75 K, is well above the trapping depth $E_{Well} = 0.54$ K. The anti-atoms are then allowed to randomize in the electric-field-free trap for 1.25 s. Almost all the anti-atoms with energy above E_{Well} will be lost during this time, however, a few will remain on quasi-bound orbits [98, 101]. We model the null trials by propagating the anti-atoms in the electric-field-free trap for a further 114.9 s. We model the stochastic trials by assigning the anti-atoms a putative Q, and propagating the anti-atoms with the stochastic electric fields present for this same 114.9 s. For each case studied, we propagate at least 100 anti-atoms.

For each $|Q|$, we determine the maximum likelihood estimate of the survival probability s by calculating the ratio of the number of anti-atoms that survive the stochastic simulations to the number that survive the null simulations; the results are shown in Fig. 7.1. Because the stochastic drive is unbiased on average, reversing direction every half-cycle, s depends only on the magnitude $|Q|$, not the sign of Q.

7.3 A statistical bound on the antihydrogen charge

The observations showed no difference between null and stochastic trials, and thus the simulation results in Fig. 7.1 would suggest that the bound on the antihydrogen charge is $|Q| \lesssim 10^{-9}$. We describe here the construction of a statistically rigorous limit, with a confidence level, that was used by ALPHA to find a bound.

We use an objective Bayesian approach to place a probabilistic limit on $|Q|$ from the counts of trapped antihydrogen atoms with and without stochastic driving [102–107]. A Bayesian method provides various advantages over a frequentist approach. It allows us to easily integrate out r, the capture-and-detection rate, as a nuisance parameter, without having to determine a value for it. It allows us to easily accommodate for systematic effects as well as random sampling errors from both experimental counts and Monte Carlo simulations. Furthermore, frequentist methods can raise questions about coverage probabilities when the method of analysis is not adopted prior to making measurements, as was the case with this measurement [6, 7].

To be fully consistent with the Bayesian approach, we emphasize that all probabilities are conditioned, if only implicitly, on background information $\mathcal{I}$. This includes knowledge of the trap geometry and operation, experimental protocols, previous simulation studies, basic facts

of classical and quantum physics, etc. To find a $(1-\alpha)$ posterior credible interval for $|Q|$, given our observed data $\mathcal{D}$, and Monte Carlo samples $\mathcal{M}$, ultimately we solve (numerically) for an upper limit Q^* such that

$$\int_0^{Q^*} d|Q| \;\; p(|Q|\,|\mathcal{M},\mathcal{D},\mathcal{I}) = 1-\alpha. \tag{7.4}$$

Our measurement described in Sec. 7.1 does not measure $|Q|$ directly. The measurement can be directly used to determine the posterior distribution for the probability $(1-s)$ of additional loss of initially-trapped anti-atoms, due to heating from the stochastic drive fields during the hold period. s is then related to $|Q|$ via the Monte Carlo simulations that relates the survival rate to $|Q|$. As the empirical measurements are independent of the Monte Carlo simulations and vice versa, the posterior probability density $p(|Q|\,|\mathcal{M},\mathcal{D},\mathcal{I})$ can be decomposed into two terms

$$p(|Q|\,|\mathcal{M},\mathcal{D},\mathcal{I}) = \int_0^1 ds \;\; p(s|\mathcal{D},\mathcal{I}) p(|Q|\,|s,\mathcal{M},\mathcal{I}). \tag{7.5}$$

The next two subsections will provide details on how to compute each of the probability densities in Eq. 7.5. Ultimately, uncertainties at both stages are convolved together and contribute to the bound on $|Q|$.

Estimating s from the data $(p(s|\mathcal{D},\mathcal{I}))$

The capture-and-detection rate, r, is the average number of anti-atoms surviving until, and detected at, the end of a single null trial. The (unknown) value of r is independent of Q, but is assumed fixed across all trials, and combines into one overall mean the initial average rate of anti-atom production as well as trapping and retention probabilities and overall detection efficiencies.

As r has, by assumption, no information of Q, r is a nuisance parameter. Thus, to obtain the posterior of s given the posterior from s and r, we simply integrate over all possible values of r [7]. The first term in Eq. 7.5 can be expressed as

$$p(s|\mathcal{D},\mathcal{I}) = \int_0^\infty dr \;\; p(s,r|\mathcal{D},\mathcal{I}). \tag{7.6}$$

Because we integrate over r, the actual value of r need never be explicitly estimated from the data, nor do we ever need to worry about factorizing r into its separate contributions from synthesis, trapping, or detection of anti-atoms.

From Bayes' theorem we infer that the joint posterior density of s and r can in turn be written as

$$p(s,r|\mathcal{D},\mathcal{I}) = \pi(s,r|\mathcal{I})\,\frac{P(\mathcal{D}|s,r,\mathcal{I})}{P(\mathcal{D}|\mathcal{I})}, \tag{7.7}$$

in terms of the prior distribution $\pi(s, r|\mathcal{I})$ for the parameters s and r, the likelihood $P(\mathcal{D}|s, r, \mathcal{I})$ of the data given s and r, and,

$$P(\mathcal{D}|\mathcal{I}) = \int_0^1 ds' \int_0^\infty dr' \;\; \pi(s', r'|\mathcal{I})\, P(\mathcal{D}|s', r', \mathcal{I}), \tag{7.8}$$

which ensures the normalization of the posterior probability density.

The data $\mathcal{D}$ consist of the observed number of antihydrogen detections events $n_1, \ldots, n_K$ in each of $K = 10$ null trials, together with the observed number of detection events $n'_1, \ldots, n'_K$ in each of the $K' = K = 10$ stochastic trials. It will also be useful to define the total number of observed anti-atoms in null and stochastic trials, $N_n = \sum_{i=1}^{K} n_k$ and $N_s = \sum_{i=1}^{K'} n'_k$.

Given this, the likelihood $P(\mathcal{D}|s, r, \mathcal{I})$ is straightforward to evaluate. All trials are conditionally independent given s and r, so the overall likelihood will just be the product of the likelihoods of all null and all stochastic trials. Probabilities for the possible observed number of anti-atoms in each null trial are all Poissonian with the same mean r, while in the stochastically driven trials, the probabilities involve an additional binomial sub-sampling (with "success" probability s) of the same underlying Poissonian process. As the Poisson distribution is the maximum entropy distribution with a specified rate parameter, there is no reason to expect that the likelihoods should not be Poissonian (see Sec. C.1) [108]. The aggregate data from the 2014 run do confirm that the distribution is Poissonian. Since

$$\sum_{z>x}^{\infty} s^x (1-s)^{z-x} \frac{r^z}{z!} e^{-r} = \frac{(rs)^x}{x!} e^{-rs}, \tag{7.9}$$

the subsampling is equivalent to a Poisson process with rate sr. Thus, the overall likelihood for the $K = 10$ null trials and the $K' = K = 10$ interleaved stochastic trials can be written as

$$P(\mathcal{D}|s, r, \mathcal{I}) = \prod_{k=1}^{K} \frac{(r)^{n_k}}{n_k!} e^{-r} \frac{(sr)^{n'_k}}{n'_k!} e^{-sr} \propto r^{N_n + N_s}\, s^{N_s}\, e^{-Kr(1+s)}, \tag{7.10}$$

where we have dropped overall constant factors which will cancel out upon normalization of the posterior distribution, see Eq. 7.8. In particular, only the total counts over the null and over the stochastic trials, N_n and N_s, are relevant, justifying the manner in which we reported results in Sec. 7.1.

To determine the posterior density $p(s, r|\mathcal{D}, \mathcal{I})$, we do require a definite choice for the joint prior density $\pi(s, r|\mathcal{I})$. We have no information that would lead to any logical or physical dependence between these parameters *a priori*, so we take the priors for the parameters to be independent: $\pi(s, r|\mathcal{I}) = \pi(s|\mathcal{I})\, \pi(r|\mathcal{I})$. To try to maximize how much we learn from the data themselves, we follow conventional Bayesian practice in choosing the priors to be

uninformative. As discussed in Appendix C, various distinct notions of what it means to be uninformative have been introduced with various rationales, but fortunately in our case, most of the widely-adopted definitions all lead to the same functional form for the priors for both the baseline rate r and the survival probability s.

The uninformative prior for a Poisson rate parameter such as r is

$$\pi(r|\mathcal{I}) \propto \frac{1}{\sqrt{r}}. \tag{7.11}$$

This prior is improper (non-normalizable) over the domain $0 \leq r < \infty$, due to the behavior of its integral for $r \to \infty$, but the posterior will be perfectly normalizable. The uninformative prior for a binomial probability, such as s is

$$\pi(s|\mathcal{I}) = \frac{1}{\pi\sqrt{s(1-s)}}. \tag{7.12}$$

This expression is normalized over the domain $0 \leq s \leq 1$.

We consider the effect of the choice of prior by replacing the uninformative priors of Eq. 7.11 and 7.12 with their conjugate priors. These are priors with the same functional form, but with adjustable hyperparameters. For the Poisson rate we consider

$$\pi(r|\mathcal{I}) \propto r^{\alpha}, \tag{7.13}$$

and for the binomial probability we consider the form

$$\pi(s|\mathcal{I}) \propto s^{a}(1-s)^{b}. \tag{7.14}$$

The uninformative choice is $\alpha = a = b = -1/2$. For $1 \geq a, b \geq -\frac{1}{2}$ and $0 \geq \alpha \geq -\frac{1}{2}$, we find our final bounds on $|Q|$ do not change up to the first significant digit. For $-\frac{1}{2} \geq a, b > -1$, the final bounds can change significantly, but this is a perverse choice of hyperparameters, as the posterior on s becomes non-normalizable when $a, b = -1$.

Using the uninformative priors and completing the integrals, which can be done analytically, we find

$$p(s|\mathcal{D},\mathcal{I}) \propto \frac{s^{N_s-\frac{1}{2}}\Gamma\left(N_n+N_s+\frac{1}{2}\right)}{(K(1+s))^{N_n+N_s+\frac{1}{2}}\sqrt{1-s}}. \tag{7.15}$$

The proportionality constant is determined by ensuring $\int_0^1 ds\, p(s|\mathcal{D},\mathcal{I}) = 1$.

Relating s to Q ($p(|Q||s,\mathcal{M},\mathcal{I})$)

Now we turn to the other contribution to Eq. (7.5), namely the probability density function $p(|Q|\,|s,\mathcal{M},\mathcal{I})$ that encodes our knowledge about the correlation between the charge magnitude $|Q|$ and the survival probability s, determined from the Monte Carlo simulations, $\mathcal{M}$.

Function	Form	Free parameters	ΔBIC	ΔF
error function	$1+\mathrm{erf}(\frac{x-x_0}{\sigma})$	2	0	0
logistic	$1+\tanh(\frac{x-x_0}{\sigma})$	2	0.8	2
Weibull	$e^{-(x/\sigma)^k}$	2	2	-
Gudermannian	$\pi-2\int_0^{\frac{x-x_0}{\sigma_0}}\frac{dt}{\cosh(t)}$	2	4	5
Gumbel	$1-e^{-e^{-\frac{x-x_0}{\sigma}}}$	2	7	8
Fisk	$\frac{1}{(x/\sigma)^\beta+1}$	2	47	49
Fréchet	$1-e^{-\left(\frac{\lvert x-x_0\rvert}{\sigma}\right)^{-\alpha}}$	3	104	-
arctan	$\pi-2\arctan(\frac{x-x_0}{\sigma})$	2	120	123

Table 7.1: Comparison of model selection criteria BIC and estimated logarithm of the Bayes factor, F, for fitting sigmoid functions to the Monte Carlo results. All criteria are compared relative to the error function, which was found to be the best. The "Form" column gives the form of the distribution as a function of x up to a normalization constant. Any other variable in the expression is a free parameter. Uninformative priors were used to compute F. Fréchet and Gumbel do not have entries in the third column as the cutoffs necessary for the priors for the parameters of those distribution are not the same as the cutoffs on the priors for the error function. Thus, ΔF in these two cases is dependent on the choice of cutoff of prior.

We performed simulations at each of a sequence of different charge values Q_i, $i = 1 \ldots, \mathcal{C}$, distributed approximately uniformly in logarithm over the relevant range where the survival probability s shows appreciable variation. For each Q_i, we numerically propagated $\mathcal{N}_i \sim 10^3$ anti-atoms to determine the number $\mathcal{X}_i$ that escaped the trap during the hold interval when subjected to the stochastic forcing.

We determine $p(|Q|\,|s, \mathcal{M}, \mathcal{I})$ by curve-fitting a parameterized fit function $s = \mathcal{S}(|Q|\,|\boldsymbol{a}, \mathcal{I})$ to the survival results from the Monte Carlo simulations, where the $\boldsymbol{a}$ are some set of adjustable parameters.

Table 7.3 compares a number of different functional forms, including those that commonly arise in survival analysis and the theory of extreme values, such as the logistic, Weibull, Gumbel, as well as other cumulative distributions with a "sigmoidal" shape. Based on comparisons of the Bayesian Information Criterion (BIC), and estimated Bayes' factors [109] between best fits within each functional family, a clear winner emerged, so we settled on a survival function of the following simple form:

$$\mathcal{S}(|Q|\,|Q_0, \sigma_0, \mathcal{I}) = \frac{1 + \mathrm{erf}\left(\frac{Q_0 - |Q|}{\sigma_0}\right)}{1 + \mathrm{erf}\left(\frac{Q_0}{\sigma_0}\right)}, \tag{7.16}$$

involving an adjustable "location" hyperparameter Q_0 and "scale" hyperparameter $\sigma_0 > 0$.

See Appendix C for details on model selection. As we are considering a small number of fit points (simulations were run for 14 different values of Q) and considering roughly the same number of fit parameters for all fits, the Akaike information criterion (AIC) [110] performs nearly identically to the BIC. Although the error function relation cannot be derived theoretically from first principles, such a functional form involving the error function is not implausible as an approximation, given that the stochastic drive essentially induces a diffusion in energy space, and this choice evidently provides a very close fit to the Monte Carlo simulations, as seen in Fig. 7.1. Note that the survival probability s is a monotonically decreasing function of the charge magnitude $|Q|$, as would be anticipated, so the function can in principle always be inverted, although it does flatten out considerably for sufficiently large or small $|Q|$.

With this functional shape, the survival curve fitting is similar to a Bayesian "probit" regression [111]. Note that we must also introduce one additional nuisance parameter, ϕ_0, which measures the survival probability for simulated trajectories in the absence of stochastic forcing. This is necessary for correct inference of the survival probability conditioned on $|Q|$, because many of the sampled initial conditions do not in fact lead to trapped anti-atom trajectories, even in the absence of stochastic forcing. But after integrating over ϕ_0, the precise value of ϕ_0 does not directly contribute to our posterior estimate for $|Q|$ as a function of s, and the overall effect is just to slightly broaden the distribution for the hyperparameters which do determine the shape of the survival or calibration curve connecting $|Q|$ and s. Taking this into account gives

$$p(Q_0, \sigma_0|\mathcal{M}, \mathcal{I}) = \int_0^1 d\phi_0 \;\; p(Q_0, \sigma_0, \phi_0|\mathcal{M}, \mathcal{I}). \tag{7.17}$$

Bayes' theorem can then be used to find the posterior for the fit parameters Q_0, σ_0, and ϕ_0, based on the priors and the likelihood function for the Monte Carlo output

$$p(Q_0, \sigma_0, \phi_0|\mathcal{M}, \mathcal{I}) = \pi(Q_0, \sigma_0, \phi_0|\mathcal{I}) \frac{p(\mathcal{M}|Q_0, \sigma_0, \phi_0, \mathcal{I})}{p(\mathcal{M}|\mathcal{I})}. \tag{7.18}$$

Recall that our simulations $\mathcal{M}$ are $\mathcal{C}$ simulations, with differing values of the antihydrogen charge, of $\mathcal{N}_i$ anti-atom trajectories to determine the number $\mathcal{X}_i$ of the trajectories that escape the trap. As the survival is a binomial process, the likelihood for the Monte Carlo trajectory samples is

$$p(\mathcal{M}|Q_0, \sigma_0, \phi_0, \mathcal{I}) = \prod_{i=1}^{\mathcal{C}} \binom{\mathcal{N}_i}{\mathcal{X}_i} (\phi_0 \mathcal{S}(Q_i|Q_0, \sigma_0, \mathcal{I}))^{x_i} (1 - \phi_0 \mathcal{S}(Q_i|Q_0, \sigma_0, \mathcal{I}))^{\mathcal{N}_i - \mathcal{X}_i}. \tag{7.19}$$

We do, in principle, have to choose a prior density $\pi(Q_0, \sigma_0, \phi_0|\mathcal{I})$, but given the large number of Monte Carlo samples, essentially any diffuse or uninformative prior will lead to nearly the same posterior density. Because of the large sample size we can approximate the integrals via a Laplace expansion, which treats the likelihood as a Gaussian function [109].

More details on this approximation are presented in Appendix C. Define a_ν as a vector of the fit parameters, Q_0, σ_0, and ϕ_0. We find the maximum likelihood estimate of these fit parameters by optimizing Eq. 7.19 and call the result a^*_ν. Under this approximation, we can write

$$p(\mathcal{M}|Q_0,\sigma_0,\phi_0,\mathcal{I}) \propto e^{\frac{1}{2}(a_\mu - a^*_\mu)(a_\nu - a^*_\nu)\partial_\mu\partial_\nu \ln(p(\mathcal{M}|Q^*_0,\sigma^*_0,\phi^*_0,\mathcal{I}))}, \tag{7.20}$$

The proportionality constant will drop out upon the normalization of the posterior in Eq. 7.18.

