The Pennsylvania State University
The Graduate School
Eberly College of Science

DEVELOPMENT OF SCALABLE APPROACHES TO NEUTRINO MASS MEASUREMENT WITH THE PROJECT 8 EXPERIMENT

A Thesis in
The Physics Department
by
Andrew Douglas Ziegler

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2023

16 The thesis of Andrew Douglas Ziegler was reviewed and approved* by the following:

Luiz de Viveiros Souza
Assistant Professor of Physics
Thesis Advisor
Chair of Committee

Carmen Carmona
Assistant Professor of Physics

17 Doug Cowan
Professor of Physics and Astrophysics

Xingjie Ni
Associate Professor of Electrical Engineering

Stephanie Wissel
18 Associate Professor of Physics

19 *Signatures are on file in the Graduate School.

Abstract

Some shit goes here.

Table of Contents

Acknowledgments

Shout out to all the haters.

Dedication

Something heartfelt.

Chapter 1
Introduction

1.1 Summary

Neutrinos are one of the fundamental particles that comprise the standard model of particle physics and account for a significant fraction of the matter in the universe. Neutrinos are the most abundant fermions in the universe, but due to their weak interactions neutrinos seldom interact with other particles. Regardless, neutrinos play a unique role in the evolution of the early-universe, therefore, a detailed understanding of the properties of the neutrino is important to understanding the cosmology of the universe as well as understanding the universe at the fundamental particle physics scale.

Unlike other fermions it was unclear that neutrinos had nonzero mass until neutrino flavor oscillations were definitively observed in the late 90's and early 00's. Flavor oscillations require that neutrinos experience time so that when acted upon by the time-evolution operator the initial neutrino state can evolve to a new flavor state. This implies that the neutrino flavor states are really a superposition of at least three separate neutrino states with well-defined masses.Measurements of neutrino oscillations that have taken place over the past couple of decades have measured the differences between neutrino mass eigenstates with increasing precision. However, oscillation measurements cannot tell us the mass scale of the neutrinos, which is required in order to measure the absolute neutrino masses.

The neutrino mass scale remains an unknown quantity in the standard model of particle physics. The value of the neutrino mass influences the evolution of the early universe and is likely relevant to the energy-scale of new physics responsible for the factor of 10^{-6} difference between the neutrino and electron masses. A model-independent way to measure the neutrino mass is to measure the tritium beta-decay spectrum near its endpoint. Energy conservation requires that the neutrino mass carry away some kinetic energy from the beta-decay electron in the form of its mass, which causes a distortion in the shape of the tritium beta-decay spectrum near the endpoint. The isotope tritium has many advantages for this measurement, and has been used by the KATRIN collaboration to perform the most sensitive direct neutrino mass measurement to date.

KATRIN represents the state-of-the-art experiment in the current generation of neutrino mass direct measurement experiments and has a final projected sensitivity to neutrino masses $m_\nu < 200$ meV. This sensitivity does not fully exhaust the allowed parameter space of neutrino masses under the normal and inverted neutrino mass ordering scenarios, which motivates the development of a next generation of neutrino mass measurement experiments.

The Project 8 collaboration is developing a next-generation neutrino mass direct measurement experiment designed to be sensitive to $m_\nu < 40$ meV. This sensitivity is sufficient to exhaust the range of neutrino masses allowed under the inverted mass ordering regime. Project 8 intends to achieve its sensitivity goal utilizing two technologies that are novel to the space of direct neutrino mass measurement — atomic tritium and cyclotron radiation emission spectroscopy (CRES). Atomic tritium is required in order to avoid systematic broadening the tritium beta-decay spectrum caused by the final state of the ^{3}He$^+$-T molecule, and the CRES technique enables a differential measurement of the tritium spectrum that is background-free and able to be directly integrated with the atomic tritium source.

The Project 8 collaboration is currently engaged in a research and development program intended to simultaneously develop the atomic tritium and CRES technologies so that they can be combined in a next-generation experiment. This past year (2022) Project 8 has used the CRES technique to measure the molecular tritium beta-decay spectrum and place an upper limit on the neutrino mass: $m_\beta \leq 152$ eV. This measurement, while not competitive scientifically, represents the first proof-of-principle that the CRES technique can be used to measure the neutrino mass.

The future goals of the Project 8 collaboration are to develop the technologies and techniques necessary to scale-up the volumes in which CRES measurements can be performed. Project 8's first neutrino mass measurement with CRES utilized a measurement volume on the cubic-centimeter scale, however, sensitivity calculations estimate that an experiment sensitive to neutrino masses of 40 meV will require several tens of cubic-meters of experiment volume filled with atomic tritium. Developing a new approach to performing CRES measurements that can be successfully scaled to these volumes is a necessary step towards Project 8's neutrino mass measurement goal, and is the primary topic of my dissertation research.

A parallel development is the technology necessary to produce, cool, trap, and recirculate a supply of atomic tritium that is compatible with CRES measurements. The atomic tritium system is equally important as the large-volume CRES measurement technology, but it will not be the focus of this dissertation since I did not contribute significantly to this effort.

The Project 8 collaboration has identified two scalable approaches to neutrino mass measurement using the CRES technique. One approach is to use an array of antennas that surrounds a volume of trapped atomic tritium that can perform CRES measurements by collection the cyclotron radiation emitted by beta-decay electrons into free-space. The other approach uses a resonant cavity filled with atomic tritium to perform CRES by measuring the excitation of resonant cavity modes caused by the motion of electrons trapped inside the cavity volume.

The cavity and antenna approaches to CRES have been studied in detail over the past five years, and, while both approaches offer a physically viable path towards a 40 meV neutrino mass measurement the collaboration has elected to pursue the cavity approach for the foreseeable future. The major advantage of the cavity approach is a significant reduction in the cost and complexity of the experiment design and data analysis, which provides a less risky path towards Project 8's scientific goals.

In this dissertation I summarize my most impactful contributions to the research and development of antenna array and cavity CRES. In short these contributions are

- the development and analysis of signal reconstruction algorithms for antenna array CRES, which provided key inputs to sensitivity analyses of antenna array CRES experiments,
- the development of a specialized antenna designed to synthesize fake CRES radiation, which enabled bench-top testing and validation of the antenna array CRES technique,
- the development of an open-cavity design for CRES measurement whose mode structure can be tuned using perturbations that modify the impedance of the cavity walls. The development of this cavity concept was one of many developments that eventually lead to the adoption of cavities as the CRES technology of choice for the future of Project 8.

1.2 Outline

The outline of this dissertation is as follows. In Chapter 2 I provide an introduction to the basic physics of neutrinos and beta-decay, which provides context for a discussion of various methods to measure the neutrino absolute mass scale.

Chapter 3 is an overview of the CRES technique and the Project 8 collaboration. I highlight the Project 8 Phase II experiment, which was the first measurement of the tritium beta-decay spectrum with CRES, and I discuss the planned research and development for an antenna array CRES experiment in Phase III of the Project 8 collaboration's experiment plan. I end Chapter 3 with a discussion of the pilot-scale and Phase IV experiments, that will combine a scalable CRES measurement technology with atomic tritium and measure the neutrino mass with 40 meV sensitivity.

Chapter 4 discusses the first of the contributions mentioned above, which is the development of signal reconstruction techniques for antenna array CRES and an antenna array demonstrator experiment called the FSCD. I discuss the important tools that Project 8 uses to simulate antenna array CRES before introducing three signal reconstruction algorithms that can be used to detect CRES signals using the array. I end Chapter 4 with a paper that summarizes a detailed analysis and comparison of the signal detection performance of each algorithm.

Chapter 5 describes my contributions to the development of antennas and an antenna measurement system for Project 8, which is the second major contribution of this dissertation. I begin with a general overview of basic principle of antennas and antenna measurements, before including a paper that describes the development of unique antenna designed to mimic the cyclotron radiation emitted by electrons in free-space when trapped in a magnetic field. I call this antenna the synthetic cyclotron radiation antenna (SYNCA) and its main purpose is to serve a fake electron for laboratory validation measurements of Project 8's antenna array CRES simulations. Chapter 5 ends with an overview of laboratory measurements of a prototype antenna array that were compared with simulations to provide upper bounds on reconstruction errors caused by imperfections in real-life measurements.

Chapter 6 discusses the cavity approach to CRES, which was adopted as the preferred CRES technology for Phase IV late into my dissertation work. The chapter stars by discussing resonant cavities in general before introducing the operating principles of the cavity approach to CRES. I end the chapter by discussing a study of and open-cavity design that could be used for CRES measurements and integrated with atomic tritium and an electron gun calibration source for the pilot-scale and Phase IV experiments.

Finally, in Chapter 7 I conclude by briefly discussing the future directions of the Project 8 collaboration as we continue towards a direct measurement of the neutrino mass.

Chapter 2
Neutrinos and Neutrino Masses

2.1 Introduction

In this chapter I provide a cursory overview of background information relevant to neutrinos and neutrino mass measurements.

In Section 2.2 I provide some background information on the history of neutrinos and beta-decay. In Section 2.3 I describe the discover of neutrino oscillations, which proved unambiguously that neutrinos have non-zero masses. In Section 2.4 I discuss the current state of the theoretical understanding of neutrino masses in the standard model. Lastly, in Section 2.5 I discuss methods for measuring the absolute scale of the neutrino mass.

2.2 Neutrinos and Beta-decay

Late in the 19th century the phenomena of radioactivity was first observed in experiments performed by Henri Becquerel with uranium, and further studied using thorium and radium by Marie and Pierre Currie [1, 2]. Early work in radioactivity classified different forms of radiation based on it's ability to penetrate different materials. Rutherford was the first to separate radioactive emissions into two types, alpha and beta radiation [3]. Alpha rays can be easily stopped by a piece of paper or thin foil of metal, whereas beta radiation could penetrate metals several millimeters thick. Later a third form of radiation was identified by Villard [4], which was still more penetrating, and was eventually termed gamma radiation by Rutherford.

When these forms of radioactivity were first discovered it was unclear what physically constituted an alpha, beta, or gamma particle. Experiments with radioactivity in magnetic fields was eventually able to identify the charge composition of different forms of radiation. In particular, experiments by Becquerel identified that beta radiation had an identical charge-to-mass ratio to the electron discovered by Thompson in his work on cathode rays [5]. This was strongly suggestive that beta particles were indeed electrons.

Further studies of beta radiation lead to the discovery that radioactivity resulted in the transmutation of elements [6] caused by the decay of a heavier nucleus to a lighter species. One feature of beta radiation, which we now properly call beta-decay, that was different from alpha-decays and gamma radiation is that the electrons produced by beta-decay have a continuous spectrum of kinetic energies, whereas, alpha and gamma particles are emitted with discrete energies. This feature of beta-decay was first observed by Chadwick in 1914 [7], and was extremely puzzling at the time since the continuous spectrum apparently violates energy conservation [8].

Famously, in 1930 Pauli proposed the existence of a new neutral particle, which he termed the "neutron", that was also produced during beta-decay in order to resolve the missing energy problem posed by the beta-decay spectrum [9]. Because this particle carried no charge, it was hypothesized at the time that it had simply not been observed in any experiments up to that time. This "neutron", which was initially estimated to have a mass no larger than that of an electron, was eventually renamed the "neutrino" by Fermi [10] after the discovery of the neutron by Chadwick in 1932 [11]. Later, in 1933, Fermi developed a quantum mechanical theory for beta-decay in which both an electron and neutrino are produced by the decay of a neutron to a proton inside the radioactive nucleus [12].

Little more than a speculation when first introduced, indirect evidence for the existence of neutrinos was obtained in 1938 by the simultaneous observation of the electron and recoiling nucleus in cloud chambers by Crane and Halpern [13]. However, it wasn't until the Cowan-Reines experiment [14] in 1956 that direct evidence for the existence of neutrinos was observed by detecting the inverse beta-decays caused by neutrinos from a nuclear reactor interacting with protons contained in water molecules. The difficulty in detecting neutrinos is caused by their weak interactions with other particles. Further, experiments revealed that different types of neutrinos existed based on the nature of the leptons produced in neutrino charged-current interactions [15], but the existence of a neutrino mass remained an open question that would take more than 40 year to resolve.

2.3 Neutrino Oscillations

The first hint of neutrino flavor transitions or neutrino oscillations was indicated by the solar neutrino problem, which referred to discrepancies between the predicted flux of ν_e from the standard solar model and measurements of the solar neutrino flux such as the famous experiment at the Homestake mine by Ray Davis Jr. and collaborators in the 1960's [16]. Essentially, fewer electron-type neutrinos than expected were being observed from the sun. Finally, in the early 2000's the SNO experiment was able to resolve the solar neutrino problem by identifying neutrino oscillations as the cause of the observed deficit [17]. Furthermore, measurements of the atomspheric flux of neutrinos by the Super-Kamiokande experiment and others revealed that fewer muon-type neutrinos survived passage through the earth than expected providing strong evidence for neutrino oscillations for both flavors [18].

The origin of neutrino oscillations is that the weak eigenstates are distinct from the mass eigenstates [19]. The neutrino mass eigenstates represent physical particles in the sense that they are solutions to the free-particle Hamiltonian, whereas, the neutrino weak eigenstates correspond to the neutrino states that interact via the weak charged-current interaction. The neutrino weak eigenstates are a linear superposition of the neutrino mass eigenstates

$$\nu_\ell = \sum_i U_{\ell i} \nu_i, \tag{2.1}$$

where $\ell = e, \mu, \tau$ and $i = 1, 2, 3$. The matrix elements $U_{\ell i}$ are the elements of the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix that describes the mixing between the neutrino flavor and mass states.

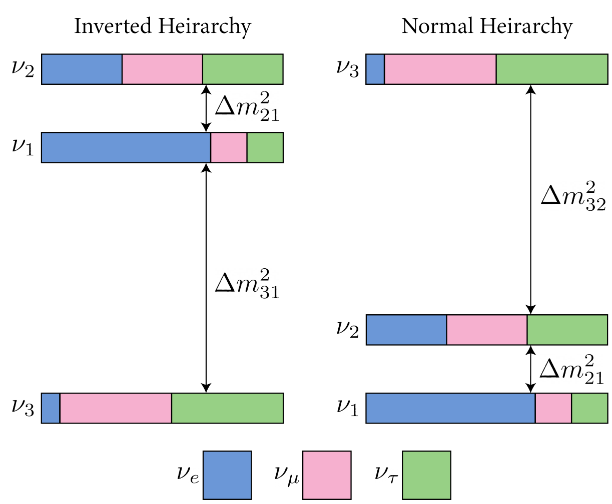


Figure 2.1: A diagram of two different neutrino mass ordering scenarios. In the inverted hierarchy (inverted mass ordering) the lightest neutrino mass is m_3, whereas,in the normal hierarchy (normal mass ordering) m_1 is the lightest neutrino. What cannot be measured by neutrino oscillations is the neutrino absolute mass scale, which is essentially the mass of the lightest neutrino mass eigenstate.

One standard parameterization of the PMNS matrix is

$$\begin{aligned} U_{PMNS} &= \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &\times \begin{bmatrix} e^{i\alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned} \tag{2.2}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The parameters α_1 and α_2 are only included in the PNMS matrix if neutrinos are Majorana particles, something which represents a current area of research in neutrino physics. The phase δ quantifies the degree of CP-violation in the neutrino sector. Including the Majorana phases the PMNS matrix contains six independent parameters. In addition, neutrino oscillation probabilities depend on the squared mass differences between neutrino mass eigenstates

$$\Delta m_{ij}^2 = m_i^2 - m_j^2, \tag{2.3}$$

where $ij = 12, 32, 31$ respectively. Because $\Delta m_{32}^2 = \Delta m_{31}^2 - \Delta m_{21}^2$, this adds an additional two parameters that must be constrained by neutrino oscillations.

A giant experimental effort over the past couple of decades has greatly contained the majority of parameters in the PMNS matrix, many to relative uncertainties of only a few percent. However, some parameters still remain relatively unconstrained, which is the origin of the current uncertainty in the ordering of the neutrino masses (see Figure 2.1). The neutrino masses can be organized by their relative mass. The current neutrino oscillation data can confirm that $m_2 > m_1$, however, the sign of Δm_{32}^2 is still unknown. This leads to two scenarios where neutrino masses follow the ordering $m_3 > m_2 > m_1$, which is called the normal mass ordering (NMO), or alternatively neutrino masses may be ordered $m_2 > m_1 > m_3$, which is called the inverted mass ordering (IMO). Next-generation neutrino oscillation experiments such as JUNO [20], Hyper-Kamiokande [21], and DUNE [22] are poised to resolve this ambiguity in the coming years.

Neutrino oscillation probabilities are only sensitive to the neutrino masses via the squared mass differences. Therefore, oscillation probabilities are unaffected by the absolute scale of the neutrino mass. However, oscillations can be used to obtain a lower bound on the neutrino masses by setting the mass of the lightest neutrino mass state to zero. This results in different lower limits depending on the ordering of the neutrino mass states. Current best-fit values [23] with 1σ-uncertainties for the squared mass differences are

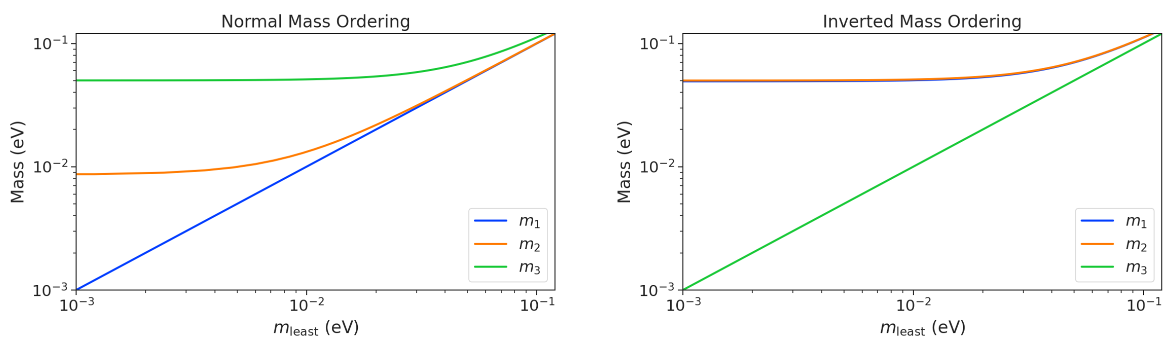
$$\Delta m^2_{21} = (7.42^{+0.21}_{-0.20}) \times 10^{-5} \text{ eV}^2, \tag{2.4}$$

$$\Delta m^2_{31} = (2.5176{+}0.026_{-0.028}) \times 10^{-3} \text{ eV}^2 \text{ (NMO)}, \tag{2.5}$$

for the normal mass ordering, and in the case of the inverted ordering we have

$$\Delta m^2_{32} = (-2.498^{+0.028}_{-0.028}) \times 10^{-3} \text{ eV}^2 \text{ (IMO)}. \tag{2.6}$$

By letting the lightest neutrino mass in each ordering scenario (m_{least}) take on a range of values one can visualize the relative masses of the neutrinos as a function of m_{least} (see Figure 2.2).



(a) (b)

Figure 2.2: The masses of the neutrinos as a function of the lightest neutrino mass in both the normal (a) and inverted (b) mass ordering regimes.

2.4 Neutrino Masses in the Standard Model

In this section, I briefly summarize the current theoretical understanding of neutrino masses in the standard model [24–26]. Neutrinos are spin 1/2 particles, which are described using the Dirac equation.

$$(i\hbar\gamma^\mu\partial_\mu - mc)\psi(x) = 0, \tag{2.7}$$

where the field that describes the particle is denoted as $\psi(x)$. In the standard model fermions acquire mass through the Yukawa interaction, which add to the standard model Lagrangian terms of the form

$$\mathcal{L}_{\text{Yukawa}} = -Y^{\ell}_{ij}\bar{L}_{Li}\phi E_{Rj} + \text{h.c.}, \tag{2.8}$$

where Y^{ℓ}_{ij} is an element of the 3×3 Yukawa coupling matrix for leptons, L_{Li} is the left-handed lepton doublet for generation i, ϕ is the Higgs doublet, and E_{Rj} is the right-handed lepton field for generation j. Neutrinos are represented only as left-handed neutrinos and right-handed antineutrinos in the standard model, which is consistent with experimental observations. Since there are no right-handed neutrino singlet fields, there are no Yukawa interaction terms, thus neutrinos in the standard model are strictly massless. Therefore, non-zero neutrino mass is evidence for physics beyond the standard model.

For the charged leptons, the Yukawa interaction leads to masses of the form

$$m^{\ell}_{ij} = Y^{\ell}_{ij}\frac{v}{\sqrt{2}}, \tag{2.9}$$

where v is the Higgs vacuum expectation value. The observation of massive neutrinos motivates the extension of the standard model to explain the origin of neutrino masses, which can be approached in different ways, but all approaches add additional degrees of freedom to the standard model. One approach is to introduce to the standard model a right-handed neutrino field that allows one to include Yukawa terms of the form

$$\mathcal{L}_{\nu\text{Yukawa}} = -Y^{\ell}_{ij}\bar{L}_{Li}\phi\nu_{Rj} + \text{h.c.} \tag{2.10}$$

where ν_{Rj} is the right-handed neutrino singlet. Because experimental evidence strongly predicts only three active neutrinos, these additional neutrinos are sterile and do not interact via the strong, weak, or electromagnetic interactions. After spontaneous symmetry breaking, the Yukawa interaction leads to mass terms given by

$$\mathcal{L}_D = -M_{Dij}\bar{\nu}_{Ri}\nu_{Lj} + \text{h.c.}, \tag{2.11}$$

which is called a Dirac mass term. One of the issues with constructing neutrino masses in this way is that the required Yukawa couplings are at least a factor of 10^6 smaller than that of an electron, which begs the question: why are the Yukawa couplings so small for the neutrinos?

An alternative approach is to allow the neutrinos to have a Majorana mass, which is possible because neutrinos are electrically neutral particles. The Majorana mass terms for the neutrino have the form

$$\mathcal{L}_M = -\frac{1}{2}(M_{Rij}\bar{\nu}_{Ri}\nu^c_{Rj}M_{Lij}\bar{\nu}_{Li}\nu^c_{Lj}) + \text{h.c.}, \tag{2.12}$$

where M_{Rij} and M_{Lij} are right-handed and left-handed Majorana mass matrices. A consequence of neutrinos being Majorana particles is lepton number violation, which predicts the occurrence of neutrino-less double beta-decay at a rate proportional to the neutrino mass.

In the most general case neutrinos have both Dirac and Majorana mass terms, which allows one to generate neutrino masses with Yukawa couplings similar to the rest of the standard model. Considering just one generation of neutrinos for illustration, the combined Lagrangian can be written as

$$\mathcal{L}_{D+M} = -m_D\bar{\nu}_R\nu_L - \frac{1}{2}(m_L\bar{\nu}_L\nu^c_L + m_R\bar{\nu}_R\nu^c_R) + \text{h.c.}, \tag{2.13}$$

or equivalently,

$$\mathcal{L}_{D+M} = -\frac{1}{2}\begin{bmatrix}\bar{\nu}_L & \bar{\nu}^c_R\end{bmatrix}\begin{bmatrix}m_L & m_D \\ m_D & m_R\end{bmatrix}\begin{bmatrix}\nu^c_L \\ \nu_R\end{bmatrix} + \text{h.c.}. \tag{2.14}$$

An example mass generation mechanism with this approach is the Type-I see-saw mechanism [27], in which we take $m_L = 0$ and $m_R \gg m_D$. By diagonalizing Equation 2.14 one obtains the mass eigenvalues that represent the physical masses of the neutrinos. The light neutrino mass eigenstate, which represents the observed neutrino mass, has a mass given by

$$m_1 \approx \frac{m_D^2}{m_R}, \tag{2.15}$$

and the heavy neutrino mass eigenstate, which represents the unobserved sterile neutrino, has a mass

$$m_2 \approx m_R. \tag{2.16}$$

For m_D similar to the other quark or lepton masses, one obtains physical neutrino masses consistent with observations from sterile neutrino masses of $m_R \approx O(10^{15})$ GeV. This mass scale is well beyond the capabilities of modern particle accelerators.

2.5 Neutrino Absolute Mass Scale

The neutrino absolute mass scale or simply "neutrino mass" cannot be probed with neutrino oscillations, since oscillation probabilities are determined by the squared mass differences between neutrino mass eigenstates, therefore, alternative techniques are needed to perform an effective measurement of the neutrino mass.

2.5.1 Limits from Cosmology

The ΛCDM model summarizes our current cosmological understanding of our universe [28]. ΛCDM predicts that the universe originated from a single expansion event colloquially called the "Big Bang". During the Big Bang, tue universe originated as a hot spacetime singularity, which abruptly experienced rapid expansion in a process known as inflation. After expansion the inflationary field eventually decayed into a population of quarks, gluons, leptons, and photons, which were kept in thermal equilibrium by the high-temperatures of the early universe.

As the universe continued to expand it's density and temperature decreased until the formation of neutral atoms, primarily hydrogen, was possible. At which point the population of photons produced during the Big Bang thermally decoupled. A direct prediction of the ΛCDM model is that this population of photons should still be present, but with a significantly reduced temperature due to the expansion of the universe. This is consistent with the observation of the CMB (cosmic microwave background), which is a population of microwave radiation with a blackbody temperature of 2.7 K. The CMB is extremely uniform in all directions with slight anisotropies that can be analyzed to study the evolution of the early universe. A series of experiments have measured the CMB with increasing levels of precision, which has lead to a significant increase in our current understanding of cosmology.

In addition to the CMB, inflation predicts the existence of a CνB (cosmic neutrino background) [29], which are the remnant neutrinos produced during the Big Bang. Since neutrinos only interact via the weak force, they decouple from the hot Big Bang plasma at an earlier time than the CMB radiation. The temperature at which the CνB decouples depends on the neutrino rest mass. Neutrinos play a unique role in the ΛCDM model, due to the fact that neutrinos act as radiation early in the universe but as matter in the late universe. This leads to specific signatures that impact the expected anisotropies of the CMB as well as the distribution of matter in the universe [30]. By combining measurements of the CMB with measurements of the large-scale structure (LSS) of the universe one can constrain the neutrino mass scale by fitting these datasets with the ΛCDM model. This analysis results in the some of the most stringent constraints on the neutrino mass. Recent analyses [28] have been able to constrain the neutrino mass scale to

$$\Sigma_{m_\nu} \equiv \sum_i m_i < 0.11 \text{ eV}, \tag{2.17}$$

where m_i are the neutrino mass eigenstates.

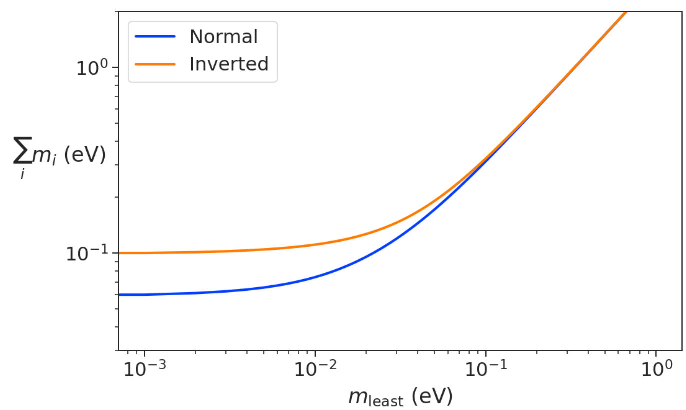


Figure 2.3: The neutrino mass observable measured by cosmology as a function of the lightest neutrino mass eigenstate.

The observable Σ_{m_ν} constrains the neutrino mass by setting the mass of the lightest neutrino mass eigenstate (m_{least}). In the normal mass ordering Σ_{m_ν} can be rewritten in the form

$$\Sigma_{m_\nu} = m_{\text{least}} + \sqrt{\Delta m_{21}^2 + m_{\text{least}}^2} + \sqrt{\Delta m_{32}^2 + m_{\text{least}}^2}, \tag{2.18}$$

where it is clear that a measurement of Σ_{m_ν} effectively sets the neutrino mass scale through m_{least}. The analogous formula for the inverted mass ordering is

$$\Sigma_{m_\nu} = m_{\text{least}} + \sqrt{-\Delta m_{32}^2 + m_{\text{least}}^2} + \sqrt{-\Delta m_{31}^2 + m_{\text{least}}^2}. \tag{2.19}$$

In figure 2.3 we plot the observable Σ_{m_ν} as a function of m_{least}.

