An Improved Antihydrogen Trap

A thesis presented
by

Rita Rani Kalra

to
The Department of Physics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Physics

Harvard University
Cambridge, Massachusetts
October 2014

Thesis advisor

Gerald Gabrielse

Author

Rita Rani Kalra

An Improved Antihydrogen Trap

Abstract

The recent demonstration of trapped atomic antihydrogen for 15 to 1000 seconds is a milestone towards precise spectroscopy for tests of CPT invariance. The confinement of a total of 105±21 atoms in a quadrupole magnetic trap was made possible by several improved methods. Improved accumulation techniques give us the largest numbers of constituent particles yet: up to 10 million antiprotons and several billion positrons. A novel cooling protocol leads to 3.5 K antiprotons, the coldest ever observed. Characterizing and controlling the geometry and density of these confined antimatter plasmas allow for consistency in antihydrogen production. Continued use of these methods along with the larger trap depth of a unique second-generation magnet are expected to yield greater numbers of trapped antihydrogen. The new magnet generates both quadrupole and octupole trap geometries, which should make it possible to reduce charged particle loss and will prove useful for laser cooling and spectroscopy. The ultra-low inductances of the magnet have been shown to vastly reduce turn-off times, which will optimize single-atom detection. Finally, improved detector characterization already makes us sensitive to smaller numbers of trapped antihydrogen atoms than before.

Contents

Acknowledgments

I am grateful for the opportunity to have spent my graduate years working on such an intriguing, exciting, and character-building experiment. In the ATRAP collaboration I worked with exceptionally talented people who have served as role models. First, I owe a debt of gratitude to my adviser Gerald Gabrielse for taking me into the group, for his support and innovative ideas, and for pushing me to do my best. My Harvard ATRAP predecessors Steve Kolthammer, Rob McConnell and Phil Richerme could not have done a better job in their tutelage and showing me the ropes of the experiment. Elise Novitski's contributions to ATRAP during our initial overlap became a valuable part of my thesis work. Later on I had the pleasure of working with Stephan Ettenauer, Eric Tardiff, Nate Jones, Martin Kossick, and Chris Hamley, all of whose dedication was unmistakable. Former postdocs Jonathan Wrubel and Ben Levitt were always interested in hearing ATRAP updates and were happy to provide advice even many years after leaving the group, for which I am thankful.

Our York University collaborators Eric Hessels, Cody Storry, Matt George, Matt Weel and Dan Fitzakerley were pivotal in keeping the positrons coming, and also in their willingness to jump into whatever projects needed to get done. Our University of Mainz collaborators Jochen Walz and Andi Müllers provided laser systems and kept them working very reliably through my first beam run. Our Forschungszentrum Jülich collaborators Walter Oelert, Dieter Grzonka and Marcin Zielinski provided moral support and vigilantly kept the detectors in order. I'm especially grateful to Dieter for the considerable time he spent discussing the detection system with me and for looking over parts of this thesis.

I would like to thank the rest of my Gabrielse group labmates, both past and present, for the company and good conversation that made working in lab a pleasant experience. These include Nick Guise, Josh Goldman, Yulia Gurevich, Shannon Fogwell, Jack DiSciacca, Josh Dorr, Paul Hess, Ben Spaun, Kathryn Marable, Mason Marshall, Cris Panda, Ronald Alexander, and Maryrose Barrios.

I am grateful to Stan Cotreau for teaching me the way of machining and for making the machine shop a fun haven with his banter. Jim MacArthur has been a most excellent resource for everything electronics-related, and could always be relied on for a touch of humor. I am thankful to the SEAS machine shop, particularly Louie DeFeo and Rich Anderson (sadly no longer with us) for bending over backwards to help ATRAP meet urgent deadlines. Jan Ragusa, Carol Davis, Lisa Cacciabaudo, and Jacob Barandes provided warm support and kept my PhD moving along. I also thankfully acknowledge my former adviser and committee member Ron Walsworth and committee member Masahiro Morii for their time and advice.

Finally, I would like to thank my family. Without my mother's enduring patience, sacrifices and inspiration throughout my life, none of my accomplishments would have been possible. I am also grateful to Paul Coote for persevering through my PhD and his never-ending help, support and encouragement.

For my mother, Radha Kalra

Chapter 1

Introduction

Antihydrogen ($\overline{\mathrm{H}}$), the simplest antimatter atom, is the bound state of an antiproton ($\overline{\mathrm{p}}$) and a positron (e^+). Studies on trapped antihydrogen bring us closer to testing more precisely some of the most intriguing questions in modern science: Is reality really invariant under CPT? Why is there more matter than antimatter? Does antimatter fall up or down?

A comparison of antihydrogen with its matter counterpart, hydrogen (H), may hold the key to answering these questions. The standard model of physics predicts that matter and antimatter have the same mass and lifetime, and equal but opposite charges and magnetic moments. It also predicts that matter and antimatter atoms have the same internal structure. This is a consequence of the CPT theorem for quantum electrodynamics (QED), which states that the combined effect of charge conjugation (C), parity inversion (P), and time reversal (T) is a perfect symmetry of nature. The CPT theorem is one of the most profound and fundamental theories in physics. It results from quantum field theory and the assumptions of locality and invariance under Lorentz transformations. Any local Lorentz-invariant theory, including the standard model and QED, manifests CPT-invariance.

No violation of the Standard Model has been found so far, but we know it is an

incomplete theory. It does not, for example, include a renormalizable quantum theory of gravity, leaving open the question of whether a theory of gravity is CPT-symmetric. A standard model extension has been developed that allows for CPT and Lorentz breaking [1] while keeping desirable features such as energy-momentum conservation, gauge-invariance, renormalizability, and microcausality. This extension leads to a modification of the Dirac equation by the addition of extra Lorentz-violating terms, some of which are also CPT-violating. Since neither CPT nor Lorentz breaking has been observed so far, these terms are presumably very small, suggesting that experiments to detect these violations will need to be exceptionally sensitive. The potential discovery of CPT violations, or else improving the bounds within which CPT is known to hold, is one of the main motivations of this work.

Fig. 1.1 shows the bounds placed on CPT violations by experiments to date. CPT and Lorentz breaking has been searched for in mesons, leptons, baryons, and gauge bosons. By comparing the H-$\overline{\text{H}}$ 1s-2s energy levels, we hope to extend this search to higher precision tests with a mixed baryon-lepton system. The best bound on CPT thus far comes from a high-energy particle physics experiment in the meson sector–the comparison of K^0-$\bar{K}^0$ mass [2]. However, the best bounds on CPT for lepton and baryon systems have been found in low-energy experiments using single particles in Penning traps. A comparison between the electron and positron g-factors provided the best test for leptons [3] and should soon be improved by a factor of 15 or more in our research group. A comparison between the proton-antiproton charge to mass ratios made by an early measurement in our group yielded the best test for baryons [4] by many orders of magnitude.

It is attractive to pursue this search in the comparison of H-$\overline{\text{H}}$, because the narrow 1.3 Hz linewidth of the metastable 1s-2s transition in H could give rise to a very small 5×10^{-16} fractional uncertainty. The best measured frequency of this line is 2 466 061 413 187 035(10) Hz, with a fractional uncertainty of 4.2×10^{-15} [5]. This was done via two-photon spec-

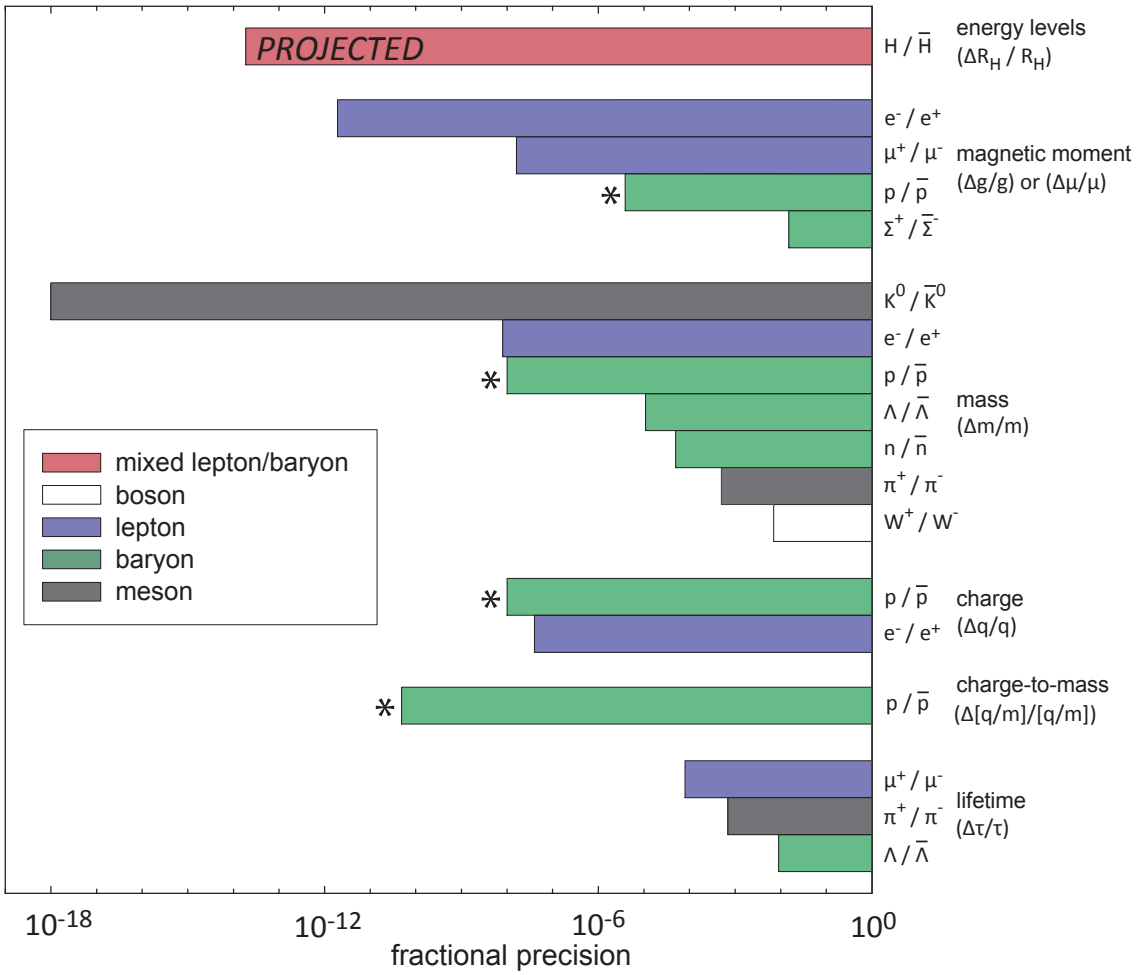


Figure 1.1: CPT tests done to date in various particle sectors. A * identifies tests that are based upon Gabrielse group measurements.

troscopy on a 5.8 K atomic beam. Making an atomic beam of antihydrogen is unlikely due to the 10 orders of magnitude fewer antihydrogen available in a typical experiment, as $\overline{\mathrm{H}}$ atoms are produced from rare constituent particles. However, the hydrogen 1s-2s frequency has also been measured in a magnetic trap [6] and attained a fractional uncertainty of 1.2×10^{-12}, with the stability of the laser system being the major limitation. Magnetic trapping of neutral antihydrogen is feasible, and the uncertainty attained in a magnetic trap could in principle be made better.

In 1986 Gabrielse proposed [7] to create and confine antihydrogen atoms in a neutral particle trap for precision study of their properties. To this end, the ATRAP collaboration was formed from its predecessor TRAP, which successfully slowed, captured, and cooled antiprotons for the first time [8]. The ATRAP collaboration includes researchers from Harvard University, York University, Forschungszentrum Jülich and University of Mainz. Antihydrogen experiments are now being pursued by others who joined the quest. Four international collaborations now work at the CERN Antiproton Decelerator (AD) to study antihydrogen, the only place in the world where this work can be done. Recently, the confinement of antihydrogen was demonstrated by both the ALPHA and ATRAP collaborations [9, 10], a critical step towards precision studies. Both used the $\overline{\mathrm{H}}$ formation process of three-body recombination [11], which involves collisions of an antiproton with two positrons.

This thesis describes work done with two iterations of ATRAP antihydrogen traps, called BTRAP and CTRAP. BTRAP was used in the 2011 beam run to trap on average 5 ± 1 antihydrogen per trial, numbers greater than previously reported. CTRAP is the next-generation apparatus, very similar to BTRAP, except with a greatly improved magnetic trap for antihydrogen. CTRAP has been under construction for several years, and was finally completed in 2012. An attempt was made to commission it for the first time during the 2012 beam run, but a helium leak in the Ioffe magnet enclosure prevented us from doing so. Thus 2013 was spent taking apart the enclosure, designing a new leak-tight enclosure and electrode stack, building a new cryogenic test apparatus just for the magnet, and carrying out the successful testing of the magnet and its quench-protection system.

ATRAP is a collaboration composed of many members, many of whom made important contributions to the work discussed here. My role can be divided up into 4 parts: antihydrogen trapping in BTRAP, improvements in $\overline{\mathrm{H}}$ detection analysis, partial construction and extensive wiring of CTRAP, and the electrical testing of the most important compo-

nents of the new antihydrogen trap: the magnetic Ioffe trap along with its quench-protection system.

Chapter 2 describes the apparatus and infrastructure needed to trap $\overline{\mathrm{H}}$: the AD, the positron accumulator, the Penning trap, the Ioffe trap, and the annihilation detector. The AD and positron accumulator provide the constituent antiprotons and positrons. The Penning trap confines the charged particles. Single particle motion in a Penning trap is explained in order to characterize the Penning trap, although we work with plasmas for $\overline{\mathrm{H}}$ experiments. The Ioffe trap, used to trap neutral antihydrogen, is introduced. The detection system is described, along with the procedure used to determine detector positions.

The next chapter continues by detailing some specific pieces of next-generation anithydrogen apparatus. In addition to an improved Ioffe trap, CTRAP also contains many other improved parts. Thus Chapter 3 motivates and presents some of the CTRAP apparatus projects I worked on throughout my time here along with descriptions of the improvements made.

In order to use the apparatus to perform antihydrogen experiments, the dynamics of plasmas in Penning traps must be understood. In practice, trapping antihydrogen is a plasma experiment–we confine not single particles but clouds of particles which exhibit collective behavior. Chapter 4 provides background on plasmas which will be referred to throughout this thesis.

Chapter 5 presents previously established methods used to trap antihydrogen, including particle loading, detection, and manipulation. It also includes recently published advances in plasma control and cooling [12, 13]. Chapter 6 describes the methods used to form antihydrogen. My contributions include studies to understand and overcome our biggest challenge: losses of charged particles in the quadrupole Ioffe field. They also include a series of detection-well experiments to explore the parameter space, with the goal of using

the optimal drive settings for $\overline{\mathrm{H}}$ experiments.

Chapter 7 reports our 2011 trapped antihydrogen results in BTRAP [10], done together with Phil Richerme and also the subject of his thesis [14]. The analysis that follows was based on an analysis reported in the thesis of Robert McConnell [15]. Limitations of the BTRAP Ioffe magnet used in these experiments are also discussed, along with the problem of trial-to-trial annihilation detection sensitivity precluding us from finding the optimal way to trap antihydrogen.

In Chapter 8 I present multivariate statistical methods that improve our understanding of measurement data obtained during the successful BTRAP antihydrogen experiments. Analysis of the distribution of calibration data obtained before the experiments reveals highly favorable statistical properties, and has already been shown to surpass the $\overline{\mathrm{H}}$ sensitivity found from previous methods. Exploiting these properties to distinguish between measurement signals associated with genuine $\overline{\mathrm{p}}$ annihilations and interference from cosmic rays yields the ability to observe fewer $\overline{\mathrm{H}}$ per trial than before. The higher confidences will help gauge the effectiveness of different experimental parameters used from trial-to-trial. This method may also prove useful at the level of precision measurements on single atoms.

Chapter 9 recounts the 2012 attempted commissioning of the next-generation antihydrogen trap at CERN and its initial cooldown failure, followed by the design and construction of a dedicated cryogenic test apparatus just for the Ioffe magnet. The Ioffe magnet design was reported in the thesis of Steve Kolthammer [16]. A description of the quench-protection and voltage tap data-acquisition system, designed largely by Elise Novitski, follows. The magnetometer data acquisition system was designed and set up by Eric Tardiff. My contributions include the test apparatus design, construction and implementation, quench-system calibration, development of testing methods, low-current and initial high-current testing, determination of magnet turn-off times, and analysis of the quench

signals. Quench-training of the magnet was completed by Stephan Ettenauer and Martin Kossick after I returned to Harvard.

Chapter 10 concludes with a summary and future directions.

Chapter 2

Apparatus

At CERN, the world's only source of low-energy antiprotons, ATRAP has two cryogenic antihydrogen traps: one that has trapped antihydrogen (BTRAP) [10] (Fig. 2.1), and an improved second-generation apparatus (CTRAP). Both utilize Penning-Ioffe traps to confine charged antiprotons and positrons and magnetically trap the resulting neutral antihydrogen. In addition to the $\overline{\mathrm{H}}$ trap itself, the infrastructure needed to carry out these experiments includes an antiproton decelerator (AD), a positron accumulator, and a pion detector. These and the necessary cryogenic and vacuum environment needed to perform antimatter experiments will be described in this chapter.

2.1 Antiparticle sources at CERN

The creation of antiprotons starts with relativistic 26 GeV/c protons produced at CERN's Proton Synchrotron. These bunches of 10^{13} high-energy protons are sent to the AD Hall, where they are made to collide with an iridium wire target. For every million protons, about three 3.5 GeV/c (kinetic energy of 2.7 GeV) antiprotons are obtained from the collision [17]. These antiprotons are then electromagnetically-guided into the AD ring

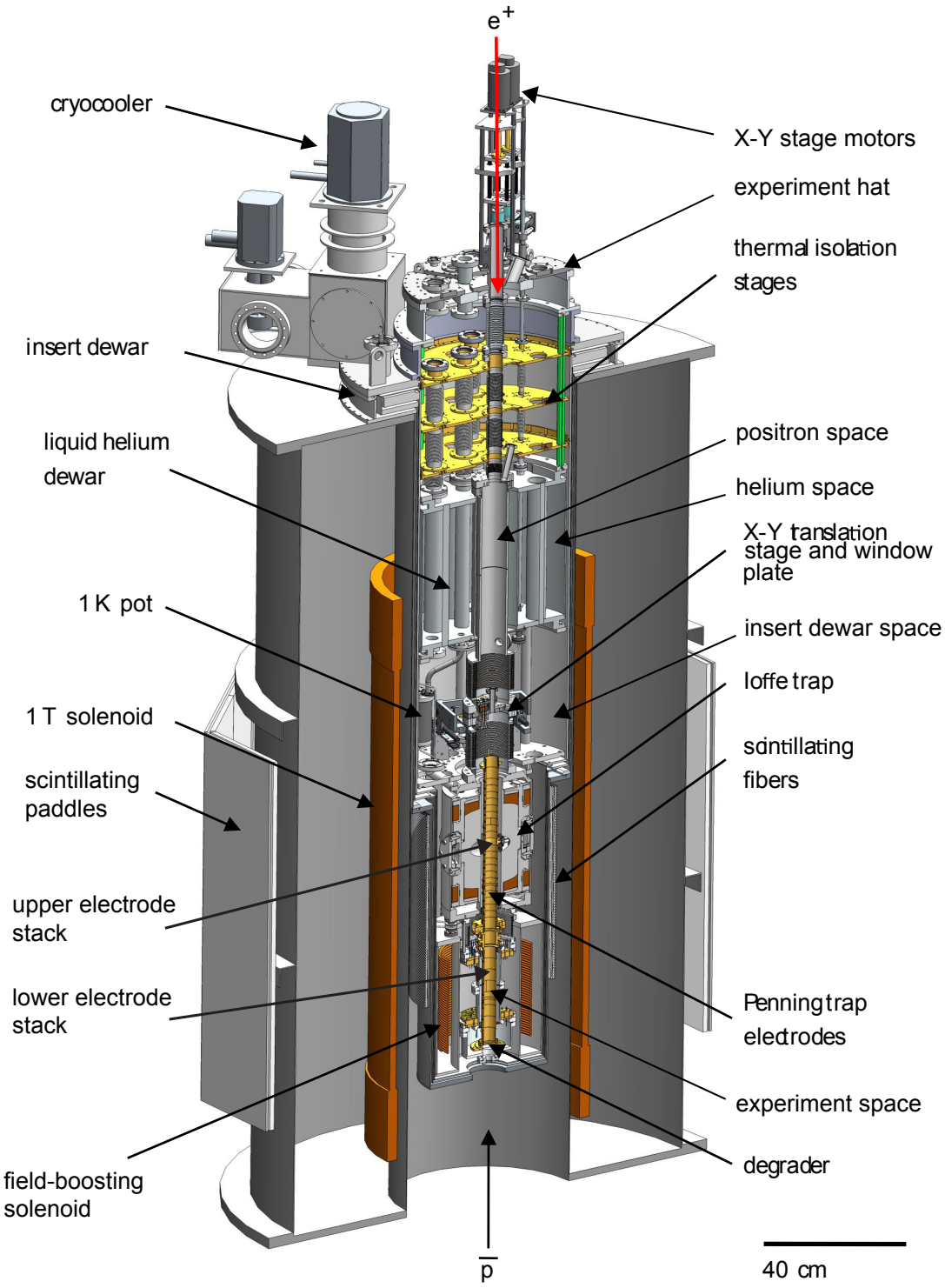


Figure 2.1: BTRAP apparatus in insert dewar and 1 T magnet bore.

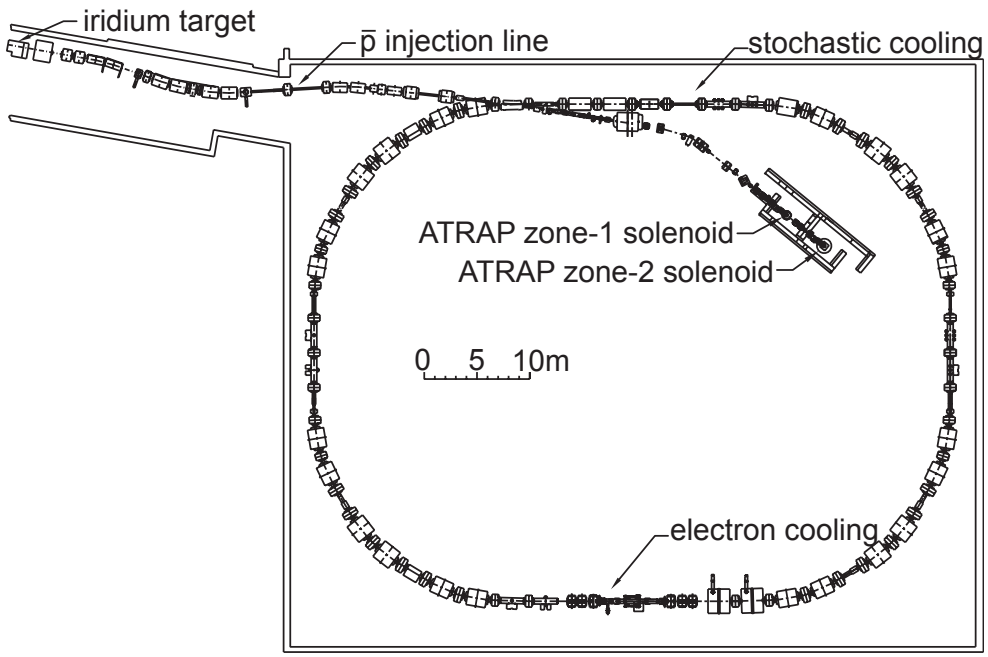


Figure 2.2: Antiproton decelerator ring. BTRAP or CTRAP are in operated in Zone 2 while a new experiment that measures the antiproton magnetic moment is located in Zone 1.

(Fig. 2.2), where their spatial emittance and momentum spread are reduced in two stochastic cooling steps [18] to 2 GeV/c and then to 300 MeV/c. Electron cooling [19] then further reduces the momentum to 100 MeV/c (kinetic energy 5.3 MeV), at which point a 200 ns bunch of 30-35 million $\overline{\mathrm{p}}$ is ejected towards our trap. Antiproton bunches are received approximately every 90 seconds.

Positrons are the other ingredient of $\overline{\mathrm{H}}$. They are produced from the β-decay of a radioactive ^{22}Na source, moderated by solid Ne to a few eV of energy, and then transferred to a room-temperature positron accumulator [20], built and operated by our York University collaborators [21]. The e^+ are then buffer-gas cooled by collisions with (low-pressure) N_2 and SF_6 as they fall into a Penning trap. A rotating wall (discussed in Chap. 5) reduces their radius to 3 mm. The positrons are accumulated for 30 seconds, and then rapidly ejected into a 10 meter long transfer line which magnetically guides the positrons into the

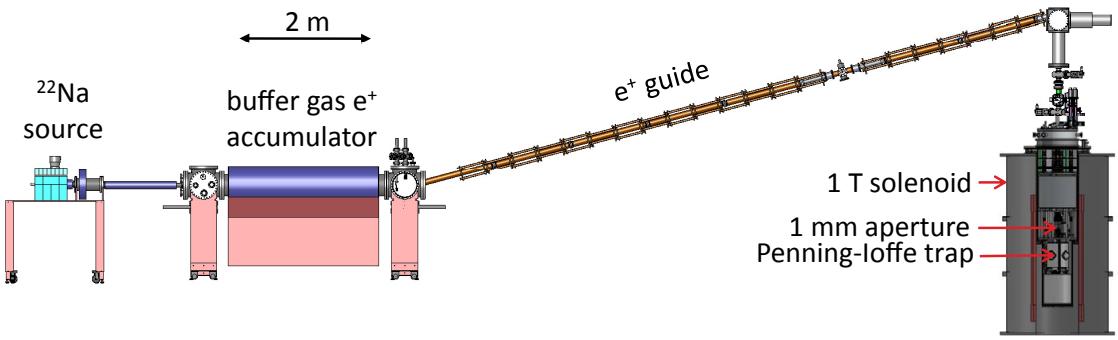


Figure 2.3: Positron Accumulator. Positrons are produced, cooled, accumulated, and launched through the 10 meter transfer line into the experiment from above.

antihydrogen trap [21] (Fig. 2.3). The transfer is nearly 100% efficient, despite the 15° and 105° upward and downward bends of the positron paths, and despite the fringing field of the 1 T bias magnet for the antihydrogen trap. Seventy steering magnets are used to achieve this. Every 30 seconds, 3 million positrons with 60 eV can be sent to the trap.

2.2 Penning traps

2.2.1 Hyperbolic Penning trap theory

The charged constituent particles of antihydrogen are confined and brought together via electrostatic and magnetostatic potentials in a Penning trap [22, 23]. Because the trap frequencies of a single particle in a Penning trap characterizes the trap, we begin with a description of single particle trajectories. The dynamics of collections of charged particles in a Penning trap will be explained in Chapter 4.

An idealized Penning trap with hyperbolic electrodes is shown in Fig. 2.4a. When the ring and endcap electrodes are biased appropriately, this configuration provides a potential minimum in the vertical direction. From Laplace's equation $\nabla^2\Phi = 0$, where Φ is the electric potential, it follows that there can be no 3D electrostatic minima or maxima in

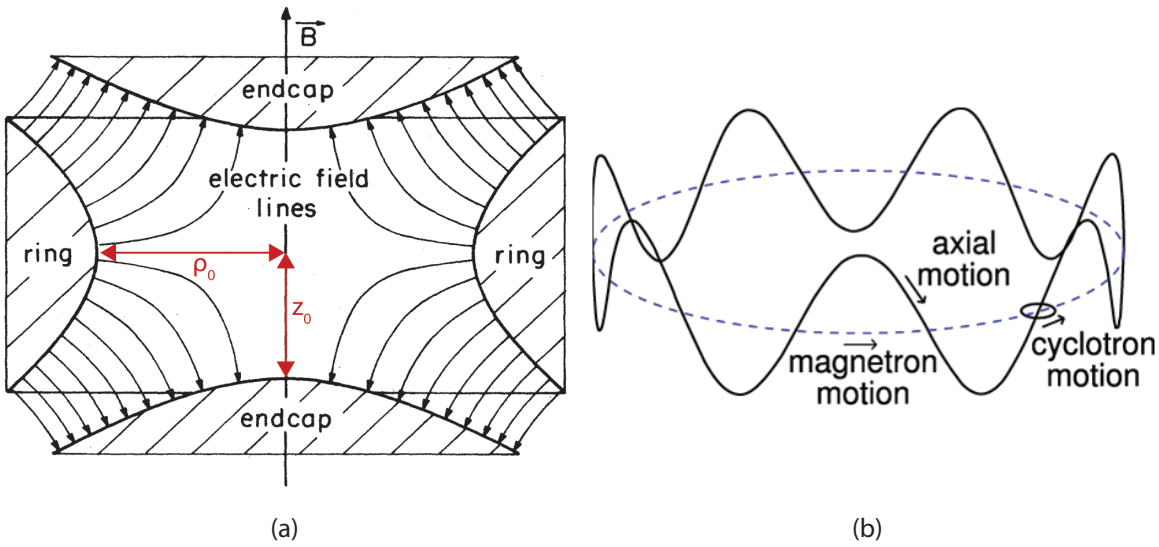


Figure 2.4: (a) Cross section of an ideal Penning trap. The hyperbolic electrodes make a quadrupole electrostatic potential that confine the particles axially and the bias uniform magnetic field confines the particles radially. (b) Single particle trajectories. The dashed line shows the guiding center motion. Adapted from [23]

free space. Therefore an electrostatic trap can only be used to confine charged particles in either one or two directions. The hyperbolic shape of the electrodes in the trap geometry shown provides an axially (vertically) confining quadrupole potential given by

$$\Phi(\rho, z) = \frac{V_0}{2d_0^2}(z^2 - \rho^2/2) + C \tag{2.1}$$

where

$$d_0^2 = \frac{1}{2}(z_0^2 + \rho_0^2/2) \tag{2.2}$$

is a geometrical constant that comes from boundary conditions and ensures that the potential difference between the ring and endcap electrodes is V_0. The lengths z_0 and ρ_0 are shown in Fig. 2.4a, and C is a constant. Because this saddle potential is radially repulsive, a uniform magnetic field $\vec{B} = B\hat{z}$ is applied to confine the particles radially.

A charged particle in these Penning trap fields experiences three distinct motions: a fast cyclotron orbit in the radial plane, a slower axial oscillation, and an even slower

magnetron orbit in the radial plane. The three motions are decoupled for a single charged particle in an ideal Penning trap. Fig. 2.4b shows these trajectories and Table 2.1 shows the hierarchy of frequencies.

To find the trap frequencies, we start with the force on a charged particle in these fields, given by the Lorentz force law:

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \tag{2.3}$$

In the absence of an electric field, the radial equation for a charged particle is

$$\begin{aligned} \ddot{\vec{\rho}} &= \frac{q}{m}(\dot{\vec{\rho}} \times \vec{B}) \\ &= \vec{\omega_c} \times \dot{\vec{\rho}} \end{aligned} \tag{2.4}$$

where $\vec{\omega_c} \equiv -\frac{q}{m}\vec{B}$. Eq. 2.4 describes a rotation of the velocity vector $\dot{\vec{\rho}}$ in the radial plane with constant angular velocity $\omega_c = |qB|/m$. This is the cyclotron frequency, and it depends only on the B-field for a given particle.

To find the equation of motion in the axial direction, we plug $\vec{B} = B\hat{z}$, $\vec{E} = -\nabla\Phi(z)$, and $\vec{v} = \dot{\vec{z}}$ into Eq. 2.3 to yield

$$\ddot{z} + \frac{qV_0}{md^2}z = 0$$

and extract the axial frequency

$$\omega_z^2 = \frac{qV_0}{md^2} \tag{2.5}$$

The axial frequency for a given particle and trap geometry depends only on the E-field.

The magnetron motion is a force-free drift in the azimuthal direction that depends on both $\vec{E}$ and $\vec{B}$. If we now plug in the radial component of the velocity $\dot{\vec{\rho}}$, $\vec{B}(z)$ and $\vec{E} = -\nabla\Phi(\rho)$ into 2.3 and express this in terms of ω_c and ω_z we get the full radial equation of motion

$$\ddot{\vec{\rho}} = \frac{1}{2}\omega_z^2\vec{\rho} - \dot{\vec{\rho}} \times \omega_c\hat{z} \tag{2.6}$$

Frequency	$\mathbf{e^+}$, $\mathbf{e^-}$	$\mathbf{\overline{p}}$
Magnetron	3.7 kHz	3.7 kHz
Axial	14 MHz	335 kHz
Cyclotron	28 GHz	15 MHz

Table 2.1: Hierarchy of frequencies of motion in Penning traps for our typical trap parameters. Note that the magnetron frequencies are identical for e^+/e^- and $\overline{p}$.

Solving Eq. 2.6 yields two radial frequencies:

$$\omega_\pm = \frac{1}{2}(\omega_c \pm \sqrt{\omega_c^2 - 2\omega_z^2}) \tag{2.7}$$

The smaller frequency ω_- is often called the magnetron frequency ω_m. The larger frequency $\omega_+ = \omega_c - \omega_m$ for an ideal trap, and is the slightly modified cyclotron frequency ω_c'. This slight reduction in cyclotron frequency is caused by the interaction of the radial E-field with the cyclotron motion.

The magnetron motion is unstable when the expression under the square root in Eq. 2.7 is negative, which can happen when the force from the repulsive radial E-field term gets larger than the confinement from the B-field. However, for the field strengths used in our traps, $\omega_c \gg \omega_m$. Furthermore, ω_m can be expressed in terms of the axial and cyclotron frequencies by noting that $\omega_+\omega_- = \omega_z^2/2$ for a perfect trap, so

$$\omega_m = \omega_z^2/2\omega_c' \approx \omega_z^2/2\omega_c \tag{2.8}$$

The magnetron frequency is independent of charge and mass, and only depends on the magnitudes of $\vec{E}$ and $\vec{B}$. We also observe that the azimuthal velocity $\vec{v} = \vec{E} \times \vec{B}/B^2$ satisfies Eq. 2.6 when the LHS = 0. That is, charged particles with this velocity move through the fields unimpeded by any force, with the same sense of rotation as the cyclotron orbit.

Synchrotron radiation

The cyclotron motion of a charged particle in a Penning trap is damped by synchrotron radiation. The energy radiated away is given by the Larmor formula

$$-\frac{dE}{dt} = \frac{q^2}{6\pi\epsilon_0 c^3}\ddot{\vec{\rho}}^2 \tag{2.9}$$

Cyclotron motion in the trap is only slightly modified from ideal cyclotron motion, so we use the cyclotron equation 2.4 for a charge in a magnetic field and its kinetic energy $E = \frac{1}{2}m\dot{\rho}^2$. This gives a solution to Eq. 2.9

$$E(t) = E_0 e^{-t/\tau_c}$$

where

$$\tau_c = \frac{3\pi\epsilon_0 mc^3}{q^2\omega_c^2} = \frac{3\pi\epsilon_0(mc)^3}{q^4B^2} \tag{2.10}$$

is the time for the energy to decay away to 1/e of E_0. For our typical B-fields of 1 and 3.7 T, the time constants for e^- (or e^+) are 2.6 s and 0.19 s, respectively. The same fields for $\overline{p}$ yield cooling times of 507 and 37 years. Thus synchrotron damping is too slow for cooling $\overline{p}$, but useful for the lighter e^- and e^+. The magnetron decay for an electron takes on the order of 10^{15} years, so there is effectively no synchrotron cooling from magnetron motion. Note that for a single particle in an ideal trap the radial cyclotron energy is decoupled from the axial energy, which could still be large despite effective synchrotron damping.

2.2.2 Cylindrical Penning trap theory

Precision hyberbolic electrodes are difficult and expensive to machine, and have the disadvantage that only particles of the same charge can be trapped within the electrostatic potential. More importantly, a large hole added to admit particles would destroy much of the hyperbola. Alternative Penning trap geometries incorporate compensated cylindrical electrodes with flat endcaps [24] and compensated cylindrical electrodes with open endcaps

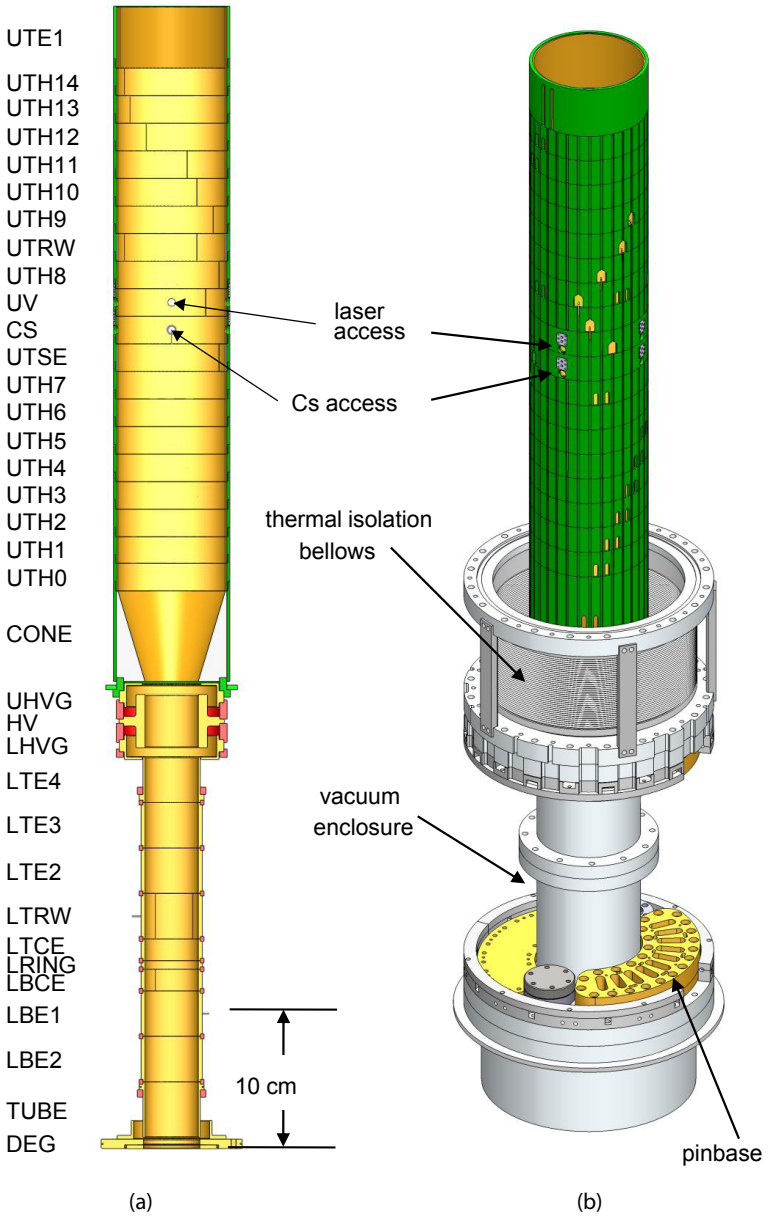


Figure 2.5: (a) Cross-sectional view of the CTRAP electrode stack. (b) Electrode stack enclosure surrounding the experiment vacuum space.

[25]. We use the latter to allow free access for the charged particles to enter the top and bottom of the trap, and to create nested wells for both antiprotons and positrons. Fig. 2.5 shows our Penning traps–a stack of ∼40 independently-biased cylindrical electrodes with 180 micron gaps in between.

Cylindrical electrodes have many advantages, but the challenge for precision experiments is to produce a sufficiently harmonic potential with this configuration. Deviations from the electric quadrupole potential cause anharmonic particle oscillations, resulting in undesired amplitude-dependent axial oscillations. Fortunately, compensation electrodes and a judicious choice of electrode length and/or applied voltage to each stack electrode can provide a nearly perfect harmonic potential at the trap center without affecting the axial oscillation frequency.

The solution to Laplace's equation for a single electrode with cylindrical boundary conditions can be expanded in Legendre polynomials. In spherical coordinates, this is

$$\Phi_0(r,\theta) = \frac{1}{2}V_0 \sum_{\substack{k=0 \\ (\text{even})}}^{\infty} C_k (\frac{r}{d_0})^k P_k(cos\theta) \tag{2.11}$$

The scale is set by d_0, given by Eq. 2.2 and z_0 and ρ_0 are the half-length and radius of the electrode. Only the even k terms are included because the electrodes are centered at $z = 0$ and only potentials symmetric about $z = 0$ are applied. The lowest-order terms dominate for particles close to the center, with $r \ll d$. The C_k coefficients depend on relative trap dimensions, but not potential. C_0 is an overall constant and can be ignored. C_2 is the strength of a perfect quadrupole potential. All terms with $k > 2$ represent undesirable trap anharmonicities, C_4 being the largest of these. The anharmonicities can be minimized by appropriate choice of the ratio of electrode length and diameter.

The first anharmonic term C_4 can be canceled out altogether by biasing additional, neighboring compensation electrodes with potential V_c, thereby superimposing the potential

$$\Phi_c(r,\theta) = \frac{1}{2}V_c \sum_{\substack{k=0 \\ (\text{even})}}^{\infty} D_k (\frac{r}{d_c})^k P_k(cos\theta)$$

where $d_c^2 = \frac{1}{2}(z_c^2 + \rho_0^2/2)$. The total potential inside the trap is the sum of potentials Φ_0 and Φ_c with the new coefficient $E_k = C_k + D_k \frac{V_c}{V_0}$. The coefficients E_k thus depend on

relative applied potentials in addition to geometry, and V_c can be adjusted to make E_4 zero. However, adjusting V_c leads to changes in C_2, changing the well depth and axial oscillation frequency. This problem is circumvented by choosing the electrode geometry to make D_2 equal to zero, in which case changing V_c will have no effect on C_2. In practice, setting $D_2 = 0$ is done by picking the appropriate trap radius ρ_0/z_0 for the compensation electrode size z_c/z_0 [24].

For applications that need precise harmonic potentials, the tuned ring and compensation electrodes LRING and LTCE/LBCE can be used. Though most of the stack electrodes are the same length, the nearest neighbor electrodes can also be used for anharmonicity compensation by adjusting their potentials V_c, and in principle more electrodes minimizes more anharmonic terms. The short upper stack electrode lengths allow for this tuning and also provide flexibility in creating multiple-well or nested-well potentials. If the goal is to simply contain particles without doing precision frequency measurements or resonant drives, it is sufficient to use single-electrode wells (assuming all other electrodes are grounded).

Fig. 2.6 shows the potentials generated by different configurations of cylindrical stack of electrodes. Near the center of each well, there is good agreement with a quadrupole potential. However, if a good harmonic potential over a larger range is needed then the hyperbolic electrode geometry is preferred.

2.3 Ioffe traps

The previous section described the confinement of charged particles in Penning traps. To confine the neutral antihydrogen, we follow the course proposed by Gabrielse many years ago [7] to trap cold antihydrogen with a Ioffe trap. The magnetic Ioffe trap

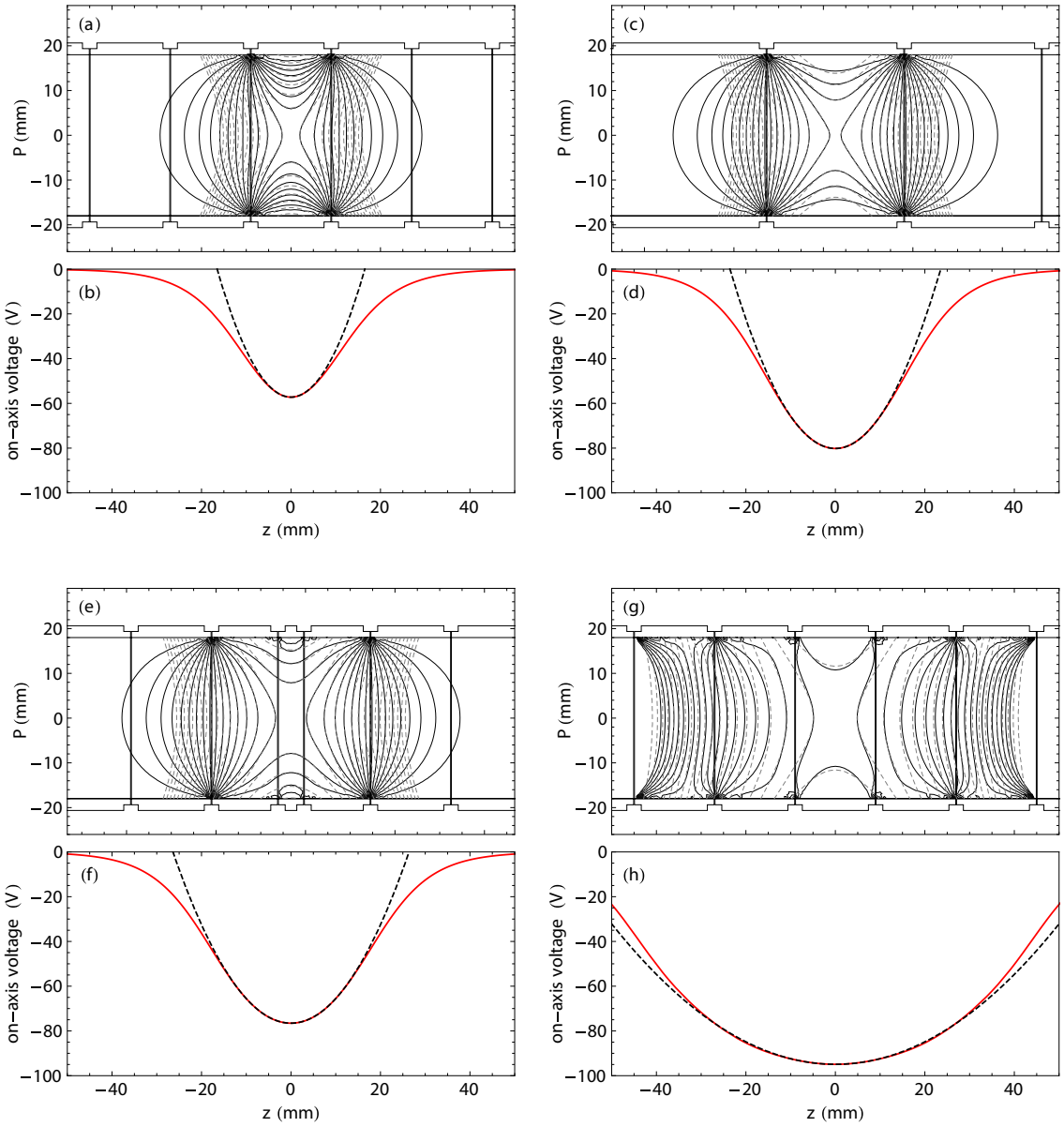


Figure 2.6: Examples of real on-axis electric potentials in the Penning trap (solid lines), compared with the the ideal quadrupole potentials (dashed). Equipotentials and on-axis potential for (a) and (b): a single electrode with $z_0 = 0.5\rho_0$, (c) and (d) a single electrode with $z_0 = 0.849\rho_0$, (e) and (f): a 3-electrode harmonic well with $z_1 = 2.52z_0$ and $V_1 = 0.881V_0$, and (g) and (h), a 5-electrode harmonic well with $V_1 = 0.931V_0$ and $V_2 = 0.706V_0$. Taken from [14]

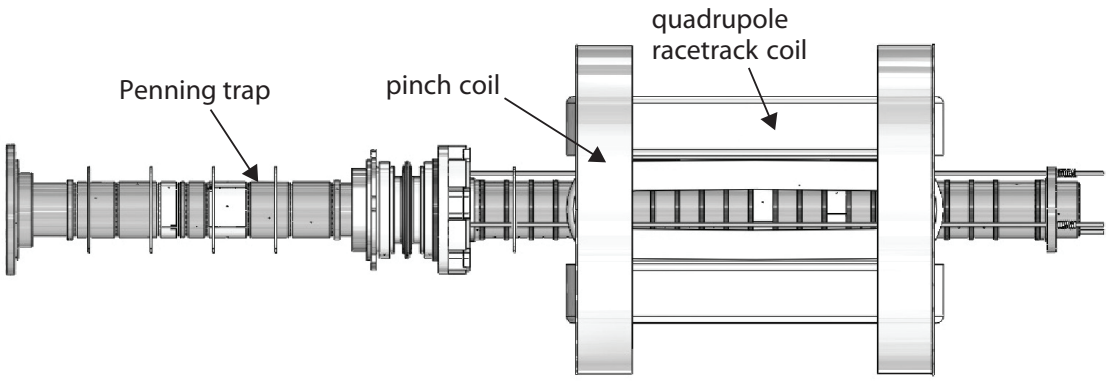


Figure 2.7: BTRAP nested Penning-Ioffe trap

[26] is superimposed with the Penning trap in a nested Penning-Ioffe trap configuration [11] shown in Fig. 2.7. The charged particles are first brought together in adjacent wells, and then the Ioffe magnetic field is turned on before $\overline{\mathrm{H}}$ production. The minimum of the Ioffe trap confines atoms that are in low-field seeking states. These atoms have magnetic moments that are anti-parallel to the B-field: $\vec{\mu} \cdot \vec{B} < 0$. For a hydrogen atom this corresponds to an electron with m_j=1/2, so for an antihydrogen atom the low-field seeking state has a positron with $m_j = -1/2$.

The trap depth for ground-state atoms in temperature units is determined from $T = \mu_B \Delta B / k_B$ where μ_B is the Bohr magneton and ΔB is the minimum trap depth in Tesla. BTRAP has a magnetic depth of 0.56 T, corresponding to a temperature trap depth of 375 mK. The B-field depth and contours are shown in Fig. 2.8.

A Ioffe quadrupole trap consists of a quadrupole magnetic field generated by alternating currents through 4 current bars (in the simplest scheme) and an axial field generated by 2 mirror ("pinch") solenoids operated with the same current sense. The quadrupole and pinch fields provide confinement in the radial and axial directions, respectively. The BTRAP quadrupole magnet was used in the trapped antihydrogen studies (Chaps. 6 and 7), and the CTRAP Ioffe trap is the next-generation magnet recently tested (Chap. 9). The geometries

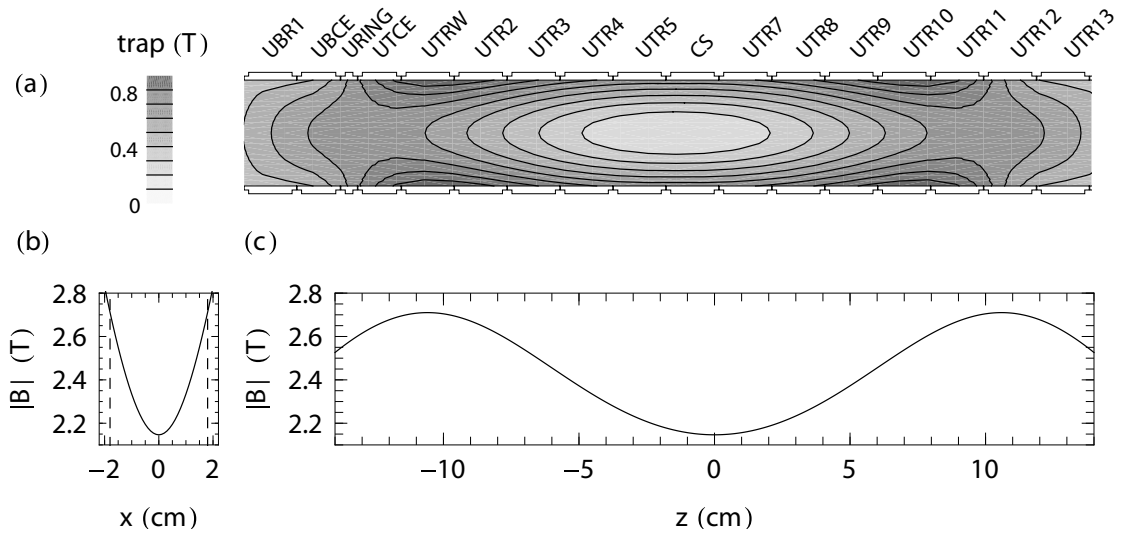


Figure 2.8: BTRAP magnetic trap contours along a cross-section of the Penning trap (a), magnetic field magnitude in radial direction at z=0 (b), and axial direction along ρ=0 (c). Taken from [16].

for the BTRAP and CTRAP Ioffe traps are shown in Fig. 2.9. Both magnets have side-port windows for laser access, and the octupole windings of CTRAP was wound around these, leading to the kinks seen at the bottom of Fig. 2.9e.

The total B-field for an ideal quadrupole trap is given by

$$\vec{B} = B_z\hat{z} + \frac{B_{\rho_0}}{\rho_0}(x\hat{x} - y\hat{y})$$

where B_z is the combination of the axial field given by the 1 T Penning trap bias field and that of the mirror coils. B_{ρ_0} is the magnetic field at the electrode wall radius ρ_0. In the absence of the magnetic quadrupole field, the field lines through the Penning trap are radially uniform. However, the addition of the quadrupole field causes field lines starting from $z = 0$ to exponentially diverge toward or away from the trap walls as they traverse the z-direction:

$$\begin{aligned} x(z) &= x_0 e^{\frac{B_{\rho_0}}{B_z \rho_0} z} \\ y(z) &= y_0 e^{-\frac{B_{\rho_0}}{B_z \rho_0} z} \end{aligned} \tag{2.12}$$

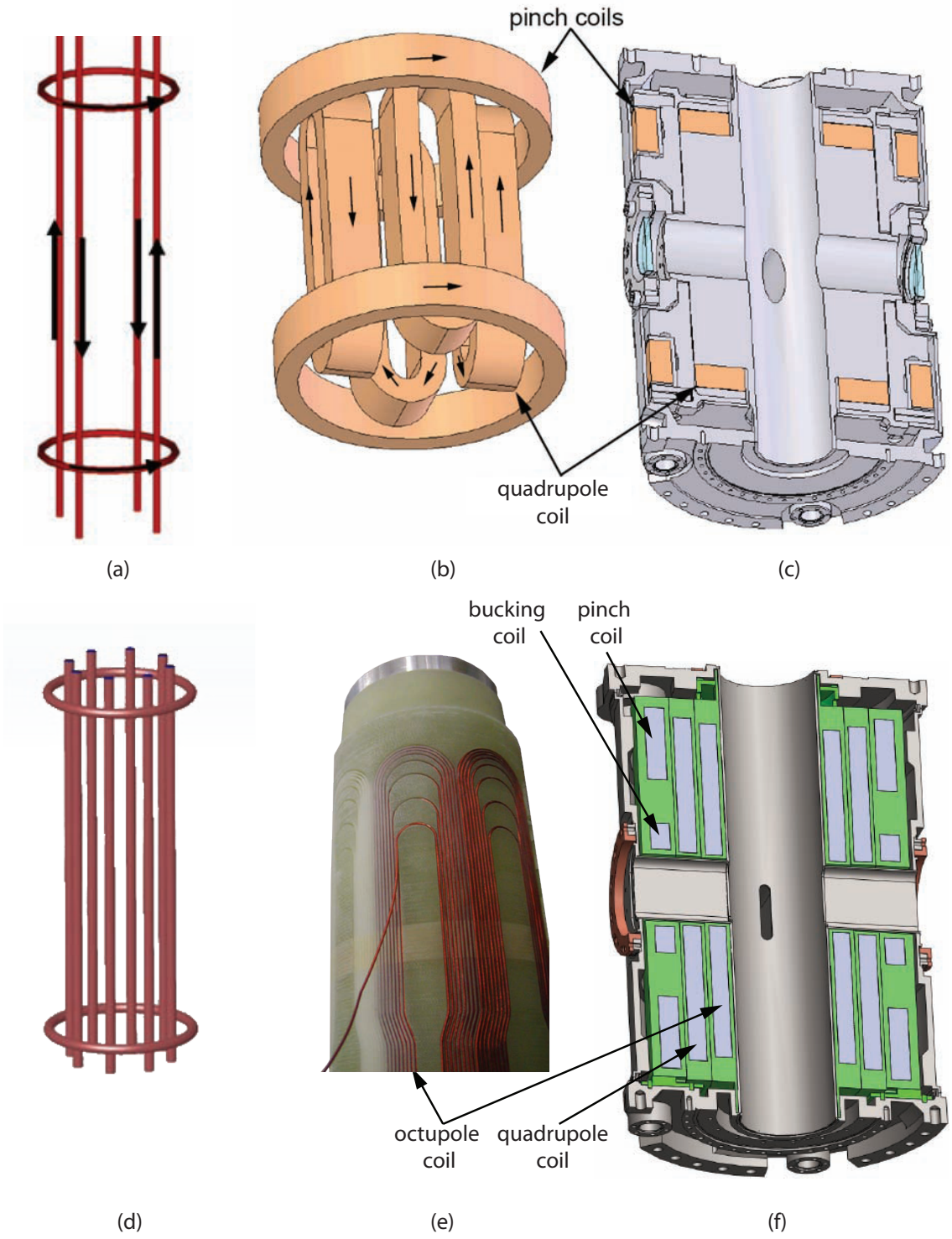


Figure 2.9: (top) BTRAP and (bottom) CTRAP magnet geometries. From left to right: simplified radial geometry, actual geometry, full magnet in helium enclosure. Both magnets contain side-port windows for laser access.

This is shown in Fig. 2.10c. The field lines take on a twisted-bowtie shape, and charged particles inside a specific radius will be ballistically lost if they are allowed to axially traverse the distance Δz needed to reach the trap walls. The well depth for an applied 50 V potential inverts at 12 mm, becoming the cutoff radius for this particular axial potential. The cutoff radius for a quadrupole field is given by r_{cut} when the following condition is satisfied:

$$\rho_0 \leq r_{cut} e^{\frac{B\rho_0}{B_z \rho_0}\Delta z} \tag{2.13}$$

The divergent trajectories past the cutoff radius posed a problem in our efforts to produce antihydrogen in the quadrupole magnetic field–significant numbers of constituent particles could be lost, rendering them unable to participate in $\overline{\mathrm{H}}$ formation. This challenge and efforts to overcome it is described in more detail in Sec. 6.5.

A higher-order multipole-field [28] with gentler field-line gradients could allow better recombination of antiprotons and positrons. To this end, a new magnet and an entirely new cryogenic apparatus (CTRAP) was designed and constructed to provide either an octupole (see Fig. 2.9) or a quadrupole field, along with fast turn-off times. The radial magnetic field given by a multipole Ioffe trap of order n with infinitely long current bars is [29]:

$$\vec{B}(\rho,\theta) = B_{\rho_0}(\frac{\rho}{\rho_0})^{n-1}(cos(n\theta)\hat{\rho} - sin(n\theta)\hat{\phi})$$

and the magnitude is

$$|\vec{B}(\rho,\theta)| = B_{\rho_0}(\frac{\rho}{\rho_0})^{n-1}$$

For a quadrupole field, n=2 and the field magnitude grows linearly with radius. For an octupole field, n=4 and the field grows cubically with radius. The BTRAP and CTRAP Ioffe traps are discussed in more detail in Sec. 7.2 and Chap. 9, respectively.

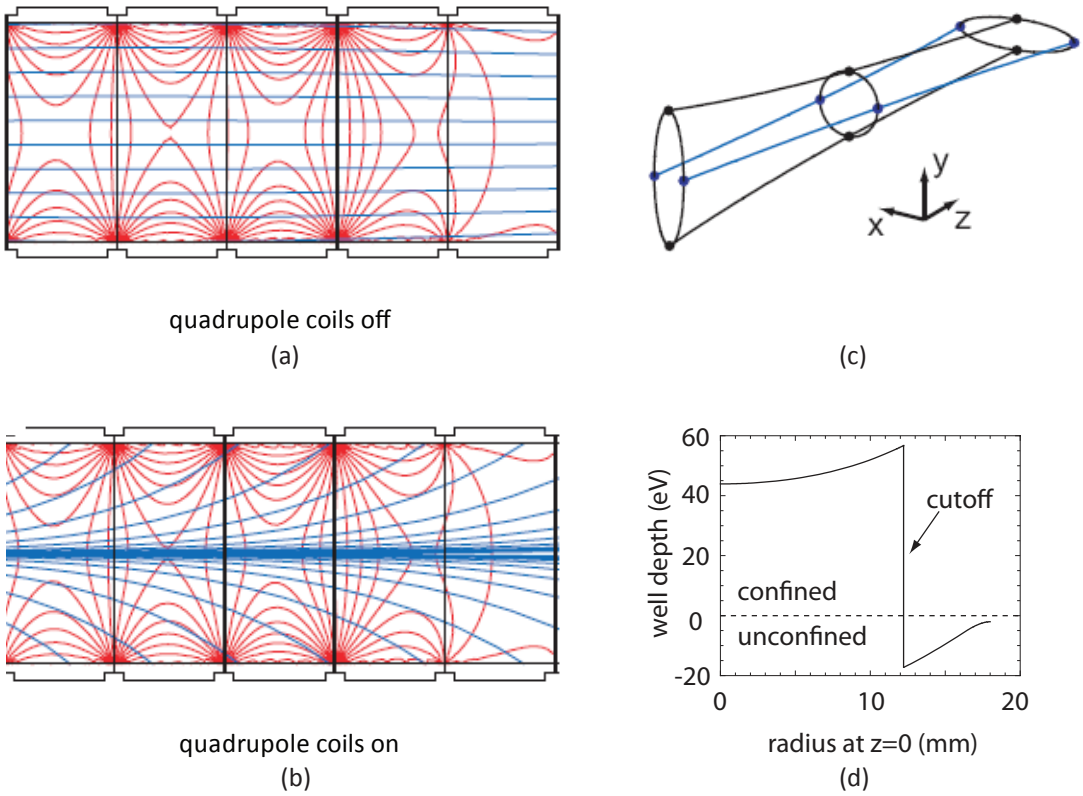


Figure 2.10: Magnetic field lines through the electrode stack in blue with (a) just the 1 T Penning trap bias field on and (b) the quadrupole field superimposed with the 1 T field. (c) Twisted bowtie shape of field lines in trap walls. (d) Cutoff radius for confinement. Adapted from [27]

2.4 Cryogen and vacuum spaces

Cold plasmas are necessary for antihydrogen trapping. If the trapped particles are cold, they have small oscillation amplitudes which make them less sensitive to trap imperfections. Additionally, three-body recombination rates for $\overline{\mathrm{H}}$ formation increase rapidly with decreasing temperature. Finally, trappable, low energy $\overline{\mathrm{H}}$ require constituents to be as cold as possible for confinement in a $\sim$ 500 mK trap. Because the cooling of our particles is limited by the temperature of the trap, antihydrogen trapping requires a cryogenic environment.

The apparatus is kept cold in several ways. First, it is isolated from room temperature by containment in an evacuated dewar, known as the "insert dewar," with 3 layers of shielding. The outer layer is at room temperature, and the inner two are held at 20 K and 60 K using pulse-tube refrigerators. The vacuum space in between the insert dewar and the apparatus is known as the insert dewar space (vacuum and cryogen spaces are labeled in Fig. 2.1).

The next level of cooling is via thermal conduction with a liquid helium reservoir (48 L capacity for CTRAP, 44 L for BTRAP) that is filled with liquid helium on a daily basis. Open pipes from the bottom of this helium dewar (Sec. 3.4) allow liquid helium to flow from the dewar into either helium enclosures housing superconducting magnet coils or into the "1 K pot"–a pumped refrigeration system that uses superfluid helium to keep the electrode stack at 1.2 K (Sec. 3.1). All containers and pipes filled with liquid helium is known as the helium space.

One of the biggest challenges for antimatter experiments is that antimatter annihilates with its matter environment. Thus we require an extremely high vacuum (XHV) to contain the antiparticles. The antiparticles are held and probed within the electrode stack, and the XHV vacuum space is between the bottom of the stack and a window plate located a few inches above the top of the stack (Fig. 2.1). This vacuum space is known as the experiment space. The pressure in this region has been previously measured by holding a known number of $\overline{\mathrm{p}}$ for 15.3 hours and observing $<3\%$ loss, corresponding to a background gas pressure of less than 6×10^{-17} Torr. This exceptionally good vacuum is remarkable given that there is a hole in the window plate that allows access for positrons into the Penning trap, and that this space is coupled to a room temperature, higher-pressure positron apparatus upstream. The space above the window plate is known as the positron space, and this UHV vacuum is only 10^{-7}-10^{-8} Torr, as measured outside the cryostat close to the hat.

The experiment space vacuum is upheld by cryopumping from activated charcoal surrounding the 1.5 mm positron hole. The hole is also 20 mm long which provides a good pumping restriction. This vacuum with its upper limit of 6×10^{-17} Torr is the best-measured vacuum for any system that also contains room temperature components and is comparable to the record low vacuum [30] of less than 5×10^{-17} Torr, which was obtained in the all-cryogenic precursor $\overline{\mathrm{p}}$ apparatus [4] to the current one.

2.5 Annihilation detectors

To detect and count the $\overline{\mathrm{p}}$, $\overline{\mathrm{H}}$, and e^+, we use a detection system built by our Jülich collaborators. It consists of 784 flexible scintillating plastic fibers immediately surrounding the insert dewar over the Ioffe trap region and 24 plastic scintillating paddles immediately surrounding the 1 T magnet bore (Fig. 2.11). Antiprotons are detected destructively by releasing them from the Penning trap either radially or axially and letting them annihilate with protons and neutrons in the trap walls or degrader. The annihilation produces on average 3.0 ± 0.2 charged pions per $\overline{\mathrm{p}}$ (as well as 2 neutral pions) [31], and these charged pions trigger the detection system. Since the pions have mean kinetic energies of 190 MeV, greater than their rest energy of 140 MeV, they are minimum-ionizing particles (MIPs) and undergo minimum energy loss upon interaction with the scintillators, which measure this energy loss. $\overline{\mathrm{H}}$ is measured via $\overline{\mathrm{p}}$ annihilations on the trap walls when we turn off the Ioffe trap radially.