We can then decompose the probability density $p(|Q|\,|s,\mathcal{M},\mathcal{I})$ in terms of a convolution over the distribution representing our uncertainty as to the correct values of the fit hyperparameters

$$p(|Q|\,|s,\mathcal{M},\mathcal{I}) = \int_{-\infty}^{\infty} dQ_0 \int_0^{\infty} d\sigma_0 \;\; p(Q_0,\sigma_0|\mathcal{M},\mathcal{I})\, p(|Q|\,|s,Q_0,\sigma_0,\mathcal{I}). \tag{7.21}$$

In decomposing the probability density in this way, we have used the fact that the fit parameters are not logically dependent on the value of s, while Q_0 and σ_0 are assumed to exhaust all of the information contained in $\mathcal{M}$ relevant to predicting Q from the survival probability s. The first probability density $p(Q_0,\sigma_0,\phi_0|\mathcal{M},\mathcal{I})$ in the integrand of (7.21) encodes all of our uncertainty due to sampling error, given a finite number of Monte Carlo simulations. If we truly believed that the functional family used in our curve-fitting reflected the actual underlying functional relationship between s and Q for an appropriate (if imperfectly known) choice of fit parameters, then we would take the second density in the integrand to be a Dirac delta function

$$p(|Q|\,|s,Q_0,\sigma_0,\mathcal{I}) = \delta\big(|Q| - \mathcal{Q}(s|Q_0,\sigma_0,\mathcal{I})\big), \tag{7.22}$$

where $\mathcal{Q}(s|Q_0,\sigma_0,\mathcal{I}) = \mathcal{S}^{-1}(s|Q_0,\sigma_0,\mathcal{I})$ is the functional inverse of the chosen fit function. But we can also incorporate systematic effects by broadening this delta function into, say, a truncated Gaussian with a width $\sigma(s)$ corresponding to the expected size of all the combined systematic uncertainties

$$p(|Q|\,|s,Q_0,\sigma_0,\mathcal{I}) \propto \frac{1}{\sigma(s)} e^{-\frac{(|Q|-\mathcal{Q}(s|Q_0,\sigma_0,\mathcal{I}))^2}{2\,\sigma(s)^2}}, \tag{7.23}$$

where the distribution is truncated and renormalized as necessary to avoid negative values of $|Q|$.

As is usually the case, these sorts of systematic uncertainties can often only be quantified approximately. Our analysis is relatively insensitive to systematic errors. The analytic expression Eq. 7.1 show that field errors enter the bound on $|Q|$ linearly. Details of the magnetic field structure beyond the magnetic field magnitude maxima and minima are likewise unimportant as they do not effect the trap depth. Since these fields are known to $\sim 1\%$, they have a negligible effect to the level we report our limits. Simulation convergence studies

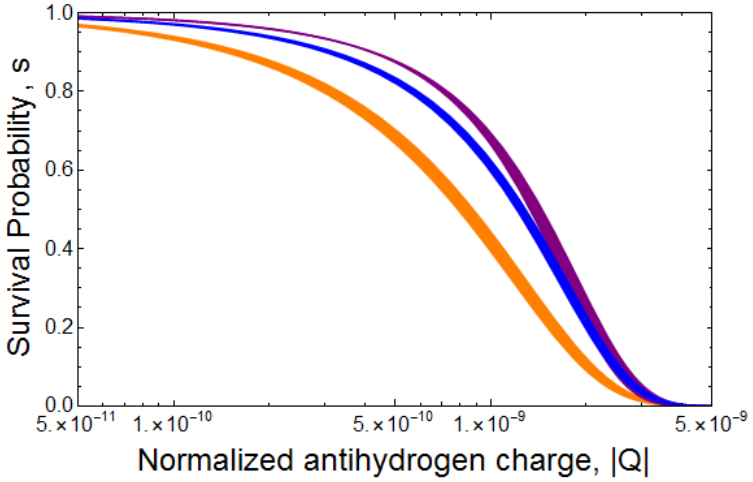

Figure 7.2: Simulated survival probability s as a function of $|Q|$ for the stochastic trials for three different distributions of initial energy. The bands of varying thicknesses are the 68.3% confidence region from a Bayesian fit to the data. The orange band assumes a linear initial distribution, the blue band assumes a Maxwellian, and the purple assumes a uniform initial energy distribution. The width of each band represents Monte Carlo sampling error, given a well-defined distribution of initial conditions, while the separation between the bands reflects systematic uncertainty regarding which distribution of initial conditions is adopted.

show that the step size need not be shorter than 4 μs (we used 0.8 μs for the simulations reported here). They also indicate that the electrode drive filter time response need only be known on the order of 4 μs; this is well within our system accuracy. Simulations show that s is not significantly influenced by the spatial region from which the anti-atoms are initially launched. All the data was taken over adjacent half shifts. We know of no mechanism that could have significantly altered the trapping rate r in sync with the strict alternation of the Stochastic and Null trials. Likewise studies of independent detector properties and detected untrapped anti-atoms find no in-sync (or otherwise) detector variations. The expected number of cosmics incorrectly identified as anti-atoms is $b = 0.15$ over all twenty trials. This can be included into the analysis by replacing the Poisson rates, r and rs in Eqs. 7.10,7.11 with $b+r$ and $b+rs$. Including this has no effect on the result to the two significant figures reported here.

We find that the dominant source of systematic error lies in the uncertainty as to the appropriate distribution of initial conditions for the energy of trappable antihydrogen atoms. Fig. 7.2 shows the difference in the fitted survival curve $s(|Q|)$ based on initial conditions described by a Maxwellian, uniform, or linear energy distributions, which span the range of plausible variation in the choice of an initial energy distribution. By setting $\sigma(s) \approx 10^{-9}\sqrt{s(1-s)}$, we find that we can account for the discrepancies. This will also account for

the other systematic effects, such as electromagnetic field errors, and ambiguities in choice of the functional form of the survival function, as we find those to be much smaller.

Bounding Q

Completing all of the integrals (several of which can be performed analytically, which cuts down on the numerical complexity of the calculation), we arrive at the posterior density for $|Q|$, from which we can find credible bounds on the magnitude $|Q|$ at any chosen level of credence. The cumulative distribution is plotted in Fig. 7.3, while the credible limits for various common choices of confidence levels are summarized in Table 7.2. In particular, we find no appreciable difference when Monte Carlo sampling errors are included. This is not surprising given the large number of Monte Carlo samples, and apparent goodness of fit of the error function. Accounting for the systematic effects degrades the limit slightly. With all effects taken into account, we find $|Q| < 7.1 \times 10^{-10}$ at 68.3% confidence. The result, close to the analytic estimate of Eq. 7.1, is a factor of 20 improvement over the best previous bound [27]. Since the charge anomaly of the antiproton is bound [7] within 7×10^{-10}, our result and the assumption of charge superposition limits the charge anomaly of the positron to 10^{-9}, an improvement over the best previous bound on the positron anomaly by a factor of 25 [93, 94].

If we take $Q = 0$ and change the equation of motion, Eq. 7.2 to

$$M\ddot{\vec{x}}(t) = \vec{\nabla}[\vec{\mu} \cdot \vec{B}(\vec{x})] + \alpha\vec{\nabla}|\vec{E}(\vec{x}, t)|^2 \tag{7.24}$$

we can interpret our experiment as setting a bound on the polarizability, α, of the antihydrogen under the assumption that antihydrogen has no charge. This sets a limit (without incorporating systematic uncertainties) of $\alpha < 3 \times 10^6\ \alpha_0$, at 68.3% confidence, where $\alpha_0 = 4\pi\epsilon_0\ 0.667 \times 10^{-30}\ \mathrm{m}^3$ is the known polarizability of hydrogen [112].

If instead we take $Q = 0$ and $\alpha = 0$, and change the equation of motion to

$$M\ddot{\vec{x}}(t) = \vec{\nabla}[\vec{\mu} \cdot \vec{B}(\vec{x})] + \vec{\nabla}[\vec{p} \cdot \vec{E}(\vec{x}, t)], \tag{7.25}$$

we can interpret our experiment as setting a bound on the permanent electric dipole moment of antihydrogen, $\vec{p}$, under the assumption that antihydrogen has no charge and polarizability. This gives a bound (again, without incorporating systematic uncertainties) of $10^{-9}\ e$ cm at 68.3% confidence. If this moment was to come from, for instance, a positron dipole moment, quantum corrections loosen this bound to about $10^{-5}\ e$ cm [113]. This is many orders of magnitude greater than the current bound on the electron dipole moment of $9.7 \times 10^{-29}\ e$ cm [114]. In principle, our analysis should be done to set a joint bound on the charge, polarizability, and permanent electric dipole moment, but as these bounds are far from the expected values of the quantities (by invoking the CPT theorem), we do not feel this is an informative analysis with this data set.

A new measurement could improve the charge (and dipole) limit, guided by the scaling in Eq. 7.1. Increasing the number of potential transitions that an antihydrogen experiences

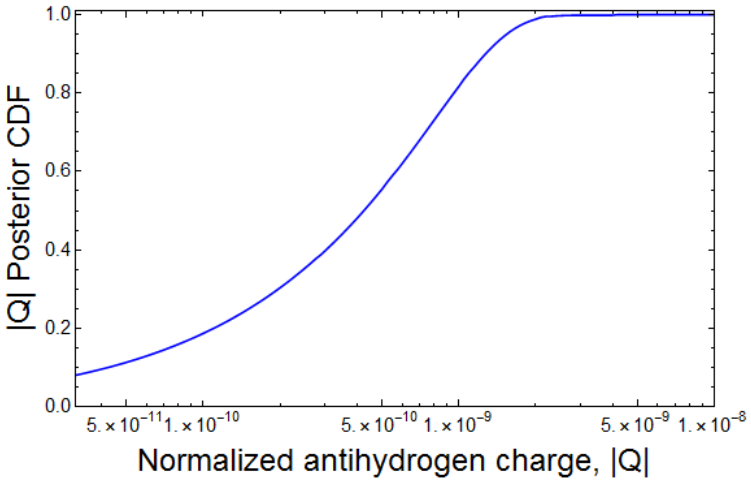

Figure 7.3: The cumulative distribution function (CDF) of the posterior prediction of $|Q|$. The value of $|Q|$ where the CDF is $1-\alpha$ gives the upper limit on $|Q|$ given the measurement and analysis (with statistical and systematic uncertainties incorporated) at $1-\alpha$ confidence.

will improve the limit, but this improvement scales as a square root. The potential $\Delta\Phi$ is limited by the electrodes and is not expected to be able to be increased. The neutral well depth E_{well} is limited by the temperature of the antihydrogen such that sufficiently many anti-atoms are trapped. With improved mixing techniques and using methods such as adiabatic or laser cooling, E_{well} could be reduced [115]. A better characterization of the initial energy distribution of antihydrogen would reduce the largest systematic effect and reduce the bound. Increased statistics in the null and stochastic trials would allow for a slight improvement in the charge bound, but is not expected to improve the measurement significantly.

Acknowledgements

I would like to thank Andrew E. Charman for useful discussions and guidance on the statistical analysis presented here.

	Q^* ($\times 10^9$)		
Uncertainties Incorporated	$1\,\sigma$	90%	99%
Data	0.59	1.0	1.5
Data+MC	0.60	1.0	1.5
Data+MC+Systematics	0.71	1.3	2.0

Table 7.2: Summary of probabilistic limits, determined from $P(|Q| \leq Q^*|\mathcal{D}, \mathcal{M}, \mathcal{I})$, on the normalized charge magnitude $|Q|$, expressed in parts per billion. The three rows reflect various sources of uncertainty incorporated into the posterior bound on $|Q|$. The three columns represent different confidence levels, meaning probabilities for the Bayesian posterior credible regions. The first row (Data) includes only the posterior uncertainty upon conditioning on the limited number of actual observations, and assumes perfect knowledge of the survival (s versus $|Q|$) curve. The second row (Data+MC) also incorporates uncertainty due to the fact that the survival curve has been fit using a large but finite number of Monte Carlo samples at a finite number of Q's. The third row (Data+MC+Systematics) incorporates additional systematic uncertainty as to the shape of the survival curve, due primarily to the choice of the ensemble of initial conditions.

Chapter 8

Bounding the antihydrogen mass

In this chapter we describe design considerations for ALPHA-g, a new antihydrogen trap optimized to make measurements of f, the gravitational to inertial mass ratio of antihydrogen. Here, we focus on a toy model of the trap and consider analysis of potential data from the trap. Significant design changes have been considered by C. So and J. Fajans, and ultimately our toy model should be replaced with their new design to confirm the results.

Since the gravitational to inertial mass ratio is known to be extremely close to 1 for hydrogen, by the CPT theorem we would expect $f = 1$ [7]. However, no precise measurements of the gravitational antihydrogen mass have been made. The ALPHA experiment placed a loose bound of $-65 < f < 110$ at 95% confidence [18]. With a new trap, ALPHA-g, designed specifically for mass measurements, we hope to improve this limit so that the uncertainty in f is less than 1%.

We find that, in our simplified model of the experiment, a $\sim$1% measurement with antiatoms with temperature 100 K and 500 annihilations, is achievable.

8.1 ALPHA-g apparatus

The ALPHA-g apparatus is quite similar to the ALPHA-II apparatus described in Sec. 2.2. The primary difference is that ALPHA-g will be oriented vertically so that there is a gravitational potential gradient between the two ends of the trap. In addition to the octupole, mirror coils, and solenoid, ALPHA-g will have a gradient coil that will cancel out the effect of the gravitational gradient to make the potential flat inside the trap if the mass of antihydrogen is the same as hydrogen. However, if the antihydrogen mass differs from the hydrogen mass, the potential will not look flat and the effect will be observable as a difference in the number of antihydrogen that leave the top versus bottom of the trap. The gradient coil has field

$$\vec{B} = \frac{\chi M g}{\mu}\left(\vec{z} - \frac{\vec{x}+\vec{y}}{2}\right), \tag{8.1}$$

where g is the acceleration due to gravity (assumed constant over the extent of the trap), M is the inertial mass of hydrogen (which we equate with the inertial mass of antihydrogen), and μ is the magnitude of the magnetic moment of antihydrogen. Here, χ is an adjustable parameter, used to tune the number of particles escaping the trap through the top versus the bottom. While the sensitivity of the simulated measurements are largely unaffected by χ, the measurements are more symmetric about $f = 1$ for χ tuned such that an equal number of antihydrogen escape from the top and bottom. $\chi = 1$ does not have equal numbers of anti-atoms going up and down as there are other magnetic fields that also contribute a z gradient. This also means that a tuned value of χ is dependent on the magnitude of the confining fields.

8.2 Simulations

The simulation used here is similar to the one described in Sec. 7.2, and we will only highlight the different aspects. The equation of motion, Eq. 7.2, is modified to

$$M\ddot{\vec{x}} = \vec{\nabla}[\vec{\mu} \cdot \vec{B}] - fM\vec{g}. \tag{8.2}$$

Variations in g over the trap have negligible contributions for 1% measurements, but could be important for precision measurements at later stages of the experiment.

In the simulations, we choose values for experimental parameters that would be reasonable for ALPHA-g. We choose the solenoid field to be $B_{sol} = 0.5$ T. The length of the octupole is chosen to be 1.6 m, and the distance between the mirror coils is chosen to be 1.4 m. Ultimately, as ALPHA-g designs are finalized, these simulations will need to be repeated with actual design parameters.

Simulated anti-atoms are launched in a volume corresponding to the volume of the positrons in the experiment. The antihydrogen have an energy distribution obeying a 3D Maxwellian distribution at a temperature T. In this simulation, the temperature is an adjustable parameter. We study what T is necessary to achieve a desired sensitivity in the antihydrogen mass. We modify the octupole and mirror fields correspondingly with temperature, such that $\mu\Delta|\vec{B}| = T$, where Δ denotes a difference in magnetic between the edge of the trap and the center. The anti-atoms are then allowed to randomize for 1 s. After the randomization, the mirror fields are shut down with an exponential shape with a time constant of 10 s. The times at which the anti-atoms leave the trap are then stored.

For each value of f, we simulate $N \sim 10000$ annihilation events $\mathcal{Z} = \{(z_i, t_i)|i \in 1, ..., N\}$, where z_i is the vertical annihilation position and t_i is the annihilation time. To simulate an experimental run, η annihilation events are randomly drawn (without replacement) from $\mathcal{Z}$ to form a set $\xi_j = \{(z_k, t_k)|k \in 1, ..., \eta\}$ [63]. We assumed experimental runs with $\eta = 100, 500$, or 1000 annihilations. We generate $K = 1000$ of these experimental runs (subsamplings) for the set $\Xi = \{\xi_j|j \in 1, ..., K\}$. Note that since N is generally smaller than $K\eta$, many samples from $\mathcal{Z}$ can be found in more than one collection from Ξ.

8.3 A statistical bound on the antihydrogen mass

We would like to determine a confidence belt at $1-\alpha$ confidence for f, the ratio of the gravitational to inertial mass of antihydrogen. We form the confidence intervals by a Neyman construction [116]. See details on the Neyman construction in Sec. C.4.

An issue that plagues the construction of a confidence interval for f is systematic effects. The obtained limits are sensitive to temperatures, particle distributions, and the magnetic fields in ALPHA-g. The anticipated mass measurement sensitivities are estimated assuming perfect knowledge, and thus an experimental measurement will have poorer sensitivity. These simulation studies should be repeated for a variety of possible scenarios to determine how each systematic factor affects the final result. In particular, if the temperature is unknown, it will need to be treated as a nuisance parameter, and more sophisticated statistical methods may be necessary. Further, one must justify and employ properly devised experimental techniques to best ensure that a reverse cumulative average over the period during which the data is acquired is justified. In particular, it is vital to rule out wild fluctuations in temperature and field configurations during data acquisition.