Upcoming experiments [31] are planned to refine measurements of the CMB, LSS, and other cosmological observables. With this additional data it is possible that in the near future cosmological measurements will be able to positively constrain the neutrino absolute mass scale. However, the strength of these limits strictly depend on the accuracy of the ΛCDM model, which highlights the need for direct experimental measurements of the neutrino mass to confirm the predictions of cosmology and to fix the neutrino mass parameter in future cosmological analyses.

2.5.2 Limits from Neutrinoless Double Beta-decay Searches

If neutrinos are Majorana fermions, then the neutrino is equivalent to its own antiparticle and lepton conservation is not an exact law of nature [32]. Limits on the rate of neutrinoless double beta-decay ($0\nu\beta\beta$), are some of the most powerful current tests of lepton number conservation [28]. If $0\nu\beta\beta$ were observed it would direct evidence that neutrinos are Majorana fermions, and provide a method for measuring the neutrino mass scale.

Standard double beta-decay occurs when two neutrons contained in the nucleus spontaneously decay into two protons, which results in the production of two electrons and two neutrinos (see Figure 2.4). However, during $0\nu\beta\beta$ the two neutrinos self-annihilate

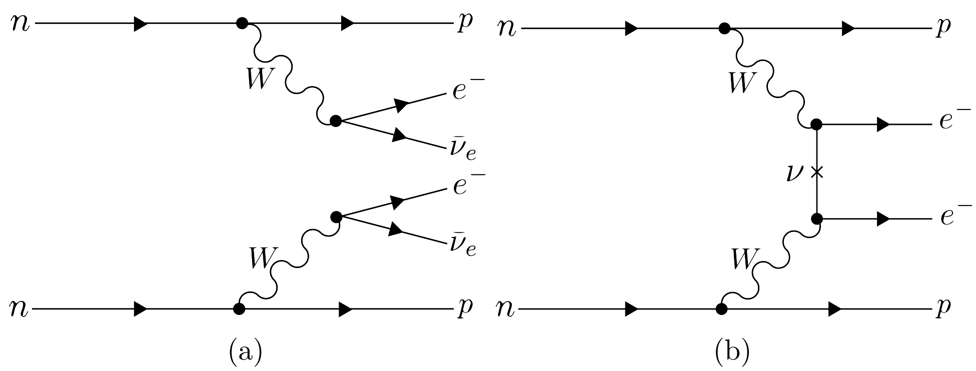


Figure 2.4: Feynman diagrams for double beta-decay (a) and $0\nu\beta\beta$(b).

producing only two electrons, which violates lepton number by two.

Assuming that the exchange of two Majorana neutrinos is the dominant channel for $0\nu\beta\beta$, then a measurement of the $0\nu\beta\beta$ half-life for a particular isotope can be used to set the neutrino absolute mass scale [33]. The half-life is written in terms of the effective neutrino mass for $0\nu\beta\beta$ ($m_{\beta\beta}$) using the equation

$$T^{0\nu}_{1/2} = \frac{1}{G|\mathcal{M}|^2 m^2_{\beta\beta}}, \tag{2.20}$$

where G is the phase-space factor for the decay and $\mathcal{M}$ is the relevant nuclear matrix element. $m_{\beta\beta}$ is given by an incoherent sum of the neutrino mass eigenstates weighted by the PMNS mixing matrix parameters,

$$m_{\beta\beta} = \left| \sum_i U_{ei}^2 m_i \right|. \tag{2.21}$$

The information provided from $0\nu\beta\beta$ on the neutrino mass scale can be visualized by expressing the value of $m_{\beta\beta}$ in terms of m_{least} and two relative Majorana phases [34]. The allowed regions for $m_{\beta\beta}$ as a function of m_{least} are shown in Figure 2.5 as the regions bounded by the black curves overlayed with the discovery probabilities of future $0\nu\beta\beta$ decay experiments based on current neutrino data.

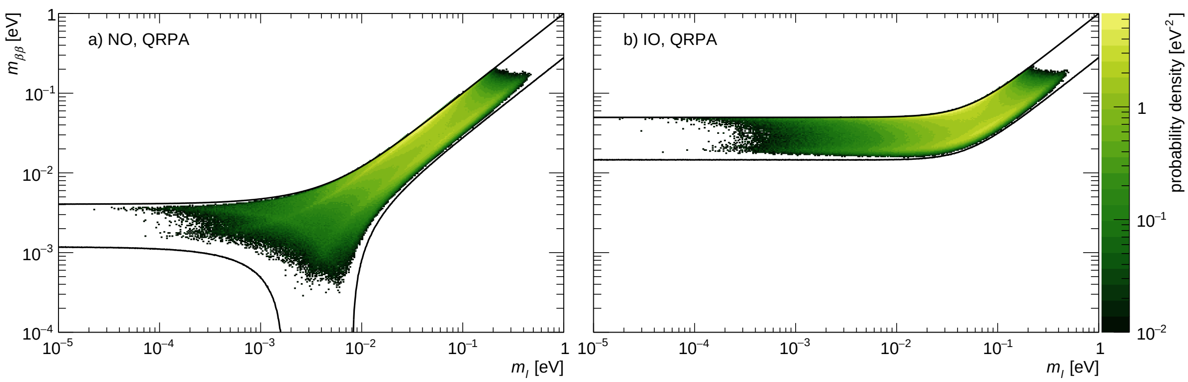


Figure 2.5: The discovery probabilities for the future generation of $0\nu\beta\beta$ experiments as a function of $m_{\beta\beta}$ and m_{least}. Figure from [34].

Because of the possibility of cancellation due to the unknown Majorana phases included in the sum specified by Equation 2.21, the information gained is necessarily imperfect. Additionally, theoretical uncertainties in the calculation of the nuclear matrix elements complicates the calculation of $m_{\beta\beta}$ from a measurement of $0\nu\beta\beta$ half-life. Similar to cosmology, there is a high degree of complementarity between direct measurements of the neutrino mass and $0\nu\beta\beta$. In particular, a measurement of m_{least} to less than than 0.1 eV sensitivity provides significant information for $0\nu\beta\beta$ searches based of the discovery probabilities of Figure 2.5.

2.5.3 Limits from Beta-decay

Certain processes involving neutrinos, in particular beta-decay (see Figure 2.6), have initial states with well-defined total energies and final states that can be measured with high accuracy and precision. Beta-decay involves the decay of an unstable isotope where a neutron spontaneously converts to a proton and emits and electron and anti-neutrino

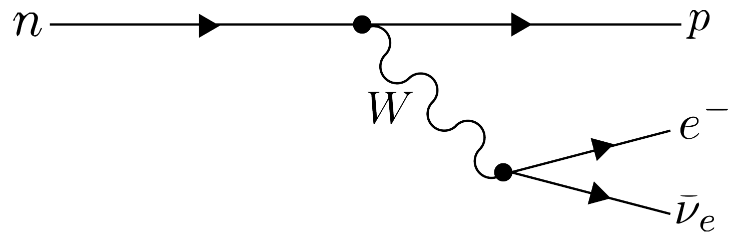


Figure 2.6: A Feynman diagram of beta decay

("neutrino" for brevity) to conserve charge and lepton number [1]. Therefore, by applying the principles of energy and momentum conservation, a measurement of the kinematics of the final state can be used to constrain the neutrino mass [35].

Using beta-decay to measure the neutrino mass can be tied back to Fermi's original 1934 theory of nuclear beta-decay [12] (see Figure 2.7). Because the constraints on the

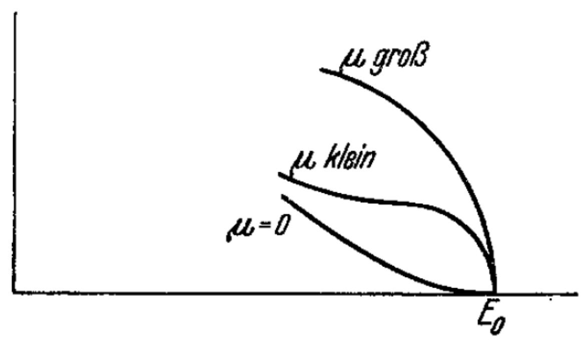


Figure 2.7: A figure from Fermi's 1934 paper on a theory of beta-decay depicting the kinetic energy spectrum of the emitted electron. The effect of the neutrino mass, written as μ, is to distort the shape of the spectrum near the endpoint from the zero-mass spectrum.

neutrino mass from beta-decay depend only on the final state measurement capabilities and the principles of energy and momentum conservation, neutrino mass measurements with beta-decay are sometimes called direct measurements. A direct measurement like beta-decay contrasts with other neutrino mass measurements approaches that are model-dependent such as cosmology and $0\nu\beta\beta$, which provide complementary ways to study the physics of massive neutrinos.

The isotope of choice for direct neutrino mass measurements with beta-decay has been tritium ($^3H_2$) for many decades, because it conveniently fulfills many experimental requirements. Of upmost importance is a decay with a low Q-value, which is the available kinetic energy based on the mass difference between the initial and final states. The effect of a massive neutrino on the shape of the spectrum is magnified for low Q-values and tritium decays have an unusually low Q-value of 18.6 keV.

Additionally, tritium beta-decay is a super-allowed decay, which results in a relatively short half-life of 12.3 years. Therefore, it is relatively easy to obtain a high-activity using a small source mass. High-activity is desirable because of the low-activity near the tritium spectrum endpoint. For tritium beta-decays only a factor of 3×10^{-13} of the decays occur in the last 1 eV of the spectrum. Isotopes with Q-values lower than tritium are known [35], but this is outweighed by exceedingly long half-lives leading to unobtainable source masses.

The measurement involves quantifying the effect of the neutrino's mass on shape of the electron's kinetic energy spectrum near the endpoint. The shape of the kinetic energy spectrum (see Figure 2.8) is given by

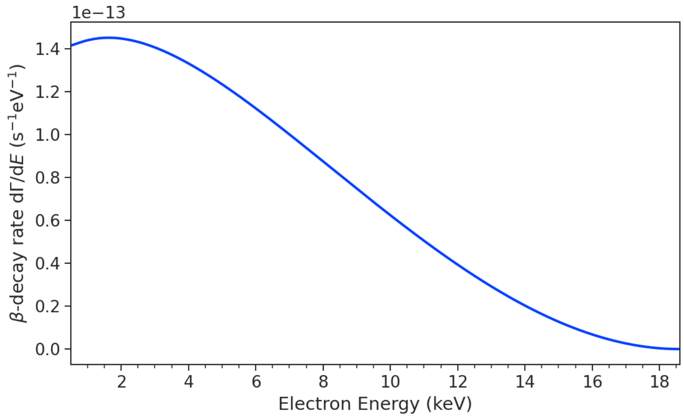
$$\begin{aligned} \frac{d\Gamma}{dE} =& \frac{G_F^2 |V_{ud}|^2}{2\pi^3} (G_V^2 + 3G_A^2) F(Z,\beta) \beta (E + m_e)^2 (E_0 - E) \\ & \times \sum_{i=1,2,3} |U_{ei}|^2 [(E_0 - E)^2 - m_i^2]^{1/2} \Theta(E_0 - E - m_i), \end{aligned} \tag{2.22}$$

where G_F is the Fermi coupling constant, V_{ud} is an element of the CKM matrix, E is the kinetic energy of the electron, β is the velocity of the electron divided by the speed of light, E_0 is the endpoint energy assuming zero neutrino mass, $F(Z, \beta)$ is the Fermi function, and $\Theta(E_0 - E - m_i)$ is the Heviside function, which enforces energy conservation. One can see that the decay spectrum is actually a combination of three spectra with different endpoints based on the actual values of the neutrino mass eigenstates, m_i. This results in "kinks" in the spectrum shape due to the overlapping spectra, but such an effect would likely be impossible to resolve given the finite energy resolution of a real experiment and low statistics.

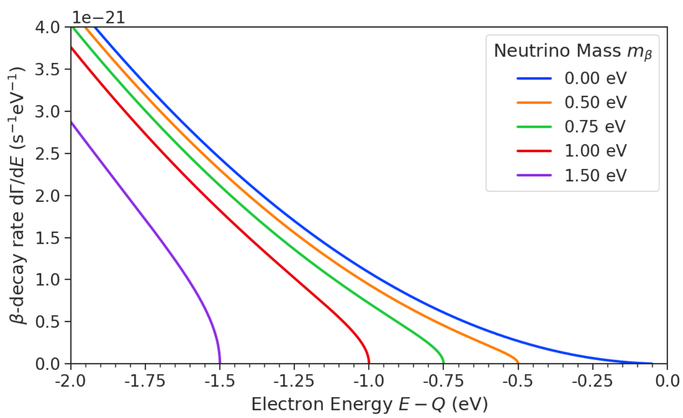
The neutrino mass scale variable measured by beta-decay is given by

$$m_\beta^2 = \sum_i |U_{ei}|^2 m_i^2, \tag{2.23}$$

where m_β is the electron-weighted neutrino mass or simply "neutrino mass" for brevity. m_β corresponds to a particular weighted sum of the neutrino masses, which is distinct from effective neutrino masses such as $m_{\beta\beta}$ [35]. Assuming unitarity, the neutrino mass can be expressed in terms of the PMNS matrix elements, squared mass differences, and



(a)



(b)

Figure 2.8: The tritium beta-decay spectrum. The affect of a massive neutrino on the spectrum is to change it's shape near the endpoint by an amount proportional to the size of the neutrino mass. This suggests that a sufficiently high-statistic and high-resolution measurement of the spectrum endpoint would be able to measure the neutrino mass.

the lightest neutrino mass eigenstate. For the normal mass ordering the equation is

$$m_\beta^2 = m_{\text{least}}^2 + |U_{e2}|^2 \Delta m_{21}^2 + |U_{e3}|^2 \Delta m_{31}^2, \tag{2.24}$$

and for the inverted ordering the equation is

$$m_\beta^2 = m_{\text{least}}^2 + |U_{e1}|^2(-\Delta m_{32}^2 - \Delta m_{21}^2) + |U_{e2}|^2(-\Delta m_{32}^2). \tag{2.25}$$

Therefore, a measurement of the neutrino mass in combination with neutrino mixing parameters is effectively a measurement of m_{least}.

Since the neutrino mass is small (< 1 eV), it's effect on the spectrum is limited to the endpoint region. The affect of a non-zero neutrino mass on the endpoint spectrum is plotted for the reader in Figure 2.8. Resolving the small changes in the spectrum shape requires an experimental technique with high statistics, excellent energy resolution, and low background activity.

Chapter 3

Direct Measurement of the Neutrino Mass with Project 8

3.1 Introduction

A promising technique for direct measurements of the neutrino mass beyond the projected limit of the ongoing KATRIN experiment [36] is tritium beta-decay spectroscopy with an atomic tritium source [37]. Atomic tritium, combined with a large-volume, high-resolution energy measurement technique, is capable of measuring the neutrino mass with sensitivity below the 50 meV limit allowed by neutrino oscillations.

Cyclotron Radiation Emission Spectroscopy or CRES is a high-resolution energy measurement technique compatible with atomic tritium production and storage that can enable the next-generation of neutrino mass direct measurement experiments [38]. The Project 8 collaboration is currently engaged in a program of research and development (R&D) aimed at developing the technology necessary for a 40 meV sensitivity measurement of the neutrino mass using CRES and atomic tritium [39].

In Section 3.2 I provide an introduction to the basics of the CRES technique as well as the goals of the Project 8 experiment. Additionally, I sketch out the phased experiment development plan being implemented by Project 8 to build towards a next-generation neutrino mass experiment.

In Section 3.3 I give a brief overview of Phase II of the Project 8 experiment [40, 41], which completed early in 2023. Although the bulk of the work presented in this thesis is relevant to designs of future Project 8 experiments, a description of the work in Phase II provides useful context for the rest of the work.

In Section 3.4 I introduce a CRES measurement concept based on antenna arrays [42], which could be the basis for the ultimate Project 8 neutrino mass experiment. A significant portion of the R&D efforts of Project 8 in Phase III were directed towards simulating and modeling this experimental concept in order to understand the achievable sensitivity to the neutrino mass.

Lastly, in Section 3.5 I introduce conceptual designs of pilot-scale experiments and Phase IV that combine atomic CRES with a large-volume CRES detection technique. This includes a design concept for an antenna array based experiment, but also a design for a resonant cavity based experiment. Resonant cavities are discussed in more depth in Chapter 6 and have become the default choice for the Phase IV experiment.

3.2 Cyclotron Radiation Emission Spectroscopy and Project 8

3.2.1 Cyclotron Radiation Emission Spectroscopy — CRES

Time and frequency are two of the most precisely measured quantities in physics. It is often advantageous to convert measurements of other physical quantities like mass or length into frequency measurements due to the digital nature of frequency measurements that make them immune to many sources of noise. Atomic clocks, which operate by measuring the frequencies of various atomic transitions, have been used to measure time with astounding relative uncertainties of 10^{-18} seconds [43]. The extreme precision possible with frequency measurements is often summarized using the a quote from the Physicist Arthur Schawlow who said advise his students to "Never measure anything but frequency!" [44].

Neutrino mass measurements using tritium beta-decay require us to measure perturbations of the 18600 eV tritium endpoint to a precision as low as 0.1 eV, therefore, a spectroscopic technique with extremely high resolution is required for this measurement. The intuitive explanation for why frequency measurements are capable of such high resolutions is that they are essentially counting measurements, which average the number of oscillations of a physical system over time. By observing a rapidly oscillating system over a sufficient length of time one can obtain essentially arbitrary precision on a frequency limited only by the time available for measurement and the SNR of the system.

What is required is that one translate the kinetic energy of the electron into a frequency, and a straightforward way to accomplish this is to place a gaseous supply of tritium into a magnetic field. When an atom decays the resulting electron will immediately begin to orbit around a magnetic field line at the cyclotron frequency, which is proportional to its kinetic energy (see Figure 3.1). The acceleration caused by the orbit leads to the emission of cyclotron radiation that can be detected using an array of antennas or a different RF sensor such as a resonant cavity. The frequency of the radiation gives the electron's kinetic energy, which is used to build the beta-decay spectrum and measure the neutrino mass. The name for this measurement technique is Cyclotron Radiation Emission Spectroscopy or CRES [38].

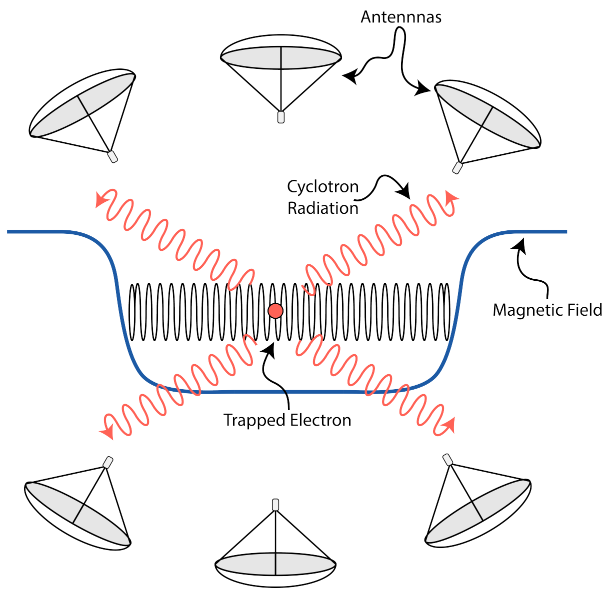


Figure 3.1: A cartoon illustration of the CRES technique. An electron is contained in a magnetic trap so that it's cyclotron radiation can be detected by an array of antennas. Detecting the cyclotron radiation allows us to measure its cyclotron frequency and determine its kinetic energy.

For non-relativistic particles the cyclotron frequency is simply a function of the charge-to-mass ratio of the particle, however, from the relativistic form of the cyclotron frequency

$$f_c = \frac{qB}{2\pi m_e \gamma} = \frac{1}{2\pi} \frac{qB}{m_e + E_{\text{kin}}/c^2}, \tag{3.1}$$

one can see that the kinetic energy (E_{kin}) of the electron is directly proportional to the inverse of the cyclotron frequency (f_c). Electrons with kinetic energies of 18.6 keV are in the weakly relativistic regime with $\beta = \frac{v}{c} = 0.263$ and $\gamma = 1.036$.

The required frequency resolution needed for neutrino mass measurement can be obtained by differentiating Equation 3.1,

$$\frac{df_c}{dE_{\text{kin}}} = \frac{1}{2\pi} \frac{-qBc^2}{(m_e c^2 + E_{\text{kin}})^2}, \tag{3.2}$$

from which we can obtain the relationship between fractional differences in energy and frequency,

$$\frac{df_c}{f_c} = \frac{1-\gamma}{\gamma}\frac{dE_{\mathrm{kin}}}{E_{\mathrm{kin}}}. \tag{3.3}$$

Therefore, an energy precision of 1 eV for an 18.6 keV electron requires a frequency precision of approximately 2 ppm.

The minimum observation time required to achieve this resolution can be estimated using the uncertainty principle as formulated by Gabor [45]. Electrons from tritium beta-decay experience random collisions with the background gas particles, which limits the uninterrupted radiation lifetime. The time between collision events, referred to as track length in the context of CRES measurements, is an exponentially distributed variable. Differences in the track lengths of a population of mono-energetic electrons leads to uncertainty or broadening in the distribution of measured frequencies proportional to the mean track length, τ_λ. The resulting frequency distribution has a Lorentizan profile, whose width is given by the Gabor limit,

$$\tau_\lambda \Delta f_c = \frac{1}{2\pi} \implies \Delta f_c = \frac{1}{2\pi\tau_\lambda}. \tag{3.4}$$

The cyclotron frequency for a 18.6-keV electron in a 1 T field is approximately 27 GHz, from which one can estimate the minimum observation time for 2 ppm frequency resolution at approximately 3 μsec. The Gabor limit is not the true lower bound on the frequency resolution for a CRES signal, since it is based on the details of the Fourier representation of a time-series with a fixed length. If one takes the approach of fitting the CRES signal in the time-domain, then one finds that the limit on frequency precision is given by the Cramér-Rao lower bound (CRLB) [46], which depends on the track length and SNR. The CRLB allows for better precision on the cyclotron frequency, however, the Gabor limit provides an intuitive limit with the correct order of magnitude.

Ensuring that an electron remains under observation long enough so that it's frequency can be precisely measured requires a magnetic trap. A magnetic trap is a local minimum in a background magnetic field generated an appropriate configuration of electromagnetic coils. Since magnetic fields can do no work, there is no danger of the magnetic trap affecting the kinetic energy electron after it is emitted from the beta-decay. One common approach to creating a magnetic trap is the "bathtub" trap configuration, which in it's simplest form consists of two high magnetic field pinch coils aligned on a central axis that are well separated (see Figure 3.2). This configuration produces a trap with a flat uniform bottom and relatively steep walls, which is ideal for CRES measurements.

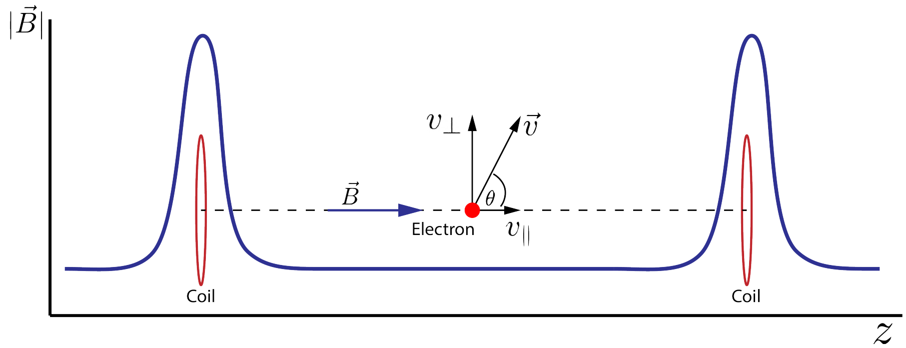


Figure 3.2: An illustration of an electron in a bathtub magnetic trap generated by two well-separated coils.

Electrons produced in the trap oscillate back and forth between the trap walls at a frequency that depends upon the pitch angle, unless they are produced with pitch angles too small to be contained in the trap. Pitch angle is defined as the angle between the component of the electron's velocity perpendicular to the magnetic field and the component parallel to the magnetic field,

$$\tan\theta = \frac{v_\perp}{v_\|}. \tag{3.5}$$

The axial motion of the electron leads to variation in the cyclotron frequency due to the changing value of the magnetic fields. This leads to frequency modulation that generate sidebands in the cyclotron radiation spectrum. Resolving these sideband frequency components is necessary for a complete reconstruction of the CRES signal in the experiment.

Electrons trapped in a cylindrically symmetric trap have three primary components of motion (see Figure 3.3). The dominant component, typically with the highest frequency, is the electron's cyclotron orbit, which encodes information on the electron's kinetic energy. Axial motion from the electron's pitch angle leads to frequency modulation but also a shift in the average magnetic field experienced by an electron. This leads to a correlation between the kinetic energy of the electron and the pitch angle depending on the particular shape of the magnetic trap, which can negatively impact energy resolution. To reduce this correlation one must engineer the trap to have a flat bottom with very steep wall both of which are more easily achieved with a small aspect ratio bathtub trap. Radial gradients in the trap leads to a third component of motion called grad-B drift [47].

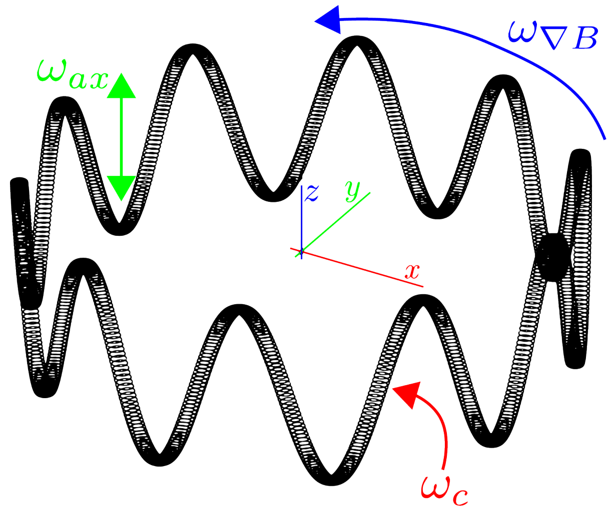


Figure 3.3: A plot of the main components of an electron's trajectory in a cylindrically symmetric trap.

The equation for the drift velocity is

$$\mathbf{v}_{\nabla B} = \frac{m_e v_\perp^2}{2qB} \frac{\mathbf{B} \times \nabla B}{B^2}. \tag{3.6}$$

These additional components of motion all influence the shape of the CRES signal so modeling their effects is critical to proper measurement of the kinetic energy.

The total power of the radiation emitted by an electron in a free-space environment is given by the Larmor equation [48]

$$P(\gamma, \theta_p) = \frac{1}{4\pi\epsilon_0} \frac{2}{3} \frac{q^2 \omega_c^2}{c} (\gamma^2 - 1) \sin^2 \theta_p, \tag{3.7}$$

where ω_c is the cyclotron frequency multiplied by 2π and θ_p is the pitch angle to distinguish it from the spherical angle coordinate. A single electron with a 90° pitch angle and 18.6 keV of kinetic energy in a 1 T magnetic field emits a total radiation power of 1.2 fW, which is quite small compared with typical RF systems, furthermore, one is typically only able to receive a fraction of this total power with an antenna or other detection system. Therefore, RF systems in CRES experiments must be operated at cyrogenic temperatures to limit the noise power such that adequate SNR can be achieved for signal detection and reconstruction. Alternatively, longer tracks enable detection of weaker signals due to the increase in the total signal energy available for the detection algorithm.