The detector is in principle only sensitive to charged particles, but a neutral particle can be seen if it produces a charged particle in an interaction within the detector. Therefore the 511 keV gammas resulting from positron annihilation on the trap walls can also be detected with the fibers at low efficiency ($\sim$1/200). The detectors are also unfortunately

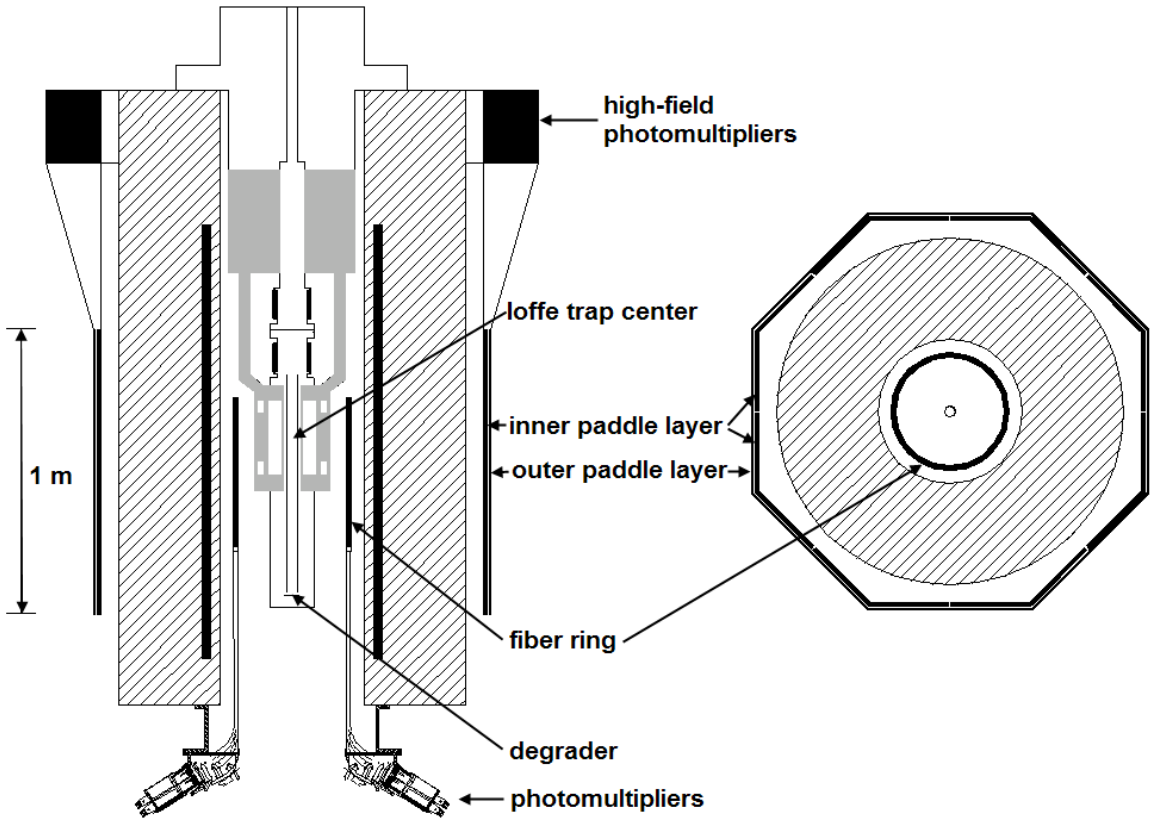


Figure 2.11: Detector Schematic.

sensitive to charged cosmic ray background particles at hundreds of Hz, primarily made up at ground level of (charged) muons with mean kinetic energy of 4 GeV [32] (compared with a rest energy of 108 MeV). The muons are also MIPs, so the distinction between $\overline{\mathrm{p}}$ and cosmic events measured on the detector is not obvious.

The 3.8 mm diameter, 523 mm long fibers (BICRON BCF-12) are arranged in a 4-layer ring, shown in 2.12. The inner 2 layers with 224 fibers each are oriented vertically straight and are shifted from each other azimuthally by half a fiber in order to avoid spatial gaps in detection. The 2 outer layers are helical and span 158°. These layers are also displaced from each other by half a fiber, and each layer has 168 fibers. The displacement between the helical fibers and straight fibers is discussed later in this section. The combination of vertical and helical fibers provide a coordinate system for detector hits. Before the Ioffe

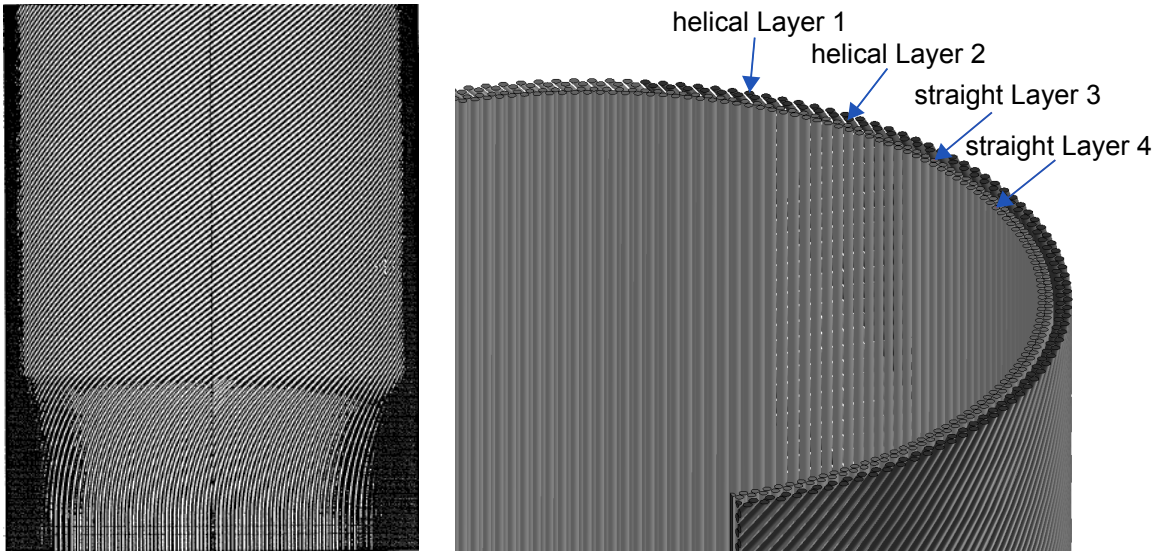


Figure 2.12: (a) Outermost helical fiber layer that straightens out towards the bottom. (b) Arrangement of inner straight and outer helical fibers.

traps were installed there were 3 sets of such fiber rings but the 2 inner sets were removed to accommodate the girth of the Ioffe magnet. These rings were meant to allow reconstruction of the antiproton annihilation vertex. Even though such reconstruction capability is no longer available, trapped antihydrogen can still be detected.

The 1 m-tall rectangular paddles (BICRON BC404) are arranged in a double layer surrounding the 1 T magnet. A layer of 16 paddles is surrounded by a layer of 8 larger paddles, with a separation distance of 7 mm. These layers are operated in 40 ns coincidence. That is, an inner paddle hit only counts if its adjacent outer paddle is also hit within 40 ns, in order to suppress noise and background hits. This inner-outer pair hit is defined to be a single paddle unit hit. Unless otherwise stated, a "paddle hit" actually refers to a paddle unit hit.

There are two types of data acquisition systems. A fast one uses a LeCroy scaler module to record scaler counts (Table 2.2) for combinations of detector hits that are read-

out and displayed constantly. These are coincidences that are measured via hardware. The read-out rate is normally set to 1 Hz, though the bin size can be as small as 100 ms. Four of the 12 scalers are fed into a multi-channel scaler (MCS) unit (2.2) which stores the most recent values after being triggered by either an $\overline{\mathrm{H}}$ or $\overline{\mathrm{p}}$ dump. The time resolution is up to 100 ns, and the fast scaler detection system only records hit rates.

An interactive slow detection system is used during $\overline{\mathrm{H}}$ trapping experiments and records events at 1000 Hz (with less than 20% dead time). It stores detailed event information such as which scintillators got hit, at what time, and with what energy, and also records all of the scaler values for that event. An event is only recorded if it satisfies the trigger condition of at least 1 inner paddle hit (in coincidence with a the adjacent large paddle) and at least 2 fiber hits. For this particular trigger condition (Scaler 9), the efficiency is 54%. Efficiencies for all the scaler values were determined from Monte Carlo studies by our Jülich collaborators.

Data acquisition

When a charged particle hits a detector, photons emitted from the scintillating material are guided into a photomultiplier (Hamamatsu H6568 for the fibers and Hamamatsu R2238 for paddles). There are a total of 14 photomultipliers (PMs) per fiber layer, each with 16 channels which get individual fiber inputs. The straight layers have 16 fibers per PM, and the helical layers have 12 fibers per PM. Each of the paddles has its own PM. The output of each PM channel is a current signal created from its photocathode.

The 12 or 16 analog fiber signals from each PM are input to a 16-channel RAL module which contains a discriminator (threshold 470 keV) for each channel. Among the outputs of the RAL module are the discriminator outputs for individual fibers (0s and 1s corresponding to a hit or not), which are saved in the data file generated from the slow

Scaler #	**Value (hit rate)**
0	fiber Layer 4
1	fiber Layer 3
2	fiber Layer 2
3	fiber Layer 1
4	# of fiber layers with $\mu \geq 1$
5	# of fiber layers with $\mu \geq 2$
6	start signal for quench
7	paddle $\mu \geq 1$
8	paddle $\mu \geq 2$
9	Scaler 7 AND Scaler 5
10	Scaler 8 AND Scaler 5
11	Scaler 8 AND Scaler 4
MCS Channel #	**Value (hit rate)**
1	Scaler 7
2	Scaler 4
3	Scaler 9
4	Scaler 10

Table 2.2: Scaler trigger values. μ is multiplicity.

detection system. It also gives the sum of the discriminator outputs, as a multiple of the 50 mV signal amplitude per fiber hit. These multiplicity signals from all the RAL modules in a fiber layer are summed up and given to another discriminator which outputs the multiplicity as a number of fibers. The multiplicity numbers are saved in the slow detection data file for each fiber layer, and become the first four scaler values shown in Table 2.2. To generate the

scaler values that are logical combinations (e.g. the last three scalers), the discriminated signals with 40 ns widths are input to coincidence modules. The widths of the discriminated signals were set to account for jitter in the electronics. All scalers are saved in the slow detection data file.

The RAL unit also outputs a sum signal of all the analog PM channel signals. This sum signal (from either a paddle PM or fiber RAL) flows into the input line of a charge-integrating ADC module (LeCroy FERA module) for 300 ns (much longer than it takes the MIPs to generate the charge) after the trigger. A capacitor charges up during this time, and at the end of the collection period the capacitor is discharged at a fixed rate while the discharge time is measured by a high-frequency clock. The number of clock oscillations is proportional to the amount of charge on the capacitor, and the time lapse is output as an ADC channel number with a resolution of 0.25 pC. This digitized signal represents the total energy loss of all the particles that hit detectors coupled to the corresponding PM. The ADC channel number ranges from 0 to 2047 and is recorded in a data file from the slow detection system.

The calibration between ADC channel number and energy deposited by the MIP has been done for every PM in the detector system using the known muon energy loss spectra obtained from Monte Carlo simulations [33], and is linear. The peak of the ADC spectrum for muons is typically at Channel 180, corresponding to 2 MeV energy loss in the scintillators. ADC channel numbers of below 100 are most likely from noise which is present even when there is no light from the scintillator. These include signals from electrons which get thermally excited and freed from the photocathode or dynodes in the PM. Therefore, paddle hits with ADC values below this threshold are generally excluded from analysis.

TDC modules provide digitized timing information relative to the trigger for every photomultiplier. The TDCs consist of a combination of a TFC module and a FERA ADC

module. At the start signal (the event trigger) a current source in the TFC begins charging up a capacitor for a maximum of 300 ns and stops when there is a stop signal (ADC sum signal), converting time to charge. The time is counted by an ADC module in the process described above, and is output as a TDC channel number with 50 picosecond resolution. The TDC channel number also ranges from 0 to 2047, and is also stored in the slow detection data file.

The 300 ns acquisition time is more than sufficient to capture the ∼3 charged pion signals from each annihilation. After creation, the pions only take on the order of 1 ns to hit the detectors, triggering the data acquisition system and opening the ADC inputs. After a 25-50 ns delay corresponding to scintillator decay time and light collection, charge begins to flow in the ADC units. After the 300 ns charge-collecting period, all inputs are closed to further triggers. The buffered data is written to the raw data file and the system is prepared for the next event. The total cycle takes up to 1 ms (depending on user-set event rate). Typically the data acquisition is set to 1000 events/sec, which is sufficient because triggers happen less frequently than 1000 Hz. Since the frequency of both $\overline{\mathrm{p}}$ and cosmic hits are small compared to the acquisition frequency, we can assume that every event corresponds to either 1 cosmic signal or 1 $\overline{\mathrm{p}}$ annihilation signal without worrying that multiple events are being recorded as one event.

Detector positions

Since knowledge of particle trajectories is useful for distinguishing between $\overline{\mathrm{p}}$ and cosmic hits, it is important in post-analysis of an $\overline{\mathrm{H}}$-trapping trial to know which detectors were malfunctioning or turned off, as well as the relative positions of the detectors. A large data set consisting of only cosmic background events were used to do this, with cuts determining patterns of detector hits.

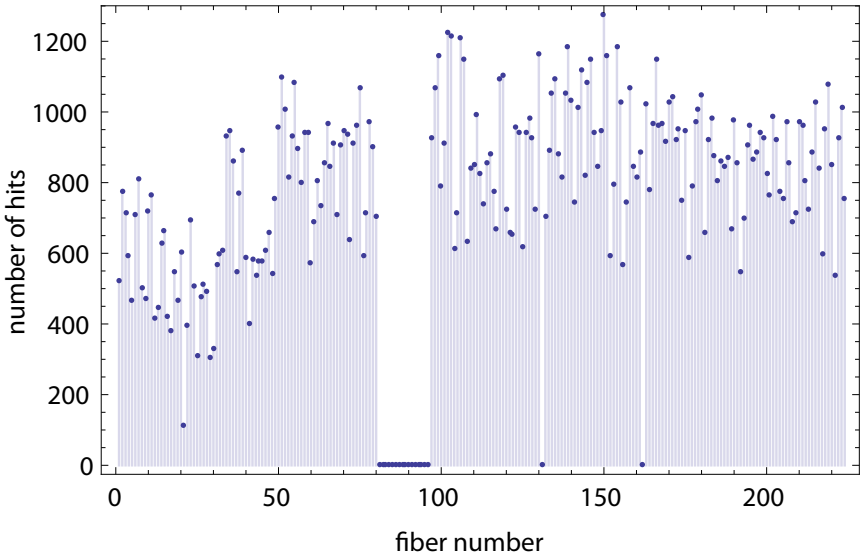


Figure 2.13: Distribution of fiber hits taken from cosmic background data for fiber Layer 3. Fibers 131 and 162 are malfunctioning, Fiber 21 has a lower-than-normal firing rate, and photomultiplier #6 is turned off.

Fibers that were malfunctioning or turned off were determined by looking at the distribution of fiber hits from the cosmic data, and observing which ones never fire or have a much lower rate of firing than their neighbors. Figure 2.13 shows an example of this for the straight fibers in Layer 3. In this example, Fibers 131 and 162 are malfunctioning, Fiber 21 has a lower-than-normal firing rate, and photomultiplier #6 is turned off. It is good practice to check for broken or turned-off fibers before analyzing data.

Fig. 2.14 shows how the rotation offset between any two fiber layers is determined, using the example of the two middle layers – helical Layer 2 and straight Layer 3. The criteria is that only 1 fiber in Layer 2 and 1 fiber in Layer 3 get hit, with no restrictions on other layers. This indicates which fibers are adjacent in the two layers. The event distributions for cosmics and $\overline{\mathrm{p}}$ are compared in Fig. 2.14a and b. Both plots show a wide band of $\sim$75 helical fibers that comes from the overlap of each straight fiber with a range of helical fibers.

The stray hits outside this band can come from noise or from multiple particles. An

antiproton annihilation creates multiple pions in an event that go off in different directions, resulting in wrongful combinations of hits. For this particular cut, it is possible for example that one particle in the event hits a straight fiber but misses a helical fiber, while another particle going in a different direction hits a helical fiber far away but misses a straight fiber. This can also happen for cosmics. Even though each event from a cosmics data set usually only contains one cosmic, it can also sometimes produce secondary electrons and photons going in different directions. A single cosmic can also cause stray hits by, for example, passing through a helical fiber, missing the overlapping straight fibers, and hitting the straight fiber on the other side of the fiber ring. This can also be seen in Fig. 2.15, which shows the straight fiber hit correlations. Pairs of adjacent straight fibers are expected to always be hit together, but there are many stray hits. The stray hits happen more frequently for $\overline{\mathrm{p}}$ than for cosmics.

The darker band with about $\sim$10 fiber width on the cosmics plots comes from the fact that the bottoms of the helical fibers must eventually turn into vertically straight sections in order to reach the photomultiplier tubes, as shown in Fig. 2.12a. The straight sections have more overlap with the straight fibers adjacent to it, so every straight fiber has a particular helical fiber or small range of helical fibers ($\sim$10) that will get more hits than the rest, giving direct relative position information. The sharper band is not seen in the $\overline{\mathrm{p}}$ distribution, because the $\overline{\mathrm{p}}$ annihilations are from radial losses that occur (by turning up the quadrupole field) near the axial center of the fiber ring, far from the straight section of the helical fibers. This results in a relatively more even distribution of hits over the 75 fiber band.

In order to isolate the data points just around the sharp band to do a linear fit (to determine relative fiber positions), only coordinates with repeated data points of repetition $>$6 were selected, the smallest repetition which got rid of outliers. The intercept represents

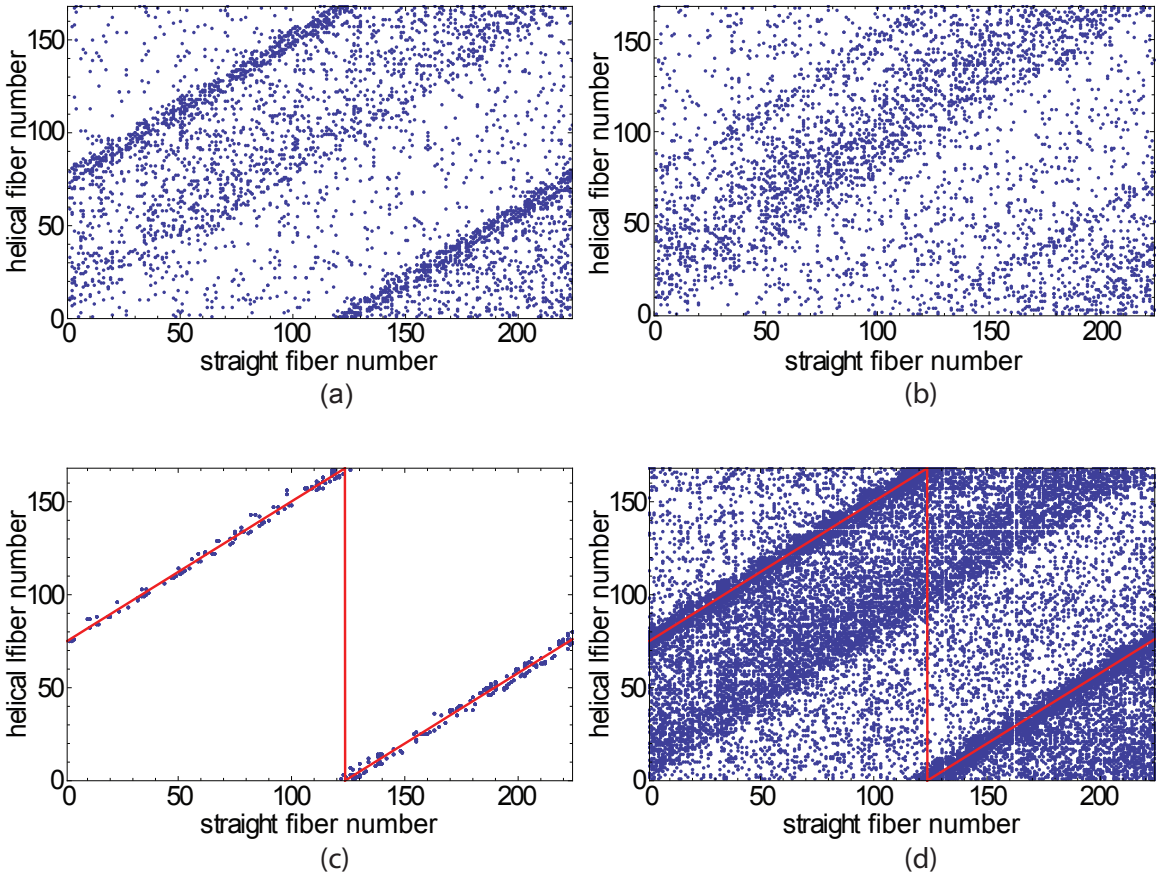


Figure 2.14: Helical (Layer 2) and straight (Layer 3) fiber relative positions. Events were selected for which there was only 1 Layer 2 hit and 1 Layer 3 hit, with no restrictions on hits in other detectors. (a) Cosmics and (b) radial $\overline{\mathrm{p}}$ data. Both data sets each contained a total of 43 000 events, of which 4 000 $\overline{\mathrm{p}}$ hits satisfied the cut while 5 000 cosmics satisfied the cut. 4 000 points for each are shown plotted, and a difference between the hit patterns is observed. (c) Cosmics data with repetition >6 events selected, in order to facilitate a linear fit. The fit gives helical = Mod(a+b×straight, 168), with $a = 74.348 \pm 1.1$ and $b = 0.758$. The intercept gives the offset between the straight and helical layers, and the error bar on the intercept comes from the deviation between intercepts when the slope is fixed to be the expected 0.75, and also fit error. (d) Fit obtained from (c) plotted on a larger cosmics data set (22 000 events) with the same cuts criteria.

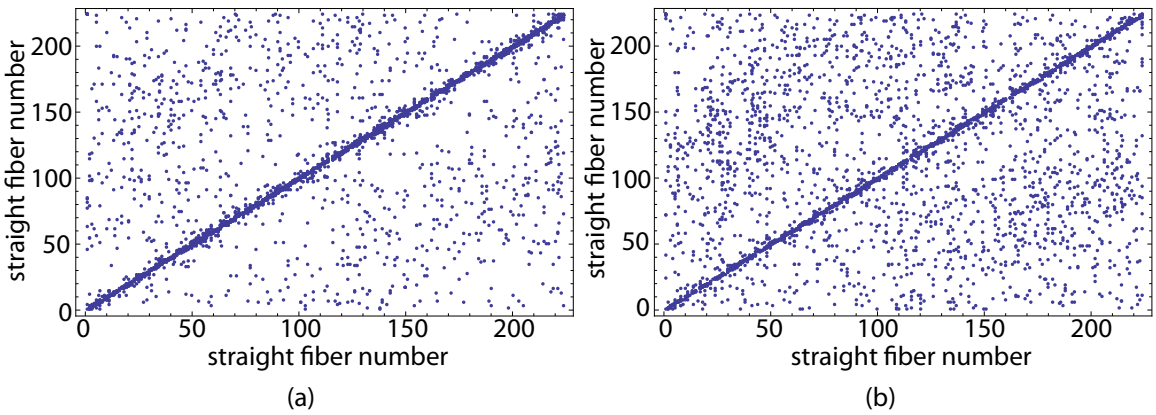


Figure 2.15: Straight (Layer 3) and straight (Layer 4) fiber relative positions. Events were selected for which there was only 1 Layer 3 hit and 1 Layer 4 hit, with no restrictions on hits in other detectors. (a) Cosmics and (b) radial $\overline{\mathrm{p}}$ data. 3 700 points for each are shown plotted.

the offset between helical and straight fiber numbers. The fit gives an offset of 74 fibers, but a slope of 0.76 instead of the expected 168/224=0.75. A fit was also done forcing the slope to be 0.75, leading to the 1 fiber error bar on the offset (along with fit error). This method of noise-reduction by rejecting low-repetition points has the disadvantage that too many events were discarded, leading to gaps in the data and many straight fibers unrepresented. A conceptually similar but more accurate way is to take the peak of the helical fiber distribution for each straight fiber and fit the peaks as a linear function of straight fiber. Alternatively, algorithms from multivariate statistics, notably the RANSAC [34] algorithm, could be used to robustly fit the data in the presence of the high volume of outliers. The latter two techniques have not been tried.

A similar procedure was used to determine relative positions of paddles and fibers. A cut was made in which there was 1 hit in an outer paddle and 1 hit in a Layer 3 straight fiber, with no restrictions on hits in other detectors. Additionally, only paddles with ADC values above 100 were chosen to filter out the noise, as described earlier. Because cosmic muons enter the apparatus mostly vertically with an angular distribution that goes as $\cos^2\theta$,

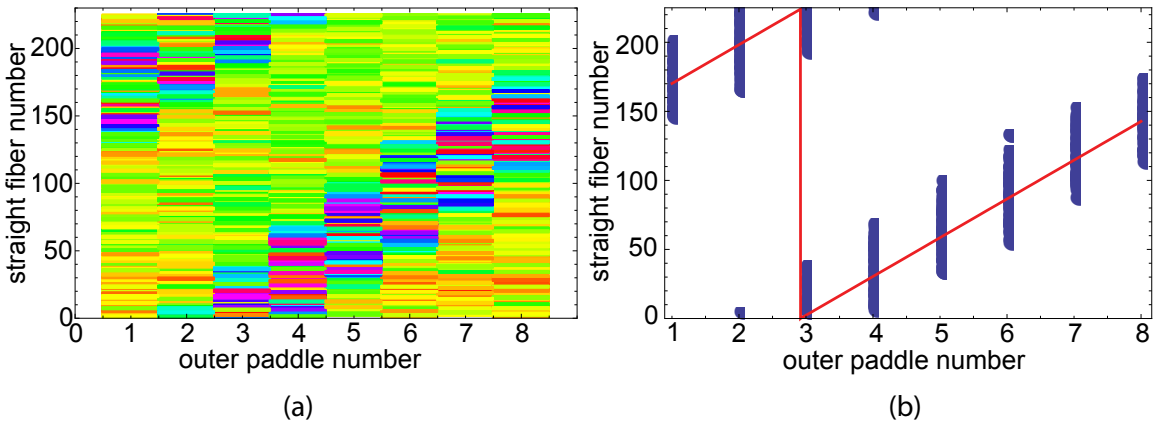


Figure 2.16: (a) Straight fibers (Layer 3) and outer paddle relative positions. Events were selected from a cosmics data set in which there was only 1 outer paddle and 1 Layer 3 hit, with no restrictions on hits in other detectors. (a) Blue represents high repetition. (b) Data from (a) with only repetition μ>50 events selected, in order to facilitate a linear fit. The fit gives fiber = Mod(a+b×paddle, 224), with $a = 142.09 \pm 0.44$ and $b = 28.08$. This equation gives the position of the fiber closest to the paddle. The error bar on the intercept comes from the deviation in intercept when the slope is fixed to be the expected 28, and also fit error.

the fibers that get hit can be assumed to be close to the paddle. The results are shown in Fig. 2.16. In Fig. 2.16b, events with repetition >50 were selected from the data in Fig. 2.16a in order to do the linear fit. There was no shortage of events for this cut, and the resulting slope was the expected 224/8=28. This line gives the fiber that is centered on each outer paddle.

This fiber-paddle correlation analysis is sufficient to determine the position of every fiber relative to a paddle. However, if more fine-tuning is needed, an additional cut could be imposed in which only 2 adjacent inner paddles get hit, and 1 fiber gets hit. The center of the range of fibers would correspond to the inner scintillator edges, giving another 8 paddle reference points. The equations presented in this section give relative positions of every detector in the detection system.

Chapter 3

CTRAP Construction Projects

This chapter describes a selected set of design and implementation projects for CTRAP, including components that are improved relative to their BTRAP counterparts.

3.1 1 K pot

Pumped refrigeration system

Producing $\overline{\mathrm{H}}$ that can be confined in our $\sim$0.5 K magnetic trap requires a thermal distribution of charged particles that is as cold as possible. Though synchrotron damping of electrons and positrons is a very effective cooling mechanism (Sec. 2.2.1), their equilibrium temperature is limited by radiative heating from the electrode stack, a significant effect due to the plasmas' small heat capacity. To minimize the electrode stack temperature, we use superfluid helium-4 to cool the stack to 1.2 K [35]. The superfluid helium is made and contained in a soda can-sized "1 K pot," shown in Fig. 3.1, by decreasing the pressure in the pot until a phase transition to superfluidity occurs.

The 4.2 K liquid helium from the helium reservoir gets to the 1 K pot via a needle valve and then a high-impedance capillary system. The needle valve's opening is controlled

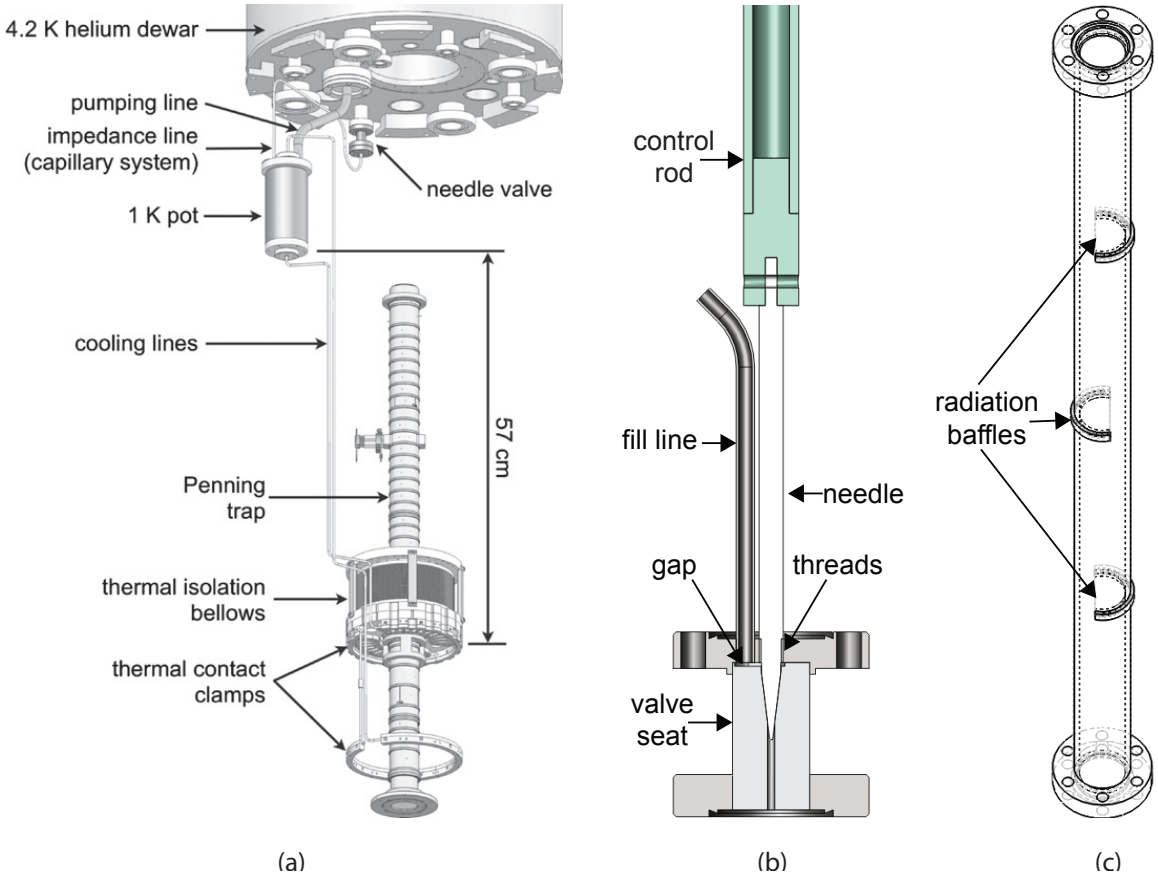


Figure 3.1: (a) Full BTRAP 1 K pot assembly, taken from [35]. The CTRAP assembly is nearly identical. (b) Close-up of the CTRAP needle valve. Control rod and fill line are inside the He dewar, directly submerged in LHe. (c) 4.2 K upper section of the CTRAP pumping line, with radiation baffles. This section of the line extends through the helium dewar, but is vacuum-isolated from LHe.

at the hat by turning a control rod–a G10 tube (for CTRAP) which minimizes the heat load from the hat. The high impedance line is made from parallel capillaries with tens of micron ID. These thermally isolate the 1.2 K pot from the 4.2 K reservoir and maintain a pressure differential between the pot and the reservoir. The pot is pumped down to 0.4 Torr (measured near the pump) using an Edwards GVSP30 scroll pump. The decrease in pressure decreases the helium temperature to 1.2 K, creating superfluid helium, and a copper sleeve in the interior of the pot prevents thermal gradients during pumpdown. The highly thermally-conductive superfluid helium then circulates through copper cooling tubes to the

electrode stack via thermal clamps on the electrode pinbases (bottom pinbase hidden in the figure) and back to the 1 K pot. The electrode stack is in good thermal contact only with with the 1 K cooling lines and pinbases, despite it being bolted to and otherwise surrounded by a 4 K environment from 2 superconducting magnets only a few millimeters away from it. The 1 K electrode stack is suspended but thermally isolated from from the 4 K Ioffe trap (not shown) via titanium edge-welded thermal isolation bellows (whose ends are rigidly fixed by G10 supports to prevent stretching).

Fig. 3.1b shows a cross-section of the custom-made needle valve for CTRAP. When the control rod is turned by hand at the hat, it rotates a threaded titanium needle into or out of a tapped hole in the top titanium conflat. The valve is closed when the needle presses into the valve seat. The seat for the BTRAP 1 K pot is made of silver, a softer metal than titanium, in order to make a good seal when the valve is closed. However, because of the challenges associated with silver-titanium welding, discussed in Sec. 3.4, we opted to make the CTRAP needle seat from titanium since a leak-tight seal is not necessary for normal operation. To make the fit as good as possible, the 7° taper of the needle was machined together with the seat without changing the tooling, to preserve the compound angle. To test the seal between the titanium needle and seat, a background helium leak check was done with the fill tube blocked and the needle screwed in tightly. The background leak rate was found to be 5×10^{-5} mbar-L/s. This is a factor of 40 times smaller than the background rate of the impedance line, indicating that the seal is good enough.

When the valve is open, 4.2 K helium filtered with 3 micron silver filter paper (AG30SP4 from Sterlitech Inc., not shown) glued to the top of the titanium fill line is allowed to flow through a small gap before spilling into the valve seat. The helium is filtered once more, necessary to prevent clogging, before it passes through the high-impedance capillary system. When the needle valve is closed, the flow rate of helium into the 1 K pot is

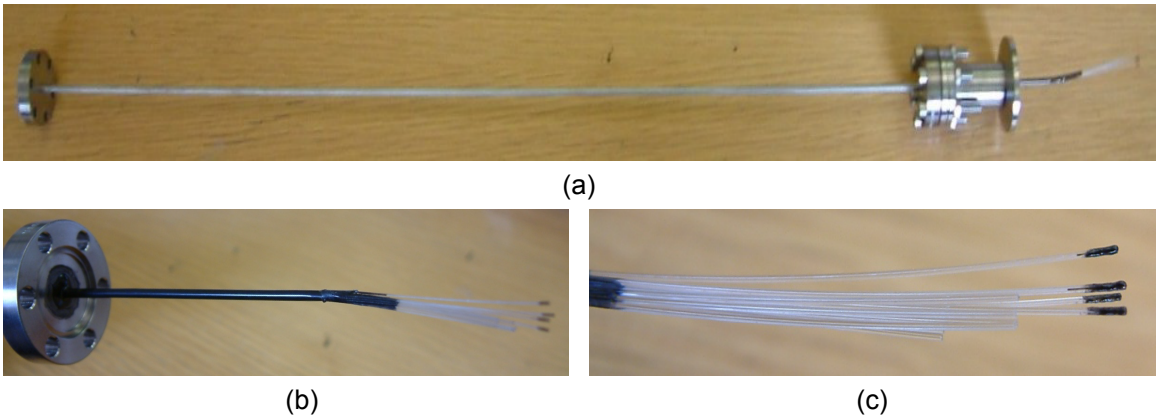


Figure 3.2: (a) Full CTRAP impedance line. (b) and (c) Close-up of the capillary ends. Although only 4 capillaries are shown blocked, a 5th was blocked later on.

limited by the leak between the needle and the seat. When the needle is all the way open, the flow rate is limited by the the impedance of the capillary system. Fig. 3.2 shows pictures of the capillary system.

Capillary system design and construction

The CTRAP capillary bundle consists of 12 pieces of 20 cm-length capillaries (Paradigm Optics CTPC050-360-1: 50 micron ID, 360 micron OD polycarbonate tubes) held together in a bunch with Stycast 2850 epoxy. Cutting and gluing the capillaries was a delicate procedure, since it was very easy for the capillaries to get clogged or deformed during the cutting. After some practice the cleanest cuts were obtained by using a fresh scalpel heated to 240°C to make quick slices of 4 capillaries at a time in a clean environment. The cuts were examined with a 10× loupe magnifier to make sure the cross sections were round and that the interiors were clear, and then immediately covered. The gluing was done by holding the capillaries vertically in a 14 cm long teflon tube (TTI-S12-1100-NAT from Newark) and dripping Stycast 2850 epoxy through the teflon tube around the capillaries. Problems with the epoxy taking too long to wick down the full length of the tube before

it started hardening were remedied by increasing the diameter of the teflon tube to AWG 12 and decreasing the viscosity of the epoxy (to a catalyst/resin ratio of 10% rather than the specified 7%). Epoxy that was too runny would flow too quickly and cause undesirable air bubbles to form. After the epoxy cured, the teflon tube was removed from the epoxy stick, as epoxy bonds very poorly with teflon. The rigid capillary-epoxy structure was then epoxied into a titanium tube, making the leak-tight seal shown in Fig. 3.2b.

The design goal for CTRAP's capillary system was to match its impedance and leak rate with those of the BTRAP system since they worked well. BTRAP's capillaries had a total laminar flow impedance of $1.57\times\ 10^{11}$ cm^{-3} (typical for 1 K pot systems) due to 4 tubes of 20 cm length and 60 micron ID (300 micron OD). This particular capillary diameter was no longer available, so the CTRAP capillary line was made with 50 micron ID tubes. Because the impedance scales as L/D^4, the decreased diameter necessitated about twice as many capillaries for the same impedance. In addition, five extra capillaries were included and intentionally blocked off with epoxy (Fig. 3.2c). These can be used as spares in case any of the other capillaries become blocked in an inaccessible place inside the tube, or for modification of the total impedance later on.

When the CTRAP capillary system was completed, its impedance was tested with a leak check. A 1 atm gaseous helium pressure differential was applied across the impedance line by inflating a punctured plastic bag taped to one end with gaseous helium, and pumping on the other end with a leak checker. The leak rate was the same as BTRAP's: 2×10^{-2} mbar-L/s. Since clogged capillaries can happen all too easily, a full spare impedance line was also made.

The 1 K pot has been tested on CTRAP without the electrode stack, $\overline{\text{p}}$-solenoid, or Ioffe trap. It was found to work properly, at least without the heat loads from the aforementioned components. After the 1 K pot was filled with 4 K helium (indicating that the

capillaries were not blocked), the needle valve was closed and the pump was turned on. This almost immediately caused the pressure to drop and the temperature to go from 4 K to 1 K. Fig. 3.3 shows that when the valve was closed again all the way while the pot was still being pumped on, the temperature increased but the 1 K pot did not go back to 4 K, indicating that there was still superfluid helium in the pot. This is presumably because in this setup there wasn't a heat load from the electrode stack to boil off the 1 K helium.

3.2 Filter boards and electrode stack wiring

Electrical signals to the electrode stack

The Penning trap electrodes have a variety of geometries and circuits that serve different purposes. The high-voltage electrodes (HV and DEG) in the lower electrode stack (Fig. 2.5), for example, are used to catch antiprotons, and pulses sent to LTE2 are used to eject electrons from the $\overline{\mathrm{p}}$ plasma (Sec. 5.1.2). All the electrodes can receive DC voltages and most can also receive either slow or pulsed RF voltages. All signals travel through feedthroughs at the hat, through the insert dewar vacuum space to the electrode stack pinbase flanges, which are located in the bore of the $\overline{\mathrm{p}}$-solenoid. A schematic and photograph of the pinbases are shown in Fig. 2.5 and 3.4. The pinbases contain leak-tight electrical feedthroughs which enter the experiment vacuum space containing the electrodes. Electrical connections to the electrodes are made inside the experiment space via these feedthrough pins.

Limiting noise to the electrodes is crucial, since noise can be detrimental to particle and plasma stability, behavior and detection, and can be a source of heating. To alleviate noise, just before signals reach the electrodes, they are filtered at the pinbase. A printed circuit board stuffed with surface-mount filter components contains tinned vias at the filter

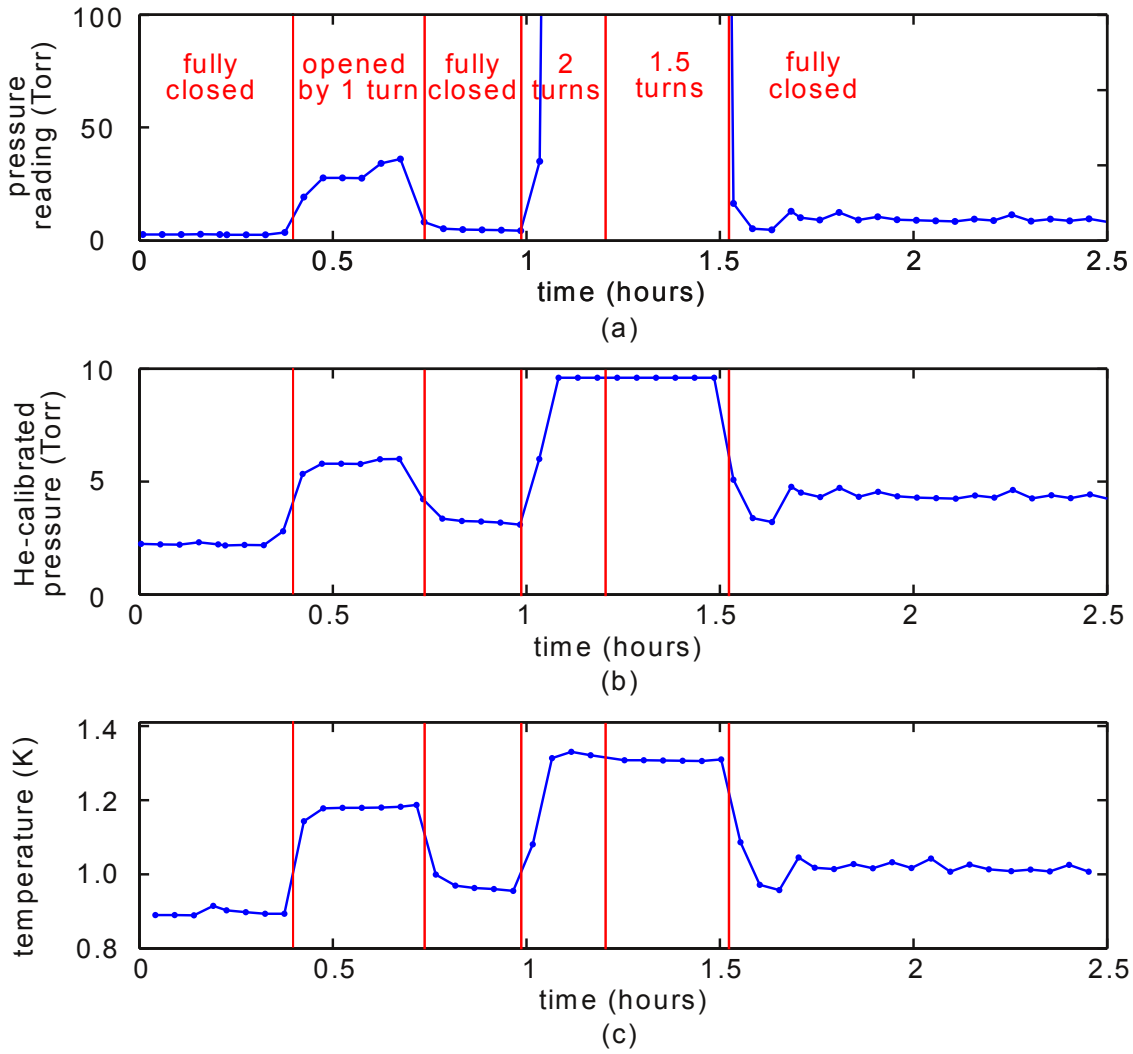


Figure 3.3: (a) 1 K pot pressure as the needle valve is opened and closed. These are the raw readings from a convectron guage calibrated for nitrogen rather than helium. (b) Calculated 1 K pot helium pressure from He-N_2 calibration. (c) 1 K pot temperature as the needle valve is opened and closed.

outputs that are soldered directly to the feedthrough pins of the pinbase. A 1 cm gap is left between the filter boards and the pinbases for screwdriver access to an indium vacuum seal connection separating the insert dewar and experiment space. Depending on how the board is stuffed, the board also accommodates other circuits for specialized electrodes. The board is grounded through grounding straps soldered to similar vias and connected to the experiment

ground. These filter boards, the 19 circuits printed on each board, and connections to the pinbases on CTRAP are shown in Figs. 3.4 and 3.5.

The circuits for each electrode and corresponding pinbase pins are shown in Figs. 3.6 – 3.8. There are three filter boards - two for the "upper stack" (all electrodes including the LHVG electrode and above), and the one for the "lower stack" (below the LHVG electrode). The electrodes and their wiring designations are arranged in the figures in order from the top of the stack downwards.

Most of the RF signals are pulses from DEI high voltage pulsers with 50 Ω output impedance. The pulses typically have a rise time of 6-10 ns and widths of 1-10 μs. These signals travel through stainless steel micro-coax (0.034" OD) cables (labeled "fast lines" in the schematic) which connect to the boards via SMA connectors. The signals are then 50-Ω terminated at the board and filtered with a 1000 pF DC blocking capacitor, the value of which was observed to give the best pulse shape. "Slow" RF signals (e.g. sine waves) are sent to the electrodes through twisted pairs (signal and ground) of constantan wire, which get jumpered to twisted copper wires which are directly soldered into the filter boards (see Fig. 3.4). This is to avoid additional heat loads from coax lines and power dissipation through 50 Ω terminators. These RF signals are also DC-blocked with blocking capacitors. The blocking capacitors block both DC noise on the AC lines, as well as DC signals from the DC bias lines (to prevent interference).

The low-voltage DC signals come from high-precision voltage amplifiers called "Uber-Elvis" with an output impedance of 404 kΩ and output voltages of up to ±1 kV. The signals are carried from the hat through twisted pair constantan wire bundles, jumpered to connectorized copper wires which plug into the J1 and J2 connectors on the filter board. The 1 MΩ resistor prevents interference from RF signals, analogous to the blocking capacitor in the RF circuits. The 1 MΩ resistor together with the same 1000 pF blocking capacitor form

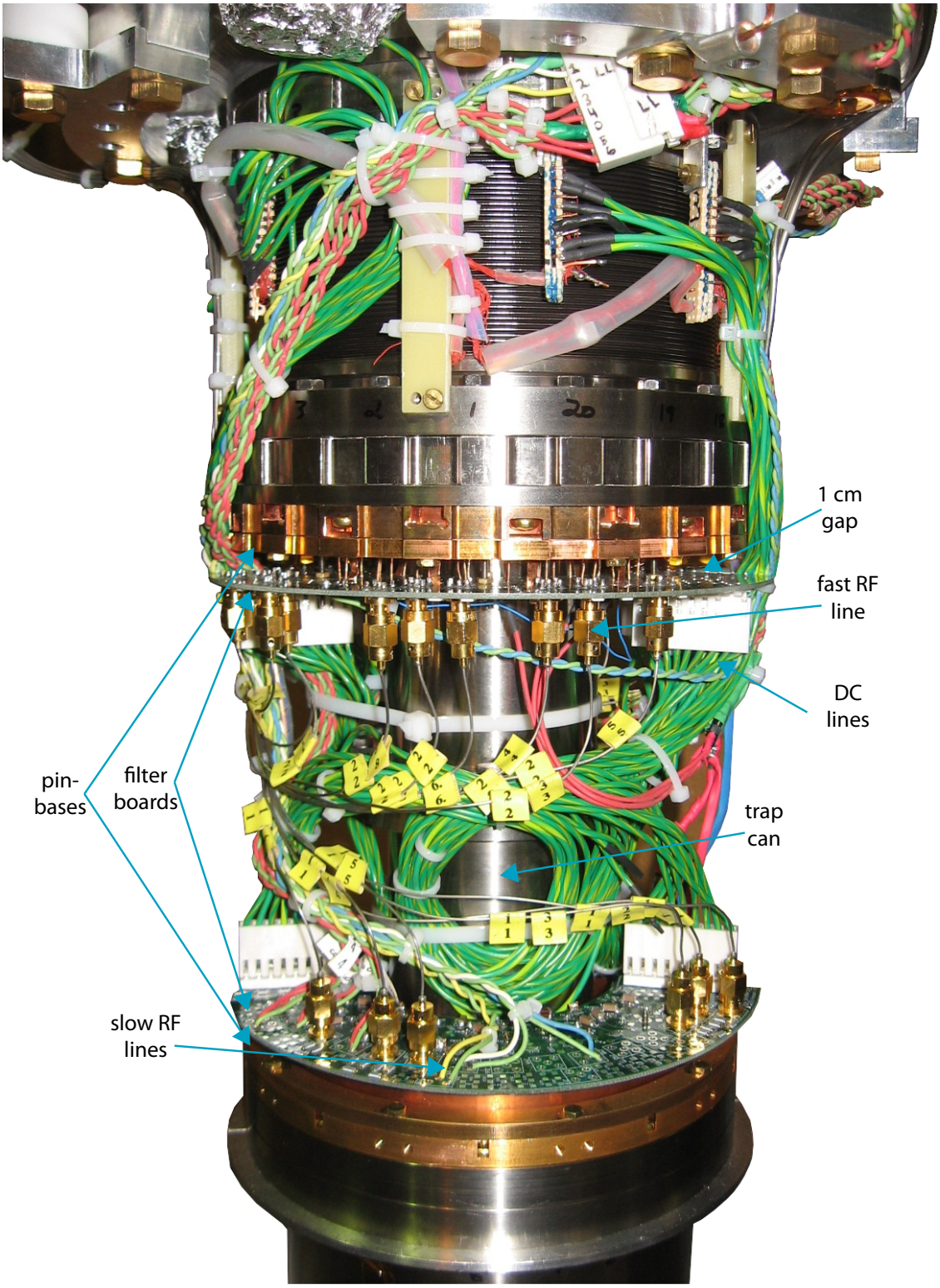
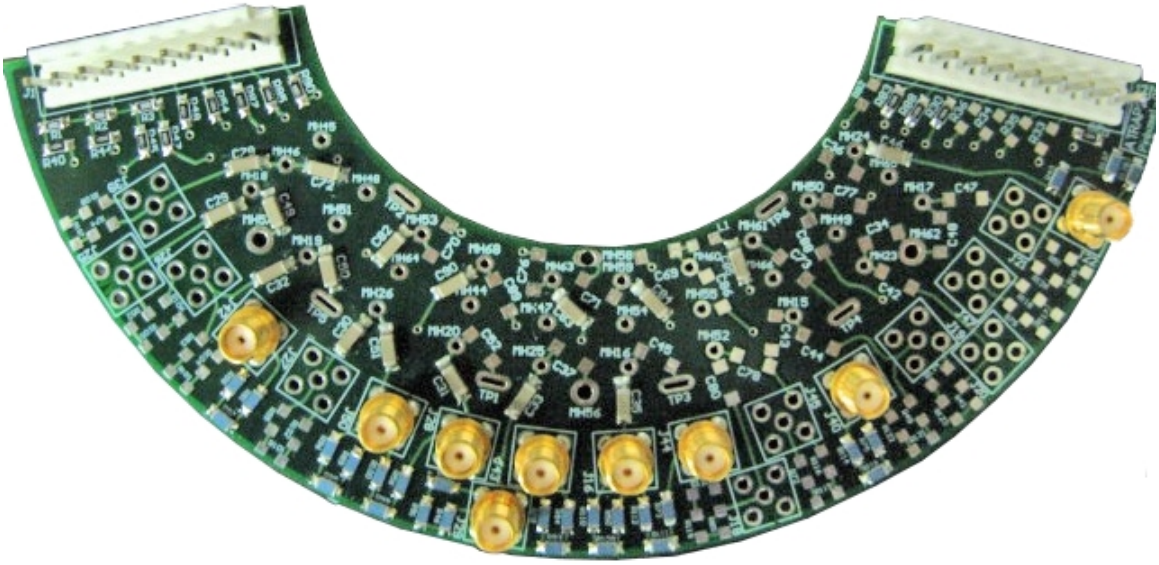


Figure 3.4: Photograph of CTRAP electrode wiring outside the trap can.

(a)



(b)

Figure 3.5: (a) CTRAP Filter board design showing the 3 of the 4 layers (ground plane not shown) (b) Photo of a final stuffed board

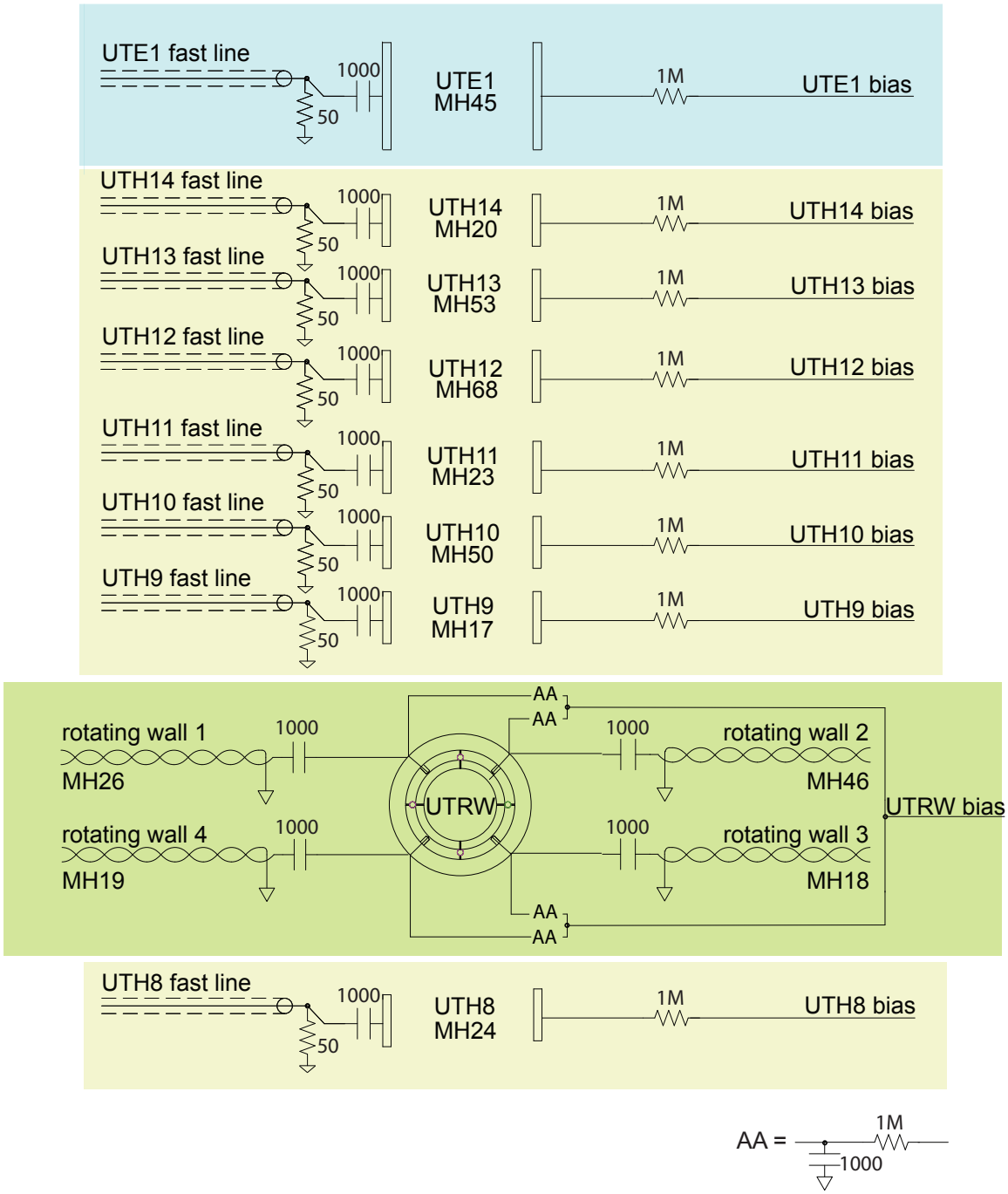


Figure 3.6: CTRAP upper stack wiring for Upper Left pinbase.

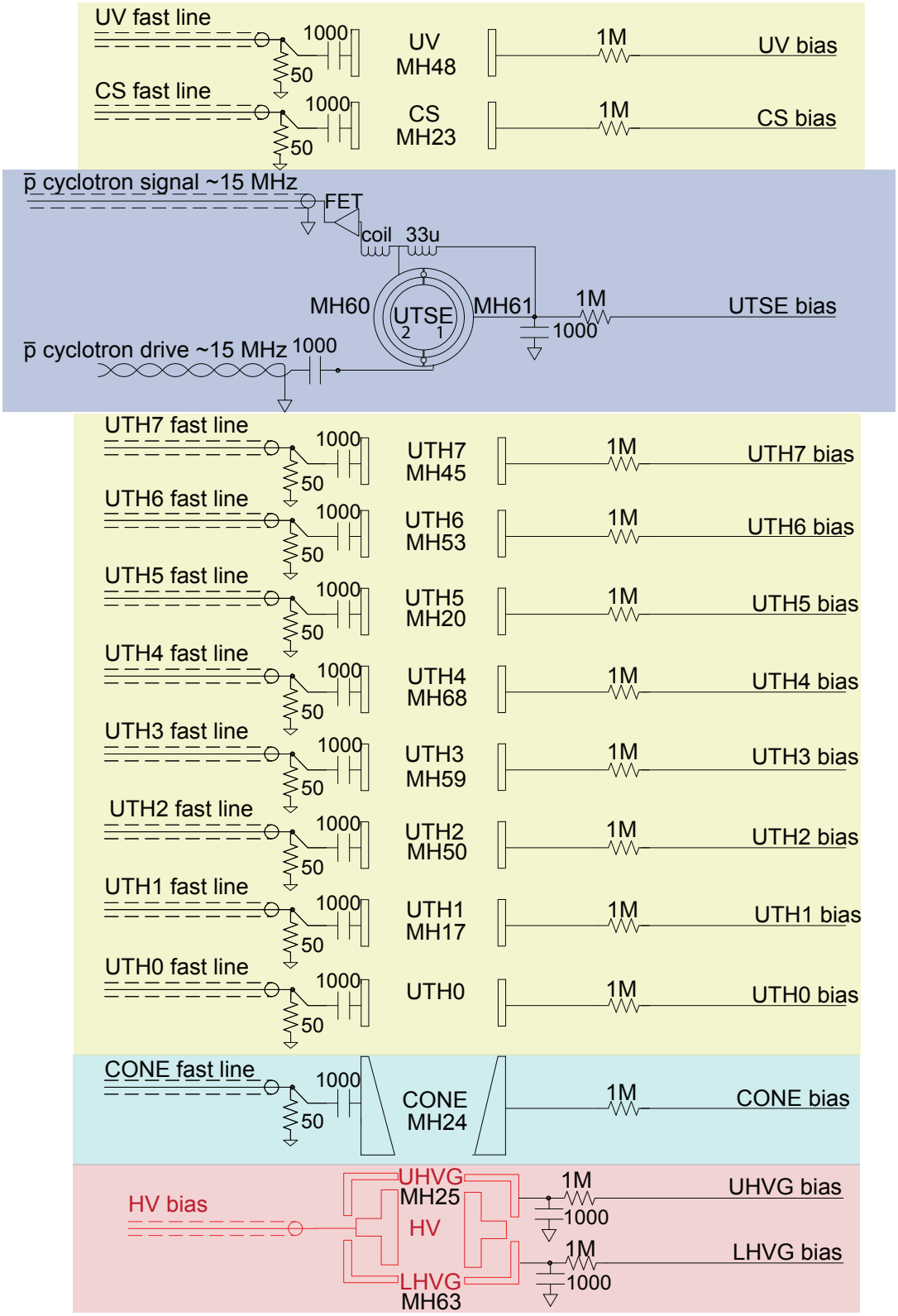


Figure 3.7: CTRAP Upper Stack wiring for Upper Right pinbase.

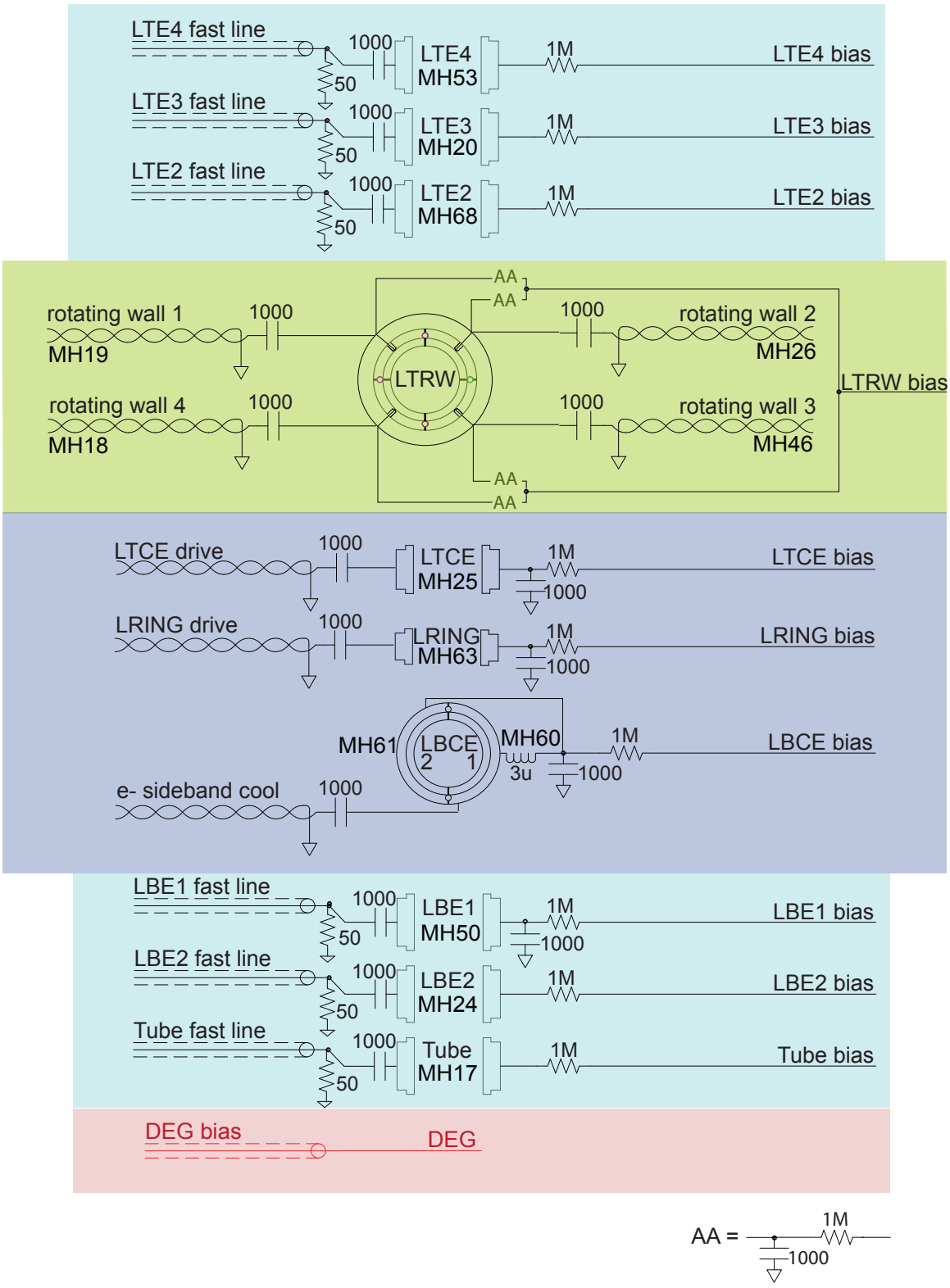


Figure 3.8: CTRAP Lower Stack wiring for Lower Left pinbase.

a low-pass filter for the DC lines with a cutoff frequency of 166 Hz.

Nearly all the electrodes are capable of being ramped, holding the particles in near-harmonic wells, forming arbitrary multi-well potentials, and moving the particles up and down the stack. There are also some specialized electrodes. The rotating walls UTRW and LTRW are split lengthwise into 4 quadrants. Each conductor has its own input RF line that receives a 9 Vp-p, 100 kHz–10 MHz sine wave generated from a Harvard-built multichannel synthesizer (incorporating digital synthesizers Analog AD9954). The signals for adjacent quadrants are 90° out of phase. These electrodes are used to impart angular momentum to the charged plasmas, spinning them up to a desired density and transverse radius. There is only 1 DC bias line that splits 4 ways for each electrode segment, so that the rotating wall can act as a single electrode with a near-uniform potential when the rotating wall mechanism is not in use.