Reverse Cumulative Average

This work below is an extension of the reverse cumulative average technique developed in Ref. [18]. We build a *reverse cumulative average* value of the vertical annihilation point for each ξ_j (a collection of η annihilation times and vertical positions) in Ξ. The reverse cumulative average $\langle z\rangle_j(t)$ is defined for each ξ_j as

$$\langle z\rangle_j(t) \equiv \frac{\sum_{(z_k,t_k)\in\xi_j, t_k\geq t} z_k}{\sum_{(z_k,t_k)\in\xi_j, t_k\geq t} 1}. \tag{8.3}$$

This is the average of the z positions of all annihilations that occur at time t or later. This is a useful definition as it captures the dynamics of the anti-atoms that are most sensitive to the gravitational field at late times, while still including the majority of anti-atoms that leave the trap before these late times. The reverse cumulative average thus highlights the temporal regions with the most gravity information in a way that minimizes the uncertainty from poor statistics.

We determine the confidence belt, corresponding to where $1-2\alpha$ of the reverse cumulate averages lie, as follows. For each time t, and each hypothetical value of the ratio of gravitational to inertial mass of antihydrogen, f, the reverse cumulative averages of all of the $\langle z\rangle_j(t)$ corresponding to each $\xi_j \in \Xi$ are sorted. Then, two error bounds $z^\alpha_<(t)$ and $z^\alpha_>(t)$ are chosen in such a way that precisely αK cumulative averages are below $z^\alpha_<(t)$ and αK cumulative averages are above $z^\alpha_>(t)$. We call this the *acceptance region*, and examples are shown in Fig. 8.1. Note that K is the number of simulated experiments, and the confidence belts are true confidence belts, in the frequentist sense, in the $K\to\infty$ limit.

To obtain a confidence interval, we first choose a value of t, called t^*. In principle t^* could be any time, but there will be certain times that distinguish masses better than other times.

Temperature (mK)	Number of annihilations	Sensitivity (at 90% confidence)
20	100	∼0.8%
20	500	∼0.3%
20	1000	∼0.2%
100	100	∼2%
100	500	∼1%
100	1000	∼0.7%
300	100	∼6%
300	500	∼2%
300	1000	∼2%

Table 8.1: The statistical sensitivity of antihydrogen mass measurement achievable on ALPHA-g as a function of antihydrogen temperature (T) and number of annihilations (η). Here, $t^* = 45$ s. This analysis ignores any systematic effects.

If t^* is chosen to be large, then the distinguishability is limited by poor statistics. Only a small fraction of the η annihilations are used to place limits, which yields large variances. If t^* is chosen to be small, the statistics are better, but many of the particles are too energetic to be noticeably influenced by the gravitational potential. Thus, an intermediate value of t^* is optimal. We report later on some choices of t^*, but our results are relatively insensitive to specific but sensible choices of t^*.

A simulated experimental run yields a measurement $\langle z \rangle_{exp}(t^*)$. A value of the mass ratio of antihydrogen to hydrogen, f, is accepted at $1 - 2\alpha$ confidence if it is true that $z^\alpha_<(t^*) < \langle z \rangle_{exp}(t^*) < z^\alpha_>(t^*)$. The collection of the accepted values of f become the frequentist confidence interval for f at $1-2\alpha$ confidence. Note that the value of t^* must be chosen prior to looking at actual data. Failure to do this will result coverage issues, with the resultant interval not being significant to the claimed level of precision [6].

As with any Monte Carlo estimate, we need to be careful that $z^\alpha_<(t)$ and $z^\alpha_>(t)$ are not affected by the finite number of simulations performed to sample the distribution. This is true when $N \gg \eta$, and by testing various values of N we have found we are in a regime where the bounds are not sensitive to N. We also found that the dependence on N is most significant at large t. Thus, as long as t^* is chosen not to be large (which is a poor choice for reasons outlined earlier), the finite numbers are irrelevant.

In Table 8.1 we show possible precision of mass measurements under this analysis for various values of temperature, T, and number of annihilations, η. We highlight some of the acceptance regions and confidence intervals used to obtain these results below.

In Fig. 8.1, we compare the 90% acceptance regions when the temperature, number of annihilations, and χ are varied. With $T = 100$ mK and $\eta = 500$, there is no overlap between the $f = 0.98$, $f = 1.00$, and $f = 1.02$ bands if $t^* < 55$ s. This means $f = 1.00$ can be distinguished from $f = 0.98$ and $f = 1.02$ (at 90% confidence), and a $< 4\%$ measurement is

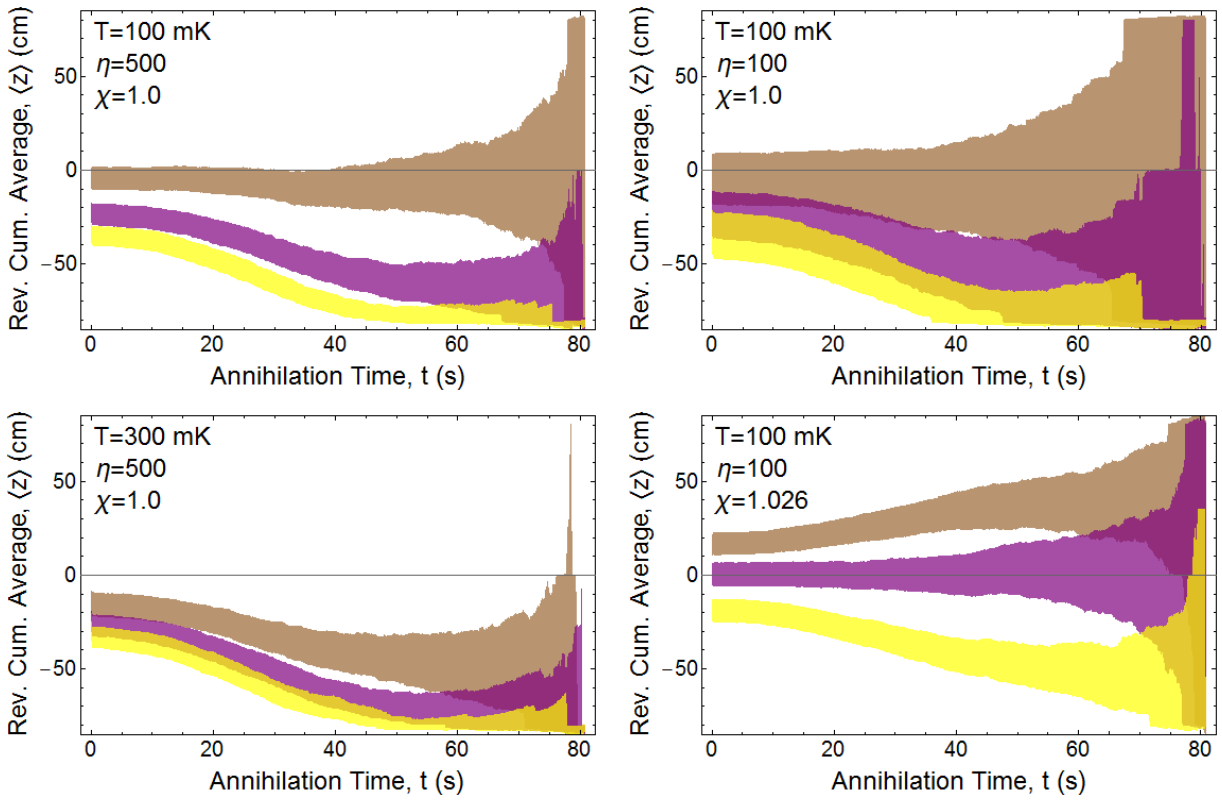

Figure 8.1: The 90% acceptance regions of the reverse cumulative average as temperature, number of annihilations, and χ are varied. The brown band is the 90% acceptance region for $f = 0.98$, the purple band for $f = 1.00$ and the yellow band for $f = 1.02$ for the parameters shown in the plots. Unless otherwise noted, $\chi = 1$. Different values of f can be distinguished for a particular t^* if there is no overlap between the bands at $t = t^*$.

feasible. The acceptance regions are also distinguishable when χ is adjusted to make an equal number of atoms go up and down when $f = 1.00$. However, in this case the distance between each of the bands is roughly equidistant whereas the $f = 1.02$ and $f = 1.00$ acceptance regions were close in the $\chi = 1$ case.

Raising the temperature to 300 mK while keeping $\eta = 500$ shifts the distributions toward lower values of $\langle z \rangle$, but does not widen the acceptance regions. The shift causes the $f = 1.02$ and $f = 1.00$ acceptance regions to overlap, meaning that these two values of f cannot be distinguished. With an appropriate choice of t^*, $f = 0.98$ can be distinguished from $f = 1.00$.

When $\eta = 100$, the acceptance regions widen compared to the $\eta = 500$ case. For $T = 100$ mK and $\eta = 100$, there is now significant overlap between the $f = 1.00$ and $f = 1.02$ bands, meaning these cannot be distinguished. There are regions where the $f = 1.00$ and $f = 0.98$ bands do not overlap, meaning these can be distinguished with an appropriate choice of t^*.

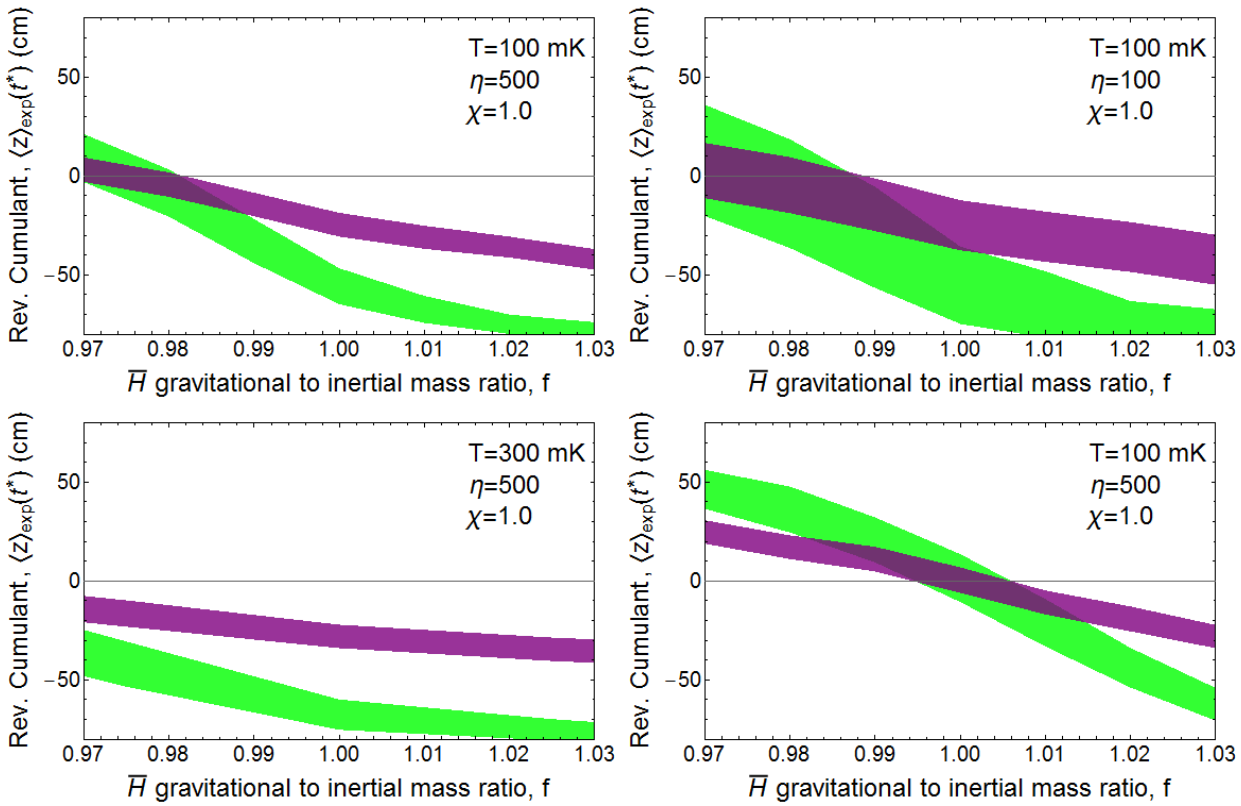

Figure 8.2: The 90% confidence belts for f for the parameters shown in the plots. The purple band has $t^* = 5$ s and the green band has $t^* = 45$ s. Unless otherwise noted, $\chi = 1$. An analysis of the experimental data would yield a value for the reverse cumulant $\langle z \rangle_{exp}(t^*)$. This is drawn as a horizontal line on the corresponding plot, and the overlap between the line and the confidence belt is the 90% confidence interval for f.

In Fig. 8.2, we construct 90% confidence belts for f for some of the scenarios in Table 1. The width of the confidence belt corresponds to the sensitivity of the experiment. The sensitivity achievable in each scenario is the width of the confidence belts near $f = 1.00$. The sensitivity with $t^* = 5$ s and $t^* = 45$ s are comparable, with $t^* = 45$ s performing slightly better. We see that adjusting χ such that the upgoing number of anti-atoms matches the downgoing number centers the confidence intervals around $f = 1.00$, but does not seem to affect the sensitivity.

The results in Table 8.1 show that, at least in this idealized model of ALPHA-g and the analysis, an up/down (100% sensitivity) measurement should certainly be feasible. In fact, with ~ 100 particles, measurements to bound the antihydrogen mass to within a few percent of the hydrogen mass seem possible.

Future work will reconsider these analyses with a more accurate model of the trap, and will consider inclusion of systematic effects such as the uncertainty in antihydrogen

temperature. Ref. [18] presents an alternate analysis method related to the Fisher criterion. The difference in statistical power of that test compared to the reverse cumulative average test presented here would be an interesting future study. As highlighted earlier, it is crucial to choose (and stick to) an analysis method *before* any data is taken such that a frequentist analysis is valid.

Acknowledgements

I would like to thank Andrey Zhmoginov for his collaboration on the content in this chapter.

Chapter 9

Conclusion

In this thesis, we have demonstrated three main ideas. First, we improve on the analysis of plasma data for diagnostics of temperature and density. Second, we analyzed the cooling of electron plasmas when trapped inside of a microwave cavity. Last, we described measurements of charge and mass of antihydrogen.

Many directions for future work in these areas are discussed in each chapter, and we only touch on the most important aspects here. For the diagnostics, when the new codes for temperature and plasma profile are written, they will need to be tested against a plethora of images and temperature dumps. Particularly, discovering use cases that will result in problems for the algorithm and rectifying these is an engineering challenge. Parameters of the method such as the bin width for the discrete, binned temperature fitter will need to be optimized. Ultimately, we will have to show that these new methods work better than the currently accepted methods.

Both the diagnostics and cavity cooling problem can benefit from careful simulations of plasma transport. The diagnostic problem has the simpler problem of determining how the plasma profile changes as the potential wells confining the plasma changes. The cooling problem must consider the effects of collisions and transport in inhomogenous magnetic fields. One approach to this problem would be to extend the validity of PIC codes to colder temperatures through approximations, such that the Debye length does not become a computational limitation. Another approach would be to develop an N-body simulation, but this has challenges such as the Green's function of the Laplace equation being complicated inside the Penning-Malmberg trap.

An extension of the analytic methods of this thesis that carefully consider collisions would also be fruitful. Experimentally, collisions do not seem to be a limitation in the temperature diagnostic presented in Chap. 5, yet our naive analysis there would suggest that this is a limitation. We assume that collisions trivially convert perpendicular cooling to parallel cooling in the cavity cooling problem, but there may be effects ignored by this simplification. In addition, the coupling of the superradiant mode and the cavity could be affected by collisions, which we have not considered.

For measurements of antihydrogen properties, a better charge measurement is an obvious

next step. The trapping rate at ALPHA has improved significantly since the measurement was conducted, and this could reduce some of the statistical uncertainties. While this yields a slight improvement, ultimately longer hold times will be necessary for order of magnitude improvements on the charge limit. Significant design work is under way for ALPHA-g, and progress is being made to completing the apparatus.

Appendix A

Hamiltonian Dynamics of Charged Particles

A.1 Adiabatic Invariants

Consider a Hamiltonian, H, depending on position coordinates $\vec{q}$ and momentum coordinates $\vec{p}$ [117]. For any pair of cyclic coordinates, consider

$$A = \oint p dq = \int_{q_t(t)}^{q_t(t+\tau)} p dq \tag{A.1}$$

Where q_t is the position coordinate at the turning point where $dq/dt = 0$, and τ is the time between periods. Now, suppose there is a parameter λ in the Hamiltonian that varies slowly with respect to timescales in the problem. Define $E = H(p(E(t), Q, \lambda(t)), q, \lambda(t))$. Look at

$$\begin{aligned} \frac{dA}{dt} &= p\frac{dq}{dt}\bigg|_{q=q_t(t)}^{q=q_t(t+\tau)} + \int_{q_t(t)}^{q_t(t+\tau)} \left(\frac{\partial p}{\partial t}\right)_q dq \\ &= \int_{q_t(t)}^{q_t(t+\tau)} \left(\left(\frac{\partial p}{\partial E}\right)_{q,\lambda} \frac{dE}{dt} + \left(\frac{\partial p}{\partial \lambda}\right)_{q,E} \frac{d\lambda}{dt} \right) dq \end{aligned} \tag{A.2}$$

We see that

$$\frac{dE}{dt} = \frac{\partial H}{\partial p}\frac{dp}{dt} + \frac{\partial H}{\partial q}\frac{dq}{dt} + \frac{\partial H}{\partial \lambda}\frac{d\lambda}{dt} = \frac{\partial H}{\partial \lambda}\frac{d\lambda}{dt} \tag{A.3}$$

where the second equality is true by Hamilton's equations. Also, since E has no explicit λ dependence,

$$\begin{aligned} 0 &= \frac{\partial E}{\partial \lambda} = \frac{\partial H}{\partial p}\left(\frac{\partial p}{\partial \lambda}\right)_{q,E} + \frac{\partial H}{\partial \lambda} \\ 1 &= \frac{\partial E}{\partial E} = \frac{\partial H}{\partial p}\left(\frac{\partial p}{\partial E}\right)_{q,\lambda} \end{aligned} \tag{A.4}$$

Plugging these relations in to Eq. A.2 gives

$$\frac{dA}{dt} = \int_{q_t(t)}^{q_t(t+\tau)} \left(\left(\frac{\partial p}{\partial E}\right)_{q,\lambda} \frac{\partial H}{\partial \lambda} + \left(\frac{\partial p}{\partial \lambda}\right)_{q,E}\right) \frac{d\lambda}{dt} dq. \tag{A.5}$$

Which is zero by Eq. A.4. Thus, $dA/dt \approx 0$ and A is approximately conserved.