3.2.2 The Project 8 Collaboration

The Project 8 collaboration[1] is a group of institutions in the United States and Germany aiming to measure the neutrino mass by developing a novel spectrometer technology based on CRES. In the ultimate Project 8 experiment the CRES technique will be used to measure the beta-decay spectrum using a large source of atomic tritium sufficient to achieve the required statistics in the last $O(10)$ eV of the decay spectrum. Project 8 is targeting a neutrino mass sensitivity below 50 meV [49], which exhausts the range of possible neutrino masses under the inverted hierarchy and is a factor of four less than sensitivity projections for the ongoing KATRIN experiment.

Project 8's proposed experiment requires the development of two novel technologies: the production and trapping of a source of atomic tritium on cubic-meter scales and technology to enable CRES measurements of individual electrons in the same volume.

Atomic Tritium

Previous measurements of the tritium beta-decay spectrum for neutrino mass measurements have relied on sources of molecular tritium for their measurements [36, 50, 51] due to the technical challenges associated with the production and storage of atomic tritium.

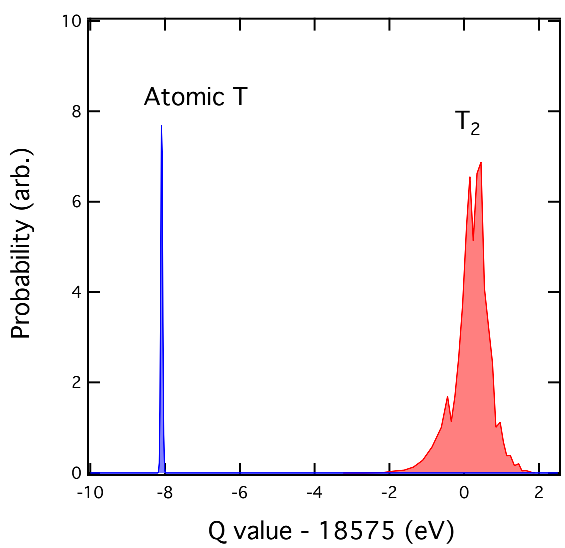


Figure 3.4: A plot of the final state distributions of atomic and molecular tritium. The final state distribution provides the primary contribution to the width of the molecular spectrum whereas thermal doppler broadening is responsible for the width of the atomic spectrum.

[1] https://www.project8.org/

One must supply sufficient energy to the tritium molecules to break the molecular bond and create atomic tritium. Common approaches to this include the use of hot coaxial filament atom crackers as well as plasma atom sources. Both involve heating the tritium atoms to temperatures of > 2500 K, which must then be cooled to temperatures on the order of a few mK so that the tritium atoms can be trapped. Cooling the atoms requires the construction of a large tritium infrastructure and cooling system that can supply a source of cold atoms to the trap.

Once cold tritium atoms are produced they cannot make contact with any surfaces to avoid recombination of the atoms to molecules. Therefore, a magnetic trap is required to store the atoms for a sufficient length of time that they have a chance to decay before escaping the trap. Trapping the atoms requires the construction of a large and complex magnet system that must be cooled to cryogenic temperatures.

The significant experimental complexity caused by atomic tritium makes a molecular source the obvious choice from practical considerations. However, the drawback of molecular tritium for neutrino mass measurement is the irreducible broadening in the electron's kinetic energy due to the final state spectrum of molecular tritium (see Figure 3.4). The broadening of the final state spectra has a RMS amplitude of 436 meV [52, 53] caused by variation in the final vibrational state of the daughter molecule. For atomic tritium the primary sources of broadening in the final state spectrum are magnetic hyperfine splittings (magnitude of $O(10^{-5})$ eV) and thermal Doppler broadening caused by the motion of the trapped atom. For atomic tritium at a temperature of 1 mK thermal broadening is the dominant contribution, providing about 1 meV RMS of broadening to the electron's kinetic energy.

The larger energy broadening with molecular tritium leads to an irreducible statistical uncertainty that limits the achievable sensitivity to approximately 100 meV at 90% confidence. For previous direct measurements of the neutrino mass this uncertainty is an insignificant contribution to the overall uncertainty budget, however, for experiments like Project 8 atomic tritium is a key component to the success of the experiment.

CRES for Neutrino Mass Measurement

Several features of the CRES technique make it an attractive choice for a next generation neutrino mass measurement experiment. For example, with a CRES experiment the volume of the source gas can be the same as the volume of the CRES spectrometer. This is due to the fact that CRES is a remote-sensing technique that can observe the energy of the electron without altering its trajectory or directly interacting with the electron. Given that tritium gas is transparent to cyclotron radiation the kinetic energies of electrons can be measured with an appropriate sensing technology, such as a cavity or antenna array, located directly outside the atom trapping volume.

The current state-of-the-art tritium beta-decay spectroscopy experiment, KATRIN, utilizes the magnetic adiabatic collimation with an electrostatic filter (MAC-E filter) technique to measure the beta-decay spectrum of molecular tritium. In this approach, a source of molecular tritium is located outside the spectrometer. When a beta-decay occurs the electron must exit the tritium source and travel through the MAC-E filter before it can be detected on the other side of the filter using a charge sensor. The measurement statistics of the MAC-E filter are limited by the transverse areas of the tritium source and the filter due to the need to travel through the experiment without scattering. This scaling is less favorable than the volumetric scaling of CRES due to the ability to co-locate source and detector.

Another promising aspect of the CRES technique is the inherently high precision of frequency based measurements. The endpoint of the molecular tritium beta-decay spectrum is approximately 18.6 keV, which dwarfs the neutrino mass scale of < 1 eV/c^2 by at least a factor of 10^5. Measuring the effect of such a small mass on a high energy electron requires excellent energy resolution. Since frequency measurements are essentially counting measurements they are intrinsically quite accurate due to the ability to measure the cyclotron frequency by effectively averaging over millions of cyclotron orbits. Using off-the-shelf RF components its is possible to achieve part-per-million accuracy on the kinetic energy with the CRES technique.

CRES is also nearly immune to typical sources of backgrounds that plague other experiments. Since CRES operates via non-destructive measurements of the electron's cyclotron frequency potential sources of background electrons are effectively filtered out by limiting the frequency bandwidth of the measurement. The fiducial volume of the experiment is free from any surfaces that could introduce stray electrons and electrons from sources outside the fiducial volume can be prevented from entering the experiment.

Neutrino Mass Sensitivity Goals

Project 8's ultimate goal is to combine CRES with atomic tritium to measure the neutrino mass with 40 meV sensitivity at the 90% confidence level (see Figure 3.5). This sensitivity is sufficient to fully exhaust the range of allowable neutrino masses under the inverted neutrino mass ordering regime and is approximately an order of magnitude less than the projected final sensitivity of the KATRIN experiment. Excluding the full neutrino mass

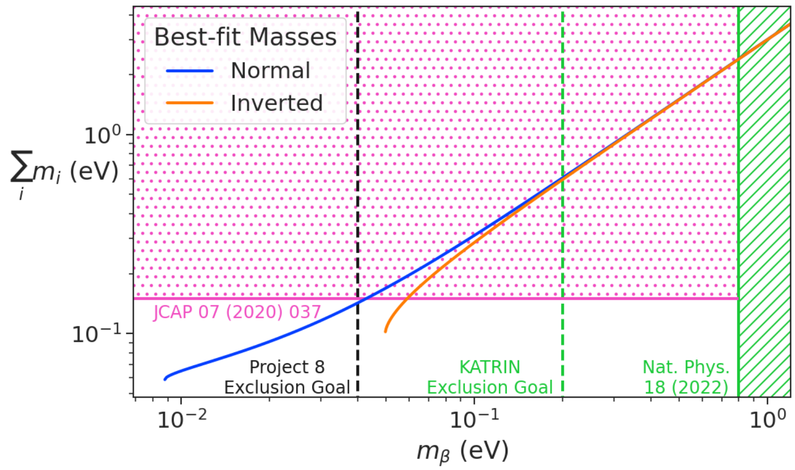


Figure 3.5: Neutrino mass exclusion plot including limits from cosmological measurements and the KATRIN experiment. Allowed ranges for neutrino masses under the normal and inverted hierarchies are shown as the blue and orange lines respectively. The black dashed line shows Project 8's goal neutrino mass sensitivity for the Phase IV experiment.

parameter space would require a sensitivity an order of magnitude lower than what is proposed by Project 8, which would require an experiment whose size and complexity are currently well beyond proposals for the next-generation of neutrino mass direct measurement experiments.

3.2.3 The Project 8 Phased Development Plan

Reaching 40 meV sensitivity requires the simultaneous development and eventually combination of CRES and atomic tritium. These technologies require a significant up-front research and development (R&D) investment to build-out the required capabilities for a 40 meV CRES experiment. Therefore, Project 8 is following a phased experiment plan in which incremental progress can be made towards the ultimate goal of a 40 meV neutrino mass measurement with CRES.

Phase I and II: Proof of Principle and First Tritium Measurements

The earlier phases of the Project 8 experiment, Phase I and II, were focused on demonstration and development of the CRES technique itself as well as a proof-of-principle measurement of the neutrino mass using the CRES technique.

In Phase I, Project 8 performed a proof-of-principle measurement of the ^{83m}Kr

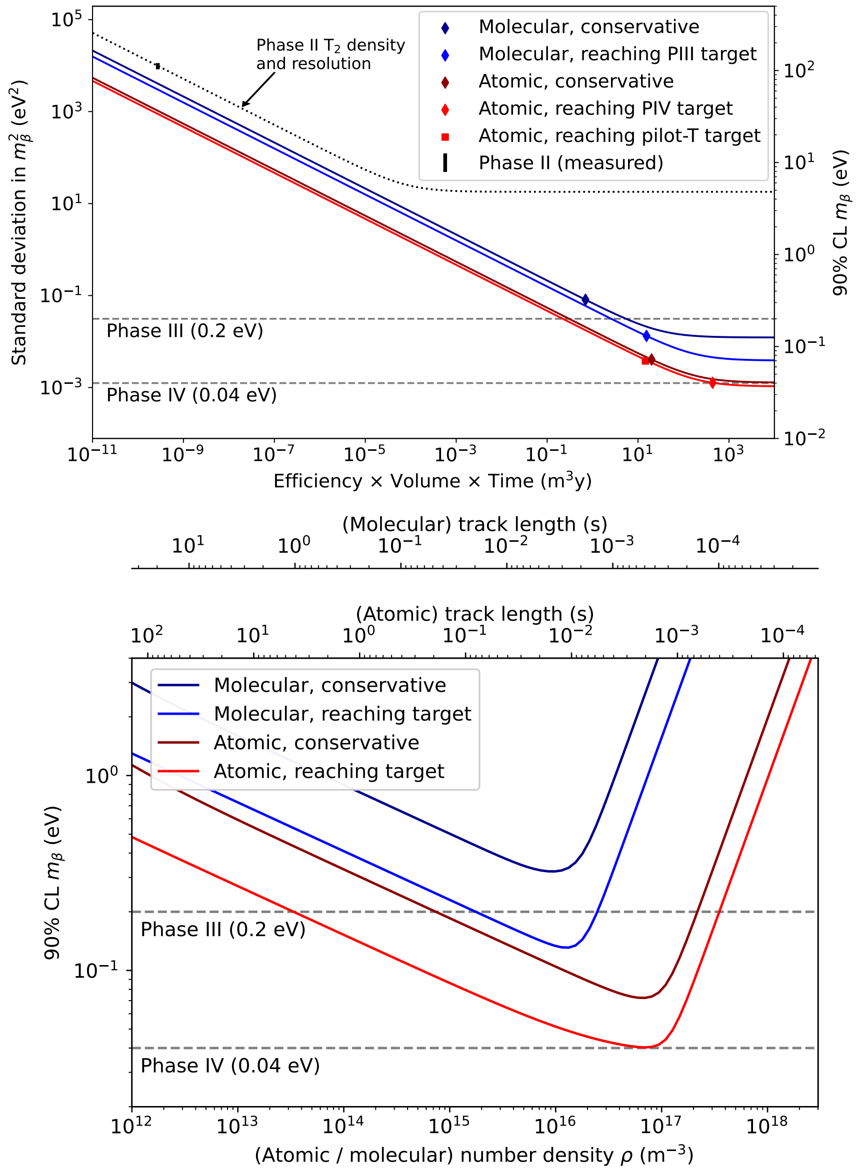


Figure 3.6: Sensitivity calculations for a cavity based CRES experiment that demonstrate the neutrino mass measurement goals of the Project 8 collaboration throughout the phased development plan. The blue curves indicate molecular tritium sources and the red curves indicate atomic tritium sources. In the current plan Phase III contains two tritium experiments. The first is the Low-frequency Apparatus (LFA) which is a molecular tritium experiment and the second is the atomic tritium pilot-scale experiment that ends Phase III. The sensitivity of these experiments is primarily a function of statistics, however, there is a critical density beyond which CRES electrons do not have enough time to radiate between collisions for a high-resolution frequency measurement leading to worse sensitivity.

spectrum using CRES, which marked the first ever energy spectrum measurement with CRES. The experiment included all of the main components expected for the full-scale version of the experiment. An electron source consisting of a gas of ^{83m}Kr was supplied to a waveguide gas cell constructed out of a segment of WR-42 waveguide and sealed with Kapton windows at the top and bottom. A magnetic trapping region was created in the waveguide cell using a single electromagnetic coil wrapped around the waveguide which provided a trapping volume on the order of a few cubic-millimeters. Detection of the cyclotron radiation was performed by connecting the waveguide cell to an additional segment of waveguide that transmitted the radiation to a cryogenic amplifier.

Success in Phase I was achieved with the 2014 publication of the measured ^{83m}Kr conversion spectrum [54], which contains a mono-energetic 17.8-keV as well as several other conversion lines at higher energies. Publication of this result marked the official end of Phase I and the start of Phase II in which Project 8 shifted its focus to the demonstration of the first tritium beta-decay spectrum using CRES. Phase II successfully concluded in 2023 with the submission of the papers demonstrating the first tritium beta-decay spectrum endpoint and neutrino mass measurement using CRES. For more information on Phase II please see Section 3.3.

Phase III: Research and Development and a Pilot-scale Experiment

After Phase II Project 8 has shifted focus to R&D towards the construction of an experiment that demonstrates all the technologies required for a 40 meV measurement of the neutrino mass. The goal for this pilot-scale experiment is to successfully retire all technological and engineering risks associated with the Phase IV experiment, while being a scientifically interesting experiment in its own right that has sensitivity to neutrino masses on par with KATRIN's final projected sensitivity.

Phase III R&D is divided into two equally important efforts — atomic tritium and CRES detection techniques. Atomic tritium development in Phase III includes the development of all aspects of the tritium system. This includes the production of tritium atoms, atomic cooling and recirculation systems, purity and isotope concentration monitoring, and atom trapping. Currently, Project 8 is operating small scale demonstrator systems developing atom crackers to show that atom production at the estimated rates needed for Phase IV is achievable. Future efforts will continue the current developments on atom production and expand to include demonstrations of atomic cooling with an evaporative beam line as well as atom trapping using Halbach magnet arrays.

The need for new CRES detection techniques is driven by the drastic increase in scale from Phase II to the Phase IV and the pilot-scale experiments. The physical volume used for CRES in Phase II was on the order of a few cubic-centimeters, and achieving Project 8's sensitivity target of 40 meV requires an experiment volume on the multi-cubic meter scale. Therefore, the waveguide gas cell CRES detection technique used in Phase II is not a feasible option for the future of Project 8 due to it's inability to scale to the required size.

Two alternative CRES detection techniques have been proposed for the pilot-scale experiment — antenna arrays and resonant cavities (see Section 3.4 and Chapter 6). Both approaches have relative advantages and disadvantages, however, the improved understanding of the antenna array and cavity approaches to CRES in the recent years has led to cavities being the preferred technology for the pilot-scale experiment and Phase IV due to the estimated reduced cost and complexity of this approach. Since a large degree of the work presented in this thesis is focused on the development of the antenna array CRES technique as well as the design of demonstrator experiments, we described the proposed R&D plan for antenna array CRES in Phase III in Section 3.4.

Cavity CRES R&D in Phase III consists of a series of demonstrator experiments intended to demonstrate cavity CRES at a variety of scales and magnetic fields using electrons from ^{83m}Kr, an electron gun, and potentially molecular tritium sources. The near-term cavity effort in Project 8 is the cavity CRES apparatus (CCA), which is a small-scale cavity experiment operating near 26 GHz, that will perform the first CRES measurements using a small cavity. This experiment will pave the way towards larger scale cavity experiments in preparation for the eventual pilot-scale tritium experiment.

The pilot-scale experiment is the first experiment, which will combine atomic tritium and large-volume CRES detection in the same experiment. It will directly demonstrate all the technologies required for Phase IV such that no technical risks remain for scaling the experiment to required scale. A robust approach to scaling the pilot-scale experiment is to simply build multiple copies of it for the Phase IV experiment.

Phase IV: Project 8's Ultimate Neutrino Mass Experiment

The design of Phase IV should be a direct extension of the pilot-scale CRES experiment that marks the official end of Phase III (see Section 3.5). The Phase IV experiment represents the final experiment in the Project 8 neutrino mass measurement experiment plan and will have sensitivity to neutrino masses of 40 meV.

3.3 Phase II: First Tritium Beta Decay Spectrum and Neutrino Mass Measurement with CRES

In Phase II Project 8 demonstrated the first ever measurement of the tritium beta-decay spectrum endpoint using the CRES technique, which lead to the first neutrino mass measurement by the Project 8 collaboration. This milestone was made possible by many improvements in the CRES technique and in the understanding of CRES systematics, which takes an important first step towards larger scale measurements of the tritium beta-decay spectrum with CRES. In this section, I briefly describe some important elements of the Phase II experiment, with the goal of contextualizing the research and development efforts for Phases III and IV of Project 8. For more complete descriptions of the work that lead to Project 8's Phase II results please refer to the relevant publications by the collaboration [40, 41].

3.3.1 The Phase II CRES Apparatus

Magnet and Cryogenics

The magnetic field for the the Phase II experiment is provided by a nuclear magnetic resonance (NMR) spectroscopy magnet with a central bore diameter of 52 mm (see Figure 3.7). The magnet produces a background magnetic field with an average value of 0.959 T and a 10 ppm variation across the bore diameter achieved using several shim coils built into the magnet. Using an external NMR field probe the variation of the magnetic field along the vertical axis of the magnet bore was measured to obtain an accurate model of the magnetic field so that the CRES cell could be positioned for optimal magnetic field uniformity.

An external solenoid magnet was installed inside the magnet bore to provide the ability to shift the magnitude of the background magnetic field by values on the order of a few mT. The solenoid has inside diameter of 46 mm and a length of 350 mm, which terminates in a vacuum flange that allows it to be inserted into the NMR magnet bore from the bottom. By shifting the value of the magnetic field by a few mT, the cyclotron frequencies of electrons produced by the 17.8 keV ^{83m}Kr internal-conversion line [55] can be shifted over a range of frequencies on the order of 100 MHz. This allows one to study the frequency dependent behavior of multiple CRES systematics such as detection efficiency that directly affect the measured shape of the tritium spectrum.

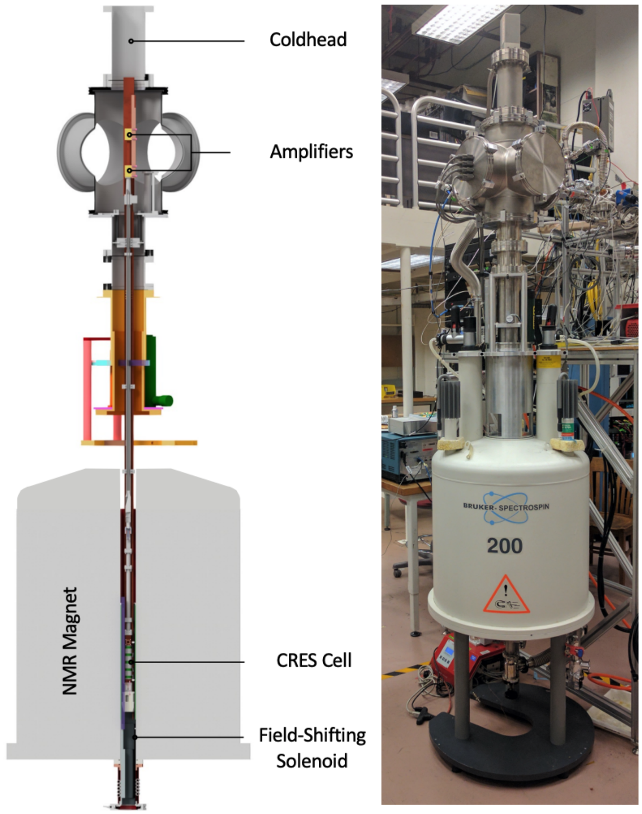


Figure 3.7: The Phase II CRES apparatus used to perform the first measurement of the tritium beta-decay spectrum using CRES.

The inside of the magnet bore diameter was pumped down to a vacuum of less than 10 μtorr using a turbomolecular pump, which allows for cryogenic cooling of the CRES cell and RF system. Cooling power was supplied to the Phase II apparatus using a cryopump with its coldhead mounted above the primary magnet and CRES cell. This arrangement allowed for sufficient cooling power to be delivered to the amplifiers to cool them to a temperature of $\approx$ 40 K, while keeping the amplifiers far enough from the magnet so as not to be damaged by the large field strength. Thermal contact between the coldhead, amplifiers, RF system, and CRES cell is achieved using a copper bar that runs the full length of the apparatus. To prevent freeze-out of ^{83m}Kr on the walls of the CRES cell a separate heater was installed to keep the CRES cell near a temperature of 85 K during the operation of the experiment.

CRES Cell

Located in the most uniform region of the magnetic field is the CRES cell, which is the region of the apparatus where radioactive decays of ^{83m}Kr and T_2 emit electrons that can be trapped and measured using CRES (see Figure 3.8). The CRES cell is manufactured

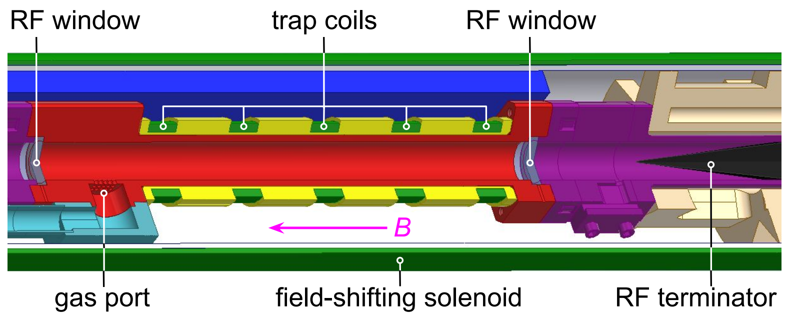


Figure 3.8: Diagram of the CRES cell portion of the Phase II apparatus.

from a segment of cylindrical waveguide designed to operate at K-band frequencies near 26 GHz. The diameter of the waveguide determines which resonant modes of the waveguide will couple to the electron and transmit its radiation to the amplifiers. For Phase II a waveguide diameter of 1 cm was selected, which allows electrons to couple to the TE_{11} and TM_{01} cylindrical waveguide modes. To reduce complexity in modeling and analyzing the CRES data, it is ideal to select a diameter that prevents electrons from coupling to higher-order waveguide modes beyond the fundamental TE and TM modes.

Around the exterior of the cylindrical waveguide are several magnetic coils used to produce magnetic traps inside the CRES cell volume. Without a magnetic trap electrons produced from decays inside the CRES cell quickly impact the cell wall, which prevents a measurement of their cyclotron frequency using CRES. Each coil along the length of the waveguide produces a separate trap that is approximately harmonic in shape. By independently controlling the currents provided to each coil the traps can be configured to have equal values of the magnetic field at the trap bottom despite a non-uniform field from the NMR magnet.

Two primary magnetic trap configurations were used during the Phase II experiment. The first was a shallow trap configuration used primarily for it's high energy resolution to study systematics using ^{83m}Kr decays, and the second was a deeper trap that could trap a higher percentage of pitch angles. The trade-off with this trap is that the higher trapping efficiency comes at the cost of lower energy resolution due to the greater variation in pitch angle. The deep trap was the trap used to measure the tritium beta-decay spectrum in Phase II.

The source gases were delivered into the CRES cell through a gas port located near the top end of the cylindrical waveguide. To prevent the gases from escaping the cell, vacuum tight RF transparent windows are needed to contain the tritium and krypton source gas across a 1 atm pressure differential, while still transmitting the cyclotron radiation without distortion. The crystalline material, CaF_2, which has a thermal expansion coefficient similar to that of copper, was used for this purpose in the CRES cell. Two windows, each 2.4 mm thick, were used to seal off the ends of the CRES cell. The thickness of 2.4 mm corresponds to half of a cyclotron wavelength when one accounts for the permittivity of CaF_2.

RF System

The RF system in the Phase II apparatus transferred the cyclotron radiation from the CRES cell to the receiver chain. The receiver chain performs the down-conversion and digitization required to obtain signals that can be analyzed to determine the cyclotron frequencies of electrons in the CRES cell (see Figure 3.9).

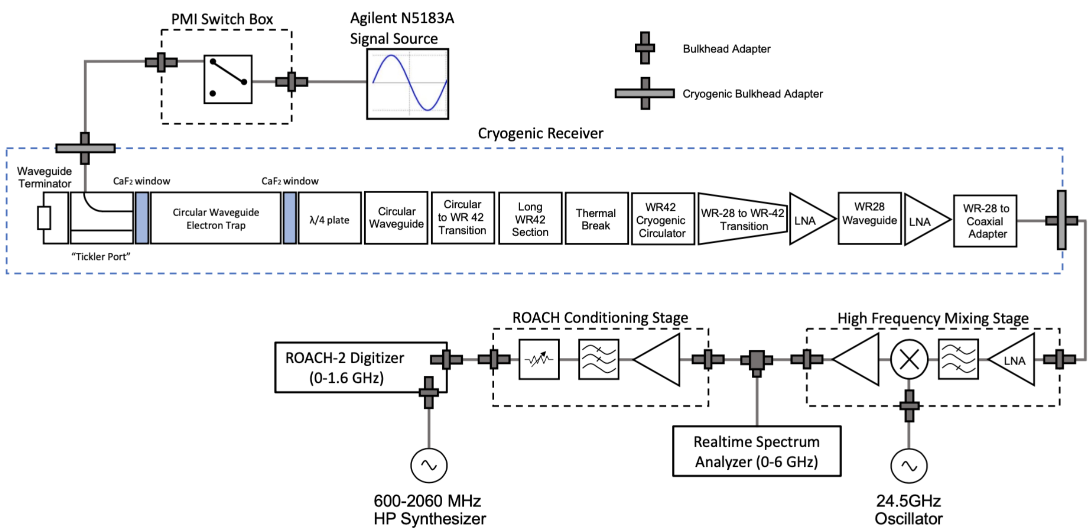


Figure 3.9: RF system diagram for the Phase II apparatus.

Below the CRES cell, at the bottom of the Phase II apparatus, is a tickler port and waveguide terminator. The tickler port is used to inject signals into the CRES cell and RF system for testing and calibration purposes. The waveguide terminator is designed to absorb cyclotron radiation emitted by electrons that transmits out of the bottom of the CRES cell. This lowers the total power received from electrons in the CRES cell, since all the energy radiated downwards is absorbed into the terminator. Earlier iterations of the

Phase II apparatus used an RF short in this location that reflected this power up towards the amplifiers, however, interference between the upward traveling and reflected radiation led to a disappearance in the signal carrier that made reconstruction impossible.