LRING and LTCE/LBCE are ring and compensation electrodes that can be used to make a very harmonic well. UTSE and LBCE are half-split electrodes for cyclotron drives and sideband cooling, respectively.

UHVG, HV, LHVG, and DEG only receive DC signals. All the electrodes except for the HV electrode and DEG (the degrader) can only receive up to 1 kV (limited by their wiring), but HV and DEG can receive up to -5 kV for antiproton-catching. These high voltages are passed through thicker (0.085" diameter) high-voltage stainless steel coax lines that are directly soldered to much larger copper feedthrough pins (rated for up to 20 kV) on the pinbase, rather than passing through the filter boards. The stainless steel signal wire had to be specially prepared with acid flux before the soldering was done, because it is difficult to heat stainless steel enough for solder to wet to it.

Filter board design

The general design of the CTRAP filter boards was based on that of the BTRAP boards, and most of the filter circuits for both traps are identical. One difference is the geometry of the CTRAP boards. This modification was made in order to provide better access to the aforementioned screws for the pinbase indium seals, which are located very close to the trap can, making screwdriver access difficult. Making these seals properly is critical for maintaining the experiment vacuum, which, as described in Sec. 2.4, is nearly perfect. With this in mind the CTRAP board was made with larger inner and outer radii to create more space inwards and all board components and traces were rerouted to accommodate this. One constraint in the new design was that the 1 K filter boards not touch the 4 K inner can of the pbar solenoid, so there is a minimum 0.062" radial gap between the two objects.

The new dimensions of the board required a complete topological redesign, and this opportunity was taken to make a few improvements. These include better power dissipation for high-voltage signals, increased distance between traces leading to less stray capacitance and chance of arcing, and shorter trace lengths to limit them from acting as antennas or causing the RF signals to reflect post-termination.

The BTRAP boards were not designed for high-voltage pulsing, and the lesson learned from burned-out BTRAP terminator resistors was to make the CTRAP board components able to handle higher voltage RF signals. The DEI pulser is capable of sourcing 800 V and 17 A. Since for certain frequencies the impedance of the 50 Ω resistor is much lower than that of the capacitor, the resistor could potentially draw a large amount of current, necessitating good power dissipation. In some circuits one way to get around this problem is to terminate with a DC-blocking capacitor in series with the resistor, but that does not help us here since we send fast pulses. Because the CTRAP board has a larger area, there was

enough space to divide up the 50 Ω terminator into three 16.9 Ω, 1/3 Watt, 200 V resistors, which gives higher power dissipation. The resistors were wired in series rather than parallel, in order to maximize the voltage rating. The 1 MΩ filter resistors can also receive up to a 1 kV from the Uber-Elvis, but either open circuits or DC blocking capacitors make the current through them tiny, so the power generated from these is small. Another potential concern is that the capacitors do not effectively block DC while they are charging or discharging (with an RC of 1 ms), which would happen when the signals are turned on and off. However, the Uber-Elvis takes about a quarter of a second to turn on and off, which is much slower than the RC time. Additional filters upstream outside the cryostat also help with power dissipation.

The conductive track routes were also based on HV and RF considerations. The filter boards are 4-layer FR-4 PCB boards from Advanced Circuits, with 0.00135" thick layers and 0.009" between each layer. The 2nd layer from the top is a ground plane. The fast RF traces were put on the top layer, adjacent to the ground plane for shielding and isolation, and the slow RF lines were placed on the 3rd layer, also adjacent to the ground plane. This was the next best thing to striplining the RF traces, which could not be done because these are not controlled-impedance boards. All DC lines were put on the bottom layer.

Because of the tiny distances between each layer, care was taken to have no two traces on the board travel parallel paths, to limit stay capacitances. The dielectric constant of FR-4 is 4.8 and the width of the traces is 0.020", so the worse-case capacitance for a maximum 3 inch long leakage capacitor between adjacent board layers would have been 7 pF (which is much smaller than the 1000 pF capacitors used in the circuits). Every micron of the board was utilized to maximize the spacing between traces on the same layer. Again the larger area of the CTRAP filter boards allowed the spacing between all components and

traces on the board to be increased to 0.06" – 0.07", from the BTRAP minimum spacing of 0.05". The FR-4 dielectric breakdown of 20 MV/m gives a breakdown voltage of 4.5 kV between layers and 30 kV between traces on the same layer, which are much higher voltages than we will ever apply to these electrodes.

Additionally, each track on the filter boards between the external wires from the hat and the electrode pin was made as short as possible, and filter components were placed at the end of all tracks, immediately before the electrode pins. Kinks in the traces were avoided, because this can cause RF signals to reflect, or for high-voltage signals, cause electrons to accumulate in a corner and eject. The CTRAP filterboards have so far been checked with low-voltage AC and DC signals and leakage resistances were checked with an electrometer. The boards work as expected.

3.3 X-Y movable stage

Access to the interior of the electrode stack is necessary to perform antihydrogen experiments. Antiprotons, for example, enter the Penning trap from the bottom through a series of thin metal windows designed to slow them down while maintaining the experiment vacuum. From the top, access is needed for electrons, positrons, and laser beams, and each requires a specialized window. Figs. 3.9 and 3.10 show a 4.2" OD window plate with 5 different windows. Only one can be centered above the stack at a time, so a motorized custom precision translational stage was designed and built [36] to bring the desired window on-axis with the electrode stack. The window plate sits 6.4" above the top electrode on a 6" OD window flange, which bolts to flexible bellows sections above and below it.

The window plate separates the experiment space from the positron space, except for a 1.5 mm diameter hole located in its center. This hole is the most commonly used

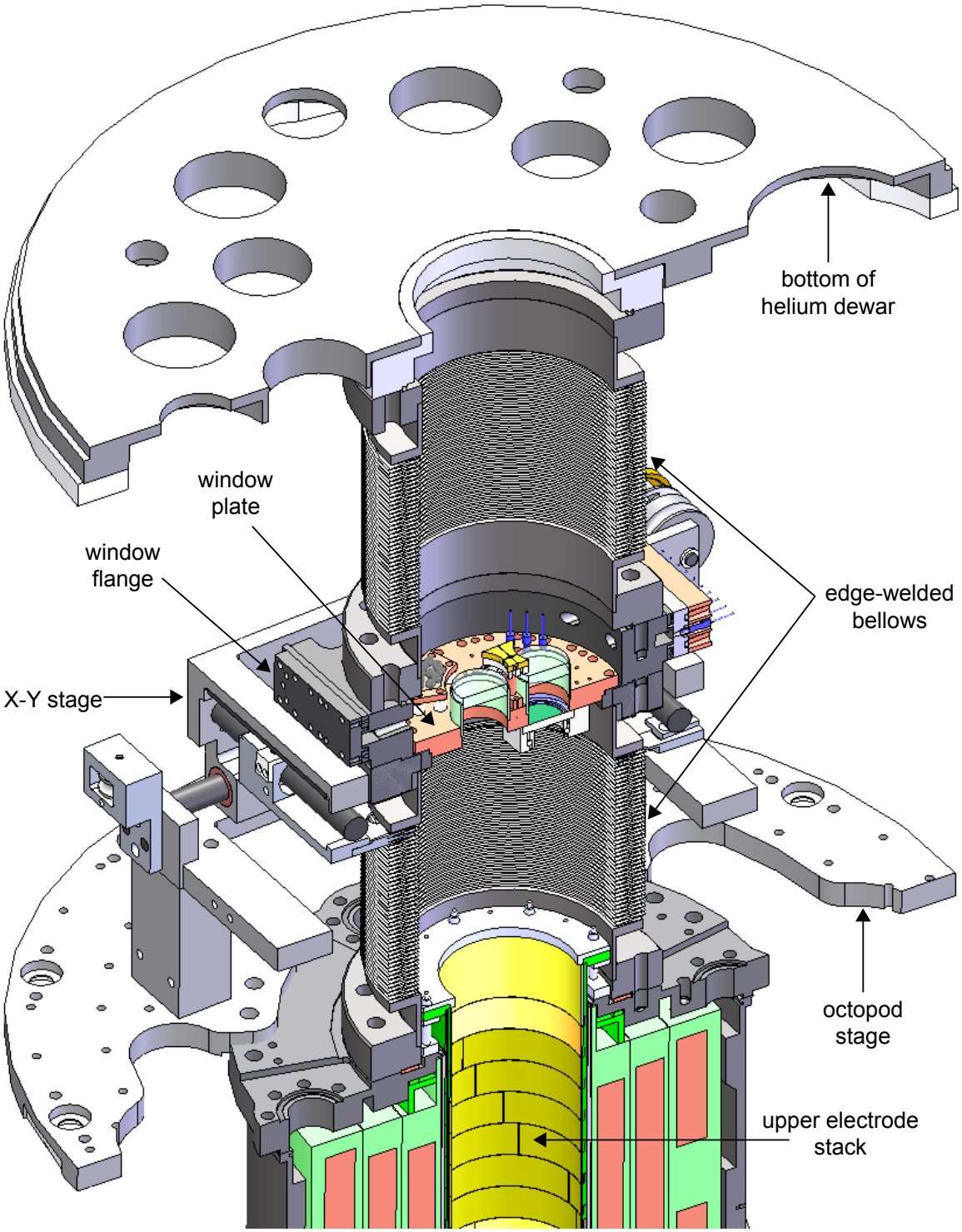


Figure 3.9: Bellows, XY Stage, Window Plate and electrode stack Assembly.

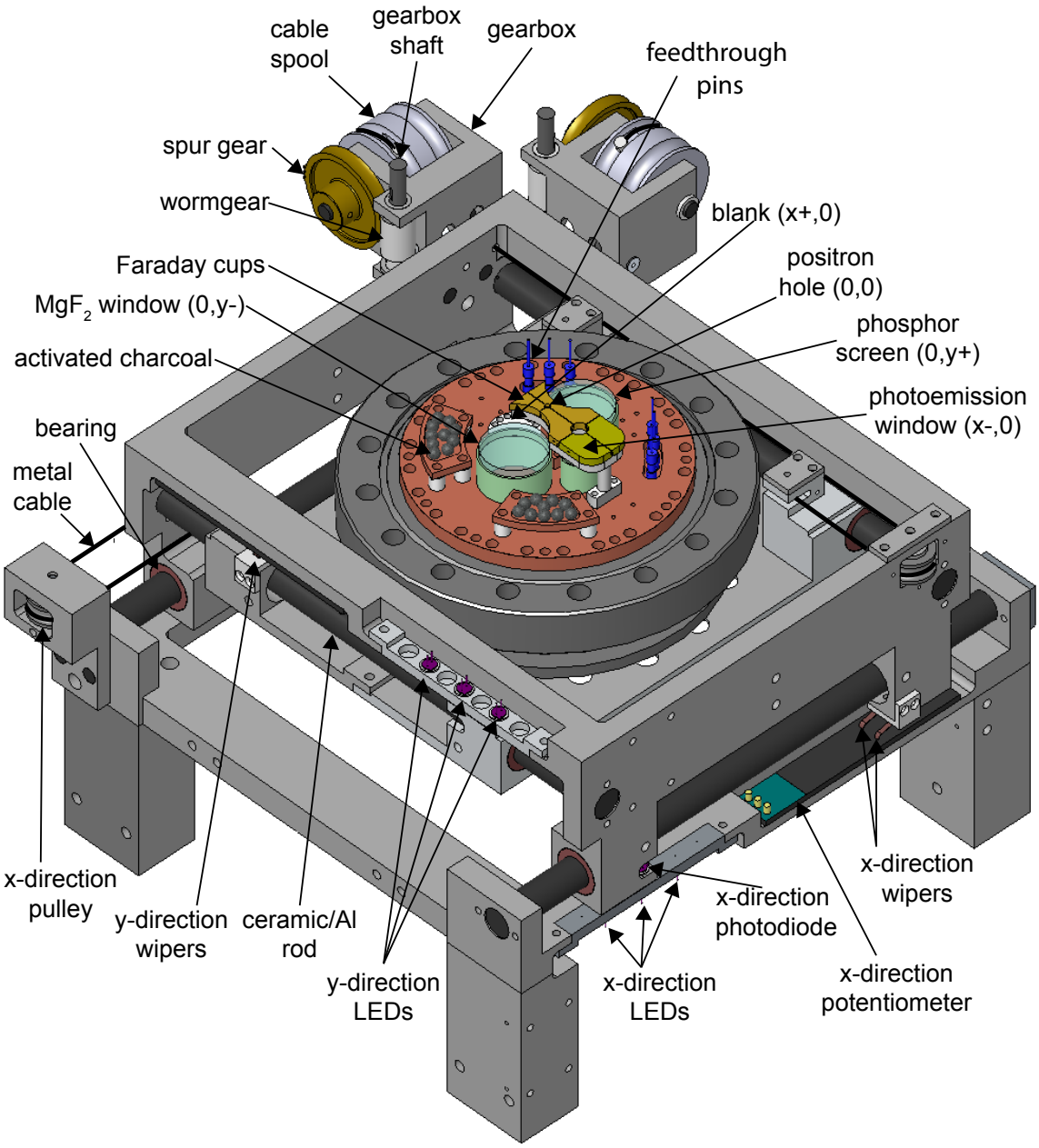


Figure 3.10: XY stage and Window Plate Assembly.

window, and allows positrons to pass through. Activated charcoal on the window plate helps maintain the experiment vacuum. Faraday cups surrounding the positron hole are used to track the positron beam for steering purposes. Once the beam is properly steered, the beam is focused tightly enough to pass through the aperture with essentially no positrons hitting the cups.

A MgF_2 window as well as a port that is currently sealed off will be used for UV laser or microwave access for cooling and/or spectroscopy. A gold-foil photoelectron loading window can be used as a source of electrons when irradiated with a UV laser. Lastly, a phosphor screen will be used to image the geometry of the charged plasmas.

The edge-welded stainless steel bellows sections (made by Metal Flex) provide separation between the experiment/positron space and insert dewar space. The 0.005" wall thickness and 70 convolutions of the 4" diameter bellows make them flexible enough to move laterally with the window plate. We normally avoid using even slightly magnetic materials such as stainless steel below the helium dewar, since this area is close to the trapping region, but the original edge-welded titanium bellows (manufactured by Mewasa) could not handle the full translation and broke repeatedly on BTRAP. The stainless steel bellows was found to be more durable than the original after cold-testing. Additionally, the window plate was modified so that the windows are spaced more closely together, minimizing the necessary travel distance of the stage. The total possible travel range in the X direction is 2.3", but limit switches (C & K Model ZMA) and spacers were installed to prevent the stage from moving more than 1.5" to prevent strain on the bellows and the motors. Similarly, the total possible travel range in the Y direction is 2.5", but the actual range is limited to 1.4". The stage can move just far enough to center the outermost windows. When the moving stage hits the limit switch the motors are inhibited. To protect the bellows and minimize its extension, the stage must not be moved in one direction unless it is centered in the other

direction, and this has been hard-coded in the control software.

The motors for the stage are located 15 inches above the hat (see 2.1). The motors contain coil windings that, when energized, cause a permanent magnet rotor to align itself with the applied field. Stepping the motor sequentially alters the direction of the applied field, generating rotary motion. Therefore, if the motors are located too close to the 1 T field they may malfunction. The rotor turns a drive shaft made up of G10 tubes and custom-machined telescoping universal joints made of metal (replacing store-bought plastic joints which did not survive the cryogenic environment). The drive shaft ends at the gearbox shaft on the stage, shown in Fig. 3.10. When this shaft rotates one way, a wormgear rotates with it, which turns a spur gear, which rotates a cable spool, which winds up a metal cable and puts it under tension, which pulls the stage towards the spool. If the drive shaft is rotated the other way, a 2nd cable on the spool that wraps around a pulley on the opposite end of the stage gets wound up instead, and pulls the stage away from the spool. The stage slides along ceramic-coated aluminum rods. All surfaces which translate or rotate against each another are mediated by either freelon-coated aluminum bearings or all-nylon bearings.

Custom electronics were designed (Fig. 3.11) and built to control the motors with an AMS Max 420 controller and read out the stage position with Opto 22 modules. Previously, these tasks were done manually by turning the drive shafts by hand above the hat and applying currents and measuring voltages through individual pins in a breakout box. Simple stage tasks took hours. Now the stage can be controlled remotely by selecting the desired window via a dropdown menu in a control Labview VI and the window is centered in a minute or less. The electronics box was designed to be interchangeable between BTRAP and CTRAP.

The stage position along one axis (e.g. X-direction) is determined by measuring the resistance of a 5 kΩ precision potentiometer (ETI Systems LCPL200-5K) that is glued to a

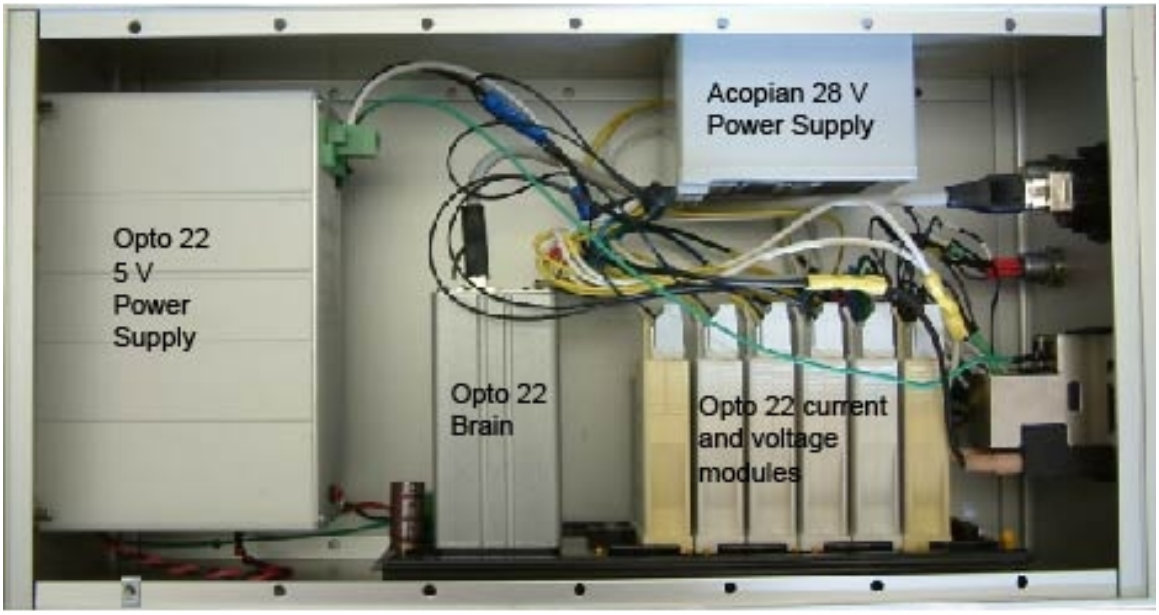


Figure 3.11: Control electronics for X-Y translational stage

fixed part of the stage. The moving part of the stage has BeCu spring contact brushes that short the conductive plastic strips of the potentiometer at different positions as the stage moves along it, shown in Fig. 3.10. The resistances are calibrated against known stage positions at room temperature. The calibration between resistance and number of motor steps is also measured, and both calibrations are used to set the stage position. However, since the resistance of the potentiometer changes with temperature and materials contract, the resistance-position calibration must be redone for different temperatures. This can be done using LEDs as position markers.

For each axis, 3 LEDs (Opto Diode Corp. OD-880W) are mounted to the fixed part of the stage, while the moving part of the stage holds a photodiode (Hamamatsu S2386-18K) that scans across the LEDs. The middle LEDs in the X and Y directions are positioned such that their photodiode signals are maximized when the central positron window is centered above the electrode stack. The LEDs along each axis are spaced the

same distance apart as the windows in the X and Y directions–0.65" at room temperature, allowing a direct correspondence between LED illumination and centered windows. That is, at other temperatures, we know if the windows are centered even if we don't know the spacing between the windows or the temperature-dependence of the potentiometer, and this enables us to find the new resistances corresponding to centered windows.

The alignment between window position and LED position does not stray much with temperature due to the similar thermal contractions of the copper window plate and the aluminum XY stage. The total contraction of the outer LEDs and the outer windows is only about 0.005" from 293 K to 4 K and the differential contraction is only 0.001", an order of magnitude smaller than the width of the LED peak. The width of a peak is 0.009"±0.002" (5.8±1.1Ω), smaller than the smallest window (the positron window) size by a factor of 6. The peak widths are limited by 0.0135" diameter pinholes mounted in front of both the LEDs and photodiode for more precise positioning. The desired window can be centered within ±0.0045" of its LED peak using the potentiometer, within just 2 moves of the stage, and this can be reduced to 1 move by incorporating backlash into the routine. After bringing the window roughly on center with the potentiometer, the centering is then fine-tuned by maximizing the LED signal. How precisely the stage can be centered on the LED maximum depends on the fineness of the motor step size, which in our case was measured to be 0.00061" (15 microns) (described in more detail below). One can keep track of motor steps to directly measure position as well. In order to actually know the stage position at arbitrary places (not marked by LEDs) to the precision of the motors, at 1.1663 inch/k Ω (measured at room temperature in the X-direction, for instance) the corresponding potentiometer resolution must be 0.52 Ω. This is not the case, as the resolution is actually about 9 Ω, corresponding to a position resolution of 0.01" with the just potentiometer, without using the LEDs for increased precision.

Calibrations

Fig. 3.12 shows some examples of the calibration data, which was mostly taken at room temperature and atmospheric pressure. Fig. 3.12a and Fig. 3.12b show the photodiode signals as the stage scanned along the X and Y axes. The LED positions are known to be -0.65", 0", and 0.65" (with 0.001" machining tolerance), allowing determination of the potentiometer resistances corresponding to centered windows. The 3 data points were fit to a line to determine stage position from resistance at arbitrary positions. The data for the first Y-axis LED peak looks messy due to broken wipers that were not making good contact with the low end of the potentiometer, and these will be replaced. This LED peak is in fact entirely accessible when the signal is recorded as a function of position.

Fig. 3.12c shows the photodiode signal as a function of number of motor moves, set to be 25 steps each. Since the positions of the LED peak positions are known, this gives the calibration between distance and number of motor steps. The potentiometer resistance was measured simultaneously, allowing a calibration between resistance and number of motor steps. A linear fit was done for the resistance data. The maximum deviation from linearity, assuming that the motor was functioning properly and actually stepping by a constant 25 steps with each move, was 9 Ω. The fact that the LED peaks are almost exactly equally-spaced indicates that on average the motor was functioning properly. Figs. 3.12a-c show data for a one-way trip across the stage. The peak resistances going forward and then backward along the X-axis were found to be shifted by 3-4 ohms (data not shown).

Cold-testing of the CTRAP stage was done during the warmup after the last CTRAP cooldown attempt in 2012. Fig. 3.12d shows Y-axis calibration data for when CTRAP was between 140-147 K, under vacuum, and in a 1 T magnetic field. This is shown together with the warm data in Fig. 3.12c. The resistances corresponding to LED peaks increased by about 20 Ω for the cold temperature but the resistance-position slope is the

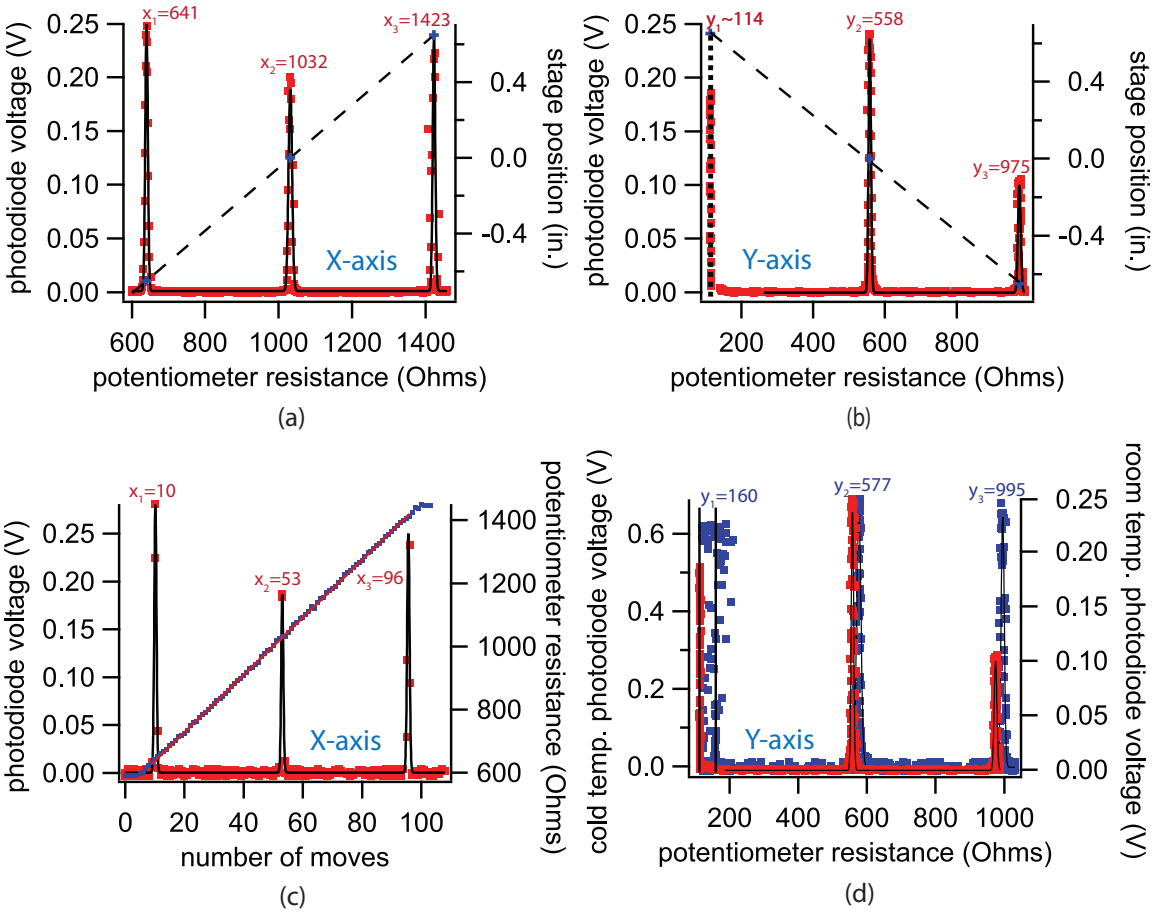


Figure 3.12: Single-pass room-temperature position-resistance calibrations for (a) X-axis and (b) Y-axis. The red markers are LED photodiode signals with black Gaussian fits and blue markers are known stage positions at each LED, with dashed linear fit. The first peak for Y calibration was not fit. (c) Single-pass room-temperature resistance-number of moves (blue) with red linear fit and distance-number of moves calibrations for X-axis. Each move is 25 motor steps. (d) Multiple-pass position-resistance calibration for Y-axis for (blue) 140 K and (red) room-temperature position-resistance calibration, with black Gaussian fits. The first peak was not fit.

same. Cold calibration data for the X-axis was not obtained, because the motors stalled before reaching the outer LEDs. It was possible, however, to move the stage entirely across the X-axis by hand when the motor was disconnected.

Challenges in stage movement

Moving a metal structure in a high magnetic field and cryogenic temperatures is challenging. Thermal contraction can cause gap sizes between parts to decrease, friction to increase, and components to re-align themselves. Even thermal cycling without moving the stage can cause screws to loosen and parts to go out of alignment. To limit these problems, the stage was designed to have as few separate components as possible and the large parts that are separated are thermally coupled with copper braid to minimize temperature gradients. Nonetheless, some difficulties remained in getting consistent motion from the stage.

Since the time that the X-Y stage was installed on CTRAP, the stage moved easily and reproducibly every time it was operated for about 2 years. In the following year, the window plate had been removed and reinstalled (with no other known change to the X-Y stage system) and a CTRAP cooldown had been attempted. When it was tested again the stage could no longer traverse the full range, evident from the motors stalling, and also from the increased difficulty in turning the drive shafts by hand. In addition, a noticeable new asymmetry in the force needed to move the stage in the + and - directions arose, especially in the X-axis. An attempt was made to find and fix these problems, and although the underlying problem was never found, many improvements were made as a result. The following describes these attempts and improvements.

When the stage is off-center, the bellows sections above and below the window flange balance each other's torques on the window flange. The bellows sections also balance each other's vertical spring forces. However, there is still a net force downwards due to the weight of the window flange, window plate, and top section of the stage. This weight puts more force on the tops of the aluminum/ceramic rods than the bottom, increasing friction during lateral movement. Additionally, the rods were observed to be slightly bowed,

though this bending may fall within the manufacturing tolerance. For the X-axis, both these problems cause a third problem of closing a 0.01" gap that should be present between the fixed and moving part of the stage, creating more friction. The bowed rod problem was partially remedied by placing support blocks underneath one of the lower rods. On the opposite side of the stage where support blocks could not be installed, washers to increase the gap between the photodiodes and LEDs were installed. This eliminated contact between the fixed and moving parts of the stage.

There was still the problem of asymmetric torques required to move the stage between X+ and X-. Force measurements were done with a spring scale at every accessible juncture of the system to find the source of asymmetry. Starting from the movable stage outwards towards the drive shafts, the first place an asymmetry arose was at the pulley. As described earlier in this section, when the stage is pulled towards the cable spool, it is pulled directly, but when the stage is moved in the opposite direction, it is pulled by a cable that loops around a pulley. The force needed to move the stage in the latter direction was 26% greater than the force needed for the former, due to friction in the pulley. A similar discrepancy was observed for the Y-axis. Other asymmetries that propagated from this were also measured and recorded.

To mitigate these problems with rotational friction, bearings were replaced. The original freelon-coated bearings (Stock Drive Products) have only a 0.020" thickness of lubricating surface and, though these worked well in the first two years of operation, they were found to wear out quickly when motional problems with the stage began. Most of these bearings have been replaced with high-performance nylon bearings (Part number 58315K104) whose entire body is lubricating, and which have a factor of 20 higher PV (pressure, veclocity) ratings. The rest, which are harder to access, will be replaced shortly.

To check the loads on the bearings, the pressure and velocity actually applied in

our system was investigated. Force measurements were used to estimate the pressure on the original freelon bearings, which was found not to exceed specifications. However, the motors had been running at the default velocity of 900 rpm, a factor of 6 too high, so the velocity was reduced by a factor of 30. Reducing the motor velocity also increased the motor's torque output, which is limited to the zero-velocity holding torque of 210 oz-in (15.1 kg-cm) at the maximum current setting. The torque increases with current, and the current setting is 36% of its maximum. Since the measured torque at the top of the drive shaft needed to move the stage was a factor of 8 less than the maximum torque of the motors, this current setting was sufficient. Longer step sizes were also implemented to reduce the number of times the motors needed to overcome static friction every time they were stepped, and to make best use of the velocity slew. These new settings (which are all now changeable in the software routine) eliminated the motor stalling at room temperature and atmospheric pressure up to the LED positions.

Summary and further improvements

Several changes were made on the X-Y stage, and these improvements incrementally eased the motion of the stage in room temperature and atmosphere, with changes in the motor settings resulting in the most obvious improvement. However, the remedies described above treated the symptoms of the stage problems but not the cause, which is still unknown. The stage still did not move as easily or as far as it did one year prior, without those improvements. Additionally, the motorized stage motion along the X-axis was unsuccessful when it was tested for the first time in the cryogenic, 1 T environment, though it worked in room temperature.

For certain pieces of the stage, unbolting and bolting them back together caused noticeable changes in the ease of stage movement. Many parts of the stage were disassembled

and reassembled for troubleshooting, but one thing that was not tried was unbolting and reinstalling the bellows sections.

The motors we used (Advanced Micro Systems AM23-210-3EFB) have the drawback that they are now a legacy model that cannot be upgraded. They were also not set up for encoder feedback capability, which allows the abortion of a motion command when the motor stalls. Thus far, the motors were stopped manually by the user when they were observed to stall from position tracking in the custom Labview program. New motors (BM23-262-S) along with the encoder hardware and cabling were purchased but not yet implemented.

Although different parts of the stage are thermally coupled, there are currently no temperature sensors on the stage for gradient monitoring. This can be implemented. Other pending mechanical improvements are replacing the the Y-direction wipers, the remaining bearings, and the spur gears with a design that does not use set screws.

On the software end, backlash should be investigated and incorporated into the motion routine, and if desired can also be minimized by decreasing the cable length on the stage. Also, the stage calibration numbers should eventually be entered in the experiment database for fast access–currently they are built into the back panels of the VIs.

Many CTRAP XY stage hardware and wiring improvements were made on systems that previously broke. Mechanical parts such as the drive shafts and universal joints are more robust and photodiodes and LEDs are more stably mounted. BTRAP problems with constantan wires and termite pins breaking was remedied on CTRAP with thorough strain-relieving and bundling, and better electrical connectors. Additionally, all wires on CTRAP terminate at the fixed part of the stage rather than the moving part, which eliminates movement of the fragile constantan wires and decreases their chance of breaking. Contact and possibility of contact between the stage parts which should be isolated has been eliminated

and friction has been reduced with better bearings. Better velocity and current settings were implemented that reduced stalling and increased torque, and better motors will be implemented. For the first time the XY stage was controlled completely remotely, making the operation of the stage much easier than previously.

3.4 Redesign of helium flow pipes

One of the major lessons learned throughout this work is that cryogenic vacuum joints are fragile. "Cryogenic" epoxy must not be used for these applications, and braze and weld joints must be tested diligently and preferably on-site. Additionally, the prep geometry for brazing and welding is crucial. An example of a problematic vacuum system on CTRAP is the helium pipes that carry liquid helium from the helium dewar to the superconducting magnets (Ioffe trap and $\overline{\text{p}}$-solenoid).

The original pipes making up the flow path through the helium space were made up of short Be-Cu bellows sections (needed to absorb thermal contractions during cooldowns) that are hard-soldered to copper tubes, which were glued to G10 tubes also glued to titanium conflats at the ends (all non-magnetic materials). These G10 glue joints were made with Stycast 1266 epoxy, and were later found to leak helium.

The G10 pieces were discarded and replaced with silver tubes, also non-magnetic. Previous failures deemed glue joints between these pieces no longer an option, so we opted for soft-soldering and brazing. Since copper cannot be welded or soldered to titanium, the silver tube was a necessary intermediary. The silver tube was soft-soldered to the copper tube on one end, and brazed with an electron beam to a titanium conflat on the other. However, an unfortunate combination of a poor braze prep design and an inexperienced technician caused almost all the Ag-Ti joints to break or leak after cold-testing.

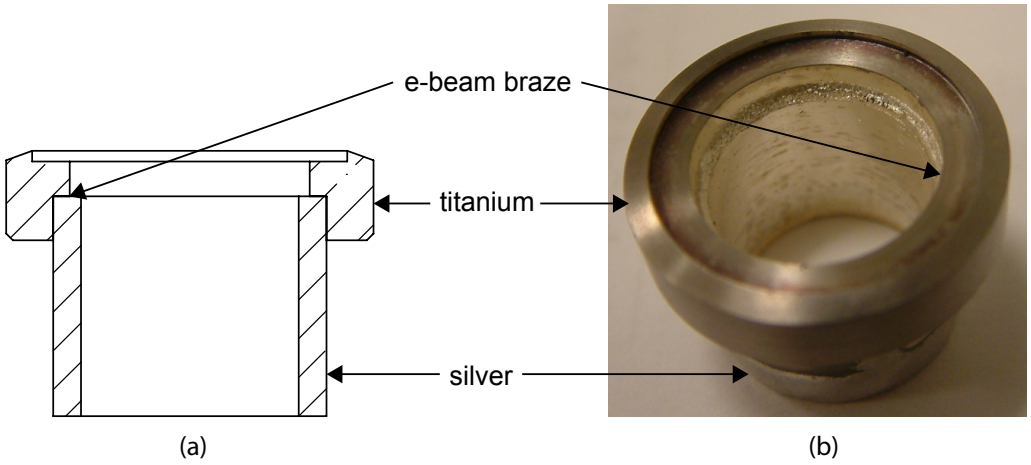


Figure 3.13: Problematic (a) braze prep and (b) resulting braze.

Fig. 3.13 shows the problematic braze prep design and the resulting braze. In this geometry, the silver plug makes a convenient seat on which the titanium conflat can be correctly positioned during the braze. However, to make this joint, the electron beam must be positioned at a 45°, resulting in a small penetration depth. Additionally, this seat on the silver plug makes for a high-stress joint because the materials are so dissimilar. Silver has about twice the coefficient of thermal expansion of the Ti (18.9 x 10^{-6} K^{-1} at room temperature and 8.6 x 10^{-6} K^{-1} at all temperatures up to the melting point [37], respectively). When heat is applied with the electron beam, the silver expands twice as much as the titanium and the weld is made in this condition. When they cool back down to room temperature the silver tends to shrink back to its original size and location but is unable to do so due to the force from the weld. The poor quality of the weld shown in Fig. 3.13b can be seen by its grainy texture.

Titanium and silver are tricky to join not only because of their tendency to make high-stress joints, but also because of their dissimilar melting points and thermal conductivities. The melting point of silver (962°C) is much lower and its thermal conductivity (429 W m^{-1} K^{-1}) is much higher than that of titanium (1665°C and 21.9 W m^{-1} K^{-1},

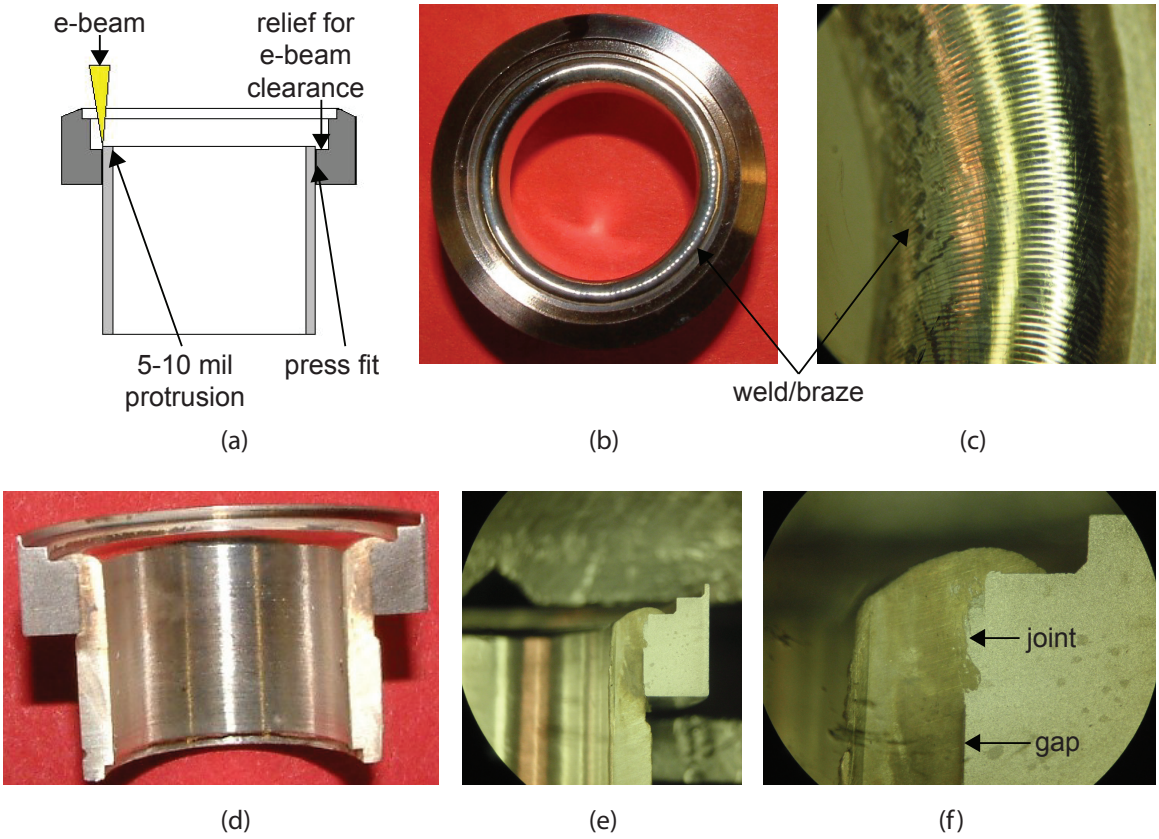


Figure 3.14: (a) A better weld prep geometry, with clearance for the electron-beam, a small protruding lip that can be melted, and a press-fit to the desired height rather than a shelf. (b) and (c) Successful weld and close-up view of expected ripples. (d) Cross-section of weld on a test-piece. The right side is acid-etched with a titanium etchant. (e) and (f) Close-up of right side to see weld-penetration. (f) The joint is about 0.080" deep and the dark line underneath is the gap between the silver and titanium pieces

respectively). As such, silver and titanium cannot be conventionally welded together, since this involves both metals melting and mixing at the same time. Instead the silver must be melted over the titanium in a process that is something in between a weld and a braze (but more of a braze).

Fig. 3.14 shows an example of the new geometries and their successful joints. This time they were done by Bruce Gold (Joining Technologies), who together with us, developed the methods and settings for these unconventional Ag-Ti joints using test pieces that we

provided. To alleviate some of the stress on the joints, the pieces were redesigned to be press-fit together, eliminating the need for extra alignment geometry and putting the weld fixed point at the edge of the silver cylinder rather than the middle. A small 0.005"-0.010" silver protrusion provides just enough silver to melt over the titanium without requiring large amounts of energy to melt. The silver is melted onto the titanium by applying 90% of the energy on the silver and 10% on the titanium. Different purities of silver were also tested, and in the end we used 99.95%+ purity silver stock from Goodfellow. Purer grades than this welded worse. The goodness of these silver-titanium joints also depended on the protrusion, relief, and wall thickness of the silver.

The final pieces were done using a penetration weld and then a cosmetic pass at the end. Cosmetic passes use about 1/4 the power and barely penetrate. These are only surface touch-ups at low power, giving rise to the smooth shapes shown in Fig. 3.14, and do not affect the rest of weld, which penetrate about 0.06 - 0.08" deep. The advantage is to smooth out any visible nooks and crannies to ensure leak-tightness, and to prevent outgassing from these holes. The resulting pieces were all cold-tested and leak-checked and found to be robust.

Chapter 4

Plasmas in Penning Traps

Thus far single charges in Penning traps have been described (Sec. 2.2). Throughout this thesis, we work with collections of millions of charged particles–single-species and multi-species plasmas. The additional electric field from the charges substantially changes the total potential, requiring modification of single-particle results. The collective dynamical behavior of plasmas [38] must also be understood in order to efficiently control them in experiments, and is thus reviewed in this chapter.

Weakly-correlated clouds of charges with density n are called single-component plasmas when the cloud is larger in all dimensions than its Debye length, $\lambda_D = \sqrt{\epsilon_0 k_B T/nq^2}$. The Debye length is the radius inside which the potential from a test charge of opposite sign is screened out by the cloud of charges surrounding it, and in the case of a Penning trap, the charge cloud shields the trap potential. Particularly, the charges rearrange themselves until the total potential is no longer harmonic at the center of the trap but instead constant throughout the plasma. Plasmas qualify as weakly-correlated if the coupling parameter $\Gamma = q^2/(4\pi\epsilon_0 a k_B T) \ll 1$, where a is the Wigner-Seitz radius, or inter-particle spacing, defined by $(4\pi n a^3)/3 = 1$. Γ is the ratio of the electrostatic energy of neighboring charges

to the random thermal energy $k_B T$. For our experiments, Γ ranges between 0.01 and 0.1.

Penning-trapped plasmas in thermal equilibrium have universal bulk properties– their rotation is shear-free (rigid), their shape is spheroidal (for small plasmas compared to the trap), and their density is nearly constant up to some spheroidal surface after which it drops to zero on the scale of the Debye length. Additionally, multi-species plasmas in thermal equilibrium can be centrifugally separated.

4.1 Plasma-modified Penning trap theory

To analyze the new effective potential caused by the plasmas we can go into a frame that rotates with the plasma at angular frequency $-\omega_r$. The effective potential in the rotating frame is [38]:

$$\Phi_R(\rho, z) = \Phi_T(\rho, z) + \frac{m}{2q}\omega_r(\omega_c - \omega_r)\rho^2 \tag{4.1}$$

$\Phi_T(\rho, z)$ is the trap potential in the lab frame already introduced in Eqs. 2.1 or 2.11. The 2nd and 3rd terms come from the fictitious Coriolis and centrifugal force experienced in the rest frame of the plasma. To solve for the plasma space-charge potential Laplace's equation is no longer valid and Poisson's equation $\nabla^2\Phi_p(\rho, z) = -qn(\rho, z)/\epsilon_0$ must be satisfied, subject to the boundary condition that Φ_p vanishes everywhere on the trap wall (this holds for our traps since the 2-9 mm plasma radius is smaller than the minimum electrode radius of 18 mm). The charges rearrange themselves until the total potential is constant inside the plasma. Therefore

$$\Phi_p(\rho, z) + \Phi_R(\rho, z) \approx \text{const} \tag{4.2}$$

The effect that the space-charge potential has on the total potential is shown in Fig. 4.1. Since $\Phi_R(\rho, z)$ is quadratic in z and ρ, for Eq. 4.2 to hold, $\Phi_p(\rho, z)$ must also be quadratic. It is known that a uniformly-charged spheroidal plasma produces such a potential inside the

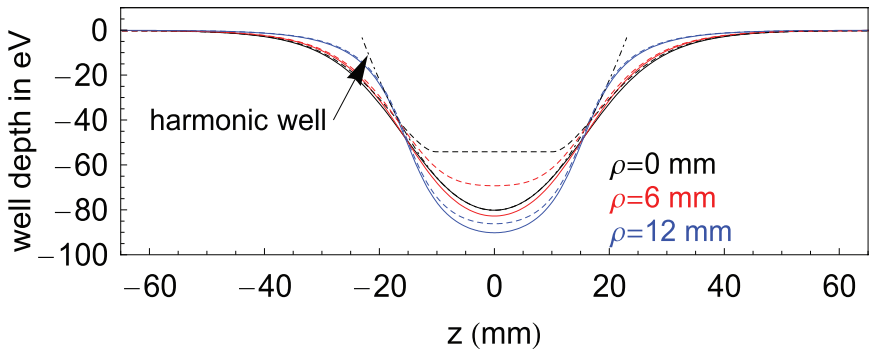


Figure 4.1: The effect of space-charge potential on the total potential. 100 V are applied on an electrode, and the trap potential without any charge (solid lines) are compared against the total potential in the presence of 100 million electrons (dashed lines), for different radii. On-axis ($\rho = 0$), the potential is nearly harmonic. The potentials get deeper and less harmonic off-axis with increasing radii. Taken from [12].

spheroid and an exterior potential that approaches zero at ∞ [39]. This implies that the shape of the plasma is spheroidal. Applying the Laplace operator to Eqs. 4.1 and 4.2 and combining them yields

$$\nabla^2\Phi_p = -\nabla^2\Phi_R = -\frac{2m}{q}\omega_r(\omega_c - \omega_r) = -\frac{qn}{\epsilon_0} \tag{4.3}$$

where $\nabla^2\Phi_T$=0 has been used. Eq. 4.3 shows several important things. First is that the density $n(\rho, z)$ is in fact constant inside the plasma (and at the edge, drops to zero on the scale of the Debye length). We can also see that increasing the plasma rotation frequency ω_r causes the density to rise, necessitating a decrease in plasma radius. Eq. 4.3 can be rearranged to give

$$2\omega_r(\omega_c - \omega_r) = \frac{q^2 n}{m\epsilon_0} \equiv \omega_p^2 \tag{4.4}$$

where ω_p is the well-known plasma frequency. Using the good approximation that $\omega_c \gg \omega_r$ yields an expression for the plasma rotation frequency ω_r:

$$\omega_r \approx \frac{qn}{2\epsilon_0 B} \tag{4.5}$$

The center of mass axial oscillation of the plasma may still be described by the axial frequency ω_z, given by Eq. 2.5.

4.2 Single-species plasmas

4.2.1 Radial confinement

The motion of N charges in the Penning trap electromagnetic fields is governed by the Hamiltonian given in Ref. [38], with total canonical angular momentum (in cylindrical coordinates)

$$P_\theta = \sum_{j=1}^{N} m\rho_j^2\dot{\theta}_j + \frac{q}{2}B\rho_j^2 \tag{4.6}$$

If the Hamiltonian is cylindrically symmetric, then the total canonical angular momentum is conserved: P_θ=L, a constant. For a sufficiently large magnetic field, the first term in Eq. 4.6 is negligible compared to the second, so

$$L \approx \frac{1}{2}qB\sum_{j=1}^{N}\rho_j^2 = \frac{1}{2}qBN\langle\rho^2\rangle = \frac{1}{5}qBN\rho_p^2 \tag{4.7}$$

where ρ_p is the plasma radius, and the last equality is true for a spheroidal shape. This makes the total angular momentum independent of mass and velocity. Eq. 4.7 is a confinement theorem for trapped plasmas [40]. As the B-field increases, the mean square radius of the plasma must decrease to conserve angular momentum. For real traps, slight imperfections and field errors break the cylindrical symmetry and cause drag torques on the plasma. This can cause the plasma to expand slightly over time. Even though the timescale for an appreciable increase can be large compared to the thermalization time, measures are taken to address this problem: a rotating wall, described in Sec. 5.1.3, and an experimental protocol designed to minimize the $\overline{\text{p}}$ storage time in their potential wells.

4.2.2 Plasma geometries

For efficient $\overline{\mathrm{H}}$ production the radii of the $\overline{\mathrm{p}}$ and e^+ clouds need to be well-matched. Since we are currently unable to directly image our plasmas (this is work in progress) we need another means to determine the plasma geometry. The total particle number in a spheroid can be calculated in terms of its geometrical properties:

$$N = \frac{4}{3}\pi n z_p \rho_p^2 = \frac{4}{3}\pi n \alpha \rho_p^3 \tag{4.8}$$

where z_p is the half-length of the plasma and $\alpha \equiv z_p/\rho_p$ is the plasma aspect ratio. The aspect ratio is related to the plasma frequency ω_p and the plasma center-of-mass axial frequency ω_z by [41]

$$\frac{\omega_z^2}{\omega_p^2} = \frac{Q_1^0\left(\frac{\alpha}{\sqrt{\alpha^2-1}}\right)}{\alpha^2-1} \tag{4.9}$$

where $Q_l^m(x)$ is an associated Legendre polynomial of the second kind, and the (l,m)=(1,0) mode represents the ω_z oscillation. The (1,0) mode can be calculated or measured. However, because ω_p and thus density (Eq. 4.4) are typically unknown, more information is needed to find the plasma aspect ratio from Eq. 4.9. A second equation, provided by a measurement of the (2,0) mode can be used with the (1,0) mode to determine the plasma's aspect ratio and plasma frequency. Knowing the plasma frequency allows us to calculate the plasma density.

The (2,0) mode is the plasma's quadrupole mode and represents the oscillation of its aspect ratio. During this oscillation the plasma remains spheroidal and its density remains uniform. We are interested in the m=0 states, corresponding to axially symmetric plasmas. For the case of a magnetized ($\omega_c \gg \omega_p$) zero-temperature spheroidal plasma, the eigenmodes for arbitrary l are given analytically by [42]:

$$1 - \frac{\omega_p^2}{\omega_l^2} = \frac{k_2}{k_1}\frac{P_l^0(k_1)Q_l^{0\prime}(k_2)}{P_l^{0\prime}(k_1)Q_l^0(k_2)}$$

where

$$k_1 = \frac{\alpha}{\sqrt{\alpha^2 - 1 + \frac{\omega_p^2}{\omega_l^2}}} \qquad\qquad k_2 = \frac{\alpha}{\sqrt{\alpha^2 - 1}}$$

and P_l^0 and $P_l^{0'}$ are the Legendre function of the first kind and its derivative, respectively. $Q_l^{0'}$ is the derivative of Q_l^0. The quadrupole frequency ω_2 can be measured (Sec. 5.1.3), enabling the determination of plasma density and the aspect ratio. The number of particles in a plasma can also be measured (Chapter 5), and these three parameters can be plugged into Eq. 4.8 to finally determine the plasma radius ρ_p. This method is used to determine electron and positron plasma radius, while the $\overline{\text{p}}$ radius is deduced by other methods (Sec. 5.2.4). It should be noted that the analytical treatment in this section only applies to spheroidal plasmas. There are some situations in which our plasmas deviate from spheroidal geometries (large plasmas, for example), and in these cases the plasma geometries can be numerically calculated.

4.2.3 Isotropization rates

As described in Chapter 2, radial and axial single particle motion in Penning traps are decoupled. This however is not true for plasmas due to collisions between its N oscillation modes, where N is the number of particles in the plasma. These collisions cause the axial and radial motions to mix.

In order to design experiments with the correct timescales (elaborated on in Chapters 5 and 6), we must have knowledge of plasma thermalization time and rate of directional energy exchange, also known as the isotropization rate. This rate is suppressed in the presence of a magnetic field due to the increased difficulty in radial transport, and decreases with decreasing temperature, both of which effectively decouple the two motions again. For

plasmas in a magnetic field, the magnetization can be parameterized by $\kappa = b_{min}/r_c$, where

$$b_{min} = \frac{1}{2\pi\epsilon_0}\frac{q^2}{k_B T} \tag{4.10}$$

is the classical minimum impact parameter and $r_c = \frac{\sqrt{2mk_BT}}{qB}$ is the cyclotron radius. A plasma is strongly magnetized when $\kappa \gg 1$ and weakly magnetized when $\kappa \ll 1$. For our 3.5-35 K (measured) antiprotons in a 1 T field, κ ranges from about 0.1 to 4, putting us in the intermediate $\kappa \approx 1$ magnetization regime.

The isotropization rate for the intermediate magnetization regime was calculated in Ref. [43] and is shown plotted in Fig. 4.2 for a typical plasma density of 1×10^6 cm^{-3}. For our values of κ, the axial-radial energy transfer times for $\overline{\mathrm{p}}$ are between 2.5 and 10 ms. Lepton isotropization times are even smaller because of their larger densities in our experiments. We allow much more thermalization time than this in typical plasma preparation. However, in some cases we want to perform operations on the $\overline{\mathrm{p}}$ before energy can be exchanged between the modes. One such example is during the axial mixing of antiprotons and positrons to create $\overline{\mathrm{H}}$, which is discussed in Chapter 6.

4.3 Multispecies plasmas

4.3.1 Collisional cooling

Sec. 2.2.1 described the cooling of electrons via synchrotron radiation and noted that $\overline{\mathrm{p}}$ cannot cool themselves this way in a useful amount of time. Cooling of $\overline{\mathrm{p}}$ is instead achieved via repeated collisions with cold electrons. In our $\overline{\mathrm{p}}$ cooling and trapping scheme, cold electrons wait in a potential well for the arrival of higher-energy $\overline{\mathrm{p}}$. $\overline{\mathrm{p}}$ bounce through the cold electron cloud repeatedly until they fall into the bottom of the potential well with the electrons [44, 45], forming a multi-species plasma (Chap. 5). Collisions between electrons in a plasma transfer energy between the different plasma motions on a timescale of much

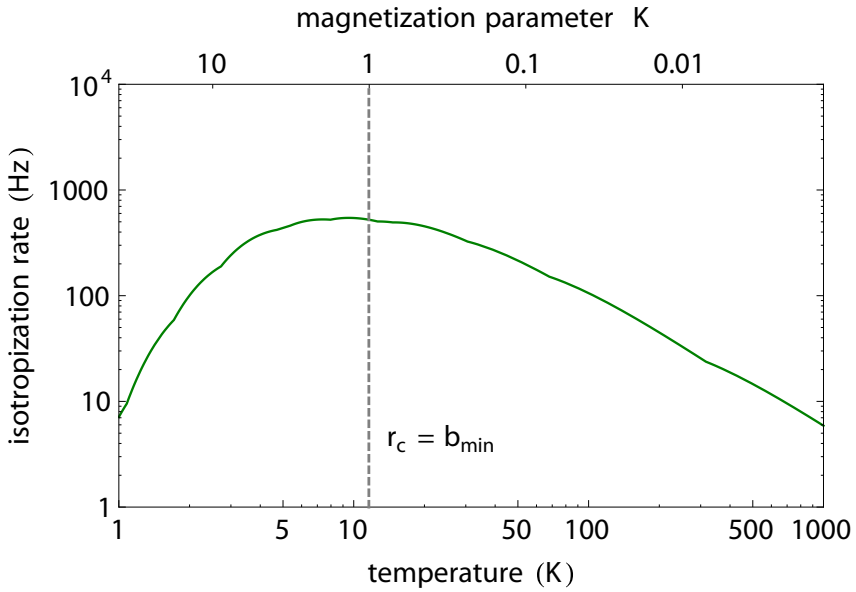


Figure 4.2: Predicted axial-radial energy isotropization rate for a $\overline{\mathrm{p}}$ plasma with central density $n = 1 \times 10^6$ cm^{-3}, in the intermediate magnetization regime.

less than a second for our electron numbers. As a result they can also effectively cool all the $\overline{\mathrm{p}}$ motional degrees of freedom.

The $\overline{\mathrm{p}}$ enter the trap with very high temperatures ($\sim$5 keV after being slowed by the degrader electrode). Since the electrons are not an infinite cold reservoir with fixed temperature, the $\overline{\mathrm{p}}$ will heat them up, as will radiation from the (fixed temperature) trap walls as they are synchrotron-cooling themselves and collisionally cooling the $\overline{\mathrm{p}}$. This can all be represented by the coupled rate equations:

$$\begin{aligned} \frac{d}{dt}T_{\overline{\mathrm{p}}} &= -\frac{1}{\tau_{e\overline{\mathrm{p}}}}(T_{\overline{\mathrm{p}}} - T_e) \\ \frac{d}{dt}T_e &= \frac{N_{\overline{\mathrm{p}}}}{N_e}\frac{1}{\tau_{e\overline{\mathrm{p}}}}(T_{\overline{\mathrm{p}}} - T_e) - \frac{1}{\tau_c}(T_e - T_w) \end{aligned} \tag{4.11}$$

where T_e, $T_{\overline{\mathrm{p}}}$, and T_w are the electron, $\overline{\mathrm{p}}$ and and electrode wall temperatures, respectively, $\tau_{e\overline{\mathrm{p}}}$ is the time constant for collisional electron cooling of $\overline{\mathrm{p}}$, N_e and $N_{\overline{\mathrm{p}}}$ are the electron and $\overline{\mathrm{p}}$ numbers, and τ_c is the electron synchrotron cooling rate given in Eq. 2.10. Eq. 4.11 can be solved for $\tau_{e\overline{\mathrm{p}}}$ for different ratios of $\overline{\mathrm{p}}$ and e^- density. The zero-B-field cooling

time for our initially 5 keV $\overline{\mathrm{p}}$ plasma, with $\frac{N_{\overline{\mathrm{p}}}}{N_e} = 10^{-2}$, takes about 5 seconds. Of special relevance to this theory is the low-electron density case during "embedded-electron" cooling in a magnetic field, which will be explained in Sec. 5.2.5.

4.3.2 Centrifugal separation theory

As the $\overline{\mathrm{p}}$-e^- plasma comes into thermal equilibrium, it exhibits the phenomena of centrifugal separation [46]. The first experimental demonstration of centrifugal separation of antiprotons and electrons [12] is described in Sec. 5.2.4.

Centrifugal separation between 2 rotating species of different masses is a result of the different centrifugal forces $F = m\omega_r^2\rho$ on the species. Its effect can be masked however, if the thermal energies are large compared to the centrifugal energy difference. That is, in order for centrifugal separation to be evident, the following ratio must be large:

$$\frac{\frac{1}{2}\omega_r^2\rho^2(m_{\overline{\mathrm{p}}} - m_e)}{k_BT} \approx \frac{\frac{1}{8}\frac{q^2}{\epsilon_0^2}(\frac{n^2\rho^2}{B^2})m_{\overline{\mathrm{p}}}}{k_BT} \tag{4.12}$$

where the substitution for ω_r using Eq. 4.5 and $m_p \gg m_e$ has been used. From this we can observe the controllable parameters that would lead to a visible separation: large density and radius (or equivalently, large density and number of particles), and low magnetic field and temperature. We can use the ratio of centrifugal force to thermal energy to define a class of scale lengths [46]:

$$\frac{1}{l(\rho)} = \frac{d}{d\rho}\left[\frac{1}{2}(m_{\overline{\mathrm{p}}} - m_e)\frac{\omega_r\rho^2}{k_BT}\right] \tag{4.13}$$

For $l(\rho_{\overline{\mathrm{p}}}) < \rho_{\overline{\mathrm{p}}}$, the separation is large. For our typical parameters, $l(\rho_{\overline{\mathrm{p}}})$ is $\sim$1 mm, and $\rho_{\overline{\mathrm{p}}}$ is 4-8 mm. If $l(\rho_{\overline{\mathrm{p}}}) < \lambda_D$, the separation is complete. This however is not satisfied by our plasmas, which have a λ_D of 10-100 μm. The plasmas separate until they reach thermal equilibrium, at which point they rotate rigidly with one frequency. With this plasma background we are now ready to discuss its application in the following chapters.

Chapter 5

Methods

The methods used to prepare particles for the production and trapping of antihydrogen are described in this chapter. These include the use of recent apparatus additions and improvements to experimental protocols in the last few years. For example, the new field-boosting solenoid, known as the "$\overline{\mathrm{p}}$-solenoid," has allowed for a 14-fold increase in the accumulation of $\overline{\mathrm{p}}$, giving us the largest number ever accumulated. Improved positron stacking techniques give us large gains in the number of positrons. As we will see in the next chapters, the increased number of these $\overline{\mathrm{H}}$ constituents lead to an increased number of produced and trapped $\overline{\mathrm{H}}$. New instrumentation for a "rotating wall" and mode measurements allow for control of plasma geometry, critical for consistency between antihydrogen-trapping trials. Finally, ATRAP's recently demonstrated methods of centrifugal separation [12] and adiabatic cooling [13] were used in $\overline{\mathrm{H}}$-trapping experiments, and exemplify our good control and characterization of plasma temperature and geometry.

The experimental protocol was designed to minimize the experiment time per trapped antihydrogen trial, which is about 2 hours. An outline of the procedure will now be described, and details of each technique are provided in the following sections. For reference

to the pieces of apparatus mentioned here, see Fig. 2.1.

First, electrons are accumulated in the lower electrode stack (Sec. 5.1.1), transferred to the upper stack to cool and catch positrons, and then ejected out of the trap. After loading positrons (Sec. 5.3.1), the $\overline{\text{p}}$-solenoid is ramped up to prepare for $\overline{\text{p}}$-catching (Sec. 5.2.3). Electrons are again loaded into the lower stack to cool and catch $\overline{\text{p}}$. A rotating wall drive (Sec. 5.1.3) is applied to the electron plasma in order to set a radius optimal for $\overline{\text{p}}$-catching, and the radius is confirmed afterwards with plasma mode frequency measurements (Sec. 5.1.3). Antiprotons are then caught in the lower stack in the same potential well as the electrons, while positrons wait in a harmonic well in the upper stack. The rotating wall spins first the $\overline{\text{p}}$ to as tight a radius as possible, and then the positrons to a matching radius. The centrifugally-separated antiproton-electron plasma (Sec. 5.2.4) then allows efficient ejection of most of the electrons, while some embedded electrons (Sec. 5.2.5) remain to continue cooling the $\overline{\text{p}}$ plasma. Next the $\overline{\text{p}}$ solenoid is ramped down and the $\overline{\text{p}}$ are brought to the upper stack in proximity to the positrons. The end of the positron spin is timed to coincide with the arrival of the $\overline{\text{p}}$. When the $\overline{\text{p}}$ arrive, a deep nested e^+-$\overline{\text{p}}$ potential well is formed (Sec. 5.4), and the particles wait there while the Ioffe trap is ramped up. Just before the antiprotons and positrons are made to interact to form $\overline{\text{H}}$, the plasmas are adiabatically cooled (Sec. 5.5) by making the deep nested wells shallower.

Noise and chirped drives acting on the $\overline{\text{p}}$ plasma and potential well ramps of the positrons are then used to make the $\overline{\text{p}}$ and positrons interact. The low-field seeking $\overline{\text{H}}$ with small enough energy fall into the Ioffe trap. Once the $\overline{\text{H}}$ formation is complete, the electrode stack is swept in order to ramp all remaining charged particles out, and the Ioffe trap is purposefully quenched to release and count the trapped $\overline{\text{H}}$. These final steps of $\overline{\text{H}}$ formation, $\overline{\text{H}}$ trapping, and $\overline{\text{H}}$ release will be described in more detail in Chaps. 6 and 7.

5.1 Electrons

5.1.1 Electron loading

Electrons are vital to antihydrogen experiments for the purpose of cooling both antiprotons [45] and positrons. Photoelectrons are obtained [47] using 10 ns pulses of light produced from a 248 nm UV KrF excimer laser (GAM Laser EX-5/250) hitting the DEG electrode at the bottom of the electrode stack. The 1 Hz pulses reflect off a sequence of mirrors that lead them into the positron space, through the 1.5 mm positron window in the window plate, and through the electrode stack until they finally strike the Be foil of the degrader electrode. The work function of Be ranges from 3.6 eV to 5.08 eV, enabling the liberation of electrons by the $hc/\lambda = 5$ eV photons.

The number of electrons accumulated from excimer laser pulses is linear with number of pulses, and 12-22 pulses are typically needed to acquire 100 million e^- (when the $\overline{p}$-solenoid is on). The exact number depends on laser alignment and when the last KrF gas refill was done. Fig. 5.1 shows the electrode potentials applied to catch the liberated electrons. An initial positive voltage of 1 V is applied to the electrode LTE3 to create a shallow confining well for the electrons. LBE1 and LTE4 are negatively biased to form potential barriers. 2 μs after the excimer laser fires, the LBE1 front-door potential is pulsed upward as shown and the large 400 V positive voltage on TUBE accelerates the electrons towards the well in LTE3. The electrons are too energetic at this point to fall into the attractive well made by the TUBE electrode. The front door is then shut by bringing the pulser electode LBE1 back to its original blocking potential of -10 V. The electrons then bounce between LBE1 and LTE4 as they lose energy through synchrotron radiation (Sec. 2.2.1). The cyclotron and axial motions couple through collisions between electrons, causing energy loss in all directions.