A is actually conserved to a high-degree of accuracy [30]. Assuming the Hamiltonian can be separated into a cyclic part and a perturbation parameterized by a small parameter ϵ, A is conserved to a time that is exponential in ϵ, as long as the potentials are well-behaved [118]. This can be calculated analogously to calculating the quantum reflection coefficient in the WKB approximation [119, 120].

In Strongly Magnetized Plasmas

In magnetized plasmas, adiabatic invariants are particularly useful to characterize plasma properties across changing properties of the system. Often, they can offer insights into the behavior of a system without doing any calculations.

First, Consider a plasma in a static, strong magnetic field pointing in the z direction. Plugging in the canonical momentum $p_\theta = mr^2\dot{\theta} - er^2B/2$, we see that

$$A = \oint p_\theta d\theta = \pi m \frac{u_\perp^2}{\omega_c}. \tag{A.6}$$

Thus, for a constant magnetic field (as in a solenoid) the kinetic energy in the perpendicular direction is an adiabatic invariant [30].

Considering the motion parallel to the magnetic field gives

$$J = \oint p_z dz.$$

For a Penning-Malmberg trap, this invariant is useful in characterizing bounce motion.

A.2 Effective Hamiltonian

Consider a Hamiltonian system depending on a variable ξ that is cyclic over a fast timescale, and η that evolves over a slower timescale. As before, define [121, 122]

$$\mathcal{J} = \oint d\xi \; p_\xi \tag{A.7}$$

Note that this definition is only reasonable if ξ is fast compared to other timescales in the problem. Otherwise the orbits would not close and the integral would be ill-defined. We see that

$$\frac{\partial \mathcal{J}}{\partial H} = \oint d\xi \; \frac{\partial p_\xi}{\partial H} = \oint dt \equiv \Delta t. \tag{A.8}$$

This is the period of one cycle in ξ. We also see

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial \eta} &= \oint d\xi \; \frac{\partial p_\xi}{\partial \eta} = \oint d\xi \; \frac{\partial p_\xi}{\partial H}\frac{\partial H}{\partial \eta} = -\oint dt \; \frac{dp_\eta}{dt} \equiv -\Delta p_\eta \\ \frac{\partial \mathcal{J}}{\partial p_\eta} &= \oint d\xi \; \frac{\partial p_\xi}{\partial p_\eta} = \oint d\xi \; \frac{\partial p_\xi}{\partial H}\frac{\partial H}{\partial p_\eta} = \oint dt \; \frac{d\eta}{dt} \equiv \Delta \eta \end{aligned} \tag{A.9}$$

Thus, this gives a prescription for a guiding center Hamiltonian. $\mathcal{J}$ acts like a Hamiltonian with time redefined by Δt given above.

Drifts

Consider a particle in an azimuthally symmetric magnetic field that can vary in r and z. The hamiltonian for a particle in this field is

$$H = \frac{p_r^2}{2m} + \frac{(p_\theta + eA_\theta(r,z))^2}{2m} + \frac{p_z^2}{2m} - e\Phi(\vec{x}) \tag{A.10}$$

We define the guiding center coordinate to be the solution to

$$p_\theta + eA_\theta(r_{gc}(p_\theta, z), z) = 0. \tag{A.11}$$

Expand the scalar and vector potentials to first order about r_{gc} and evaluate all quantities at the guiding center coordinate. Now consider the integral of motion in the r direction

$$\begin{aligned} \mathcal{J} &= \oint dr \; p_r \\ &\approx \oint dr \; \sqrt{2mH - p_z^2 - e^2B_z^2\delta r^2 + 2me\Phi + 2meE_r\delta r} \\ &= \oint dr \; \sqrt{2m\left(H - \frac{p_z^2}{2m} + e\Phi + \frac{m}{2}\frac{E_r^2}{B_z^2}\right) + e^2B_z^2\left(\delta r - \frac{mE_r}{eB_z}\right)^2} \\ &= \frac{4\pi m}{eB_z}\left(H - \frac{p_z^2}{2m} + e\Phi + \frac{m}{2}\frac{E_r^2}{B_z^2}\right) \end{aligned} \tag{A.12}$$

The last line is clear as this is the area of a circle with radius squared equal to the first term. Assuming an azimuthally symmetric potential and using the negative of this as an effective hamiltonian gives the following velocity equations

$$\begin{aligned}\frac{\Delta\theta}{\Delta t} &\approx \frac{H_\perp}{eB_z^2}\frac{\partial B_z}{\partial r} - \frac{E_r}{B_z} - \frac{m}{eB_z}\frac{E_r}{B_z}\frac{\partial}{\partial r}\left(\frac{E_r}{B_z}\right)\\ \frac{\Delta r_{gc}}{\Delta t} &\approx 0\\ \frac{\Delta z}{\Delta t} &= \frac{p_z}{m}.\end{aligned} \tag{A.13}$$

The terms in the first equation are the $\vec{\nabla}|\vec{B}|$ drift and the $\vec{E}\times\vec{B}$ drifts. The third term is a $\vec{\nabla}|\vec{E}|$ drift that can be attributed to the definition of r_{gc} [123]. We see the radial and z velocity are simple in this level of approximation. This also shows that in strong magnetic fields, particles will stick to field line as their radial guiding center coordinate is constrained.

Higher order drifts can be obtained by keeping further terms in these expansions as well as expanding the scalar and vector potential to higher orders around the guiding center coordinate in Eq. A.12.

Example: Particles in a changing field

Suppose charged particles are in a strong magnetic field. These particles start in a region of high field and travel to a region of lower field (though still large enough to be strongly magnetized). As shown above, the particles will follow the magnetic field lines. Assuming no azimuthal dependence in the magnetic field, Gauss's law for magnetism requires

$$\frac{\partial B_z}{\partial z} + \frac{1}{r}\frac{\partial}{\partial r}\left(rB_r\right) = 0. \tag{A.14}$$

Note that with the substitution

$$B_r = -\frac{r}{2}\frac{\partial B_z}{\partial z}, \tag{A.15}$$

Eq. A.14 is satisfied, as long as $r\frac{\partial^2 B_z}{\partial z \partial r} \ll \frac{\partial B_z}{\partial z}$. For example, this is true in an ideal solenoid [72]. A field line satisfies $dz/dr = B_z/B_r$, so plugging this form in gives

$$\begin{aligned}\frac{dz}{dr} &= -B_z\frac{2}{r}\left(\frac{\partial B_z}{\partial z}\right)^{-1}\\ \frac{dB_z}{B_z} &= -\frac{2dr}{r}\\ \frac{B_0}{B_1} &= \frac{r_1^2}{r_0^2}.\end{aligned} \tag{A.16}$$

Thus, the radial position of the charged particles increases by a factor of $\sqrt{B_0/B_1}$ in traveling from the region of strong field to lower field. This means that the radius of a plasma will expand by the same factor when a plasma travels from a region of strong field to lower field.

Appendix B

Vlasov Equation and Equilibria

B.1 Derivation of the Vlasov equation

Here we will derive the Vlasov equation, the PDE for the time evolution of the plasma distribution. We will follow [124], and further details are given in the reference. Assume a single component plasma of particles with charge $-e$ subjected to homogenous applied electric and magnetic fields, $\vec{E}$ and $\vec{B}$. Consider one of the particles in the plasma. Call the 6-vector of this particle's position and velocity $x^\mu = (\vec{r}, \vec{v})$. The equation of motion for this particle is

$$\frac{dx^\mu}{dt} = \begin{pmatrix} \vec{v} \\ -\frac{e}{m}(\vec{E} + \vec{v} \times \vec{B}) - e\sum_i G(\vec{r} - \vec{r}_i) \end{pmatrix} \tag{B.1}$$

where the sum is over all of the other particles in the plasma. G is the Green's function for the Laplace equation with appropriate boundary conditions. Now, define $f(\vec{r}, \vec{v}) = \sum_i \delta^3(\vec{r} - \vec{r}_i)\delta^3(\vec{v} - \vec{v}_i)$. The above equation becomes

$$\frac{dx^\mu}{dt} = \begin{pmatrix} \vec{v} \\ -\frac{e}{m}(\vec{E} + \vec{v} \times \vec{B}) - e\int d^3\vec{r}' \int d^3\vec{v}' \;\; f(\vec{r}', \vec{v}')G(\vec{r} - \vec{r}') \end{pmatrix} \tag{B.2}$$

Assume that the ensemble average of $f(\vec{r}, \vec{v})$ is sufficient in describing the plasma, and thus replace the $f(\vec{r}, \vec{v})$ in the expression above with its ensemble average. Note that this is the key simplification and departures from this simplification lead to the Fokker-Planck terms not present in the Vlasov equation. Then, we can write

$$\frac{dx^\mu}{dt} = a^\mu = \begin{pmatrix} \vec{v} \\ -\frac{e}{m}(\vec{E} + \vec{E}_{self} + \vec{v} \times \vec{B}) \end{pmatrix}, \tag{B.3}$$

where $\vec{E}_{self}$ is the integral over the Green's function. Define the probability density of the particle coordinate $x^\mu(t)$ as

$$P = \langle \delta^6(x^\mu - x^\mu(t)) \rangle, \tag{B.4}$$

where the brackets denote an ensemble average. Let $\tilde{P}$ be the Fourier transform of P. We see

$$\tilde{P} = \langle e^{-ik_\mu x^\mu(t)} \rangle. \tag{B.5}$$

Using the equation of motion, we get that the change in $\tilde{P}$ by a first order perturbation in dt is

$$\begin{aligned} \delta\tilde{P} &= \langle e^{-ik_\mu x^\mu(t+dt)} \rangle - \langle e^{-ik_\mu x^\mu(t)} \rangle \\ &\approx \langle e^{-ik_\mu x^\mu(t)}(1 - ik_\nu a^\nu dt) \rangle - \langle e^{-ik_\mu x^\mu(t)} \rangle \\ &= -ik_\nu dt \langle e^{-ik_\mu x^\mu(t)} a^\nu \rangle \\ &= -dt\ \mathcal{F}(\partial_\mu(a^\mu P)). \end{aligned} \tag{B.6}$$

where $\mathcal{F}$ denotes the Fourier transform. Taking the inverse transform gives

$$\frac{\partial P}{\partial t} = -\partial_\mu(a^\mu P).$$

Rewriting in terms of $\vec{r}$ and $\vec{v}$ gives

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial \vec{r}} \cdot (\vec{v}\ P) - \frac{\partial}{\partial \vec{v}} \cdot \left(-\frac{e}{m}(\vec{E} + \vec{E}_{self} + \vec{v} \times \vec{B})P \right).$$

Note that

$$\begin{aligned} \frac{\partial}{\partial \vec{v}} \cdot (\vec{v} \times \vec{B})\ P &= \vec{B}\ P \cdot \left(\frac{\partial}{\partial \vec{v}} \times \vec{v} \right) - \vec{v} \cdot \frac{\partial}{\partial \vec{v}} \times (\vec{B}\ P) \\ &= \vec{v} \cdot \vec{B} \times \frac{\partial P}{\partial \vec{v}} \\ &= (\vec{v} \times \vec{B}) \cdot \frac{\partial P}{\partial \vec{v}}. \end{aligned}$$

Thus

$$\frac{\partial P}{\partial t} + \vec{v} \cdot \frac{\partial P}{\partial \vec{r}} - \frac{e}{m}(\vec{E} + \vec{E}_{self} + \vec{v} \times \vec{B}) \cdot \frac{\partial P}{\partial \vec{v}} = 0$$

Since no particle in the plasma is distinguishable, it must be that the distribution of one particle is equivalent to the distribution of any other particle. Hence, we identify $P \propto f$, and the above equation becomes the Vlasov equation

$$\frac{\partial f}{\partial t} + \vec{v} \cdot \frac{\partial f}{\partial \vec{r}} - \frac{e}{m}(\vec{E}_{tot} + \vec{v} \times \vec{B}) \cdot \frac{\partial f}{\partial \vec{v}} = 0. \tag{B.7}$$

$\vec{E}_{tot} = \vec{E} + \vec{E}_{self}$ and must satisfy Maxwell's equations, such that

$$\vec{\nabla} \cdot \vec{E}_{tot} = -\Delta V = -e \int d^3\vec{v} \;\; f(\vec{r}, \vec{v}), \tag{B.8}$$

which is of the form in Eq. B.2 if the Green's function is known. We also assumed here that the magnetic fields produced by the plasma are negligible compared to the applied fields.

B.2 Equilibrium Solutions

We see that plugging in any distribution of the form $f(H)$ satisfies the Vlasov and Laplace equations because of Hamilton's equations (the coordinates are still for the one particle in the previous section, so the argument is valid) [121]. Similarly, if the Hamiltonian does not depend on θ (it is azimuthally symmetric), the canonical angular momentum p_θ will be a conserved quantity. For example, in an ideal solenoid field, $p_\theta = mr(v_\theta - r\omega_c/2)$ is conserved. A general solution will depend on the linear combination of the two, and hence will be of the form $f(H - \omega p_\theta)$, which is called a "rigid rotor" plasma [125].

Appealing to the second law of thermodynamics, we would expect the maximum entropy distribution of this class of distributions to be the physical solution to the Vlasov equation (see details in Appendix C.1) [91]. Assuming the energy has mean T, the form is

$$f(\vec{r}, \vec{v}) \propto \exp\left(-\frac{H - \omega p_\theta}{T}\right). \tag{B.9}$$

Integrating over all velocities (and ignoring relativistic effects) gives

$$n(r) = \int d^3\vec{v}\, f(\vec{r}, \vec{v}) = n_0 \exp\left(\frac{m\omega(\omega - \omega_c)r^2}{2T} + \frac{eV(\vec{r})}{T}\right), \tag{B.10}$$

where n_0 is a normalization constant related to the plasma density. Including the effects of collisions ignored earlier yields the full Fokker-Planck equation with friction and diffusion terms. These terms are zero when the distribution has the form of Eq. B.9, so this form is justifiable without a maximum entropy principle as well [30].

Debye Shielding

Let there be a point charge with charge $-e$ at a position $\vec{r}'$ inside of a non-rotating electron plasma in a weak magnetic field. To find the form of $V(\vec{r})$, we look at the Laplace equation for the potential, plugging in the form of Eq. B.10

$$-\Delta V(\vec{r}) = -e\left(n_0 e^{\frac{eV(\vec{r})}{T}} + \delta^3(\vec{r} - \vec{r}')\right), \tag{B.11}$$

Under the condition $|eV(\vec{r})| \ll T$, the equation becomes

$$\Delta V(\vec{r}) = e\left(n_0\left(1+\frac{eV(\vec{r})}{T}\right)+\delta^3(\vec{r}-\vec{r}')\right). \tag{B.12}$$

As a constant in the potential is irrelevant, we redefine the potential to absorb the constant. Then, finding the potential becomes the problem of finding the Green's function of the Helmholtz operator. This is discussed in many texts, including [72]. If there are no boundary conditions, and assuming the plasma is localized, the viable solution should decay far away from the point charge. Thus, the potential is

$$V(\vec{r}) = \frac{e}{4\pi|\vec{r}-\vec{r}'|}e^{-\frac{|\vec{r}-\vec{r}'|}{\lambda_D}}, \tag{B.13}$$

where $\lambda_D = \sqrt{T/n_0e^2}$ is the *Debye length* of the plasma. The charges in the plasma screen the potential of the point charge when $|\vec{r}-\vec{r}'| \gg \lambda_D$.

Plasmas at high temperatures

Consider an extremely hot electron plasma, and take the origin of the coordinate system to be at the center of the plasma. Then, the Laplace equation for the potential becomes

$$-\Delta V(\vec{r}) = -en_0\exp\left(\frac{m\omega(\omega-\omega_c)r^2}{2T}+\frac{eV(\vec{r})}{T}\right) \approx -en_0. \tag{B.14}$$

Assuming the vacuum solution allows for a harmonic potential over the extent of the plasma, we have

$$V(\vec{r}) = -\frac{en_0}{2}z^2 - \frac{en_0}{4}r^2 \tag{B.15}$$

The solution to this, assuming boundary conditions are irrelevant to the shape of the plasma, is

$$f(\vec{r}) = -en_0e^{-\frac{|\vec{r}|^2}{2\lambda_D^2}}. \tag{B.16}$$

So we can expect hot plasmas to have a Gaussian radial profile.

Plasmas at low temperatures

Consider an extremely cold electron plasma. Then from Eq. B.10 it is clear that the following criterion needs to be satisfied for the plasma density to be nonzero anywhere [126].

$$-\frac{m}{2}\omega(\omega-\omega_c)r^2 + eV(\vec{r}) = \Omega(T) \tag{B.17}$$

Let us see if we can find a form where the right hand side is independent of spatial coordinates. Taking a derivative with respect to z shows that V is independent of z. This means that inside

the plasma the plasma "cancels out" the applied external potential to make the potential flat. Taking derivatives with r in the appropriate way to get the Laplacian yields

$$e\Delta V(\vec{r}) = -2m\omega(\omega - \omega_c) = -e^2 n(\vec{r}) \tag{B.18}$$

Solving for the density gives

$$n(\vec{r}) = \frac{2m\omega(\omega - \omega_c)}{e^2} \tag{B.19}$$

and thus the constant density assumption was a consistent one. Intuitively, the plasma is constant so that its self potential will be quadratic in r to cancel out the effective rotation potential. Obviously, since there is finite charge, this relation only holds out to a boundary. Determination of this boundary is discussed later in Sec. 4.1. This means that the plasma has a hard edge.