Radiation traveling upward passes through the CaF_2 window passes through a $\lambda/4$ plate, which transforms the circularly polarized cyclotron radiation into linear polarization. The linearly polarized fields next travel through a segment of circular waveguide that transitions into a long segment of WR-42 waveguide that carries the fields out of the high magnetic field region. A thermal break segment is included, which consists of a a segment of gold-plated stainless steel WR-42 waveguide, to help thermally isolate the relatively warm CRES cell from the colder amplifiers. The radiation then passes through a cryogenic circular, which prevents signals reflected from the amplifiers from interfering with the CRES cell before a WR-42 to WR-28 transition connects the waveguide to the first of the cyrogenic amplifiers. The radiation passes through two cyrogenic amplifiers before being coupled to a coaxial termination at the top of the Phase II apparatus.

The coaxial cable transfers the cyclotron radiation signals to a high-frequency mixing stage that performs an analog frequency down-conversion using a 24.5 GHz LO. Two forms of digitization can be used at this stage to readout the CRES data. One is a real-time spectrum analyzer that digitizes the CRES signal data in time-domain and computes the frequency spectrum in real-time, which allows for direct visualization of CRES signal spectrograms as the experiment is running. The real-time spectrum analyzer is most useful for taking small amount of streamed data for debugging and analysis of the system. The other method, which was used to collect the majority of the CRES data in Phase II, is a ROACH-2 FPGA and digitizer system. The ROACH system consists of a fast ADC that samples the CRES signal data at 3.2 GSps. Internal digital down-conversion stages implemented in the FPGA perform a mixing operation that reduces the bandwidth of the CRES signals to 100 MHz. The FPGA implements a 8192 sample FFT and packetizes time and frequency domain records in parallel. The packetized data is then transferred from the ROACH to be analyzed by the data-processing pipeline.

3.3.2 CRES Track and Event Reconstruction

Time-Frequency Spectrogram

The online data-processing is intended to identify interesting data that could contain CRES signals using a software real-time triggering algorithm. Interesting segments of data identified by this algorithm are collected into files that are transferred to a server for offline processing and analysis. The data files contain a continuous series of time-domain samples, broken into a set of records, which are 4096 samples long. The time-series is made up of 8-bit IQ samples acquired at 100 MHz.

Each time-series record is accompanied by an associated frequency spectrum consisting of 4096 frequency bins approximately 24.4 kHz wide, which is represented as a power spectral density. The individual frequency spectra can be organized temporally to create a time-frequency spectrogram that represents the evolution of the cyclotron frequency spectrum over the course of the CRES event (see Figure 3.10). The time-frequency

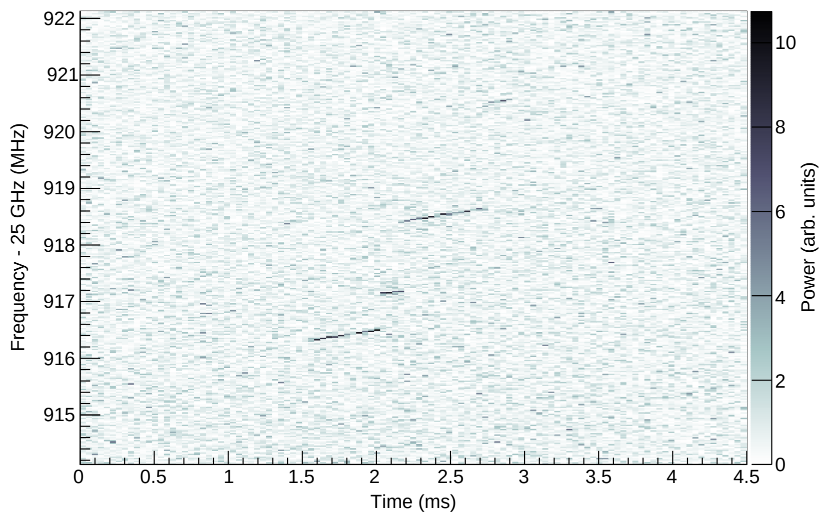


Figure 3.10: The time-frequency spectrogram of a tritium CRES event in the Phase II apparatus.

spectrogram is represented as a two-dimensional image where the color of each pixel is proportional to the power spectral density. Each vertical slice of pixels in the image represents a frequency spectrum, therefore, each horizontal bin represents the data obtained over a duration of $4096 \times 0.01\ \text{MHz}^{-1} = 40.96\ \mu$sec.

CRES Event Data Features

Phenomenologically, a CRES signal appears as a sinusoidal signal whose frequency slow increases ("chirps") over time. Axial motion of the electron in the trap leads to the formation of frequency sidebands that surround the more powerful carrier frequency, due to Doppler modulation of the electron's frequency as it bounces between the walls of the magnetic trap. The critical piece of information that must be extracted from the track and event reconstruction procedure is the carrier frequency, since it is this frequency that gives the cyclotron frequency and thus the kinetic energy. While axial motion from non-90° pitch angles does change the average magnetic field experienced by an electron and, therefore, changes the cyclotron frequency. Because of low-SNR sidebands were unable to be observed in Phase II, so a correction for the effect of the pitch angle on the cyclotron frequency was not possible.

In the time-frequency spectrogram representation the chriping carrier frequency appears as a linear track of high-power frequency bins (see Figure 3.10). The vertical slope of the tracks is caused by the emission of energy from the electron in the form of cyclotron radiation, therefore, the size of the slope parameter is directly proportional to the Larmour power. The continuous track is periodically interrupted by random jumps to higher frequency and lower energy caused by random inelastic collisions with background gas molecules. The length of a track is an exponentially distributed variable whose mean value is inversely proportional to the gas density. The size of the frequency discontinuities is directly proportional to the energies of the rotational and vibrational states of background gas molecules.

A CRES event refers to the collection of tracks produced by a trapped electron until it inevitably scatters into a pitch angle that can no longer be trapped. The goal of track and event reconstruction is to first identify the set of tracks present in a time-frequency spectrogram that represents a segment of data acquired in the Phase II apparatus. These tracks must then be clustered into events from which we can determine the first track produced by the electron and thus estimate it's starting cyclotron frequency and kinetic energy.

Track Reconstruction

The first step in CRES event reconstruction is the identification of tracks in the time-frequency spectrogram, which is essentially an image processing task. Track finding starts by normalizing the power spectral density based on the average noise power. Next a power threshold is applied to the normalized spectrogram where only bins that have a signal-to-noise ratio greater than five are selected to build tracks. In this case signal-to-noise ratio is defined as the ratio between the normalized, unitless power of a bin divided by the average normalized power across the full frequency spectrum.

The sparse spectrogram produced by this power cut consists only of a sparse collection of high-power frequency bins that could be part of a CRES signal track (see Figure 3.11). In this form is it much easier to identify tracks "by eye", however, for the Phase II analysis Project 8 developed its own custom-made track finding algorithm, called the

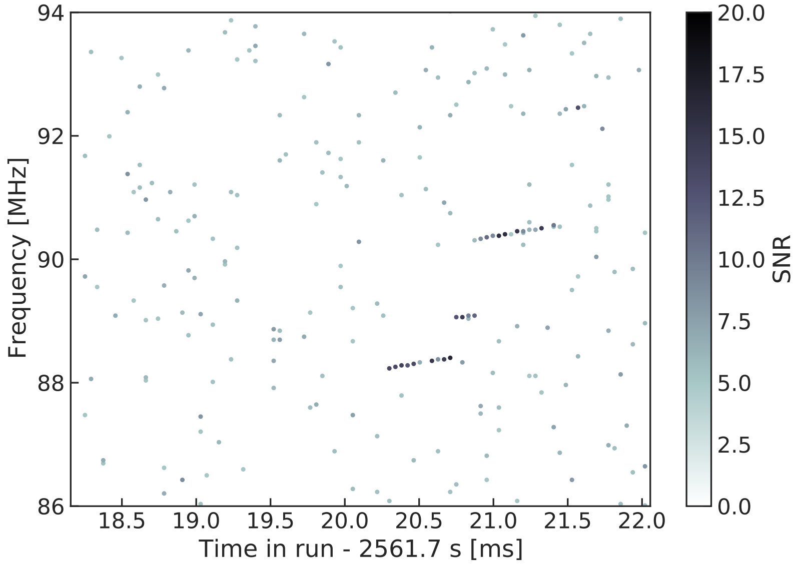


Figure 3.11: The sparse spectrogram obtained by placing a power cut on the raw spectrogram shown in Figure 3.10.

sequential track finder (STF).

The STF algorithm processes the sparse spectrogram in sequential fashion, processing each time-slice one-by-one until the end of the spectrogram is reached. Tracks are found by searching for points in the sparse spectrogram that appear to fall on a straight line. Multiple configurable parameters are built into the STF algorithm that allow the user to tune the criteria for adding a point to an existing track or creating a new track. These include parameters such as maximum time and frequency differences between subsequent points in a track as well as minimum SNR values for the start and endpoints of the track. Additionally, tracks are required to have a minimum length and slope to be considered potential CRES tracks rather than random noise fluctuations.

The resulting output of the STF is a collection of track objects that consist of the track point objects and their properties. The final step is to calculate track-level properties and apply cuts to reject false tracks found by the STF. This involves the fitting of a line to the collection of track points as well as the total and average power of the track obtained by computing the sum and mean of the points powers. The starting frequency of the track is determined by calculating the time coordinate that intersects with the linear fit. A cut is the performed to remove all tracks that do not have a specified average power over their duration, which helps to remove the majority of noise fluctuations that have passed all previous cuts up to this point.

Event Reconstruction

After track reconstruction comes event reconstruction where the identified tracks are grouped into events the correspond to the trjectory of a single electron in the trap. This procedure attempts to match tracks head to tail by checking if the start and end times of a pair of tracks falls withing a certain tolerance. This tolerance is a configurable parameter that can be tuned to an optimal value using Monte Carlo simulations of events in the Phase II apparatus.

After the event building procedure has completed there remains a small likelihood that false tracks have made it through to the event reconstruction stage. Typically, cuts at the track level are able to remove 95% of the false tracks identified by the STF, which leads to a significant number of false tracks at the event building stage. However, the additional event-level information makes it possible to reject events that contain these false tracks with a high degree of confidence.

Two event level features are associated with events caused by real electrons — the duration of the first track as well as the number of tracks in the event. Real electrons tend to have event structures with longer first tracks and a higher number of total tracks. Based on the values of these two criteria, a minimum threshold on the average power in the first track was configured to reject false events. The average power in the first track was chosen due to the critical nature of the starting frequency of the first track in an event to the krypton and tritium spectrum analyses.

3.3.3 Results from Phase II

The main result from Phase II was the measurement of the tritium beta-decay spectrum using CRES, which lead to the first neutrino mass limit with CRES. However, Phase II also included a significant ^{83m}Kr measurement campaign to understand important systematics relevant to the tritium spectrum measurement, but also to understanding the fundamentals of the CRES technique itself. This required high-resolution measurements of the ^{83m}Kr internal-conversion spectrum [55], which is an interesting science result in its own right.

The results from Phase II represents a significant effort from the entire Project 8 collaboration over several years. Because the focus of my contributions to Project 8 is directed towards the research and development efforts for the Phase III experiments, the goal in this section is not to provide a detailed description of the analyses that lead to the Phase II results. Rather, I will provide brief descriptions of a few plots representative of the main results from Phase II.

Measurements with Krypton

Measurements with krypton were a key calibration tool for Phase II of the experiment and will continue to be useful in Phase III. In the context of Project 8 krypton measurements refers to CRES measurements of the internal-conversion spectrum of the metastable state of krypton-83, ^{83m}Kr, produced by electron capture decays of ^{83}Rb. A supply of ^{83}Rb was built into the Phase II apparatus gas system that supplied the CRES cell with ^{83m}Kr via emanation.

The ^{83m}Kr internal-conversion spectrum consists of several lines based on the orbital of the electron ejected during the decay. The conversion lines useful to Project 8 are those that emit electrons with kinetic energies that fall inside the detectable frequency bandwidth of the Phase II apparatus. These are the K; L2 and L3; M2 and M3; and N2 and N3 lines with kinetic energies of 17.8 keV, $\approx$ 30.4 keV, $\approx$ 31.9 keV, and $\approx$ 32.1 keV, respectively. The different energies of the lines allowa onw to test the linearity of the relationship between kinetic energy and frequency across the range of frequencies covered by the continuous tritium spectrum.

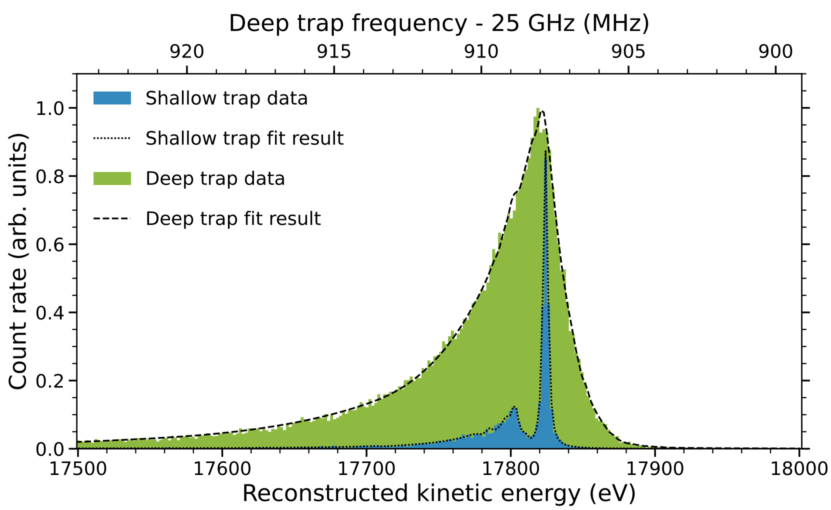


Figure 3.12: Fits to the measured 17.8-keV ^{83m}Kr conversion line using the deep and shallow trap configurations.

Numerous detector related effects relevant to the tritium analysis can be characterized by measuring the shape of the krypton spectrum. Specific examples include variations in the magnetic field as a function of the radial position of the electron, variation in the magnetic field caused by the trap shape, variation in the average magnetic field for electrons with different pitch angles, and the effect of missing tracks due to scattering. These spectrum shape measurements focused on the 17.8-keV krypton line and utilized different trap geometries based on the particular goal of the dataset (see Figure 3.12).

Krypton measurements with a shallow trap allow for high energy resolution, since variation in frequency due to pitch angle differences is sharply reduced in the shallow trap configuration. With this trap the main 17.8-keV peak of the conversion spectrum is clearly visible along with additional satellite peaks at lower energy, which correspond to the shakeup/shakeoff spectrum of the decay. The high accuracy of the fit demonstrates a high degree of understanding of the CRES systematics.

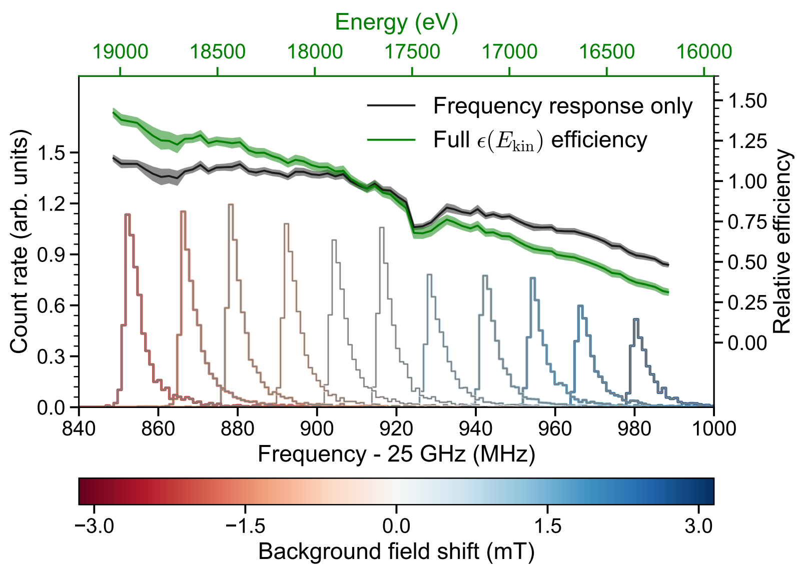


Figure 3.13: Measurements of the 17.8-keV ^{83m}Kr line using the deep trap configuration for different values of the magnetic field from the field shifting solenoid.

The broadening of the krypton spectrum seen for the deeper track is due to the large range of electron pitch angles that can be trapped. Furthermore, with a deeper trap there is a larger parameter space of electron that could be produced with pitch angles that are trappable but not visible in the time-frequency spectrogram. These electrons live in the trap and can scatter multiple times before randomly scattering to a visible pitch angle. This leads to one or more missing tracks earlier in the event, which leads to a misreconstruction of the true starting frequency. By measuring the krypton spectrum shape in the same trap used to detect tritium events, the effect this has on the spectrum shape can be characterized to mitigate its impact on the tritium measurements.

Changes in the Krypton spectrum shape as a function of CRES frequency were

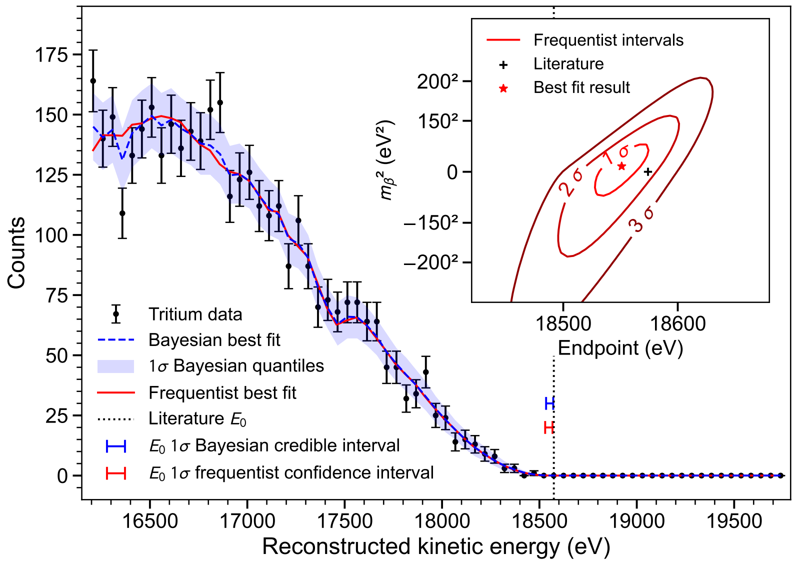


Figure 3.14: The measured tritium spectrum from Phase II with Bayesian and frequentist fits.

used to study the detection efficiency of the Phase II apparatus. Variations in the detection efficiency as a function of frequency directly influences the measured shape of the continuous tritium spectrum, which can lead to errors in the neutrino mass estimate if not modeled appropriately. Using the field shifting solenoid the cyclotron frequency of the krypton 17.83 keV line was shifted across the full frequency range of the tritium spectrum data (see Figure 3.13). Variations in the deep trap krypton spectrum shape can be used to infer the detection efficiency as a function of frequency and correct for this affect in the tritium measurements.

Tritium Spectrum and Neutrino Mass Results

The tritium measurement campaign resulted in the collection of 82 days of detector live time during which 3770 total tritium events were detected. The track and event reconstruction analysis extracted the starting frequencies of these tritium events, which were used to build a frequency spectrum of tritium beta-decays. The resulting frequency spectrum was then converted to an energy spectrum using the information gleaned from the krypton measurement campaign to obtain the tritium beta-decay spectrum (see Figure 3.14).

CRES is inherently a very low background technique with the dominant source of noise being random RF fluctuations. Monte Carlo simulations backed validated using measurements of the RF noise background were used to set track and event cuts to guarantee that zero false events would occur over the duration of the experiment with 90% confidence. Notably, the measured spectrum has zero events beyond the tritium spectrum endpoint, which allows us to constrain the background rate in the Phase II apparatus to less than 3×10^{-10} counts/ev/s. Achieving a low background is critical for future neutrino mass experiments that seek to measure the neutrino mass with less than 100 meV sensitivity.

Bayesian and frequentist based fits to the measured tritium spectrum, incorporating information gained about CRES systematics from the krypton measurements, were performed to extract upper limits on the tritium beta-decay spectrum endpoint as well as the neutrino mass. The estimated spectrum endpoints are 18553^{+18}_{-19} eV for the Bayesian analysis and 18548^{+19}_{-19} eV for the frequentist analysis. The quoted uncertainties are 1-σ, and both results are within 2-σ of the literature endpoint value of 15574 eV. The estimated neutrino mass for both results is consistent with $m_\beta^2 = 0$. The 90% confidence upper limits for the Bayesian analysis is $m_\beta < 155$ eV/c^2 and $m_\beta < 152$ eV/c for the frequentist analysis.

Though the neutrino mass results from Phase II are not competitive with KATRIN the experimented was a promising first step towards the development of more precise neutrino mass measurements using CRES. The low-background and high-resolution achievable with krypton measurements are promising features of the technique that were demonstrated with the Phase II apparatus. As new technologies are developed to enable CRES measurements in larger volume, many of the lessons learned from Phase II will continue to influence the operation and design of future experiments.

3.4 Phase III R&D: Antenna Array CRES

The goal of Phase III in the Project 8 experimental program is to develop the technologies and expertise required to build an experiment that uses CRES to measure the neutrino mass with a target sensitivity of 40 meV. One of the key technologies is a method for performing high resolution CRES measurements in a large volume, which allows one to observe a sufficient quantity of tritium to measure the low-activity endpoint region of the tritium spectrum.

3.4.1 The Basic Approach

One possible approach, suggested in the original CRES publication [38], is to use many antennas to surround a volume of tritium gas in a magnetic field (see Figure 3.15). When a decay occurs the electron will begin to emit cyclotron radiation that can be collected by the array and used to perform CRES. Each antenna in the array collects only a small

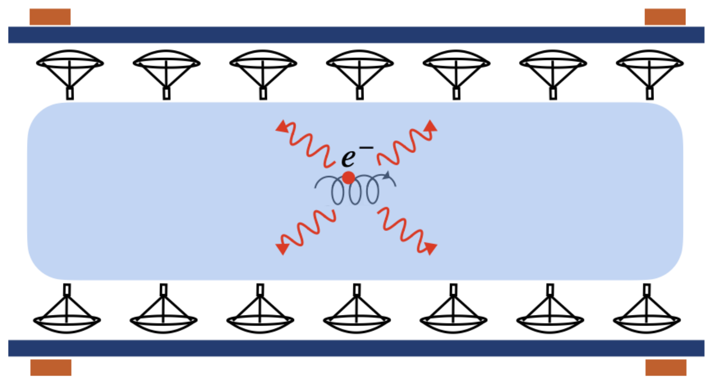


Figure 3.15: A cartoon illustration of the basics of the antenna array CRES technique.

fraction of the electron's signal power, which is less than 1 fW for a 18.6 keV kinetic energy electron in a 1 T magnetic field. Scaling to large volumes with the antenna array approach is accomplished by increasing the number of antennas in the array, which increases the volume under observation proportionally, so that a sufficient population of tritium atoms can be observed to measure the tritium spectrum endpoint shape.

Several features of the antenna array approach make it an attractive candidate technology for a large volume experiment. One example is the accurate position reconstruction made possible by the multichannel nature of the array. Using techniques like digital beamforming it is possible to estimate the radial and azimuthal positions of the electron in the magnetic trap with a precision significantly less than the size of the cyclotron wavelength. This capability allows one to perform event-by-event estimations of the magnetic field experienced by an electron, which is crucial to achieving high energy resolution with the CRES technique.

The easy availability of position information with the antennas array approach is potentially a unique advantage that provides significant flexibility in the magnetic field uniformity requirements compared to other proposed approaches to large volume CRES (see Chapter 6). Spatial discrimination using digital beamforming leads to pileup reduction, which helps to reduce the potential of background events caused by missing tracks or by incorrectly clustering a group of tracks into an event. Limits on the background rate for a neutrino mass measurement with 40 meV sensitivity are stringent and the total activity of the tritium source for such an experiment is gigantic relative to the activity near the endpoint. Thus, pileup discrimination could be an important tool for a large scale CRES experiment.

Another beneficial quality of the antenna array approach is that the volume of the experiment can be scaled independent of frequency by simply adding more antennas to the array (see Figure 3.19). Resonant cavities, the proposed alternative large volume CRES technology, are ideally operated in magnetic fields that cause electrons to move with cyclotron frequencies near the fundamental cavity resonance, to avoid complex coupling of the electron to many cavity modes simultaneously. This leads to a coupling between the cavity volume and the magnetic field magnitude, which forces one to lower the magnetic field in order to increase the experiment scale. Whereas, for antenna arrays, in principle there is no physical limitation on the size of the antenna array that can be used at a particular magnetic field. However, the nature of scaling an antenna array based experiment leads to rapidly increasing cost and complexity due to the large number of antennas, amplifiers, and data streams that require substantial computer processing power to effectively analyze.

3.4.2 The FSCD: Free-space CRES Demonstrator

The complexity of the antenna array CRES technique requires the construction of a small scale demonstration experiment to develop an understanding of technique itself and relevant systematics. Without a demonstrator experiment it is not possible to sufficiently retire the technical risks associated with the full-scale experiment. Therefore, Phase III of the Project 8 experimental program is primarily focused on the development and operation of demonstrator experiments to inform the design of the Phase IV experiment.

The demonstrator experiment developed for antenna array CRES in Phase III is called the Free-space CRES Demonstrator or FSCD. The FSCD is intended as a demonstration of antenna array CRES, but is also a capable neutrino mass measurement experiment in its own right, with a target neutrino mass sensitivity of a few eV using a molecular tritium source.

Magnetic Field

The background magnetic field for the FSCD is provided by a hospital-grade MRI magnet (see Figure 3.16). The magnet produces a magnetic field of approximately 0.958 T, which corresponds to a tritium spectrum endpoint frequency of approximately 25.86 GHz. The

Figure 3.16: An image of the MRI magnet installed in the Project 8 laboratory at the University of Washington, Seattle.

magnet is installed in the Project 8 laboratory located at the University of Washington, Seattle, and is shimmed to produce a uniform magnetic field with variations on the ppm-level. Measurements of the magnetic field non-uniformities are performed using a NMR probe and rotational gantry to capture measurements of the magnetic field around an elliptical surface in the center of the MRI magnet. During the operation of the FSCD an array of Hall or NMR magnetometers would be used to periodically measure the magnetic field to monitor its time stability.

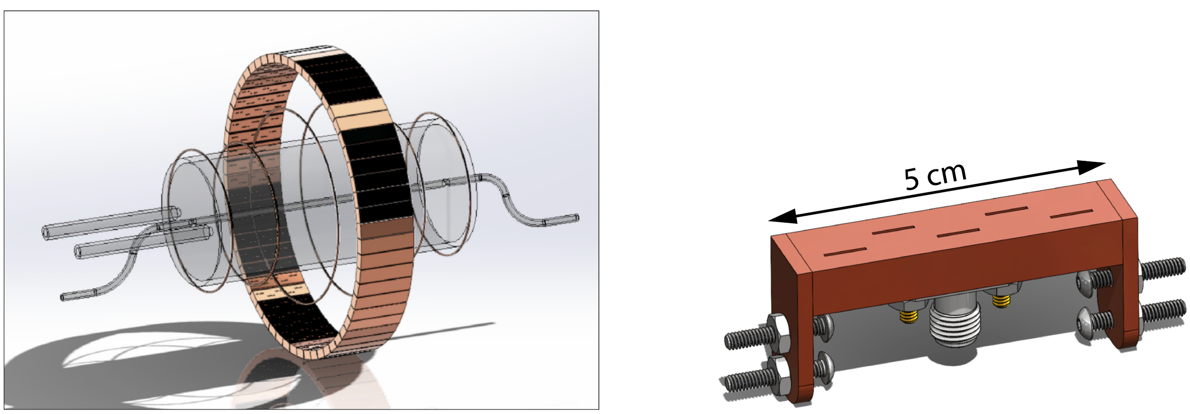
Inside the field of the MRI magnet additional electromagnets would be installed that provide the capability to shift the value of the background magnetic field and produce a magnetic trap. Shifting the background magnetic field by a few μ T lets one control the cyclotron frequencies of electrons with a fixed kinetic energy, which is key to an effective calibration of the FSCD. The preferred calibration method for the FSCD is a mono-energetic electron gun that can inject electrons into the magnetic trap with a known kinetic energy. In combination with the field shifting magnet, one can vary the cyclotron frequencies of the electrons to measure the response of the antenna array as a function of the radiation frequency and electron position. This procedure characterizes the response of the antenna array and provides further information on magnetic field uniformity, which important to achieving good energy resolution.