Every pulse of electrons adds an increased space-charge potential, decreasing the well depth for subsequent pulses of electrons. To compensate for this, the well depth also gets incremented by 1 V with every pulse. This method achieved higher loading rates than starting out with a deep potential, possibly because a shallow well reduces the amount of kinetic energy that the electrons must lose to sink to the bottom of the well. The kinetic energy of the electrons finally reduces to $\ll$1 eV. At the end of electron loading, a much larger 100 V is applied to LTE3 to hold the electrons. This corresponds to a shallower on-axis holding well of 80 eV, as shown in Fig. 5.1b. The electrons can then either be counted, moved to the upper stack in preparation for positron loading, or prepared in the lower stack for $\overline{\text{p}}$ loading. The $\overline{\text{p}}$-solenoid surrounds the lower electrode stack region where both electrons and antiprotons are initially captured, and electrons can be loaded with the $\overline{\text{p}}$-solenoid on or off. Positrons are loaded with the solenoid off, and $\overline{\text{p}}$ are loaded with the solenoid on.

5.1.2 Electron ejection and counting

Before any experiments can be done, the number of electrons must be calibrated with the number of excimer laser pulses. This is done by loading a certain number of electron pulses as just described, and then releasing the electrons toward a conducting plate to count them. Typically throughout an 8-hour beam shift the calibration remains constant (5-10% reproducibility error), so this destructive counting is only done once at the beginning of a shift.

The electrons are initially in LTE3. They are counted by pulsing the voltage on LTE2 upwards as shown in Fig. 5.2, which releases a fraction of the highest-energy electrons. A positive voltage ramp towards the degrader accelerates the electrons toward the degrader, which serves as Faraday cup. The degrader is hooked up to a charge-sensitive amplifier,

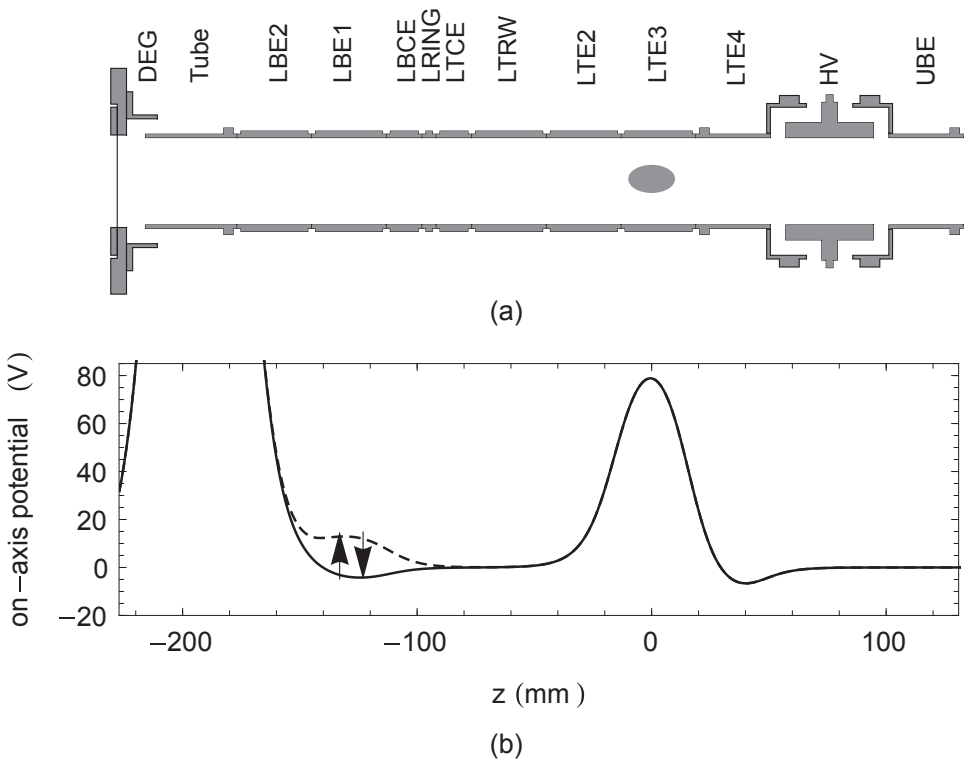


Figure 5.1: (a) Cross-sectional schematic of the lower electrode stack, showing the location of the accumulated electron plasma. Left to right electrodes in the schematic represent bottom to top in reality. (b) Electron loading potentials. A positive bias on the Tube electrode accelerates the electrons from the degrader into the lower stack, while a pulsed voltage on LBE1 allows electrons to enter or prevents them from escaping. A voltage on LTE4 forms the other end of the barrier, and electrons settle into LTE3.

and is biased to +100 V during this process in order to prevent the emission of secondary electrons which would provide a false count. After each pulse, the well depth of LTE3 is also lowered by 2.5 V to push the electrons closer to the brim of the potential. The electron counting must be done in concentrated pulses so that they can be detected above noise levels. However, the pulses must be small enough to prevent saturation of the preamplifier, which occurs for about 50 million electrons. The pulse-out happens until the on-axis potential on LTE3 is inverted. The number of electrons detected from each pulse as the potential is reduced from 80 V is shown in Fig. 5.2a. The pulse-out also gives information about

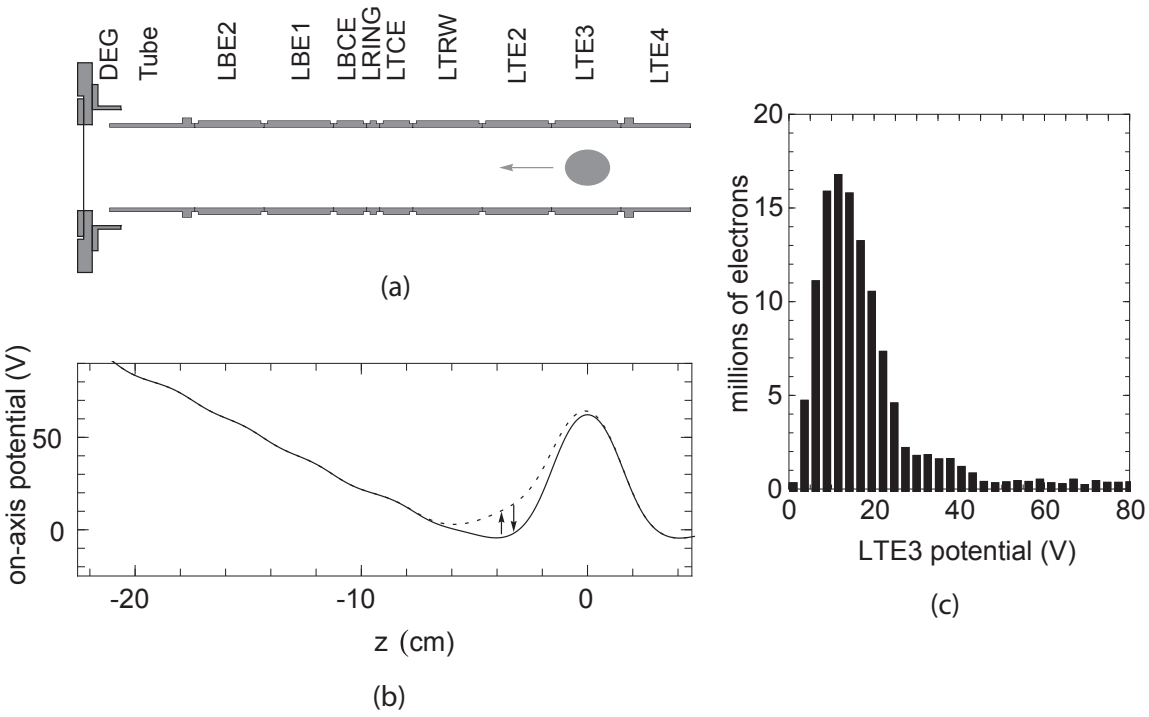


Figure 5.2: (a) Cross-sectional schematic of the lower electrode stack, showing the trajectory of the released electron plasma. Left to right electrodes in the schematic represent bottom to top in reality. (b) Electron counting potentials. The barrier on LTE2 is pulsed down to allow electrons to spill out. A positive voltage ramp accelerates the electrons towards the degrader where each pulse is counted with a charge-sensitive amplifier, leading to the distribution shown in (c).

the energy distribution of the electron plasma–for example, in the distribution shown the highest-energy electrons have around 45 eV. Knowing the number of electrons N in the electron plasma is necessary to make trials reproducible, as the electrons directly affect the $\overline{\mathrm{p}}$ radius and temperature. Knowing N can also be used to determine other properties of the plasma, as discussed in Sec. 4.2.2.

5.1.3 Electron preparation: rotating wall and mode measurements

Once the electron calibration has been done, we can set the plasma to a desired radius. The N electrons are loaded on LTE3 with an initial radius of 6-9 mm, and are given

30 seconds to equilibrate. The equilibrated plasma rotates with the E×B frequency ω_E. We then expand the well as shown in Fig. 5.3 and apply a rotating dipole potential to the 4-way split electrode LTRW [48] to spin up the plasma. This is done by applying sine waves that are 90° out of phase between adjacent quadrants. Applying a torque with frequency ω_{RW} that is larger than ω_E counters the natural expansion of the plasma (eventually causing particle loss) due to drag torques from small field imperfections. Since cylindrical symmetry is broken by the azimuthally asymmetric rotating field, the plasma is now free to compress (or expand, if $\omega_{RW}<\omega_E$).

When the drive frequency applied is much higher than $\omega_E/2\pi$, $\omega_r/2\pi$ increases along with the plasma density, and the plasma compresses as it reaches dynamical equilibrium as explained in Sec. 4.2.1. Therefore the rotating wall not only mitigates particle loss but also allows the plasma radius to be a controllable parameter that depends on the frequency, amplitude and duration of the drive. However, because drag forces also increase with density, the plasma rotation frequency never actually reaches the applied $\omega_{RW}/2\pi$. The rotation rate also depends on plasma temperature, but for our purposes this is fixed. The radii achievable for an electron plasma is anywhere between 2-8 mm. 2 mm is the limit we could reach with our plasmas, presumably because at 2 mm the electron plasma has already reached its torque-balanced state.

Fig. 5.3 shows the electrode potentials for a 3-electrode well structure, which was found to give the minimum plasma radius with minimal particle loss. Other schemes may be used to achieve larger radii. For example, to prepare electrons for $\overline{\text{p}}$ loading, a 2-electrode flat well is used to compress the electrons to a 6 mm radius. This larger radius is required to overlap with the incoming $\overline{\text{p}}$ beam, described in Sec. 5.2, and was found to optimize $\overline{\text{p}}$-catching. To obtain this radius, a 9 V p-p amplitude, 500 kHz, 240-300 s duration drive is used. The drive duration needed can vary depending on laser alignment and therefore

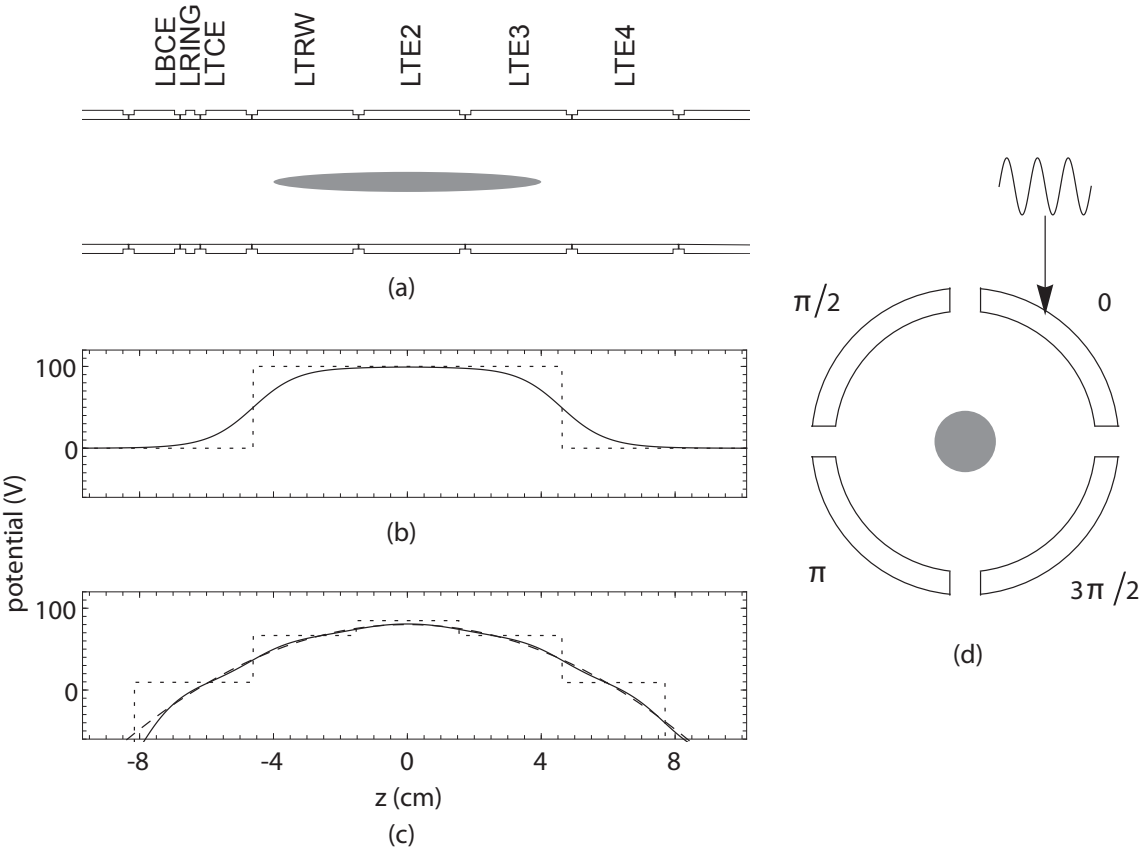


Figure 5.3: (a) Cross-sectional schematic of the lower electrode stack, showing the plasma shape in the rotating wall potential. Left to right electrodes in the schematic represent bottom to top in reality. (b) and (c): Rotating wall potentials for 3-electrode wells. Dashed lines represent applied electrode potentials and the solid lines are the resulting on-axis potentials for a (b) flat well and (c) harmonic well. The 3-electrode well was found to achieve the minimum radial compression (2 mm) with minimal particle loss. (d) Cross-sectional top view of the split-rotating wall electrode LTRW. The rotating drive is applied only to this electrode, spinning the entire plasma by its tail.

spatial emittance of the electrons when they come off the degrader.

The drive necessarily results in some plasma heating (though not as much as when electrons are pulsed out). Once the rotating wall drive has ended, the plasma is brought back to LTE3 and the plasma modes ω_z and ω_2 are measured non-destructively in order to determine the plasma geometry. The modes are measured using a pulsed-drive ring-

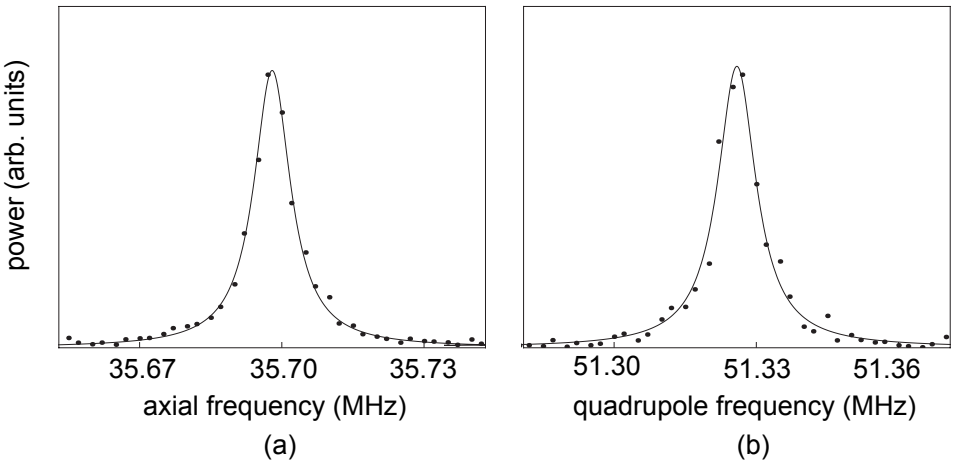


Figure 5.4: Plasma mode measurements. (a) Axial center-of-mass mode ω_z. (b) Quadrupole mode ω_2.

down technique. A 10 to 100 MHz frequency sweep is generated (Programmed Test Sources PTS 250) and gated by a pulse with 1 μs duration (Stanford Research System DG535, Mini-circuits ZYSWA-2-50DR) on a trap electrode adjacent to the plasma. The resulting image current induced on the opposite adjacent electrode is recorded. The plasma resonance frequencies are shown in Fig. 5.4. At times, the plasma density would be unusually low, and this would result in an inability to measure the modes. The frequencies ω_z and ω_2 are used along with the particle number N to calculate the plasma radius as described in Sec. 4.2.2. The rotation continues at the same frequency for hours, preserving the plasma shape and density, while the drag forces still act as a perturbation. Rotating wall drives can also be applied in the upper stack to the UTRW electrode for positron spinning, but the drives must be applied with the opposite rotation sense.

Electron plasma parameters used for antihydrogen trapping are shown in Table 5.1. Particle number N is measured as described previously and magnetic field B is known. Axial frequency $\omega_z/2\pi$ and quadrupole frequency ω_2 are measured. The rest of the parameters in the table are calculated from these inputs, giving a full characterization of the plasma.

After spinning, the electrons sit in their holding well, ready to catch $\overline{\mathrm{p}}$.

particle number N	10^8
magnetic field B	3.7 T
axial frequency $\omega_z/2\pi$	36.6 MHz
quadrupole frequency ω_2	49 MHz
radius ρ	6 mm
aspect ratio α	0.155
half-length z_p	0.93 mm
density n	$7{\times}10^7$ cm^{-3}
rotation frequency ω_r	27.5 kHz

Table 5.1: Electron plasma parameters.

5.2 Antiprotons

5.2.1 Antiproton steering

The pbar are injected into our trap from the bottom with 5.3 MeV. Silicon detectors and a Parallel Plate Avalanche Counter (PPAC) are used to monitor and steer the $\overline{\mathrm{p}}$ beam to be on-axis with the electrons. The silicon detector is located about 1 m away from the Penning trap and consists of a segmented 15 μm thick silicon diode that can be rotated in or out of the beam path. This detector is used for rough upstream steering, done by adjusting currents on AD beamline steering magnets. The four steering magnets closest to the experiment are called DHZ45, DHZ85, BVT10, and BVT25. DHZ45 and BVT10 are typically adjusted for centering the beam on the silicon detector, while DHZ85 and BVT25 are used for centering the beam on the PPAC for fine-tuning of the beam closer to the

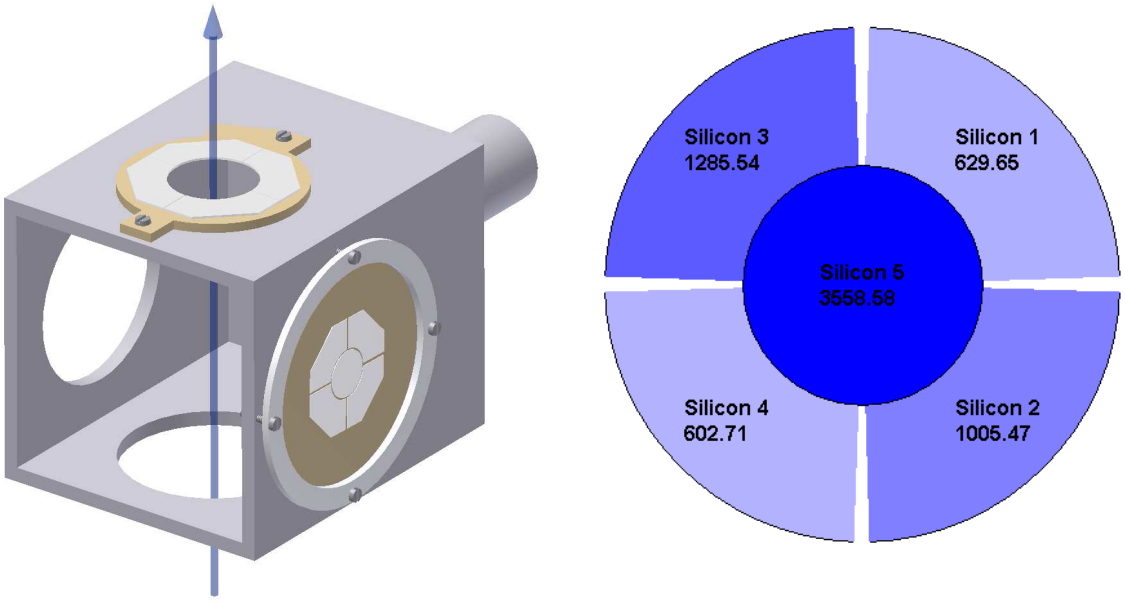


Figure 5.5: (a) Silicon detector. (b) Typical signals on the silicon detector for a centered $\overline{\mathrm{p}}$ beam.

experiment. Fig. 5.5 shows the silicon detector and an example of signals for a centered beam. These settings are stored for ATRAP and looking at the silicon detector is usually not necessary in a shift. Instead it is normally used when the beam emittance seems bad, in combination with the PPAC, to check if the beam is at an angle to the Penning trap. If the steering is particularly bad, one can also look at the last wire chamber before the experiment, called MW17.

The PPAC is used on a daily basis at the start of the shift to fine-tune the $\overline{\mathrm{p}}$ steering, and is located about halfway between the silicon detector and the Penning trap. It consists of 2 sets of anode plates rotated by 90 degrees with respect to each other, with a shared cathode as shown in Fig. 5.6. The plates correspond to either the X or Y direction. The anode plates have 5 aluminum strip electrodes with 2 mm width that are separated by 0.5 mm, while the cathode is an aluminized Mylar foil. The PPAC cell is filled with pure

argon at atmospheric pressure and as the antiprotons pass through they strip electrons from the argon atoms. The PPAC is operated in linear-type mode rather than avalanche mode, so the 150 V applied to the anodes is high enough to collect a measurable amount charge, but low enough not to produce an avalanche. The beam is centered when the middle anodes show maximum signals as shown in the Figure. The incident $\overline{\mathrm{p}}$ beam waist is seen to be $\sim$2 mm at the position of the PPAC.

5.2.2 Antiproton slowing

After passing through the $\sim$ 65 μm worth of Kapton and Mylar foils in the PPAC, the $\overline{\mathrm{p}}$ enter a 15 mm thick energy-tuning gas cell consisting of a mixture of helium and SF_6. The ratio of these gases can be used to change the $\overline{\mathrm{p}}$ energy by a maximum of 0.5 MeV from outside the apparatus in order to optimize trapping efficiency. Alternatively, 100% N_2 can be used in place of the optimal He-SF_6 mixture altogether.

As shown in Fig. 5.6a, the $\overline{\mathrm{p}}$ then proceed through a 10 μm titanium foil, which separates the insert dewar vacuum space from the magnet bore space at atmospheric pressure. It then passes through 6 layers of 6.4 μm of Mylar foil, which serves to insulate the insert dewar layers (which are at different temperatures) from each other. They then pass through another 10 μm titanium foil, which separates the insert dewar space from the experiment space. Finally, the $\overline{\mathrm{p}}$ pass through the 100 μm beryllium degrader, where most of their energy is reduced [49]. Antiprotons that emerge with less than 5 keV axial energy get confined over the length of the lower electrode stack.

5.2.3 Antiproton loading

The 2.7 T $\overline{\mathrm{p}}$-solenoid surrounding the lower stack is turned on for $\overline{\mathrm{p}}$ catching (and for loading electrons intended for $\overline{\mathrm{p}}$), and turned off when $\overline{\mathrm{p}}$ loading and preparation are

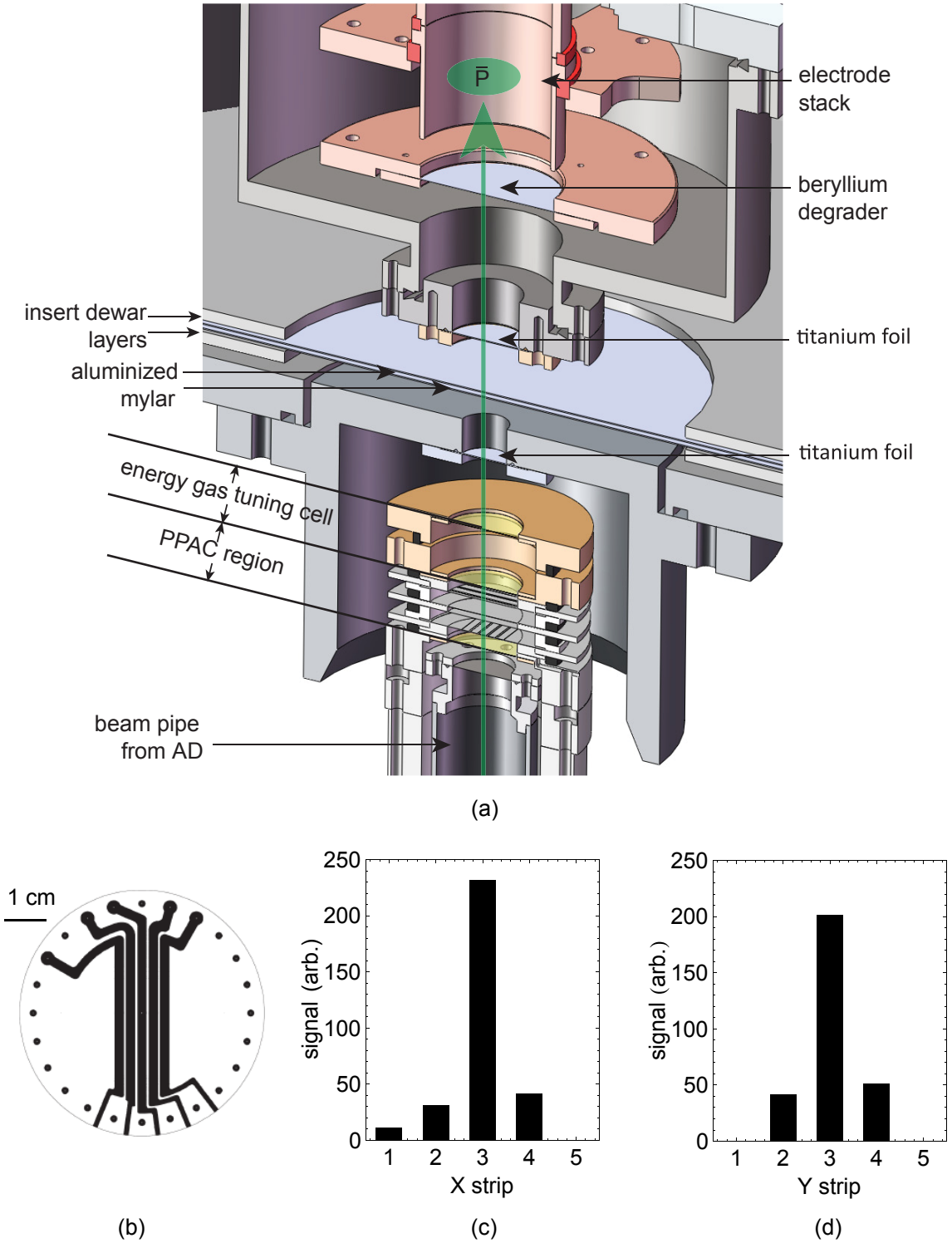


Figure 5.6: (a) Components used for steering and slowing the antiprotons. (b) Closer view of one of the anodes. Typical signals on the PPAC for (c) the X-direction and (d) the Y-direction.

done. The 2.7 T field adds to the 1 T bias field and provides a large gain in capture efficiency, due to the fact that the radius of the cyclotron orbit $r_c = \sqrt{2E_c m_p}/qB$ is reduced, where E_c is the radial energy. This allows more $\overline{\text{p}}$ to fit within the 18 mm radius of the lower stack electrodes. The potentials applied to catch $\overline{\text{p}}$ are shown in Fig. 5.7. To let the $\overline{\text{p}}$ in [8], the "front-door" degrader voltage is held at an attractive high positive value of 1.2 kV, while the "back-door" HV electrode is held at ~ -5 kV. The positive voltage on the degrader also serves the purpose of suppressing secondary electrons from being emitted and filling up the well on LTE3.

Pulses of $\overline{\text{p}}$ come in 90 second cycles consisting of an AD injection, prewarning signal, warning signal, and ejection into the experiment. An injection refers to when a high-energy $\overline{\text{p}}$ pulse has arrived in the AD ring. We use this pulse to synchronize our $\overline{\text{p}}$-loading electronics as well as positron pulses during positron loading. A prewarning pulse from the AD to our experiment arrives 4 seconds before the ejection. This pulse is used to inhibit the detector PMTs to prevent saturation. A warning pulse comes 4 μs before the ejection, and at this point the PPAC scopes get set up. 5 μs after ejection the front door of the trap is shut by application of -5 kV to the degrader. This 5 μs delay is called the HV switch delay. It gives enough time to capture the whole 200 ns $\overline{\text{p}}$ pulse and has been optimized for $\overline{\text{p}}$ catching efficiency (see Fig. 5.8). The $\overline{\text{p}}$ that were caught between the -5 kV barriers bounce between them and in the process collide with the cold well of 6 mm electrons sitting on LTE3, whose radius and density provide good spatial overlap with the now-wider 3.5 mm radius $\overline{\text{p}}$. This collisionally cools the $\overline{\text{p}}$ [45], as described in Sec. 4.3.1, and they settle into the same well as the electrons in LTE3. 4 seconds after the ejection, the detectors are enabled again.

60 seconds after the ejection trigger, the front door is opened and the uncaught high-energy $\overline{\text{p}}$ get dumped on the positively-biased degrader. This is called the HV dump,

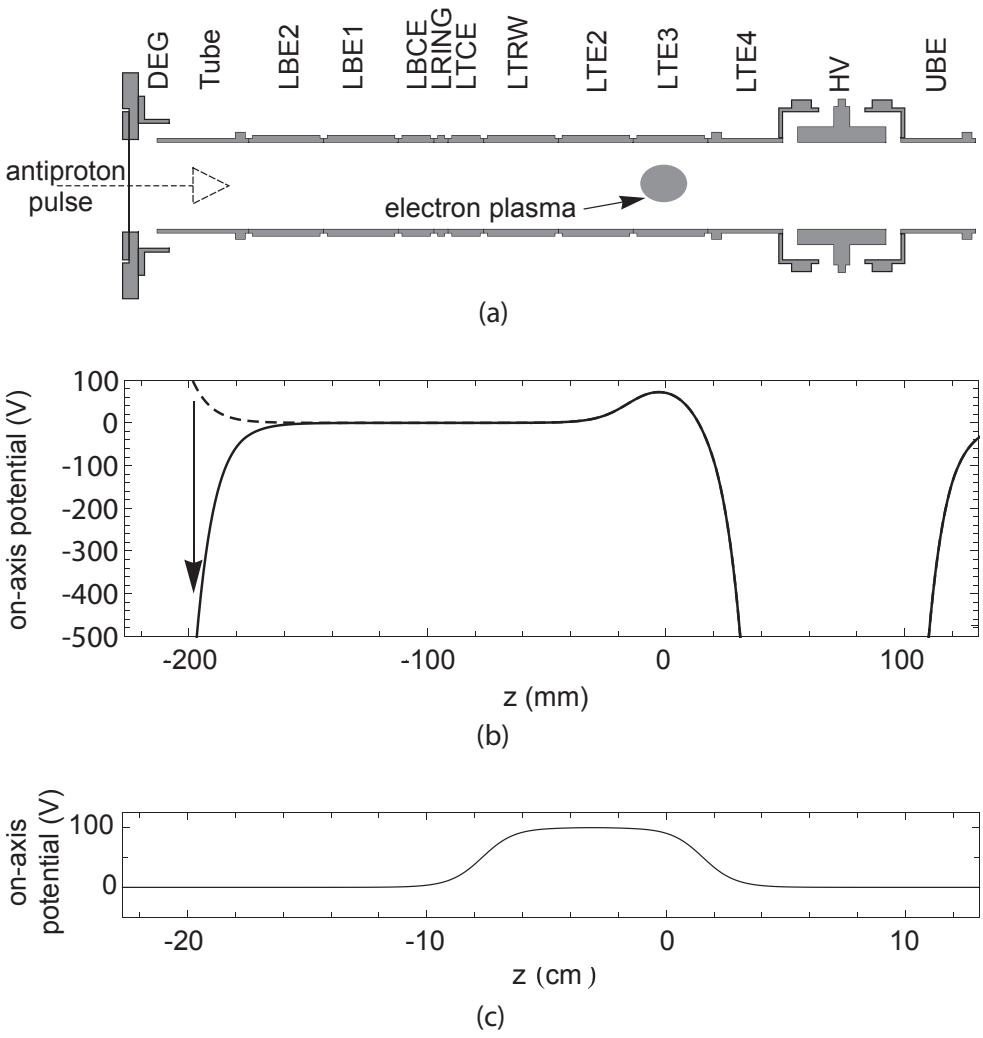


Figure 5.7: (a) Cross-sectional schematic of the lower electrode stack, showing the electron plasma waiting to catch $\overline{\text{p}}$. Left to right electrodes in the schematic represent bottom to top in reality. (b) Potentials for accepting (dashed line) and trapping (solid line) $\overline{\text{p}}$. Electrons are waiting in their 80 V on-axis potential in LTE3. (c) 3-electrode flat well for rotating wall drive on the $\overline{\text{p}}$-electron plasma.

and the 60 second delay is the HV ramp delay time. If the $\overline{\text{p}}$ signal measured by the detectors exceeds some threshold value, it counts as a shot (sometimes AD pulses are too weak to be useful). On a good day, 150k $\overline{\text{p}}$ per shot are trapped in the electron well, and each shot gives a HV dump of ∼20k counts for 6 mm catching electrons. The counts are displayed on the MCS-HSADC ramp readout, which shows all 4 of the MCS channel counts as discussed

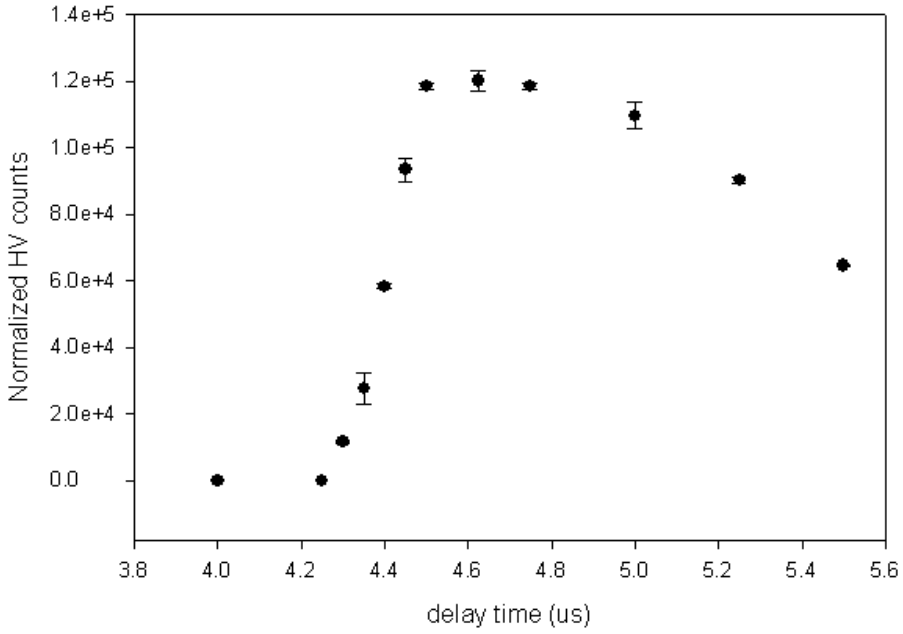


Figure 5.8: HV counts vs HV switch delay time.

in Sec. 2.5. The HV dump counts have previously been found to be directly proportional to the number of $\overline{\mathrm{p}}$ caught in the potential well, allowing a calibration between number of $\overline{\mathrm{p}}$ shots and number of $\overline{\mathrm{p}}$ caught. This allows one to adjust the number of shots needed on the fly, without the need for any destructive $\overline{\mathrm{p}}$ counting and calibration, which would be very time-consuming. Beam quality can be deduced at the start of a shift, even before $\overline{\mathrm{p}}$-catching, by checking the HV count during steering when there are no electrons. In this case a good HV count is $\sim$120k. Sometimes beam quality is poor and shots are missed altogether. This is not ideal, as this increases the time that the $\overline{\mathrm{p}}$ are sitting in their wells, expanding (as they have not been rotated yet).

About 30 seconds after the HV dump, another ejection pulse of $\overline{\mathrm{p}}$ arrives, and the cycle starts again. For $\overline{\mathrm{H}}$ trapping experiments we stack 6-9 shots of $\overline{\mathrm{p}}$ [50], depending on beam quality, in order to capture 1 million antiprotons in the 100 million electron well. By

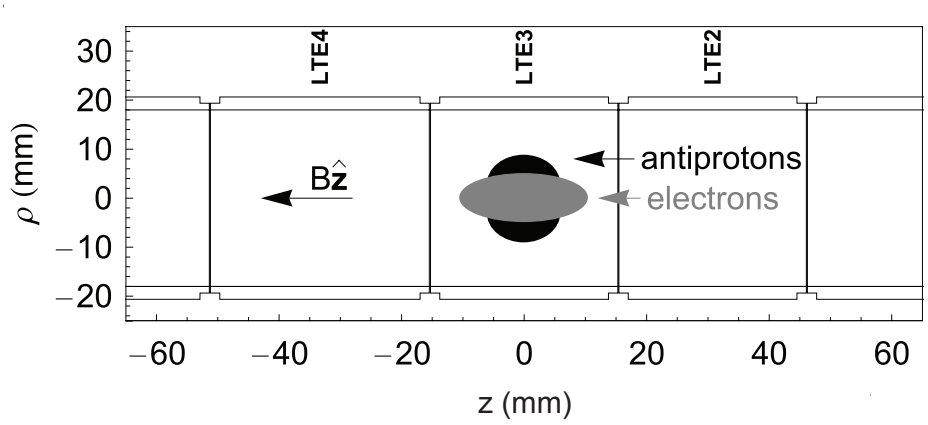


Figure 5.9: Cross-sectional schematic of the lower electrode stack, showing centrifugal separation of antiprotons and electrons. Left to right electrodes in the schematic represent top to bottom in reality.

the end of $\overline{\mathrm{p}}$ loading, the $\overline{\mathrm{p}}$ have cooled to ∼1 meV, 9 orders of magnitude colder than when they first arrived at the trap. At this point, the multispecies electron-$\overline{\mathrm{p}}$ plasma centrifugally separates as it comes to thermal equilibrium, which is believed to make a later step of electron removal from the plasma more effective and lead to less heating. We will pause now in the $\overline{\mathrm{H}}$ trapping protocol to describe the first observation of centrifugal separation between antiprotons and electrons.

5.2.4 Observed centrifugal separation of antiprotons and electrons

We recently observed the first centrifugal separation of antiprotons and electrons [12], whose theoretical basis has been explained in Sec. 4.3.2. With our experimental parameters of N_e=10^8, $n_e \approx 10^8$, and B=1-3.7 T, we showed that our $\overline{\mathrm{p}}$ cloud surrounds the electrons as shown in Fig. 5.9. This was determined in two different ways: reducing the axial potential which spills out the innermost particles first, and reducing the radial magnetic field confinement which causes loss of the outermost particles first.

Axial potential reduction

Using the methods just described, 100 million electrons were prepared with a radius of 5 mm and held in an 80 V on-axis well on LTE3. 1 million $\overline{\mathrm{p}}$ were then loaded into the same well. After 2 minutes of equilibration time, sufficient to allow radial transport of $\overline{\mathrm{p}}$ across the B-field [51], the voltage on LTE3 was reduced by either 2-V steps or by adiabatic ramping. Electrons were observed to spill out first in either situation, as shown in Fig. 5.10. Because the ρ=0 axial potential is the shallowest axial potential (see for example Fig. 4.1), this indicates that the electrons were concentrated around ρ=0 while the $\overline{\mathrm{p}}$ are not. The results do not change when the time over which the potential was reduced was increased by a factor 100, and when delays were added between voltage steps. This indicates that the electrons did not simply escape first because they are lighter–they are actually in a different axial potential well region (central).

Magnetic field reduction

The converse situation was also studied. The particles were loaded and prepared in a similar way, but electron radius was varied from trial to trial. After a 2 minute thermalization time between electrons and $\overline{\mathrm{p}}$ the magnetic field was reduced from 3.7 T to 1 T. By the confinement theorem in Eq. 4.7, this causes the total 2-component plasma to expand. $\overline{\mathrm{p}}$ were observed to annihilate on the 18 mm radius trap walls at different B-field values depending on electron radius. Since ΔB is known and the final $\overline{\mathrm{p}}$ radius is known (18 mm), the initial $\overline{\mathrm{p}}$ radius can be determined. Fig. 5.11 shows that the initial $\overline{\mathrm{p}}$ radii were always found to be larger than the initial electron radii, which were determined by mode measurements. Additionally, no electrons were lost in any of the plasma expansion studies. This is due to a resolution limit: the B-field could not be set to less than 1 T, and at this field the electrons are at their maximum radius of $\sim$15 mm, less than the 18 mm electrode radius.

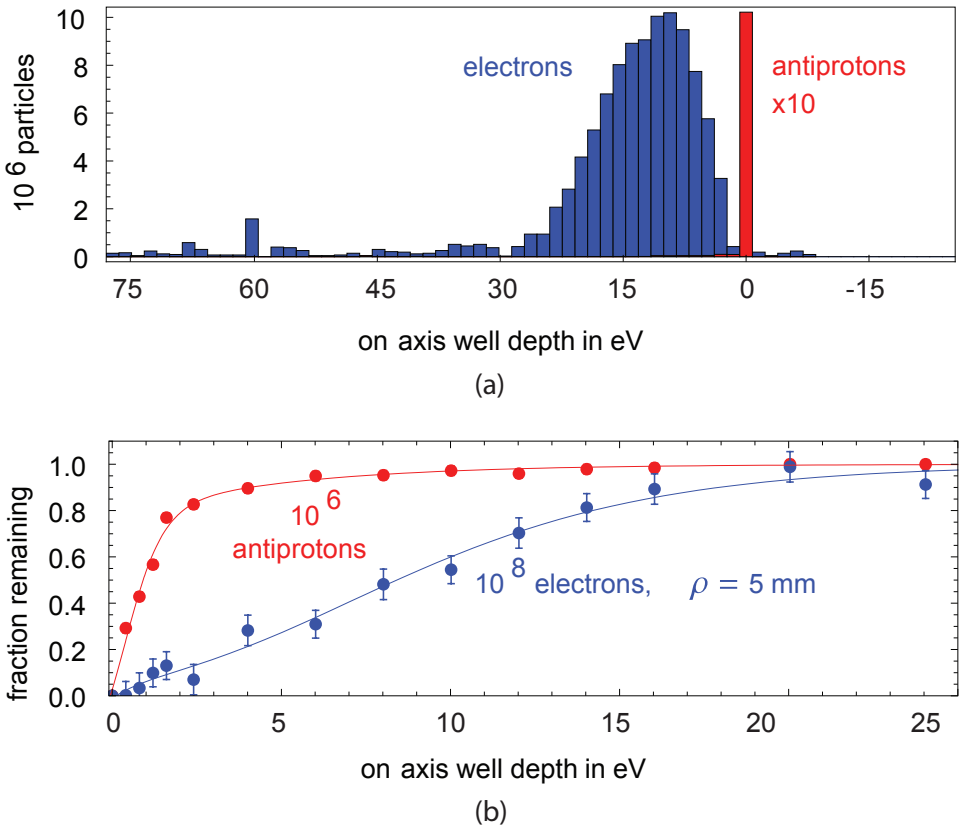


Figure 5.10: Centrifugal separation detection using axial reduction. (a) Number of particles that spill when the well voltage is reduced non-adiabatically in 2-V steps. (b) Fraction of particles remaining in the well when the well voltage is ramped adiabatically.

Discussion

These results demonstrate our good control over the plasmas and indicate that we do indeed have a low temperature plasma, since the separation could not have been observed otherwise. The upper temperature limit of the $\overline{\mathrm{p}}$ plasmas can be deduced by equating the centrifugal separation energy with thermal energy in Eq. 4.12. We find that for our parameters, the $\overline{\mathrm{p}}$ temperature must have been below $T_{sep} \approx 75$ K. We normally pulse out the electrons once we have used them to catch $\overline{\mathrm{p}}$, which causes some of the $\overline{\mathrm{p}}$ to accelerate towards the center axis to fill the void. This heats the 75 K $\overline{\mathrm{p}}$ to up to

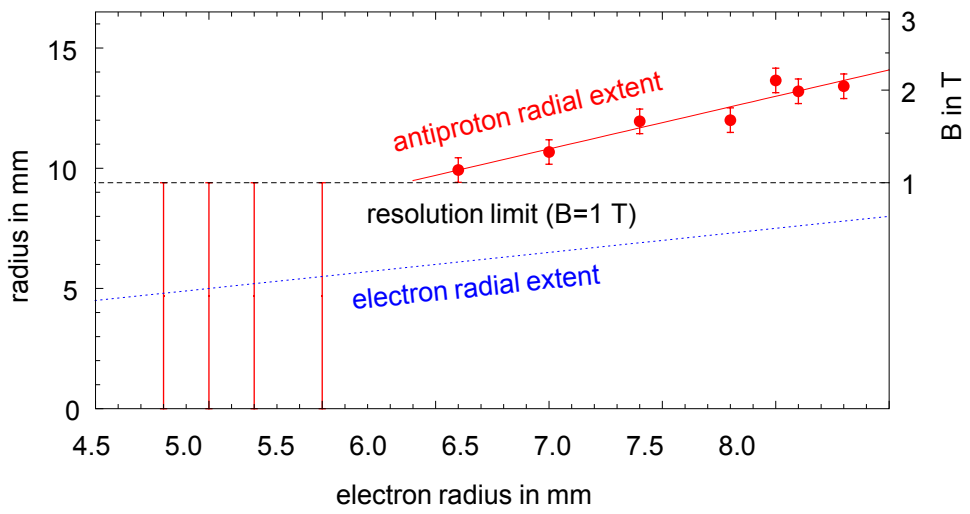


Figure 5.11: Centrifugal separation detection using magnetic field reduction. The outer initial plasma radius, deduced from the B-field at which the first annihilations were seen, is shown as a function of initial electron radius. No annihilations were seen for the 4 measurements on the left, indicating that the $\overline{\mathrm{p}}$ outer radius was less than 9.4 mm.

thousands of Kelvin. Centrifugal separation can make this pulsing process more efficient, so fewer pulses are required. Alternatively, the separation should also enable us to produce controlled fractions of $\overline{\mathrm{p}}$ and electrons as shown in Fig. 5.10 by gradual ramping out some of the electrons. This has not been implemented yet in our $\overline{\mathrm{H}}$ trapping experiments.

There is also a somewhat negative consequence of centrifugal separation. When the electrons are pulsed out, some $\overline{\mathrm{p}}$ move inwards. However, to conserve angular momentum and preserve the mean square radius (Sec. 4.2.1), some of the $\overline{\mathrm{p}}$ must also move radially outward. The final radius of the $\overline{\mathrm{p}}$ after electron pulseout is given by

$$\rho_f = \sqrt{\frac{5}{4}(\rho_1^2 + \rho_2^2)} \tag{5.1}$$

where ρ_1 and ρ_2 are the inner and outer initial radii, respectively. This causes the problem that the outer $\overline{\mathrm{p}}$ radius actually expands after pulseout, whereas we want to keep the $\overline{\mathrm{p}}$ radius as small as possible. This problem is dealt with by making the $\overline{\mathrm{p}}$-electron plasmas as radially small as possible before ejection, described next.

5.2.5 Antiproton preparation with a rotating wall, electron pulseout, and embedded electron cooling

Radial expansion of plasmas can occur over the course of an experiment in several ways. These include heating from electron ejection, noisy electrodes, or movement up and down the electrode stack (especially through the HV electrodes). Radial expansion can also result from turning off the $\overline{\text{p}}$-solenoid, drag torques, or the quadrupole field. Therefore, before removing the electrons from the $\overline{\text{p}}$-electron plasma, we spin the entire 2-component plasma before electron pulse-out, to make the initial $\overline{\text{p}}$ radius as small as possible. We continue now with $\overline{\text{H}}$-trapping methodology to explain how this is done.

After the $\overline{\text{p}}$ have been caught in LTE3 with the electrons, the $\overline{\text{p}}$ and electrons are then together moved into a 3-electrode rotating wall (Fig. 5.7c) and spun for 783 seconds at 1 MHz and 9 V p-p amplitude which causes them to radially compress. They are given 2 minutes to equilibrate and separate. The plasmas are then returned to LTE3.

The radius of the spheroidal electron plasma can be measured again using mode measurements. This gives accurate results even in the presence of the $\overline{\text{p}}$ since the 2 order of magnitude difference in electron and antiproton number makes the antiprotons a small perturbation on the electrons. The electron radius is found to be 2 mm after the rotating wall. Since the $\overline{\text{p}}$ form a shell around the electron plasma after the rotating wall, they are also assumed to be around 2 mm in radius, though this cannot be directly measured.

After the electrons have been used to catch and prepare $\overline{\text{p}}$, they actually become a hindrance if they are kept in the $\overline{\text{p}}$ well. They increase the space-charge potential, requiring deeper wells for manipulation. Their presence causes most of the centrifugally separated $\overline{\text{p}}$ to be located on larger radii, which is undesirable. They also make it impossible to perform operations directly on the $\overline{\text{p}}$, such as measuring their temperature. Additionally, the e^- can interfere with $\overline{\text{H}}$ production during three-body recombination [52] (discussed in Sec. 6.1),

potentially stripping the e^+ from weakly-bound $\overline{H}$. Therefore, at the end of the spin and thermalization we pulse out the electrons in a sequence of 4 pulses, which leaves the $\overline{p}$ at a radius of just over 3 mm (estimated from Eq. 5.1). After the pulse-out ~1000 e^- embedded in the $\overline{p}$ plasma are left behind. As mentioned in Sec. 5.2.4, the pulse-out undesirably heats the $\overline{p}$ up to ~1000 K.

The embedded electrons however, cool the $\overline{p}$ back down, although with a much longer time constant. After pulsing out the electrons, we wait at least 10 minutes to let the $\overline{p}$ cool from the embedded electrons. In this time embedded electrons cool the $\overline{p}$ to the baseline electrode stack temperature of 10-30 K. To find the thermalization time, we can use the rate equations 4.11. For electron number $N_e \ll N_p$, the $\overline{p}$ number, the two plasma components have the same temperature T as long as the time between $\overline{p}$-electron collisions $\tau_{col} \ll \tau_{\overline{p}}$, the $\overline{p}$ cooling time from embedded electrons. $\tau_{\overline{p}}$ can be expressed as $\tau_c \frac{N_p}{N_e}$. Then the rate equations simplify to

$$\frac{d}{dt}T = -\frac{1}{\tau_{\overline{p}}}(T - T_i)$$

where T_i is the steady-state temperature for $t \gg \tau_{\overline{p}}$. The assumption that $\tau_{col} \ll \tau_{\overline{p}}$ is satisfied by 6 orders of magnitude in the best case (zero B-field limit) and 3 orders of magnitude in the worse case (strong B-field and suppressed isotropization rate). For our electron and $\overline{p}$ numbers the thermalization time for embedded electrons is 34 s.

Later we cool the $\overline{p}$ further to below T_i using adiabatic cooling (Sec. 5.5). If the $\overline{p}$ are below T_i, the embedded electrons are detrimental–they tend to heat the $\overline{p}$ back up to T_i by thermalizing with the trap electrodes. Therefore the number of electrons chosen to leave behind must strike a balance between initial cooling and later heating. In order to have the $\overline{p}$ be as cold as possible when they interact with the e^+ to form $\overline{H}$, the $\overline{H}$ formation process must take place immediately after adiabatic cooling, before the embedded electrons have a chance to heat the $\overline{p}$ up again.

5.3 Positrons

5.3.1 Positron loading, counting and preparation

Positron production is described in Sec. 2.1. Every 30 seconds, a bunch of positrons with 60 eV enter the trap, 2 μs after they are ejected from the positron accumulator. The e^+ trapping potentials are shown in Fig. 5.12. First, electrons are accumulated on LTE3 as usual, but not spun up. This is done with the $\overline{p}$-solenoid off, for two reasons. After accumulation these electrons are moved to the upper stack to catch positrons, and if the solenoid were on, its fringing field would cause particle loss. Additionally, positron steering and accumulation have been found to be optimized with the $\overline{p}$-solenoid off. In this configuration, typically 44 shots of electrons are required to accumulate 150 million electrons. The electrons are then moved to UTR8.

In each positron shot cycle there is a constant back-door barrier on UBR4 and the front-door barrier on UTE2 is pulsed down to let the positrons in. The e^+ thermalize with the e^- and fall into the surrounding wells UTR7 and UTRW in about 1 second. Although e^+ can synchrotron-cool themselves just like the e^-, this cooling time goes as B^{-2}, and the positrons are only in a 1 T field in the upper stack as opposed to the 3.7 T typically used for electron loading for $\overline{p}$. Without electrons, it would take $>$100 s for the e^+ to convert their axial energy into cyclotron energy, which would then be synchrotron radiated away. Electron cooling of low-energy positrons in a Penning trap was first demonstrated at ATRAP and works very efficiently: 85% of the incoming e^+ get accumulated in the well. Additionally, the production of positronium and e^+-e^- annihilation are suppressed at high e^+ initial energies. As the e^+ cool down, they fall into deep wells which ensures that they are out of contact with the e^-. Once the desired number of shots are taken, after a 10 s thermalization time the electrons are smoothly ejected by decreasing the potential on UTR8.

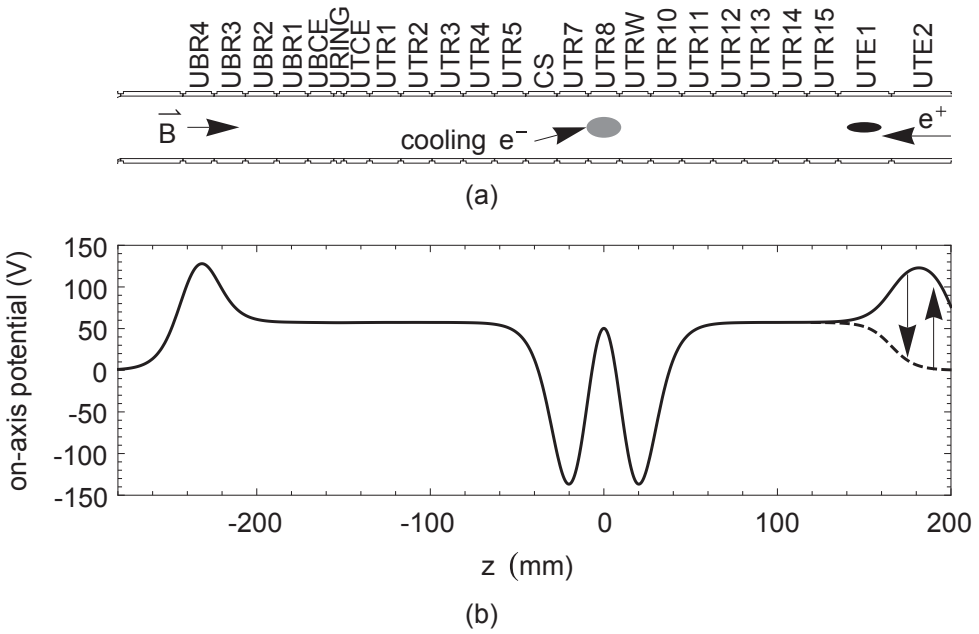


Figure 5.12: (a) Cross-sectional schematic of the upper electrode stack, showing the electron plasma waiting to catch e^+. Left to right electrodes in the schematic represent bottom to top in reality. (b) Positron loading potentials. Electrons wait in UTR8. The front-door on UTE2 is pulsed to let positrons in. As the positrons thermalize they settle into the nested well surrounding the electrons. After accumulation the electrons are ejected and the double positron potential is combined into one at UTR8.

This also brings the positrons together on a single electrode well on UTR8. They are now able to cool themselves from synchrotron radiation and collisions.

Just like the electrons, positrons must be initially calibrated with number of shots by destructive charge-counting at the degrader, before being loaded for experiments (though for positrons, this method requires some background compensation due to the creation of secondary annihilation electrons). Typically 3 million e^+ per shot are trapped. Because the 90 million positrons used for $\overline{H}$ experiments can take up to 15 minutes to load the calibration is typically done with only 9 shots, and the number of shots required for the experiment is linearly extrapolated. Up to billions of e^+ can be stacked in the Penning-Ioffe trap, the largest numbers accumulated for any system studying $\overline{H}$.

Once the positron calibration and loading is done, the $\overline{p}$-solenoid is ramped back up for $\overline{p}$ catching. The positrons are moved into a 5-electrode harmonic well centered on CS to wait while the $\overline{p}$ are caught and prepared. The harmonic well is used to prevent e^+ from expanding during this wait time.

When the $\overline{p}$ spin is done, the rotating wall drive is then switched to the upper stack to spin the positrons. The e^+ are moved to a 5-electrode flat well where a rotating wall drive on UTRW is applied at 200 kHz for 800 s. Positrons are initially loaded with 8-10 mm radius, and mode measurements after the spin reveal a radius of 2-3 mm. The radius is made as small as possible for retention in the quadrupole field later on. The positron radius also roughly matches the final $\overline{p}$ radius. This to maximize interaction with the $\overline{p}$ later on for positron-cooling of $\overline{p}$ and $\overline{H}$ formation. While the positrons are spun in the upper stack, electrons are being ejected from the $\overline{p}$ in the lower stack, followed by embedded electron cooling, followed by a 6-minute $\overline{p}$-solenoid rampdown, followed by the $\overline{p}$ adiabatically making their way up to the lower stack to meet the positrons. The end of the positron rotating wall and mode measurements coincide with arrival of the $\overline{p}$ in the upper

stack. Due to the reduction in B-field upon the $\overline{\text{p}}$-solenoid rampdown, the $\overline{\text{p}}$ radius increases by about a factor of 2, followed by a reduction again at a later step in the presence of the Ioffe pinch field.

5.4 Nested Well and Preparation for $\overline{\text{H}}$ Formation

Since the antiprotons and positrons must be in proximity in order to interact, after the positrons are spun up, they are moved to UTR8 while $\overline{\text{p}}$ are brought from LTE3 to UTR3 adiabatically in a 50 V transfer well. The positrons are then moved to the CS electrode, the center of the Ioffe trap. A deep nested well [53] is formed around the centered positrons from UTR5, CS, and UTR7 (see Fig. 5.13a), and the potential on UTR3 is then raised, spilling the $\overline{\text{p}}$ into UTR5. The wells are made deep to limit the charged particles' axial extent, in anticipation of radial loss from the Ioffe field (a problem described in Secs. 2.3 and 6.5). The Ioffe quadrupole and pinch coils are then ramped up to full field at 69 A and 80 A, respectively, and make a 375 mK trap (for ground state $\overline{\text{H}}$) with equipotentials shown in Fig. 5.13b. Just before $\overline{\text{p}}$-e^+ mixing begins, the nested well potentials are swept to shallower values in Fig. 5.13c for adiabatic cooling and $\overline{\text{H}}$ formation. Additionally, the number of positrons loaded on the CS well are such that the positrons saturated the well once made shallower. The signature of positrons being filled to the brim of the potential (at the same energy level as the $\overline{\text{p}}$) was a sudden loss of positrons immediately after setting the final well potentials. The final nested well potentials and number of positrons were chosen based on $\overline{\text{H}}$ detection well studies that will be described in the next chapter.

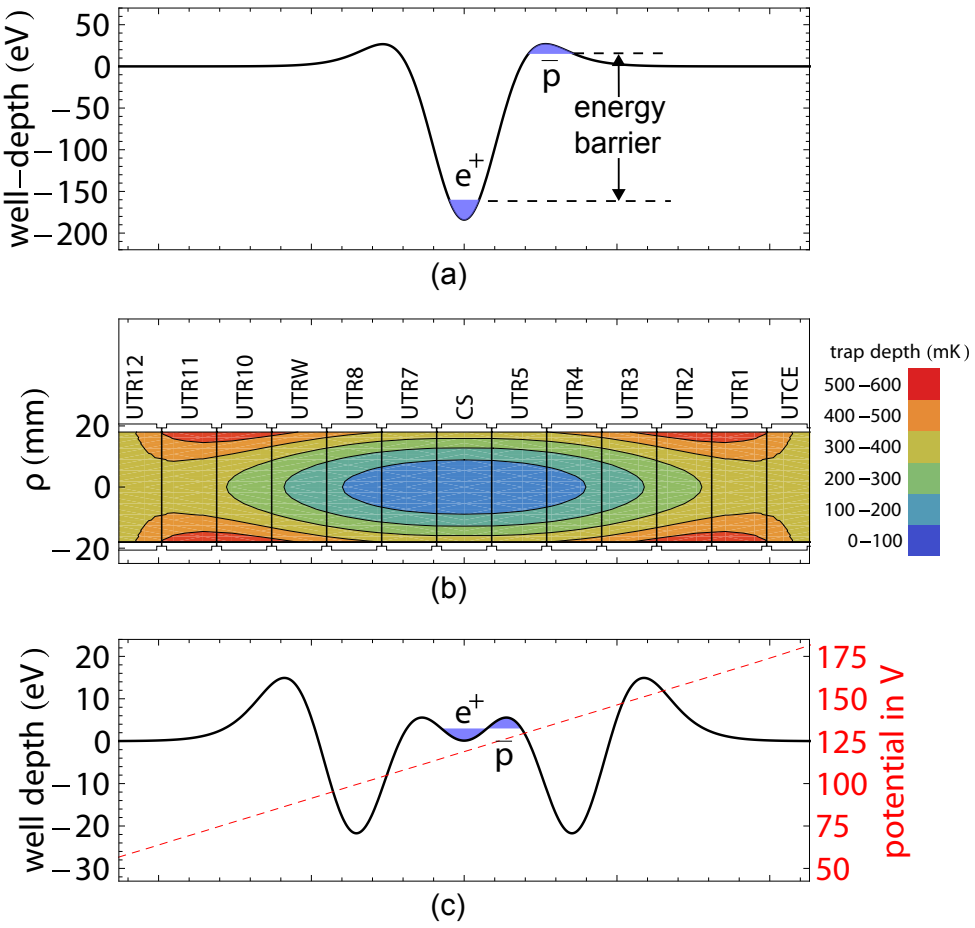


Figure 5.13: (a) Deep nested well for holding particles before the Ioffe field is ramped up. (b) Ioffe trap potential at full field. (c) Once the Ioffe field is on the deep well is made shallower for adiabatic cooling and $\overline{\mathrm{H}}$ formation.

5.5 Adiabatic cooling

Immediately after the Ioffe trap has been ramped to full field and just before the positrons and antiprotons are driven to interact, the deep nested well potentials are made shallower in order to adiabatically cool the charged particles from about 30 K to 3.5 K, well below the embedded electron cooling limit (Sec. 5.2.5). As described in Sec. 6.1, cold constituent particles are highly desirable because the three-body $\overline{\mathrm{H}}$ formation rate sharply increases with inverse temperature. Cold antiprotons are also necessary to attain cold,

trappable antihydrogen atoms.

We have recently demonstrated the largest numbers (3×10^6) of the coldest (3.5 ± 0.7 K) trapped $\overline{\mathrm{p}}$ yet, using the new method of adiabatic cooling [13]. Because of the enormous effort needed to create and trap rare antiprotons, adiabatic cooling has a big advantage over evaporative cooling in that it is lossless. We will now pause again in the $\overline{\mathrm{H}}$ trapping protocol to describe this experimental result.

Adiabatic cooling principle

Adiabatic cooling is achieved when the potential well depth is adiabatically decreased. This decreases the frequency f (the empty well axial frequency characterizing the trap) and thus the energy E. The adiabatic condition is satisfied when the change in frequency during an oscillation period is much less than the frequency, or when the final plasma temperature is independent of the rate of change of f. In this case, if f_i is reduced to f_f, then T_i reduces to T_f. This process is reversible, since there is no change in entropy.

Comparisons with an ideal gas can be made if the inter-particle spacing in a plasma is larger than the distance of closest approach (found from equating electrostatic energy with thermal energy). In this case, inside the plasma, the free particles move and collide in a similar way to an ideal gas. As the potential decreases, the plasma does work on the potential walls to increase its volume V. Since $TV^{\frac{2}{3}}$ of an ideal gas is constant for reversible, adiabatic processes, the temperature of the plasma must decrease.

Other point of views and predictions can also be considered. If we treat the plasma as a particle oscillator, then we can use the familiar adiabatic invariant E/ω to find the relationship $T_f = (f_f/f_i)T_i$. If coupled oscillatory motion and the heat capacity of the $\overline{\mathrm{p}}$ is taken into account, the prediction for an individual particle becomes $T_f = (f_f/f_i)^{1/2}T_i$.