Appendix C

Statistics

In this appendix we give an overview of statistical ideas that are referred to in the thesis.

C.1 Maximum Entropy Distributions

The second law of thermodynamics states that systems tend toward states with higher entropy. Then, if there is a state that has maximized entropy, this must be the eventual state of the system. Thus, given no other information than a set of constraints for a system, we expect that the system will be in the distribution that maximizes the entropy under those constraints.

The *Shannon entropy* for a probability distribution $p(x)$ is defined, as in statistical physics (with $k_B = 1$) as [127],

$$H = -\int dx\ p(x) \ln(p(x)) \tag{C.1}$$

The Shannon entropy is defined similarly for discrete and multivariate distributions.

Suppose the constraints are expectation of functions $f_j(x)$ set to take values a_j. We wish to extremize

$$\mathcal{L} = -\int dx\ p(x) \ln(p(x)) - \lambda_j \left(a_j - \int dx\ f_j(x) p(x) \right), \tag{C.2}$$

where the λ_j are Lagrange multipliers. Note that $f_0(x) = 1$ and $a_0 = 1$ is always a constraint as probability distributions are normalized. Taking the functional derivative of $\mathcal{L}$ with respect to p gives [121]

$$\frac{\delta \mathcal{L}}{\delta p} = -1 - \ln(p(x)) + \lambda_j f_j(x). \tag{C.3}$$

Setting this to zero extremizes the functional. Solving for $p(x)$ gives

$$p(x) \propto e^{\lambda_j f_j(x)}, \tag{C.4}$$

and the constraint equations can be used to solve for the λ_j.

Fixed mean on $[0, \infty)$

Consider the class of probability distributions $p : [0, \infty) \to [0, 1]$. Denote a random variable drawn from this distribution as H and specify $\mathbb{E}(H) = T$. Then, by Eq. C.4

$$p(H) \propto e^{\lambda H} \tag{C.5}$$

Normalizing and using the condition on the mean gives

$$p(H) \propto e^{-H/T}. \tag{C.6}$$

Thus, the exponential distribution is the maximum entropy distribution over the positive reals with a constrained mean.

Fixed mean and variance on $(-\infty, \infty)^n$

Consider the class of probability distributions $p : \mathbb{R}^n \to [0, 1]$ for $n \in \mathbb{Z}^+$. Denote a random variable drawn from this distribution as $\vec{x}$ and specify $\mathbb{E}(\vec{x}) = \vec{\mu}$ and $\mathbb{E}(|\vec{x} - \vec{\mu}|^2) = \sigma^2$. Then, by Eq. C.4

$$p(\vec{x}) \propto e^{\lambda_1 \vec{x} + \lambda_2 |\vec{x} - \mu|^2} \tag{C.7}$$

Then, normalizing and using the condition on the mean and variance gives

$$p(\vec{x}) \propto e^{-\frac{|\vec{x} - \vec{\mu}|^2}{2\sigma^2}}. \tag{C.8}$$

Thus, the Gaussian distribution is the maximum entropy distribution over $\mathbb{R}^n$ with a constrained mean and variance.

Fixed rate on $\mathbb{N}$

We will show the Poisson distribution is the maximum entropy distribution for a fixed rate over $\mathbb{N}$. While this is possible to show directly, it requires a different entropy function to account for the independence of the events [103, 105]. Thus, we will show the maximum entropy nature of the Bernoulli process, and build up the Poisson distribution from this result.

Consider a process with two possible outcomes. Constrain the first outcome to happen with probability p. This is constrained enough to only allow one probability distribution (the one where the first outcome happens with probability p and the second with probability $1 - p$), and thus this is the maximum entropy distribution.

Now, consider summing up $\mathcal{N}$ Bernoulli processes. Then, for distinguishable events, the probability of k successes would just be $p^k(1-p)^{\mathcal{N}-k}$. However, because the events are indistinguishable, we must account for the equivalence with the binomial coefficient and thus

$$p(k) = \binom{\mathcal{N}}{k} p^k (1-p)^{\mathcal{N}-k}. \tag{C.9}$$

We will reinterpret the first outcome as an event happening and the second outcome as an event not happening. Further, we assume that the first outcome has a low probability of occurring. We take the limit $\mathcal{N} \to \infty$, but with a fixed Poisson rate $\mathcal{N}p \to \lambda$ in the previous equation. Through some manipulation, the above equation becomes [128]

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}. \tag{C.10}$$

So, for a sequence of indistinguishable events, the Poisson distribution is the maximum entropy distribution for a fixed rate over the positive integers.

C.2 Uninformative priors

In this section we briefly discuss the notion of uninformative priors in Bayesian analysis. There are many notions of how "uninformative" should be defined, but we only focus on the Jeffreys notion here. For one parameter distributions like the ones discussed in this thesis, many notions of uninformative are equivalent [129–131].

Jeffreys prior

While a uniform prior may seem to weight all possible values of a parameter equally, it suffers from the fact that under a coordinate transformation, the prior may no longer weight all possible values equally. For example, suppose a probability, $s \in [0, 1]$, is being measured as in Sec. 7.3. One might argue that it was not the probability, but the odds $s/(1-s)$ being measured. A uniform distribution in s is no longer uniform in $s/(1-s)$.

Jeffreys' solution is to construct the prior such that it is invariant under coordinate transformations. For a likelihood function $f(\vec{x}|\boldsymbol{\theta})$, ($\vec{x}$ and $\boldsymbol{\theta}$ can be vectors in different spaces), this is defined as

$$\pi(\boldsymbol{\theta}) = \sqrt{\det \mathbb{E}\left(\frac{\partial^2}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \ln(f(\vec{x}|\boldsymbol{\theta})) \right)} \tag{C.11}$$

The matrix inside of the determinant is the *Fisher information* matrix [102].

Consider the transformation of the prior $\pi(\boldsymbol{\theta})$ into a function of $\boldsymbol{\phi}$. We see

$$\begin{aligned}\pi(\boldsymbol{\phi}) &= \left|\det \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\right| \pi(\boldsymbol{\theta}) \\ &= \sqrt{\left(\det \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\right)^2 \det \mathbb{E}\left(\frac{\partial^2}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \ln(f(\vec{x}))\right)} \\ &= \sqrt{\det \mathbb{E}\left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}} \frac{\partial^2}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \ln(f(\vec{x})) \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\right)} \\ &= \sqrt{\det \mathbb{E}\left(\frac{\partial^2}{\partial \boldsymbol{\phi}_i \partial \boldsymbol{\phi}_j} \ln(f(\vec{x}))\right)}\end{aligned} \tag{C.12}$$

And thus the prior is independent of parameterization.

Consider a Poisson rate λ. That is, in the above $f(x|\lambda) = e^{-\lambda}\lambda^x/x!$. We have

$$\left|\frac{\partial^2 \ln(f(x|\lambda))}{\partial \lambda^2}\right| = \frac{x}{\lambda^2}. \tag{C.13}$$

For a Poisson distribution, $\mathbb{E}(x) = \lambda$, so this gives $\pi(\lambda) \propto 1/\sqrt{\lambda}$.

Consider a binomial process of $\mathcal{N}$ trials with rate p. This means, $f(x|p) \propto p^x(1-p)^{\mathcal{N}-x}$. We have

$$\left|\frac{\partial^2 \ln(f(x|p))}{\partial \lambda^2}\right| = \frac{Np^2 - 2px + x}{(1-p)^2 p^2}. \tag{C.14}$$

For a binomial distribution, $\mathbb{E}(x) = Np$, so this gives $\pi(p) \propto 1/\sqrt{p(1-p)}$.

C.3 Model Selection

Given multiple valid models, it is not trivial to select which model represents the data best. While a fit with more parameters will tend to reduce the total residuals of the fit and the data, it is prone to overfitting and lose predictive power. The interplay between these two is characterized by *information criteria*. There are various information criteria used, but here we will consider the Bayes factor, the Bayesian information criterion (BIC), and cross-validation.

In this section, $\mathcal{D}$ will represent the data with $\mathcal{N}$ points that is being fit, and $\mathcal{M}_i$ will represent models used to fit the data. As in Sec. 7.3, $\mathcal{I}$ represents all background information.

A Bayesian approach

The Bayesian approach is to directly compute the probability of the models given the data, that is $p(\mathcal{M}|\mathcal{D},\mathcal{I})$. This is intractable as $p(\mathcal{D}|\mathcal{I})$ in Bayes' theorem can be hard to compute

in this case. However, by taking a ratio of posteriors, the term cancels, and we see how one model fares *relative* to another model. Calling the log of the ratio of the posteriors F, we find, using Bayes' theorem

$$F = \ln\left(\frac{p(\mathcal{M}_1|\mathcal{D},\mathcal{I})}{p(\mathcal{M}_2|\mathcal{D},\mathcal{I})}\right) = \ln\left(\frac{p(\mathcal{D}|\mathcal{M}_1,\mathcal{I})}{p(\mathcal{D}|\mathcal{M}_2,\mathcal{I})}\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}\right). \tag{C.15}$$

If the prior probabilities of the models are taken to be the same, this reduces to the log of the ratio of the likelihoods given the models. This is the log of the Bayes factor. This provides a systematic way to weight models over other models from *a priori* information.

For simplicity, let the model, $\mathcal{M}_i$, be a functional form that depends on a set of n fit parameters, $\vec{\mu}$. We are only able to determine $p(\mathcal{D}|\mathcal{M}_i,\mathcal{I})$ through $\vec{\mu}$. This is just the evidence integral (the denominator in Bayes' theorem), and thus

$$p(\mathcal{D}|\mathcal{M}_i,\mathcal{I}) = \int d^n\vec{\mu}\ \ p(\mathcal{D}|\vec{\mu},\mathcal{M}_i,\mathcal{I})\pi(\vec{\mu}|\mathcal{M}_i,\mathcal{I}). \tag{C.16}$$

$p(\mathcal{D}|\vec{\mu},\mathcal{M}_i,\mathcal{I})$ is easily computable once $\mathcal{M}_i$ and $\vec{\mu}$ are given. Uninformative priors can be chosen for $\pi(\vec{\mu}|\mathcal{M}_i,\mathcal{I})$ (see previous section), or this can actually incorporate prior knowledge of the parameters. Thus, the problem of computing F in Eq. C.15 has been reduced to computing integrals. There are techniques like Markov-chain Monte Carlo that have been developed for this purpose [63].

If the fit is well-constrained, further simplifications can be made. As the fit specification improves, the likelihood $p(\mathcal{D}|\vec{\mu},\mathcal{M}_i,\mathcal{I})$ looks like a δ function and make the integral trivial. The *Laplace approximation* quantifies this intuition. As in Sec. 7.3, the best-fit parameters can be found by maximizing

$$l(\vec{\mu}) = \ln(p(\mathcal{D}|\vec{\mu},\mathcal{M}_i,\mathcal{I})) \tag{C.17}$$

with respect to its argument. Call the maximum likelihood value of the fit parameters $\vec{\mu}^*$. We can expand around this maximum to obtain

$$p(\mathcal{D}|\vec{\mu},\mathcal{M}_i,\mathcal{I}) \approx e^{l(\vec{\mu}^*)}e^{\frac{1}{2}(\vec{\mu}-\vec{\mu}^*)_\rho(\vec{\mu}-\vec{\mu}^*)_\sigma\partial_\rho\partial_\sigma l(\vec{\mu}^*)}, \tag{C.18}$$

where there is obviously no term linear in $\vec{\mu}$ as $\vec{\mu}^*$ is an extremal point. If the fit is well-specified, the curvature, $\partial_\rho\partial_\sigma l(\vec{\mu}^*)$, should have large entries in it. Thus, this makes the problem suitable for the Laplace approximation. Assume that all parameters of $\vec{\mu}$ are defined on $(-\infty,\infty)$. If this is not the case, we can usually extend the parameter to take on all values of this range without affecting the result. This is true when the parameter is constrained enough such that without imposing a strict cutoff for the parameter values, the probability of the parameter taking on a value outside its range of validity is extremely small. We expand the prior around the maximum likelihood and only keep the 0th order term to get

$$\begin{aligned} p(\mathcal{D}|\vec{\mu},\mathcal{M}_i,\mathcal{I}) &\approx \int d^n\vec{\mu}\ \ e^{l(\vec{\mu}^*)}e^{\frac{1}{2}(\vec{\mu}-\vec{\mu}^*)_\rho(\vec{\mu}-\vec{\mu}^*)_\sigma\partial_\rho\partial_\sigma l(\vec{\mu}^*)}\pi(\vec{\mu}|\mathcal{M}_i,\mathcal{I}) \\ &\approx \sqrt{\frac{(2\pi)^n}{-\det\partial_\rho\partial_\sigma l(\vec{\mu}^*)}}e^{l(\vec{\mu}^*)}\pi(\vec{\mu}^*|\mathcal{M}_i,\mathcal{I}), \end{aligned} \tag{C.19}$$

provided the prior grows slower than the Gaussian term. This provides a tractable way to approximate F without the need to compute any multi-dimensional integrals. Note that as we need to evaluate the prior at a certain point, we do have to have regular priors for Eq. C.19 to make sense. This is no problem if the two fit models have parameters with the same prior, as then the factor cancels out. However, if the two fit models have parameters with different (non-normalizable) priors, then cutoffs need to be chosen for those priors and the selection criterion can unfortunately be affected by this cutoff.

Corrections to Eq. C.19 can come from either approximation. First, In expanding the likelihood, one should check that the Eq. C.17, is a good approximation of the likelihood and that the third order term is irrelevant. Second, it should be checked that higher order terms in the expansion of the prior in Eq. C.19 are also not relevant. Note that linear terms in the prior are irrelevant (they integrate to zero), so the relevant term to check is the second order term in the expansion of the prior.

As the number of observation points becomes large, the result simplifies even further. Consider the determinant of the curvature of the log-likelihood in Eq. C.19. Note that $l(\vec{\mu}^*)$ is just a sum of the log of the likelihood function evaluated at each data point. Assuming that these data points are uncorrelated, this is just a sum of uncorrelated variables. Thus, the law of large numbers says that in the $\mathcal{N}\to\infty$ limit, the sum should converge to $\mathcal{N}$ times the expectation of the log-likelihood for one point over the data. Thus, in this limit

$$\ln(p(\mathcal{D}|\vec{\mu},\mathcal{M}_i,\mathcal{I})) \overset{\mathcal{N}\to\infty}{=} l(\vec{\mu}^*)+\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\mathcal{N})+\frac{1}{2}\ln(-\det\partial_\rho\partial_\sigma l_1(\vec{\mu}^*))+\ln(\pi(\vec{\mu}^*|\mathcal{M}_i,\mathcal{I})). \tag{C.20}$$

Where $l_1(\vec{\mu})$ is the expectation of the log-likelihood of one point over the distribution. When $\mathcal{N}\to\infty$, $l(\vec{\mu}^*)$ scales as $\mathcal{N}$ as it is a sum of $\mathcal{N}$ uncorrelated variables. The only other term that grows with $\mathcal{N}$ is the $\ln(\mathcal{N})$, so it is kept. Then, multiplying by -2 gives

$$\text{BIC} = -2l(\vec{\mu}^*) + n\ln(\mathcal{N}). \tag{C.21}$$

This is a simple expression that can be evaluated easily and is relatively robust [109] . Dependence on the prior has dropped out of this expression. While the Bayes factor certainly has more information, for a quick comparison of models, the BIC is good for these reasons.

For our model selection in Sec. 7.3, $\mathcal{N}$ was not large, as the $n\ln(\mathcal{N})$ is about the same as the $n\ln(2\pi)$ term that is thrown out in the asymptotic expansion. However, we still found that it was useful in model selection.

Models without a functional form

Though not discussed in the thesis, one may be interested in models to describe the data that do not have a functional form. This includes, for example, decision-tree and clustering machine learning methods. For these models, the AIC and BIC are not applicable as the number of fit parameters that goes into calculating AIC and BIC are not well-defined. Thus, it is extremely common to use *cross-validation* as a means of model selection [132]. The idea of cross-validation is to split the model into training and validation sets, where the validation sets are disjoint. Fits are performed on the training set and the performance is directly measured on the validation set. Since no elements of the validation set are used in the fit, this method automatically prevents overfitting.

The number of sets to choose depends on the problem. For small problems like the ones discussed in this thesis, each validation set could contain one element. This is called *hold one out validation.* While computationally expensive, hold one out validation best emulates the performance of the model if it were fit on all of the data, as the training sets look very similar to the whole dataset. For large datasets, it is more common to choose n disjoint sets, so that the same statistic can be approximated but with less computational cost.

Cross-validation is also simple to implement in that no extra work beyond fitting is necessary. This contrasts, for example, computing a Bayes factor which can involve numerical integrals or a bit of analysis to get an amenable expression.

C.4 Confidence Intervals

Frequentist

Frequentist methods construct a $1-\alpha$ confidence intervals such that, if the experiment were repeated many times, the interval would contain the true value $1-\alpha$ of the time. Ideally, this would be exactly $1-\alpha$ of the time, but overcoverage (when the true value is contained greater than $1-\alpha$ of the time) is usually inevitable, especially when the likelihood distribution is discrete.