The design of the magnetic trap is absolutely critical to the success of a CRES experiment. The ideal shape is the perfect magnetic box, which has a flat bottom and step function walls. Any variation in the average magnetic field experienced by an electron leads to changes in the cyclotron frequency that can make determining the true starting kinetic energy more difficult. This includes changes in the magnetic field caused by the walls of the magnetic trap as well as radial magnetic field variations.

The ideal box trap is completely uniform and has infinitely steep walls that cause no change in the electron's cyclotron frequency as it is reflected from the trap wall, however, such a trap cannot be made from any combination of magnetic coils since it violates Maxwell's equations. One of the goals of magnetic trap design is to identify the configuration of coils that produces a trap that approximates the perfect box trap as closely as possible.

Antenna Array

The canonical antenna array design for CRES is a uniform cylindrical array of antennas that surrounds the magnetic trap volume. Since the FSCD is a demonstrator experiment, the antenna array design is the simplest form of the uniform cylindrical array, which is a single circular ring of antennas with a diameter of 20 cm (see Figure 3.17). Along this



(a) (b)

Figure 3.17: (a) A model of the FSCD antenna array, magnetic trap, and tritium containment vessel design.(b) A more detailed model of a prototype design for the 5-slot waveguide antenna design.

circle are sixty slotted waveguide antennas that fully populate the available space around the array circumference. In order to maximize the power collected from each electron it is optimal to cover as large a fraction of the solid angle around the magnetic trap as possible.

The distance between antennas around the circumference of the array is proportional to the wavelength of the cyclotron radiation. Therefore, maximizing the solid angle coverage of the array, while minimizing channel count to keep the hardware and data acquisition costs manageable, biases one towards smaller array diameters. Antenna near-field effects limit the minimum diameter of the array for a given antenna design since the radiation from electrons that are too close to the array cannot be detected due to destructive interference caused by path-length differences from the electron to different points on the antenna surface.

Slotted waveguide antennas are used in the FSCD antenna array due to their high efficiency and low loss, which comes from the lack of dielectric materials in the antenna structure. Coupling to the waveguide can be performed with a coaxial cable connected at the center or on either end of the waveguide. One of the drawbacks of waveguide antennas is the large amount of space required to fit them inside the limited MRI magnet volume. Alternative antenna designs, constructed from microstrip printed circuit boards require significantly less space at the cost of slightly higher energy loss in the antenna structure.

The FSCD antenna design is a 5 cm long segment of WR-34 waveguide with 5 vertical slots cut into the side. The distance between slots along the length of the waveguide is a half wavelength for optimal power combination between the individual antenna slots. Each slot is offset from the center of the antenna face a small distance in order to most effectively couple the slot to waveguide modes inside the antenna.

The passive power combination achieved by placing 5 slots in a single waveguide is a compromise intended to reduce the cost and complexity of the antenna array system. Each additional channel in the array requires it's own cryogenic amplifier and also increase the required computer power to process the raw data collected by digitizing each channel. Passive summation, achieved by combining antennas into arrays axially, reduces the array channel count at the cost of losses from imperfect passive combination. Imperfect passive combination is caused by effects such as re-radiation of energy from and destructive interference between slots in the waveguide antenna.

Interference and re-radiation eventually limit the achievable the axial extent of passive power combination. The 5-slot designed developed for the FSCD is optimized to minimize the impact of these losses while achieving the maximum amount of axial coverage with a single ring of antennas. Scaling beyond the volume covered by a single ring of antennas is achieved by stacking additional rings of antennas together to cover a larger trap volume for a higher statistics measurement of the tritium spectrum endpoint region. A likely scenario for the FSCD experiment involves a staged experiment approach, where first a series of measurements is performed using only a single ring of antennas followed by experiments that add additional rings to the FSCD. The goal would be to first understand the principles of antenna array CRES using the simplest possible experiment, before attempting to scale the technique by expanding the antenna array size.

Tritium Source

While the primary purpose of the FSCD is as a technology demonstrator, it is unlikely for the collaboration to gain the required confidence in the antenna array CRES technique to perform neutrino mass measurements at the 40 meV sensitivity level without an intermediate scale measurement of the neutrino mass using antenna array CRES. Therefore, the FSCD has an additional scientific goal of measuring the neutrino mass with a rough sensitivity goal of a few eV. This level of precision is achievable using a source of molecular tritium with a volume of approximately 1 L at a density comparable to potential Phase IV scenarios.

Unlike previous CRES experiments, where the tritium source could be co-located with the receiving antenna inside a waveguide transmission line, the tritium source in the FSCD is thermally isolated from the antenna array to avoid freeze-out of the tritium molecules. The tiny radiation power emitted by electrons requires a system noise temperature of ≈ 10 K or less, in order to detect events at a high enough efficiency to reach the neutrino mass sensitivity goals of the experiment. Achieving a system noise of 10 K requires that the antenna array and amplifiers operate at cryogenic, liquid helium temperatures of ≈ 4 K, which significantly lowers the vapor pressure of molecular tritium. By keeping the molecular tritium isolated in an RF-transparent vessel the tritium gas can be kept at a relatively warmer temperature in the range of 30 K to avoid the accumulation of tritium on the experiment surfaces.

Data Acquisition and Reconstruction

A fundamental change in the data acquisition system for the FSCD is the shift from single to multi-channel reconstruction. This transition results in a significant increase in the data-generation rate, which is linearly related to the number of independent channels in the array. The larger data volume coincides with an increased demand for computer processing power based on the need for more precise signal reconstruction algorithms driven by the FSCD and Phase IV sensitivity goals. Therefore, the data acquisition system for the FSCD is likely to represent a significantly larger fraction of the experiment cost and complexity than previous CRES experiments.

Each antenna in the array is connected to a cryogenic amplifier and down-converted from the 26 GHz CRES frequency using an IQ-mixer to reduce the size of the analysis window in which the tritium spectrum is measured. Using an LO with a frequency of approximately 25.80 GHz the antenna array signals can be digitized at a rate of 200 MHz, which is sufficient bandwidth to resolve the complete sideband spectrum produced by axial oscillations of electrons in the FSCD magnetic trap.

Direct storage of the raw FSCD antenna array data is undesirable, since the estimated amount of raw data generated is $O(1)$ exabyte per year. The management and storage of such a large dataset is infeasible for a demonstrator experiment on the scale of the FSCD and would represent a large fraction of the budget for a Phase IV scale antenna array based CRES experiment. Therefore, a sub-goal of the FSCD experiment is the development of real-time reconstruction methods that could reduce the raw data volume by detecting and reconstructing CRES events in real-time. The ultimate goal would be a complete real-time reconstruction pipeline that takes raw voltages samples from the antenna array and returns estimates for the starting kinetic energies of CRES events in the data.

The feasibility of a real-time reconstruction pipeline rests on the development of computationally efficient algorithms that can be implemented without the need for enormous computing resources. One challenge with the antenna array approach is that the small radiation power of a single electron is distributed between each channel in the array, such that reconstruction using only the information in a single channel is not possible. Therefore, the simply performing the initial step in reconstruction — signal detection — requires orders of magnitude more computational power than previous CRES experiments. This operation will then be followed by other, potentially more expensive, reconstruction steps that are required in order to determine the kinetic energy of the electron.

3.5 Pilot-scale Experiments

3.5.1 Choice of Frequency

The optimal CRES frequency for Project 8 is that which can reach our target sensitivity of 40 meV, while minimizing the cost and complexity of the overall experiment. The magnitude of the background magnetic field determines the cyclotron frequency, which affects the entirety of the CRES detection system design, specifying the operating frequency of the CRES experiments is one of the first steps towards developing a full design.

Scaling Laws

The Phase I and II experiments utilized a background magnetic field of 0.959 T provided by an NMR magnet. This magnetic field was selected primarily for convenience, however, the cyclotron frequencies for electrons near the tritium endpoint in a 0.959 T field ranges from 25 to 26 GHz, which is within the standard RF Ka-band. Therefore, microwave electronics specialized for these frequencies are easily obtainable for relatively low cost. The operating frequency for the large-scale experiments must be selected in a more rigorous manner due to the increased scale and complexity of the systems as well as the requirements of the 40 meV neutrino mass science goal.

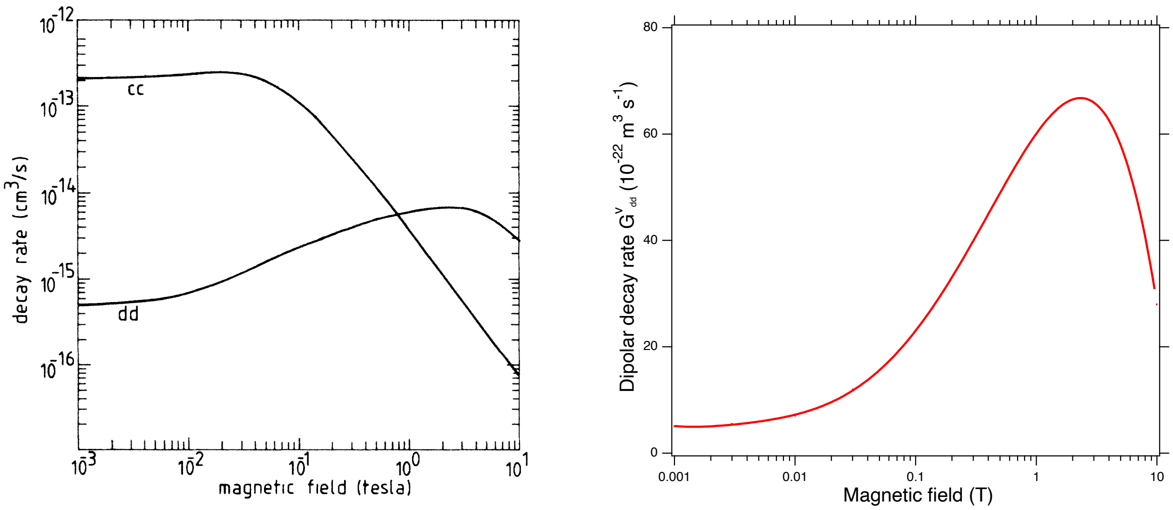
There is a bias towards lower frequencies in a large-volume experiment, due to the direct relationship between wavelength and the physical size of the compatible RF components like antennas and cavities. With a longer wavelength more volume can be surrounded by an array with fewer antennas, which reduces hardware and data-processing costs. Additionally, the size of a cavity experiment is directly proportional to the wavelength since this sets the physical dimensions of the cavity. Furthermore, it is easier to engineer a magnet that provides a uniform magnetic field across several cubic-meters of space at lower magnetic fields, which provides advantages in terms of cost-reduction and field uniformity.

A concern with lower magnetic fields and frequencies is the scaling of the Larmour power equation, which is proportional to the square of the frequency. Naively, one would predict that the SNR would decrease with lower fields, however, two additional scaling laws that affect the noise power also come into play. Noise power is directly proportional to the required bandwidth, which decreases linearly with the magnetic field. Furthermore, at lower frequencies it is possible to purchase amplifiers with lower noise temperatures until approximately 300 MHz at which point this relationship tends to flatten. Therefore, it is expected that the SNR remains approximately constant as the frequency decreases.

The SNR directly impacts the overall efficiency of the experiment through its effects on signal detection and energy resolution. Thus, the expectation that SNR remains the same at lower frequencies clearly biases large-scale experiments in this direction. One drawback of lower magnetic fields is the increased influence of external magnetic fields on the experiment. This includes magnetic fields from the building materials as well as variations in the earth's magnetic field. To deal with these affects a suitable magnetic field correction system will need to be devised, which includes constant monitoring of external fields.

Atomic Tritium Considerations

The pilot-scale experiments will be the first Project 8 experiments to combine CRES with atomic tritium, therefore, the optimal frequency should take into account the affect of the background magnetic field on the atom trap. The primary influence of the background



(a) (b)

Figure 3.18: (a) A plot of the decay rate for the two-body dipolar spin exchange interaction for cc and dd state. (b) A plot of the decay rate of the dipolar spin exchange interaction for d+d states as a function of magnetic field magnitude. Lowering the magnetic field is key for reducing the losses from this interaction.

field magnitude is through the rate of dipolar spin-flips caused by a spin exchange interaction between trapped atoms [56].

Atomic tritium is a simple quantum system with a hyperfine structure given by the addition of the nuclear and atomic spins. The addition of two spins leads to a hyperfine structure with four states in the (m_s, m_I) basis [57]. The states with atomic spins directed anti-parallel to the magnetic field have $m_s = -1/2$ and are labeled as the a and b states. The a and b states are colloquially known as high-field seeking states, since their energy is minimized when in regions of higher magnetic field. This leads to losses in the magnetic trap as these atoms are drawn to higher fields away from the trap center. Alternatively, the c and d states, with atomic spin $m_s = +1/2$, minimize their energy in low magnetic fields because of the parallel alignment between spin and the magnetic field. Therefore, these low-field seeking states tend to stay trapped significantly longer than the high-field seeking states.

It would be advantageous to prepare tritium atoms in purely c and d states before trapping, however, even in this case losses still occur due to dipolar interactions between pairs of c and d states leading to flipped atomic spins and subsequent losses from high-field seeking atoms. The rate of these interactions depends on the magnitude of the background magnetic field and is maximal for dd interactions around 1 T (see Figure 3.18). The rate of losses from these interactions at 1 T requires atomic tritium production at a rate two orders of magnitude larger than at 0.1 T, thus, requirements on the whole atomic tritium system are significantly relaxed at lower magnetic fields, which provides an additional argument for transitioning to lower frequencies with the pilot-scale experiments.

3.5.2 Pilot-scale Experiment Concepts

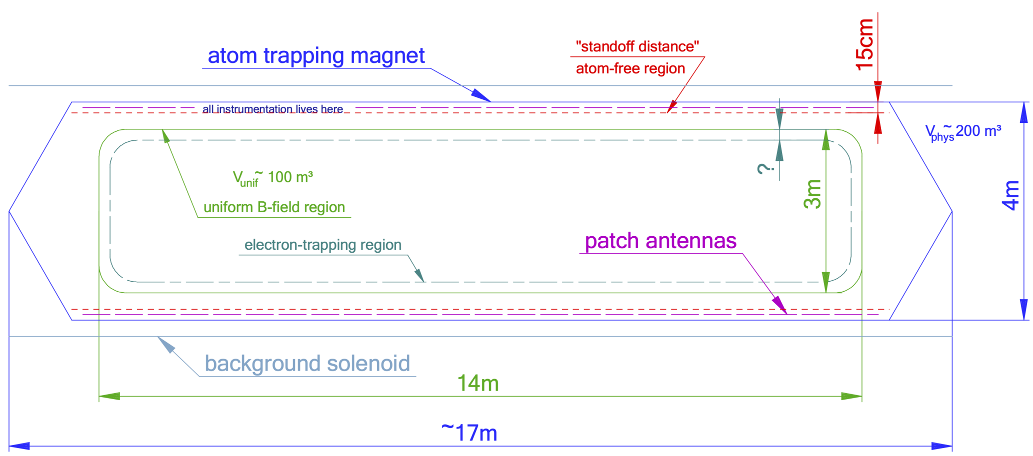


Figure 3.19: A conceptual sketch of a large-volume antenna array based CRES experiment to measure the neutrino mass.

While the pilot-scale experiments are still in the early stages, enough is known to sketch the general features of these experiments at the conceptual level.

Pilot-scale Antenna Array CRES Experiment Concept

A conceptual design for an antenna-based CRES experiment is shown in Figure 3.19. A large solenoid magnet provides a uniform background magnetic field less than 0.1 T in magnitude. Inside this region is the atom trapping magnet that generates a high magnetic field at the walls, which decays exponentially towards the central region. Known magnet designs that produce suitable atom trapping fields include Ioffe-Prichard traps, which use conducting coils, as well as a Halbach array made from permanent magnets. Either magnet choice produces a region of high magnetic fields, which excludes atoms and allows for the placement of antennas inside the experiment.

Inside this region an array of microstrip patch antennas is inserted to collect the cyclotron radiation without providing a surface for atomic tritium recombination. Due to the lower frequency of cyclotron radiation antennas of a larger size can be used, which lowers the total number of antennas required to observe the experiment volume. Because of this scaling, the lower frequency experiment uses a similar number of antennas compared to a much smaller demonstrator experiment with a 1 T magnetic field.

The atomic tritium beamline that supplies fresh tritium atoms to the experiment is not shown in the figure. The general configuration would matches the one shown for the pilot-scale cavity experiment (see Figure 3.20).

Pilot-scale Cavity CRES Experiment Concept

The pilot-scale cavity experiment includes both an atomic tritium system and cavity CRES system. The atomic system consists of a thermal atom cracker located at the start of an evaporatively cooled atomic beamline. The atomic tritium system provides a supply of tritium atoms to the trap with temperatures on the order of a few mK. Atoms at this temperature can be trapped magneto-gravitationally, which is the reason for the vertical orientation of the cavity. At these low magnetic fields the trapping requirements for electrons and atoms differ enough such that it is advantageous to decouple the the trapping potentials to avoid radioactive heating of the tritium atoms from excess trapped electrons. Electron trapping is provided by a set of magnetic pinch coils at the top and bottom of the cavity and a multi-pole Ioffe or Halbach magnet serves to contain the atoms.

The cavity design for the pilot-scale experiment consists of a large cylindrical cavity with a TE011 resonance of 325 MHz. Such a cavity is truly enormous, with a diameter of approximately 1.2 m and a height of 11 m. When an electron is produced inside the cavity with a cyclotron frequency that matches the TE011 resonant frequency it's cyclotron orbit couples the electron to the TE011, which drives a resonance in the cavity. These resonant fields can be read-out using an appropriate cavity coupling mechanism located at the center of the cavity. For more information on the cavity approach to CRES see Chapter 6.

The bottom of the cavity has a cone termination to match the contour of the atom

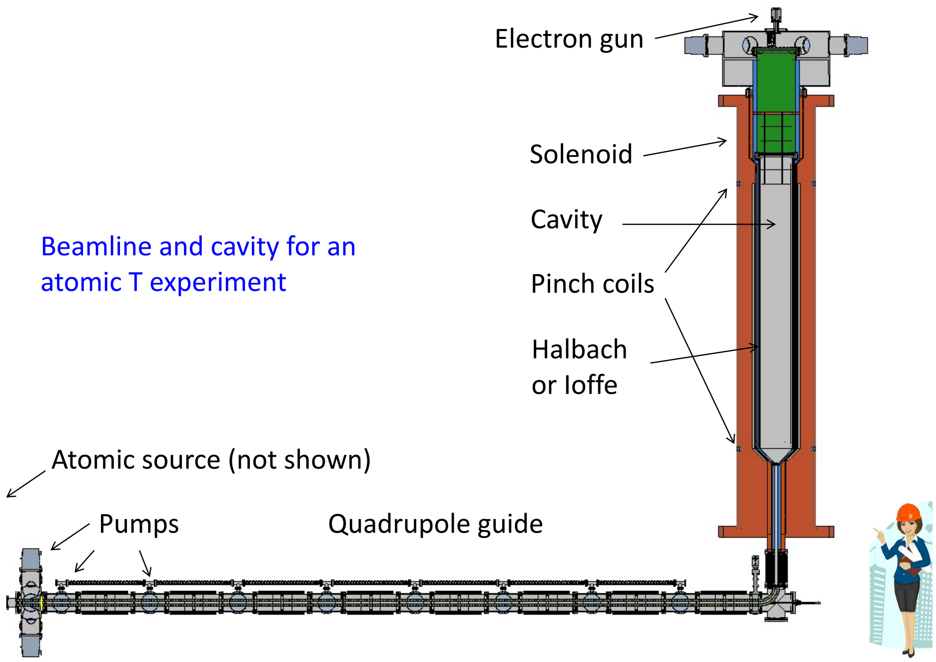


Figure 3.20: A conceptual sketch of a pilot-scale cavity CRES experiment with an atomic tritium beamline.

trapping magnet. This shape still allows for TE011 resonances with high internal Qs, which are required for good SNR in the cavity experiment. A small opening in the bottom of the cone serves as an entry point for the tritium atoms. To allow for calibration of the magnetic field inhomogeneities with an electron gun, the top of the cavity is left nearly completely open. Normally, this would drastically lower the Q-factor of the TE011 mode, but a specially configured coaxial partition is inserted at the top. This termination scheme is designed to act as a perfect short for the TE011 mode since the circular shape of the partition matches the electric field boundary conditions for the TE011 mode. Simulations with HFSS have confirmed that this design results in a high quality TE011 resonance despite the nearly completely open end.

3.6 Phase IV

The baseline CRES technology being pursued by the Project 8 collaboration are resonant cavities, which, due to their geometric properties, simple CRES signal structure, and low channel count, appear to be the better option for Phase IV. The current knowledge of the antenna array CRES approach reveals no technical obstacles that would preclude it as a baseline technology for Phase IV though it would most certainly be significantly more expensive. Therefore, antenna arrays represent a fallback approach if resonant cavities prove infeasible.

The sensitivity of the pilot-scale atomic tritium experiment is estimated to be on the order of 0.1 eV, which means that increasing the sensitivity to reach the Phase IV goal will require an even larger experiment. Because of the direct coupling between the RF characteristics of a cavity and its geometry, the baseline plan is to build multiple copies of the pilot-scale experiment (see Figure 3.21) to obtain the required amount of volume rather than increase the size of the cavity beyond the pilot-scale. The built-in redundancy of this approach is useful in the sense that the experiment has no single point of failure, additionally, building several copies of the a pilot-scale experiment will minimize new engineering and design effort.

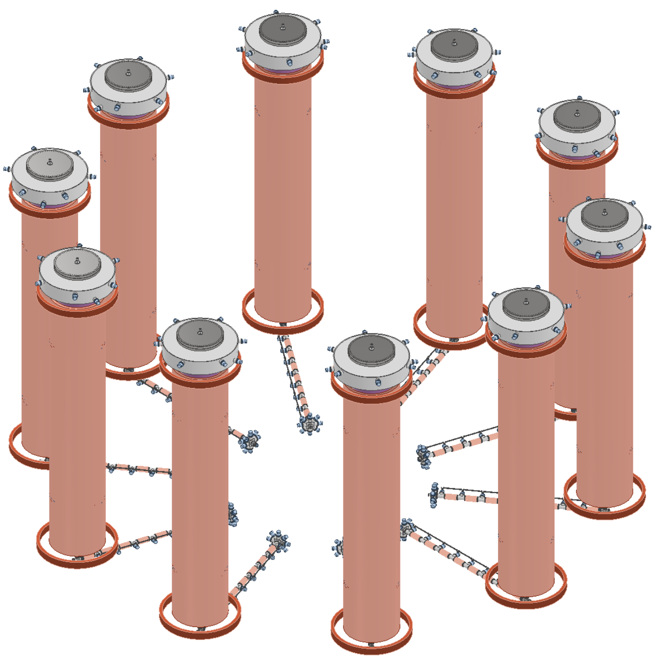


Figure 3.21: An illustration of a possible arrangement of ten pilot-scale cavity experiments for Phase IV. The experiments are arranged in a circle with an approximate diameter of 50 meters. Each atomic beamline connected to the bottom of each cavity is approximately 10 m in length. The cavities themselves are designed to operate at 325 MHz and are approximately 11 m tall. The circular arrangement of cavities has some advantages when it comes to cancellation of fringe fields from neighboring magnets, which is important due to the small magnetic field magnitudes consistent with these CRES frequencies. The advantage of ten independent atomic sources and cavities is that there is no single point of failure for the experiment. If an experiment goes down for repairs the other nine may continue running. Figure courtesy of Michael Huehn at UW-Seattle.

Chapter 4

Signal Reconstruction Techniques for Antenna Array CRES and the FSCD

4.1 Introduction

The transition from a waveguide CRES experiment to an antenna array CRES experiment introduces new challenges related to data acquisition, signal detection, and signal reconstruction caused by the multi-channel nature of the data. The development of signal reconstruction algorithms is crucial to the design of antenna array based experiments like the FSCD, because these algorithms directly influence the detection efficiency and energy resolution of the CRES experiment. In this Chapter I summarize my contributions to the development and analysis of signal reconstruction and detection algorithms for the FSCD experiment.

In Section 4.2 I discuss the primary tool for this work, which is the Locust simulations package developed by the Project 8 experiment. Locust is used to simulate CRES events in the detector. Locust uses Kassiopeia to calculate particle trajectory solutions for electrons in the magnetic trap. The trajectories are then used to calculate the response of the antenna array to the cyclotron radiation produced by the electron, which results in signals that can be used to analyze the performance of different signal reconstruction algorithms. More recently, Project 8 has developed CREsana, which is a new simulations package that takes a more analytical approach to CRES signal simulations for antenna arrays. Although CRESana signals were not used for the signal reconstruction algorithm development detailed here, we introduce the software as it plays a role in the antenna array measurements presented in Section 5.5.

In Section 4.3 I discuss the signal reconstruction and detection approaches analyzed for the FSCD experiment. In general there are two steps to signal reconstruction — detection and parameter estimation. With signal detection one is primarily concerned only with distinguishing between data that contains a signal versus data that contains only noise, whereas, with parameter estimation one is interested in extracting the kinematic parameters of the electron encoded in the cyclotron radiation signal shape. Due to the low signal power of electrons near the spectrum endpoint in the FSCD experiment, signal detection is a non-trivial problem. This is magnified by the need to maximize the detection efficiency of the experiment in order to achieve the neutrino mass sensitivity goals. My contributions to signal reconstruction analysis fof the FSCD are focused on this signal detection component of reconstruction.

After the discussion of various signal detection approaches, in Section 4.4 I present a more detailed analysis of the detection performance of three algorithms, which could be used to signal detection in the FSCD. This section was originally prepared for publication in JINST as a separate paper. The algorithms include a digital beamforming algorithm, a matched filter algorithm, and a neural network algorithm, which I analyze in terms of classification accuracy and estimated computational cost.

4.2 FSCD Simulations

Antenna array CRES and the FSCD requires a combination of different capabilities not often found in a single simulation tool. First of all, accurate calculations of the magneto-static fields produced by current-carrying coils are required in order to accurately model the magnetic trap and background magnets. The resulting magnetic fields must then be used to calculate the exact relativistic trajectory of electrons, which is required in order to calculate the electro-magnetic (EM) fields produced by the acceleration of the electron. Finally, the simulation has to model the interaction of the antenna and RF receiver chain with these EM-fields in order to produce the simulated voltage signals produced by the antenna array during the CRES event. At the time when Project 8 was developing this simulation capability, no single available simulation tool was known to adequately perform this suite of calculations, which prompted the development of custom simulation framework to simulate the FSCD. This simulation framework includes custom simulation tools developed by Project 8 as well as other open-source and proprietary software developed by third-parties.

4.2.1 Kassiopeia

Kassiopeia[1] is a particle tracking and static EM-field solver developed by the KATRIN collaboration for simulations of their spectrometer based on magnetic adiabatic collimation with an electrostatic filter [58]. Due to the measurement technique employed by the KATRIN collaboration, Kassiopeia is not designed to solve for the EM-fields produced by electrons in magnetic fields. However, it does provide efficient solvers for static electric and magnetic fields and charged particle trajectory solvers. Because of this, Project 8 has incorporated parts of Kassiopeia into it's own simulation framework.

Magnetostatic Field Solutions

The solutions to the electric and magnetic fields generated by a static configuration of charges and currents is given by Maxwell's equations in the limit where the time-dependent terms go to zero. In their static form Maxwell's equations [47] are

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{4.1}$$

$$\nabla \times \mathbf{E} = 0 \tag{4.2}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{4.3}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}, \tag{4.4}$$

where we can see that the electric and magnetic fields are now completely decoupled from each other. The solution for the magnetic field in this boundary value problem is given by the Biot-Savart law

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int dr'^3 \frac{\mathbf{J}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|^3}, \tag{4.5}$$

which Kassiopeia uses a variety of numeric integration techniques to solve for a user defined current distribution.