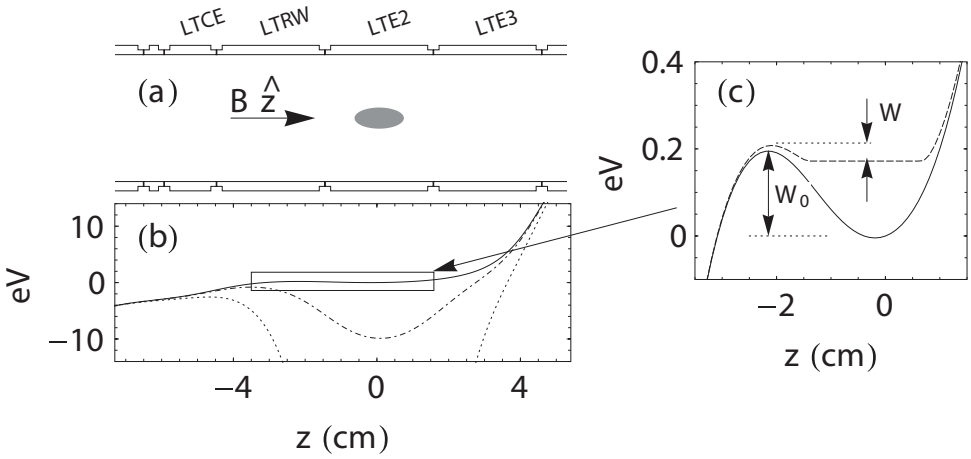


Figure 5.14: (a) Cross section of the trap electrodes with the plasma on LTE2. (b) On-axis potential energies for well depths W_0 of 0.2, 9, and 77 eV. (c) Expanded view showing the well depth for an empty trap (solid curve) W_0 and the space-charge modified well depth W (dashed curve) for 5×10^5 antiprotons. Taken from [13].

Adiabatic cooling experiment

To demonstrate adiabatic cooling, we load various numbers of antiprotons on LTE2, and pulse out the catching electrons with either 3 or 4 pulses, leaving behind either 6 000 or 900 embedded electrons. After embedded electron cooling takes place, adiabatic cooling begins with a f_i of between 3 MHz and 90 kHz, corresponding to an on-axis well depth of between 800 and 0.4 eV. The frequency of the potential is then lowered to f_f, which is also varied, over a time duration of hundreds of ms. The final temperature was found to be independent of $\dot{f}$ when the duration was varied by a factor of 5. Embedded electron cooling has a negligible effect on this timescale since embedded electron cooling takes place over tens of seconds. The location of the plasma and potential energies are shown in Fig. 5.14.

The final temperature of the plasma is measured immediately after the adiabatic cooling by reducing the potential well further until the $\overline{\mathrm{p}}$ spill out. The energies of the first few thousand $\overline{\mathrm{p}}$ correspond to the tail of a Boltzmann distribution, and this can be fit in

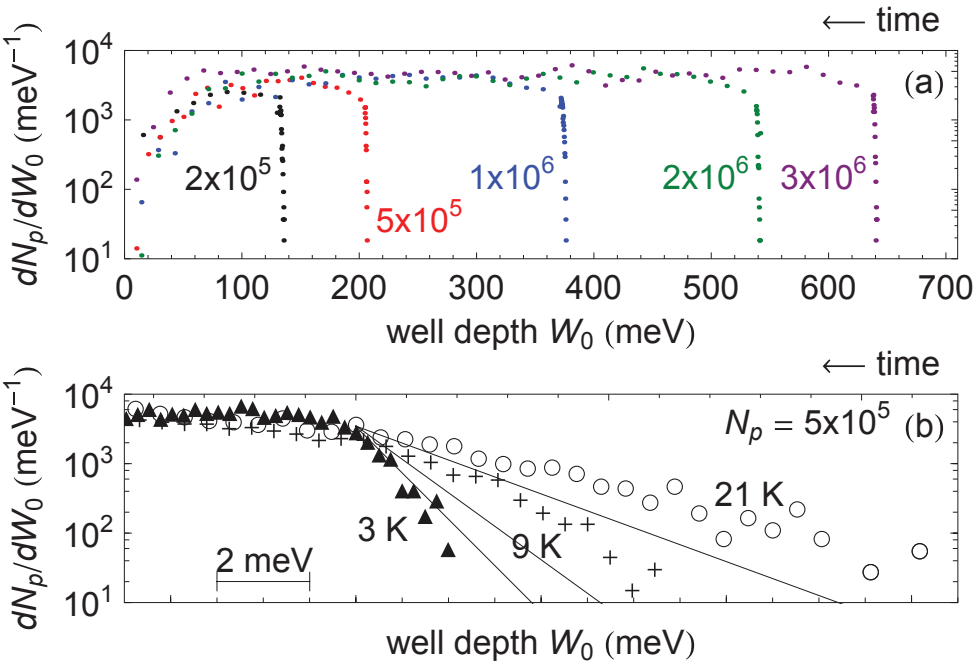


Figure 5.15: Temperature measurements. (a) $\overline{\text{p}}$ loss spectra for various numbers of $\overline{\text{p}}$ as the well depth W_0 is reduced linearly in time. (b) An exponential fit to the tail end of the distribution is used to determine the final temperature T_f.

order to extract the temperature. It is important to note that we cannot use the energy distribution of all the $\overline{\text{p}}$, as the plasma's space-charge energy continuously adjusts as the $\overline{\text{p}}$ are being spilled. The first few thousand lost $\overline{\text{p}}$ only make a negligible difference in the space-charge. Fig. 5.15 shows how temperature measurements are done. The rate of number of $\overline{\text{p}}$ $dN_{\overline{\text{p}}}$ escaping the well as the well depth is reduced by dW_0 is measured as a function of the well depth W_0. Fig. 5.15a shows the resulting $\overline{\text{p}}$ loss spectra for different $N_{\overline{\text{p}}}$. The sharp nearly-vertical lines represent the first few, highest-energy $\overline{\text{p}}$ that spill out. The exponential slope is determined from the first thousand of these, as in Fig. 5.15b, to determine the temperature.

The results of the adiabatic cooling temperature measurements are shown in Fig. 5.16. The initial and final well frequencies are varied and the temperature is shown as a function of both f_i and f_i/f_f. The lowest temperature observed was 3.5 ± 0.7 K. Why even

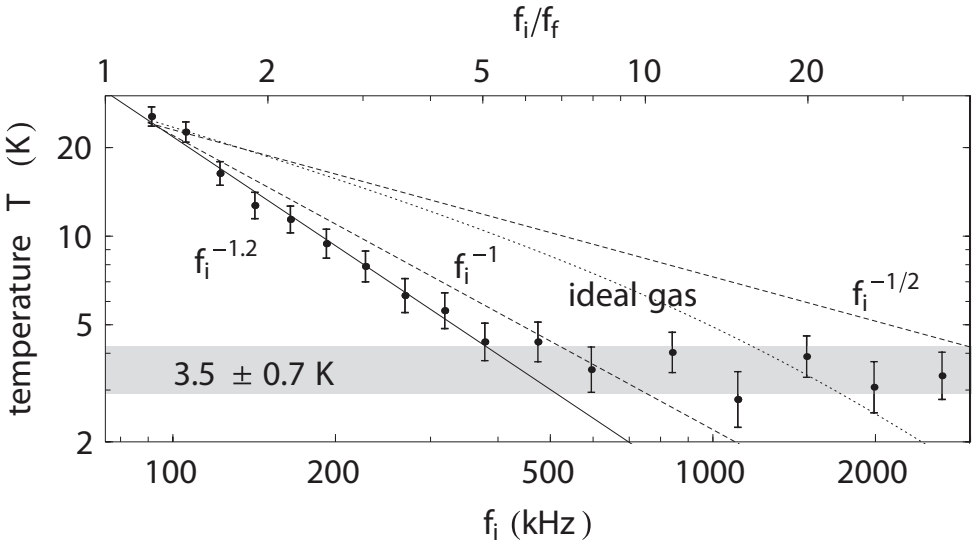


Figure 5.16: Measured temperatures of 5×10^5 $\overline{\mathrm{p}}$ after adiabatic cooling. The best fit power law is found to be $f_i^{-1.2}$. Predictions for an ideal gas and adiabatic invariants of an oscillator are also plotted for comparison. The measured temperatures are lower than the predicted temperatures.

lower temperatures were not observed is not understood. One possibility is that we simply hit a measurement limit given our methods and apparatus. In this case, if we extrapolate T from the best-fit power law $f_i^{-1.2}$, the actual temperature could be as low as 0.4 K. Another possibility is that we have a noise source (as of yet unidentified) which prevents the $\overline{\mathrm{p}}$ from cooling any further. It is also feasible that the measurements at higher f_i are indeed correct and a better theoretical understanding is needed to fit the curve. The prediction from the ideal gas law is shown in the Figure. Our $\overline{\mathrm{p}}$ plasma is not in fact an ideal gas due to the plasma boundary, where the plasma density decays on the scale of a Debye length. Also shown are the predictions arising from the adiabatic invariants of an oscillator. The measured temperatures are lower than all the predictions.

Fig. 5.17 shows the effect of embedded electron *heating* of the plasma after the adiabatic cooling has been performed. The electrons come back into equilibrium with the trap electrodes, and thermalize with the $\overline{\mathrm{p}}$ in a time $\tau_{\overline{\mathrm{p}}}$, defined in Sec. 5.2.5. The equi-

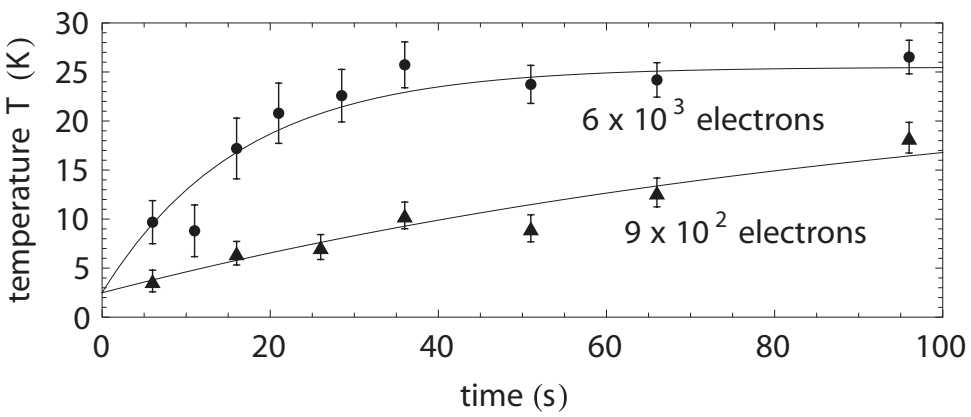


Figure 5.17: After the 5×10^5 $\overline{\mathrm{p}}$ have been adiabatically cooled, thermal equilibrium with the trap electrodes is slowly reestablished via embedded electron heating. The time constant $\tau_{\overline{\mathrm{p}}}$ depends on the number of embedded electrons present.

libration curves are shown for 900 and 6 000 embedded electrons. As expected, the larger number of electrons heat up the $\overline{\mathrm{p}}$ more quickly, and both curves converge to the same value, corresponding to the electrode temperature.

Embedded electron and adiabatic cooling were crucial to our $\overline{\mathrm{H}}$-trapping efforts. Many more $\overline{\mathrm{p}}$ were cooled to much lower temperatures and no $\overline{\mathrm{p}}$ were lost in the process. If embedded electron and adiabatic cooling are followed by evaporative cooling, even smaller $\overline{\mathrm{p}}$ temperatures should be possible. With the ingredients for $\overline{\mathrm{H}}$ cooled and ready in their final nested wells, we can proceed to $\overline{\mathrm{H}}$ formation, detection, and trapping in the following two chapters.

Chapter 6

Antihydrogen Production

Once positrons and antiprotons are in proximity in a nested Penning-Ioffe trap and have been adiabatically cooled, they are ready to form $\overline{\mathrm{H}}$. This requires a driving mechanism to mix the two species. Fig. 6.1 illustrates mixing methods that have historically been used. These include pulsing $\overline{\mathrm{p}}$ into the nested well through the positrons until they cool into the lobes of the nested well [54], driving the $\overline{\mathrm{p}}$ into the positron clouds with fixed-frequency [55], noise [55, 27, 10], or chirped drives [56, 10], and ramping positrons axially out of their nested wells [27]. This chapter will describe the formation methods used to trap antihydrogen, the exploration and optimization of parameter space for producing antihydrogen, and the challenges faced.

6.1 Three-body recombination

The dominant mechanism for $\overline{\mathrm{H}}$ formation at cryogenic temperatures is three-body recombination (TBR) [11, 57], in which a collision between two positrons and an antiproton results in an antihydrogen atom and a positron that carries away the excess momentum and energy. Other $\overline{\mathrm{H}}$ formation processes include radiative recombination [58]

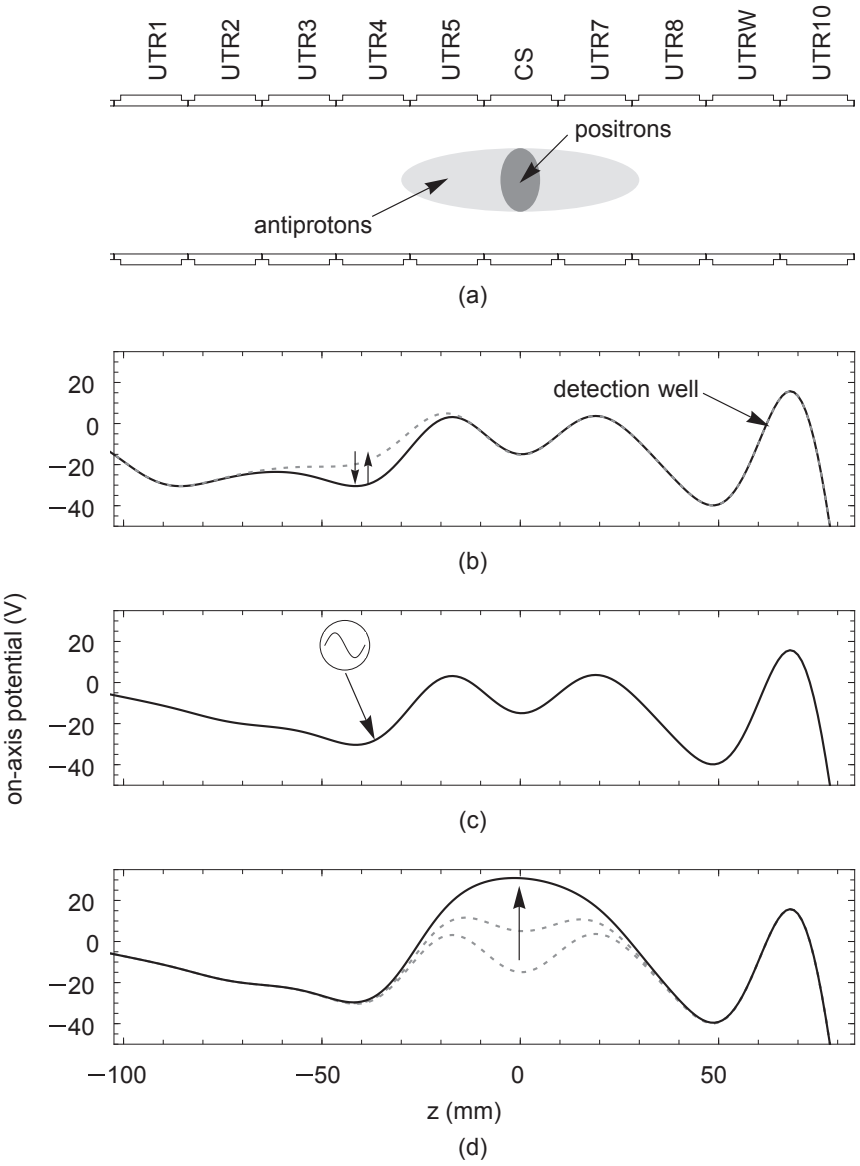


Figure 6.1: Antiproton-positron mixing methods for $\overline{\mathrm{H}}$ production. In order to count a fraction of the Rydberg $\overline{\mathrm{H}}$ produced, a detection well with a highly ionizing electric field is formed near the nested well. (a) Cross-sectional schematic of the upper electrode stack, showing the nested positron-antiproton well. Left to right electrodes in the schematic represent bottom to top in reality. (b) The potential on UTR3, where $\overline{\mathrm{p}}$ are located, is raised to launch pbars into the nested well through the positrons. (c) $\overline{\mathrm{p}}$ are located on UTR5 while a noise, fixed-frequency or chirp drive applied on UTR4 drives the $\overline{\mathrm{p}}$ into the e^{+}. (d) The positron well on CS is ramped until it is inverted, allowing positrons to escape through the $\overline{\mathrm{p}}$. The procedures in (c) and (d) have also been combined for some studies.

in which a radiated photon carries away the excess momentum and energy after a two-body $\overline{\text{p}}$-e$^+$ collision, and charge-exchange [59] of a positron between a $\overline{\text{p}}$ and positronium atom upon collision. However, the rate for three-body recombination dominates over these by many orders of magnitude for our temperatures and positron densities. The three-body recombination rate is given by:

$$\Gamma = C n_{e^+}^2 v_{e^+} b_{min}^5 \propto n_{e^+}^2 T_{e^+}^{-\frac{9}{2}} \tag{6.1}$$

where C is a constant that depends on B-field, n_{e^+} is the positron density, v_{e^+} is the positron velocity proportional to $T^{\frac{1}{2}}$, b_{min} is the classical distance of closest approach as defined in Eq. 4.10, and T_{e^+} is the positron temperature. C is of order unity for zero B-field, and about 0.07 for strong B-field. TBR occurs during the positron cooling of $\overline{\text{p}}$ [60] when the antiprotons and positrons are mixed, and Eq. 6.1 shows that the rate scales strongly with positron density and inverse temperature. Rydberg antihydrogen atoms are created at $n \approx 50$ (though n is not a good quantum number for high B-fields), and these de-excite through collisions to states that do not allow any more collisional de-excitation due to insufficient thermal energy and large spacing between levels. From here the atoms can further de-excite via spontaneous emission.

6.2 Detection wells

A fraction of the $\overline{\text{H}}$ atoms that are produced by the methods listed at the start of this chapter can be detected and counted using field-ionizing "detection wells" [54]. These wells are depicted along with the nested potentials in Fig. 6.1. If a neutral $\overline{\text{H}}$ encounters the large electric field of the well, the $\overline{\text{p}}$ is stripped and captured in the well. The fraction of atoms that can fall into this potential include weakly-bound atoms which are traveling towards and are within the solid angle subtended by the ionization well. Once the $\overline{\text{p}}$-e$^+$

mixing is complete, the remaining particles in the nested well are discarded away from the detection well, leaving only the stripped antiprotons behind. These are then released and counted.

One might worry about nested-well $\overline{\mathrm{p}}$ entering the detection well and becoming confounded with the ionized $\overline{\mathrm{p}}$. The large blocking potential on UTR8 and UTRW prevents this from happening, and positron cooling of nested-well $\overline{\mathrm{p}}$ [60] deeper into their wells makes it even more unlikely. Additionally, if a $\overline{\mathrm{p}}$ did somehow make it into the detection well region, it would have no mechanism of cooling into the well. This method of counting the produced $\overline{\mathrm{H}}$ is therefore background-free.

Detection well studies have proved very useful: they were used to demonstrate the first background-free observation of $\overline{\mathrm{H}}$ [54] and measurements of $\overline{\mathrm{H}}$ states [55] and velocities [61]. In 2008, detection well studies in a Ioffe field yielded the first observation of $\overline{\mathrm{H}}$ produced in a quadrupole field [27], resolving a debate over whether this could be done at all, and initiating efforts to trap antihydrogen in a quadrupole trap. One drawback of this method, however, is that it only enables the detection of the $\overline{\mathrm{H}}$ that are too excited to trap. Additionally, these $\overline{\mathrm{H}}$ have high enough axial to radial velocity ratios to reach the detection well before annihilating on the trap wall, whereas cold $\overline{\mathrm{H}}$ are expected to have comparable axial and radial velocities.

The first velocity measurements showed that the most weakly-bound $\overline{\mathrm{H}}$ atoms traveled at an axial speed that was 20 times larger than expected for a 4.2 K temperature. It was surprising that these fast $\overline{\mathrm{H}}$ were formed, despite collisional cooling from cold positrons [60]. One speculation was that these $\overline{\mathrm{H}}$ are both fast and weakly-bound because of decreased interaction with positrons, which could happen for $\overline{\mathrm{p}}$ at a larger radius than the positrons, for example. Another suggestion came from a theoretical study [62] showing that slow, trappable $\overline{\mathrm{H}}$ are also formed in the mixing process, but that some of these charge-exchange

with fast $\overline{\mathrm{p}}$ in the side wells of the nested Penning trap. Thus $\overline{\mathrm{H}}$ production studies using detection wells may still gauge qualitatively whether low-energy $\overline{\mathrm{H}}$ are also being formed with a given driving method, though this has not been proven. There is currently no known technique to directly detect low-energy $\overline{\mathrm{H}}$, other than trapping them in a shallow magnetic potential that is centered around the region where they are formed, followed by destructive annihilation detection.

6.3 Parameter optimization using detection wells

In order to begin our antihydrogen-trapping attempts, we needed a starting point for experimental settings. A parameter space consisting of number of particles, radius, nested well depths, magnet settings, and drive settings had to be optimized. Given the low number of $\overline{\mathrm{H}}$ trapping trials that could be done in a day (1-2 trials, due to a recovery time needed for the magnet after quenching), we did detection well studies (up to 4 in a day) to learn how to make $\overline{\mathrm{H}}$, and later used these settings in trapping trials.

Examples of parameter optimization using detection wells include determining the optimal final nested well potential (after the adiabatic-cooling sweep) and finding the optimal number of positrons for the particular well. The positron number must be matched to the final potential in order to achieve the smallest possible energy barrier between the positrons and antiprotons (see Fig. 5.13) before the mixing, in order to minimize the amount of energy needed to drive the $\overline{\mathrm{p}}$. Final $\overline{\mathrm{p}}$-lobe potentials and $\overline{\mathrm{p}}$ numbers were not varied for these trials. Fig. 6.2a shows the number of $\overline{\mathrm{H}}$ produced and detected as a function of final positron nested well voltage (CS voltage). In all cases, the starting CS voltage was -35 V and the number of positrons initially loaded was 90 million. We found that the deeper CS potentials produced more $\overline{\mathrm{H}}$, because they were able to retain larger numbers of positrons

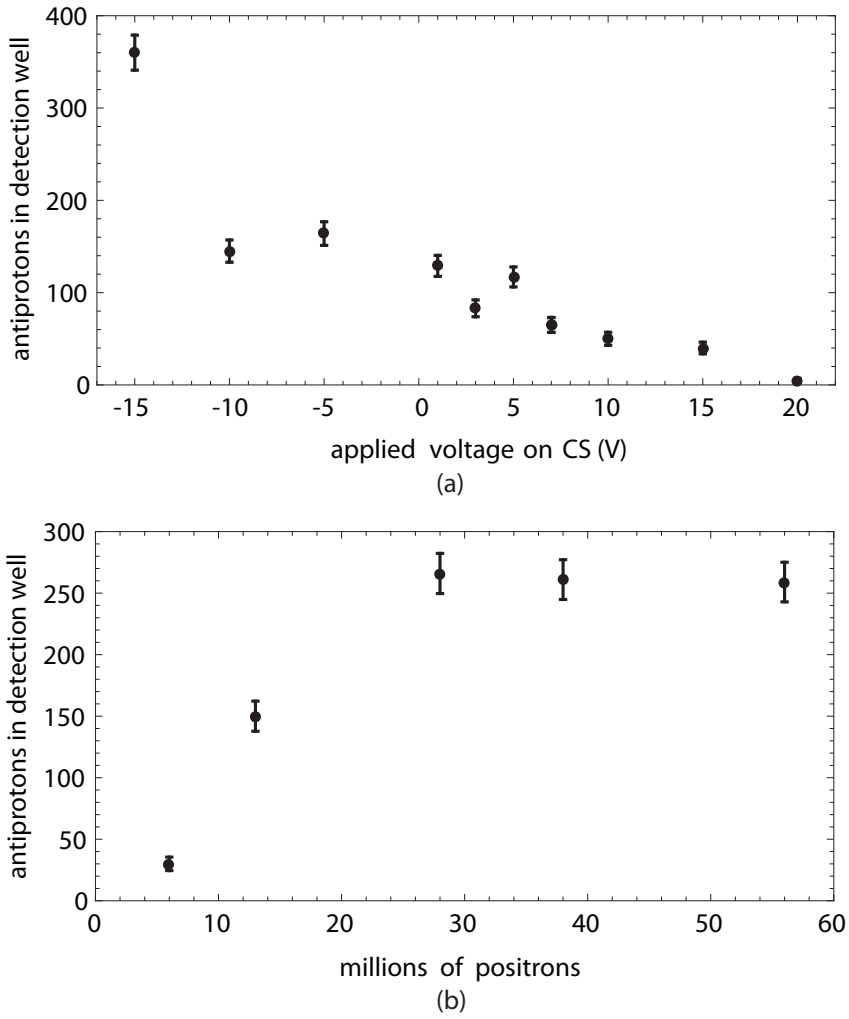


Figure 6.2: Detection well studies with the Ioffe field off, and application of a -3 dBm attenuated noise drive on the $\overline{\mathrm{p}}$ for 600s. 1 million $\overline{\mathrm{p}}$ were initially loaded into the nested well. (a) Detected $\overline{\mathrm{H}}$ produced increases with deeper CS final well. The positron holding well is swept from the initial -35 V to the final CS voltage indicated to make a shallower well. The nested well was initially loaded with 90 million positrons. (b) Detected $\overline{\mathrm{H}}$ produced increases and then saturates with increasing positron number. In this experiment, the final CS voltage after nested well sweep was -7 V.

in the Ioffe field (see Sec. 6.5).

The largest numbers of $\overline{\mathrm{H}}$ produced using 90 million positrons were found with a -15 V final CS potential. It was clear that 90 million positrons were enough to saturate the

−15 V well, due to the observable sudden loss of excess positrons when the positron well was swept to its final value. This loss was not seen for smaller numbers of positrons in the −15 V well. The "loss spike" became a signature that was used throughout the beam run to indicate that we had a sufficient number of positrons just before TBR, despite positron losses that could happen before this point (discussed in more detail in Sec. 6.5). This helped to ensure trial-to-trial reproducibility in positron number before $\overline{\mathrm{H}}$ production, independent of the exact number of positrons initially loaded. A −20 V CS potential was also investigated for even larger number of positrons (135 million) but gains in detection well $\overline{\mathrm{H}}$ were not observed.

Fig. 6.2b illustrates that for a particular positron well depth, detected $\overline{\mathrm{H}}$ increases with increasing positron number but then levels off, consistent with the positron wells getting saturated. The data corresponds to a final positron voltage of −7 V. The maximum number of $\overline{\mathrm{H}}$ achieved for −7 V was lower than that for −15 V, showing the beneficial effect of a larger density of positrons. The saturation is expected because the space-charge of the $\overline{\mathrm{p}}$ reduces the effective well depth for the positrons, causing excess positrons to be lost despite the fact that the trap potential on its own is large enough to hold all the positrons. Based on these studies we chose to use a −15 V positron well with 90 million loaded positrons for all the trapped antihydrogen experiments. Detection wells were also used to optimize noise drive parameters, discussed in Sec. 6.4.1.

6.4 Mixing methods

The methods used for the optimization and/or trapped antihydrogen studies were noise, fixed-frequency and chirped resonant drives applied to $\overline{\mathrm{p}}$ and inverted-well ramps applied to positrons. The $\overline{\mathrm{p}}$ are located on UTR5 (Fig. 6.1) and driven by a signal applied

from a SRS DS345 function generator on the adjacent electrode UTR4. As the $\overline{\mathrm{p}}$ are driven out, they can either become radially lost, oscillate across the entire nested well and cool from positrons with each pass until they are confined in one of the lobes, or form $\overline{\mathrm{H}}$ which are then trapped or lost. The goal for all the drives was to provide the minimum energy needed to allow the $\overline{\mathrm{p}}$ to escape out of their well and into the positron cloud. The positron ramps were done slowly compared to their axial bounce times to prevent heating. It was also experimentally confirmed that the drives applied did not heat the positrons out of their well. Thus care was taken to create the lowest-energy $\overline{\mathrm{H}}$ possible.

Since the $\overline{\mathrm{p}}$ plasma must be axially driven far from the center of its well to reach the positron cloud, it experiences an anharmonic potential. Anharmonic oscillators have resonance frequencies that depend on their amplitude, in contrast to harmonic oscillators whose amplitude and frequency are independent. If a single resonant frequency is applied, as the amplitude increases, the plasma then shifts out of resonance and the amplitude decreases. The cycle repeats. For this reason, a fixed-frequency axial drive is not the most efficient way to excite the entire $\overline{\mathrm{p}}$ cloud into the positrons. Instead, for the trapped antihydrogen trials reported here, we use noise drives, which contain a broad range of frequency components, and coherent drives chirped from high to low frequency. The chirped drives are designed to chase the plasma's resonance frequency as the amplitude changes.

Fig. 6.3 shows the calculated dependence of the center-of-mass, on-axis axial bounce frequency as a function of energy or amplitude. As the frequency decreases, the amplitude approaches zero because the particle is driven to the central turning point of the nested well and is at an unstable equilibrium. This is the point at which the $\overline{\mathrm{p}}$ has become unbound. At higher energies (lower, non-singular frequencies), the freed $\overline{\mathrm{p}}$ oscillates along the entire length of the nested well. This curve shifts to the right and spans a higher frequency range for the deeper off-axis potentials, as in Fig. 3 of Ref. [55].

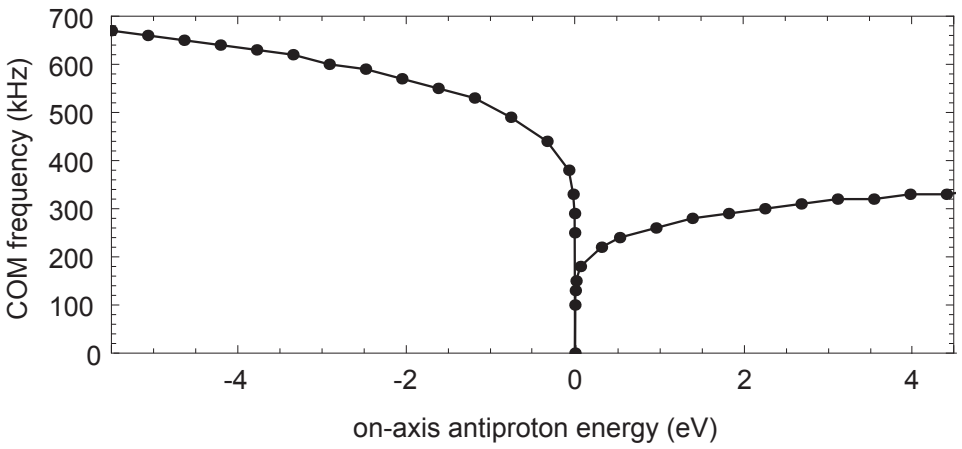


Figure 6.3: Axial antiproton center-of-mass bounce frequency as a function of its energy, calculated with respect to its unbound energy at 0 eV at the center of the nested well. Points to the right of 0 eV indicate that the $\overline{\mathrm{p}}$ has left its lobe of the nested well and is oscillating across the entire double-well structure. As the frequency is decreased, the energy and amplitude grow until the $\overline{\mathrm{p}}$ leaves the well.

6.4.1 Noise drives

We used the SRS DS345 frequency generator, capable of generating white noise with a spectrum between 0 and 10 MHz, to access the changing axial frequencies of the $\overline{\mathrm{p}}$. Fig. 6.3 shows that at low frequencies $\overline{\mathrm{p}}$ are excited too high in energy (out into the entire nested well), and at high frequencies the drives interact very little with the $\overline{\mathrm{p}}$. A bandpass filter was installed to restrict the generator output to between 300 kHz and 1 MHz, with 3 dB attenuation.

The optimal noise drive parameters were determined via field-ionization studies. Fig. 6.4 shows a study in which a noise drive was applied on 1 million $\overline{\mathrm{p}}$, with 90 million positrons initially loaded in the nested well. During the noise drive, positrons were also spilled out axially by ramping the well shallower until it was inverted. The duration of the mixing was varied and as expected, the longer interaction times gave higher numbers of detected $\overline{\mathrm{H}}$. Ramp and noise drives were tested separately and for our parameters the noise drive had the dominating effect. From these results we set relatively long noise drive times

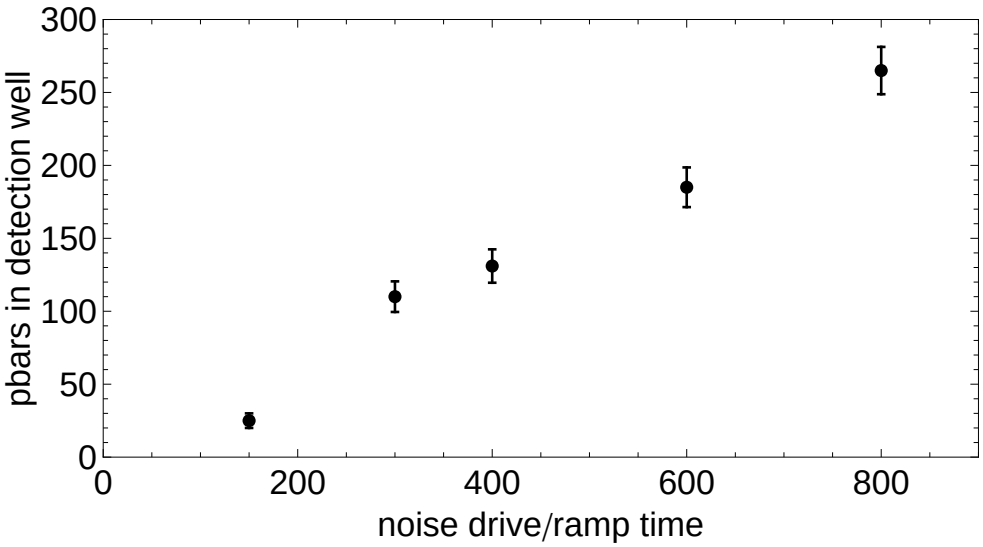


Figure 6.4: Detected $\overline{\mathrm{H}}$ produced increases with interaction time. A noise drive was applied while the CS voltage was simultaneously increased until the positron well inverted at +43 V. Nested well loaded with 90 million positrons and 1 million $\overline{\mathrm{p}}$. Ioffe field is off here.

of 10 minutes for all the trapped antihydrogen noise drive trials.

Detection well studies were also done to find the optimal noise drive strengths. Fig. 6.5 shows $\overline{\mathrm{H}}$ detected as a function of the noise drive attenuation, both with and without the Ioffe field on. Detection wells are known to only detect high-energy $\overline{\mathrm{H}}$, but actually creating these high-energy $\overline{\mathrm{H}}$ is undesirable. Thus the peaks in Fig. 6.5 set an upper limit on the strength of noise drive, and for all trapped antihydrogen trials we used a larger attenuation than -9 dBm, resulting in lower amplitude noise drives.

Fig. 6.6 shows the $\overline{\mathrm{p}}$ energy distribution in UTR5, immediately after the $\overline{\mathrm{p}}$ were driven and energy-broadened with a noise drive. This was measured destructively by ramping UTR4 until $\overline{\mathrm{p}}$ spilled out from the well and were accelerated towards the degrader. Positrons were not loaded at all in the nested well, but the calculated expected positron distribution is shown for comparison. In this example, 11% of the $\overline{\mathrm{p}}$ in the well overlap with the expected positron energy (calculated numerically), though we have seen up to 20%

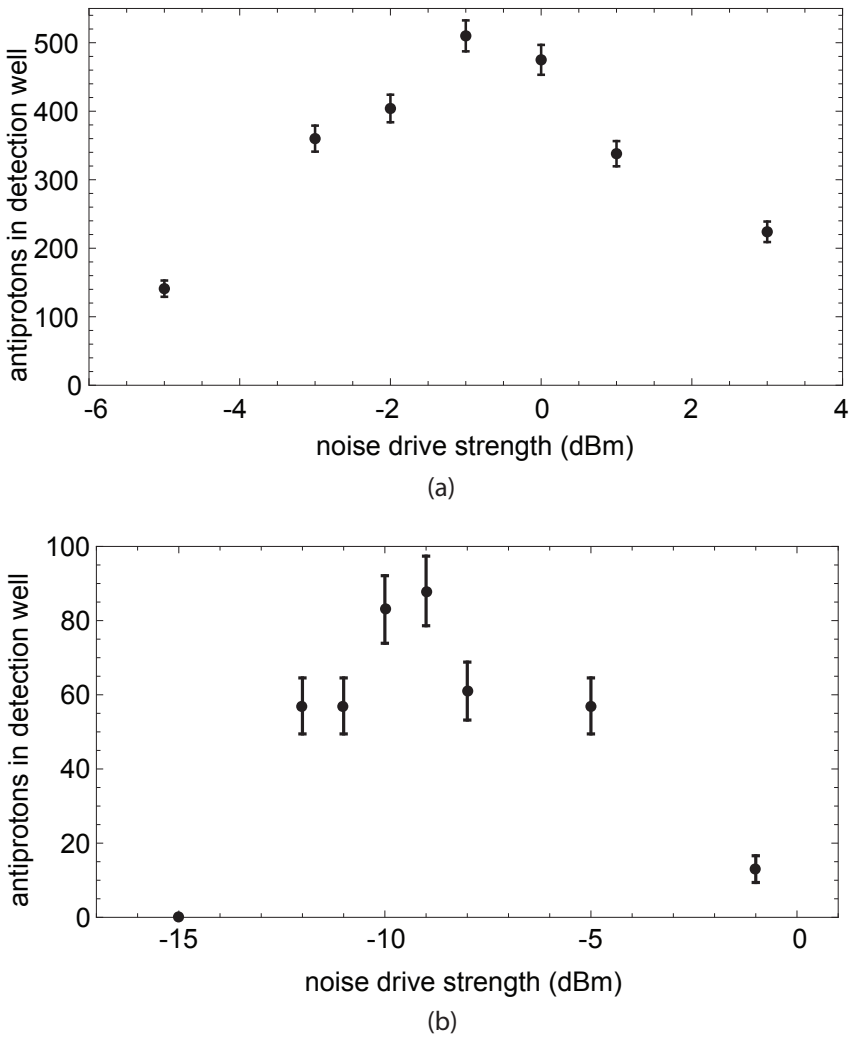


Figure 6.5: Noise drives applied for 600 s with varying drive strengths, using a −15 V final e^+ well. The nested well was loaded with 90 million positrons and 1 million $\overline{p}$. (a) Detected $\overline{H}$ peak at −1 dBm with Ioffe field off. (b) Detected $\overline{H}$ peak at −9 dBm with Ioffe field on.

overlap with the positron energies. This is comparable to the best obtained overlap using coherent drives, discussed next. Eventually, these $\overline{p}$ energy distributions were used as the figure of merit for optimizing the coherent drives.

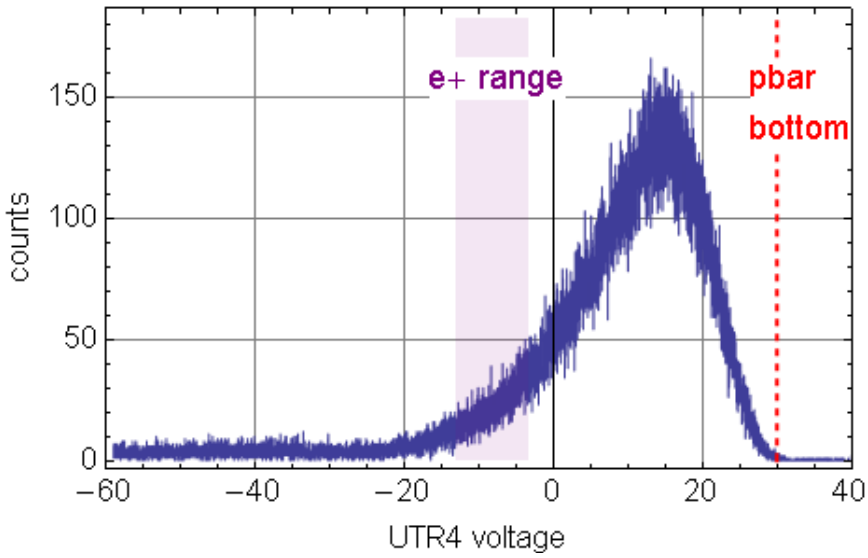


Figure 6.6: Antiproton energy distribution due to a noise drive, with no positrons in the nested well. The energy width is broadened and the peak is shifted from the bottom of the well. This particular example shows that 11% of the $\overline{\mathrm{p}}$ overlap with the expected positron energy.

6.4.2 Coherent drives

A cold, dense $\overline{\mathrm{p}}$ plasma can behave as a single particle and be excited coherently with a drive of sufficient amplitude. This is true despite the fact that the $\overline{\mathrm{p}}$ are initially in a non-zero temperature, thermally-broadened distribution centered around 3.5 K, and despite the fact that most of the individual particles in the plasma see a constant rather than confining potential. The space-charge repulsion of the plasmas actually keeps the plasma together, allowing it to act as a single particle [63]. In principle, chirping the frequency downwards to stay in resonance with the plasma allows oscillation amplitudes to increase without bound (until the particles exit the well) as long as there is no damping, and thus no heating of $\overline{\mathrm{p}}$ plasma. The end frequency determines the final energies of the $\overline{\mathrm{p}}$.

Fig. 6.3 shows the calculated range of frequencies expected to raise the $\overline{\mathrm{p}}$ energies. This ignores the effect of positrons, which would have had an effect at higher energies where the $\overline{\mathrm{p}}$ and positrons overlap. For comparison, an experimental study using fixed-frequency

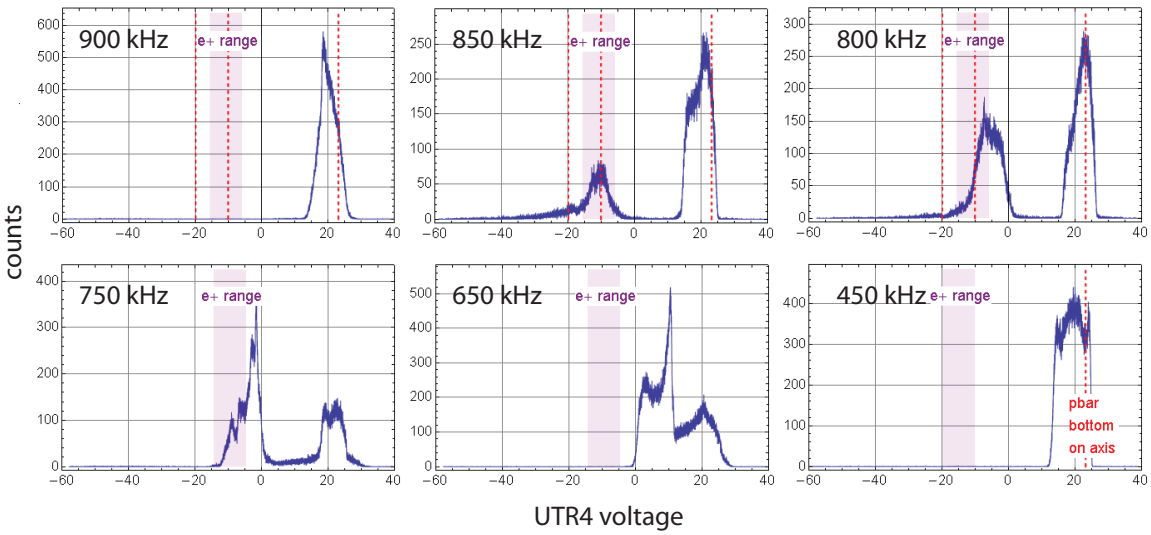


Figure 6.7: Fixed-frequency $\overline{\mathrm{p}}$ drives of 600 s duration at 1 Vpp. Positrons were not present in this study, in order to test the effect of drives on the $\overline{\mathrm{p}}$ alone. As the frequency is decreased, more $\overline{\mathrm{p}}$ are excited into resonance and into the expected positron range of energies, until the drive is no longer locked to the plasma at the lowest frequency.

drives on the $\overline{\mathrm{p}}$ was done to determine the $\overline{\mathrm{p}}$ energies and how much they overlap with the expected positron energies, with no positrons actually loaded. The $\overline{\mathrm{p}}$ reservoir in UTR5 was driven for 10 minutes with a fixed-frequency that was varied from trial to trial. Immediately after application of the drive the $\overline{\mathrm{p}}$ were ramped out and their energy distributions were measured. The sequence of distributions is shown in Fig. 6.7. Overlaid are also the expected range of positron energies if 90 million positrons had been sitting in the CS well.

The highest frequency at 900 kHz shows little interaction with the $\overline{\mathrm{p}}$. As the frequency is decreased, a clear divided response among the $\overline{\mathrm{p}}$ emerges. Some $\overline{\mathrm{p}}$ are left behind at the bottom of the $\overline{\mathrm{p}}$ well, while a significant fraction are excited into a higher energy lobe. Decreasing the frequency further results in a higher fraction of $\overline{\mathrm{p}}$ in the higher energy lobe. At 800 kHz, the best overlap of $\overline{\mathrm{p}}$ with the expected positron energies is achieved at over 25%, corresponding to 160 000 $\overline{\mathrm{p}}$. Decreasing the frequency even more results in the distribution in the higher lobe beginning to shift back to low energy. Finally the lowest

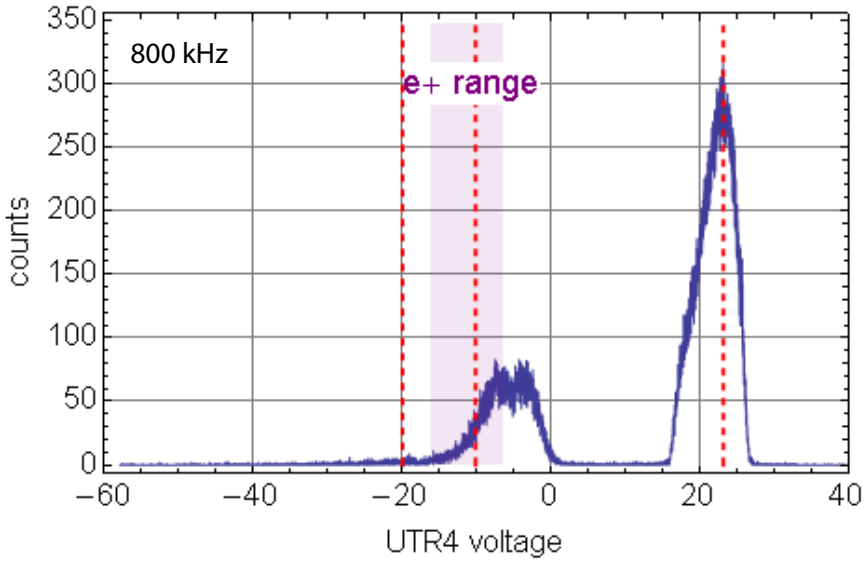


Figure 6.8: 800 kHz fixed-frequency drive. 600 s drive at 1Vpp. Positrons are present in CS well. The presence of positrons reduces the number of $\overline{p}$ in the positron range to 11%.

frequency at 450 kHz once again exhibits a complete loss of resonance. An analogous study using actual positrons is shown in Fig. 6.8 for the best case of 800 kHz. Here positron cooling makes it more difficult for $\overline{p}$ to get excited to higher energies, and only 11% of the $\overline{p}$ overlap with the e^+ energies.

In contrast, noise drives did not result in narrow divided lobes but rather a single broadened energy distribution, as would be expected (see Fig. 6.6). The goal of all the coherent drive studies was to both populate the upper lobe and shift the entire distribution into the positron energy range. Although the presence of actual positrons changes the $\overline{p}$ distributions, these studies were used to roughly find the appropriate range of frequencies for chirps, and confirms the resonant response of $\overline{p}$ when coherent drives are applied. These studies also employed a detection well, which was ramped out and analyzed separately, but appreciable numbers of $\overline{p}$ in these wells were not observed. Based on the results of the fixed-frequency studies, we applied chirped drives starting at 900 kHz and ending at 600 or

700 kHz. This range is shifted to higher frequencies than expected from Fig. 6.3, possibly due to the higher off-axis $\overline{\mathrm{p}}$ oscillation frequencies.

The amplitude of the drives was also varied to ensure that the threshold amplitude to maintain coherence was being met (see Ref. [63]). A balance had to be struck between keeping the $\overline{\mathrm{p}}$ behaving as a macroparticle and driving them too hard, resulting in radial loss. Therefore in addition to studying the energy distribution of the $\overline{\mathrm{p}}$, we also examined the number of $\overline{\mathrm{p}}$ driven into UTR7 (the other nested well $\overline{\mathrm{p}}$ lobe) as a measure of driving effectiveness, and compared the total number of $\overline{\mathrm{p}}$ left behind in the nested well with the number of $\overline{\mathrm{p}}$ lost. At times we would lose half the $\overline{\mathrm{p}}$ at the very start of the drive, resulting in an increased barrier between positrons and $\overline{\mathrm{p}}$ due to space-charge depletion. This further necessitated multiple or changing frequencies rather than just one. To overcome this issue we tried variations of experiments: slowing down the chirp towards the end to avoid blasting all the $\overline{\mathrm{p}}$ out of the trap when they near the top of the barrier, and restarting the chirp multiple times during the same trial.

In the trapped $\overline{\mathrm{H}}$ experiments, we used frequency chirps (again using the SRS DS345) on three timescales: 2 ms, 2.4 s, and 7.5-15 min durations. Chirped drives inject the antiprotons into the positrons with minimal excess kinetic energy. Applying the entire chirp quickly enough leaves no time for collisional transfer from axial to radial energy, which is a decoherence mechanism that would result in either antiproton loss or in $\overline{\mathrm{H}}$ with high kinetic energy. Fig. 4.2 shows an isotropization rate of $\sim$300 Hz for our initially 3.5-30 K $\overline{\mathrm{p}}$. As the $\overline{\mathrm{p}}$ heat up this rate goes down, making our shortest drive timescale equal to a collision time of a couple milliseconds. This was also safely above the microsecond scale axial bounce time, below which the drives would no longer be adiabatic. These fast chirps did not result in the expected lobe-splitting observed with fixed-frequency drives. This may have been due to the axial energy coupling to radial modes in the radial quadrupole Ioffe

field gradient, despite our efforts to avoid collisional coupling by using short drive times. For future experiments, the fast chirp duration could also be varied up to 10 ms, the longest isotropization time for our $\overline{\mathrm{p}}$ temperatures.

Because the 10-minute fixed-frequency drives did result in an excited class of $\overline{\mathrm{p}}$, we tried a variation on this technique: slow chirps on the 10 minute timescale. The idea was to access both classes of high and low energy $\overline{\mathrm{p}}$ shown in Fig. 6.7 and slowly drag the entire distribution up to higher energies via collisions. On short timescales compared to the slow chirped drive, but longer than the thermalization time, the drive looks like a fixed-frequency drive and the plasma beats in and out of resonance as described in the beginning of this section. This results in a broadened $\overline{\mathrm{p}}$ distribution like the one for noise drives, except that the mean of the distribution is at higher energy.

6.4.3 Ramps

Ramping the CS well shallower until it inverts is another method that was used in combination with drives. This is meant to compensate for the increased barrier between the $\overline{\mathrm{p}}$ and e^+ due to space-charge depletion, and results in positrons spilling out and combining with the $\overline{\mathrm{p}}$. Ramping the CS well was not used on its own but rather combined with other methods. Some of our experiments using the slow chirped drive for example, involved positrons being ramped out at the end of the drive, after the entire $\overline{\mathrm{p}}$ distribution had been moved up in mean energy to nearly that of the positrons. This gives many more $\overline{\mathrm{p}}$ the opportunity to mix with the positrons when they are ramped out, compared with the case in which the $\overline{\mathrm{p}}$ distribution spans energies all the way to the bottom of its well. This method also uses up all the positrons, giving all of them the opportunity to participate in $\overline{\mathrm{H}}$ formation.

6.5 Charged particle losses

Our biggest challenge in trapping antihydrogen was the retention of constituent charged particles in the quadrupole Ioffe field. As discussed in Sec. 2.3, because of the twisted-bowtie shape of the field lines (Fig. 2.10 and Eqns. 2.12), charged particles outside the cutoff radius given by Eq. 2.13 follow the field lines into the trap walls where they annihilate. The particles on the $\hat{x}$ and $\hat{y}$ axes are lost most quickly (in the static example shown in Fig. 2.10), but plasma rotation through these axes eventually causes all the particles to be lost.

The amount of plasma deformation depends on the ratio ω_r/ω_z [28]. The rotation frequency ω_r comes from $\vec{E}\times\vec{B}$ motion in the plasma self-fields and magnetron motion from the plasma edges, or from the rotating wall drive. ω_z is the axial bounce frequency. If the plasma is rotating rapidly compared to the axial frequency, it averages over the distorted field lines and does not become twisted itself. If the ratio is small however, as is the case for our trap parameters, then the plasmas follow the field lines and annihilate on the trap walls.

The loss can be limited if the axial extent of the plasmas is made smaller by holding them in deeper wells before $\overline{\mathrm{H}}$ production. The deeper wells compress the plasma axially, leading to a smaller Δz and larger or less stringent cutoff radius, as evident from Eq. 2.13. Fig. 6.9 shows the effect of deepening the holding well in retention of $\overline{\mathrm{p}}$. Antiprotons were loaded into a deep 50 V well and the Ioffe trap was then ramped up to full field. The well voltage was swept to a shallower value indicated, and a 10 minute wait time was implemented to determine the fraction of $\overline{\mathrm{p}}$ retained. This study was done both to find $\overline{\mathrm{p}}$-well operating points and also to check for a time limit over which we could do Ioffe trials before seeing losses. We found that deeper wells could retain 100% of the $\overline{\mathrm{p}}$ despite the Ioffe field.

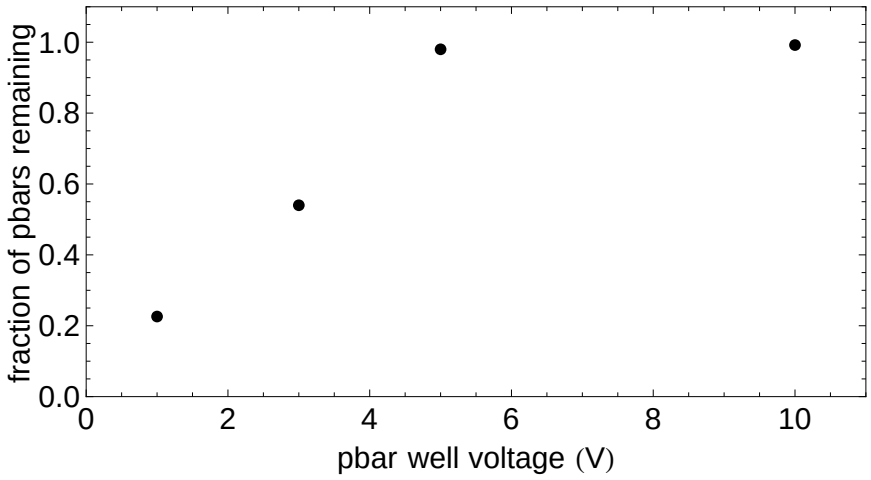


Figure 6.9: Effect of well depth on $\overline{\mathrm{p}}$ losses in the full Ioffe field. The $\overline{\mathrm{p}}$ were initially loaded into a deep 50 V well, and after the Ioffe trap was ramped up to full field, the voltage was changed to the value indicated. The fraction of $\overline{\mathrm{p}}$ retained was measured after a 10 minute wait time.

Even so, deepening the holding wells cannot help us during $\overline{\mathrm{H}}$ production, when the wells must be made shallower for adiabatic cooling (Sec. 5.5), and when one or both of the constituent particles are driven out of their axial well for interaction (Sec. 6.4). Since the quadrupole field couples the radial and axial directions, driving the particles out of their axial wells essentially drives them into the walls.

To reduce the initial plasma radius to below the cutoff for following a field line into the trap wall, we spin the plasmas to as tight a radius as possible before ramping up the Ioffe trap. Once the magnet is turned on, however, since the magnetic quadrupole field breaks cylindrical symmetry, the confinement theorem discussed in Sec. 4.2.1 no longer holds and the now highly-distorted non-spheroidal plasma expands. As it expands, it reaches the cutoff radius for annihilation. These two detrimental issues of axial-radial coupling and broken cylindrical symmetry cause a runaway effect of particle loss.

Fig. 6.10 shows the expansion of the positron radius as the full Ioffe field (both quadrupole and pinch) is increased. Halfway through the ramp-up, the plasma radius more

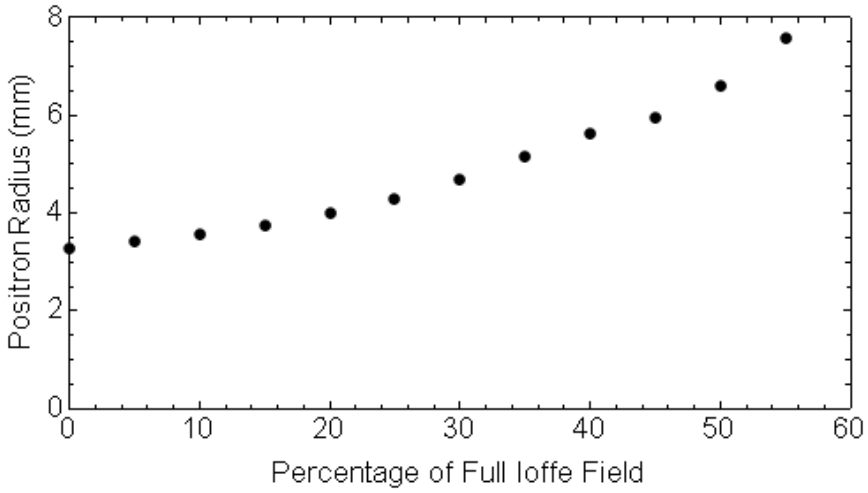


Figure 6.10: Positron radius as a function of the full Ioffe field. The plasma oscillation modes used to determine radius could no longer be measured after 55% of the full Ioffe field. However, it is evident that the plasma radius more than doubles halfway through the ramp-up. The cutoff radius in the Ioffe field was found to be 8 mm for this case of a 0 V potential well.

than doubled. After 55% of full field, the positron oscillation modes used to determine the plasma radius (Sec. 5.1.3) could not be measured, presumably due to a distorted plasma shape or decreased plasma density. This also coincided with the start of positron losses, allowing us to measure the positron cutoff radius of 8 mm for the 0 V potential well used here. Even after pausing the Ioffe-field ramp-up, positron losses continued at a constant rate. This is consistent with the runaway effect just described–the positron plasma continued to expand due to broken cylindrical symmetry and so losses continued. Deepening the positron well during the losses caused them to stop.

By the start of antiproton and positron mixing, the number of positrons reduced from 90 million to 30 million. The vast majority of these losses were purposeful, due to the creation of the final nested well, but tens of millions of positrons could also be lost before this during the rotating wall. We initially had the problem of also losing a large fraction of positrons during the Ioffe ramp-up, but for both $\overline{\mathrm{p}}$ and positrons this problem was eliminated

using our tactics of small initial radii and deep holding wells. However, once the nested well was swept to its final shallower values, the positron cutoff radius decreased and within 5 minutes, only 7 million of the 30 million positrons were typically left in the well. The loss of so many positrons is not necessarily bad if they were involved in making $\overline{\mathrm{H}}$, but the fact that these losses were uncontrolled were a problem. The $\overline{\mathrm{p}}$ however, were stably confined in the shallower potentials in the Ioffe fields, and were only lost during the application of drives. That positrons were more readily lost in the Ioffe field is not understood, given their very small initial 2-3 mm radii (the $\overline{\mathrm{p}}$ are slightly larger in radius).

No solution was found for positron losses during this mixing period, but it is thought that the lost positrons had radii so large that they couldn't have interacted with the $\overline{\mathrm{p}}$ anyway. A detection well study supported this idea–a noise drive on the $\overline{\mathrm{p}}$ was only started after all the observable positron losses died down (about 5 minutes after the final nested well sweep), and it was observed that the same number of $\overline{\mathrm{H}}$ were detected as in an almost identical study in which the noise drive was started just after the nested well sweep.

Retaining as many usable positrons is crucial not only because of the strong TBR dependence on positron density, but also because they heavily suppress $\overline{\mathrm{p}}$ losses when the $\overline{\mathrm{p}}$ are driven. We have seen at least a factor of 5 $\overline{\mathrm{p}}$ loss suppression between trials in which there was a factor of 10 difference in the number of positrons. Additionally, the temperature difference between the remaining nested-well $\overline{\mathrm{p}}$ in these two trials was tens of thousands of Kelvin.

If a quadrupole Ioffe trap is used again to trap antihydrogen, the positron loss problem remains to be investigated. One suggestion is to center the nested well on an electrode other than the CS electrode, because this electrode has holes bored into it for cesium and laser access. One of the positron loss studies (without $\overline{\mathrm{p}}$) showed that centering the nested well on UTR8 instead of CS resulted in 28 million positrons retained by the end

of the trial, as opposed to 6 million with CS as the center. One consequence of centering the nested Penning trap elsewhere is that it would no longer be centered with the Ioffe trap.

Other options include loading even more positrons in even deeper final wells with with expense of time, or loading the same number of positrons (90 million) but making the final positron well deeper so as to retain them. The tradeoff with the latter option is that the energy gap between the $\overline{\mathrm{p}}$ and e^+ will increase, necessitating higher-energy drives on the $\overline{\mathrm{p}}$.

After exploring driving methods, parameters, and particle loss reduction, we were in a position to start trapping antihydrogen, described in the next chapter.

Chapter 7

Trapped Antihydrogen

Trapping large numbers of antihydrogen is critical for detailed study of its structure, the stepping stone to testing fundamental symmetries. The minimum number of antihydrogen atoms needed for high-resolution spectroscopy (at least 1 part in 10^{12}) is estimated to be on the order of 1000 [64]. We demonstrate progress towards this goal with an average of 5 ± 1 trapped atoms per trial [10], the largest number of simultaneously trapped antihydrogen yet. The antihydrogen that were confined in our quadrupole trap had energies less than the 375 mK magnetic trap depth, and were stored for long enough (15-1000 s) to reach their ground state. This successful confinement was done despite the claim that quadrupole fields cannot be used to trap antihydrogen [65], the reasons for which are discussed in Sec. 2.3 and Sec. 6.5. Our result, achieved with 10^6 $\overline{\mathrm{p}}$ and 3×10^7 e^+, compares favorably to the ALPHA collaboration's reported result of 1 atom per trial [66] from 1.5×10^4 $\overline{\mathrm{p}}$ and 10^6 e^+. By using the 10^7 $\overline{\mathrm{p}}$ and many more e^+ available, it may be possible to further scale up the number of simultaneously trapped $\overline{\mathrm{H}}$. This chapter is a review of our result [10].

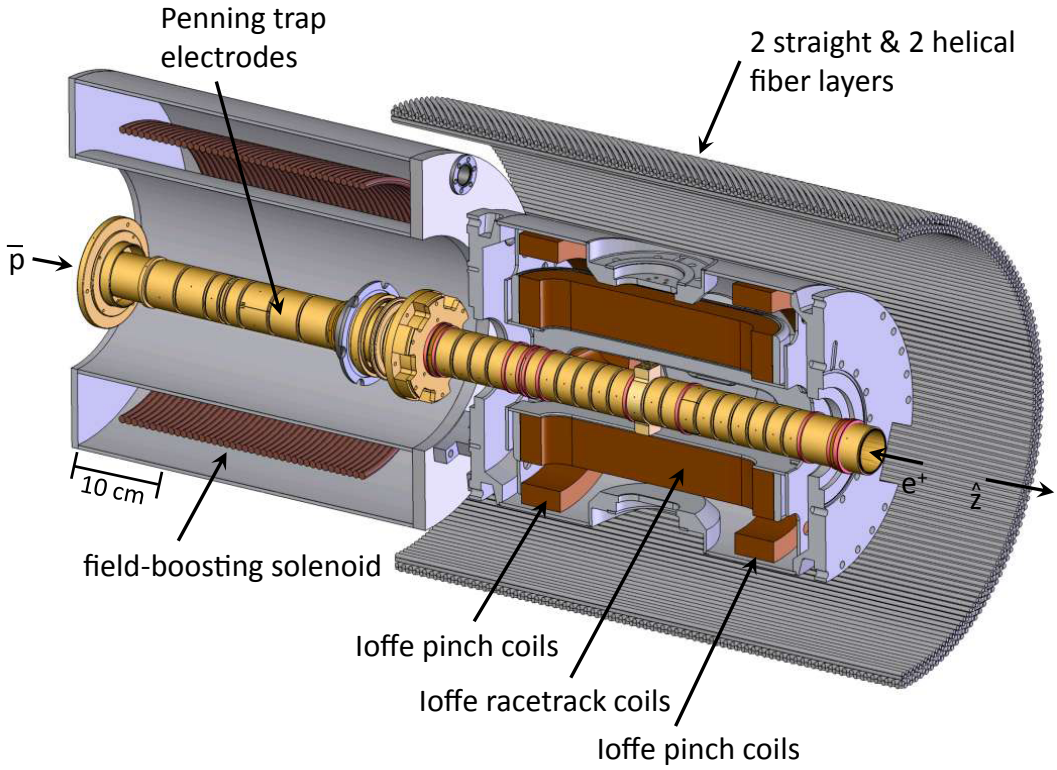


Figure 7.1: The essential components for antihydrogen creation, trapping, and detection. Taken from [10].