Confidence intervals are constructed using the Neyman construction [6, 7, 116]. For simplicity, we will describe the construction with one measured parameter x, and one quantity to create an interval for, θ. We will discuss the multi-dimensional case later. Let $f(x|\theta)$ be the likelihood function of x. If $x \in \mathbb{R}$ (or a subset of $\mathbb{R}$), for each value of θ, an interval $[a_\theta, b_\theta]$ is chosen such that

$$1-\alpha = \int_{a_\theta}^{b_\theta} dx \;\; f(x|\theta). \tag{C.22}$$

Or, if $x \in \mathbb{Z}$ (or a subset of $\mathbb{Z}$), for each value of θ, an interval $[a_\theta, b_\theta]$ is chosen such that

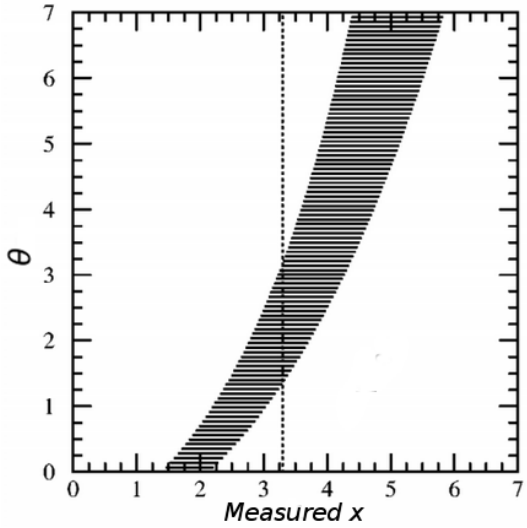

Figure C.1: Construction of confidence intervals. For each θ (the quantity to set a limit on), a horizontal interval of x (the measured quantity) is chosen, such that it satisfies Eq. C.22 or Eq. C.23. The union of these horizontal intervals forms the 90% confidence belt. When a measurement is made, a vertical line is drawn at the measurement (shown here as a dotted line). The intersection of this line and the confidence belt is corresponding confidence interval for θ that the measurement allows [6, 7]. Figure is modified from Ref. [6].

$$1 - \alpha \leq \sum_{a_\theta}^{b_\theta} f(x|\theta), \tag{C.23}$$

as equality cannot be guaranteed when discrete packets of f are added.

The collection of these lines for each θ form a *confidence belt.* When an experiment measures a value x, all values of θ such that $x \in [a_\theta, b_\theta]$ are accepted, and the collection of accepted θ becomes the $1 - \alpha$ confidence interval on θ. Fig. C.1 graphically depicts this process.

The choice of the upper and lower limits for each θ are arbitrary in the construction of the confidence belt as long as the method used to determine them is not influenced by the data itself, and Eq. C.22 or Eq. C.23 is satisfied for all values of θ within the physical bounds on θ. The choice may depend on whether one desires a lower bound or an upper bound, though there are methods such as the Feldman-Cousins method [6] that smoothly transition between one and two-sided confidence intervals.

Choosing and committing to a method before conducting an experiment is imperative [6]. Suppose an experimenter decides to quote an upper limit if he or she observes less than or equal to the expected number of background events and quotes a two-sided limit if he observes more than the expected number of background events. Then, even if in each case

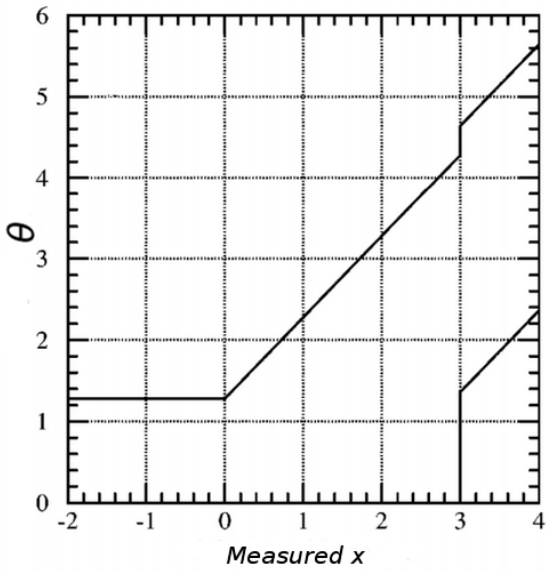

Figure C.2: The 90% confidence belts that a flip-flopping physicist, as described in the text, would produce for Gaussian distributed data with an expected background rate $\bar{b} = 3$. The kink in the plot results at $x = 3$ from the arbitrary decision made by the experimenter. For the signal means in the range $1.36 < \bar{s} < 4.28$, the coverage, or the probability contained in the horizontal belts, is 85%, which is undercovering [6]. Figure is modified from Ref. [6].

the flip-flopping physicist plans to use a method that will return $1 - \alpha$ confidence intervals, there will be regions where the coverage is less than $1 - \alpha$, as shown in Fig. C.4. Then, the intervals the experimenter publishes are not confident to $1 - \alpha$. In terms of the definitions given earlier, if the experiment were repeated many times, there are some values of the true signal value θ such that the limits set on θ will contain the true value less than $1 - \alpha$ of the time, and thus is not a valid method to construct confidence intervals.

In multiple dimensions, the construction is identical. The integration (or summation) regions in the multi-dimensional analog of Eq. C.22 or Eq. C.23 can be any multi-dimensional region such that the condition is satisfied for a given $\vec{\theta}$, as long as the choice is independent of the measurement. The $\vec{\theta}$ is accepted if the measured $\mathbf{x}$ is contained within that multi-dimensional region.

Bayesian

In a Bayesian formalism, building a confidence interval[1] is straightforward. To construct a confidence interval at $1 - \alpha$ confidence on a quantity θ based on an observation $\mathcal{D}$, one computes the posterior for θ. The interval $[\theta_1, \theta_2]$ must satisfy [7]

[1]Sometimes a Bayesian confidence interval is called a *credible interval* to distinguish it from a frequentist confidence interval.

$$1 - \alpha = \int_{\theta_1}^{\theta_2} d\theta \;\; p(\theta|\mathcal{D}). \tag{C.24}$$

As with the frequentist case, any choice of θ_1 and θ_2 that satisfies this is a valid confidence interval. If one would like to place an upper bound on θ (assuming $\theta > 0$), one could set $\theta_1 = 0$ and solve for θ_2. One way to construct a two-sided interval would be to find the value of θ where $p(\theta|\mathcal{D})$ is maximal, and then take a symmetric interval about that point that satisfies Eq. C.24.

$p(\theta|\mathcal{D})$ can be determined from the likelihood using Bayes theorem. This gives

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)\pi(\theta)}{p(\mathcal{D})}. \tag{C.25}$$

$p(\mathcal{D})$ can be determined by requiring that the integral of $p(\theta|\mathcal{D})$ over all physical θ must be 1. $p(\mathcal{D}|\theta)$ is the likelihood function of the data. $\pi(\theta)$ can be chosen to be uninformative (see Sec. C.2), or can actually take into account prior knowledge of θ.

Note that in the Bayesian formalism, no choices (except for the prior) are made to construct the posterior, $p(\theta|\mathcal{D})$. This means that the posterior distribution is insensitive to whether the method to construct it was decided before or after a measurement was made, unlike frequentist methods. Any interval satisfying Eq. C.24 is a valid confidence interval at $1 - \alpha$ confidence, so the confidence interval is also insensitive to whether the construction method was decided before or after a measurement was made.

If θ is n-dimensional quantity (a joint bound on more than one parameter is desired), the equation is the same, but the integration region can be any curve in n-dimensional space that satisfies the above constraint.

Appendix D

Codes

In this appendix we provide the code used in this thesis.