Kassiopeia Simulation of the FSCD Magnetic Trap

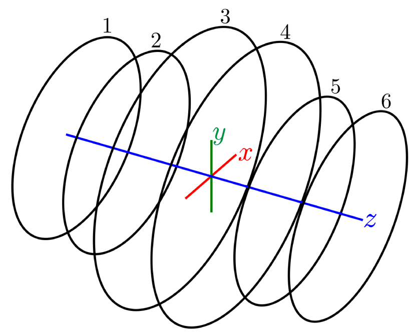
The trap developed for the FSCD experiment utilizes six current carrying coils, which surround a cylindrical tritium containment vessel (see Figure 4.1). Some critical aspects of the trap design include the total trapping volume, the maximum trap depth, the

[1] https://github.com/KATRIN-Experiment/Kassiopeia

steepness of the trap walls, as well as the radial and azimuthal uniformity of the magnetic fields.

The volume of the FSCD trap is a cylindrically shaped region with a radius of 5 cm and a length of 15 cm resulting in a roughly 1 L total trap volume. The trap volume is an important design feature, because it sets the volume of the experiment that is potentially usable for CRES measurements. Trapping a larger volume allows one to observe a larger number of tritium atoms, which increases the statistical power and sensitivity of the neutrino mass measurement. Due to the cost of constructing magnets with large and uniform magnetic fields it is important that the trap use as much of the available volume as possible to limit the overall cost of the experiment.

Coil	Radius (mm)	Z Pos. (mm)	Current (A×Turns)
1	50.0	-92.3	750.0
2	50.1	-56.9	-220.3
3	68.5	-19.5	-250.0
4	68.5	19.5	-250.0
5	50.1	56.9	-220.3
6	50.0	92.3	750.0



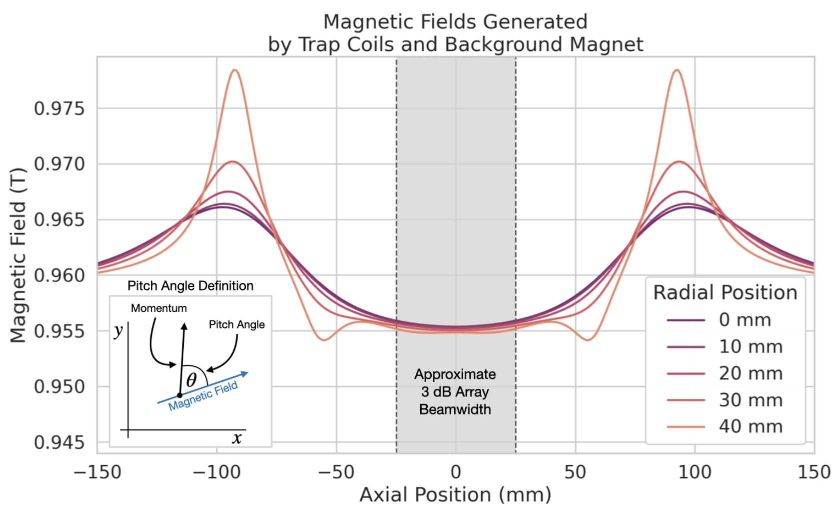


Figure 4.1: The geometry and parameters of the coils used to simulate the FSCD magnetic trap in Kassiopeia. Some axial profiles of the magnetic trap at different radial positions are show to demonstrate the shape of the magnetic field and trap depth as a function of position. Calulation of the magnetic field profiles was graciously done by René Reimann.

The depth of the FSCD trap is approximately 10 mT when measured along the central axis, which is sufficient to trap electrons with pitch angles as small as 84°. The trap depth factors into the efficiency of the experiment by directly controlling the range of electron pitch angles that can be trapped. If a higher fraction of pitch angles are trapped then, in principle, more decay events can be observed. However, the signals from electrons with small pitch angles are typically significantly harder to detect than larger pitch angles when using an antenna array, which increases the likelihood of not detecting the first track of the CRES event and harms the energy resolution of the experiment.

The steepness of the trap walls as well as any non-uniformities in the magnetic field contribute to the total energy resolution of the CRES measurement by causing uncertainty in the relationship between an electron's kinetic energy and it's cyclotron frequency. When an electron is trapped, it oscillates back and forth along the trap z-axis (see Figure 4.1) unless it is produced with a pitch angle of exactly 90° [59]. As the electron is reflected from the trap walls it experiences a change in the total magnetic field, which causes a modulation in the cyclotron frequency. This change in magnetic field from the trap introduces a correlation between the pitch angle and kinetic energy parameters of the electron that can reduce energy resolution. In order to mitigate this effect it is important to make the trap walls as steep as possible.

Particle Trajectory Solutions

The magnetic fields solved by direct integration of the electron's current density can be used by Kassiopeia to solve for the trajectory of electrons based on user specified initial conditions. Various distributions are available within Kassiopeia that can be sampled in order to replicate realistic event statistics, including uniform, Gaussian, and Lorentzian among others. In general, an electron has six kinematic parameters that define its trajectory, which are the three-dimensional coordinates of the initial position and the three components of the electron's momentum vector. However, when simulating CRES events it is more common to parameterize the electron's trajectory in terms of it's initial position, the kinetic energy, the pitch angle, and the initial direction of the component of the electron's momentum perpendicular to the magnetic field. This parameterization is completely equivalent to specify each component of the electrons initial position and momentum vectors.

From the initial parameters of the electron and the magnetic field, Kassiopeia solves for the trajectory of the electron. The direct approach proceeds by solving the motion of the electron using the Lorentz force equation, which takes the form of a set of differential equations

$$\frac{d\mathbf{r}}{dt} = \frac{\mathbf{p}}{\gamma m} \tag{4.6}$$

$$\frac{d\mathbf{p}}{dt} = e(\mathbf{E} + \frac{\mathbf{p} \times \mathbf{B}}{\gamma m}), \tag{4.7}$$

where $\mathbf{r}$ is the position of the electron, $\mathbf{p}$ is the electron's momentum, e is the charge of the electron, m is the electron's mass, and γ is the relativistic Lorentz term. To account for kinetic energy losses from radiation Kassiopeia includes an additional term in the momentum differential equation, which calculates the change in the electron's momentum induced by synchrotron radiation. Kassiopeia solves this pair of differential equations using numerical integration, however, the exact trajectory can be computationally intensive to solve. If the adiabatic approximation can be applied, then Kassiopeia can make use of a simpler set of equations that can be more readily solved numerically.

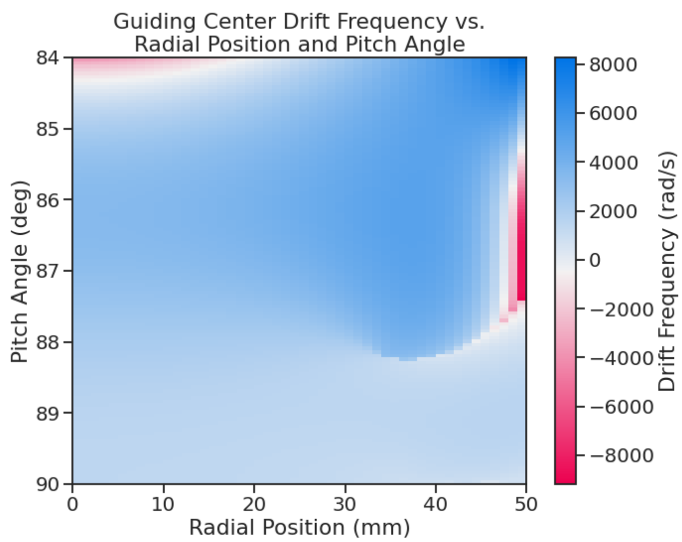


Figure 4.2: A map of the average ∇B-drift frequency for electrons trapped in the prototype FSCD trap shown in Figure 4.1. Negative drift frequencies indicate electrons that are drifting opposite to the standard direction, which means that they are close to escaping the magnetic trap.

Even though Kassiopeia is not directly capable of simulating the cyclotron radiation, it is still an invaluable CRES simulation tool, due to the accurate trajectory solutions for electrons in magnetic traps. With Kassiopeia it is possible to test the efficiency of a particular trap design and analyze features of the electron trajectories that are important to the position, track, and event reconstruction algorithms (see Section 4.3). One example of this for the FSCD is the analysis of the average ∇B-drift frequency as a function of the electrons radial position and pitch angle in the magnetic trap (see Figure 4.2). Radial gradients in the trap cause the guiding center of the electron to drift around the center of the magnetic trap with an average frequency on the order of 10^3 rad/s. This frequency, while slow compared to the length of a typical CRES time-slice, is large enough to cause a significant loss in efficiency of certain signal reconstruction algorithms. Therefore, it is important to model the drift of the electron in the reconstruction algorithm in order to mitigate the effects of this motion on the reconstruction.

4.2.2 Locust

The Locust[2] software package [60] is the primary simulation tool developed and used by the Project 8 collaboration for CRES experiments. Locust simulates the responses of antennas and receiver electronics chain to rapidly time-varying electric fields using a flexible approach that allows one to choose from a variety of electric field sources and antennas. Similarly, one can simulate the receiver chain using a series of modular generators that include standard signal processing operations such as down-mixing and fast Fourier transforms (FFT). Since the primary focus of this chapter is the application of Locust to analyses of the FSCD, we shall describe only the most relevant aspects of the software rather than provide a comprehensive description.

Cyclotron Radiation Field Solutions

Simulating CRES events in the FSCD requires that we calculate the electric fields produced by the acceleration of the electron. In the general case, this can be a complicated question to answer, due to back-reaction forces on the electron from it's own electric fields that occur when the electron is surrounded by conductive material such as a waveguide or cavity. However, in the case of the FSCD it is possible to ignore such effects and approximate the electron as radiating into a free-space environment.

The equations that describe the electromagnetic fields from a relativistic moving point particle are the Liénard-Wiechert field equations [61, 62], which are obtained by differentiating the Liénard-Wiechert potentials. In their full form the Liénard-Wiechert field equations are

$$\boldsymbol{E} = e\left[\frac{\hat{n}-\boldsymbol{\beta}}{\gamma^2(1-\boldsymbol{\beta}\cdot\hat{n})^3|\boldsymbol{R}|^2}\right]_{t_\mathrm{r}} + \frac{e}{c}\left[\frac{\hat{n}\times[(\hat{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}]}{(1-\boldsymbol{\beta}\cdot\hat{n})^3|\boldsymbol{R}|}\right]_{t_\mathrm{r}} \tag{4.8}$$

[2] `https://github.com/project8/locust_mc/tree/master`

$$\boldsymbol{B} = [\hat{n} \times \boldsymbol{E}]_{t_{\mathrm{r}}} , \tag{4.9}$$

where e is the charge of the particle, $\hat{n}$ is the unit vector pointing from the particle to the position where the fields are calculated, $\boldsymbol{\beta}$ and $\dot{\boldsymbol{\beta}}$ are the velocity and acceleration of the particle divided by the speed of light (c), $\boldsymbol{R}$ is the distance from the particle to the field calculation position, and γ is the relativistic Lorentz term. The subscript t_{r} indicates that the equations must be evaluated at the retarded time so that the time-delay from the travel time of the electromagnetic radiation is correctly accounted for.

The only required input to calculate the electric field at the position of an FSCD antenna is the velocity and acceleration of the electron, which can be obtained from Kassiopeia simulations. Therefore, when simulating a CRES event Locust first runs a Kassiopeia simulation of the electron and calculates the electric field incident on the antenna. The only difficulty with this approach is the determination of the retarded time. The retarded time corresponds to the time that a photon, which has just arrived at an antenna at the space-time position $(t, \boldsymbol{r})$, was actually emitted by the electron at the space-time position of $(t_{\mathrm{r}}, \boldsymbol{r}_e(t_{\mathrm{r}}))$. Defined in this way, finding the retarded time requires solving

$$c(t - t_{\mathrm{r}}) = |\boldsymbol{r} - \boldsymbol{r}_e(t_{\mathrm{r}})|, \tag{4.10}$$

where the distance traveled by the photon between the measurement and retarded times is equal to the distance between the antenna and the electron at the retarded time. Locust solves Equation 4.10 using a built-in root finding algorithm to find the retarded time, and thus the electric field produced by the electron at the position of each antenna in the FSCD array.

Antenna Response Modeling

With the electric field it is possible, in principle, to calculate the resulting voltages produced in the antenna. However, direct simulation of the antenna itself is computationally expensive since it would require the modeling of complex interactions of the electron's electric fields with charge carriers in the conductive elements of the antenna. Direct simulation of the antenna in Locust can be avoided by modeling the antenna response using the antenna factor, or antenna transfer function, approach. The antenna factor defines the voltage produced in the antenna terminal for an incident electric field [63],

$$A_{\mathrm{F}} = \frac{V}{|\boldsymbol{E}|}, \tag{4.11}$$

where V is the voltage and $|\boldsymbol{E}|$ is the magnitude of the incident electric field. To obtain the antenna factor for the antennas developed for the FSCD Project 8 employs Ansys HFSS. HFSS is a commercially available finite element method electromagnetic solver widely used throughout the antenna engineering industry [64]. HFSS is capable of calculating the antenna factor and gain patterns for complex antenna designs and outputting the resulting quantities in the form of a text file that can be used as an input to the Locust simulation.

The antenna factor defines the steady-state response of the antenna to electromagnetic plane waves and is a function of the frequency of the radiation. Therefore, in order to apply the transfer function for the calculation of the antenna voltage response in the time domain, Locust models the antenna as a linear time-invariant system [65]. In this formalism the response of the system to the driving force is given by

$$y[n] = h * x = \sum_k h[k]x[n-k], \tag{4.12}$$

where $y[n]$ is the discretely sampled response, x is the driving force stimulus, and h is the finite impulse response (FIR) filter. When applied to the FSCD array, this formalism calculates the voltage time-series produced in each antenna by convolving the electric field time-series with the antenna FIR filter, which is obtained by performing a inverse Fourier transform on the transfer function from HFSS.

Radio-frequency Receiver and Signal Processing

After obtaining the voltage time-series by computing the electron trajectory and antenna response, Locust simulates the signal processing associated with the radio-frequency receiver chain. The standard receiver chain used in Locust simulations of the FSCD attempts to mimic the operations that would actually occur in hardware (see Figure 4.3).

Frequency down-conversion is used in the FSCD to reduce the digitization bandwidth required to read-out CRES data. According to the Nyquist sampling theorem [66], the minimal sampling rate that guarantees no information loss for a signal with a bandwidth Δf is given by

$$f_{\mathrm{Nyq}} = 2\Delta f. \tag{4.13}$$

The total bandwidth of CRES signal frequencies from tritium beta-decay ranges from 0 to 26 GHz in a 0.95 T magnetic field, therefore, direct digitization of CRES signals from the FSCD would require sampling frequencies greater than 50 GHz, which is infeasible for

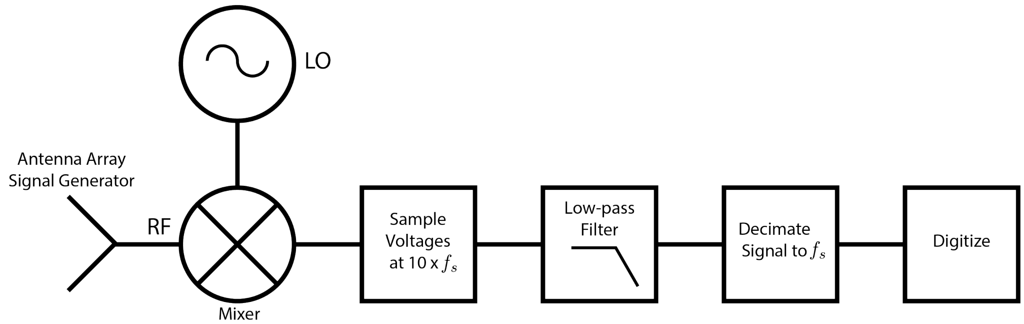


Figure 4.3: The receiver chain used by Locust when simulating CRES events in the FSCD.

a real experiment. However, for the purposes of neutrino mass measurement we are only interested in measuring the shape of the spectrum in the last 100 eV, which corresponds to a frequency bandwidth of 5 MHz. Down-conversion is a technique for reducing the base frequencies of signals in a bandwidth given by $[f_{\rm LO}, f_{\rm LO} + \Delta f]$ to the bandwidth $[0, \Delta f]$, by performing the following multiplication

$$x(t) \to x(t)e^{-2\pi f_{\rm LO}}. \tag{4.14}$$

In down-conversion the signal $(x(t))$ is multiplied by a sinusoidal signal with frequency $f_{\rm LO}$ to reduce the absolute frequencies of the signals in the bandwidth. In the FSCD this allows us to detect events in the last 100 eV of the tritium spectrum while sampling the data far below 50 GHz. The standard bandwidth used in the FSCD is 200 MHz, which allows for higher frequency resolution than the minimum sampling frequency for 100 eV of energy bandwidth.

Trying to directly simulate down-conversion with a frequency multiplication in Locust would require the sampling of the electric fields at each antenna in the FSCD array with a period of $\approx$ 20 ps, which is extremely slow computationally. To avoid this Locust performs the down-conversion by intentionally under-sampling the electric fields with a frequency of 2 GHz. Sampling below the Nyquist limit causes the higher frequency components of the CRES signal to alias, however, Locust can remove these aliased frequency peaks using a combination of low-pass filtering and decimation to recreate frequency down-conversion. After filtering and decimation, Locust simulates digitization by an 8-bit digitizer at a sampling frequency of 200 MHz to recreate the conditions of the FSCD. The voltage offset and the digitizer range must be configured by the user based on the characteristics of the simulation.

Data

The output of Locust simulations for the FSCD primarily consists of two data files. The first is the electron trajectory information calculated by Kassiopiea, which is output in the form of a `.root` file [67]. This file contains important kinematic information about the electron such as it's position and pitch angle as a function of time. The other file is produced by Locust and it contains the digitized signals acquired from each antenna in the FSCD array. The Locust output files conform to the Monarch specification developed by Project 8, which is based on the commonly used HDF5 file format, and matches the format of the files produced by the Project 8 data acquisition software. This makes it possible to use the same data analysis code to analyze both simulated and real data.

4.2.3 CRESana

Locust is the primary simulation tool used by Project 8 in the development and simulation of the FSCD. However, simulations of CRES events in larger antenna arrays (≥ 100 antennas) using Locust can take several hours to complete, which is prohibitively long when one is performing a sensitivity analysis for a large scale antenna experiment. One of the reasons for Locust's slow operation is that the electric fields from the electron must be solved numerically for each time-step for each of the antennas in the array. These numerical solutions allow Locust to accurately simulate the electric fields from arbitrarily complicated electron trajectories at the cost of more computations and slower simulations. Therefore, an additional simulation tool that sacrifices some accuracy for computational efficiency would be extremely useful simulations and sensitivity analyses of larger antenna array experiments.

To fill this need, Project has developed a new simulations package called CRESana[3], specifically designed to perform analytical simulations of antenna array based CRES experiments. CRESana is not as flexible as Locust, but it provides a significant increase in simulation speed. It does this by using well-justified analytical approximations of the electrons motion in the magnetic field and the resulting electric fields from the electron's acceleration. The electric fields and signals generated by CRESana are consistent with theoretical calculations of the electron's radiation, and are test for accuracy using well-known test-case simulations and consistency checks.

[3] `https://github.com/MCFlowMace/CRESana`

4.3 Signal Detection and Reconstruction Techniques for Antenna Array CRES

Antenna Array CRES Signal Reconstruction

A robust set of FSCD simulation tools are vital to the development of the analysis algorithms necessary for antenna array CRES to succeed. In order to perform CRES measurements using an antenna array, one must develop an algorithm that uses the multi-channel time-series obtained by digitizing the array to estimate the starting kinetic energies of electrons produced in the magnetic trap. This procedure consists of a multi-stage process of detecting a CRES signal then estimating the parameters of the electron that produced and is often referred to as simply CRES signal reconstruction.

Compared with the signal reconstruction approaches of the Phase I and II CRES experiments, antenna array CRES requires a significantly different approach to signal reconstruction. In Phase I and II, CRES was performed using a waveguide gas cell that could be directly connected to a waveguide transmission line. The transmission line efficiently transmits the cyclotron radiation along it's length to an antenna at either end of the waveguide. However, with an antenna array the electron is essentially radiating into free-space, therefore, the cyclotron radiation power collected by the array is directly proportional to the solid angle surrounding the electron that is covered with antennas. Because it is not practical to fully surround the magnetic trap with antennas, some of the cyclotron radiation power that would have been collected by the waveguide escapes into free-space. Furthermore, the power that is collected by the antenna array is split between every channel in the antenna array, which significantly lowers the signal-to-noise ratio (SNR) of CRES signals in a single antenna channel compared to a waveguide apparatus. Therefore, a suite of completely new signal reconstruction techniques are needed in order to perform CRES in the FSCD.

Changes to the approach to CRES signal reconstruction are also motivated by the more ambitious scientific goals of the FSCD experiment. A measurement of the tritium beta-decay spectrum that is sensitive to neutrino masses as small as 40 meV requires that we measure the kinetic energies of individual electrons with a total energy broadening of 115 meV [68]. This resolution includes all sources of uncertainty in the electron's kinetic energy such as magnetic field inhomogeneities. This level of energy resolution is compatible only with an event-by-event signal reconstruction approach where the kinetic energies, pitch angles, and other parameters of the CRES events are estimated before constructing the beta-decay spectrum.

The event-by-event approach is distinct from the analysis done for the Phase I and Phase II experiments where only the starting cyclotron frequency of the event was estimated by analyzing the tracks formed by the carrier frequency in the time-frequency spectrogram. These frequencies were then combined into a frequency spectrogram, which was converted to the beta-decay energy spectrum using an ensemble approach that averaged over all other event parameters. The ensemble approach to signal reconstruction results in poor energy resolution because other kinematic parameters such as pitch angle change the cyclotron carrier frequency due to changes in the average magnetic field experience by the electron, and it is therefore incompatible with the future goals of the Project 8 collaboration.

Components of Reconstruction: Signal Detection and Parameter Estimation

CRES signal reconstruction can be viewed as a two-step procedure consisting of signal detection followed by parameter estimation. In the former, one is concerned with identifying CRES signals in the data regardless of the signal parameters, whereas, in the latter one operates under the assumption that a signal is present and then estimates it's parameters.

More formally, signal detection is essentially a binary hypothesis test between the signal and noise data classes and parameter estimation describes a procedure of fitting a model to the observed data. While both of these processes are required for a complete reconstruction (see Figure 4.4), the focus of my work and this chapter is on the signal detection aspect of antenna array CRES signal reconstruction.

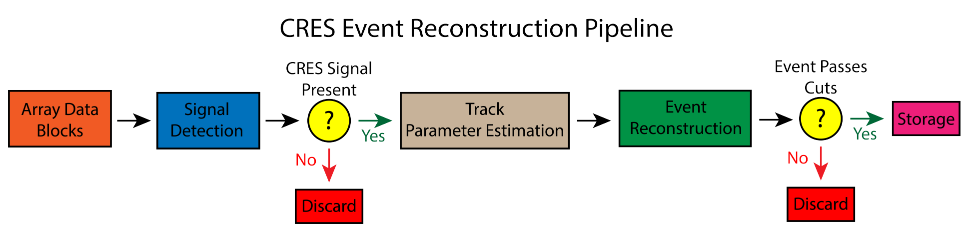


Figure 4.4: A high-level diagram depicting the process of CRES event reconstruction. The first step consists of identifying the presence of a signal in the data. This step is necessary to avoid the danger of performing a reconstruction of a false event, which would constitute a background contribution to the tritium spectrum measured by CRES.

Detection Theory

The problem of signal detection can be posed as a statistical hypothesis test [69]. For CRES signals, which are essentially vectors with added white Gaussian noise (WGN), one needs to choose between two hypotheses

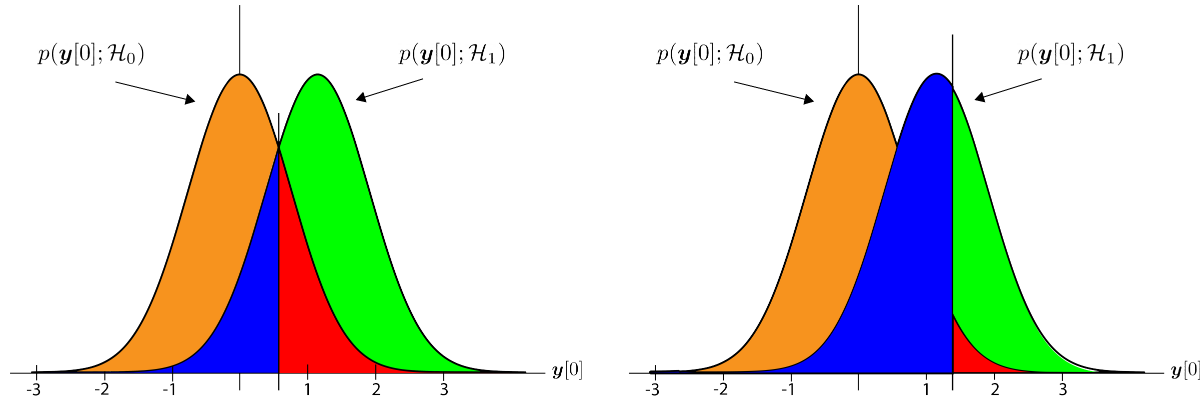
$$\mathcal{H}_0 : \quad \boldsymbol{y} = \boldsymbol{\nu} \tag{4.15}$$

$$\mathcal{H}_1 : \quad \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\nu}, \tag{4.16}$$

where $\boldsymbol{y}$ is the CRES data vector, $\boldsymbol{\nu}$ is a sample of WGN, and $\boldsymbol{x}$ represents the CRES signal. The hypothesis that the data contains only noise is labeled $\mathcal{H}_0$ and the hypothesis that the data contains a signal is labeled $\mathcal{H}_1$.

For illustrative purposes one can examine the case where one the first sample of data is used to distinguish between $\mathcal{H}_0$ and $\mathcal{H}_1$. The value of the first data sample is distributed according to two gaussian distributions corresponding to $\mathcal{H}_0$ and $\mathcal{H}_1$ (see Figure 4.5). By setting a decision threshold on the value of this sample, one can choose the correct hypothesis with a probability given by the areas underneath the probability distributution curves. A true positive corresponds to correctly identifying that the data contains signal, whereas, a true negative means that one has correctly identified the data as noise. The rate at which the detector performs a true positive classification is given by the green region underneath $p(\boldsymbol{y}[0]; \mathcal{H}_0)$, and the rate at which the detector performs a true negative classification is given by the orange region underneath $p(\boldsymbol{y}[0]; \mathcal{H}_1)$. Two types of misclassifications are possible. Either we declare noise data as signal, which is call a false positive, or we declare signal data as noise, which is a false negative. Note that it is only possible to trade off these two types of errors by tuning the detection threshold. One cannot simultaneously reduce the rate of false positives without also increasing the rate of false negatives.

The approach taken with CRES signals is to fix the rate of false positives by setting a minimum value for a detection threshold. The rate of false positives that is acceptable at the detection stage depends upon the rate of background events compatible with the sensitivity goals of the experiment. The ultimate goal of a neutrino mass measurement with 40 meV sensitivity in general has strict requirements on the number of background events, which requires a relatively high detection threshold to achieve. Consequently, the ideal signal detection algorithm is the one that achieves the maximum rate of true positives for a fixed rate of false positives, so that the detection efficiency of the experiment is maximized and potential sources of background are kept to a minimum.



(a) (b)

Figure 4.5: An illustration of two PDFs associated a binary hypothesis test. The decision threshold is represented by the vertical line that partitions both distributions. The orange and red areas correspond to the true negative and false positive probabilities and the blue and green areas correspond to the false negative and true positive probabilities respectively. To decided between the two hypotheses we perform the likelihood ratio test specified by the Neyman-Pearson theorem. This approach achieves the highest true positive probability for a given false positive probability.