7.1 Releasing and detecting trapped antihydrogen

Fig. 7.1 shows the Penning-Ioffe trap (without the 1 T bias magnet) used to make and trap antihydrogen, along with the fiber rings of the detection system and the $\overline{\text{p}}$-solenoid which also plays a critical role in the detection of atoms. A typical experimental trial is described in detail in the previous two chapters. To summarize, a nested well containing antiprotons and positrons is created in the upper stack at the center of the Ioffe trap (right side of the Figure), and this is where the mixing and antihydrogen production occurs for 2 ms-15 min with the Ioffe trap already on. Low-field seeking antihydrogen atoms with energies less than 375 mK are trapped in the magnetic minimum of the Ioffe trap. We then stop the mixing drive/ramp, wait 1 s to ensure that all atoms fall to their ground state, and

apply a ± 5 V/cm gradient along the entire electrode stack to accelerate all charged particles out of the trap. These include remaining particles in the nested well, charged particles that could have been caught in stray fields of the trap, and any potential mirror-trapped antiprotons. This latter possibility is addressed in Sec. 7.3.

Once all the charged particles are eliminated, we continue the experiment by quenching one of the four racetrack coils of the quadrupole Ioffe trap. This was the quickest way to turn off the trap (described in more detail in Sec. 7.2), needed to limit the number of background counts from cosmic rays while the $\overline{\mathrm{H}}$ are spilling out. The time elapsed between the clearing fields and quench ranged from 15-60 seconds, representing the minimum $\overline{\mathrm{H}}$ storage time after the end of production. However, in many trials, as long as 1000 s had elapsed between the start of the $\overline{\mathrm{p}}$-e^+ mixing and the quench, making 1000 s the longest possible $\overline{\mathrm{H}}$ storage time in our experiments.

After the start of the quench, the trap takes 1 s to de-energize, and the sharp change in magnetic flux induces an e.m.f. in the $\overline{\mathrm{p}}$-solenoid just underneath the Ioffe trap (see Fig. 7.1). The e.m.f. trace is used to precisely monitor the turn-off of the Ioffe trap–allowing us to locate the time of quench, and reveals the trap depth over the turn-off time span. Fig. 7.2a-c shows the magnetic trap contours, induced e.m.f., and trap depth during a quench. The release of the $\overline{\mathrm{H}}$ causes annihilation pions to trigger the detection system, while a cosmic ray background also triggers the system. In order to identify antihydrogen atoms, we look for counts above background during the 1 s interval in which the magnetic field of the trap decays after a quench.

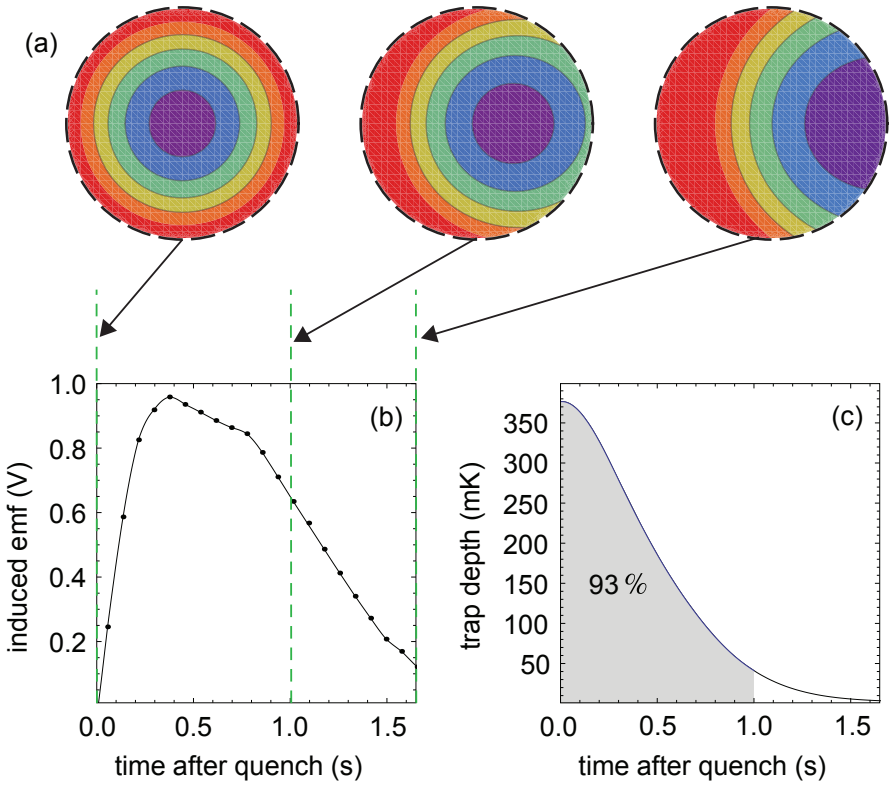


Figure 7.2: Releasing $\overline{\mathrm{H}}$ from the Ioffe trap. (a) Cross-section of the magnetic trap depth contours before and after quenching one of the quadrupole coils with a heater. The dotted circle indicates the electrode boundary. (b) An e.m.f. is induced on the $\overline{\mathrm{p}}$-solenoid due to the rapid flux change of the quadrupole field. (c) Trap depth for ground-state atoms as a function of time, deduced from the $\overline{\mathrm{p}}$-solenoid e.m.f. Within 1 second, 93% of the $\overline{\mathrm{H}}$ are released.

7.2 BTRAP magnet operation and quenching

Quenching the BTRAP Ioffe magnet on demand is necessary to achieve the fastest turnoff time of 1 s. This was done in one of two ways. The first, usual way was to apply 14 W pulse to a resistive heater that was installed in a sideport near the quadrupole coil. When the heater failed close to the end of the beam run, we ramped up the current in the Ioffe trap at a high rate of 0.5 A/s in order to exceed the coils' superconducting critical current. The latter method resulted in quench-training the magnet to well past its nominal current–in the last trapped antihydrogen trial of the year we quenched at 100 Amps, 145% of the nominal maximum field. As soon as the quench happens and the trap is released, we ramp down the current in all magnet coils.

Quenching the trap on demand results in a turnoff time significantly shorter than what could be achieved by ramping down the current using the power supplies. The limiting factor in the BTRAP magnet ramp rate is its quench protection system, required to prevent the large amount of stored magnetic energy from dissipating in the magnet and destroying it. Fig. 7.3 shows BTRAP's passive quench protection system. 5 V diodes are connected in parallel to each coil. If the ramp rate is large enough to induce a voltage drop greater than 5 V across the diodes, the diodes begin conducting current. For the 3.3 H inductance of the BTRAP quadrupole magnet, the maximum ramp rate allowed by this maximum allowed voltage is 1.5 A/s. Ramping down 69 A of current at 1.5 A/s would therefore take 45 s. In effect, the mechanism that prevents a resistive voltage buildup that protects the magnet from a quench is the same that inhibits a large inductive voltage, which limits the ramp rate.

In contrast, when we quench the trap, the current is forced from the magnet into the diode loop. It takes $\sim$1 s for all the current to exit the magnet, as detected by the e.m.f.

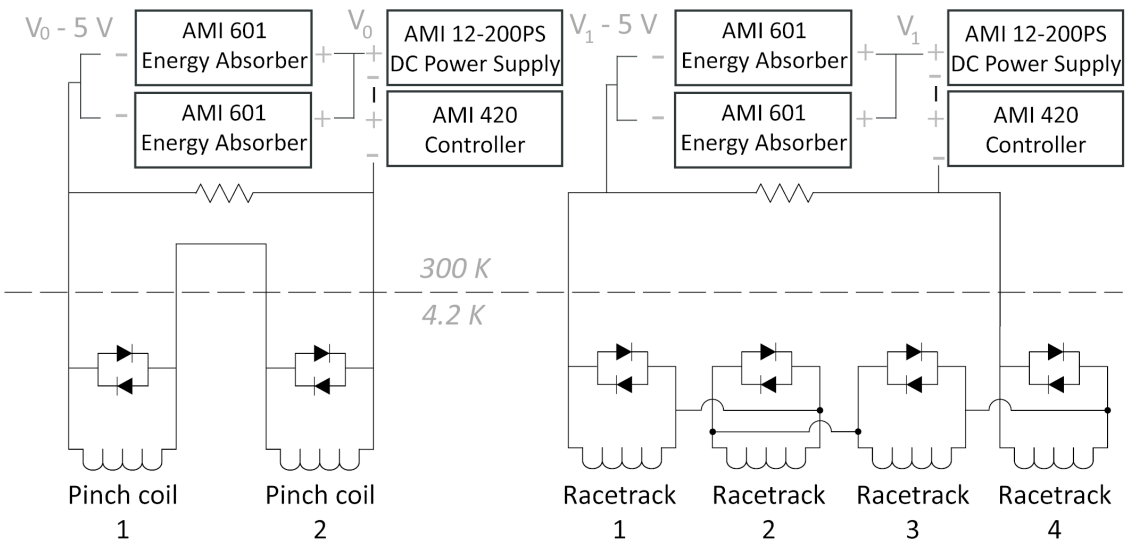


Figure 7.3: BTRAP Ioffe trap electrical setup, including quench protection diodes. Taken from [67]

change across the $\overline{\text{p}}$ solenoid during the turnoff (Fig. 7.2). During the quench, the majority of the quadrupole coil's stored energy is dissipated in the protection diodes.

There are several disadvantages to this method of turning off the Ioffe trap. One is that these quenches were often violent, easily boiling off 30% of the 44 L liquid helium reservoir in one go, and the ensuing exhaust had to be dealt with immediately (there are built-in pressure release systems but exhausting to air was an added precaution). If the quenches happened at currents past the nominal current, 50-60% of the helium could be boiled off. Additionally, it was experimentally seen (but the reason is not understood) that the BTRAP magnet needed a recovery time after quenching–the magnet could only be quenched 1-2 times in an 8-hour shift, limiting the number of experiments one could do in a day. The times of the quenches were not reproducible or predictable, leading to different trapping durations for each antihydrogen experiment.

Ramping up the BTRAP Ioffe trap had to be done carefully to avoid premature, unintentional quenches. It took some experimentation with magnet ramp rates and ramping

the coils in serial versus parallel to both avoid premature quenches and also minimize charged particle losses. The latter depends on the ratio of the pinch and quadrupole fields. For example, a serial ramp starting with the pinch coil has the benefit that the axial pinch field compresses the $\overline{\mathrm{p}}$ radius, which minimizes particle loss once the quadrupole field starts being ramped up. However, this method takes twice as long to ramp up the trap. The final settings were to ramp in parallel at the relatively slow rates of 0.1 A/s and 0.15 A/s for the quadrupole and pinch coils, respectively. The pinch coil was ramped faster in compensation for being unable to ramp it up first. These settings provided a balance between experiment time and particle loss (combined with deeper well depths), and avoided undesired quenches.

The last disadvantage to the BTRAP magnet system is that the 1 second turnoff time is still too slow. To measure our $\overline{\mathrm{p}}$ signal between cosmic ray counts would require turning off the trap in faster than 25 ms. The new CTRAP magnet, discussed in Chap. 9, is designed to turn off on this timescale, and without the need for quenching.

7.3 Mirror-trapped antiprotons

Antihydrogen annihilations are determined from the annihilation of its antiproton. Therefore, it is critical to eliminate all antiprotons before releasing the Ioffe trap [68]. This brings us to the concern of mirror-trapped antiprotons that would not be cleared away by the electric fields if their cyclotron energies were strong enough. A charged particle can be reflected via magnetic mirroring if it approaches an increased magnetic field–for example, one of the two pinch coils that confine the neutral $\overline{\mathrm{H}}$ atoms axially. To keep the magnetic moment μ of the particle constant, as the B-field strength increases, its transverse kinetic energy (cyclotron energy) E_T must also increase. In the absence of an electric field, to keep the total energy constant, the particle's longitudinal energy E_L parallel to the B-field must

decrease. Since the longitudinal velocity can go negative, the particle can not only slow down but also reflect. With two magnetic mirrors the particle can be mirror-trapped.

Mirror-trapping can be avoided if the longitudinal energy is large compared to the transverse energy, so that the $\overline{\mathrm{p}}$ can keep moving in the same direction despite a loss in longitudinal energy from an increased magnetic field. An analysis of mirror-trapping for our experiment is quickly summarized here, but can be found in greater detail in Ref. [68]. First, in the absence of any clearing fields, we find that the condition to avoid mirror-trapping is $E_L/E_T \geq 0.26$. This condition is not satisfied by 8% of $\overline{\mathrm{p}}$ in a 4 K thermal distribution, which is an unacceptably large fraction given our small numbers of detected $\overline{\mathrm{H}}$. With the addition of an electric potential Φ, however, the increase in total energy changes the condition to one that can be satisfied for any E_T and axial magnetic field B provided that Φ is large enough. The most stringent case of no longitudinal energy, $E_L = 0$, was considered. For our $B = 2.15$-2.71 T and our choice of Φ, in order for a $\overline{\mathrm{p}}$ to remain mirror-trapped, it would need to have an $E_T > 137$ eV. By considering $\overline{\mathrm{p}}$ thermal distributions before, during and after $\overline{\mathrm{p}}$-e^+ mixing, all geometrical locations in the trap, collisions, and applied electric and magnetic fields, it was concluded that this mirror-trapping condition for our trap cannot be satisfied, and therefore no $\overline{\mathrm{p}}$ in our experiment were mirror-trapped.

7.4 Trapped antihydrogen analysis

We can suppress the 40 Hz cosmic background rate with statistical analysis. Training data files consisting of only cosmics (350 000 events) or only $\overline{\mathrm{p}}$ (300 000 events, which include a small fraction of background cosmics) are recorded. Events are recorded by the slow data acquisition system if they satisfy the trigger condition of 2 fiber hits and 1 paddle hit (in coincidence). Antiprotons are detected with a 54% efficiency for this trigger condi-

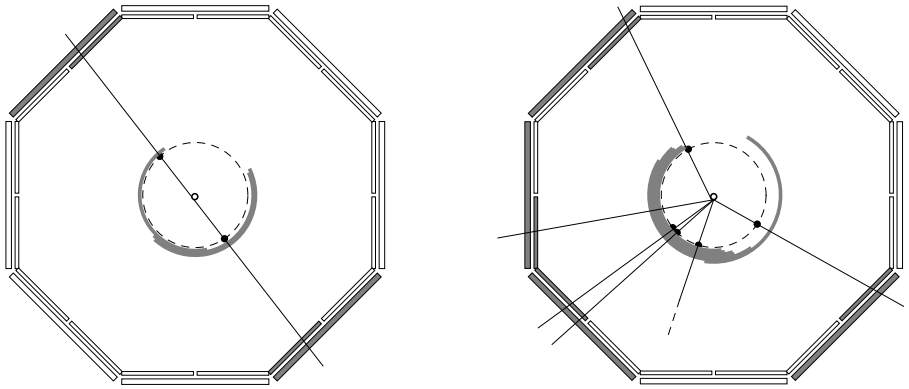


Figure 7.4: (a) An example of a straight-line trajectory traversed by a cosmic. (b) An example of $\overline{\mathrm{p}}$ annihilation hits which cannot be reconciled with a straight-line trajectory. Adapted from [15].

tion. Events recorded during $\overline{\mathrm{H}}$ trapping experiments can be compared to the training data to be classified as either a $\overline{\mathrm{p}}$ or cosmic. As explained in more detail in Sec. 2.5, among the data recorded in these files is the identification of which scintillating fibers and paddles were hit. The discrimination between $\overline{\mathrm{p}}$ and cosmics is illustrated in Fig. 7.4: cosmics are assumed to travel in straight line trajectories as in Fig. 7.4a while a hit pattern such as the one shown in Fig. 7.4b cannot be reconciled with a straight line trajectory, and so is assumed to be caused by a $\overline{\mathrm{p}}$.

The ATRAP analysis done for the published trapped-antihydrogen result [10] was detailed by McConnell [15], using the method of binning and thresholding, followed by a threshold optimization detailed by Richerme [14]. A proposal for an improved analysis using machine learning algorithms rather than binning is presented in the next chapter. The first analysis is summarized here for reference.

First, straight fibers are divided up into clusters based on nearest-neighbor and next-nearest neighbor hits, and whether each cluster consists of a single straight fiber hit or at least 2 fiber hits is determined. How many straight fiber clusters are separated by at least

$\pi/8$ radians is also determined. This process is repeated for helical fibers. Additionally, the number of paddle hits and the number of non-neighboring paddle coincidences is determined. Finally, checks are done to see whether detectors that fired are consistent with a straight line trajectory, using the following criteria:

1. At least 1 paddle coincidence
2. At least 3 helical fiber clusters of any kind
3. At least 3 straight fiber clusters of any kind
4. At least 3 non-neighboring paddle coincidences
5. The sum of the number of helical fiber clusters of any kind and the number of straight fiber clusters of any kind is at least 5
6. At least 3 multiplicity-2 helical fiber clusters separated by more than $\pi/8$, at least 4 multiplicity-2 helical fiber clusters, at least 4 helical fiber clusters separated by more than $\pi/8$, OR at least 5 helical fiber clusters of any kind
7. At least 3 multiplicity-2 straight fiber clusters separated by more than $\pi/8$, at least 4 multiplicity-2 straight fiber clusters, at least 4 straight fiber clusters separated by more than $\pi/8$, OR at least 5 straight fiber clusters of any kind
8. At least two helical fiber clusters of any kind and at least one paddle coincidence or straight fiber
9. At least two straight fiber clusters of any kind and at least one paddle coincidence or helical fiber
10. At least two paddle coincidences and at least one straight fiber or helical fiber

11. At least 2 multiplicity-2 helical fiber clusters separated by more than $\pi/8$ where a straight line through the helical fiber clusters is more than $\pi/8$ away from 1 paddle coincidence or more than $\pi/8$ away from 2 straight fiber clusters

12. At least 2 multiplicity-2 straight fiber clusters separated by more than $\pi/8$ where a straight line through the straight fiber clusters is more than $\pi/8$ away from 1 paddle coincidence or more than $\pi/8$ away from 2 helical fiber clusters

There are $2^{12} = 4096$ combinations of the listed criteria, since each one takes on binary values. Each event in a training data set is classified into one of these 4096 bins. There are two such groups of these 4096 bins: one for the $\overline{\mathrm{p}}$ training data and one for the cosmics training data, and the groups of bins are then separately normalized to the total number of training events N_p and N_c. Based on these, the likelihood of a particle in each bin being a $\overline{\mathrm{p}}$ is determined:

$$L = \frac{f_{\overline{\mathrm{p}}}}{f_{\overline{\mathrm{p}}} + f_c} \tag{7.1}$$

where $f_{\overline{\mathrm{p}}}$ and f_c are the $\overline{\mathrm{p}}$ fraction of N_p and cosmics fraction of N_c in each bin. When an $\overline{\mathrm{H}}$-trapping data set is taken, each event similarly falls into one of these bins and the probability of the event being a $\overline{\mathrm{p}}$ is read-out from the values obtained from the training data. One can then set a probability threshold, only counting the events in a bin as true $\overline{\mathrm{H}}$ counts if the $\overline{\mathrm{p}}$ probability L associated with the bin exceeds the threshold. The trade-off is that while this reduces the noise (the number of cosmics in this bin is small), fewer $\overline{\mathrm{p}}$ also fall into this category, giving less signal, even though the confidence is higher.

The trade-off is shown in Fig. 7.5b and c, reproduced from Ref. [14]. As the threshold is increased, the cosmic background rate desirably decreases, but so does the $\overline{\mathrm{p}}$ efficiency. Fig. 7.5a shows the $\overline{\mathrm{p}}$ likelihood for each bin. The maximum detector efficiency possible is the trigger efficiency of 54% (for our chosen trigger). This efficiency was de-

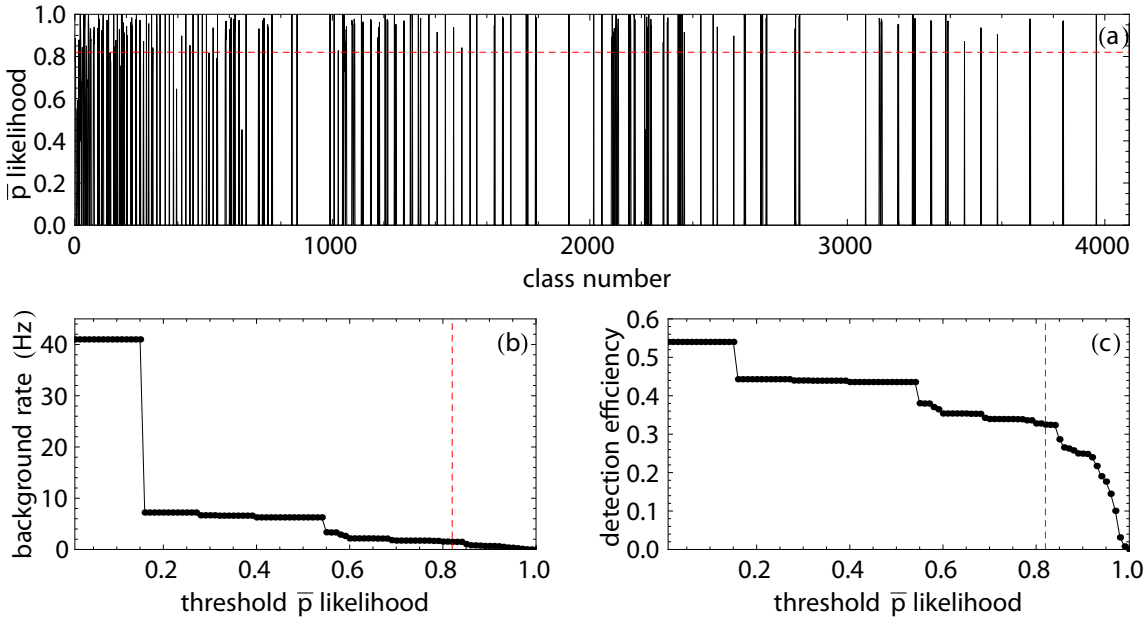


Figure 7.5: (a) $\overline{\text{p}}$ likelihood for each of the 4096 bins. (b) As the threshold increases, the background decreases. (c) As the threshold increases, the efficiency also decreases. The optimal threshold was found from Monte Carlo simulations to be 0.82, plotted in red for (a), (b), and (c).

termined from Monte Carlo simulations using the GEANT4 toolkit, and is limited by the solid angle subtended by our detection system. At this efficiency, before any analysis was done, 100 $\overline{\text{p}}$ above the 20×41 Hz background was observed (in the 20 pooled trials), corresponding to a prior $\overline{\text{p}}$ likelihood of 11%. In order to find the optimal threshold probability, Monte Carlo-simulated data sets were generated, each corresponding to 20 trials and each containing 100 $\overline{\text{p}}$ and 820 cosmics. These $\overline{\text{p}}$ and cosmics were distributed into 4096 bins according to the real training data sets. The optimal threshold probability was the one that minimized the error bar on the number of $\overline{\text{H}}$ detected, and was found to be 0.82. This threshold reduced the cosmic ray background rate from 41 Hz to 1.7 Hz, and also reduced the $\overline{\text{p}}$ efficiency from 54% to a tolerable 33%.

7.5 Trapped antihydrogen results

Twenty $\overline{\mathrm{H}}$-trapping trials were analyzed using the methods described, both on an individual basis and in combination. The individual experimental parameters and drives used for different trials varied. Lists of the drive parameters and the results are shown in Tables 7.1 and 7.2.

Certain trials were repeated with as identical conditions as possible. For example, our best trial, yielding 40 $\overline{\mathrm{H}}$ using a 600 s noise drive (Trial 3), was done 4 times (Trials 3-6), but the number of $\overline{\mathrm{H}}$ found in each trial was not the same. This is also true for repeated trials 9-12, in which the plasma interactions and thus $\overline{\mathrm{H}}$ production and trapping varied noticeably, as seen in $\overline{\mathrm{p}}$ radial loss patterns displayed by our detection system in real-time. Fig. 7.6 shows the real-time radial losses for these trials as the slow 900 s chirped drive was applied. The vertical lines through all the loss spectra indicate times at which at least one of the four trials had a "resonance," indicated by a sudden $\overline{\mathrm{p}}$ loss spike during the chirp. Sometimes the loss spike happened at the expected time and sometimes it didn't, even though all experimental parameters were nominally the same. This indicates that the plasma interactions, which we were unable to precisely control, are very sensitive to initial conditions. For example, slightly different positron radii from trial to trial could change the final numbers of positrons used in each experiment, which affects $\overline{\mathrm{p}}$ losses and $\overline{\mathrm{H}}$ formation rates and changes the effective nested well depths and optimal experimental parameters. Better control of plasma radius and number, combined with a further exploration of parameter space including for example, varying the positron radius, may yield better and more reproducible results. This reproducibility issue, along with the fact that the numbers of $\overline{\mathrm{H}}$ found from many of the individual trials were not known to high confidence, rendered us unable to determine the best $\overline{\mathrm{H}}$-trapping method. Increased numbers of $\overline{\mathrm{H}}$ trapped per trial

Trial	Type	Amplitude	f_i	f_f	t_{drive}	$e+$ sweep?	t_{sweep}
1	Noise	-9 dBm	300 kHz	1 MHz	600 s	No	
2	Noise	-10 dBm	300 kHz	1 MHz	600 s	No	
3	Noise	-12 dBm	300 kHz	1 MHz	600 s	No	
4	Noise	-12 dBm	300 kHz	1 MHz	600 s	No	
5	Noise	-12 dBm	300 kHz	1 MHz	600 s	No	
6	Noise	-12 dBm	300 kHz	1 MHz	600 s	No	
7	Chirp	0.10 Vpp	1.3 MHz	100 kHz	2.4 s	No	
8	Chirp	0.10 Vpp	1.3 MHz	100 kHz	2.4 s	No	
9	Chirp	0.06 Vpp	900 kHz	600 kHz	900 s	No	
10	Chirp	0.06 Vpp	900 kHz	600 kHz	900 s	No	
11	Chirp	0.06 Vpp	900 kHz	600 kHz	900 s	No	
12	Chirp	0.06 Vpp	900 kHz	600 kHz	900 s	No	
13	Chirp	0.06 Vpp	900 kHz	700 kHz	600 s	Yes	300 s
14	Chirp	0.06 Vpp	900 kHz	700 kHz	600 s	Yes	150 s
15	Chirp	0.06 Vpp	900 kHz	700 kHz	600 s	Yes	1.8 s
16	Chirp	0.06 Vpp	900 kHz	700 kHz	600 s	Yes	1.8 s
17	Chirp	0.06 Vpp	900 kHz	700 kHz	600 s	Yes	0.54 s
18	Chirp	0.06 Vpp	900 kHz	700 kHz	450 s	Yes	2.1 s
19	Chirp	1.00 Vpp	900 kHz	700 kHz	0.002 s	Yes	0.009 s
20	Chirp	1.00 Vpp	900 kHz	700 kHz	0.002 s	Yes	0.030 s

Table 7.1: List of trials and parameters

Trial	Raw Counts	Counts above Background	$\overline{\mathrm{H}}$ above Background	Significance (σ)
1	2	0.3	0.9	0.2
2	0	-1.7	-5.1	-1.3
3	15	13.3	40.0	10.2
4	3	1.3	3.9	1.0
5	3	1.3	3.9	1.0
6	1	-0.7	-2.1	-0.5
7	3	1.3	3.9	1.0
8	4	2.3	6.9	1.8
9	6	4.3	13.0	3.3
10	2	0.3	0.9	0.2
11	2	0.3	0.9	0.2
12	3	1.3	3.9	1.0
13	4	2.3	6.9	1.8
14	3	1.3	3.9	1.0
15	2	0.3	0.9	0.2
16	2	0.3	0.9	0.2
17	3	1.3	3.9	1.0
18	4	2.3	6.9	1.8
19	4	2.3	6.9	1.8
20	3	1.3	3.9	1.0
Total	69	35	105	6.0

Table 7.2: Trapped antihydrogen trial-to-trial results

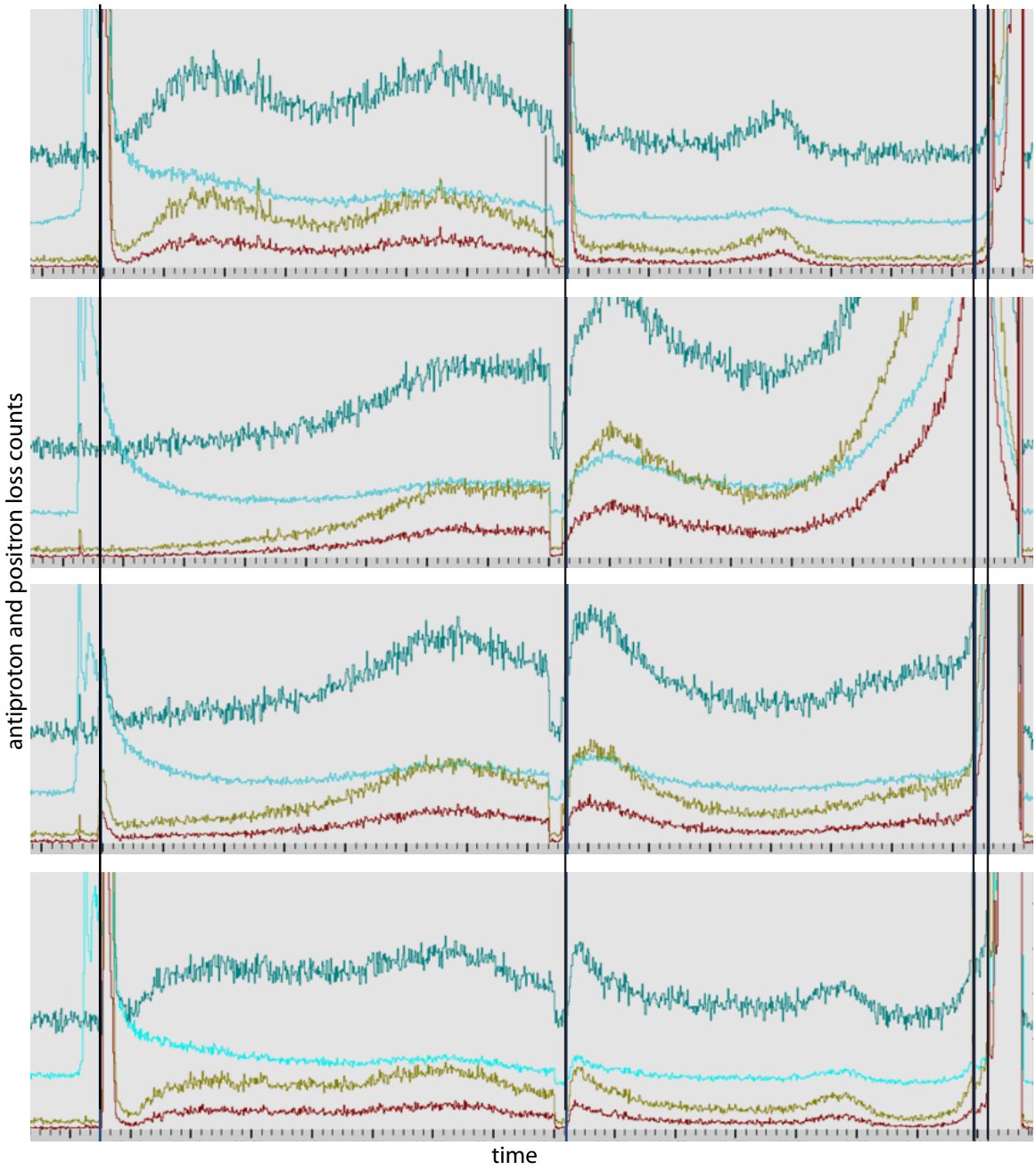


Figure 7.6: Antiproton losses during a repeated 900 s chirp drive (Trials 9-12). The green, gold, and maroon lines indicate $\overline{\mathrm{p}}$ losses with different detector efficiencies. Vertical black lines indicate "resonant" $\overline{\mathrm{p}}$ losses at various times throughout the chirped drive. These loss spikes were not reproducible from trial to trial. The light blue line shows fiber hits, mainly indicative of positron losses. The characteristic positron loss spike from reducing the nested well is shown at the start, followed by a further bump of losses from the Ioffe field.

or increases in detection sensitivity would remedy this problem.

We summed all the detector counts from the 20 trials. This allowed an $\overline{\mathrm{H}}$ signal to become visible above background when using our improved detector efficiency and background, as shown in Fig. 7.7. The central dark blue bin represents the counts during the 1 s interval that the Ioffe trap is shutting off. The surrounding bins show the measured background before and after the turnoff. The central bin shows 35±7 counts above the background. Dividing this by the 33% detector efficiency gives 105±21 trapped $\overline{\mathrm{H}}$, with a statistical significance of 6 standard deviations. This indicates that the probability of the increased central bin counts being due to a fluctuation of the Poisson-distributed cosmic background is 1 in 10^7. On average, 5±1 $\overline{\mathrm{H}}$ were simultaneously trapped per trial.

Twenty control "quench trials" were also done to test for the possibility that the rapid flux change from quenching the magnet results in false signals. In these trials we quenched the Ioffe trap without any particles loaded at all. The summed measurements from these trials are shown in Fig. 7.7c, indicating counts consistent with background and no false signals.

The best of the trapped antihydrogen trials showed a background signal consistent with all the other trials. When Fig. 7.7 is produced without this best trial (though we have no justification for removing it), we still find a significant result of 3.5±0.7 $\overline{\mathrm{H}}$ trapped per trial for the 19 trials, with a standard deviation of 4σ. This is consistent within error with the average for all 20 trials.

7.6 Ground state antihydrogen

For the trapped antihydrogen to be useful for spectroscopy, they must be in their ground state. However, almost all the $\overline{\mathrm{H}}$ produced are in highly excited Rydberg states

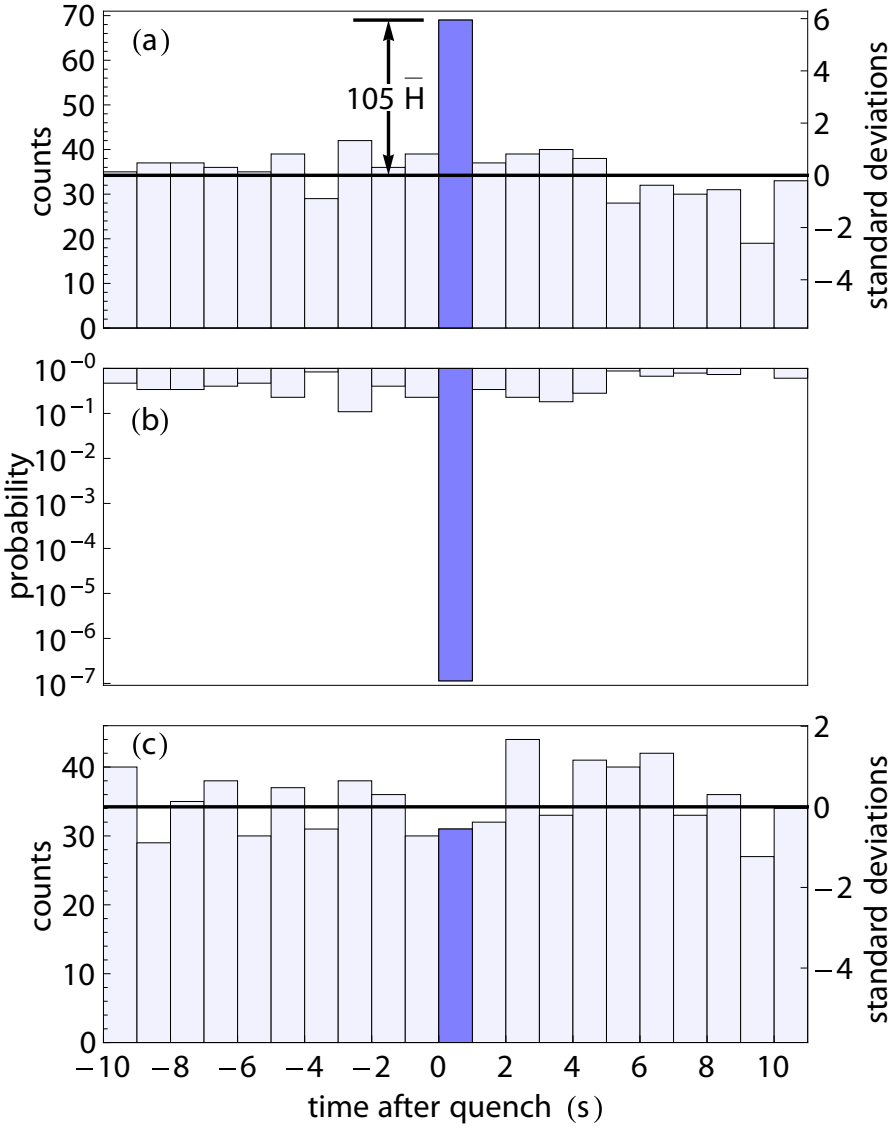


Figure 7.7: (a) 105±21 $\overline{\text{H}}$ detected, a 6σ result. (b) The probability that these counts were caused by cosmics. (c) Sum of 20 control trials in which the magnet was quenched with no particles. No false signals were generated. Taken from [10].

[54]. These highly excited, guiding center atoms [52] are in high-field seeking states that cannot be trapped. Earlier field ionization measurements [69] showed that a small fraction of more deeply-bound atoms with radii less than 0.14 μm are also produced (corresponding to $n \approx 50$, though n is not a good quantum number for high B-fields). A fraction of these are weak-field seeking.

However, these weak-field states can only be trapped if they also decay into weak-field seeking states, and simulations [70, 71] suggest this can this happen. An upper bound on the time needed for the $\overline{\mathrm{H}}$ to reach the ground state can be estimated using the slowest possible decay paths, which are between circular states. Rate equations yield a cascade from the $n = 50$ circular state to the ground state which takes about 0.5 s. The actual cascade time is shorter due to the fact that fields and collisions mix in states other than circular states that have faster decay times. Thus our demonstration of trapped antihydrogen for 15-1000 seconds is long enough to ensure that they are in their ground state.

7.7 Summary

ATRAP observed on average 5±1 trapped ground-state antihydrogen atoms per trial [10], substantially more than previously reported [66]. Implementing the larger numbers of constituent particles which are available should lead to a further increase in trapped $\overline{\mathrm{H}}$ atoms. This result was achieved not without challenges presented by the quadrupole trap, which included particle loss and the possible loss of $\overline{\mathrm{p}}$ plasma coherence when driven with chirped drives. Our particle confinement and preparation schemes succeeded in circumventing some of these problems, and the low-gradient octupole field of the next-generation magnet may help further. Another challenge for $\overline{\mathrm{H}}$-trapping was lack of reproducibility from trial to trial, indicating that the plasma dynamics are very sensitive to initial conditions.

More precise control of antiproton and positron interactions should produce the same numbers of $\overline{\mathrm{H}}$ in every trial, and is a goal for future experiments. Finally, detector post-analysis substantially reduced the background rate, with a smaller decrease in $\overline{\mathrm{p}}$ efficiency, giving an overall improvement in sensitivity. Further increases in detection sensitivity will result in the ability to observe smaller numbers of atoms and gauge the efficacy of trapping parameters from trial-to-trial. The next chapter describes an analysis that demonstrates such an improvement. We are optimistic that increases in the number of trapped and detected $\overline{\mathrm{H}}$ atoms are coming, and will pave the way for laser cooling and precision spectroscopy.

Chapter 8

Multivariate Analysis of Trapped Antihydrogen

$\overline{\mathrm{H}}$ events are detected by their $\overline{\mathrm{p}}$ annihilation signals. However, cosmic rays are a source of background noise that also trigger the detection system. As seen in the previous chapter, background events in the $\overline{\mathrm{H}}$ data can be suppressed using the different event topologies of cosmics and $\overline{\mathrm{p}}$. In the statistical method previously described (Sec. 7.4), binning the events based on 12 criteria followed by thresholding gave rise to a trade-off between background cosmic rate and $\overline{\mathrm{p}}$ detection efficiency. Ideally, we would only reject cosmic background without rejecting any $\overline{\mathrm{p}}$. In practice, we filter out cosmics as much as possible without mistakenly removing too many $\overline{\mathrm{p}}$. The probability threshold that minimized the error bar on detected $\overline{\mathrm{H}}$ in the previous analysis, determined from Monte Carlo simulations, reduced the cosmic background rate from 41 Hz to 1.7 Hz and also reduced the $\overline{\mathrm{p}}$ efficiency from 54% to 33%. Improvements on these numbers for future analyses are possible with alternative methods using multivariate statistics. Particularly, the goal is to reduce the minimum number of trapped antihydrogen atoms required for signal detection, previously

$12/\sqrt{N}$ atoms in N trials. In our trapped $\overline{\mathrm{H}}$ experiments, although we observed on average 5 ± 1 atoms per trial in a pool of 20 trials, we were not sensitive to 5 atoms on a trial-to-trial basis. The number of $\overline{\mathrm{H}}$ in most individual trials were known to 0.2 - 1.8σ. This was prohibitive in determining the best experimental protocols for trapping antihydrogen, since many trials are required to evaluate their performance. Thus higher confidences and smaller uncertainties are needed. The methods presented in this chapter lead to a better detection resolution, and this combined with the faster turnoff time of the new Ioffe trap (Chap. 9) achieves single-atom resolution that will make it possible to optimize our antihydrogen trapping methods.

8.1 Dimension reduction

ATRAP's detector system acquires 959 measurements per event. This constitutes a 959-dimensional parameter space, and we typically have hundreds of thousands of cosmic and $\overline{\mathrm{p}}$ events. In order to make this large volume of data tractable for analysis, the number of dimensions must be reduced. The previous analysis method reduced each 959-measurement event to 12 binary numbers, and allocated all events into one of $2^{12} = 4096$ bins. There are several problems associated with this method of binning, the main one being that of dimensionality. When a region of space is divided into cells, the number of cells grows exponentially with the number of dimensions, making the data exponentially sparse. The large amount of training data needed for statistical significance in all the cells is often not available. This causes many of the bins to have small numbers of data in them, leading to large uncertainties in the assigned probabilities. Additionally, there is loss of information when the 959 real-valued measurements acquired per event are reduced to 12 binary numbers.

The goal then is to reduce the number of dimensions in such a way that preserves

the maximum amount of useful (i.e. discriminatory between $\overline{\mathrm{p}}$ and cosmics) information. Multivariate analysis methods [72] using decision theory or machine learning techniques can be applied to large-dimensional spaces, exploiting the fact that real data will often be confined to a region of space with lower effective dimensionality. Particularly, linear models for classification (Sec. 8.2) can be used to project the data occupying 959 dimensions in this case down to just 1 dimension, while reducing the loss of useful information. Additionally, the reduced parameter space is continuous; the projected 1-D data has an infinite number of bins rather than 4096.

A demonstration of this with $\overline{\mathrm{p}}$ and cosmic training data is presented in Sec. 8.4, using the full raw data sets consisting of 957 of the 959 dimensions, rather than 12. Justification for throwing out two of the dimensions is presented in Sec. 8.3. Sec. 2.5 and Table 2.2 detail the kinds of information available about each event. These include hardware triggers, hit timings, energy deposited, hit multiplicities, and the identification of the detectors hit. The previously reported analysis used some combinations of detector multiplicities as well as detector hit patterns, which were processed into 12 binary variables based on physical insight into possible particle trajectories. The present analysis shows that some of the measurements that were previously excluded contain useful information that can be used to better discriminate between cosmics and $\overline{\mathrm{p}}$. A proposal for a potentially better set of variables is presented at the end of this chapter (Sec. 8.5).

This preliminary demonstration only uses raw data. A better performance could be expected if a good choice of processed variables based on physical assumptions (what we did in the previously reported analysis) were used in addition to the raw measurements. Nonetheless, the results using only raw data give a notably better efficiency/background trade-off than previously reported, along with larger likelihood that an event surpassing the threshold is a $\overline{\mathrm{p}}$. Confidence levels are also exceeded, requiring fewer atoms for detection.

The optimal threshold is calculated from Bayes' decision rule (Sec. 8.4.3) and the associated minimized misclassification rate is found to be very small. The binomial standard deviation is slightly smaller than previously reported, and even larger likelihood thresholds with better discrimination can reduce this further. These threshold and error estimates were done without the need for Monte Carlo simulations. Finally, further improvements can be made by better filtering out background events that contaminate the $\overline{\mathrm{p}}$ training data.

8.2 Fisher linear discriminant

Fig. 8.1 illustrates differences between the methods of binning and linear discrimination. Each bin in Fig. 8.1a defines a particular combination of values or range of values the variables can take. Bin sizes that are too large can lead to loss of important information. Since each variable in the previous analysis could only take values of true or false, it only used 2 bins per binary "dimension" in 12-space, while the linear projection method makes use of the continuous data spectrum in the full 957 dimensions.

If the 2 sets of training data in an n-dimensional space is distributed in a way that is linearly separable, there exists a hyperplane in the space that separates them optimally. In order to reduce the n dimensions to 1 dimension, the data must be projected onto a line. The optimal vector to project onto is the normal to the optimal separating hyperplane, and the optimization problem is to find the projection vector $\vec{a}_{op}$ that maximizes the separation between the means m (signal) compared to their widths σ (noise) after projection. The elements of $\vec{a}_{op}$ are weights that specify the relative importance of each of the n dimensions, or measurement variables. This concept is portrayed in Fig. 8.1b, with two choices of projection vectors $\vec{a}$ as examples. The left is a poor choice because the separation between the 1-D training data sets is small, while the right represents the optimal choice. An optimal

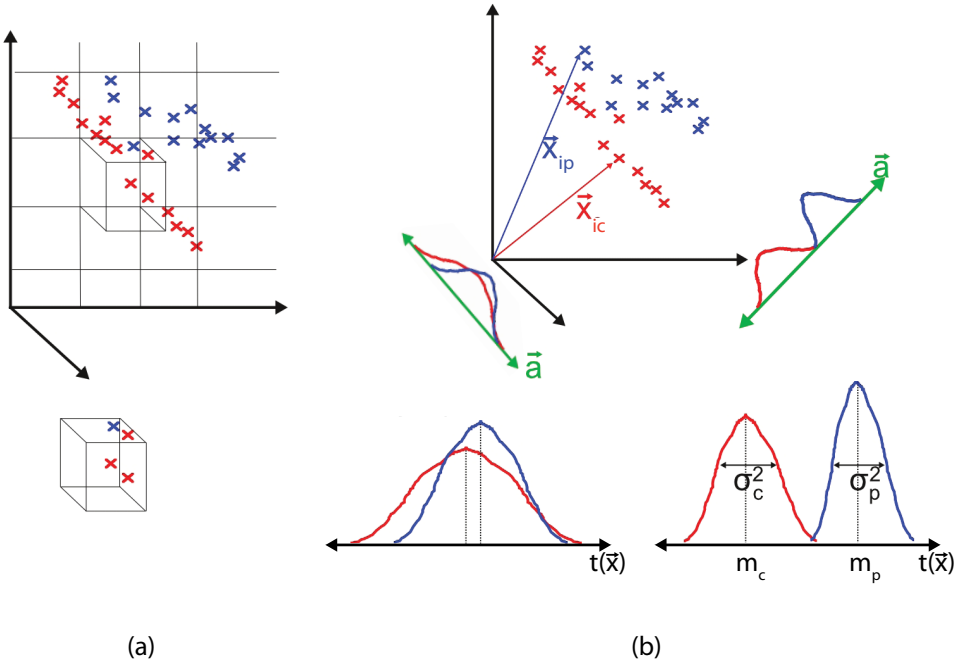


Figure 8.1: Hypothetical 3-dimensional $\overline{\text{p}}$ and cosmic training data sets and the contrast between the methods of binning and linear projections. (a) All data in a bin are treated equally, and large bin sizes can lead to a large loss of useful information. The probability of an event in each bin being a $\overline{\text{p}}$ is given by Eq. 7.1, determined purely from training data. (b) Projections make use of the continuous space of real numbers. The line onto which the data is projected is optimized, maximizing the signal to noise. The left picture shows a poor choice of projection vector, whereas the right shows the optimal vector. The probability of an event with a particular Fisher coordinate being a $\overline{\text{p}}$ is determined by Bayes' theorem, Eq. 8.7, and incorporates both the training data and prior probabilities.

threshold coordinate on this projection axis can then be determined (Sec. 8.4.3) to make decisions about whether a test event is a $\overline{\text{p}}$ or cosmic. It is often unknown a priori how the data will be distributed in space, and the only way to know if the linear projection method will work is to try it.

The following theory describes how to find the optimized vector $\vec{a}_{op}$ and is described in more detail in Refs. [73] and [74]. The simplest linear discriminant function is given by:

$$t(\vec{x}_i) = \vec{x}_i \vec{a}^T \tag{8.1}$$

where $\vec{x}_i$ is the row vector for event i, consisting of 957 elements and $t(\vec{x}_i)$ is its Fisher coordinate–the statistic of interest for each event. The superscript T indicates a transpose, so that the statistic $t(\vec{x})$ is an inner product, i.e. the projection of each vector $\vec{x}$ onto $\vec{a}$, the vector to be optimized. Additionally, Eq. 8.1 assumes that the data is centered about the origin; the global mean has been subtracted off. We define a separation to maximize:

$$J(\vec{a}) = \frac{\text{signal}}{\text{noise}} = \frac{(m_p - m_c)^2}{\sigma_p^2 + \sigma_c^2} \tag{8.2}$$

where m and σ are the means and standard deviations of a set of $t(\vec{x}_i)$ assumed to be normally distributed as shown in Fig. 8.1b, and are calculated separately for each of the two training sets ($\overline{\text{p}}$ and cosmic).

The signal to noise separation in Eq. 8.2 is the Fisher criterion [73, 74]. We now express Eq. 8.2 in terms of the vector $\vec{a}$ defined in Eq. 8.1 so that we can optimize this function with respect to $\vec{a}$:

$$J(\vec{a}) = \frac{(\vec{\mu}_p\vec{a}^T - \vec{\mu}_c\vec{a}^T)^2}{\vec{a}\Sigma_p\vec{a}^T + \vec{a}\Sigma_c\vec{a}^T} \tag{8.3}$$

Here $\vec{\mu}$ is the mean of all the training event vectors $\vec{x}_i$, and Σ is the sample covariance matrix of the training data. These sample statistics are calculated separately for the two kinds of training data. Taking the derivative of Eq. 8.3 and setting it equal to zero gives the analytical solution for the optimized vector $\vec{a}$:

$$\vec{a}_{op} \propto (\vec{\mu}_p - \vec{\mu}_c)(\Sigma_p + \Sigma_c)^{-1} \tag{8.4}$$

where we drop the scalar coefficient since only the direction of $\vec{a}_{op}$ is important. Because the expression for $\vec{a}_{op}$ contains the inverse of the sum of the covariance matrices, it is necessary that this quantity be invertible (Sec. 8.3) and thus that the variables used be linearly independent. Eq. 8.4 shows that the only information necessary to perform the optimization are the sample means and covariance matrices of the training variables, which can easily be found.

8.3 Principal component analysis

In this analysis the covariance matrix sum in Eq. 8.4 found from the 959 measurements was singular, which necessitated throwing out the problematic dimensions. In order to do this, principal component analysis (PCA) [75], was used. This technique finds the unit eigenvectors and eigenvalues of the square, symmetric covariance matrix, and orders the eigenvectors from largest to smallest corresponding eigenvalue. These normalized eigenvectors are known as the principal components, and the ones with the largest eigenvalues contribute the most information. Eigenvalues equal to zero are problematic because they make the covariance matrix non-invertible. Eigenvalues that are very close to zero are also problematic since they lead to instability in numerical computations. Additionally, these eigenvalues represent dimensions that add no or negligible information. Thus, these dimensions may be discarded with no loss of generality.

The eigenvalue spectrum of the combined $\overline{\mathrm{p}}$ and cosmic data sets is shown in Fig. 8.2. Dimensions 958 and 959 in the ordered set contain the negligible (when compared to the largest) eigenvalues of 10^{-18} and 10^{-20}, and were discarded, which resulted in a reduced invertible covariance matrix of rank 957. Furthermore, discarding all but the first 175 dimensions caused no difference in the projected histograms (Sec. 8.4.1) and error rates (Sec. 8.4.3) as in discarding only the last two dimensions. Reducing the number of dimensions below 175 resulted in a continuous shape change of the histograms and increase in error rate. The inset of Fig. 8.2 shows that the kink at dimension 175 in the ordered set of eigenvalues (with eigenvalue 0.02) is the point at which the eigenvalues become negligible, indicating that only the first 175 orthogonal dimensions are needed to distinguish $\overline{\mathrm{p}}$ from cosmic events. The two neglected components do not directly correspond to two particular measurement variables. Rather, the principal components are linear combinations of the raw

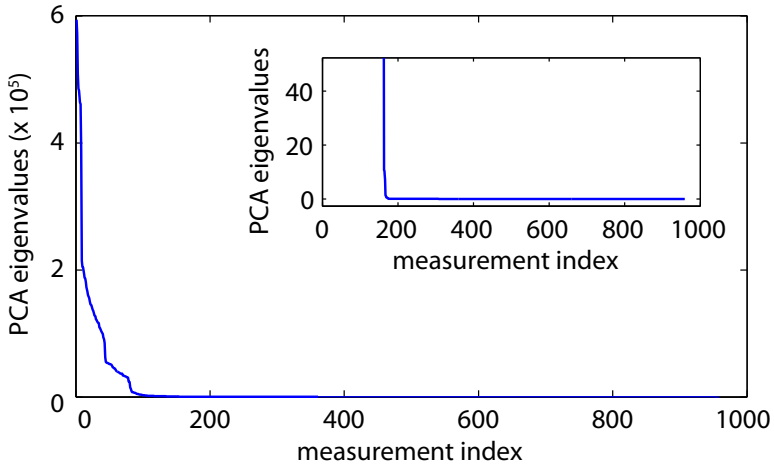


Figure 8.2: Eigenvalues of the combined training data covariance matrix, sorted from largest to smallest. The principal components with near-zero eigenvalues add no new information and may be discarded. Inset: there is a sharp kink at dimension 175 at which point the eigenvalues rapidly become negligible.

data, and the neglected components are potentially related to more than two measurement variables.

8.4 Preliminary results

8.4.1 Probability densities, efficiencies, and background rates

The method outlined in Sec. 8.2 was used to find the optimal projection vector for the raw training data. Fig. 8.3a shows the separately-normalized distributions of the Fisher coordinates $t(\vec{x})$ obtained from projecting 155 000 $\overline{\mathrm{p}}$ and 155 000 cosmic training events onto this vector. There is a clear distinction between the two particle classes, indicating that the data can indeed be separated linearly. These probability densities were used to determine $\overline{\mathrm{p}}$ likelihood, $\overline{\mathrm{p}}$ detection efficiency, and cosmic background rate. The $\overline{\mathrm{p}}$ likelihood for a given Fisher coordinate is defined as

$$L(\vec{x}) = \frac{p(\vec{x}|\overline{\mathrm{p}})}{p(\vec{x}|\overline{\mathrm{p}}) + p(\vec{x}|\mathrm{c})} \tag{8.5}$$

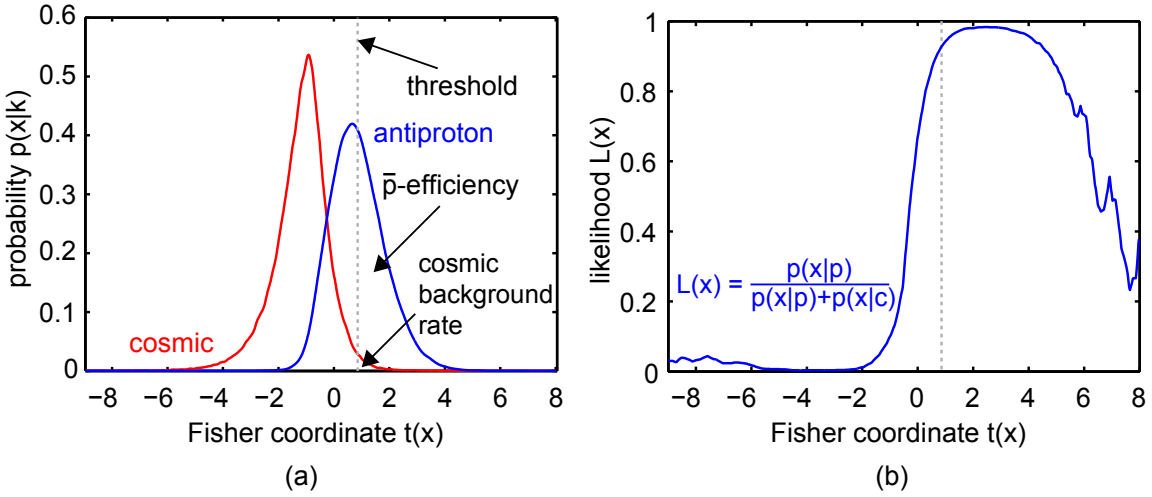


Figure 8.3: (a) Normalized frequency vs. discriminant function histograms for raw $\overline{\text{p}}$ and cosmic training data (155 000 events each). A clear distinction is seen between the $\overline{\text{p}}$ and cosmic distributions, labeled by k. (b) Mapping between antiproton likelihood $L(x)$ and Fisher coordinate t(x).

where $p(\vec{x}|k)$ is the probability of finding an event vector $\vec{x}$ given class k. The likelihood parameterizes $\overline{\text{p}}$ efficiency and background, and the mapping between this and the Fisher coordinate is shown in Fig. 8.3b. Eq. 8.5 is the continuous analogue of Eq. 7.1.

Antiproton efficiencies represent the fraction of $\overline{\text{p}}$ that are correctly identified out of the ensemble of $\overline{\text{p}}$ for a given threshold L, and are found from the area of the $\overline{\text{p}}$ probability density to the right of the chosen likelihood, as shown in Fig. 8.3a. Similarly, the cosmic background rate is given by the area of the cosmic probability density leaking into the $\overline{\text{p}}$ region to the right of L. Scaling these areas to the maximum values of 54% trigger efficiency and 41 Hz background gives the final efficiency and background rates. These are shown in Fig. 8.4 as a function of $\overline{\text{p}}$ likelihood, overlaid with the previous results obtained from binning the data. The cosmic background rates for both analyses are comparable for relevant values of L, while the $\overline{\text{p}}$ efficiencies are higher in the present analysis for nearly all likelihoods.

Particularly, the optimal likelihood threshold of 89%, calculated in Sec. 8.4.3, results in a 31% $\overline{\text{p}}$ efficiency and 0.9 Hz background, an improvement on the 33% efficiency

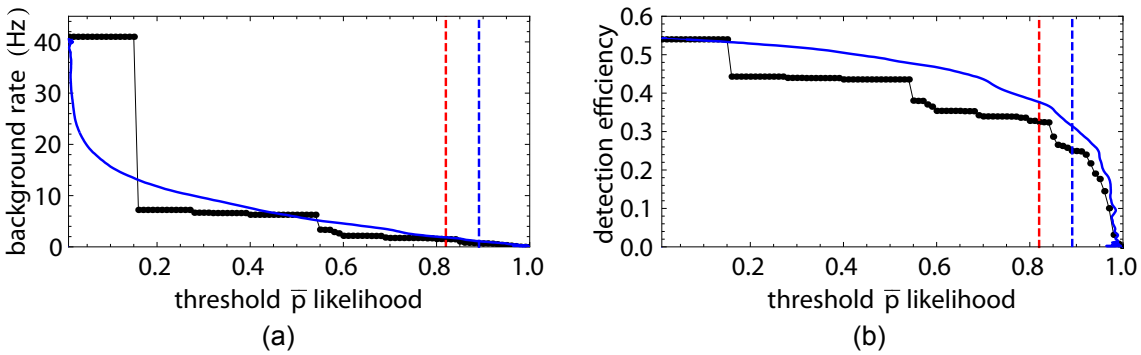


Figure 8.4: cosmic background rate (a) and $\bar{p}$ efficiency (b) for the (black) previous analysis, and (blue) present analysis. The dashed lines indicate the optimal threshold likelihoods for the (red) previous analysis and (blue) present analysis.

and 1.7 Hz background realized in our previous analysis. The efficiency to background ratio, a figure of merit which can be used to compare signal to noise trade-offs, is higher for the present analysis by a factor of 1.8. Fig. 8.4 shows another advantage of the Fisher method: it reveals the continuous nature of the cosmic and $\bar{p}$ distributions. This avoids the large, sharp discontinuities and sensitivity to small changes that arise from discrete bins.

8.4.2 Discriminatory variables

We can extract the measurements that contributed most to the projected separation between $\bar{p}$ and cosmics by looking at the relative separation of each measurement parameter– that is, its mean relative to its standard deviation. For each measurement element j we define a *sensitivity index* d'_j, similar to the Fisher criterion in Eq. 8.2 except that it is taken element-by-element:

$$d'_j = \frac{(\mu_{jp} - \mu_{jc})^2}{\sigma_{jp}^2 + \sigma_{jc}^2} \tag{8.6}$$

Here μ_j is the mean of each variable in the event vector $\vec{x}$, taken over all training events, and σ_j is the width of each variable distribution.

Fig. 8.5a shows the mean values μ_{jp} and μ_{jc} of each of the 959 measurements for the $\bar{p}$ and cosmic training data, with the global means of the data set subtracted off. The

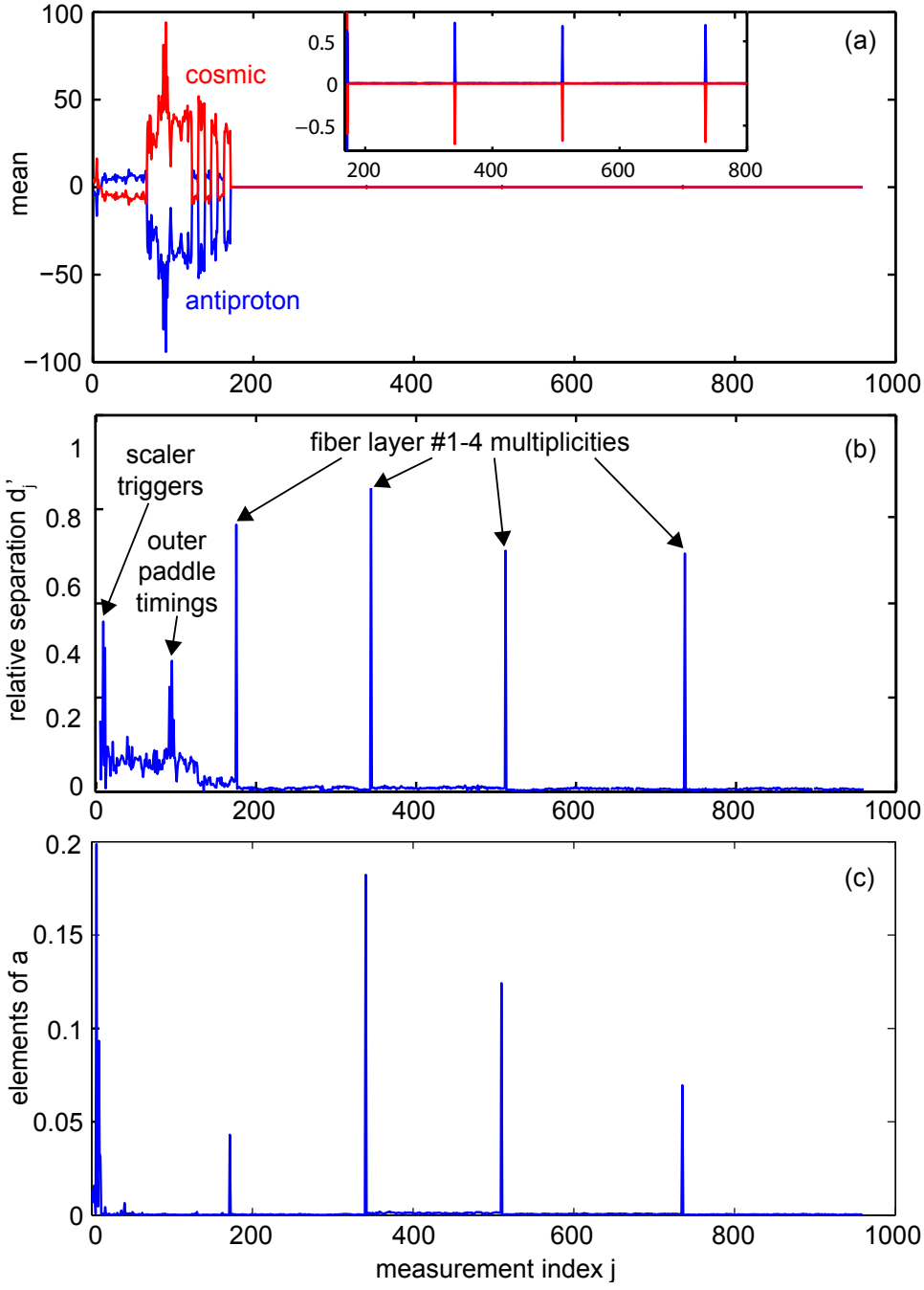


Figure 8.5: (a) The mean value of each of the 959 measurements across the entire cosmic or $\overline{\text{p}}$ data set. There is a clear difference between the means for $\overline{\text{p}}$ and cosmics. Inset: Close-up of fiber hit measurements. (b) Relative separation for each of the 959 measurements calculated from Eq. 8.6. The large spikes represent the most discriminatory detector measurements. (c) Spectrum of measurement weights found from the elements of the vector $\vec{a}_{op}$.

flat regions starting around measurement index #200 indicate the 784 fiber hits which can have values of either 0 or 1, as well as the fiber multiplicities which have values of the same order. The inset shows a close-up of this region. The μ_j can be compared to each other if they represent the same physical quantities–e.g., energies in Inner Paddle 1 vs energies in Inner Paddle 2. They are also are useful for checking which of $\overline{\mathrm{p}}$ and cosmics had the higher average measurement value. The mean vectors made up of the elements in Fig. 8.5a, when projected onto the optimal vector $\vec{a}_{op}$, yield the mean Fisher coordinates m_p and m_c of the probability distributions shown in Fig. 8.3a.