D.1 Levenberg-Marquardt Algorithm

The following is CUDA [51] code and is an implementation of the Levenberg-Marquardt algorithm for a GPU. This script can be exported as a dll to perform the modified Gaussian fits of Sec. 4.3.

```
 // Win32Project1.cpp : Defines the exported functions for the DLL application.
//

#include "stdafx.h"

#include <stdlib.h> // For: exit, drand48, malloc, free, NULL, EXIT_FAILURE
#include <stdio.h> // For: perror

#include <math.h> // For: fabs

#include <iostream>
#include <fstream>

#include "ConsoleApplication1.h"

//#include <sys/time.h>

#include <cuda.h>

#include <cublas_v2.h>
```

```
#include "cuda_runtime.h"
#include "device_launch_parameters.h"

#include <windows.h>

#define NUM_THREADS 256
#define eps2 1e-5
#define MAXITER 500

/*float read_timer()
{
  static bool initialized = false;
  static struct timeval start;
  struct timeval end;
  if (!initialized)
  {
    gettimeofday(&start, NULL);
    initialized = true;
  }
  gettimeofday(&end, NULL);
  return (end.tv_sec - start.tv_sec) + 1.0e-6 * (end.tv_usec - start.tv_usec);
}*/

//Use this method when the vector is small enough that it's not worth using the
    GPU to take the infinity norm
float infnorm(const float *A, const int len){
  /*Inputs: Small vector of floats A on host and the length (len) of it*/

  float toReturn = 0.;
  for (int i = 0; i<len; i++){
    float tmp = abs(A[i]);
    if (tmp > toReturn) toReturn = tmp;
  }

  return toReturn;
}

//Use this method when the vector is small enough that it's not worth using the
    GPU to take the 2 norm
float twonorm(const float *A, const int len){
  /*Inputs: Small vector of floats A on host and the length (len) of it*/

  float toReturn = 0.;
  for (int i = 0; i<len; i++){
```

```
    toReturn += A[i] * A[i];
  }

  toReturn = sqrtf(toReturn);
  return toReturn;
}

/*Method will use cuBLAS to calculate A^T A*/
void gpu_blas_mmul(cublasHandle_t &handle, const float *A, float *C, const int m,
     const int k) {
  /*Input A is a device columnated mxk matrix
  output C will be a device columnated kxk matrix, with only the lower diagonal
      filled*/
  int lda = m, ldc = m;
  const float alf = 1.;
  const float bet = 0.;
  const float *alpha = &alf;
  const float *beta = &bet;

  // Do the actual multiplication
  cublasSsyrk(handle,
    CUBLAS_FILL_MODE_LOWER, CUBLAS_OP_N,
    m, k,
    alpha,
    A, lda,
    beta,
    C, ldc);

}

/*Method will use cuBLAS to calculate g = Ax+bet g*/
void gpu_blas_applymat(cublasHandle_t &handle, const float *A, const float *x,
    float *g, const int m, const int n, const float bet = 0., bool trans = true){
  /*Input A is a device columnated mxn matrix, input x is a device nx1 vector
  output g will be a device columnated kx1
  vector
  trans says whether A should be transposed or not before being applied to x*/

  int lda = m;
  int incx = 1;
  int incy = 1;
  const float alf = 1.;
  const float *alpha = &alf;
  const float *beta = &bet;
```

```
  // Do the actual application, with transpose turned on or not
  if (trans) cublasSgemv(handle, CUBLAS_OP_T, m, n, alpha, A, lda, x, incx, beta,
       g, incy);
  else cublasSgemv(handle, CUBLAS_OP_N, m, n, alpha, A, lda, x, incx, beta, g,
      incy);

}

/*Prints entries of A that is mxn.*/
void print_matrix(const float *A, int m, int n) {

  for (int i = 0; i < m; ++i){
    for (int j = 0; j < n; ++j){
      std::cout << A[j * m + i] << " ";
    }
    std::cout << std::endl;
  }
  std::cout << std::endl;
}

void FindCenterMoments(const float*Image, float *center, const int width, const
    int height){
  /*This routine finds the center of the distribution as the weighted average of
      all the pixels,
  ,.with the weight being the intensity of the image at the pixel
  The input is a linearized widthxheight vector Image on the host.
  Each entry of center will be updated with the found value of the center (x and
      y coordinates)*/


  float avgx = 0.;
  float avgy = 0.;

  float totweight = 0.;

  for (int i = 0; i<width; i++){
    for (int j = 0; j<height; j++){
      float toAdd = Image[i + j*height];
      avgx += (float)i*toAdd;
      avgy += (float)j*toAdd;

      totweight += toAdd;
    }
```

```
  }

  *center = avgx / totweight;
  *(center+1) = avgy / totweight;

  /*printf("Center: (%f,%f)\n", center[0], center[1]);*/
}

float FindCenterSlopes(const float *Image, float *center, const int width, const
    int height){
  /*This routine finds the center of the distribution as the weighted average of
     all the pixels, with the weight being the intensity of the image at the
     pixel
  The input is a linearized widthxheight vector Image and a 2x1 vector center on
     the host.
  Each entry of center will be updated with the found value of the center (x and
     y coordinates)*/


  float *w = (float*)malloc(sizeof(float)*width*height);
  float *m = (float*)malloc(sizeof(float)*width*height);

  center[0] = 0.;
  center[1] = 0.;
  float dist = 10.;
  int k = 0;

  /*Finding the minimum value of the distribution is useful, so we will do it
     here*/
  float minval = 100000.;
  for (int i = 0; i<width*height; i++){
    if (Image[i]<minval && Image[i] != 0) minval = Image[i];
  }

  printf("%f\n", minval);

  /*Gradient are calculated only once, but this information is reused*/
   for(int j=0;j<height;j++){
      for(int i=0;i<width;i++){
      float unorm = 0.;
      float vnorm = 0.;
      if(Image[i+j*height]!=0 && Image[(i+1)+(j+1)*height]!=0 && Image[i+(j+1)*
          height]!=0 && Image[(i+1)+j*height]!=0){
```

```
          unorm = Image[(i+1)+(j+1)*height]-Image[i+j*height]; //Intensity
              gradient along x+y
          vnorm = Image[i+(j+1)*height]-Image[(i+1)+j*height]; //Intensity
              gradient along x-y
    }
        w[i+j*height] = unorm*unorm+vnorm*vnorm; //Initially, the weights are
            the norm of the gradient vector.
        if(unorm!=vnorm && i!=width-1 && j!=height-1) m[i+j*height] = (unorm+
            vnorm)/(unorm-vnorm); //Prevent divide by zero.
        else{
          m[i+j*height] = 0.;
          w[i+j*height] = 0.;
          }
       }
    }


/*Termination condition is when the best guess for the center changes by less
    than a pixel after applying this algorithm.
We place a hard cutoff at 100 steps
                                                                          */
while(dist>1. && k<100){
    k++;

    //Find weights
    if(k>1){
       for(int j=0;j<height-1;j++){
          for(int i=0;i<width-1;i++){
      if(Image[i+j*height]!=0 && Image[(i+1)+(j+1)*height]!=0 && Image[i+(j+1)*
          height]!=0 && Image[(i+1)+j*height]!=0){
                float unorm = Image[(i+1)+(j+1)*height]-Image[i+j*height]; //
                    Intensity gradient along x+y
              float vnorm = Image[i+(j+1)*height]-Image[(i+1)+j*height]; //
                  Intensity gradient along x-y
           float d = (((float)i-center[0])*((float)i-center[0])+((float)j-center
              [1])*((float)j-center[1])); //Distance from current pixel to the
              current best guess for the center
          w[i+j*height] = (unorm*unorm+vnorm*vnorm)/d; //Weight is the norm of
              gradient over distance.
      }
      else{
        w[i+j*height] = 0.;
      }
            }
```

```
          }
        }

     //Use the analytic expressions to calculate what the weighted center should
        be:
     float sum1 = 0.;
     float sum2 = 0.;
     float sum3 = 0.;
     float sum4 = 0.;
     float sum5 = 0.;

     for(int i=0;i<height*width;i++){
        float xk = (float)(i%height);
        float yk = (float)((i-xk)/height);
        float denom = (m[i]*m[i]+1.);
        sum1 += m[i]*w[i]*(yk-m[i]*xk)/denom;
        sum2 += w[i]/denom;
        sum3 += m[i]*w[i]/denom;
        sum4 += (yk-m[i]*xk)*w[i]/denom;
        sum5 += m[i]*m[i]*w[i]/denom;
     }

     float Delta = sum3*sum3-sum5*sum2;
     float newx = (sum1*sum2-sum3*sum4)/Delta;
     float newy = (sum1*sum3-sum5*sum4)/Delta;

     //This is the distance squared from the old center to the new one
     dist = (newx-center[0])*(newx-center[0])+(newy-center[1])*(newy-center[1]);

     //Update the center
     center[0] = newx;
     center[1] = newy;
   }

   //printf("Center: (%f,%f) in %d steps\n",center[0],center[1],k);

   free(m);
   free(w);

   return minval;
}

/*Use a Cholesky Decomposition to solve Ax=g*/
```

```
void CholeskySolve(cublasHandle_t &handle, const float *A, float *g, const int
    numparams){
  /*Input is A, a numparamsxnumparams host positive definite symmetric matrix
     filled on the lower
  diagonal
  g is device a numparamsx1 vector*/

  float *L = (float*)malloc(sizeof(float)*numparams*numparams);

  /*Use analytic expressions for the Cholesky decomposition*/
  if (numparams == 4){
    L[0] = sqrtf(A[0]);
    L[1] = A[1] / L[0];
    L[2] = A[2] / L[0];
    L[3] = A[3] / L[0];
    L[5] = sqrtf(A[5] - L[1] * L[1]);
    L[6] = (A[6] - L[1] * L[2]) / L[5];
    L[7] = (A[7] - L[1] * L[3]) / L[5];
    L[10] = sqrtf(A[10] - L[2] * L[2] - L[6] * L[6]);
    L[11] = (A[11] - L[2] * L[3] - L[6] * L[7]) / L[10];
    L[15] = sqrtf(A[15] - L[11] * L[11] - L[3] * L[3] - L[7] * L[7]);
  }

  if (numparams == 6){
    L[0] = sqrtf(A[0]);
    L[1] = A[1] / L[0];
    L[2] = A[2] / L[0];
    L[3] = A[3] / L[0];
    L[4] = A[4] / L[0];
    L[5] = A[5] / L[0];

    L[7] = sqrtf(A[7] - L[1] * L[1]);
    L[8] = (A[8] - L[1] * L[2]) / L[7];
    L[9] = (A[9] - L[1] * L[3]) / L[7];
    L[10] = (A[10] - L[1] * L[4]) / L[7];
    L[11] = (A[11] - L[1] * L[5]) / L[7];

    L[14] = sqrtf(A[14] - L[2] * L[2] - L[8] * L[8]);
    L[15] = (A[15] - L[2] * L[3] - L[8] * L[9]) / L[14];
    L[16] = (A[16] - L[2] * L[4] - L[8] * L[10]) / L[14];
    L[17] = (A[17] - L[2] * L[5] - L[8] * L[11]) / L[14];

    L[21] = sqrtf(A[21] - L[3] * L[3] - L[9] * L[9] - L[15] * L[15]);
    L[22] = (A[22] - L[15] * L[16] - L[3] * L[4] - L[10] * L[9]) / L[21];
```

```
    L[23] = (A[23] - L[15] * L[17] - L[3] * L[5] - L[11] * L[9]) / L[21];

    L[28] = sqrtf(A[28] - L[4] * L[4] - L[10] * L[10] - L[16] * L[16] - L[22] * L
        [22]);
    L[29] = (A[29] - L[4] * L[5] - L[10] * L[11] - L[16] * L[17] - L[22] * L[23])
        / L[28];

    L[35] = sqrtf(A[35] - L[5] * L[5] - L[11] * L[11] - L[17] * L[17] - L[23] * L
        [23] - L[29] * L[29]);

  }

  //Copy the Cholesky factors to GPU to be able to use cuBLAS
  float *d_L;
  cudaMalloc((void **)&d_L, sizeof(float)*numparams*numparams);
  cudaMemcpy(d_L, L, sizeof(float)*numparams*numparams, cudaMemcpyHostToDevice);

  //Use cuBLAS twice to solve the triangular system L^TLx = g
  const float alf = 1.;
  const float *alpha = &alf;

  cublasStrsm(handle,
    CUBLAS_SIDE_LEFT, CUBLAS_FILL_MODE_LOWER,
    CUBLAS_OP_N, CUBLAS_DIAG_NON_UNIT,
    numparams, 1,
    alpha,
    d_L, numparams,
    g, numparams);

  cublasStrsm(handle,
    CUBLAS_SIDE_LEFT, CUBLAS_FILL_MODE_LOWER,
    CUBLAS_OP_T, CUBLAS_DIAG_NON_UNIT,
    numparams, 1,
    alpha,
    d_L, numparams,
    g, numparams);

  free(L);
  cudaFree(d_L);
}

/*Computes the Jacobian matrix on the GPU*/
__global__ void make_jacobian(const float *values, const int *mask, const float *
    beta, const float* center, float *jacob, float *resid, const int height, const
```

```
   int width, const int numparams){
/*Input is values, a heightxwidth device matrix of image intensity values
beta, a numparamsx1 device vector
center, a 2x1 device vector (information in center not used if numparams==6)
Jacobian is an empty device matrix of size (height*width)xnumparams that will
    have the jacobian
resid is a heightxwidth device matrix that will store residuals at each point*/

int tid = threadIdx.x + blockIdx.x * blockDim.x;
if (tid >= height*width) return;
if (values[tid] == 0) return;

//Each thread looks at one point of the image
float x0 = (float)(tid%height);
float y0 = (float)((tid - x0) / height);

float dist, norm, expon, exponentiated;

//Calculate each component of the Jacobian. The analytic expressions for the
    derivative are used
if (numparams == 4){
  dist = (center[0] - x0)*(center[0] - x0) + (center[1] - y0)*(center[1] - y0);
  norm = sqrtf(dist) / beta[1];
  expon = -powf(norm, beta[2]);
  exponentiated = expf(expon);
}
if (numparams == 6){
  dist = (beta[4] - x0)*(beta[4] - x0) + (beta[5] - y0)*(beta[5] - y0);
  norm = sqrtf(dist) / beta[1];
  expon = -powf(norm, beta[2]);
  exponentiated = expf(expon);
}

//resid is filled with the value at the point minus the fit value.
resid[mask[tid]] = values[tid] - (beta[0] * exponentiated + beta[3]);

jacob[mask[tid]*numparams] = exponentiated;
jacob[mask[tid]*numparams + 1] = beta[0] * exponentiated*beta[2] * (-expon) /
    beta[1];
jacob[mask[tid]*numparams + 2] = beta[0] * exponentiated*expon*logf(norm);
jacob[mask[tid]*numparams + 3] = 1.;

if (numparams == 6){
```

```
    jacob[mask[tid]*numparams + 4] = beta[0] * exponentiated*(x0 - beta[4])*beta
        [2] * (-expon) / dist;
    jacob[mask[tid]*numparams + 5] = beta[0] * exponentiated*(y0 - beta[5])*beta
        [2] * (-expon) / dist;
  }
}

__declspec(dllexport) int __cdecl ComputeFit(const float* values, const int width
    , const int height, float *beta)
{
  //This code should be compiled with a physical connection to the GPU computer.
      If compiled over a remote connection, it will fail.

  /*Define the size of the image and how many parameters we would like to fit*/
  const unsigned int numparams = 6;

  /*Initialize variables*/
  //float *values = (float*) malloc(sizeof(float)*width*height);
  int *mask = (int*)malloc(sizeof(int)*width*height);

  float *d_resid, *d_jacob, *d_values, *d_newresid, *d_newjacob;

  int  *d_mask;

  cudaMalloc((void **)&d_values, sizeof(float)*width*height);
  cudaMalloc((void **)&d_mask, sizeof(int)*width*height);

  float center[2] = { 0., 0. };
  int totpoints = 0;

  for(int i=0;i<width*height;i++){
    if(values[i]==0) mask[i] = 0;
    else{
      mask[i] = totpoints;
      totpoints++;
    }
  }

  float *resid = (float*)malloc(sizeof(float)*totpoints);
  float *jacob = (float*)malloc(sizeof(float)*numparams*totpoints);

  cudaMalloc((void **)&d_resid, sizeof(float)*totpoints);
  cudaMalloc((void **)&d_jacob, sizeof(float)*numparams*totpoints);
  cudaMalloc((void **)&d_newresid, sizeof(float)*totpoints);
```

```
cudaMalloc((void **)&d_newjacob, sizeof(float)*numparams*totpoints);

//float start = read_timer();
//printf("I/O time: %f\n", start - begin); //I/O time is not included in
    performance analyses

//FindCenterMoments(values,center,width,height);
//float minval = 1.;
float minval = FindCenterSlopes(values, center, width, height); //Find the
    center of the distribution

//First three initial values are hard-coded, the chosen values don't matter too
     much
//float beta[6] = {10.,50.,2.0,minval,center[0],center[1]};

beta[0] = 10.;
beta[1] = 500.;
beta[2] = 2.0;
beta[3] = minval;
beta[4] = center[0];
beta[5] = center[1];

float* d_center, *d_beta;

cudaMalloc((void **)&d_center, sizeof(float) * 2);
cudaMalloc((void **)&d_beta, sizeof(float)*numparams);

//float end = read_timer();
//printf("Center Time: %f\n", end - start);

//Copy the vectors to the device
cudaMemcpy(d_values, values, sizeof(float)*width*height, cudaMemcpyHostToDevice
    );
cudaMemcpy(d_center, center, sizeof(float) * 2, cudaMemcpyHostToDevice);
cudaMemcpy(d_beta, beta, sizeof(float)*numparams, cudaMemcpyHostToDevice);
cudaMemcpy(d_mask, mask, sizeof(int)*width*height, cudaMemcpyHostToDevice);

int blks = (height*width + NUM_THREADS - 1) / NUM_THREADS;

cudaThreadSynchronize();

//Set up the problem by calculating the Jacobian, J^T J, and J^Tf
make_jacobian <<< blks, NUM_THREADS >>> (d_values, d_mask, d_beta, d_center,
    d_jacob, d_resid, height, width, numparams);
```

```
float *A = (float*)malloc(sizeof(float)*numparams*numparams);
float *g = (float*)malloc(sizeof(float)*numparams);

float* d_A, *d_g;

cudaMalloc((void **)&d_A, sizeof(float)*numparams*numparams);
cudaMalloc((void **)&d_g, sizeof(float)*numparams);

cublasHandle_t handle;
cublasStatus_t stat;

stat = cublasCreate(&handle);

if (stat != CUBLAS_STATUS_SUCCESS) {
      printf ("CUBLAS initialization failed\n");
      if(stat == CUBLAS_STATUS_NOT_INITIALIZED) return -2;
  return -1;
  }

gpu_blas_mmul(handle, d_jacob, d_A, numparams, totpoints);

gpu_blas_applymat(handle, d_jacob, d_resid, d_g, numparams, totpoints, 0.,
    false);
/*Define the size of the image and how many parameters we would like to fit*/
bool found = 0;
int iter = 0;

float mu = 1.;
float nu = 2.;

float diffnorm, xnorm, Fold, Fnew, rho;
float* h = (float*)malloc(sizeof(float)*numparams);

//float newx[4] = {0.,0.,0.,0.};
float newx[6] = { 0., 0., 0., 0., 0., 0. };

float *d_h;
float *d_newx;
cudaMalloc((void **)&d_newx, sizeof(float)*numparams);
cudaMalloc((void **)&d_h, sizeof(float)*numparams);

float *d_Acopy;
cudaMalloc((void **)&d_Acopy, sizeof(float)*numparams*numparams);
```

```
//Store the current value of the residual to see if gets better
cublasSnrm2(handle, totpoints, d_resid, 1, &Fold);

while (!found && (iter<MAXITER)){
  iter += 1;

  //Copy d_g and A since we don't want to modify it unless we take a step
  cublasScopy(handle, numparams, d_g, 1, d_h, 1);
  cudaMemcpy(A, d_A, sizeof(float)*numparams*numparams, cudaMemcpyDeviceToHost);

  //Apply dampening with mu
  for (int i = 0; i<numparams; i++){
    A[i + numparams*i] += mu*A[i + numparams*i];
  }

  //Solve for the step size, d_h
  CholeskySolve(handle, A, d_h, numparams);
  cudaMemcpy(h, d_h, sizeof(float)*numparams, cudaMemcpyDeviceToHost);

  //If step is small, we will terminate
  for (int i = 0; i<numparams; i++){
    newx[i] -= beta[i];
  }

  diffnorm = twonorm(h, numparams);
  xnorm = twonorm(beta, numparams);

  if ((diffnorm<eps2*(xnorm + eps2))) found = 1;
  else{
    int check = 0; //Will check if a parameter that shouldn't go negative goes
       negative
    for (int i = 0; i<numparams; i++){
      newx[i] = beta[i] + h[i];
      if (newx[i]<0 && i != 0 && i != 3) check++;
    }

    if (check == 0){

      /*Compute the new Jacobian and residual when the step is taken, see if it
         reduces the
      total residual*/

      cudaMemcpy(d_h, h, sizeof(float)*numparams, cudaMemcpyHostToDevice);
```

```
  cublasScopy(handle, totpoints, d_resid, 1, d_newresid, 1);
  gpu_blas_applymat(handle, d_jacob, d_h, d_newresid, numparams, totpoints,
     1., true);

  cudaMemcpy(d_newx, newx, sizeof(float)*numparams, cudaMemcpyHostToDevice);

  //int blks = (height*width + NUM_THREADS - 1) / NUM_THREADS;

  make_jacobian <<< blks, NUM_THREADS >>> (d_values, d_mask, d_newx,
     d_center, d_newjacob, d_newresid, height, width, numparams);

  cublasSnrm2(handle, totpoints, d_newresid, 1, &Fnew);

  rho = Fold - Fnew;
}
else rho = 0;

//If rho>0, the residual is decreasing, so save take this step and save the
    results
if (rho>0){
  for (int i = 0; i<numparams; i++) beta[i] = newx[i]; //Take the step

  Fold = Fnew; //Update total residual value

  cublasScopy(handle, totpoints*numparams, d_newjacob, 1, d_jacob, 1); //
     Update Jacobian
  cublasScopy(handle, totpoints, d_newresid, 1, d_resid, 1); //Update
     residual matrix

  gpu_blas_mmul(handle, d_jacob, d_A, numparams, totpoints); //Update A

  gpu_blas_applymat(handle, d_jacob, d_resid, d_g, numparams, totpoints, 0.,
      false); //Update g

  mu /= 3.; //Make mu smaller to see if we can take a big step
  nu = 2;
}
else{
  //This was a bad step

  mu *= nu; //Make mu larger to take a smaller step next time
  nu *= 2; //We'll take an even smaller step if it fails again.
```

```
      }
    }
  }

  //Display results
  printf("Number Steps: %d\n", iter);
  print_matrix(beta, numparams, 1);

  //Free up memory free(resid);
  free(jacob);
  free(g);
  free(h);
  free(mask);

  cudaFree(d_resid);
  cudaFree(d_newresid);
  cudaFree(d_values);
  cudaFree(d_jacob);
  cudaFree(d_newjacob);
  cudaFree(d_center);
  cudaFree(d_beta);
  cudaFree(d_A);
  cudaFree(d_g);
  cudaFree(d_newx);
  cudaFree(d_h);
  cudaFree(d_mask);

  cublasDestroy(handle);

    return iter;
}

int main(){

return 0;
}
```

D.2 Discrete Temperature Fitter

The following is *Mathematica* [70] code to perform fits for temperatures of binned data, as described in Sec. 5.5.

```
q = Import["C:\\Users\\lenny\\Downloads\\fwdrecode\\PMTCSV_42162_Dump-Cold_\
```

```
SIS_PMT_CATCH_OR.csv"];
q = q[[2 ;;]];
r = Import["C:\\Users\\lenny\\Downloads\\fwdrecode\\ColdDumpE4E5.dump",
   "Table"];
VT = Interpolation[r];

start = 17;
minbins = 5;
maxbins = 30;
temps = {};
pvalues = {};
For[cutoff = start + minbins, cutoff <= start + maxbins, cutoff++,
 tot = Sum[q[[i, 3]], {i, start, cutoff}];
 Vbar = 1/tot Sum[q[[i, 3]] VT[q[[i + 1, 2]]], {i, start, cutoff}];
 l = (tot - 1) Log[x[1]] - Log[x[2]] -
   x[1] (Exp[-(VT[q[[cutoff + 1, 2]]]/x[2])] -
      Exp[-(VT[q[[start, 2]]]/x[2])]) +
   Sum[q[[i, 3]]*
     Log[(Exp[- VT[q[[i + 1, 2]]]/x[2]] -
        Exp[- VT[q[[i, 2]]]/x[2]])], {i, start, cutoff}];
 m = FindMaximum[{l}, {{x[2], Vbar - VT[q[[cutoff + 1, 2]]]}, {x[1],
     tot*Exp[VT[q[[cutoff + 1, 2]]]/(
       Vbar - VT[q[[cutoff + 1, 2]]])]}}, MaxIterations -> 10000,
   Method -> "PrincipalAxis"];
 curv = -2*Table[D[D[l, x[i]], x[j]], {i, 1, 2}, {j, 1, 2}] /.
   m[[2]];
 temps = Append[
   temps, {cutoff, m[[1]], m[[2]], Vbar - VT[q[[cutoff + 1, 2]]],
    tot*Exp[VT[q[[cutoff + 1, 2]]]/(Vbar - VT[q[[cutoff + 1, 2]]])],
    curv}];
 pvalues =
  Append[pvalues,
   Table[(Gamma[q[[i, 3]] + 1,
       x[1] (Exp[- VT[q[[i + 1, 2]]]/x[2]] -
           Exp[- VT[q[[i, 2]]]/x[2]]) /. m[[2]]])/(Factorial[
       q[[i, 3]]]) , {i, 1, 60}]]]
```

Bibliography

[1] A. Povilus, *Cyclotron-cavity mode resonant cooling in single component electron plasmas*. PhD thesis, University of California, Berkeley, 2015.

[2] N. Evetts, "Cavity cooling of leptons for increased antihydrogen production at ALPHA," Master's thesis, University of British Columbia, 2015.

[3] C. So, *Antiproton and positron dynamics in antihydrogen production*. PhD thesis, University of California, Berkeley, 2014.

[4] ALPHA Collaboration, "Centrifugal separation and equilibration dynamics in an electron-antiproton plasma," *Physical Review Letters*, vol. 106, no. 14, p. 145001, 2011.

[5] A. P. Povilus, N. D. DeTal, L. T. Evans, N. Evetts, J. Fajans, W. N. Hardy, E. D. Hunter, I. Martens, F. Robicheaux, S. Shanman, C. So, X. Wang, and J. S. Wurtele, "Electron plasmas cooled by cyclotron-cavity resonance," *Physical Review Letters*, vol. 117, p. 175001, Oct 2016.

[6] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals," *Physical Review D*, vol. 57, pp. 3873–3889, 1998.

[7] K. Olive and Particle Data Group, "Review of particle physics," *Chinese Physics C*, vol. 38, no. 9, p. 090001, 2014.

[8] ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.

[9] CMS Collaboration, "Combined results of searches for the standard model higgs boson in pp collisions at $\sqrt{s} = 7$ TeV," *Physics Letters B*, vol. 710, no. 1, pp. 26 – 48, 2012.

[10] K. Hagiwara, A. Martin, D. Nomura, and T. Teubner, "Improved predictions for $g-2$ of the muon," *Physics Letters B*, vol. 649, no. 23, pp. 173 – 179, 2007.

[11] BABAR Collaboration, "Evidence for an excess of $\overline{B} \to D^{(*)}\tau^{-}\overline{\nu}_{\tau}$ decays," *Physical Review Letters*, vol. 109, p. 101802, 2012.

[12] M. Lindner, T. Ohlsson, and G. Seidl, "Seesaw mechanisms for Dirac and Majorana neutrino masses," *Physical Review D*, vol. 65, no. 5, p. 053014, 2002.

[13] C. P. Burgess and G. D. Moore, *The standard model: a primer*. Cambridge University Press, 2006.