According to the Neyman-Pearson theorem [70], the statistical hypothesis test that maximizes the probability of detection for a fixed rate of false positives is the likelihood ratio test, which is formed by computing the ratio of the signal likelihood to the noise likelihood,

$$L(x) = \frac{P(\boldsymbol{y};\mathcal{H}_1)}{P(\boldsymbol{y};\mathcal{H}_0)} > \gamma. \tag{4.17}$$

Here, the likelihood of the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are described by the probability distributions $P(\boldsymbol{y};\mathcal{H}_0)$ and $P(\boldsymbol{y};\mathcal{H}_1)$ respectively, and γ is the threshold for deciding $\mathcal{H}_1$. The decision threshold is determined by integrating $P(\boldsymbol{y};\mathcal{H}_0)$ such that

$$P_{\mathrm{FP}} = \int_{\gamma}^{\infty} P(\tilde{\boldsymbol{y}};\mathcal{H}_0)d\tilde{\boldsymbol{y}} = \alpha, \tag{4.18}$$

where α is the desired false positive detection rate given by the red colored areas shown in Figure 4.5. The true positive detection rate is given by the similar integral

$$P_{\mathrm{TP}} = \int_{\gamma}^{\infty} P(\tilde{\boldsymbol{y}};\mathcal{H}_1)d\tilde{\boldsymbol{y}}, \tag{4.19}$$

which corresponds to the green areas in Figure 4.5.

Changing the decision threshold allows one to trade-off between P_{TP} and $P_{textrmFP}$ as appropriate for the given situation. It is common to summarize the relationship between P_{TP} and P_{FP} using the receiver operating characteristic (ROC) curve, which is obtained by evaluating the true positive and false positive probabilities as a function of the decision threshold value (see Figure 4.6). The ROC curve provides a convenient way to compare

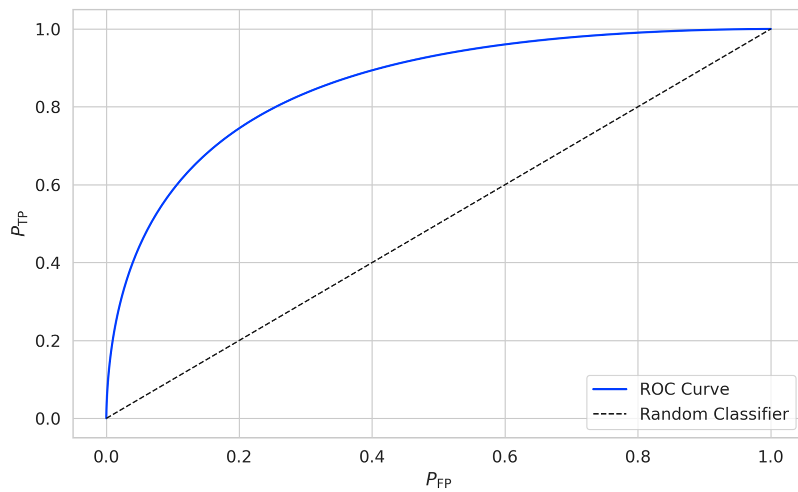


Figure 4.6: An example ROC curve formed by computing the P_{FP} and the P_{TP} for a given likelihood ratio test. As the decision threshold is increased P_{FP} decreases at the expense of a lower P_{TP}. The black dashed line indicates the lower bound ROC curve obtained by randomly deciding between $\mathcal{H}_0$ and $\mathcal{H}_1$.

the performance of different signal detection algorithms. In general, a classifier with a higher the P_{TP} as a function of P_{FP} is desirable, which corresponds to a larger area underneath the respective ROC curve. A perfect classifier has an area underneath the curve of 1.0, however, such a classifier is almost never achievable in practice.

4.3.1 Digital Beamforming

Introduction to Beamforming

Beamforming refers to a suite of antenna array signal processing techniques that are designed to enhance the radiation or gain of the array in certain directions and suppress it in other direction [63]. Beamforming is of interest to Project 8 as a first level of signal reconstruction for the FSCD and other antenna array CRES experiments, which operates at the signal detection stage of reconstruction.

Beamforming is accomplished by performing a phased summation of the signals received by the antenna array. The beamforming phases are chosen such that the signals emitted by the array will constructively interfere at the point of interest (see Figure 4.7). As a consequence of the principle of reciprocity [71], when the array is operating in receive mode, the signals emitted from a source at the same point will constructively interfere when summed. The origin of the phase delays in beamforming is the path-

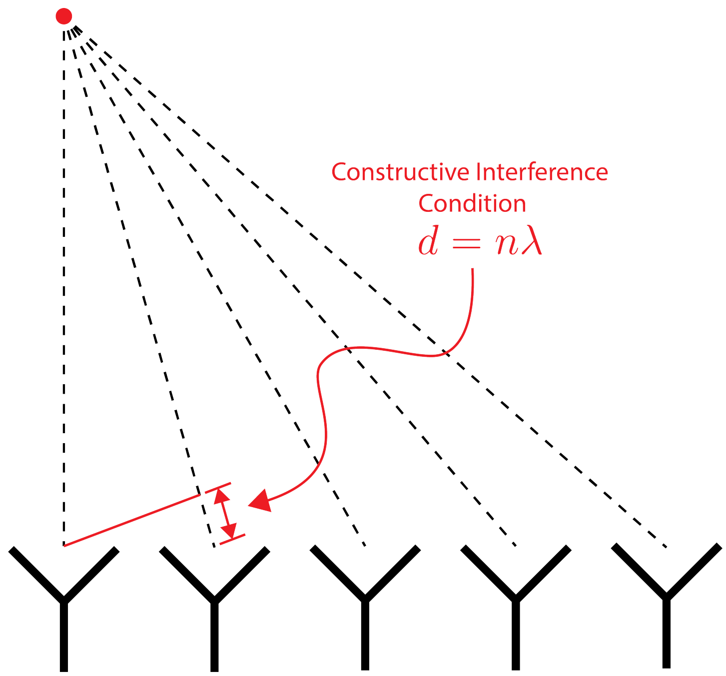


Figure 4.7: An illustration of the constructive interference condition which is the operating principle of digital beamforming using a uniform linear array as an example.

length difference to the beamforming point between different antennas in the array. The relationship between the phase delay and the path-length difference is given by the familiar equation

$$\phi = \frac{2\pi d}{\lambda}, \tag{4.20}$$

where ϕ is the phase delay, d is the path-length difference, and λ is the wavelength of the radiation. In practice, one chooses the values of d by specifying the beamforming positions of interest and then calculates the beamforming phases using Equation 4.20, which is guaranteed to follow the constructive interference condition shown in Figure 4.7.

Beamforming can be neatly expressed mathematically using the vector equation

$$y[n] = \mathbf{\Phi}^T[n]\boldsymbol{x}[n], \tag{4.21}$$

where $\boldsymbol{x}[n]$ is the array snapshot vector, $\mathbf{\Phi}[n]$ is a vector of beamforming shifts, and $y[n]$ is the resulting summed signal. The beamforming shifts consist of a set of complex numbers that contain the beamforming phase shift and an amplitude weighting factor,

$$\mathbf{\Phi}[n] = \left[A_0[n]e^{-2\pi i\phi_0[n]}, A_1[n]e^{-2\pi i\phi_1[n]}, ..., A_{N-1}[n]e^{-2\pi i\phi_{N-1}[n]}\right], \tag{4.22}$$

where the set of magnitudes $A_i[n]$ are amplitude weighting factors and $\phi_i[n]$ are the phase shifts from the path-length differences. The index i is used to denote the antenna channel number. The amplitude weighting factor is the relative magnitude of the signal received by a particular antenna to the other antennas in the array, such that the antennas that receive signals with higher amplitude, due to being closer to the source, have more weight in the beamforming summation. The input and outputs signals beamforming are naturally expected to be functions of time as indicated by the index $[n]$, however, it is also possible to use time dependent beamforming phases that shift the beamforming position of the array over time.

Digital beamforming is the type of beamforming algorithm of interest to Project 8 for CRES. Specifically, digital beamforming means that the beamforming phases are applied to the array signals in software rather than employing fixed beamforming phase shifts in the receiver chain hardware. The advantage of digital beamforming is that for a given series of array snapshots one can specify a large number of beamforming positions and effectively search for electrons by performing the beamforming summation associated with each point and applying a signal detection algorithm to identify the presence of a CRES signal.

One of the most attractive features of digital beamforming is the spatial filtering effect, which is a direct consequence of the constructive interference condition used to define the beamforming phases. Spatial filtering allows for signals from multiple electrons at different positions in the trap to be effectively separated, because the constructive interference condition will force the signals from electrons at positions different from the beamforming position to cancel. This helps to reduce signal pile-up that could become an issue for large scale CRES experiments using a dense tritium source.

The digital beamforming positions can be specified with arbitrary densities limited only by the available computational resources. This provides a very straight-forward way to estimate the position of the electron in the trap by using a dense grid of beamforming positions and maximizing the output power of the beamforming summation over this grid. This natural approach to position reconstruction is attractive due the requirements of an event-by-event signal reconstruction, which needs an accurate estimation of the exact magnetic field experienced by the electron in order to correctly estimate it's kinetic energy. Combined with an accurate map of the magnetic field inhomogeneities of the trap obtained from calibrations, beamforming allows one to apply this magnetic field correction with a spatial resolution that is a fraction of the cyclotron wavelength.

Laboratory Beamforming Demonstrations

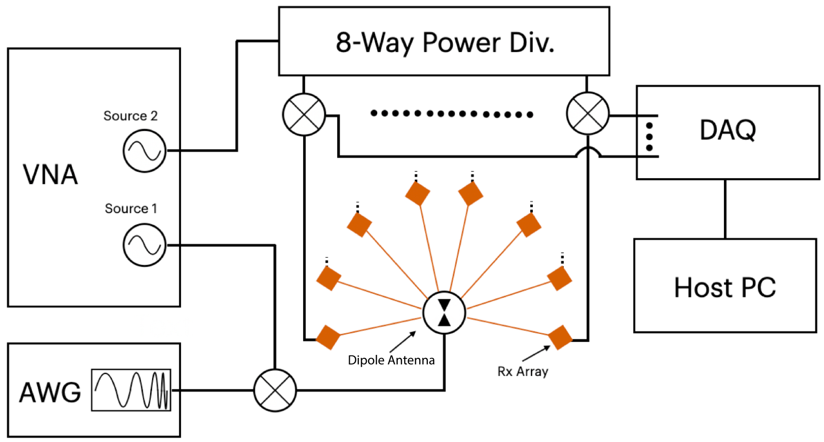


Figure 4.8: System level diagram of the laboratory setup used for beamforming demonstrations at Penn State. For more information on this system see Chapter 5. Signals near 26 GHz are fed to a dipole antenna using and arbitrary waveform generator (AWG) and vector network analyzer (VNA), which drive a mixer. The dipole radiation is collected by an array of antennas connected to the digitizer data acquisition (DAQ) system.

As part of the development of antenna array CRES for the FSCD, an antenna measurement setup was constructed at Penn State to serve as a testbed for antenna prototypes and to perform laboratory validations of array simulations. This system is discussed in more detail in Chapter 5. Early versions of the antenna measurement system (see Figure 4.8 and Figure 4.9) were used to perform beamforming reconstruction studies of a simple probe antenna to better understand the principles of beamforming and confirm the estimated beamforming performance of Locust.

Signals from an arbitrary waveform generator were up-converted to 26 GHz using a mixer and a high-frequency source from a vector network analyzer and fed to the dipole antenna through a balun. The radiation from the dipole antenna was received by an array of horn antennas. The signals from the horn antennas were then down-converted to baseband using a collection of mixers and an 8-way power divider. The signals were then digitized and saved to a host computer for analysis.

The data collected using the dipole and horn antenna array is reconstructed using the

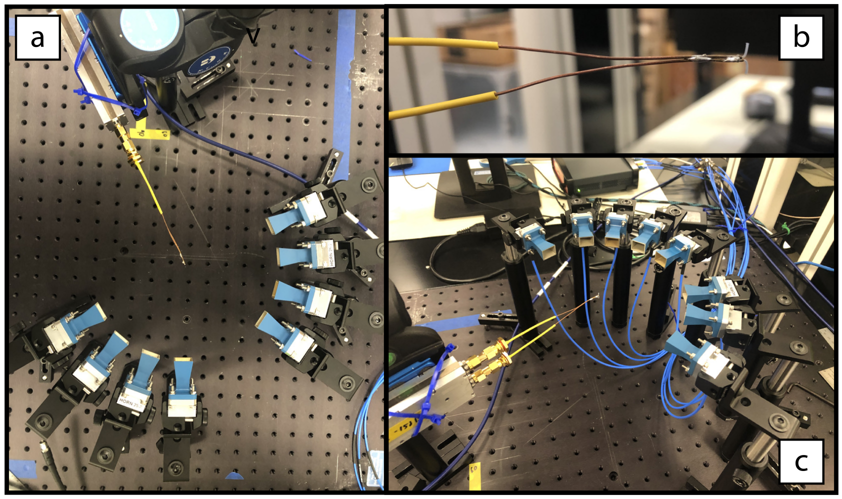


Figure 4.9: Photographs of the beamforming demonstration setup. In (a) I show a top-down view of the dipole antenna and the array of eight horn antennas. Manual repositioning of the horn antennas allows one to synthesize a full-circular antenna array. The dipole antenna is mounted on a camera tripod mount that allows for manual position tuning. (b) is a close up image of the dipole, which is manufactured from two segments of semi-rigid coaxial cable. (c) is another image of the dipole and array.

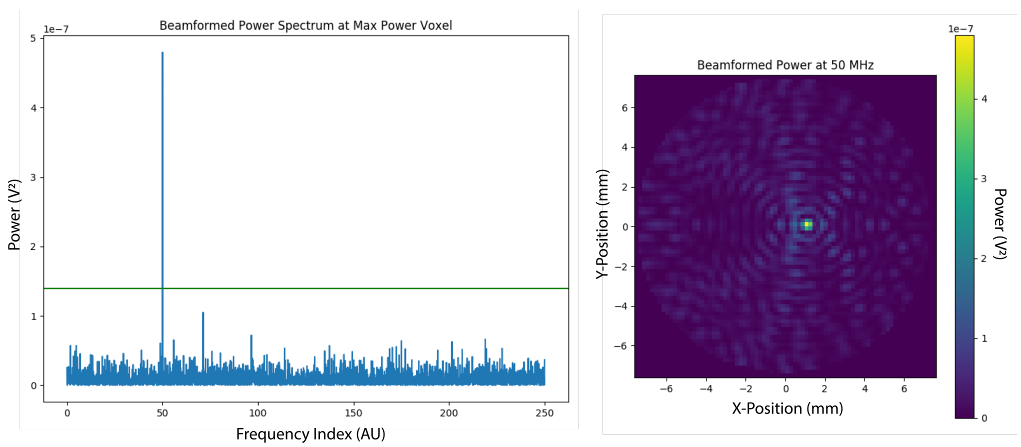


Figure 4.10: An example of digital beamforming reconstruction of a dipole antenna using a synthetic array of horn antennas. The beamforming image on the right is constructed by computing the time-averaged power of the summed signals for a two-dimensional grid of beamforming positions. In the image one can see a clear maximum that corresponds to the position of the dipole antenna. On the left I show the frequency spectrum of the time-series at the maximum power pixel. White gaussian noise is added to the signal to mimic a more realistic signal-to-noise-ratio. The signal emitted by the dipole is clearly visible as the high power peak in the frequency spectrum.

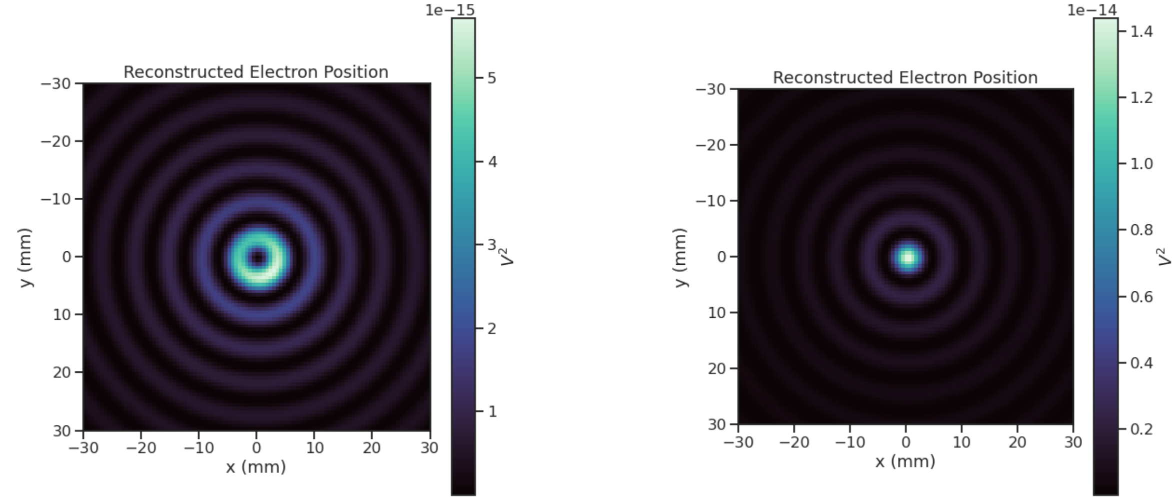
beamforming reconstruction approach specified in Section 4.3.1. A two-dimensional grid of xy-positions is defined and the beamforming phase shifts for each of these positions is calculated. The phased summation can be visualized by plotting the time-averaged power for each of the summations as a pixel in the resulting beamforming image (see Figure 4.10). White Gaussian noise (WGN) can be added to the data at this stage to simulate more realistic signal-to-noise ratios (SNR) if desired. The beamforming peak maxima is expected to have a Bessel function shape due to the circular symmetry of the array, and by analyzing the size of the beamforming maxima one can confirm that the beamforming reconstruction measurement has similar position resolution as expected from Locust simulations. Additionally, signal detection rates can be estimated from the data by comparing the magnitude of the beamforming signal peak in the frequency spectra to simulation.

FSCD Beamforming Simulations

Using Locust simulations of the FSCD one can perform beamforming reconstruction studies using the simulated CRES signal data. As we mentioned in the previous section, the beamforming procedure beings by specifying a set of beamforming positions and corresponding beamforming shifts. The beamforming positions form a grid that covers the region of interest in the field of view of the antenna array. There are effectively an infinite number of ways to specify the grid positions, however, uniform square grids are the most commonly used due to their simplicity. In the FSCD experiment the number and pattern of the grid positions would be optimized to cover the most important regions of the trap volume to maximize detection efficiency while minimizing superfluous calculations.

The beamforming grids used for signal reconstruction with the FSCD consist of a set of points that cover a region of the two-dimensional plane formed by the perimeter of the antenna array. The axial dimension is left out of the beamforming grid because the electrons are assumed to occupy only an average axial position, which corresponds to the center of the magnetic trap. This is because it is impossible to resolve the axial position of the electron as a function of time due to the rapid axial oscillation frequencies of trapped electrons relative to the FSCD time-slice duration.

After beamforming, a summed time-series is obtained for each beamforming position that can be evaluated for the presence of a signal using a detection algorithm. A beamforming image is a visualization method that is equivalent to arranging the beamforming grid points according to their physical locations to form a three-dimensional matrix where the first two dimensions encode the XY-position of the beamforming point and



(a) (b)

Figure 4.11: Beamforming images visualizing the reconstruction of an electron without (a) and with (b) the cyclotron phase correction. The images were generated using data from Locust simulations. The cyclotron phase refers to a phase offset equal to the relative azimuthal position of an antenna in the array. This phase offset is caused by the circular electron orbit and must be corrected for during reconstruction.

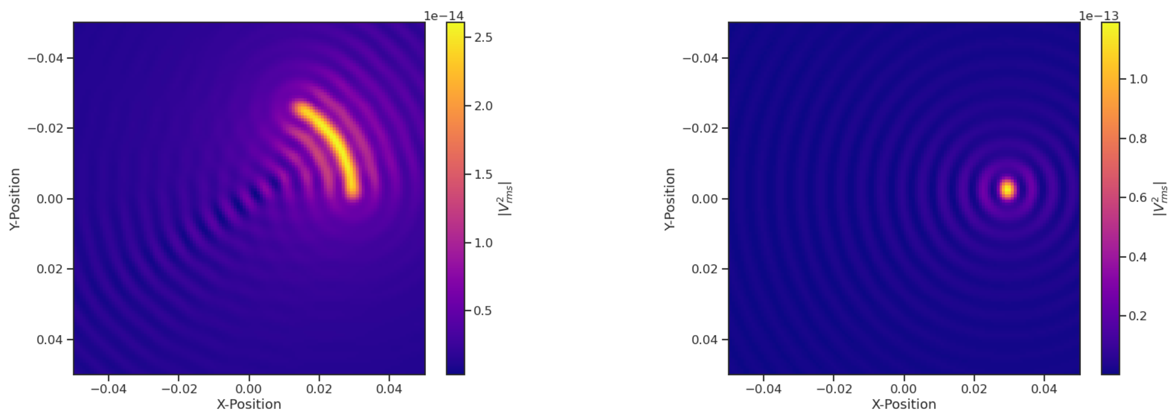
the third dimension contains the summed time-series. The image is formed by taking the time-averaged power (see Figure 4.11). Beamforming images are purely for the purposes of visualization and are not particularly useful for signal detection or reconstruction.

If the beamforming phases consist only of the spatial phase component from Equation 4.20, then the resulting beamforming image contains a relatively high-power ring-shaped region that is centered on the position of the electron (see Figure 4.11a). The origin of this shape is an additional phase offset particular to a cyclotron radiation source. Essentially, the circular motion that produces the cyclotron radiation introduces a relative phase offset to the electric fields that is equal to the azimuthal position of the field measurement point. For example, if we have two antennas, one located at an azimuthal position of 0° and another located at an azimuthal position of 90°, then the CRES signals received by these antennas will be out of phase by 90°, which is the difference in their azimuthal positions. This phase offset can be corrected by adding an additional term to the beamforming phase equation that is equal to the azimuthal position of the antenna relative to the electron,

$$\phi_i[n] = \frac{2\pi d_i[n]}{\lambda} + \Delta\varphi_i[n], \tag{4.23}$$

where $\Delta\varphi_i$ is difference between the azimuthal position of the electron and the i-th antenna channel. Using the updated beamforming phases in the summation changes the ring feature into a Bessel function peak whose maximum corresponds to the position of the electron. Including this cyclotron phase correction significantly improves the signal detection and reconstruction capabilities of beamforming by more than doubling the summed signal power and shrinking the beamforming maxima feature size.

The beamforming image examples in Figure 4.11 were produced using an electron located on the central axis of the magnetic trap, which do not experience ∇B-drift. However, for electrons produced at non-zero radial position the beamforming phases must be made time-dependent in order to track the position of the electron's guiding center over time. Without this correction the ∇B-drift causes the electron to move between beamforming positions, which effectively spreads the cyclotron radiation power over a wider area in the beamforming image (see Figure 4.12). This effect significantly



(a) (b)

Figure 4.12: Beamforming images visualizing the reconstruction of an electron located off the central axis of the FSCD trap. In (a) we performing beamforming without the ∇B-drift correction, and in (b) we include the ∇B-drift correction.

reduces the power of the beamforming maxima and increases the size of the beamforming features, simultaneously harming detection efficiency and position reconstruction.

The ∇B-drift correction simply adds a circular time-dependence to the beamforming positions as a function of time,

$$r[n] = r_0 \tag{4.24}$$

$$\varphi[n] = \varphi_0 + \omega_{\nabla B} t[n], \tag{4.25}$$

where $\omega_{\nabla B}$ is the drift frequency and $t[n]$ is the time vector. In the ideal case the ∇B-drift frequencies from Figure 4.2 for the correct pitch angle and radial position would be used, however, it is not possible to know the electron's pitch angle a priori. In principle, one could perform multiple beamforming summations for a given beamforming position using different drift frequencies and choose the one that maximizes the summed power, but this approach leads to a huge computational burden that would be impractical for a real FSCD experiment. A compromise is to use an average value of $\omega_{\nabla B}$ obtained by averaging over the drift frequencies for electrons of different pitch angle at a particular radius. This approach keeps the computational cost of time-dependent beamforming to a minimum while still providing a significant increase in the detection efficiency of digital beamforming.

Signal Detection with Beamforming and a Power Threshold

Up to this point we have neglected any specific discussion of how digital beamforming is used for signal detection and reconstruction. This is because, strictly speaking, digital beamforming consists only of the phased summation of the array signals and cannot be used alone for signal detection. The example beamforming images shown in Figure 4.11 and Figure 4.12 were produced using simulated data that contained no noise, which significantly degrades the utility of analyzing the beamforming images for signal detection and reconstruction.

Digital beamforming as a detection algorithm is understood to mean digital beamforming plus a detection threshold placed on the amplitude of the frequency spectrum obtained by applying a fast Fourier transform (FFT) to the summed time-series (see Figure 4.13). This approach is most similar to the time-frequency spectrogram analysis employed in previous CRES experiments, however, in principle any signal detection algorithm could be used after the beamforming procedure. In Section 4.4 I analyze the signal detection performance of the power threshold approach in detail.

From the example frequency spectra in Figure 4.13 it is clear that without a reconstruction technique that coherently combines the signals from the full antenna our ability to detect CRES signals will be drastically reduced. Because the CRES signals are in-phase at the correct beamforming position the summed power increases as a function of N^2 compared to a single antenna channel, where N is the number of antennas. It is true that the noise power is also increased by beamforming, but, because the noise is incoherent, it's power only increases linearly. Consequently, the signal-to-noise ratio (SNR) of the CRES signal increases linearly with the number of antennas, which greatly improves detection efficiency compared to using only the information in a single antenna.

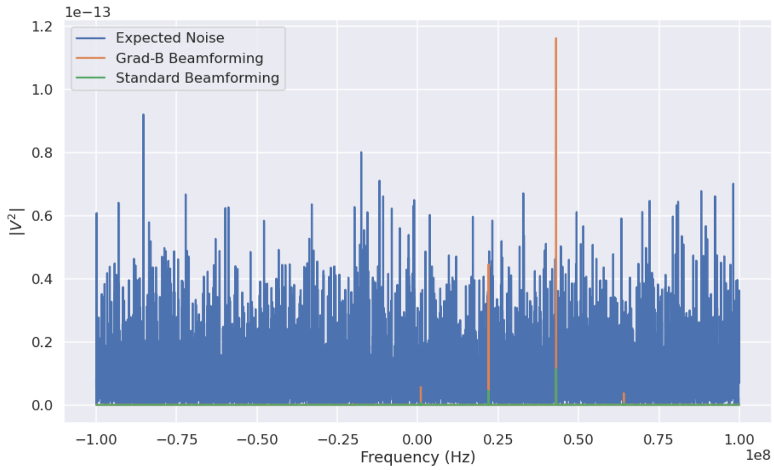


Figure 4.13: A plot of a typical frequency spectrum obtained by applying a Fourier transform to the time-series obtained from beamforming. The frequency spectra are plotted without noise on top of an example of a typical noise spectrum to visualize a realistic signal-to-noise ratio. In the example we see that without beamforming it would not be possible to detect anything since the signal amplitudes would be reduced by a factor of sixty relative to the noise. Additionally, we see that the ∇B-drift correction is needed to detect this electron since it comes from a simulation of an electron with a significant off-axis position.

The power threshold detection algorithm searches for high-power frequency bins that should correspond to a frequency component of the CRES signal. In order to prevent random noise fluctuations from being mistaken as CRES signals the power threshold must be set high enough so that it is unlikely that random noise could be responsible. A consequence of this is that many electrons that can be trapped will go undetected because the modulation caused by axial oscillations leads to the cyclotron carrier power to falling below the decision threshold. The time-dependent beamforming used to correct for the ∇B-drift increases the volume of the magnetic trap where electrons can be detected, but it is ineffective at increasing the range of detectable pitch angles (see Figure 4.14). Fundamentally, this is because the power threshold only uses a fraction of the signal power to detect electrons and ignores the power present in the frequency sidebands. In the subsequent sections I examine two other signal detection algorithms that seek to improve the detection efficiency of the FSCD by utilizing the more of the signal shape to compute the detection test statistics.