However, the means alone do not determine the most discriminatory variables, so weighting these means by the measurement widths by plugging μ_{jp} and μ_{jc} into Eq. 8.6 yields the results shown in Fig. 8.5b. The most useful features are naturally the ones with large separation of means compared with standard deviation. However, no feature is useless unless the difference in means is zero. This element-by-element spectrum, obtained without doing any optimization, qualitatively shows which measurements are expected to be weighted highly when finding the optimized projection vector $\vec{a}_{op}$.

Fig. 8.5c shows the elements of $\vec{a}_{op}$, which represent the relative weights given to each measurement. These were found using the first 175 of the 959 principal components, in order to decrease noise without decreasing useful information. The spectrum mostly follows the expected shape from Fig. 8.5b, except that the outer paddle timings spike does not show up in this spectrum. It was found that reducing the number of dimensions used in this analysis below 174 caused the fiber multiplicity spikes to disappear and the outer paddle timings spike to appear. This may indicate some degree of linear dependence between these variables.

8.4.3 Optimal threshold and error estimation

The optimal $\overline{\mathrm{p}}$ likelihood threshold, which represents the optimal boundary separating the two classes labeled by $k \in \{\overline{\mathrm{p}}, \mathrm{c}\}$, is found from Bayes' decision rule. Bayes' theorem is given by:

$$P(k|\vec{x}) = \frac{p(\vec{x}|k)P(k)}{p(\vec{x})} \tag{8.7}$$

The posterior probability $P(k|\vec{x})$ depends on the conditional probability densities $p(\vec{x}|k)$ plotted in Fig. 8.3a, and the prior probabilities $P(k)$. The denominator of Eq. 8.7 is a normalization factor given by:

$$p(\vec{x}) = \sum_k p(\vec{x}|k)P(k)$$

The prior probability factor is what differentiates posterior probability from the $\overline{\mathrm{p}}$ likelihood. The prior probabilities were found from the previous analysis to be $P(\overline{\mathrm{p}}) = 0.11$ and $P(\mathrm{c}) = 0.89$. This was obtained from finding 100 $\overline{\mathrm{p}}$ above a background of 820 cosmics in the pool of 20 one second-long experimental $\overline{\mathrm{H}}$-trapping trials, before any optimization was done–that is, using the 54% detector efficiency and 41 Hz background rate. The numerator in Eq. 8.7 is shown plotted in Fig. 8.6a. The coordinate at which the $\overline{\mathrm{p}}$ and cosmic distributions intersect is the one that minimizes the total misclassification error rate of both $\overline{\mathrm{p}}$ and cosmics, found from the area of the intersecting region. This is the decision point, or optimal threshold. The number of $\overline{\mathrm{H}}$ events from trapped $\overline{\mathrm{H}}$ experiments are determined by projecting the "test" event vectors from the experiment data set onto the $\vec{a}_{op}$ found from the training data, and counting the ones that surpass the threshold.

The posterior probabilities are shown in Fig. 8.6b. The second crossover point occurs because the $\overline{\mathrm{p}}$ and cosmic distributions intersect again around $t(x) = 5$ in Fig. 8.6b, at a region with negligible probability density. Note that the probability of finding a $\overline{\mathrm{p}}$ does not reach 1 (and neither does the $\overline{\mathrm{p}}$ likelihood), because the cosmic background

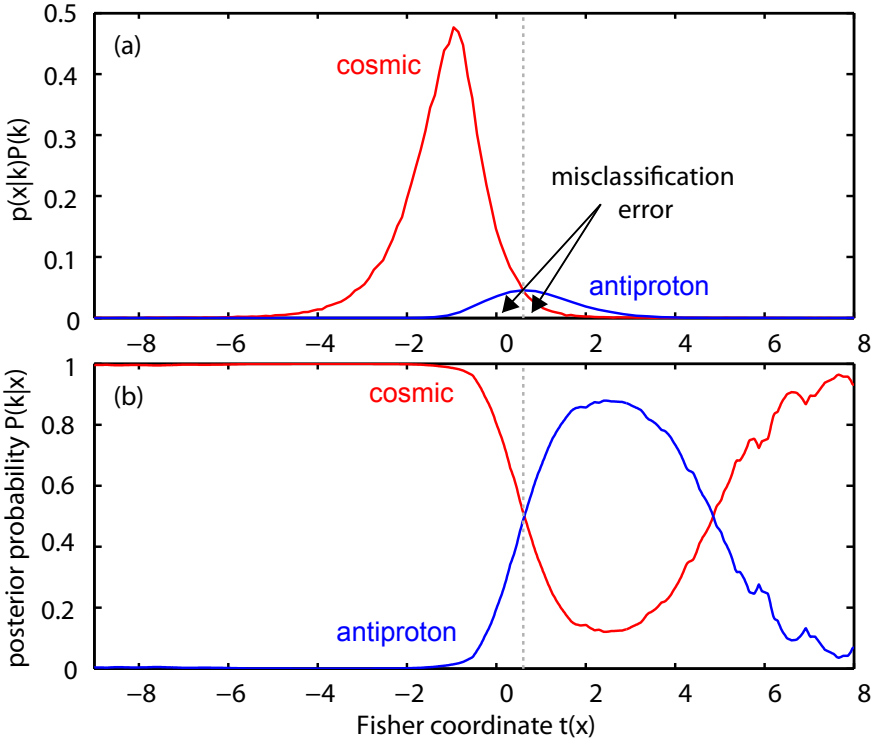


Figure 8.6: (a) Antiproton and cosmic histograms weighted by the prior probabilities of the $\overline{\mathrm{p}}$ and cosmic occurrence. (b) Posterior probabilities are determined from Bayes' theorem. The dashed line indicates the optimal threshold.

penetrates significantly into the $\overline{\mathrm{p}}$ distribution for nearly its entire distribution. Increasing the separation between the class histograms would alleviate this and increase the region of high $\overline{\mathrm{p}}$ probability.

Mapping the optimum Fisher threshold to $\overline{\mathrm{p}}$ likelihood (Fig. 8.3b) gives a threshold $\overline{\mathrm{p}}$ likelihood of 0.89 and a total misclassification error rate of 6.6%. The contribution from $\overline{\mathrm{p}}$ misclassification is 4.6% and the contribution from cosmic misclassification is 2%. This optimal likelihood gives a higher confidence in $\overline{\mathrm{p}}$ events than the previous 0.82, found from Monte Carlo simulations that minimized the uncertainty on reported $\overline{\mathrm{H}}$ events.

The standard deviation can also be calculated using the efficiency and background found from the optimal threshold. The expectation value of a binomially-distributed random

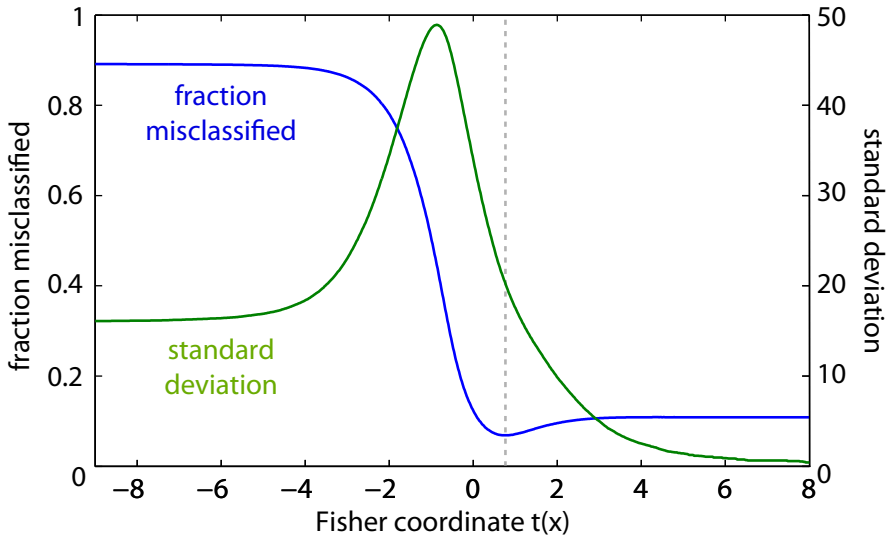


Figure 8.7: Misclassification error and standard deviation as a function of Fisher coordinate. The dashed line indicates the optimal threshold.

variable is $E = nP$ where n is the number of events and P is the probability of success. Every event is classified as either a $\overline{\mathrm{p}}$ or a cosmic, and the standard deviation on the counts reported as $\overline{\mathrm{p}}$ is found from the variance of a binomial distribution, given by $\sigma^2 = nP(1-P)$. The probability of detecting a true $\overline{\mathrm{p}}$ with this analysis is $P_{\overline{\mathrm{p}}} = 31\%$. If $n_{\overline{\mathrm{p}}} = 100$, then $E_{\overline{\mathrm{p}}} = 31$, and $\sigma^2_{\overline{\mathrm{p}}} = 21.5$. The probability of a misclassified cosmic getting through is $P_c = 2.2\%$. If $n_c = 820$, then $E_c = 18.3$ and $\sigma_c^2 = 17.9$. Adding the variances in quadrature yields a standard deviation of $\sigma = 6.3$ counts, and scaling the counts by the detection efficiency of 0.31 yields a final standard deviation of $\sigma = 20$ $\overline{\mathrm{p}}$, or 20%. A similar calculation using the efficiencies and background rates for the previous analysis yields a slightly higher standard deviation of 22.4%.

The threshold that optimizes the Bayes misclassification error, given by $P(\overline{\mathrm{p}})(1 - P_{\overline{\mathrm{p}}}(\vec{x})) + P(c)P_c(\vec{x})$, does not also minimize the binomial variance $\sigma_c^2 + \sigma^2_{\overline{\mathrm{p}}} = n_{\overline{\mathrm{p}}}P_{\overline{\mathrm{p}}}(\vec{x})(1 - P_{\overline{\mathrm{p}}}(\vec{x})) + n_cP_c(\vec{x})(1 - P_c(\vec{x}))$, as seen in Fig. 8.7. However, it is apparent that increasing the optimal threshold by increasing the separation between particle classes would lead to

smaller standard deviations in addition to further decreasing the misclassification error.

Finally, we can determine how many $\overline{\mathrm{H}}$ atoms would be needed to observe a 3σ signal above background with this optimal scheme. Assuming a Poisson distribution of background, detection of $\overline{\mathrm{H}}$ in one trial with 99.7% (3σ) confidence requires at least 2.8 counts, or 9 atoms. This permits the observation of $3\sigma/\sqrt{N} = 9/\sqrt{N}$ in N trials, where $3\sigma/\sqrt{N}$ is a standard error reported with 3σ confidence. This is an improvement over the previous $12/\sqrt{N}$ atoms in N trials. Translating the improved numbers to our trapped-$\overline{\mathrm{H}}$ results would have resulted in 3 orders of magnitude lower probability that our signal corresponding to 105 $\overline{\mathrm{H}}$ was due to background (at 1.5×10^{-10}), or equivalently, a 7.7σ result. Although our reported probability and confidence of 10^{-7} and 6σ, respectively, is already a strong result, the higher confidences obtained in this analysis will be important for experimental trials with much smaller numbers of atoms. Additionally, we can use this improved detector performance with our new CTRAP Ioffe magnet. Assuming a worst-case turn-off time of 50 ms, corresponding to the quadrupole magnet (Chap. 9), and a background rate of 41 Hz and $\overline{\mathrm{p}}$ efficiency of 54%, 8 $\overline{\mathrm{H}}$ atoms would need to be trapped in one trial to observe a signal. With this analysis, only 2 atoms would need to be trapped, and the improvements suggested in Sec. 8.5 may give us single-atom resolution for the quadrupole trap. The best case turn-off time for the octupole trap (13 ms) using this analysis already gives us single-atom resolution per trial.

The threshold selection procedure and error estimates described in this section differ from the previously reported analysis in that Monte Carlo simulations are not needed in either case. Previously, Monte Carlo simulations determined the threshold that minimized the standard deviation in resulting simulated $\overline{\mathrm{H}}$ counts. The Monte Carlo data accounted for finite statistics: not all bins were well-represented in the population of events and further artificial data was needed to facilitate statistical calculations. In this analysis, the problem

of sparsely-populated bins does not occur. The smoothness of the training data histograms reflects that the amount of training data used is sufficient. The condition that the number of training events exceed the number of parameters is met by two orders of magnitude. The ability to select a threshold and calculate error using real rather than simulated data is one of the key advantages of this analysis method.

8.5 Proposal for further analysis

The analysis presented thus far is in principle sufficient to make single-atom detection resolution possible with the CTRAP octupole magnet. However, if better sensitivity for the CTRAP quadrupole magnet or for the BTRAP Ioffe trap is desired, improvements can be made in the analysis. Specifically, it is possible to increase the separation between the probability distributions of $\overline{\mathrm{p}}$ and cosmic classes from what was obtained here using the bare minimum of information (the raw measurements). This will result in even higher efficiencies, lower background rates, higher confidences and smaller error bars.

One way to improve the discrimination between $\overline{\mathrm{p}}$ and cosmics is to "purify" the training data. For example, in the $\overline{\mathrm{p}}$ training data set used in the present analysis, 11% of the events were cosmics, leading to unwanted overlap. Fig. 8.8 shows the event rates for both training data sets used in this analysis. The cosmic event rate histogram shows a uniform 41 Hz noise, as expected. The $\overline{\mathrm{p}}$ training data event distributions however, can vary depending on how the the spill-out was done. Fig. 8.8b shows a long flat region followed by a large bump of $\overline{\mathrm{p}}$ loss at the end. If only the flat part of this data set had been used, a large fraction of cosmics would have infiltrated. If only the end of the data set were used, multiple $\overline{\mathrm{p}}$ annihilations could have been recorded in the same event. Both extremes do not well-represent actual $\overline{\mathrm{p}}$ events during antihydrogen annihilations, and this variation in

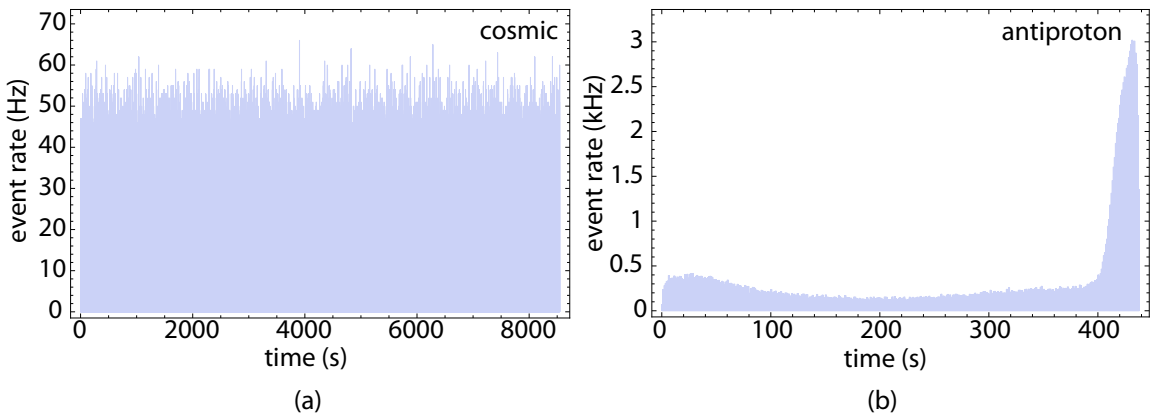


Figure 8.8: Training data event rates for (a) 300 000 cosmics and (b) 155 000 antiprotons. Only the first 155 000 cosmic training events were used in the analysis.

$\overline{\mathrm{p}}$ training data could lead to systematic errors. Using the first 29 000 training events, for example, gave a larger separation and better figures of merit than the ones reported in this chapter using the full 155 000 events shown in 8.8b.

This problem can be alleviated by using the lower-rate $\overline{\mathrm{p}}$ data and filtering out the cosmic events. One way is to split the lower-rate $\overline{\mathrm{p}}$ training data into two parts, using the first part as training data, and the second part as test data. The test data can be projected down to 1-D using the optimal projection vector found from the $\overline{\mathrm{p}}$ and cosmic training sets, and if this projection is a bimodal distribution, the known number of cosmic events that most closely match the known cosmic distribution (with the smallest Fisher coordinates) thrown out. This can be repeated using the first part of the data as test data.

In addition to improving the training data, the inclusion of additional variables based on physics (along with raw detector measurements) can only improve the separation between $\overline{\mathrm{p}}$ and cosmics or leave it the same, because variables that have little influence will be weighted less. Since the number of training events is much greater than the current number of variables, we are free to add more variables, which can be thought of as fit parameters. For example, the previous analysis selected 12 variables based on physical assumptions, but

did not use raw measurements, and some of those 12 variables could also be included in the present analysis. Note that Fisher's method gives the optimal *linear* separator, so it is only worthwhile to include new variables if they are nonlinear functions of the raw data. Suggestions for other variables are listed below.

- Particle trajectories: Cosmics arrive at the detector mostly vertically with a $\cos^2\theta$ dependence, and so will never take on a horizontal trajectory, whereas annihilation pions can. Therefore, for those events with 2 fiber hits, the angle from the vertical can be deduced and used as a separation parameter. Chi-square deviations from straight line trajectories could be another parameter.

- Detector Hits: Since there are more annihilation pions per event than cosmics (expected to only be 1 per event), they will hit more detectors. The number of detector hits per event, including both paddles and fibers, can be counted and used as a variable. Zooming into 8.5a, for example, already showed that the fiber multiplicities per fiber layer were higher for $\overline{\mathrm{p}}$ than for cosmics, leading to the spikes seen in 8.5b. Additionally, a cosmic is unlikely to hit more than 1 paddle (and also satisfy the event trigger condition), so the number of paddle hits alone can also be used as a discriminatory variable. Currently, the identification and number of paddles that were hit are not used as variables, because these are not raw detector measurements. The identification can be deduced however, based on whether or not the corresponding PM has a TDC value below the time-out value (this issue is discussed in more detail in Sec. 2.5).

- Energies: Annihilation pions may deposit a higher total energy in the detectors than cosmics, since they hit more detectors. The energy (ADC) values for all detector hits in an event can be summed up and used as a variable. Currently, the energies for each

PM, whether or not its corresponding detectors were hit, are used in the analysis since these measurements are found in the raw data. Energies for PMs only corresponding to hits should also be extracted from the raw data and analyzed.

8.6 Summary

We have demonstrated how linear methods in multivariate statistics can be used to classify $\overline{\mathrm{p}}$ and cosmic data given large training data sets. Several advantages over the old method were explored and described. The formalism for optimizing the dimension reduction, optimizing the threshold, classifying the particles, and determining error bars is clear and straightforward. The continuity of the data removes the problem of large jumps that result from discrete binning, removing sensitivity to small parameter changes. Simulated Monte Carlo data sets are not needed to deal with sparsely populated bins–errors and optimal thresholds are instead determined using real data. In this method, many additional variables can be added freely with no detriment (until we approach the training data limit), but rather with expected improvement. Because the Fisher algorithm selects for maximum discrimination, we needn't be too concerned about inadvertently adding useless variables, since these will be weighted at or near zero. In contrast, in the old method variables are not weighted. Adding more variables in that case could be detrimental if the variables are not well-chosen, thereby increasing dimensionality and making the data exponentially more sparse. This ultimately yields larger uncertainty in the analysis. Finally, the performance of the present analysis method using only raw data already surpasses the old method in signal to noise ratio, $\overline{\mathrm{p}}$ likelihood, error, and confidence levels, and methods for further improvement of detector resolution have been suggested. The minimum number of atoms needed for high statistical significance, needed to analyze data on a trial-to-trial basis has

been lowered, and combining this analysis with use of the new CTRAP Ioffe magnet yields single-atom resolution. This will allow exploration and optimization of a parameter space for trapped antihydrogen experiments that could not be done before. The next chapter describes the magnet which will give us single-atom detection capability.

Chapter 9

Next-generation Ioffe Trap

The heart of our next-generation antihydrogen trap, CTRAP, is the next-generation superconducting Ioffe magnet. We developed the new magnet in collaboration with Advanced Magnet Lab, and it has many improvements over the first-generation BTRAP magnet. The key features of both magnets are summarized in Table 9.1. First, CTRAP's magnet has the capability to produce either quadrupole or octupole radial Ioffe fields. It also contains an extra set of mirror coils, which, along with larger electrode diameters, gives the quadrupole configuration a larger trap depth than that of its predecessor. Finally, the new magnet coils have orders of magnitude lower inductance, leading to faster turn-off times which are achieved without the need to purposefully quench the magnet as before. The faster turn-off times will allow for higher atom-detection sensitivity, since the antihydrogen signal would be observed before cosmic ray background counts could interfere. This chapter describes the new magnet, attempts to use it on CTRAP, a dedicated magnet-testing apparatus, and the electrical testing results including measurement of the turn-off times.

Magnet	**Trap Depth (mK)**	**L (mH)**	**R (Ω)**	**Max. I (A)**	**Measured τ (ms)**	**Calculated τ (ms)**	**Stored Energy (J)**	**Power (W)**
BTRAP								
Quadrupole	375	3300	—	69	1000	—	7856	7856
Pinch	—	∼2000	—	80	—	—	∼6400	—
CTRAP								
Quadrupole	580	112.6	1.95	500	47.9	57.9	14125	294762
Octupole	375	18.8	1.14	680	13.2	16.7	4601	348077
Pinch	—	107.0	3.89	390	32.1	29.3	8167	254479
Bucking	—	19.8	1.85	340	10.6	11.7	1127	106632

Table 9.1: Comparison of the BTRAP and CTRAP Ioffe traps. CTRAP's listed trap depths are for its new titanium enclosure (elaborated on in text). Currents are the maximum operational currents. Stored energy and power dissipated are based on maximum current. τ is the fastest possible magnet turnoff time, measured with 10 Amps of current through the magnet, and also calculated using the full RLC magnet circuit model with the R and L values indicated and a C of 1680 μF. R comes from 4-wire resistance measurements of the dump resistor and L was measured at Harvard.

9.1 CTRAP Ioffe magnet

9.1.1 Trap fields

The next-generation magnet can generate both quadrupole and octupole fields, shown in Fig. 9.1a. BTRAP's radial quadrupole magnetic field, described in Secs. 2.3 and 6.5, was problematic because its steep gradient led to large charged particle losses as antiproton and positron clouds were brought together [65]. Because the addition of the quadrupole trap fields break cylindrical symmetry, the plasma expands, bringing the charged particles to the cutoff radius for following the field lines into the trap wall. As a result, we experienced particle losses even though the plasmas started with radii below the cutoff radius.

The smaller gradient of CTRAP's octupole magnetic field [28] provides a larger (less stringent) cutoff radius and makes the trap more cylindrically symmetric near the trap axis. This may allow the reduction of plasma radius before ramping up the Ioffe trap to be more effective, and should lead to fewer charged particle losses and therefore more efficient $\overline{\mathrm{H}}$ trapping.

On the other hand, the quadrupole field offers a larger trap depth than does the octupole. In order to maximize the octupole magnet's usable field, the magnet is positioned as close to the electrodes as possible, as shown in Fig. 9.1a. The quadrupole field also provides tighter spatial confinement, which would allow the atoms to fit within the focus of the UV laser beam that will be used for spectroscopy. Having a magnetic trap that can create both kinds of fields would give us the option to trap $\overline{\mathrm{H}}$ in the octupole field and then switch to the quadrupole for spectroscopy. It would also allow us to directly compare the performance of both fields.

The vertical field profile is also improved in the new trap. The field from the pinch

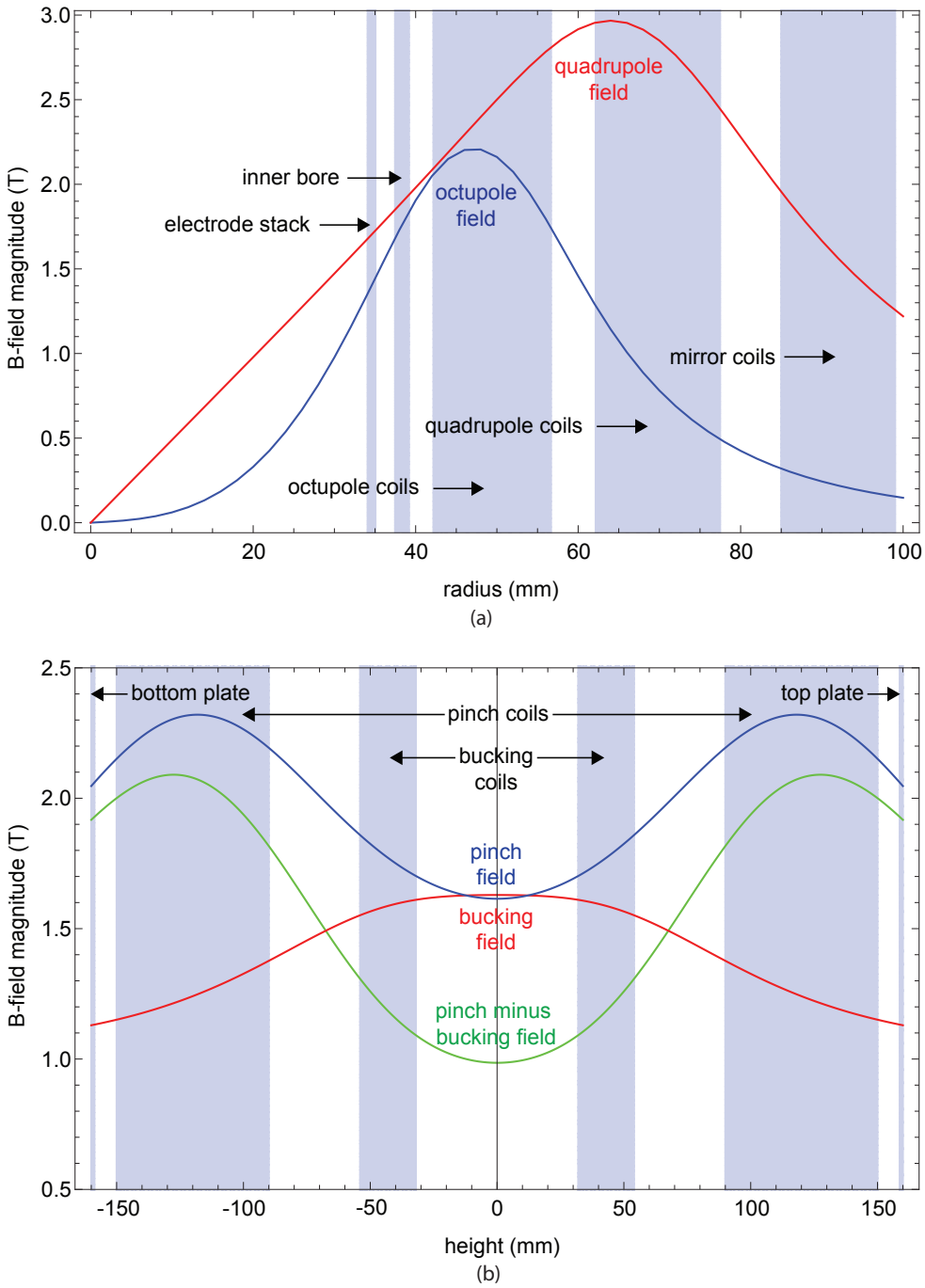


Figure 9.1: (a) Octupole and quadrupole fields. (b) Pinch and bucking fields. The shaded regions indicate the locations and extent of the electrodes, magnet coils, and the titanium helium enclosure.

coils creates a large offset in the center of the trap, due to the coils' small axial separation and large radius. The compensating bucking coils that run at the opposite sense current to the pinch coils cancels this axial field at the trap center, shown in Fig. 9.1. This in addition to the larger electrode IDs give the CTRAP quadrupole configuration a larger trap depth than that of BTRAP (see Table 9.1).

9.1.2 Operational currents

In order to achieve the tens of ms CTRAP magnet turnoff times, orders of magnitude decrease in inductance from BTRAP's was necessary. The lower inductance is achieved with fewer windings, and therefore significantly larger currents are needed to attain similar magnetic fields as BTRAP's. To accommodate the higher currents, larger cross section of wire is used. Thus, for the same B-field, the current density for the BTRAP and CTRAP magnets is similar.

Table 9.2 (top) gives the specifications of the magnet coils. The critical current refers to the short-sample critical current, found experimentally by the wire manufacturer. This is done by running current through a short strand of wire in an external field to map out the superconducting phase transition as a function of current density, external magnetic field, and temperature. This phase boundary indicates the ideal performance of the wire. However, windings in a magnet are not supported or cooled as well, so the phase transition (quench) will happen at lower current density or field.

The nominal current is the operating current that Advanced Magnet Lab claimed could be realistically met given the short-sample currents, and represent operating currents for a full trap. There are large margins (difference between the ideal critical current and the nominal current) for the nominal current. The octupole for example, has a current margin of 40% and a temperature margin of 1 K, indicating that the coil can potentially be

Operational Current Specifications		
	Critical Current (A)	**Nominal Current (A)**
Octupole	946*	676
Quadrupole	600*	500
Pinch	772.6	393
Bucking	772.6	337
Trap Currents		
	Octupole Trap (A)	**Quadrupole Trap (A)**
Octupole	680	0
Quadrupole	0	500
Pinch	330	390
Bucking	250	340

Table 9.2: (Top) CTRAP individual magnet operational currents. Critical current refers to short-sample critical current. Nominal currents are based on a 1 T superimposed external field. (Bottom) Full trap currents for Ioffe traps in the octupole and quadrupole configurations. *Deduced from specifications–this was not explicitly stated by the company.

run at higher currents than the nominal current, and slightly higher temperature, before it quenches.

The currents in Table 9.2 (bottom) are the currents used by all coils for the indicated octupole or quadrupole trap configuration. The axial and radial well depth is the same for a trap. Such balancing of the radial and axial confinement optimizes the trap depth, defined as the difference between the lowest maximum magnetic field magnitude and the minimum magnetic field magnitude. The B-field from one coil limits the maximum current achievable in another due to the critical B-field constraint for superconductivity. The trap-depth balancing was done in an iterative way: if the lowest maximum is in the radial direction, and assuming the radial coil is already at as high of current as it can take, lowering the axial coil current a little allows more current to be put into the radial coil. Iterating in this way leads to a balanced trap and a maximal trap depth. Note that the mirror coils run at lower currents for the octupole trap, which has a lower trap depth, than the quadrupole trap.

9.1.3 Windings

The magnet is put together in 3 concentric layers (Fig. 9.2) - the innermost layer contains the octupole windings, the middle layer contains the quadrupole windings, and the outmost layer contains the mirror coils [16]. As shown in Fig. 9.2, in between layers are grooves that were machined down the length of the G10/epoxy support structure, to allow liquid helium to be in as close contact to the windings as possible.

The superconducting wire is made of NbTi (a Type II superconductor) filaments embedded in a copper matrix. The multipole windings have a copper to superconductor ratio of 1:1 while the solenoids have a higher 2.7:1 ratio. The octupole and quadrupole wire diameter are 0.85 mm, and the rectangular mirror coil wires are 1 x 1.3 mm in cross section.

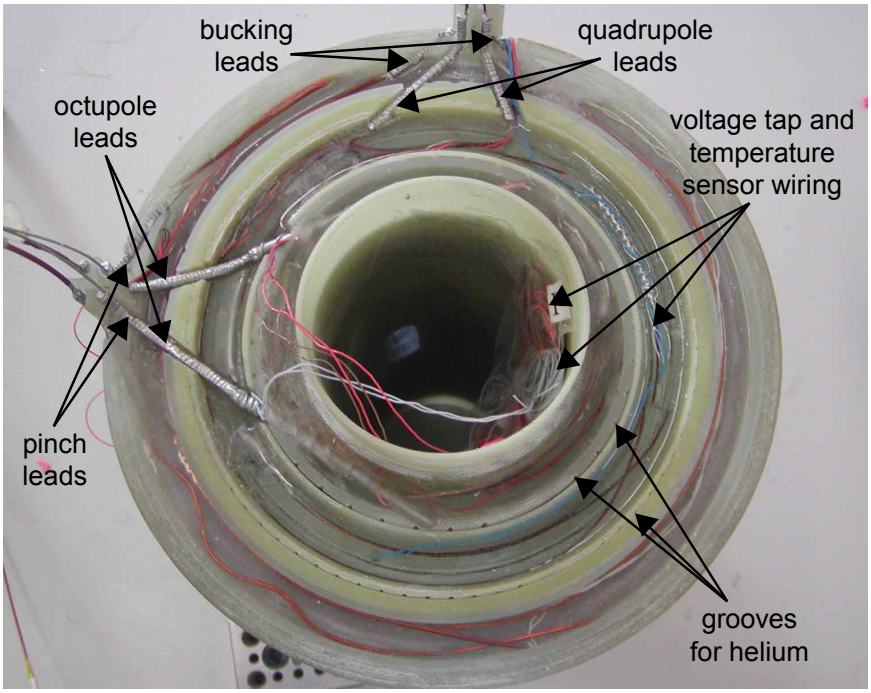


Figure 9.2: Top view of the bare magnet before final assembly inside the helium enclosure. The superconducting windings are embedded in the G10/epoxy support structure. Visible are the 4 sets of superconducting leads and grooves for liquid helium. The wires seen inside the bore are for voltage taps and temperature sensors.

The filament diameters for the radial and vertical coils are 6 and 82 microns, respectively. The very small octupole and quadrupole filament size was chosen to facilitate the stringent ramp rate requirements of the magnets.

Generally, the fact that the windings are in the form of superconducting filaments rather than bulk wire helps to prevent quenching in superconducting magnets. If there is any heating within a superconductor, which can be caused for example by penetration from an external, changing magnetic field, the critical current density drops. For Type II superconductors, this causes the screening currents which normally expel external magnetic fields to decay, allowing more magnetic field penetration. This flux motion within the superconductor generates heat, decreasing the critical current even more and causing further

flux disturbance. If the effect is strong enough, it will cause an avalanche of heat and flux motion, also known as a flux jump. In a flux jump, the effective heat capacity of the superconductor goes to zero and any small perturbation will cause the temperature to rise without limit [76], [77]. This results in quenching.

The solutions to this problem require reducing the flux motion caused by the decrease in critical current, or just conducting away the heat caused by the flux motion. Both can be accomplished by making fine divisions of the superconductor. For a given volume of superconductor, the trapped flux (and thus the magnetization and energy stored) is proportional to the thickness, so fine filaments minimize the residual field in the superconductor. By filamenting the wire and surrounding it with copper, the surface area of the superconductor is increased, and the copper can carry away heat. The copper also provides a path for current when the superconductor is driven normal, and spreads currents between the filaments if one filament gets damaged.

9.2 First magnet-testing attempts

The magnet construction with its helium enclosure was finished and shipped to CERN in 2011. As discussed in Sec. 2.3, CTRAP was built just for this new magnet. Therefore, finishing the design, wiring, and construction of CTRAP so we could install and use the new magnet for better $\overline{\mathrm{H}}$-trapping was a primary goal for the 2012 beam run. We succeeded in completing the apparatus but then suffered a series of vacuum leaks during attempts to cool it down, rendering us unable to test the new Ioffe trap that year.

The leak was found in the liquid helium enclosure, meant to house the magnet and keep it submerged in liquid helium. Most of this enclosure was made of G10, an electrical and thermal insulator, to avoid eddy currents and the tricky task of welding a metal enclosure

too close to the windings. Additionally, the enclosure geometry made possible by using G10 maximized the inner bore diameter, giving the largest possible trap depths. As such, the enclosure constituents had to be joined together using Stycast 1266, a cryogenic epoxy. This glue was chosen because its exceptionally low viscosity allowed it to penetrate into 0.003" gaps, and its coefficient of thermal expansion matched that of G10. Though the enclosure along with 2 other full-sized prototypes had passed several cold-tests on their own before installation, in the final configuration the glue joints did not survive the CTRAP cooldown attempts and leaked helium. This section describes the cooldown attempts and methods.

9.2.1 First CTRAP cooldown attempt

A typical antihydrogen trap cooldown requires pumping out the insert dewar and experiment spaces to $\sim 10^{-5}$ Torr, cooling down the insert dewar shields with compressors, purging the trap's liquid helium dewar of air with gaseous N_2, and then starting LN_2 fills in the LHe space. After cooling to 77 K, we purge the LHe space of N_2, and then fill with LHe until we reach 4 K. Most of the trap is cooled conductively by thermal contact to the 4 K LHe dewar, while the magnets are submerged directly in LHe as the liquid flows from the dewar through the helium space of the Ioffe magnet and into the helium space of the $\overline{\mathrm{p}}$ solenoid, as described in Sec. 2.4. Resistive heaters on the hat controlled by Variac voltage supplies keep it at room temperature, so as to not freeze the Oring seal between it and the insert dewar.

BTRAP cooldowns have taken as little as 2.5 days. However, with the first CTRAP cooldown, because the robustness of the glue joints in the G10 enclosure was unknown we proceeded with great caution. Temperature sensors were installed in many places on the cryostat and also on the magnet in order to monitor temperature gradients on different regions of the magnet. To start the cooldown, the objective was to gently flow LN_2 into the

LHe space at intervals rather than continuously, and cool the trap with its boiloff without letting liquid accumulate. Gradients were monitored, and whenever gradients approached 30 K the fill was stopped. The projected cooldown time for this slower cooldown was over a week.

In the first CTRAP cooldown in March of 2012, a gradual increase in vacuum pressure on Day 2 suggested that a leak had opened up, but we continued the cooldown in the hope that as the trap approached LN_2 temperature cryopumping would greatly improve the vacuum. To achieve the vacuum in the insert dewar and experiment spaces we used a roughing pump (Edwards ESDP12, dry oil-free scroll pump, which can get pressures down to 1×10^{-2} Torr) connected in series with a turbo pump (BOC Edwards Turbomolecular pump EXT 255H). We pumped on both spaces at the same time with the same pumps, which had not been done for previous BTRAP cooldowns.

On Day 6 it was confirmed that there was a leak between the helium space and the experiment space. The cooldown was aborted at 170 K. Because we had to be careful with temperature gradients, we could not cool down quickly enough to cryopumping temperatures, $<$ 120 K [78], to risk continuing to operate the turbo pump at such a high pressure. The backing pressure (between the roughing pump and the turbo pump) was at an acceptable 1 Torr pressure, but the inlet pressure (in front of the turbo pump) was 2 orders of magnitude above the maximum allowed for continuous pumping, which is tens of mTorr. Not complying with these specifications causes the turbo pump to heat up, which can liquefy the oil on the bearings and cause the lifetime of the pump (typically 3-4 years) to drop.

After warming up, a leak-check confirmed that the leak was caused by glue joints on the G10 enclosure between the helium space and the experiment space. With LHC beam time approaching quickly we decided somewhat desperately to attempt salvaging the G10

enclosure with glue patch-ups. We tested several kinds of glue and settled on Stycast 2850, which is a more flexible epoxy than Stycast 1266 and survived our rather violent cold-testing. We scraped off as much of the excess Stycast 1266 as possible, and applied the Stycast 2850 over the old glue joints while also pumping on the enclosure.

9.2.2 Second CTRAP cooldown attempt

For the second cooldown attempt in October of 2012, we required not just a slower but also more gentle way to cool down. To this end we installed a new helium flow pipe that bypassed the helium dewar and let us inject cryogen directly into the $\overline{\mathrm{p}}$ solenoid, which is suspended under the Ioffe magnet. This was to allow the $\overline{\mathrm{p}}$ solenoid's boiloff to gently vapor-cool the Ioffe trap. This time, we pumped on the vacuum spaces separately.

Additional temperature sensors were installed on either side of the glue joints to monitor gradients directly across them. There were a total of 15 Cernox temperature sensors on the cryostat (CX-1030-CU), around glue joints (CX-1030-SD-HT)), and on the Ioffe trap coils inside the magnet (calibrated CX-1050-AA-1.4L), which had already been installed by the company. An AC resistance bridge (Lakeshore Model 370) along with a scanner (Lakeshore Model 3716) were used to cycle through the temperature sensors and measure their resistances, which are temperature-dependent. The calibration between resistance and temperature for CX-1030-CU was:

$$log(T) = c_0 + \frac{c_1}{log(R)} + \frac{c_2}{log(R)^2} \tag{9.1}$$

where c_0, c_1 and c_2 are fit parameters found from 3 given temperature-resistance points provided by Lakeshore at 300, 77, and 4 K.

For the calibrated coil sensors, 100 temperature-resistance data points were provided. The 4-parameter fit function [35] normally used for CX-1030-CU sensors below 4 K,

displayed in Eq. 9.2,

$$log(T) = \frac{c_0}{R^{c_3}} + \frac{c_1}{log(R)} + \frac{c_2}{log(R)^2} \tag{9.2}$$

was used to calibrate the sensors because it gave very good agreement with the provided data points to ± 0.5 K for all coils.

For the second cooldown, the gradient threshold was lowered from 30 K to 10 K. Despite all this care, one week into the cooldown, a leak occurred again between the helium and experiment spaces. Once a leak opened up, the cooldown could no longer proceed in a controlled way. As liquid nitrogen leaked from the helium space and made contact with the experiment space, thermalization time dropped and gradients spiked. Again we continued the cooldown after the leak was detected hoping we could at least turn on the Ioffe trap, if not trap antimatter. However, the leak became so large that the pressure buildup in the experiment space became a concern. Leaving the experiment space unpumped for 10-15 minutes caused the experiment pressure to rise to 30 Torr, well past its normal 10^{-7} Torr baseline.

The cooldown was aborted, the Ioffe trap was removed from CTRAP, and the G10 vacuum enclosure was machined off.

9.2.3 Enclosure

The removal of the original G10 enclosure necessitated the design of a new metal one that could be welded together. The problem with a metal enclosure is that eddy currents can cause magnetic field distortions and heating (and therefore loss) of charged particles. Since the eddy currents decay on the order of milliseconds, they are especially worrisome for CTRAP, due to its millisecond scale turnoff times and large magnet currents. In the bare magnet, this problem is precluded by the windings' G10-epoxy composite support structure (in contrast to BTRAP, whose support for the windings was made of titanium). However,

a metal enclosure includes a non-insulating inner bore, which is close enough to octupole coils to pose a problem (see Fig. 9.1a). The previous G10 enclosure did not have a separate inner bore at all since it was glued directly to the magnet. This allowed us to have a larger-diameter upper electrode stack and increased magnet trap depths. Because of the new metal inner bore (made as thin as possible), we needed to decrease the size of the electrodes and therefore lose some trap depth. Still, the new electrodes are significantly larger than the BTRAP electrodes.

Titanium was a natural choice for the new enclosure, since it is non-magnetic. Careful measurements and calculations were done to find a titanium alloy that would not become superconducting at our temperatures and magnetic fields (and cause persistent eddy currents), and to ensure that eddy current heating would not be a problem. Grade 5 titanium, used in BTRAP, was not chosen because of its strong superconducting properties in these conditions. Grade 2 was not considered because of its large conductivity. Grade 6 and Grade 9 were close candidates, but in the end Grade 9 was chosen because of purity and availability issues. The new titanium enclosure has now been welded over the Ioffe magnet and appended to CTRAP, ready for use.

9.3 Cryogenic apparatus for magnet testing

In the rush to get the second-generation Ioffe trap into the antiproton beam in 2012, the electrical properties of the Ioffe trap remained to be carefully tested. We expected AML to perform such verifications before shipping the magnet, but this had not been done, leaving the testing up to us. The downtime at CERN the following year (2013-2014 LHC Long Shutdown 1) afforded us the opportunity to do extensive testing outside of beam time.

Testing the magnet on CTRAP rather than separately would have let us incorpo-

rate the full system with the correct heat loads, cooling mechanism, and 1 T bias magnetic field. However, this was no longer possible after the G10 enclosure was machined off, since the bare magnet could not be appended to CTRAP and the titanium enclosure was not yet available. Therefore, a separate cryogenic test jig for stabilizing and cooling the magnet was constructed.

Considerations in designing this jig included compatibility with an existing bucket dewar, heat load minimization, pressure regulation, eddy current minimization, accommodation of both the bare magnet as well as the forthcoming titanium enclosure, and accommodation of future vacuum spaces. Fig. 9.3 shows the lead and magnet assembly in the test jig.

9.3.1 Materials and heat loads

To minimize the appearance of eddy currents that might cause inaccurate magnetic-field measurements, the jig pieces in the vicinity of the magnet are almost entirely made of G10. To minimize the conductive heat load from 300 K, all structural pieces extending down from the hat are also made of G10. For electrical connections from the hat downwards, very small diameter (0.003" - 0.01") constantan wire was used, with the exception of the heater wires (made of copper). Constantan wire is often chosen in cryogenic applications because its thermal resistance is relatively high. A copper plate located 6" below the hat served as a thermal anchor to 77 K. The anchoring was done via flexible beryllium-copper fingermounts that are bolted to the perimeter of the plate. These "cold-fingers" are in good thermal contact with the liquid nitrogen sleeve of the bucket dewar. Whenever possible, wiring from the hat was heat-sunk to this anchor. Any tubes extending from the hat down into the liquid helium bath were either thin-walled stainless steel or made from insulating materials. For example, the guide tube for the liquid helium transfer stinger has 0.01" wall

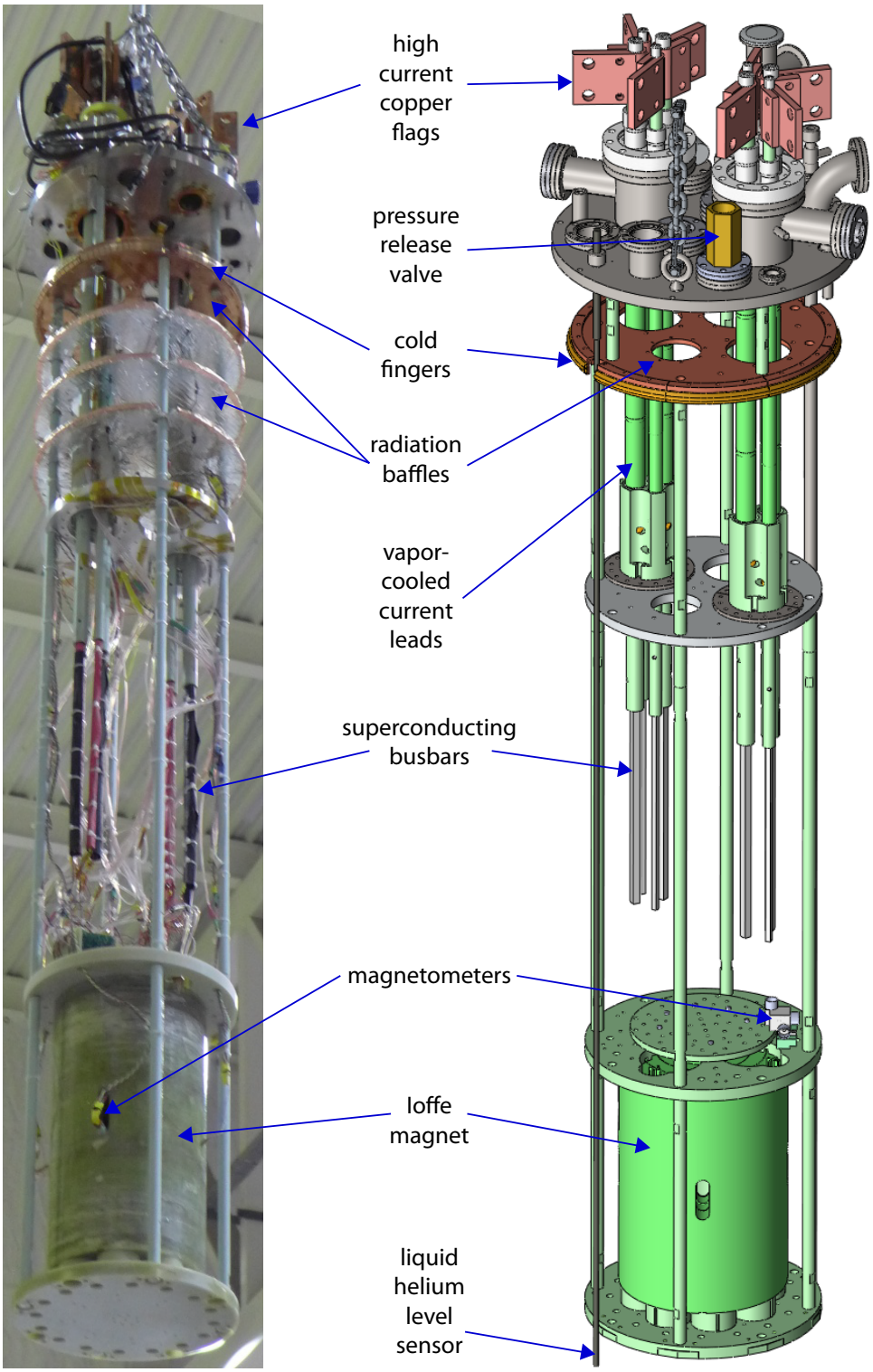


Figure 9.3: Magnet testing jig and essential components.

thickness and is made of stainless steel, a poor thermal conductor compared to commonly used metals. The liquid helium level sensor, also extending from the hat into the liquid helium, is made of G10 and superconducting wire. These two pieces, with heat loads of 50 and 30 mW, respectively, are the only pieces in the jig not anchored to 77 K and are also the biggest source of heat when the retractable vapor-cooled current leads (VCCLs) are retracted.

The VCCLs are by far the largest heat load on the system when they are engaged, with a specified boiloff of 3.3 L/hr (though we observed much higher) when three of the four sets of magnet leads (a full trap) are at full current. For comparison, the total conductive boiloff from all other parts of the jig was calculated to be ∼100 mL/hr. Because of the very high currents used in the operation of the magnet, vapor-cooling of the non-superconducting retractable leads is essential. They are detached from the busbars when not in use, minimizing the heat load on the liquid helium, whose loss rate is 1.4 L/(W h) at atmospheric pressure. The thermal mass of the retractable leads was designed such that when they are seated, the boiloff from the leads themselves provide sufficient self-cooling across the full range of operating currents. The cooling of these leads by the helium vapor flow through them in turn reduces the boiloff of liquid helium due to the leads. This includes boiloff that would arise from both thermal conduction from the 300 K hat and also resistive heating from electrical conduction. The busbars, made of a soldered laminate of Nb3Sn/Cu (Type II superconductor), are always submerged in liquid helium and so are always superconducting. Therefore there is no heat load coming from resistive heating of the busbars. The retractable leads and the busbars were made by American Magnetics.

The 77 K copper plate serves as a radiation shield, which, if no holes in the shield were present, would reduce the radiative heat load by a factor of about $(\frac{293}{77})^4$, or 210. At times, the helium boiloff vapor-cooled the plate to colder than 77 K (measured), reducing

this factor. This happened when several sets of leads were engaged, or with sustained high current. A shiny aluminum plate and three cardboard radiation baffles covered entirely in Mylar super-insulation further reduce the radiative heat load by a factor of 5. All plates in the jig have circular cutouts that allow the lead assembly to pass through. As long as the appropriate exhaust channels are opened, these cutouts force the helium vapor to flow alongside the leads for cooling on the outside as well as the inside of the leads. Furthermore, the helium vapor is guided around the perimeter of the baffles, cooling the wall of the bucket dewar and the stinger guide tube and the helium level sensor, which are both located just inside the dewar wall. The presence of these plates also break the flow of warm helium gas at the top of the dewar back down into the cold helium, limiting heating due to convection.

The 100 L capacity bucket dewar made by Kadel Engineering was also designed to minimize conductive and radiative heat loads. The helium space is only in contact with the nitrogen space through a 1.65" thick aluminum ring and is otherwise separated from it by vacuum. The aluminum ring is isolated from the liquid helium by a large G10 break, which prevents conduction of the 77 K heat load along the wall of the dewar, and is also isolated from the 300 K hat by a G10 break. The vacuum between the helium and nitrogen spaces prevents air from acting as an exchange gas between the 4 K and 77 K space, and once the spaces are filled with cryogen, cryopumping of the vacuum space via activated charcoal improves the vacuum to 10^{-5}-10^{-6} Torr. Superinsulation between the 300 K and 77 K walls of the dewar minimizes radiative heat loads from the periphery of the dewar.

9.3.2 Exhaust and pressure system

A minimum pressure differential is required to maintain a sufficient flow rate of cold helium gas through the vapor-cooled leads. Pressure was regulated by sending the exhaust through variable flowmeters and also appropriate choice of pressure relief valves.

The desired pressure in the dewar was one high enough to push the necessary amount of helium boiloff through the leads but low enough to avoid damaging the dewar.

The minimum pressure was calculated using the boiloff rates provided by American Magnetics, assuming that the leads themselves provide all the vapor needed (as claimed by the company) and Poiseuille's law for hydrodynamic laminar flow through a tube:

$$P_1 = \sqrt{\frac{256 R\nu T}{\pi}\frac{L}{D^4}n + P_2^2} \tag{9.3}$$

where P_1 and P_2 are the pressures on either side of the tube, R is the universal gas constant, ν is the viscosity of the gas for a given temperature T, D is the diameter of the tube, and n is the gas molar flow rate. The molar flow rate is determined from the boiloff rate from the vapor-cooled leads. In the limit of small pressure differentials, applicable here, Eqn. 9.4 becomes

$$\Delta P = \frac{128 R\nu T}{\pi}\frac{L}{D^4}n\frac{1}{\bar{P}} = \frac{Q}{C} \tag{9.4}$$

where $Q = nRT$ is the gas volume flow rate, $C = \frac{\pi D^4 \bar{P}}{128\nu L}$ is the conductance, and $\bar{P}$ is the average pressure.

The maximum pressure was estimated to be 2 atm above atmosphere (29 psig) using Barlow's formula for burst (or permanent deformation) pressure of pipes: $P = \frac{2ST_W}{D_O F}$, where S is the yield strength of the aluminum dewar, T_W is the wall thickness, D_O is the outer diameter of the helium vessel and F is a safety factor, chosen to be 5. The safety factor can range from about 1.5-10, with 1 giving the collapsing pressure. In the event of the pressure relief valve failing, there are also either rubber stoppers or latex hoses on all exhausts that would pop off or break before the bucket dewar bursts.

There are several exhausts on the jig: the Ioffe lead exhausts which consist of 42 vent holes in each large vapor-cooled lead and 21 holes in each small lead, and the main exhausts. The main exhausts include two large holes in the top plate (known as the hat) and

an exhaust on the helium stinger guide tube that prevents a pressure buildup inside it when filling with LHe. The vents are very thin-walled (0.005" or 0.010" wall thickness) reentrant tubes that are designed to guide the helium vapor out of the apparatus with minimal contact with the hat. This is to prevent the hat from from frosting up–indicative of inefficient use of the cold vapor or unacceptably high boiloff. The hat is kept at room temperature using heaters, so as to not freeze the O-ring making the helium-tight seal with the dewar. All helium gas exiting the exhaust ports is recovered and recycled.

The helium exhaust system for the testing jig was set up in the AD hall in the same way as it will be for CTRAP when it is commissioned. All exhaust lines were connected to a manifold that allows separate control of each Ioffe lead and main exhausts. These lines are normally sent through flowmeters and then the recovery system, but can be redirected to bypass the flowmeters and go straight into the recovery line. This is useful when doing high-pressure helium fills. All exhausts can also bypass both the flowmeters and recovery line and be vented straight to atmosphere in case of an emergency–for example, a violent quench.

9.3.3 Magnetometers

Hall probes (Cryomagnetics, HSP-A) were installed within the Ioffe trap to measure magnetic field. The data acquisition was required to be fast enough to capture the ∼10 ms turnoff of the magnets. Magnetic field measurements were done by sending a maximum DC or AC current of 100 mA through a probe and measuring the Hall voltage caused by the external B-field deflecting the current within the probe. The sensitivity of the voltage measurement is proportional to the current applied, and the measured voltage is converted to B-field using known calibrations for each Hall sensor (50-80 mV/T sensitivity). Each probe can measure the B-field component only in the direction of its axis.

Three Hall probes were installed in the sideports near the center of the magnet (see left side of Fig. 9.3), where there is good signal resolution even at low currents due to the probes' proximity to the coils. This enabled us to send DC currents (Agilent E3612A power supply) through the probes and measure voltage signals without amplification, and these signals were used to find the turnoff times of all the magnets. Two probes were positioned to measure radial and vertical fields at the inner bore tube of the magnet. The third measured the radial fields at the center of the magnet (expected to be small but nonzero since the radial coils are not infinitely far away). The positions of these magnetometers were temporary for magnet testing, since there is no space for them in the sideports of the titanium enclosure. One of these three sideport magnetometers is shown in the photo in the left side of Fig. 9.3.

The nearest possible position to the magnet for the magnetometers in the final CTRAP setup is on top of the enclosure. This is significantly further away from the coils giving rise to a much lower signal to noise ratio. To see whether resolution of the signals in this position was possible we also installed 3 other magnetometers on the jig in the final CTRAP position (see right side of Fig. 9.3). The coordinates of the three magnetometers (facing the X, Y, and Z directions) on CTRAP were chosen to optimize the signal from the octupole field given space constraints in mounting them. The octupole gives the weakest signal due to its rapid falloff with distance. For these weak-field measurements, amplification was required and the setup used is shown in Fig. 9.4. A 20 kHz AC current was applied to the probes and their output 20 kHz Hall voltage was sent to a preamplifier with a bandpass frequency range of 10-30 kHz and gain of 100. The output of this was fed into a lock-in amplifier that picked out the amplitude of the 20 kHz signal and amplified it by 10 000. An ethernet DAQ module was used to record this data at a sampling rate of 2 kHz, high enough to resolve the fast exponential turn-off curve of the magnet.

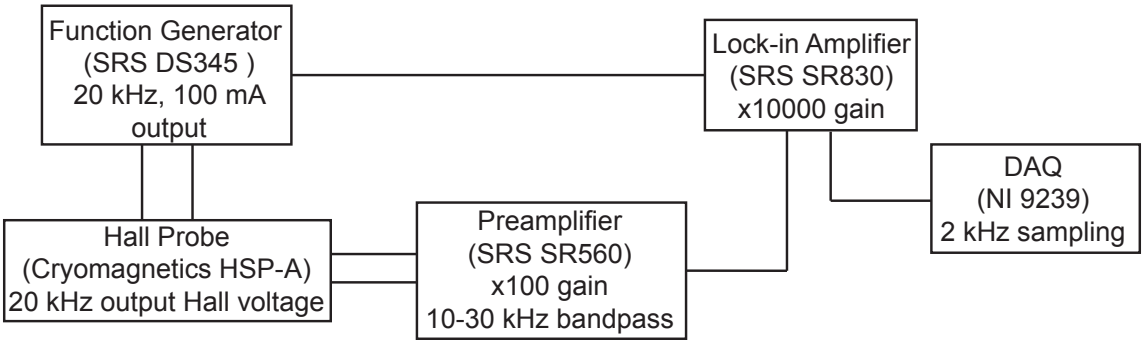


Figure 9.4: AC magnetometer measurement setup.

9.3.4 Cooldown procedure

After placing the completed testing jig with magnet into the helium space of the bucket dewar, we purged this space of air with N_2 gas in preparation for filling the outer nitrogen sleeve with liquid nitrogen. The nitrogen exhaust was vented out a latex hose. When the purging was done, a floating ball flowmeter was connected to the end of the (long) hose to prevent backflow of air.

We then purged the outer LN_2 sleeve with nitrogen gas. We filled through one quick-connect while exhausting out of another quick-connect on the sleeve. After purging for a few minutes we switched to LN_2, but just before switching we clamped the exhaust and over-pressurized the sleeve with gaseous N_2. This was to ensure an overpressure in the time span that we remove the N_2 gas line and insert the liquid nitrogen stinger, so that there is no backflow of air into the sleeve. Filling up the sleeve with LN_2 took a little over an hour.

The dewar and jig were then left to thermalize overnight. The next day the entire jig was at 77 K through contact with the cold fingers and convective cooling from the exchange gas in the helium space (N_2). Because the helium space had already been purged, we did not have to worry about water vapor freezing onto the cold fingers, or worse, getting into the small grooves in the Ioffe magnet. Water expands as it freezes and has the potential

to cause cracks and other damage in the magnet. Though all temperature sensors read 77 K, we filled the helium space with some liquid nitrogen anyway (first making sure the space was overpressurized with nitrogen gas as before), just to be certain that the relatively large thermal mass of the magnet was at 77 K. Liquid nitrogen was filled until at least the bottom plate of the jig was submerged, after which we let the jig thermalize with the liquid for an hour. Filling up the sleeve and pre-cooling the helium space of the bucket dewar with LN_2 took a little over 3/4 of a 220 L liquid nitrogen dewar. Pre-cooling with liquid nitrogen is necessary both because its cooling efficiency is significantly higher than cooling with just liquid helium, and also to conserve helium. Additionally, when the surfaces of the dewar that are in contact with cryogen are cooled with liquid nitrogen, cryopumping improves the insulating vacuum between the outermost wall of the dewar and the cryogen spaces. This prevents convective heating.

Next came cooling with liquid helium. We purged the helium space of nitrogen liquid and vapor by flowing gaseous helium through one of the exhausts, and collecting the liquid from a blow-out tube that had been inserted into the fill port. After all the liquid came out, we flowed gaseous helium for 5 more minutes, and appropriate exhausts were opened to ensure that gaseous nitrogen in trapped spaces was also forced out.

Finally we switched to LHe. Pressure gauges on the hat, the transfer dewar, and at the recovery line were monitored to ensure that there was no backflow of helium from the bucket dewar back into the transfer dewar, which can happen if the recovery pressure is too high. Before transferring the liquid helium, the transfer line was purged of air and inserted into the bucket dewar once there was a visible plume of liquid helium coming out of it. The transfer dewar was pressurized to 4.5 psig and liquid helium was filled to the maximum level of the top of the busbars, 41.5" above the bottom. To stop the fill, we stopped pressurizing the transfer dewar and once pressures went down to safe levels we removed the transfer line

first from the bucket dewar side, quickly stoppering the quick-connect, and then from the transfer dewar side.

9.4 CTRAP magnet operation and quench protection

A significant operational difference between the BTRAP and CTRAP magnets is the turn-off mechanism. While the BTRAP Ioffe magnet needed to be purposefully quenched to achieve the minimum turn-off time (detailed in Sec. 7.2), CTRAP's Ioffe magnet is turned off by switching its current into dump resistors. CTRAP's quench protection system, outlined in this section, serves the dual purpose of protecting the windings from a quench and also switching off the magnet in normal operation.

Damage to the CTRAP magnet windings in the event of an accidental quench is of grave concern. If the quench propagation is slow compared to the energy dissipation time, then the very high stored energies can cause the G10 insulation to char, or worse– a small length of the winding to be vaporized by the heat, which would destroy the entire magnet. Superconducting magnets with small stored energies typically do not require quench protection. For CTRAP, if all the power generated from the quadrupole coils (Table 9.1) were dissipated in the magnet, the coils would heat up to hundreds of Kelvin above room temperature, making quench protection critical. In contrast to BTRAP, the CTRAP Ioffe magnet uses an active rather than passive quench protection system to allow the magnetic field to be much more quickly turned off.

The CTRAP quench protection system consists of both quench detection and protection electronics, shown in Figs. 9.5-9.8. Voltage taps are attached to the center of each coil, the ends of the magnet leads, the tops of the busbars, and the tops of the vapor-cooled leads, as shown in Fig. 9.9. The tap signals from the magnet, magnet leads, and busbars

are divided down by a factor of 5 from voltage dividers (not shown) before connecting to an electrical feedthrough at the hat, to alleviate concerns about high-voltage arcing between connector pins in the helium vapor. The tap signals from the flags on top of the vapor-cooled leads are divided by 5 outside the cryostat. Exiting the hat, these signals travel to the quench protection box for further processing–there the voltages are divided down by another factor of 20, and voltage drops across symmetric sections are compared.

Each measured voltage drop across the superconducting coils consists of both a resistive and inductive component. When the magnet is in a superconducting state, the resistive part is zero, leaving only the inductive part which comes from induced e.m.f.s during charging or discharging of the magnetic field. These e.m.f.s are expected to be nearly identical for matching sections since the voltage taps are centered and each section has similar amounts of magnetic flux penetration and the same current, causing each section to have similar inductance. During steady-state superconducting operation, the measured voltage drops are expected to be zero. The voltage drop differences between matching sections are always expected to be zero.

When the magnet quenches, however, the resistive component of the voltage becomes nonzero and increases as the quench propagates. Since it is very unlikely that quenches of equal magnitude will develop simultaneously in symmetric sections of the coil, a comparison between voltage drops on these sections during a quench will show a voltage imbalance. The quench detection system checks whether this voltage imbalance exceeds a pre-set imbalance threshold, and triggers a magnet dump once this condition is met.