[14] V. Mukhanov, *Physical foundations of cosmology*. Cambridge University Press, 2005.

[15] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*. Westview Press Reading, 1995.

[16] J. DiSciacca, M. Marshall, K. Marable, G. Gabrielse, S. Ettenauer, E. Tardiff, R. Kalra, D. W. Fitzakerley, M. C. George, E. A. Hessels, C. H. Storry, M. Weel, D. Grzonka, W. Oelert, and T. Sefzick, "One-particle measurement of the antiproton magnetic moment," *Physical Review Letters*, vol. 110, p. 130801, 2013.

[17] A. I. Zhmoginov, A. E. Charman, R. Shalloo, J. Fajans, and J. S. Wurtele, "Nonlinear dynamics of anti-hydrogen in magnetostatic traps: implications for gravitational measurements," *Classical and Quantum Gravity*, vol. 30, no. 20, p. 205014, 2013.

[18] ALPHA Collaboration and A. E. Charman, "Description and first application of a new technique to measure the gravitational mass of antihydrogen," *Nature Communications*, vol. 4, p. 1785, 2013.

[19] V. A. Kostelecky, "The Status of CPT," in *Physics beyond the standard model. Proceedings, 5th International WEIN Symposium, Santa Fe, USA, June 14-19, 1998*, pp. 588–600, 1998.

[20] M. Villata, "CPT symmetry and antimatter gravity in general relativity," *European Physics Letters*, vol. 94, no. 2, p. 20001, 2011.

[21] R. Lehnert, "CPT and Lorentz-symmetry breaking: A Review," *Frascati Physics Series*, vol. 43, pp. 131–154, 2007.

[22] S. Weinberg, *The quantum theory of fields, volume 1: foundations*. Cambridge University Press, 1995.

[23] H. Greaves and T. Thomas, "On the CPT theorem," *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, vol. 45, no. 1, pp. 46–65, 2014.

[24] H. Greaves, *Spacetime symmetries and the CPT theorem*. PhD thesis, Rutgers University, 2008.

[25] S. Maury, "The antiproton decelerator: AD," *Hyperfine Interactions*, vol. 109, p. 43, 1997.

[26] ALPHA Collaboration, "Resonant quantum transitions in trapped antihydrogen atoms," *Nature*, vol. 483, pp. 439–443, 2012.

[27] ALPHA Collaboration and A. E. Charman, "Experimental limit on the charge of antihydrogen," *Nature Communications*, vol. 5, p. 3955, 2014.

[28] ALPHA Collaboration, "An improved limit on the charge of antihydrogen from stochastic acceleration," *Nature*, vol. 529, no. 7586, pp. 373–376, 2016.

[29] G. Gabrielse, "Comparing the antiproton and proton, and opening the way to cold antihydrogen," vol. 45 of *Advances In Atomic, Molecular, and Optical Physics*, pp. 1 – 39, Academic Press, 2001.

[30] R. Fitzpatrick, *Plasma physics: an introduction*. Taylor & Francis, 2014.

[31] A. Kruithof and F. Penning, "Determination of the Townsend ionization coefficient α for pure argon," *Physica*, vol. 3, no. 6, pp. 515–533, 1936.

[32] T. M. O'Neil, "A confinement theorem for nonneutral plasmas," *Physics of Fluids*, vol. 23, p. 2216, 1980.

[33] E. M. Hollmann, F. Anderegg, and C. F. Driscoll, "Confinement and manipulation of non-neutral plasmas using rotating wall electric fields," *Physics of Plasmas*, vol. 7, no. 7, pp. 2776–2789, 2000.

[34] D. Cassidy, S. Deng, and A. Mills, "Noisy emission of a slowly released single component plasma," *Nuclear Instruments and Methods in Physics Research B*, vol. 248, no. 1, pp. 121 – 126, 2006.

[35] N. Evetts, I. Martens, D. Bizzotto, D. Longuevergne, and W. N. Hardy, "Open microwave cavity for use in a Purcell enhancement cooling scheme," *Review of Scientific Instruments*, vol. 87, no. 10, 2016.

[36] E. M. Purcell, H. C. Torrey, and R. V. Pound, "Resonance absorption by nuclear magnetic moments in a solid," *Physical Review*, vol. 69, pp. 37–38, Jan 1946.

[37] ALPHA Collaboration, "Antihydrogen formation by autoresonant excitation of antiproton plasmas," *Hyperfine Interactions*, vol. 212, p. 61, 2012.

[38] Y. V. Gott, M. Ioffe, and V. Telkovskii, "Some new results on confinement in magnetic traps," tech. rep., Academy of Sciences, Moscow, 1962.

[39] ALPHA Collaboration, "A magnetic trap for antihydrogen confinement," *Nuclear Instruments and Methods in Physics Research A*, vol. 566, pp. 746–756, 2006.

[40] J. Fajans and A. Schmidt, "Malmberg-Penning and minimum-B trap compatibility: the advantages of higher-order multipole traps," *Nuclear Instruments and Methods in Physics Research A*, vol. 521, p. 318, 2004.

[41] E. P. Gilson and J. Fajans, "Quadrupole induced resonant particle transport in a pure-electron plasma," *Physical Review Letters*, vol. 90, p. 015001, 2003.

[42] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[43] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[44] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C: the art of scientific computing*. Cambridge University Press, 2nd ed., 1992.

[45] L. N. Trefethen and D. Bau, *Numerical linear algebra*. Society for Industrial and Applied Mathematics, 1997.

[46] J. W. Demmel, *Applied numerical linear algebra*. Society for Industrial and Applied Mathematics, 1997.

[47] C. T. Kelley, *Iterative methods for optimization*. Society for Industrial and Applied Mathematics, 1999.

[48] S. Lang, *Real and functional analysis*. Springer-Verlag, 1993.

[49] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics Doklady*, vol. 4, no. 4, pp. 1035–1038, 1963.

[50] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall, 1983.

[51] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with CUDA," *Queue*, vol. 6, no. 2, pp. 40–53, 2008.

[52] J. Cao, K. A. Novstrup, A. Goyal, S. P. Midkiff, and J. M. Caruthers, "A parallel Levenberg-Marquardt algorithm," in *Proceedings of the 23rd International Conference on Supercomputing*, ICS '09, (New York, NY, USA), pp. 450–459, ACM, 2009.

[53] nVidia, *CUBLAS Library User Guide*. nVidia, v5.0 ed., 2012.

[54] C. Elliott, V. Vijayakumar, W. Zink, and R. Hansen, "National Instruments LabVIEW: A programming environment for laboratory automation and measurement," *Journal of the Association for Laboratory Automation*, vol. 12, no. 1, pp. 17–24, 2007.

[55] A. J. Peurrung and J. Fajans, "A pulsed microchannel plate based nonneutral plasma imaging system," *Review of Scientific Instruments*, vol. 64, no. 1, 1993.

[56] ALPHA Collaboration, "Antiproton, positron, and electron imaging with a microchannel plate/phosphor detector," *Review of Scientific Instruments*, vol. 80, p. 123701, 2009.

[57] A. Peurrung and J. Fajans, "Experimental dynamics of 2-D vorticity rings," in *1992 International Conference on Plasma Physics*, pp. III–1899, European Physical Society, 1992.

[58] R. Landau, J. Páez, and C. Bordeianu, *A survey of computational physics: Introductory computational science*. Princeton University Press, 2011.

[59] R. Parthasarathy, "Rapid, accurate particle tracking by calculation of radial symmetry centers," *Nature Methods*, vol. 9, no. 7, pp. 724–726, 2012.

[60] L. G. Roberts, *Machine Perception of Three-Dimensional Solids*. Garland Publishing, 1963.

[61] S. M. Anthony and S. Granick, "Image analysis with rapid and accurate two-dimensional gaussian fitting," *Langmuir*, vol. 25, no. 14, pp. 8152–8160, 2009.

[62] ALPHA Collaboration, "Experimental and computational study of the injection of antiprotons into a positron plasma for antihydrogen production," *Physics of Plasmas*, vol. 20, p. 043510, 2013.

[63] L. Wasserman, *All of statistics: a concise course in statistical inference*. Springer Publishing Company, Incorporated, 2010.

[64] J. Hopcroft and R. Kannan, *Foundations of data science*. August 2014.

[65] D. L. Eggleston, C. F. Driscoll, B. R. Beck, A. W. Hyatt, and J. H. Malmberg, "Parallel energy analyzer for pure electron plasma devices," *Physics of Fluids B*, vol. 4, no. 10, 1992.

[66] B. R. Beck, *Measurement of the magnetic and temperature dependence of the electron-electron anisotropic temperature relaxation rate*. PhD thesis, University of California, San Diego, 1990.

[67] S. Chapman, *The effect of multipole-enhanced diffusion on the joule heating of a cold non-neutral plasma*. PhD thesis, University of California, Berkeley, 2011.

[68] J. Fajans, 2016. private communication.

[69] H. C. Chen and H. S. Uhm, "Diocotron instability of an intense relativistic electron beam in an accelerator," *Physical Review A*, vol. 32, pp. 1657–1662, 1985.

[70] Wolfram Research, Inc., Mathematica, Version 11.0, Champaign, IL (2016).

[71] T. R. Weber, J. R. Danielson, and C. M. Surko, "Creation of finely focused particle beams from single-component plasmas," *Physics of Plasmas*, vol. 15, no. 1, 2008.

[72] J. Jackson, *Classical electrodynamics*. John Wiley & Sons, Inc, 3rd ed., 1999.

[73] E. Hunter *et al.*, 2017. Plasma diagnostics using a photomultiplier (private communication).

[74] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes, volume I: elementary theory and methods*. Springer, 2003.

[75] C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation*. McGraw-Hill Book Company, 1985.

[76] J. P. Verboncoeur, A. B. Langdon, and N. T. Gladd, "An object-oriented electromagnetic PIC code," *Computer Physics Communications*, vol. 87, p. 199, 1995.

[77] J. M. Dawson, "Particle simulation of plasmas," *Reviews of Modern Physics*, vol. 55, pp. 403–447, Apr 1983.

[78] A. Hirose, *Electromagnetic Theory Lecture Notes*. University of Saskatchewan, 2011.

[79] J. Barnes and P. Hut, "A hierarchical O(N log N) force-calculation algorithm," *Nature*, vol. 324, pp. 446–449, Dec. 1986.

[80] L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations," *Journal of Computational Physics*, vol. 73, no. 2, pp. 325 – 348, 1987.

[81] T. M. O'Neil, "Cooling of a pure electron plasma by cyclotron radiation," *Physics of Fluids*, vol. 23, p. 725, 1980.

[82] T. M. Antonsen, A. Fliflet, J. P. Calame, and B. Levush, "Collective theory of shot noise in gyroklystrons," *Physics of Plasmas*, vol. 8, no. 10, 2001.

[83] F. Robicheaux, 2016. Simulation of oscillator equations (private communication).

[84] E. Kreyszig, *Advanced engineering mathematics*. John Wiley & Sons, Inc., 8th ed., 2000.

[85] J. Tan and G. Gabrielse, "Parametrically pumped electron oscillators," *Physical Review A*, vol. 48, pp. 3105–3122, Oct 1993.

[86] R. Burden and J. Faires, *Numerical Analysis*. Cengage Learning, 2004.

[87] C. M. Bender and S. A. Orszag, *Advanced mathematical methods for scientists and engineers*. McGraw-Hill Book Company, 1978.

[88] G. N. Watson, *A treatise on the theory of Bessel functions.* Cambridge University Press, 1944.

[89] R. Gopalan and J. Fajans, "Non-neutral plasmas in gradient fields," *Bulletin of the American Physical Society*, vol. 42, p. 1959, 1997.

[90] J. Fajans, "Non-neutral plasma equilibria, trapping, separatrices and separatrix crossing in magnetic mirrors," *Physics of Plasmas*, vol. 10, p. 1209, 2003.

[91] R. Pathria and P. Beale, *Statistical mechanics.* Elsevier Science, 1996.

[92] A. E. Charman, *Random aspects of beam physics and laser-plasma interactions.* PhD thesis, University of California, Berkeley, 2007.

[93] M. Hori, A. Sótér, D. Barna, A. Dax, R. Hayano, S. Friedreich, B. Juhász, T. Pask, E. Widmann, D. Horváth, L. Venturelli, and N. Zurlo, "Two-photon laser spectroscopy of antiprotonic helium and the antiproton-to-electron mass ratio," *Nature*, vol. 475, pp. 484–488, 2011.

[94] M. S. Fee, S. Chu, A. P. Mills, R. J. Chichester, D. M. Zuckerman, E. D. Shaw, and K. Danzmann, "Measurement of the positronium $1\,^3S_1$–$2\,^3S_1$ interval by continuous-wave two-photon excitation," *Physical Review A*, vol. 48, pp. 192–219, 1993.

[95] M. Baquero-Ruiz, *Studies on the Neutrality of Antihydrogen.* PhD thesis, University of California, Berkeley, 2013.

[96] M. Baquero-Ruiz, A. E. Charman, J. Fajans, A. Little, A. Povilus, F. Robicheaux, J. Wurtele, and A. I. Zhmoginov, "Using stochastic acceleration to place experimental limits on the charge of antihydrogen," *New Journal of Physics*, vol. 16, p. 083013, 2014.

[97] ALPHA Collaboration, "Trapped antihydrogen," *Nature*, vol. 468, pp. 673–676, 2010.

[98] ALPHA Collaboration, "Confinement of antihydrogen for 1000 seconds," *Nature Physics*, vol. 7, pp. 558–564, 2011.

[99] ALPHA Collaboration, "Discriminating between antihydrogen and mirror-trapped antiprotons in a minimum-B trap," *New Journal of Physics*, vol. 14, p. 015010, 2012.

[100] Commercial product from COMSOL, Inc., http://www.comsol.com/.

[101] J. Bowman and S. Penttila, "On the measurement of the neutron lifetime using ultracold neutrons in a vacuum quadrupole trap," *Journal of Research of the National Institute of Standards and Technology*, vol. 110, pp. 361–366, 2005.

[102] H. Jeffreys, *Theory of probability.* Clarendon Press, 1961.

[103] E. Jaynes, *Probability theory: the logic of science*. Cambridge University Press, 2003.

[104] G. D'Agostini, *Bayesian reasoning in data analysis: a critical introduction*. World Scientific, 2003.

[105] D. S. Sivia and J. Skilling, *Data analysis: a Bayesian tutorial*. Oxford University Press, 2nd ed., 2006.

[106] P. Gregory, *Bayesian logical data analysis: a comparative approach with Mathematica support*. Cambridge University Press, 2010.

[107] W. von der Linden, V. Dose, and U. von Toussaint, *Bayesian probability theory: applications in the physical sciences*. Cambridge University Press, 2014.

[108] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-Interscience, 1991.

[109] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge University Press, 6th ed., 2003.

[110] H. Akaike, "A new look at statistical model identification," *IEEE Transactions of Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[111] P. McCullagh and J. A. Nelder, *Generalized linear models*. London: Chapman & Hall, 2nd ed., 1989.

[112] W. Haynes, *CRC Handbook of Chemistry and Physics*. Taylor & Francis, 93rd ed., 2012.

[113] P. G. H. Sandars, "The electric-dipole moments of an atom II: The contribution from an electric-dipole moment on the electron with particular reference to the hydrogen atom," *Journal of Physics B: Atomic and Molecular Physics*, vol. 1, no. 3, p. 511, 1968.

[114] ACME Collaboration, "Order of magnitude smaller limit on the electric dipole moment of the electron," *Science*, vol. 343, no. 6168, pp. 269–272, 2014.

[115] H. J. Metcalf and P. van der Straten, *Laser cooling and trapping*. Springer-Verlag, 1999.

[116] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 236, no. 767, pp. pp. 333–380, 1937.

[117] P. M. Bellan, *Fundamentals of plasma physics*. Cambridge University Press, 2006.

[118] W. Craig, *Hamiltonian dynamical systems and applications*. Springer Netherlands, 2008.

[119] D. Griffiths, *Introduction to quantum mechanics*. Pearson Prentice Hall, 2005.

[120] V. Arnold, E. Khukhro, V. Kozlov, and A. Neishtadt, *Mathematical aspects of classical and celestial mechanics*. Springer Berlin Heidelberg, 2007.

[121] H. Goldstein, *Classical mechanics*. Addison Wesley, 2nd ed., 1980.

[122] R. G. Littlejohn, "Hamiltonian formulation of guiding center motion," *Physics of Fluids*, vol. 24, p. 1730, 1981.

[123] R. J. Perkins, *Experimental and analytical studies of merging plasma loops on the Caltech Solar Loop experiment*. PhD thesis, California Institute of Technology, 2011.

[124] S. Denisov, W. Horsthemke, and P. Hänggi, "Generalized Fokker-Planck equation: derivation and exact solutions," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 68, no. 4, pp. 567–575, 2009.

[125] R. C. Davidson, *Physics of nonneutral plasmas*. Addison-Wesley, 1990.

[126] J. J. Bollinger, D. J. Wineland, and D. H. E. Dubin, "Nonneutral ion plasmas and crystals, laser cooling, and atomic clocks," *Physics of Plasmas*, vol. 1, no. 5, 1994.

[127] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[128] A. Papoulis and S. Pillai, *Probability, random variables, and stochastic processes*. McGraw-Hill, 2002.

[129] J. O. Berger, *Statistical decision theory and Bayesian analysis*. Springer, 6th ed., 1985.

[130] J. Bernardo and A. F. Smith, *Bayesian theory*. Wiley, 6th ed., 2000.

[131] C. C. Rodríguez, "The ABC of model selection: AIC, BIC and the new CIC," in *25th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (K. Knuth, A. Abbas, R. Morris, and J. Castle, eds.), vol. AIP Proceedings Volume 803, pp. 80–87, Melville, NY: American Institute of Physics, 2005.

[132] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer, 2001.