At this point a fiber signal from the quench detection electronics is sent to the quench protection electronics, opening the switch shown in Fig. 9.10. Insulated-gate bipolar transistors (IGBTs) were used to do the fast switching (Semikron SkiiP 1513GB172-3DL modules). All the current is rerouted into the dump resistor within tens of ms (Table 9.1

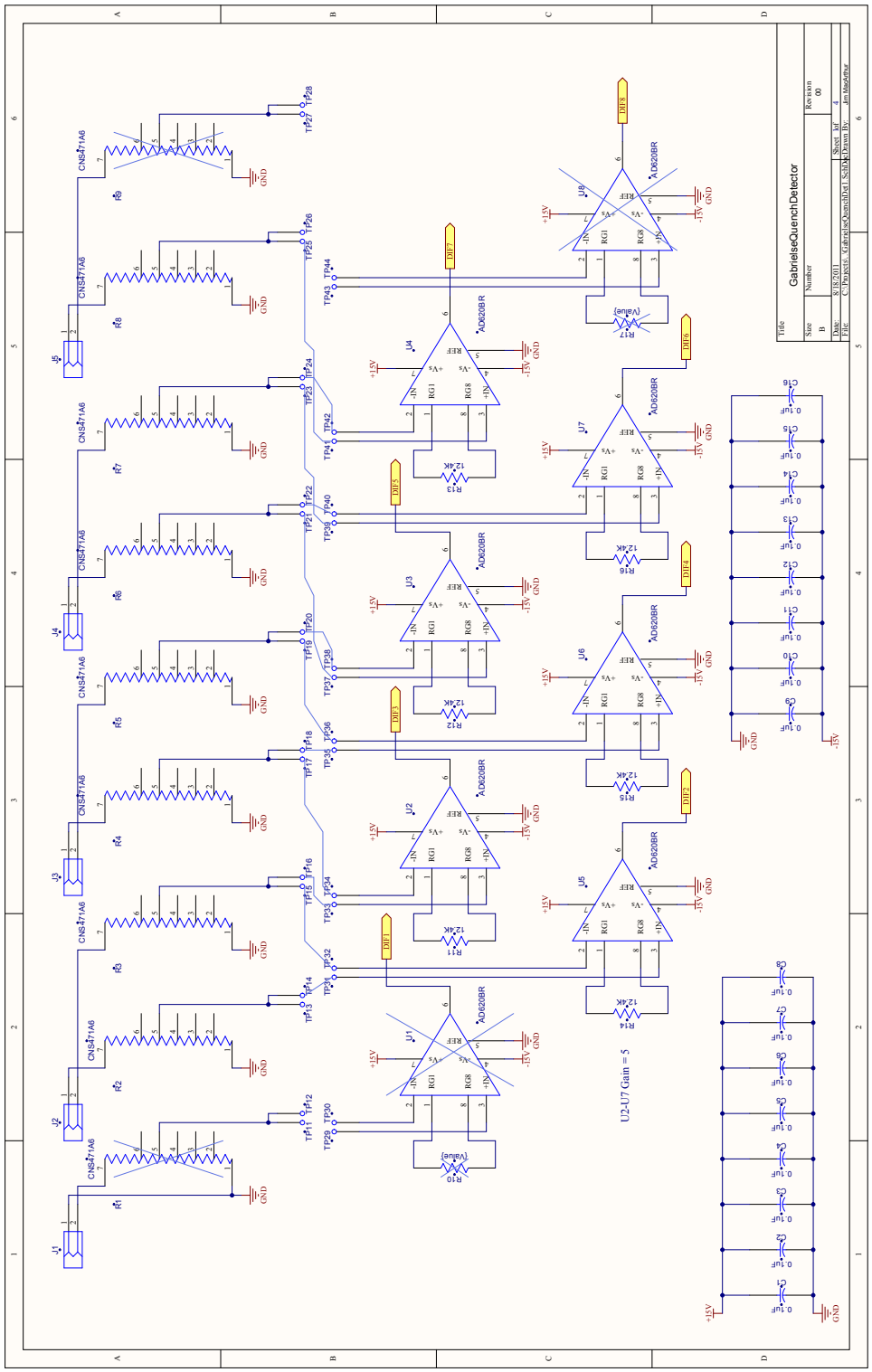


Figure 9.5: Quench wiring diagram 1: Voltage tap inputs are divided down by 20, voltage drops are determined, and half of the measured drops are negated.

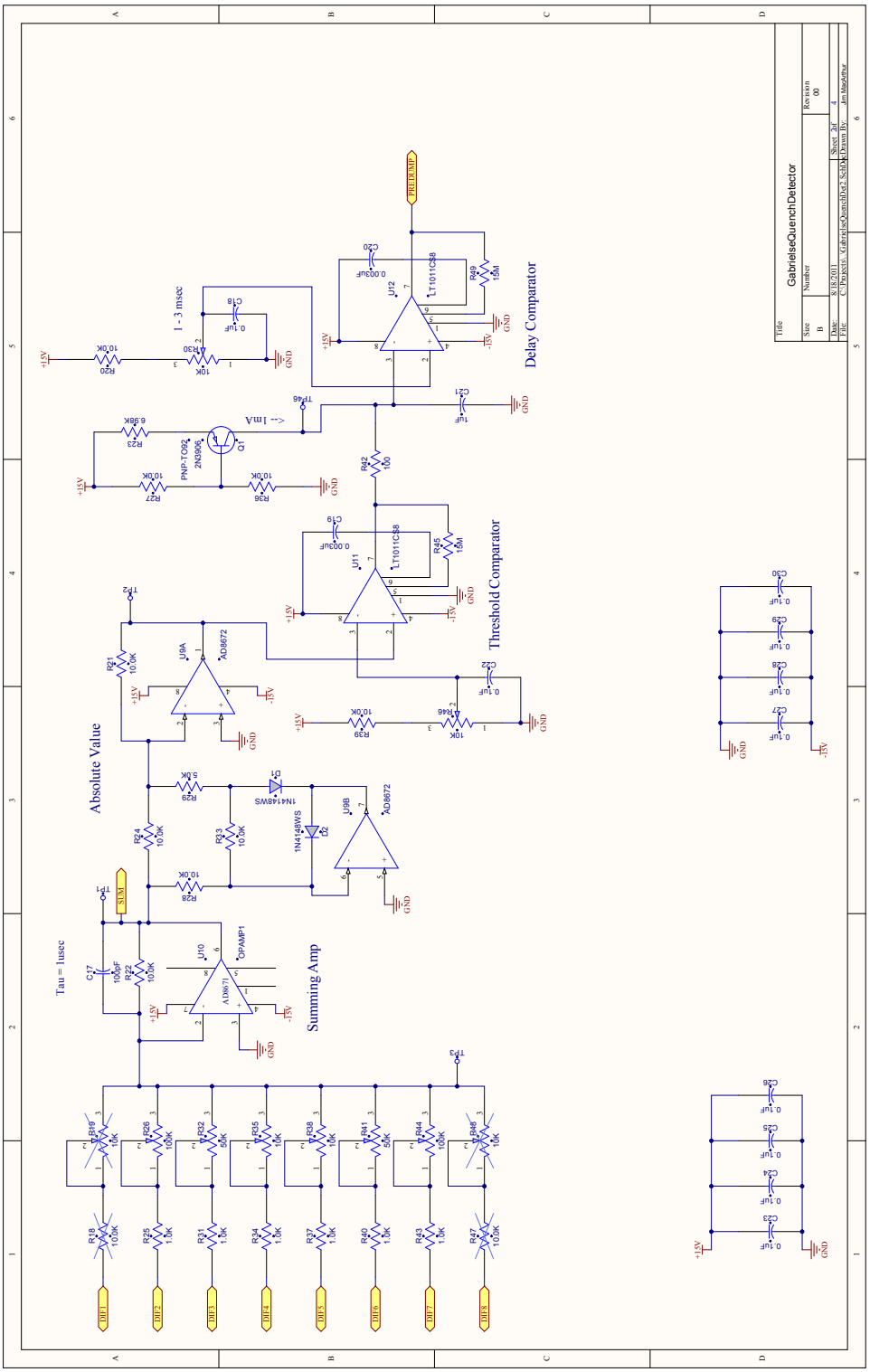


Figure 9.6: Quench wiring diagram 2: Voltage drops are weighted, amplified and summed. The weighted sum is compared with the threshold value. Holdoff time is processed.

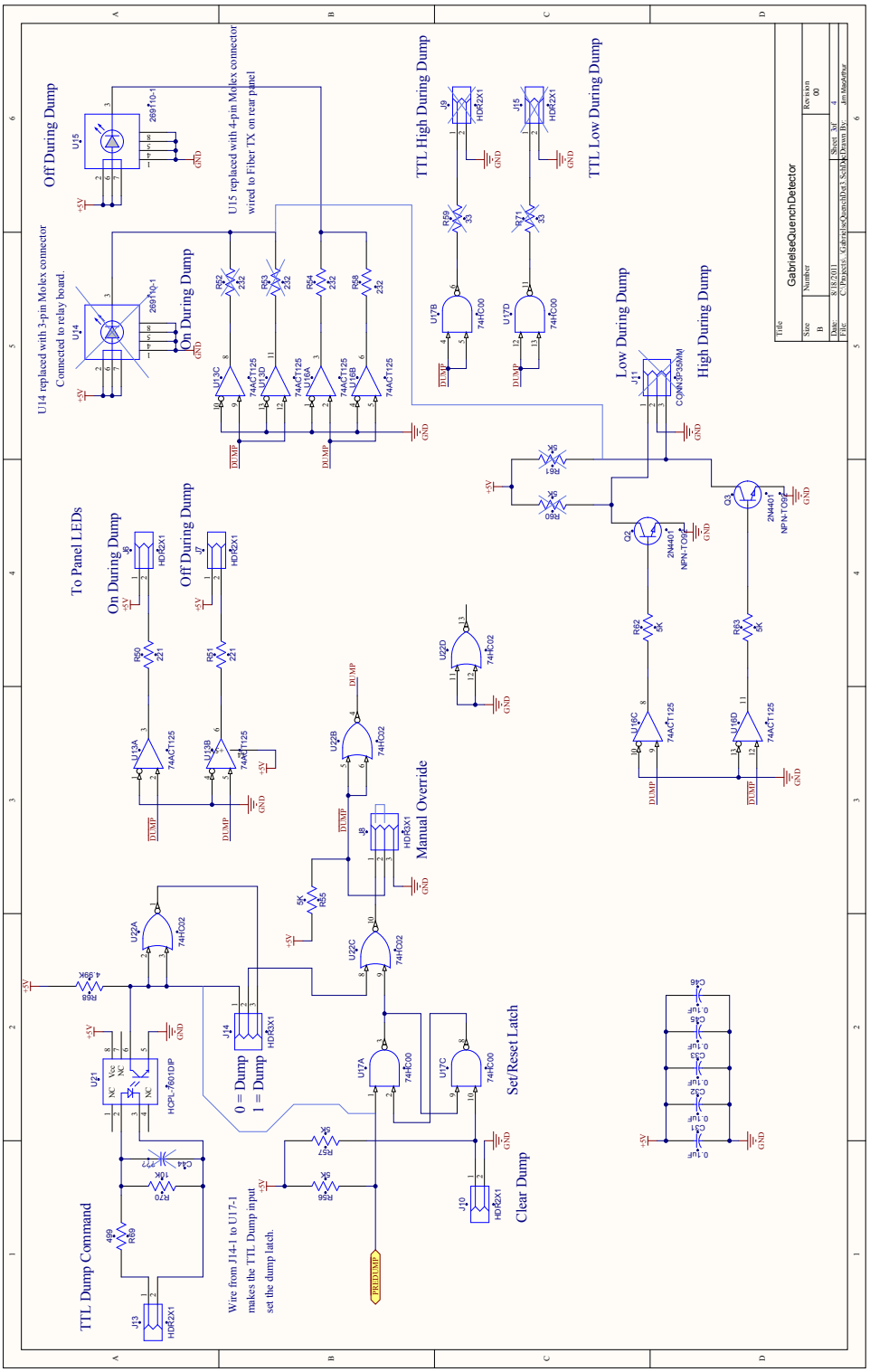


Figure 9.7: Quench wiring diagram 3: In the event of a quench, dump signals are sent to the IGBT switches, power supplies are inhibited, fast voltage data acquisition is triggered.

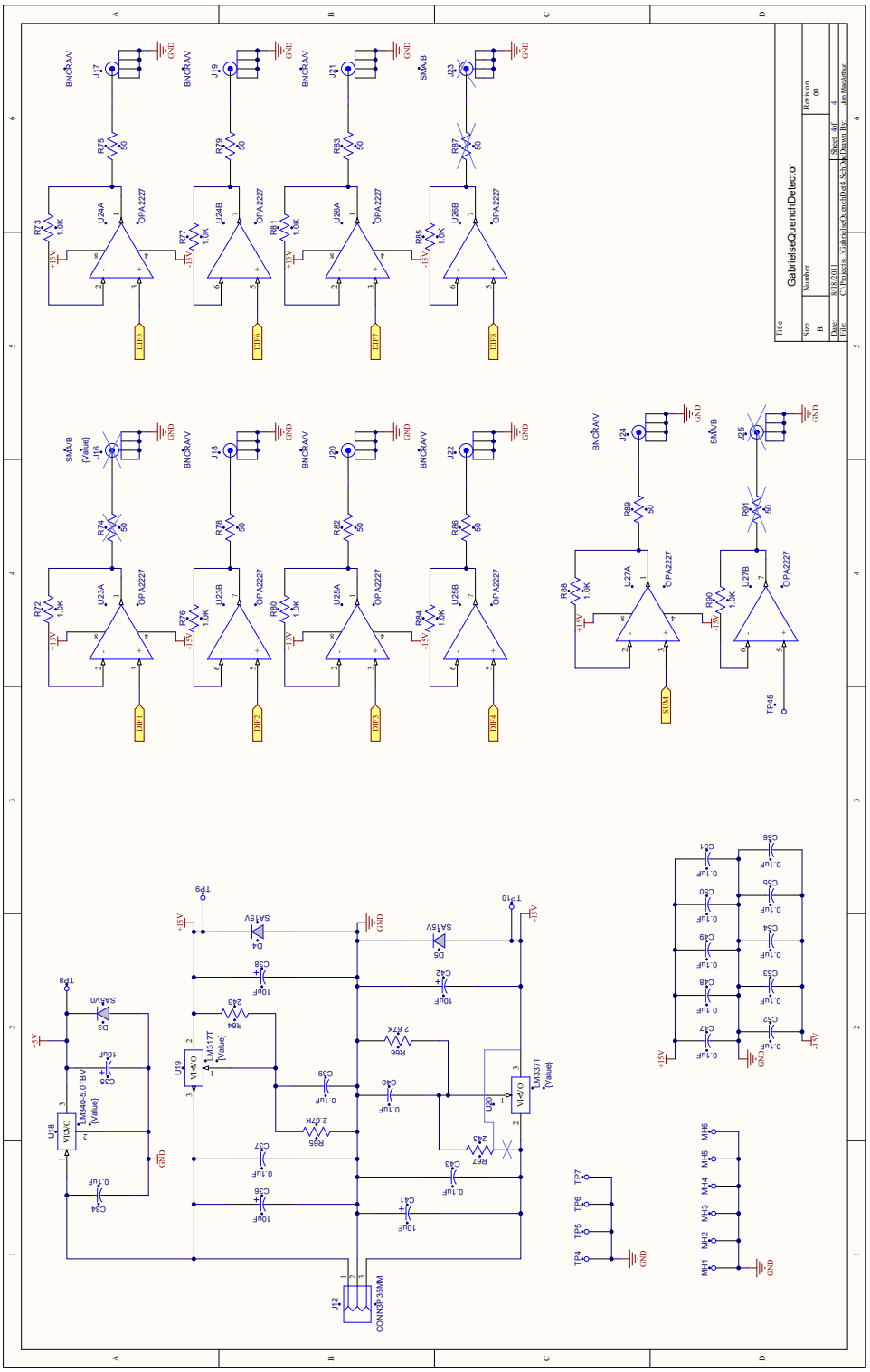


Figure 9.8: Quench wiring diagram 4: Voltage drop and weighted sum outputs are sent to DAQ and saved to disk.

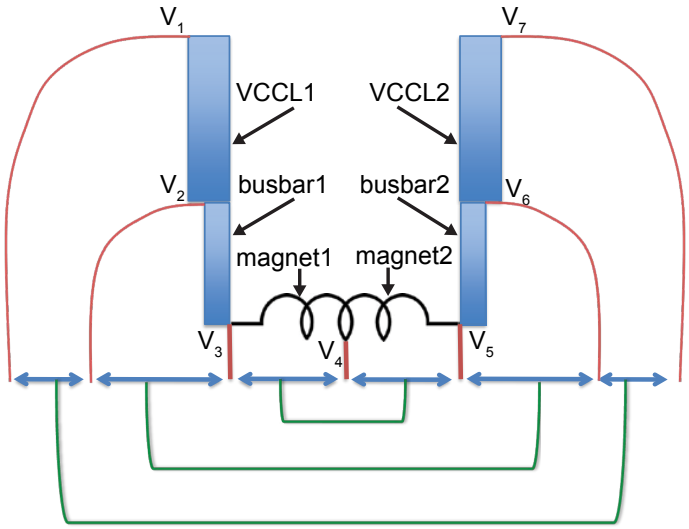


Figure 9.9: Voltage taps on 7 different positions of a single coil assembly, attached to the vapor-cooled current leads, busbars, and magnet.

shows the turnoff times for each coil), protecting the superconducting coils. At the same time, a digital signal is sent from the detection electronics to a fault input in the power supply, shutting off the output of the power supply by commanding 0 V and 0 A. The output takes 12-35 ms to shut off, and the power supplies are put in "protect" mode.

Turning off the magnet by choice, or "dumping on command," works by a similar process. In this case the user sends a command to a pulse generator, which outputs a pulse to the quench detection input to trigger a magnet dump. A fiber signal then triggers the IGBTs to switch and reroute the current into the dump resistors as just described. This is how the magnets will be routinely de-energized during future antihydrogen experiments with CTRAP.

The circuit in Fig. 9.10 should be explained. In normal operation, current only flows through the power supply, the IGBT switch (of which only one of the two drawn is actually used), and the magnet. Once the IGBT switch is opened, current flows through the capacitor for a transient amount of time until it is charged up, at which point current

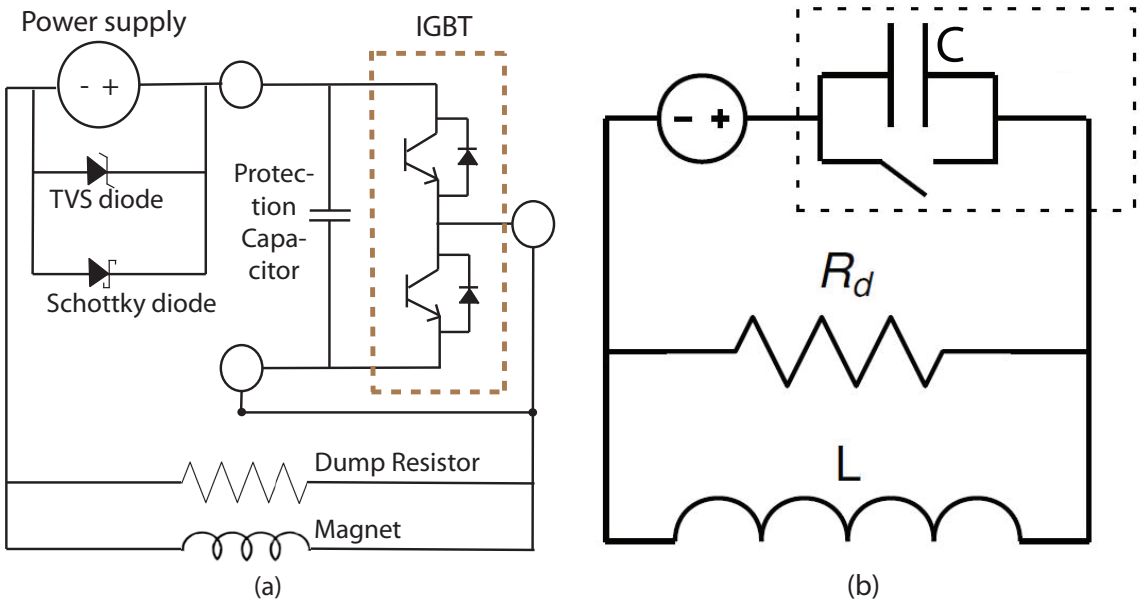


Figure 9.10: Octupole magnet quenches during the dump caused by the octupole quench.

is forced to go through the resistor-magnet loop and dissipate. The values of the dump resistors vary for the magnets and are given in Table 9.1. These were chosen to minimize the L/R turnoff time, given the potential limit of a 1 kV voltage drop across the magnet. This limitation was to avoid arcing in the helium gas where the leads and voltage taps exit towards the top of the apparatus. As mentioned earlier, voltage dividers submerged in LHe divide down the voltage of the voltage taps inside the trap for this reason. The capacitor, there to protect the IGBTs against the kickback voltage, has a value of 1680 μF. This was chosen based on the resistance values to keep the circuit critically damped. Additionally, too high of a capacitance would increase the turnoff time. In principle, the resistance values can be increased to give smaller turnoff times if arcing is not an issue, but then the capacitor would have to change too. However, we are limited by the capacitors available for our system. The diodes installed across the power supply are there to protect the supply from the large inductive voltage.

9.5 Magnet testing

The final magnet testing was done at CERN in order to operate the magnets in conditions as similar to CTRAP's as possible. Before shipping the jig, bucket dewar, and ourselves off to CERN, basic tests were done at Harvard to ensure that the magnet and jig were functioning as expected. We leak-checked the system, did a test cooldown, checked that we could apply up to 2 Amps without a quench protection system, measured the B-fields at these currents and checked that they agreed with expected values, and checked that the voltage across the magnet leads did not increase noticeably with current when it was at 4 K.

Because the quench protection system depends on voltage imbalances, it was important to check first for inherent imbalances across half-sections, which could indicate whether or not the voltage taps were properly centered or if there were any geometrical asymmetries between magnet sections. 4-wire resistance measurements were done across the full magnet and half-sections, and later on, inductances over the same regions were also measured using a Stanford SR720 LCR meter. The measurements are displayed in Table 9.3. Although these electrical measurements were not used to correct for imbalances, they gave a sense of the asymmetries before we had the full experimental setup at CERN. Induced EMF signals during the full testing were instead used to directly correct for geometrical imbalances. Our main goals for the testing at CERN were to measure the turnoff times, ensure that the quench protection system was working properly, and to quench train the magnet.

9.5.1 Experimental setup

The following describes the high-current setup at CERN. Hundreds of amps supplied by 5 kW Agilent 6681a power supplies flow through high-current cables which bolt onto the copper flags (Fig. 9.11) at the top of the lead and magnet assembly. The power

Octupole			**Quadrupole**			**Pinch**			**Bucking**		
Leads	R (Ω)	L (mH)	Leads	R (Ω)	L (mH)	Leads	R (Ω)	L (mH)	Leads	R (Ω)	L (mH)
1, 3	31.68	18.8	5, 7	54.13	112.6	6, 8	9.58	107.0	2, 4	3.44	19.8
1, center	16.49	6.0	5, center	29.55	37.1	6, center	5.52	51.7	2, center	2.34	8.7
3, center	16.05	6.5	7, center	25.9	29.8	8, center	5.65	51.2	4, center	2.32	8.8

Table 9.3: 4-wire resistance measurements and inductance measurements for the Ioffe trap magnet coils at room temperature. Measurements were made across the entire coil and half-sections of the coil to compare imbalances between sections.

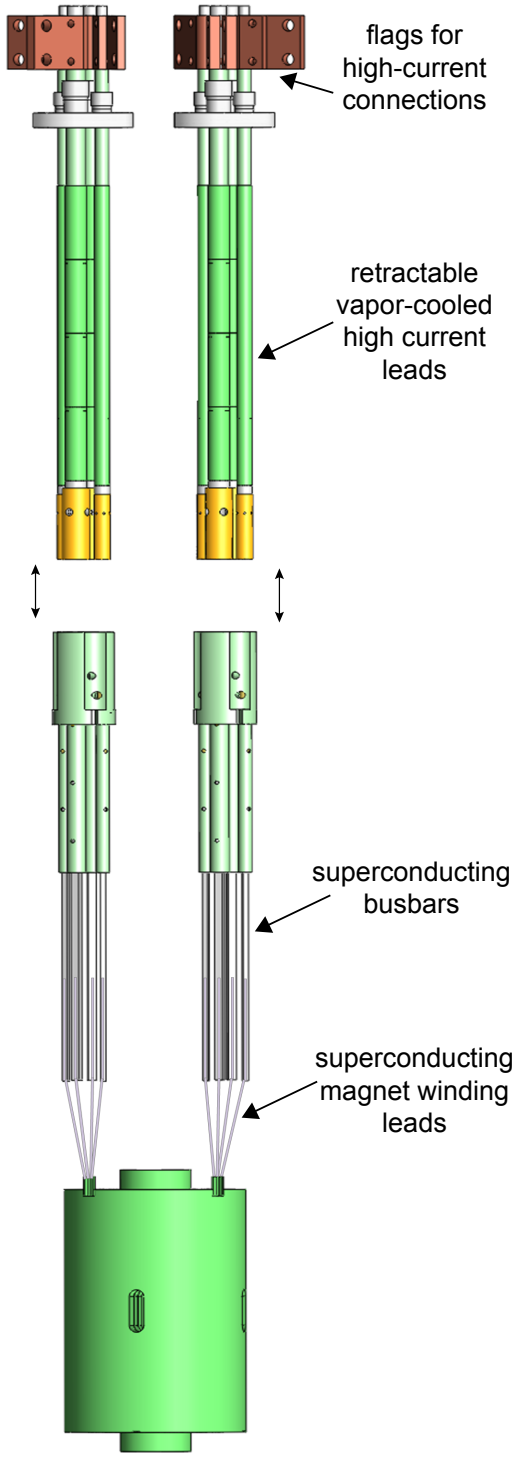


Figure 9.11: Lead and Magnet Assembly. Retractable leads (shown retracted) connect to superconducting busbars which are soldered to the magnet windings.

supplies have a maximum output of 580 A, so to supply the octupole coil with its 680 A operating current, two of these power supplies were connected in parallel in a master/slave configuration. The other coils had their individual power supplies connected directly as shown in Fig. 9.10.

As shown in Fig. 9.11, the copper flags are connected to retractable vapor-cooled leads, which connect to superconducting busbars beneath them. The superconducting busbars are soldered to the superconducting windings that extend out of the magnet. All high-current cable connections were carefully done using a procedure of cleaning, sanding, and applying a thin layer of petroleum jelly to all contact surfaces before clamping. The petroleum jelly is to make the joint air and water-tight. It is non-conductive, but the loose metal particles mixed with the jelly, and the fact that the jelly gets squeezed out upon clamping ensures that there is still a low-resistance connection.

Because of the $\sim$1 kV voltages (depending on the dump resistance) that could be induced across the circuit during a rapid turnoff, the entire lead assembly inside the test apparatus is electrically insulated to prevent arcing in the helium space. At the hat, rubber sheets are placed between the magnet flags even though electrical breakdown in air is less of a concern given the ample space between the flags. At a later stage in the testing, the superconducting magnet leads were also mechanically supported with copper rods up to the solder joints at the busbars to avoid any movement of the conductor under the influence of Lorentz forces, which would lead to heating and quenching.

9.5.2 Measuring turnoff times

The most important measurement of the new magnet was its turnoff time. The first testing was done without the quench protection system, by setting all quench-voltage thresholds to their least sensitive values. Currents on all coils were increased to 10 A by

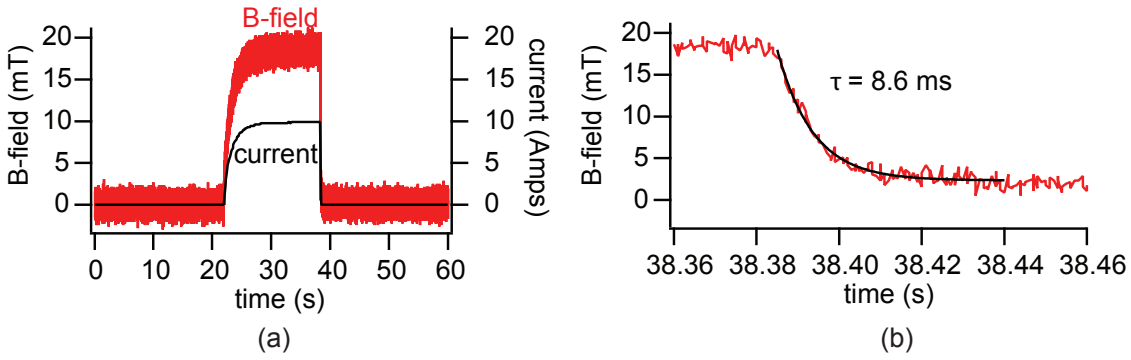


Figure 9.12: Example of turnoff measurements, illustrated with the bucking coil. (a) Magnet ramp up to 10 Amps followed by rapid turn-off. (b) Close-up of the rapid turn-off in (a) showing the exponential decay of the B-field after a magnet dump, used to find the turnoff time constant τ.

specifying a voltage limit required to achieve a desired current. In this mode, the specified voltage is applied by the power supply in less than 900 μs, and the current follows the load. This caused a smooth exponential current ramp indicative of charging an LR circuit with a constant voltage, as shown for example, in Fig. 9.12a. This sort of ramp was preferred over a current-limited ramp since the discontinuities in the current steps would cause unwanted induced e.m.f. spikes. The voltage limit was always set to be the minimum voltage necessary to achieve the desired current. To ensure that the power supply was always run in constant-voltage mode, the setpoint current was a value higher than the desired current. If a higher-than-necessary voltage was specified, depending on the setpoint current, the current would either be a smooth ramp to a higher-than-desired current or the power supply would overshoot past the specified current, switch to constant current mode, and ring down to the higher-than-desired specified current.

After reaching 10 A, the magnet was rapidly switched off by triggering a dump on command, and example measurements are shown in Fig. 9.12. For these tests, one dump command switched all 4 IGBTs. For antihydrogen experiments, we will only switch off the radial coil to release the atoms. The sharp turnoff after ramp up can be seen in Fig. 9.12a

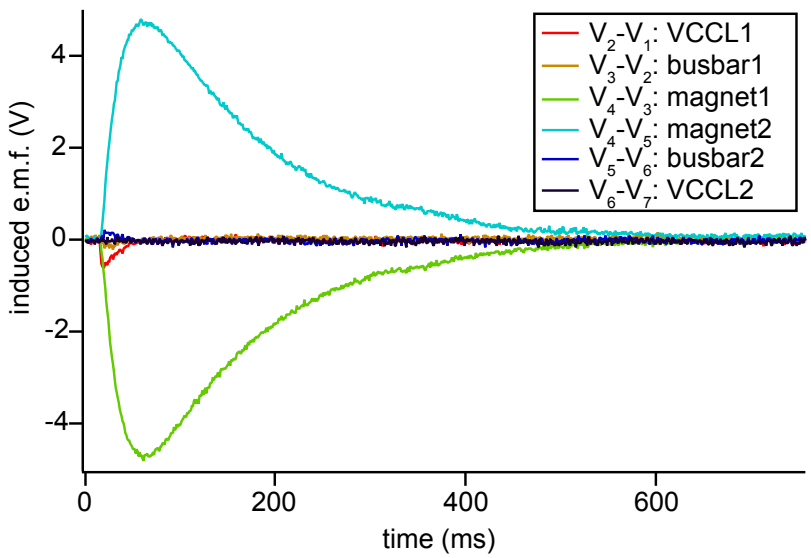


Figure 9.13: Example of a voltage imbalance measurement, illustrated with the octupole coil. The ratio magnitudes of the induced e.m.f. during a 10 Amp dump (no quench) determines the imbalance. The imbalance in the octupole coil is very small. The voltage differences labeled are as in Fig. 9.9

and a zoom-in of the dump can be seen in Fig. 9.12b. The exponential decay of the B-fields allowed us to extract the turnoff times of the magnets, summarized in Table 9.1. The turnoff times match what is expected from the LRC circuit model.

The voltage tap data acquisition system is used to measure induced e.m.f.s. Low-rate 1 Hz voltage tap and weighted sum (explained in the next section) data is always displayed and saved. Whenever there is a dump, whether caused by a real quench or manual command, the detection electronics save buffered high-rate 10 kHz data corresponding to 100 ms before the trigger and 1 s afterward. This enabled us to record the fast induced e.m.f. from the rapid turn-off, $V = -LdI/dt$, shown in Fig. 9.13. Imbalances between matching coil sections were found by comparing their induced e.m.f. amplitudes. Since the magnet currents were only 10 Amps, less than 2% of the critical currents, these imbalances were inherent and not caused by quenches.

9.5.3 Setting the quench threshold

In order to move onto higher currents, it was necessary to use the quench protection system with an appropriate threshold. The threshold refers to a weighted sum threshold of all the voltage drops across all sections. The weighted sum W is:

$$W = A \times \left[\sum_{i=1}^{3} k_{i+1}(V_{i+1} - V_i) - \sum_{i=4}^{6} k_{i+1}(V_{i+1} - V_i) \right] \tag{9.5}$$

where V_i are the voltages labeled in Fig. 9.9. The constant $A = 10(1 + x)$ comes from multiplicative factors in the electronics and voltage dividers in the apparatus. x is the value in kΩ of the source trimpot used to set the weights, k_i. The exact value of x differs for each magnet but is $\sim$4 kΩ for the nonzero weights.

The weights range from 0 to 1, and correct for inherent imbalances in the magnet/lead systems, possibly from voltage taps not being perfectly on center or other geometrical differences between sections. Ideally the two halves of the coil would have identical induced e.m.f. magnitudes and therefore have equal weights, causing the weighted sum in 9.5 to be trivially zero for all coils when there is no quench, and below the weighted sum threshold for a quench trigger. This is nearly the case for the octupole coil (Fig. 9.13). The quadrupole coil however, had the largest imbalance of 14%. If left uncorrected, the observed voltages across the 2 sections of the quadrupole coil would have erroneously caused the quench detection system to fire when the current was being increased.

To prevent false quench triggers, the voltage across section pairs had to be weighted appropriately and implemented in the quench detection hardware via the aforementioned trimpots. The weights k_4 and k_5 for the coil half-sections were found by measuring the ratios of the e.m.f. peaks for each coil section. The superconducting busbars were given equal weights k_3 and k_6 since there was no observed imbalance (no e.m.f. was induced since they are not coils so inductance is negligible). The vapor-cooled leads were given weights k_1

and k_7 as close to zero as possible, since they are not superconducting and therefore cannot quench.

A threshold of 4-5 times the maximum voltage noise level on each magnet was chosen. This corresponded to 200-250 mV above the 50 mV noise on the magnet. The threshold was set in the quench detection hardware by sending simulated quench signals into the voltage tap inputs of the hardware and adjusting the threshold trimpot until the protection system triggered at a weighted sum of 200-250 mV.

A hold-off or delay time between quench detection and trigger can also be set via trimpots on the quench detection hardware. This corresponds to an amount of time that the measured weighted sum must exceed the threshold and can be set to a maximum of 7 ms. For the 10 A measurements discussed in the previous section and the quench data displayed in this chapter this was 4±0.5 ms. For quench-training, it was 1 ms.

These settings worked well in preventing false triggers and capturing true quenches. The one specific exception is that of the pinch coil for high-current, one-step, voltage-limited ramp-ups. This triggered the quench protection system unexpectedly during this kind of ramp-up and is not understood. However, multi-step ramps for the pinch coil avoided the problem.

9.5.4 High-current testing

With the thresholds set we were ready to move to high currents. The mirror (pinch and bucking) coils were ramped to their nominal currents without the need for training. This was expected because of their large operational margins (Table 9.2) and their large copper to superconductor ratio, the significance of which is discussed in Sec. 9.1.3. The first attempt at ramping the quadrupole and octupole magnets to full current failed, however. The quadrupole and octupole magnets quenched prematurely at 600 and 350 A, 89% and

70% of their nominal values, respectively. It was surprising that the octupole did not reach full current, since it was the only coil tested and quench-trained by AML.

Fig. 9.14 shows the data from the first octupole quench. In Fig. 9.14a, the voltage tap voltages and weighted sum indicate a quench developing just before the quench trigger goes off. The voltage imbalance is seen between traces "magnet1" and "magnet2," which represent the voltage drops across each half-coil section. These voltages are beginning to rise past zero, and the fact that the voltages are not equal and opposite (as they are in Fig. 9.13) indicate that they are not inductive but instead resistive voltages coming from a section of the coil going normal. Had they been inductive because of a current fluctuation, the weighted sum also displayed would have remained zero and the trigger (TTL in the figure) would not have gone off.

The first time the threshold is exceeded is at ∼2 ms. The hold-off requirement for the quench protection system to trigger is that this be sustained for 3.5 – 4.5 ms. This condition is met starting at ∼3 ms, and at 7.5 ms the dump is triggered, 4.5 ms later. This indicates that the threshold and hold-off setting worked correctly. The dump begins 0.1 ms after the beginning of the TTL pulse.

The weighted sum in Fig. 9.14b was analyzed to check if the weights described in Sec. 9.5.3 were set correctly via the trimpots. The expected weighted sum based on input weights k_i and voltages for this quench event was calculated from Eq. 9.5 and compared against the measured data and found to match well. This indicates that the voltage drops across sections were in fact being weighted correctly.

Fig. 9.15 shows the power dissipated in the quench resistors. First, the voltage drops across all sections of the coil were summed to find the total voltage across the entire magnet and lead assembly. This is the same voltage drop across the dump resistor and the high current leads. The power is $P(t) = V(t)^2/R$. Integrating the power curve gives the

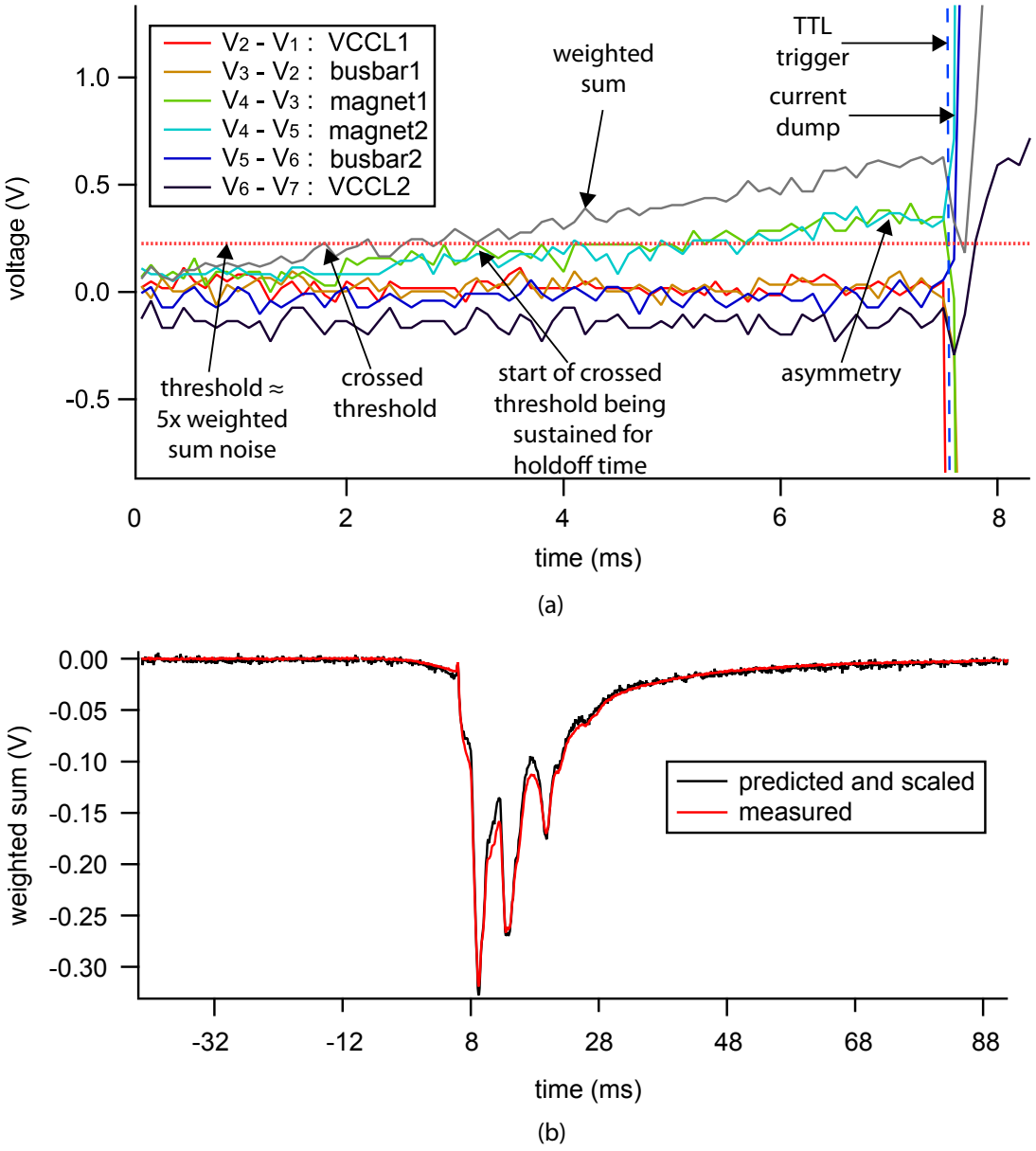


Figure 9.14: First octupole quench. (a) Voltage imbalance during a quench leads to the quench protection system being triggered. (b) Calculated and measured weighted sum trace for this quench. Time axis is relative to the scale in (a).

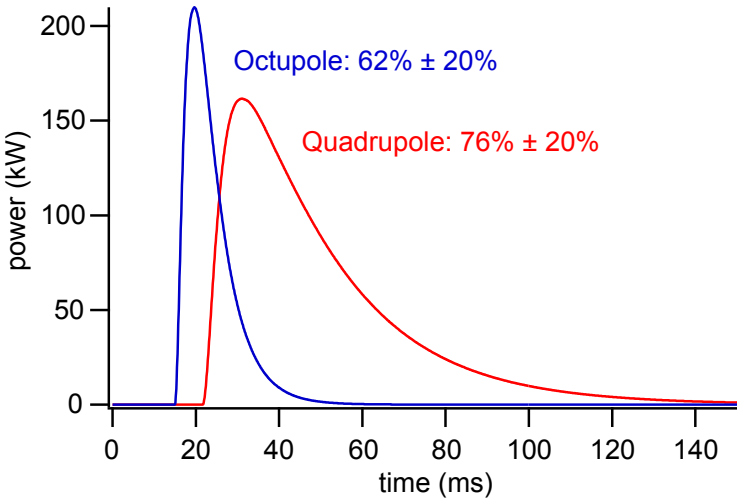


Figure 9.15: Power dissipated in dump resistors during a quench event. The total power dissipated as a percentage of stored energy is indicated.

energy dissipated in the dump resistor. It was found that $76\% \pm 20\%$ of the quadrupole's stored energy and $62\% \pm 20\%$ of the octupole's stored energy were dissipated in their dump resistors. The energy dissipated from induced e.m.f.s on other coils was negligible. A similar energy dissipation analysis was done for the same magnets at lower current, 10 Amps, and the power dissipated was found to be proportional to the current.

9.5.5 Quench training

After the first octupole and quadrupole quenches confirmed that the quench protection settings were working properly, that the IGBTs were switching at the right time, and that most of the power was dissipated in the dump resistors, we moved on to quench training the octupole and quadrupole magnets.

Two methods of ramping were investigated for the quench training: the one-step voltage-limited ramp (already described for the first 10 Amp data), and programming the voltage and current using an analog input to the current limit on the power supply. The one-step voltage ramp, while fine for initial testing which required immediate dumps on

	Maximum Current Reached (A)	**Shortest Ramp-up Times (s)**
Quadrupole	510	20
Octupole	607	175
Pinch	390	75
Bucking	350	5

Table 9.4: Quench training results achieved for ramping coils one at a time. The shortest ramp-up times listed were obtained from using a voltage input to the power supply. The maximum octupole current ever achieved was with a full trap, at 616 Amps.

command, had the limitation that if the current were sustained, after a few minutes it began to drift away from the intended current. The analog control was meant to be an alternative to this, and worked but was only tested with a periodic input. A sustained input required hardware we did not have and therefore this was not tested. Sustained and stable currents are needed for up to ∼10 minutes for antihydrogen trapping experiments, and the current in each coil was found to be sustainable for over 10 minutes. Table 9.4 shows the highest currents achieved after the quench training, and the shortest ramp-up times we achieved (not necessarily a limit). No observable recovery times were needed after a quench.

After the coils were quench-trained individually, they were then tested in full-field configurations, which required quench-training from scratch again. Fig. 9.16 shows the training plots for full-field ramp-ups. It was observed that for both the octupole and quadrupole full-trap configurations, when the radial coil quenched, the two mirror coils also got dumped within ms. It is possible that because the coils don't all have the same center, there could be an umbalanced e.m.f. from mutual inductances between coils causing the quenching. The magnets may have to again be re-trained when they are used in the 1 T

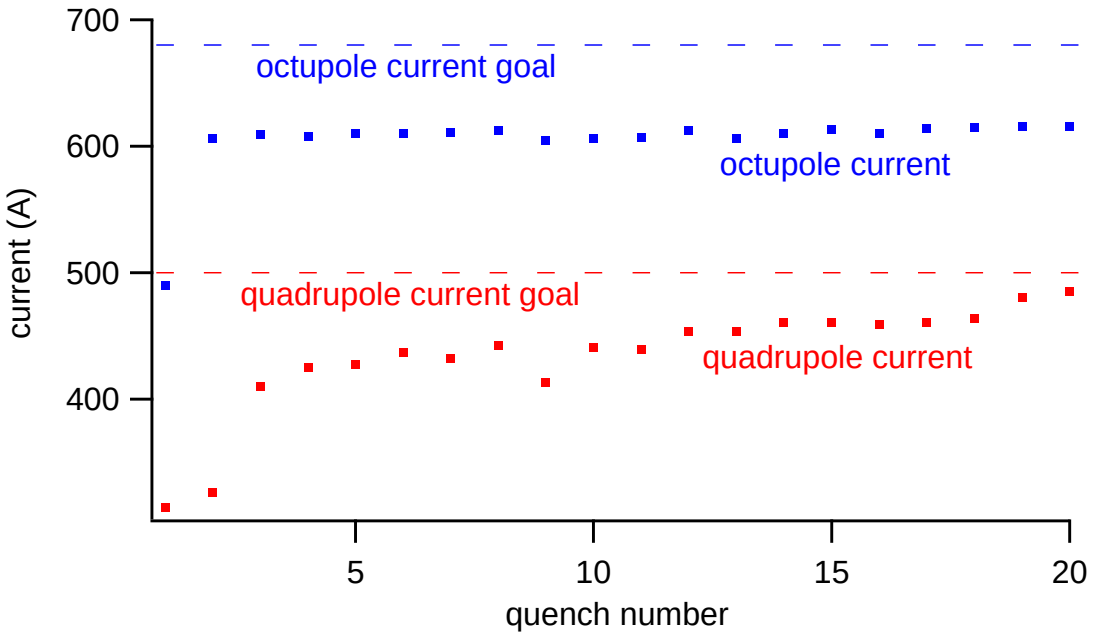


Figure 9.16: Quench training the quadrupole and octupole magnets in full-field configurations. Quadrupole full field was attained while octupole was not.

Penning trap bias field.

9.6 Summary

We had many successes in the testing of the second-generation magnet. The trap turnoff times achieved were tens of milliseconds, which will cut down our cosmic background rate by an order of magnitude during $\overline{\mathrm{H}}$-trapping experiments. Changing the values of the dump resistors could lead to even faster turnoff times, allowing for more accurate determination of $\overline{\mathrm{H}}$ events and for single atom detection in both the quadrupole and octupole configurations. The 1-2 order of magnitude faster turn-on times of the CTRAP magnet will save valuable experiment time (previously 10-15 minutes) and lead to proportionally fewer particle losses during the Ioffe trap ramp-up. The quench-protection system was found to be robust, crucial because even a single failure at high current would be catastrophic for

the magnet. Being able to turn off the magnet normally without violent quenching is a significant advantage which will make experiments easier and also save helium. No recovery time needed after a quench or a turn-off means more experiments that can be done in a day. Finally, all the magnets made it to full-field except for the octupole, but this could change with more magnet use. Even though there were many hurdles getting there, in the end the magnet was tested successfully. Its new titanium enclosure is done and the magnet is appended to CTRAP, ready to use with antiprotons this year.

Chapter 10

Conclusions

The principal motivation [7] for this thesis was the confinement of large numbers of $\overline{\mathrm{H}}$ atoms for eventual precision measurements. To this end progress was made on several fronts: trapping of the largest numbers of $\overline{\mathrm{H}}$ per trial yet, improvement of $\overline{\mathrm{H}}$ detection sensitivity, designing and building parts of the second-generation antihydrogen apparatus, and the successful electrical testing of the main feature of the new apparatus: the magnetic Ioffe trap for antihydrogen.

Better control of plasma geometry and temperature as well as improved methods to capture larger numbers of colder charged particles set the stage for $\overline{\mathrm{H}}$ trapping. There were also new developments crucial to determining $\overline{\mathrm{H}}$-trapping protocols. These include optimization of parameter space in $\overline{\mathrm{H}}$ detection studies and limiting charged particle losses in the quadrupole Ioffe field. Effective plasma preparation and energy measurements of driven particles allowed us to better control the interaction between the constituent plasmas. Various kinds of drives including noise drives, chirped drives, and ramps were used on several timescales. There were many more ideas whose implementation time would not permit, and variations on driving techniques and trapping schemes are still left to explore.

The improved methods were used in ATRAP's first demonstration of trapped antihydrogen in 2011 [10]. The trapping was done in a magnetic quadrupole field, despite a claim [65] that this could not be done. In our published demonstration, data from 20 trials using several driving methods were pooled together, and an analysis of $\overline{\mathrm{H}}$ annihilation signals revealed on average 5 atoms per trial–the most ever simultaneously trapped. Later, an alternative analysis using machine learning techniques yielded better $\overline{\mathrm{p}}$ sensitivity against cosmic background, enabling us to detect fewer $\overline{\mathrm{H}}$ atoms with high confidence. With BTRAP we will now be able to observe with 3σ confidence a minimum of 9 atoms per trial, reduced from 12 atoms, while with CTRAP this method already promises single-atom detection resolution. This will allow us to optimize parameters for $\overline{\mathrm{H}}$-trapping on a trial-to-trial basis and will be of benefit for future experiments that use a single trapped $\overline{\mathrm{H}}$ for precise spectroscopy.

After successfully trapping antihydrogen, our efforts were directed towards implementing the second-generation trap. The apparatus construction and wiring was completed in 2012, and the first cooldown attempts were done the same year. Although the cooldowns were not successful due to failed vacuum joints in the Ioffe trap's G10 enclosure, valuable lessons were learned from this experience.

The G10 enclosure was removed and a dedicated cryostat was constructed for testing the bare magnet as well the quench protection system. The quench protection system was calibrated for the magnet's inductive imbalance and found to work properly in preventing quenches. The magnet's turnoff times were measured to be tens of milliseconds, which will reduce the number of background counts in future antihydrogen studies by at least an order of magnitude. A decrease in the value of the dump resistor could further decrease the turnoff time for essentially background-free $\overline{\mathrm{H}}$ detection. Other advantages to the new Ioffe trap include a fast turn-off mechanism that does not require purposeful

quenching, and that no recovery time is needed between usages. This will enable many more trapping experiments to be done. Since the testing, a new helium enclosure made of titanium has been welded onto the Ioffe magnet, and the magnet has been mounted onto CTRAP. A successful cooldown and use in the current beam run is anticipated.

A wealth of improvements have been made in apparatus, methods and analysis since the first $\overline{\mathrm{H}}$ trapping attempts, and there are still gains to be had. The next-generation antihydrogen trap with its greater magnetic trap depths, flexibility in field geometries, and faster turn-off times promises to yield larger numbers of and greater sensitivity to trapped antihydrogen. Additionally, laser systems for cooling and spectroscopy have now been built. These advances give hope that the first precision measurement on antiatoms will happen in the very near future.

Bibliography

[1] D. Colladay and V. A. Kostelecký, Phys. Rev. D **55**, 6760 (1997).

[2] R. Carosi, P. Clarke, D. Coward, D. Cundy, N. Doble, L. Gatignon, V. Gibson, P. Grafström, R. Hagelberg, G. Kessler, and et al., Phys. Lett. B **237**, 303 (1990).

[3] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. **59**, 26 (1987).

[4] G. Gabrielse, A. Khabbaz, D. S. Hall, C. Heimann, H. Kalinowsky, and W. Jhe, Phys. Rev. Lett. **82**, 3198 (1999).

[5] Christian G. Parthey, Arthur Matveev, Janis Alnis, Birgitta Bernhardt, Axel Beyer, Ronald Holzwarth, Aliaksei Maistrou, Randolf Pohl, Katharina Predehl, Thomas Udem, Tobias Wilken, Nikolai Kolachevsky, Michel Abgrall, Daniele Rovera, Christophe Salomon, Philippe Laurent, and Theodor W. Hänsch, Phys. Rev. Lett. **107**, 203001 (2011).

[6] C. L. Cesar, D. G. Fried, T. C. Killian, A. D. Polcyn, J. C. Sandberg, I. A. Yu, T. J. Greytak, D. Kleppner, and J. M. Doyle, Phys. Rev. Lett. **77**, 255 (1996).

[7] G. Gabrielse, in *Fundamental Symmetries*, edited by P.Bloch, P. Paulopoulos, and R. Klapisch (Plenum, New York, 1987), pp. 59–75.

[8] G. Gabrielse, X. Fei, K. Helmerson, S. L. Rolston, R. L. Tjoelker, T. A. Trainor, H. Kalinowsky, J. Haas, and W. Kells, Phys. Rev. Lett. **57**, 2504 (1986).

[9] G. B. Andresen, M. D. Ashkezari, M. Baquero-Ruiz, W. Bertsche, P. D. Bowe, E. Butler, C. L. Cesar, S. Chapman, M. Charlton, A. Deller, S. Eriksson, J. Fajans, T. Friesen, M. C. Fujiwara, D. R. Gill, A. Gutierrez, J. S. Hangst, W. N. Hardy, M. E. Hayden, A. J. Humphries, R. Hydomako, M. J. Jenkins, S. Jonsell, L. V. Jorgensen, L. Kurchaninov, N. Madsen, S. Menary, P. Nolan, K. Olchanski, A. Olin, A. Povilus, P. Pusa, F. Robicheaux, E. Sarid, Nasr, D. M. Silveira, C. So, J. W. Storey, R. I. Thompson, D. P. van der Werf, J. S. Wurtele, and Y. Yamazaki, Nature **468**, 673 (2010).

[10] G. Gabrielse, R. Kalra, W. S. Kolthammer, R. McConnell, P. Richerme, D. Grzonka, W. Oelert, T. Sefzick, M. Zielinski, D. W. Fitzakerley, M. C. George, E. A. Hessels, C. H. Storry, M. Weel, A. Müllers, and J. Walz, Phys. Rev. Lett. **108**, 113002 (2012).

[11] G. Gabrielse, S. L. Rolston, L. Haarsma, and W. Kells, Phys. Lett. A **129**, 38 (1988).

[12] G. Gabrielse, W. S. Kolthammer, R. McConnell, P. Richerme, J. Wrubel, R. Kalra, E. Novitski, D. Grzonka, W. Oelert, T. Sefzick, M. Zielinski, J. S. Borbely, D. Fitzakerley, M. C. George, E. A. Hessels, C. H. Storry, M. Weel, A. Müllers, J. Walz, and A. Speck, Phys. Rev. Lett. **105**, 213002 (2010).

[13] G. Gabrielse, W. S. Kolthammer, R. McConnell, P. Richerme, R. Kalra, E. Novitski, D. Grzonka, W. Oelert, T. Sefzick, M. Zielinski, D. Fitzakerley, M. C. George, E. A. Hessels, C. H. Storry, M. Weel, A. Müllers, and J. Walz, Phys. Rev. Lett. **106**, 073002 (2011).

[14] P. Richerme, Ph.D. thesis, Harvard University, 2012.

[15] R. P. McConnell, Ph.D. thesis, Harvard University, 2011.

[16] W.S. Kolthammer, Ph.D. thesis, Harvard University, 2011.

[17] D. Möhl, Hyperfine Int. **109**, 33 (1997).

[18] S van der Meer, Technical Report No. CERN-ISR-PO-72-31. ISR-PO-72-31, CERN, Geneva (unpublished).

[19] G.I. Budker, Soviet Atomic Energy **22**, 438 (1967).

[20] C. M. Surko, M. Leventhal, and A. Passner, Phys. Rev. Lett. **62**, 901 (1989).

[21] D Comeau, A Dror, D W Fitzakerley, M C George, E A Hessels, C H Storry, M Weel, D Grzonka, W Oelert, G Gabrielse, R Kalra, W S Kolthammer, R McConnell, P Richerme, A Müllers, J Walz, and ATRAP Collaboration, New Journal of Physics **14**, 045006 (2012).

[22] F. M. Penning, Physica (Utrecht) **3**, 873 (1936).

[23] L. S. Brown and G. Gabrielse, Rev. Mod. Phys. **58**, 233 (1986).

[24] G. Gabrielse and F. Colin MacKintosh, Intl. J. Mass Spec. Ion Proc. **57**, 1 (1984).

[25] G. Gabrielse, L. Haarsma, and K. Abdullah, Hyperfine Interact. **89**, 371 (1994).

[26] D. E. Pritchard, Phys. Rev. Lett. **51**, 1336 (1983).

[27] G. Gabrielse, P. Larochelle, D. Le Sage, B. Levitt, W. S. Kolthammer, R. McConnell, P. Richerme, J. Wrubel, A. Speck, M. C. George, D. Grzonka, W. Oelert, T. Sefzick, Z. Zhang, A. Carew, D. Comeau, E. A. Hessels, C. H. Storry, M. Weel, J. Walz, and ATRAP Collaboration, Phys Rev Lett **100**, 113001 (2008).

[28] J. Fajans and A. Schmidt, Nuc. Inst. Meth. A **521**, 318 (2004).

[29] T. Bergman, G. Erez, and H. Metcalf, Phys. Rev. A **35**, 1535 (1987).

[30] G. Gabrielse, X. Fei, L. A. Orozco, R. L. Tjoelker, J. Haas, H. Kalinowsky, T. A. Trainor, and W. Kells, Phys. Rev. Lett. **65**, 1317 (1990).

[31] G. Bendiscioli and D. Kharzeev, La Rivista Del Nuovo Cimento Series 3 **17**, 1 (1994).

[32] W.M. Yao et al., Journal of Physics G: Nuclear and Particle Physics **33**, 1 (2006).

[33] Z. Zhang, Ph.D. thesis, Ruhr-Universität Bochum, 2007.

[34] Martin A. Fischler and Robert C. Bolles, Commun. ACM **24**, 381 (1981).

[35] J. Wrubel, G. Gabrielse, W.S. Kolthammer, P. Larochelle, R. McConnell, P. Richerme, D. Grzonka, W. Oelert, T. Sefzick, M. Zielinski, J.S. Borbely, M.C. George, E.A. Hessels, C.H. Storry, M. Weel, A. Müllers, J. Walz, and A. Speck, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **640**, 232 (2011).

[36] D. Lesage, Ph.D. thesis, Harvard University, 2008.

[37] In-Kook Suh, H. Ohta, and Y. Waseda, Journal of Materials Science **23**, 757 (1988).

[38] D. H. E. Dubin and T. M. O'Neil, Rev. Mod. Phys. **71**, 87 (1999).

[39] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, Oxford, 1960).

[40] T. M. O'Neil, Phys. Fluids **23**, 2216 (1980).

[41] J. J. Bollinger, D. J. Heinzen, F. L. Moore, W. M. Itano, D. J. Wineland, and D. H. E. Dubin, Phys. Rev. A **48**, 525 (1993).

[42] D. H. E. Dubin, Phys. Rev. Lett. **66**, 2076 (1991).

[43] M. Glinsky, T. O'Neil, M. N. Rosenbluth, K. Tsuruta, and S. Ichimaru, Phys. Fluids B **4**, 1156 (1992).

[44] S. L. Rolston and G. Gabrielse, Hyperfine Interact. **44**, 233 (1988).

[45] G. Gabrielse, X. Fei, L. A. Orozco, R. L. Tjoelker, J. Haas, H. Kalinowsky, T. A. Trainor, and W. Kells, Phys. Rev. Lett. **63**, 1360 (1989).

[46] T. M. O'Neil, Physics of Fluids (1958-1988) **24**, (1981).

[47] B. Levitt, G. Gabrielse, P. Larochelle, D. Le Sage, W.S. Kolthammer, R. McConnell, J. Wrubel, A. Speck, D. Grzonka, W. Oelert, T. Sefzick, Z. Zhang, D. Comeau, M.C. George, E.A. Hessels, C.H. Storry, M. Weel, and J. Walz, Physics Letters B **656**, 25 (2007).

[48] X. P. Huang, F. Anderegg, E. M. Hollmann, C. F. Driscoll, and T. M. O'Neil, Phys. Rev. Lett. **78**, 875 (1997).

[49] G. Gabrielse, X. Fei, L. A. Orozco, S. L. Rolston, R. L. Tjoelker, T. A. Trainor, J. Haas, H. Kalinowsky, and W. Kells, Phys. Rev. A **40**, 481 (1989).

[50] G. Gabrielse, N. S. Bowden, P. Oxley, A. Speck, C. H. Storry, J. N. Tan, M. Wessels, D. Grzonka, W. Oelert, G. Schepers, T. Sefzick, J. Walz, H. Pittner, and E. A. Hessels, Phys. Lett. B **548**, 140 (2002).

[51] T. M. O'Neil and C. F. Driscoll, Physics of Fluids (1958-1988) **22**, (1979).

[52] M. Glinsky and T. O'Neil, Phys. Fluids **B3**, 1279 (1991).

[53] G. Gabrielse, S.L. Rolston, L. Haarsma, and W. Kells, Physics Letters A **129**, 38 (1988).

[54] G. Gabrielse, N. S. Bowden, P. Oxley, A. Speck, C. H. Storry, J. N. Tan, M. Wessels, D. Grzonka, W. Oelert, G. Schepers, T. Sefzick, J. Walz, H. Pittner, T. W. Hänsch, and E. A. Hessels, Phys. Rev. Lett. **89**, 213401 (2002).

[55] G. Gabrielse, N. S. Bowden, P. Oxley, A. Speck, C. H. Storry, J. N. Tan, M. Wessels, D. Grzonka, W. Oelert, G. Schepers, T. Sefzick, J. Walz, H. Pittner, T. W. Hänsch, and E. A. Hessels, Phys. Rev. Lett. **89**, 233401 (2002).

[56] G. B. Andresen, M. D. Ashkezari, M. Baquero-Ruiz, W. Bertsche, P. D. Bowe, E. Butler, P. T. Carpenter, C. L. Cesar, S. Chapman, M. Charlton, J. Fajans, T. Friesen, M. C. Fujiwara, D. R. Gill, J. S. Hangst, W. N. Hardy, M. E. Hayden, A. J. Humphries, J. L. Hurt, R. Hydomako, S. Jonsell, N. Madsen, S. Menary, P. Nolan, K. Olchanski, A. Olin, A. Povilus, P. Pusa, F. Robicheaux, E. Sarid, D. M. Silveira, C. So, J. W. Storey, R. I. Thompson, D. P. van der Werf, J. S. Wurtele, and Y. Yamazaki, Phys. Rev. Lett. **106**, 025002 (2011).

[57] G. Gabrielse, S. L. Rolston, L. Haarsma, and W. Kells, Hyperfine Interact. **44**, 287 (1988).

[58] D. R. Bates and A. Dalgarno, in *Atomic and Molecular Processes*, edited by D. R. Bates (Academic Press, New York, 1962), pp. 245–271.

[59] E. A. Hessels, D. M. Homan, and M. J. Cavagnero, Phys. Rev. A **57**, 1668 (1998).

[60] G. Gabrielse, J. Estrada, J. N. Tan, P. Yesley, N. S. Bowden, P. Oxley, T. Roach, C. H. Storry, M. Wessels, J. Tan, D. Grzonka, W. Oelert, G. Schepers, T. Sefzick, W. Breunlich, M. Cargnelli, H. Fuhrmann, R. King, R. Ursin, J. Zmeskal, H. Kalinowsky, C. Wesdorp, J. Walz, K. S. E. Eikema, and T. Haensch, Phys. Lett. B **507**, 1 (2001).

[61] G. Gabrielse, A. Speck, C. H. Storry, D. Le Sage, N. Guise, D. Grzonka, W. Oelert, G. Schepers, T. Sefzick, H. Pittner, J. Walz, T. W. Hänsch, D. Comeau, and E. A. Hessels, Phys. Rev. Lett. **93**, 073401 (2004).

[62] T. Pohl, H. R. Sadeghpour, and G. Gabrielse, Physical Review Letters **97**, 143401 (2006).

[63] I. Barth, L. Friedland, E. Sarid, and A. G. Shagalov, Phys. Rev. Lett. **103**, 155001 (2009).

[64] C. Zimmerman and T. Hänsch, Hyperfine Interact. **76**, 47 (1993).

[65] Joël Fajans, W Bertsche, K Burke, Steven F Chapman, and Dirk Peter Van der Werf, Physical review letters **95**, 155001 (2005).

[66] G. B. Andresen et al. (ALPHA Collaboration), Nature Physics **7**, 558 (2011).

[67] P. Larochelle, Ph.D. thesis, Harvard University, 2009.

[68] P. Richerme, G. Gabrielse, S. Ettenauer, R. Kalra, E. Tardiff, D. W. Fitzakerley, M. C. George, E. A. Hessels, C. H. Storry, M. Weel, A. Müllers, and J. Walz, Phys. Rev. A **87**, 023422 (2013).

[69] G. Gabrielse, Adv. At. Mol. Opt. Phys. **50**, 155 (2005).

[70] T. Pohl, H. R. Sadeghpour, Y. Nagata, and Y. Yamazaki, Phys. Rev. Lett. **97**, 213001 (2006).

[71] F. Robicheaux, Phys. Rev. A **73**, 033401 (2006).

[72] M. Wolter, Physics of Particles and Nuclei **38**, 255 (2007).

[73] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification* (Wiley, USA, 2012).

[74] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006).

[75] M. Loève, *Probability Theory: Foundations, Random Sequences* (Van Nostrand, Toronto, 1955).

[76] M.N. Wilson, *Superconducting Magnets*, *Monographs on Cryogenics* (Clarendon Press, Oxford, 1987).

[77] B. Seeber, *Handbook of Applied Superconductivity* (Taylor & Francis, Institute of Physics Publishing, London, 1998), No. v. 2.

[78] C. Day, CAS-CERN Accelerator School and ALBA Synchrotron Light Facility: Course on Vacuum in Accelerators, Platja d'Aro, Spain p. 241 (2006).