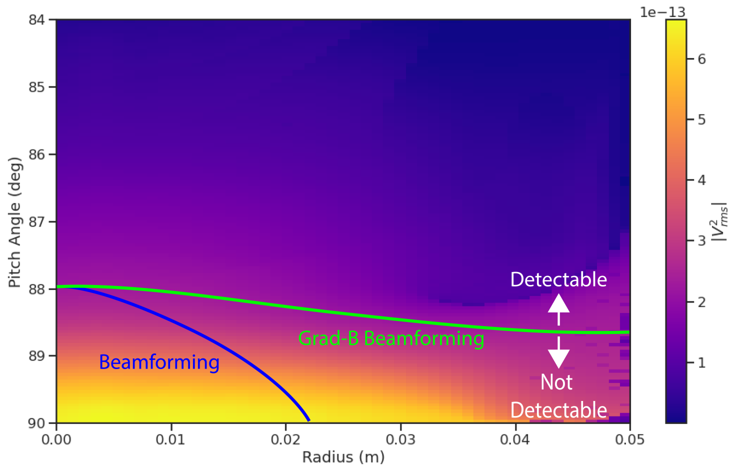


Figure 4.14: A plot of the total signal power received by the FSCD array from trapped electrons with different radial positions and pitch angles generated using Locust simulations. The lines on the plot indicate a 10 dB detection threshold above the mean value of the noise in the frequency spectrum. With static beamforming electrons with radial positions larger than about two centimeters are undetectable due to the change in the electron's position over time causing losses from beamforming phase mismatch. This is corrected by including ∇B-drift frequencies in the beamforming phases. Both beamforming techniques fail to detect electrons below $\approx 88.0°$, since these signal are composed of several relatively weak sidebands that are comparable to the noise.

4.3.2 Matched Filtering

Introduction to Matched Filtering

The problem of CRES signal detection is the problem of detecting a signal buried in WGN, which has been examined at great depth in the signal processing literature [69]. For a fully known signal in WGN the optimal detector is the matched filter, which means that it achieves the highest true positive rate for a fixed rate of false positives. The matched filter test statistic is calculated by taking the inner product of the data with the matched filter template

$$\mathcal{T} = \left| \sum_n h^\dagger[n] y[n] \right|, \tag{4.26}$$

where $h[n]$ is the matched filter template and $y[n]$ is the data. The matched filter test statistic defines a binary hypothesis test in which the data vector is assumed to be an instance of two possible data classes. By setting a decision threshold on the value of $\mathcal{T}$, one can classify a given data vector as belonging to two distinct hypotheses. Under the first hypothesis the data is composed of pure WGN, and under the second hypothesis the data is composed of the known signal with additive WGN. The matched filter template is obtained by rescaling the known signal in the following way

$$h[n] = \frac{x[n]}{\sqrt{\tau \sum_n x^\dagger[n] x[n]}}, \tag{4.27}$$

where τ is the variance of the WGN and $x[n]$ is the known signal. Strictly speaking, Equation 4.27 is only true for noise with a diagonal covariance matrix, however, in the context of the FSCD we are justified in assuming this to be true. Defining the matched filter templates in this way guarantees that the expectation value of $\mathcal{T}$ is equal to one when the data contains only noise, which is the standard matched filter normalization in the signal processing literature.

Although matched filters are canonically formulated in terms of a perfectly known signal, it is still possible to apply the matched filter technique given imperfect information about the signal provided that the signal is deterministic. From our discussion of CRES simulation tools for the FSCD (see Section 4.2) we know that the shape of CRES signals are completely determined by the initial parameters of the electron. The random collisions with background gas molecules which cause the formation of signal tracks are the only stochastic component of the CRES event after the initial beta-decay, therefore, it is possible to develop a matched filter for the detection of CRES signal tracks which are fully determined by the parameters of the electron after the initial beta-decay or subsequent collision events.

The matched filter test statistic for CRES signals is a modified version of Equation 4.26

$$\mathcal{T} = \max_{\boldsymbol{h},m} |\boldsymbol{h} * \boldsymbol{y}| = \max_{\boldsymbol{h},m} \left| \sum_k h^\dagger[k] x[m-k] \right|, \tag{4.28}$$

where the matched filter inner product has been replaced with a convolution operation and a maximization over the template and convolution delay (m). Replacing the inner product with a convolution accounts for the fact that the start time of the CRES signal is now an unknown parameter, in addition, we now perform a maximization of the matched filter convolution over a number of different templates. Because the shape of the signal is unknown we are forced to guess a number of different signal shapes to create a template bank with which we can identify unknown signals by performing an exhaustive search.

The template bank approach to matched filtering, while quite powerful, can quickly become computationally intractable. This is especially true in the case of the FSCD because of the large amount of raw data produced by the array that must be analyzed.

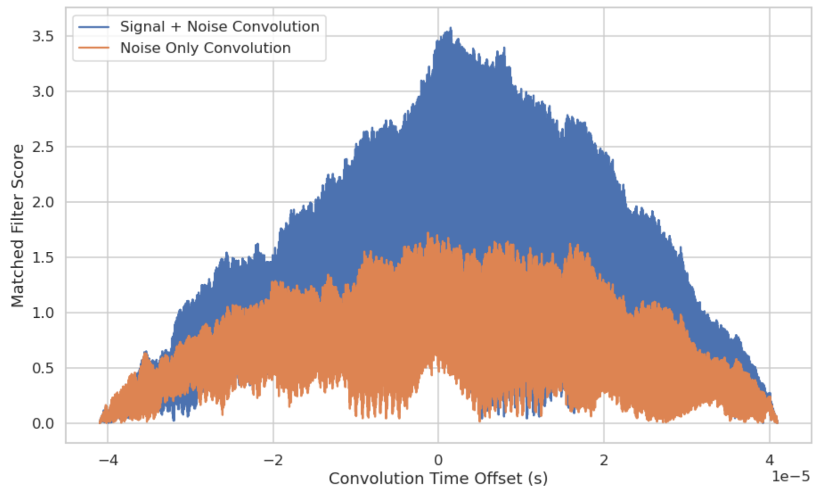


Figure 4.15: Example of a convolution of a CRES signal template with a segment of noisy data. A simulated CRES signal was simulated using Locust and normalized to create a matched filter template. When this template is convolved with noisy data the contains the matching signal the convolution output increases dramatically compared to data with only noise. The decreasing convolution output as the time offset of the convolution increases is caused by zero-padding of the data and template.

Specifically, the time-domain convolution specified by Equation 4.28 is particularly computationally intensive and is a major barrier towards the implementation of a matched filter for signal detection in an experiment like the FSCD. This can be avoided by using the convolution theorem to replace the time-domain convolution with an inner product in the frequency domain.

The convolution theorem states that

$$\boldsymbol{f} * \boldsymbol{g} = \mathcal{F}^{-1} \left(\boldsymbol{F} \cdot \boldsymbol{G} \right) \tag{4.29}$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are discretely sampled time-series, $\boldsymbol{F}$ and $\boldsymbol{G}$ are the respective discrete Fourier transforms, and $\mathcal{F}^{-1}$ is the inverse discrete Fourier transform operator. The convolution theorem allows us to perform the matched filter convolution by first computing the Fourier transform of the template and data, then performing a point-wise multiplication of the two frequency series, and finally performing the inverse Fourier transform to obtain the convolution output. Because discrete Fourier transforms can be performed extremely efficiently, the convolution theorem is almost always used in lieu of directly computing the convolution.

One thing to note here is that the convolution theorem for discrete sequences shown here, is technically valid only for circular convolutions, which is not directly specified in Equation 4.28. However, because typical CRES track lengths are much longer than the Fourier analysis window and also that the frequency chirp rates are small compared to the time-slice duration, it is relatively safe to use circular convolutions to evaluate matched filter scores for CRES signals, which allows us to apply the convolution theorem to compute matched filter scores using the frequency representation of the data and matched filter template.

Matched Filter Analysis of the FSCD

The optimality provided by the matched filter makes it a useful algorithm for analysis of CRES experiment designs for sensitivity analyses, since it indicates the best possible detection efficiency achievable by an experiment configuration. The standard approach to performing these studies involves generating a large number of simulated electron signals that span the kinematic parameter space of electrons in the magnetic trap. In general, electrons have six kinematic parameters along with an additional start time parameter.

In order to limit the number of simulations required to evaluate the detection efficiency the standard approach is to fix the starting axial position, starting azimuthal position, starting direction of the perpendicular component of the electron's momentum, and event start time to reduce the parameter space to starting radial position, starting kinetic energy, and starting pitch angle. The fixed variables are true nuisance parameters that do not affect the detection efficiency estimates for the FSCD design, because they manifest as phases which are marginalized during the calculation of the matched filter score.

Across radial position, kinetic energy, and pitch angle one defines a regular grid of parameters and uses Locust to simulate the corresponding signals (see Figure 4.16). This grid of simulated signals can be used to estimate the likelihood of detecting signals, because the matched filter score specifies the shape of the PDF that defines the detection probability and the size of the template bank influences the likelihood of a good match between a template and a random signal.

The matched filter approach can also be used to estimate the achievable energy resolution of the experiment by using a dense grid of templates generated with parameters close to the unknown signal (see figure 4.17). Because matched filter templates with similar parameters have signal shapes that are also similar, templates with incorrect parameters can have nearly identical matched filter scores as the correct template. Since only one sample of noise is included in a sample of real data, one cannot guarantee that the best matching template corresponds to the ground truth parameters of the signal.

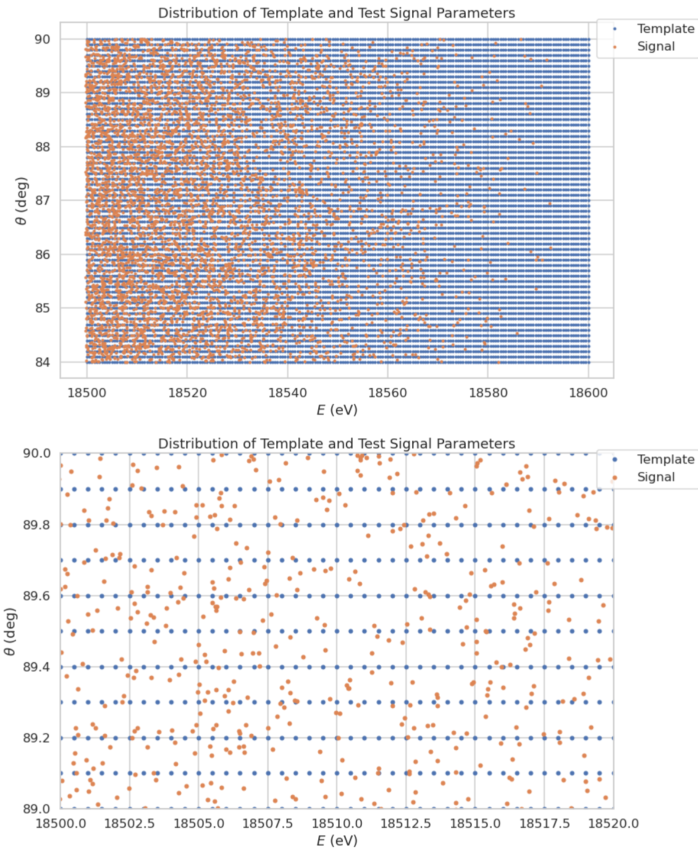


Figure 4.16: An example two-dimensional parameter distribution of a matched filter template bank and random test signals. θ refers to the pitch angle of the electron and E is the kinetic energy. The template bank forms a regular grid of in pitch angle and energy, whereas, the test signals are uniformly distributed in pitch angle and follow the tritium beta-decay kinetic energy distribution. This is why there are fewer test signals at higher energies. The need for high match across the full parameter space prevents one from reducing the density of templates in this low activity region. A zoomed in version of the template bank illustrates the relative density of templates and signals needed for match $> 90\%$.

This introduces uncertainty into the signal parameter estimation that manifests as an energy broadening. Dense grids of matched filter templates allows one to quantify this broadening by analyzing the parameter space of templates with matched filter scores close to the ground truth. This approach is analogous to maximum likelihood estimation and is one key component of a complete sensitivity analysis for an antenna array CRES experiment.

A key parameter for describing the performance of a matched filter template bank at signal detection is match, which we define as the average ratio of the highest matched

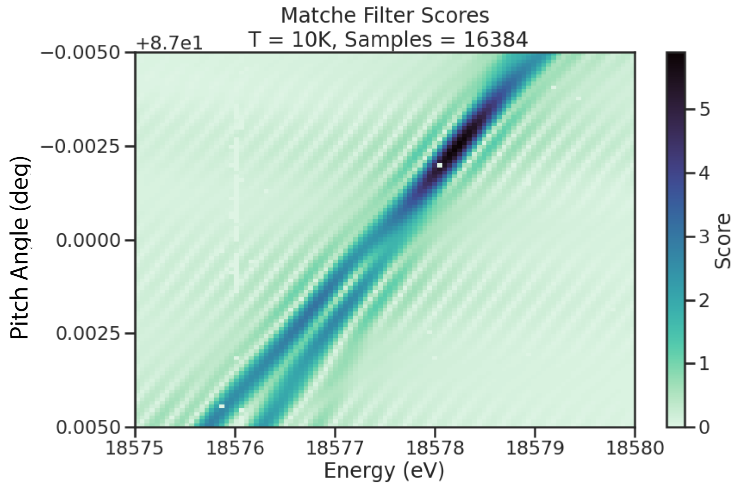


Figure 4.17: The matched filter scores of a dense grid of templates in pitch angle energy space. Dense template grids allow one to estimate the kinetic energy of the electron by identifying the best matching template. The uncertainty on this value is proportional to the space of templates that also match the test signal well. In the worst case matched filter templates can be completely degenerate where templates with different parameters match a signal with equal likelihood.

filter score for a random signal to the matched filter score for a perfectly matching template. In equation form this is

$$\text{Match} \equiv \Gamma = \frac{\mathcal{T}_{\text{best}}}{\mathcal{T}_{\text{ideal}}}, \tag{4.30}$$

where $\mathcal{T}_{\text{best}}$ is the matched filter score of the best fitting template in the bank and $\mathcal{T}_{\text{ideal}}$ is the hypothetical matched filter score one would measure if the signal perfectly matched the template. Generally, one desires an average match as close to one as possible, however, the average match value is an exponential function of the number of templates in the template bank (see Figure 4.18). This behavior is observed for dense matched filter grids like the one in Figure 4.17. A dense grid was used to calculate the average value of match for different template bank sizes shown in Figure 4.18.

The exponential relationship between match and template bank size is also evident for template banks that cover a wide range of parameters, such as the template bank visualized in Figure 4.16. Since no prior knowledge of the signal parameters is available, one has no choice but to use a template bank that covers a large range of parameters for signal detection. Achieving a high average match in this scenario can easily overwhelm the available computational resources, so in practice only a limited number of templates

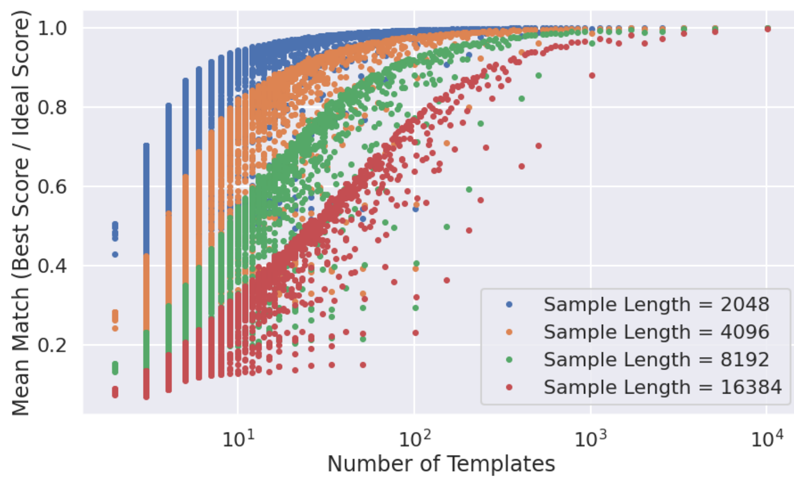


Figure 4.18: The mean match of the dense template grid shown in Figure 4.17 for different numbers of templates. Grids of different sizes were obtained by decimating a dense grid of templates and the average match for each grid was computed using the same set of randomly distributed test signals. Plotting the mean match against the size of the grid allows one to visualize the exponential relationship between match and template bank size. The noise in each curve is caused by sampling effects from the decimation algorithm. In general, longer templates are harder to than shorter templates.

could be used at the detection stage. Therefore, accurately modeling the effects of match is key to correct sensitivity calculations.

The effect of match on the detection efficiency of the matched filter template bank can be summarized using the ROC curve (see Figure 4.19). A single ROC curve is obtained by averaging over the PDFs that describe the detection probabilities of each individual template. The matched filter score for a template follows a Rician distribution with a mean value equal to the matched filter score multiplied by the match ratio between the template and signal. Therefore, the distribution that describes the average matched filter score when there is a signal in the data is obtained by averaging over the distributions for every template, whose expectation values are multiplied by the average match ratio.

The distribution of the matched filter score when there is no signal in the data follows a Rayleigh distribution. Therefore, a trials penalty, which is the statistical penalty one pays for randomly checking many templates in order to avoid a random match between noise and a template, is included by computing the joint distribution of N_{template} Rayleigh distributions, where N_{template} is the size of the template bank. For more information on the calculation of matched filter template bank ROC curves please refer to Section 4.4.

An alternative way to visualize the detection performance for each algorithm is to

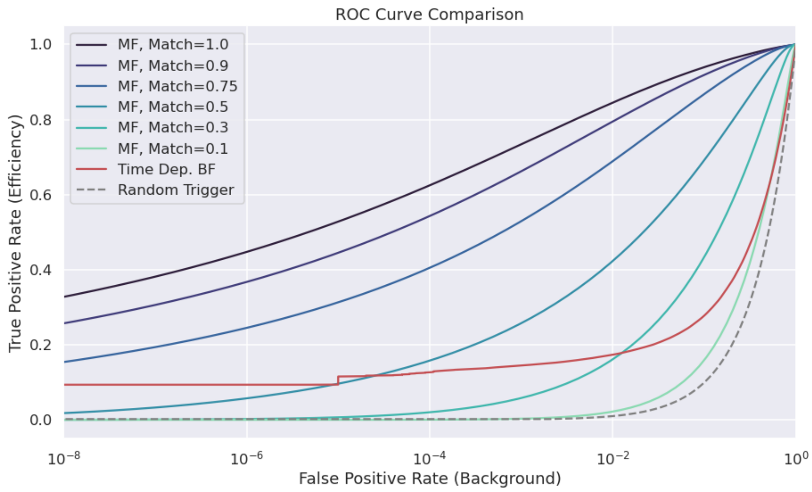


Figure 4.19: Matched filter template bank ROC curves as a function of mean match. One can see that for low match a matched filter is on average worse than the more straight forward beamforming detection approach.

specify a minimum acceptable false positive rate at the trigger level. This is equivalent to specifying a minimum threshold on the value of the matched filter score or the size of a frequency peak for a beamforming power threshold trigger. One can then draw regions of detectable signals as a function of the electron's pitch angle and radial position (see Figure 4.20). A kinetic energy shift is equivalent to an overall frequency shift of the signal and should have no effect on the detection probability assuming sufficient density of matched filter templates in the energy dimension. A electron is declared "detectable" for the regions in Figure 4.20 if the signal has at least 50% probability of falling above the decision threshold of the respective classifier. One can see that the parameter space of detectable signals is greatly expanded beyond the beamforming power threshold trigger with a matched filter (MF) or deep neural network (DNN) (see Section 4.3.3). Plots such as Figure 4.20 are useful for visualization, but, since the handling of detection likelihood is not sufficiently rigorous, the detection probability boundaries are not well-suited to sensitivity estimates.

Optimized Matched Filtering Implementation for the FSCD

The biggest practical obstacle to the implementation of a matched filter template bank detection approach is oftentimes the computational cost associated with exhaustively calculating the matched filter scores of the template bank, and the FSCD is no exception in this regard. At a basic level computing a matched filter score requires the convolution

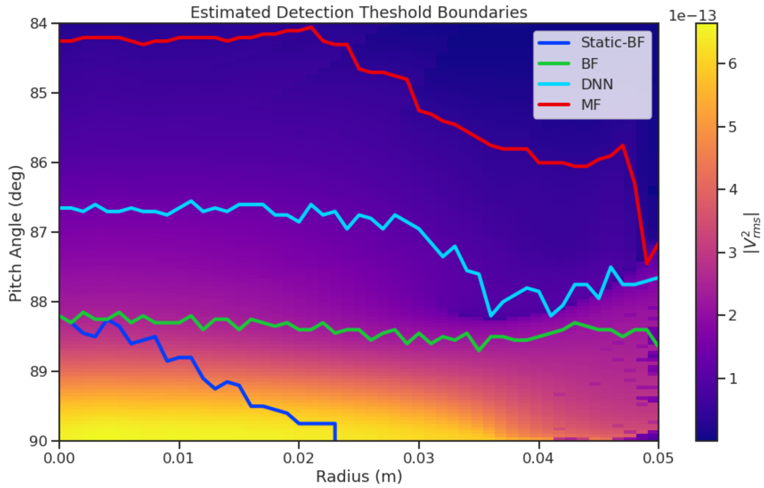


Figure 4.20: Boundaries of detectable electrons in pitch angle kinetic energy space for a series of different signal detection algorithms. A detectable signal is defined as a signal that is above a consistent decision with at least 50% probability. This non-rigorous treatment of detection probability is primarily useful for the visualization the relative increases in detection performance provided by the different algorithms. The static beamforming (Static-BF) algorithm is the digital beamforming algorithm introduced above without the ∇B-drift correction. The DNN algorithm refers to a convolutional neural network classifier trained to detect CRES signals (see Section 4.3.3).

of two vectors, which can be performed very efficiently by computers if the convolution theorem and fast Fourier transforms (FFT) are utilized. Furthermore, one can consider applying digital beamforming as a pre-processing step to reduce the dimensionality of the data before the matched filter is applied. In order to understand the relative gain in computational efficiency offered by these optimizations we analyze the total number of floating-point operations (FLOP) of several matched filter implementations in big O notation that utilize different combinations of optimizations.

A direct implementation of a matched filter as specified by Equation 4.28 involves the convolution of N_{ch} signals of length N_{s} with template signals of length N_{t}. As a uniform metric we shall compare the FLOP of the various matched filter implementations on a per-template basis, since each implementation scales linearly with the number of templates. The direct convolution approach to matched filtering costs

$$O(N_{\mathrm{ch}}) \times O(N_{\mathrm{s}} \times N_{\mathrm{t}}) \tag{4.31}$$

FLOP per-template, whose cost is dominated by the $O(M \times N)$ convolution operation.

The computational cost of the direct matched filter approach can be significantly reduced by exploiting the convolution theorem and FFT algorithms. If we restrict ourselves to signals and templates that contain equal numbers of samples then the convolution can be calculated by Fourier transforming both vectors, performing the point-wise multiplication, and then performing the inverse Fourier transform to obtain the convolution result. The FFT algorithm is able to compute the Fourier transform utilizing only $O(N \log N)$ operations compared to $O(N^2)$ for a naive Fourier transform implementation. This optimization results in a computational cost per-template of

$$O(N_{\mathrm{ch}}) \times O(N_{\mathrm{s}} \log N_{\mathrm{s}}) \tag{4.32}$$

A typical signal vector in the FSCD contains $O(10^4)$ samples in which case the FFT reduces the computational cost of the matched filter by a factor of $O(10^3)$. This large reduction in computational cost implies that a direct implementation of a matched filter is completely infeasible in the FSCD due to resource constraints.

Rather than relying solely on the matched filter it is tempting to consider using digital beamforming as an initial step in the signal reconstruction for the purposes of data reduction. The primary motivation is to reduce the dimensionality of the data by a factor of N_{ch} by combining the array outputs coherently into a single channel. One can view the beamforming operation as a partial matched filter, in the sense that the matched filter convolution contains the beamforming phased summation along with a prediction of the signal shape. By separating beamforming from the signal shape one hopes to reduce the overall computational cost by effectively shrinking the number of templates and reducing the number of operations required to check each one.

The nature of this optimization requires that we account for the number of templates used for pure matched filtering versus the hybrid approach. To first order, the total number of templates at the trigger stage is a product of the number of guesses for each of the electron's parameters

$$N_{\mathrm{T}} = N_{\mathrm{E}} \times N_{\theta} \times N_{\mathrm{r}} \times N_{\varphi}, \tag{4.33}$$

where N_{E} is the number of kinetic energies, N_{θ} is the number of pitch angles, N_{r} is the number of starting radial positions, and N_{φ} is the number of starting azimuthal positions. The starting axial position and cyclotron motion phase are not necessary to include in the template bank since these parameters manifest themselves as the starting phase of the signal, which is effectively marginalized when using a FFT to compute the matched filter convolution. Therefore, the total number of operations required by a matched filter to detect a signal in a segment of array data is on the order of

$$O(N_{\mathrm{T}}) \times O(N_{\mathrm{ch}}) \times O(N_{\mathrm{s}} \log N_{\mathrm{s}}) \tag{4.34}$$

With the hybrid approach we attempt to remove the spatial parameters from the template bank by using beamforming to combine the array signals into a single channel. Beamforming explicitly assumes a starting position, which allows us to only use matched filter templates that span the two-dimensional space of kinetic energy and pitch angle. The total computational cost of the hybrid method is directly proportional to the number of beamforming positions. For the time-dependent beamforming defined in Section 4.3.1, the number of beamforming positions is given by

$$N_{\mathrm{BF}} = N_{\mathrm{r}} \times N_{\varphi} \times N_{\omega_{\nabla\mathrm{B}}}, \tag{4.35}$$

where N_{r} and N_{φ} are the same spatial parameters encountered in the pure matched filter template bank and $N_{\omega_{\nabla\mathrm{B}}}$ is the number of ∇B-drift frequency assumptions. If a unique drift frequency is used for each pitch angle then the hybrid approach is effectively equivalent to a pure matched filter in the number of operations. The key efficiency gain of the hybrid approach is to exploit the relatively small differences in $\omega_{\nabla B}$ for electrons of different pitch angles by using only a small number of average drift frequencies.

The total number of operations for the hybrid approach can be expressed as a sum of the operations required by the beamforming and matched filtering steps,

$$O(N_{\mathrm{BF}}) \times O(N_{\mathrm{ch}} N_{\mathrm{s}}) + O(N_{\mathrm{BF}}) \times O(N_{\mathrm{E}} N_{\theta}) \times O(N_{\mathrm{s}} \log N_{\mathrm{s}}). \tag{4.36}$$

The first product in the sum is the number of operations required by beamforming, which is simply the number of beamforming points times the computational cost of the beamforming matrix multiplication, and the second product is the computational cost of matched filtering the summed signal generated by each beamforming position. To compare this to pure matched filtering we take the ratio of Equations 4.34 and 4.36 to obtain

$$\Gamma_{\mathrm{BFMF}} = \frac{O(N_{\omega_{\nabla B}})}{O(N_{\mathrm{E}} N_{\theta}) \times O(\log N_{\mathrm{s}})} + \frac{O(N_{\omega_{\nabla B}})}{O(N_{\mathrm{ch}})}. \tag{4.37}$$

This expression can be simplified by observing that $O(N_{\mathrm{E}} N_{\theta}) \times O(\log N_{\mathrm{s}}) \gg O(N_{\mathrm{ch}})$, which means that the ratio of computational cost for the two methods can be reduced to

$$\Gamma_{\mathrm{BFMF}} \approx \frac{O(N_{\omega_{\nabla B}})}{O(N_{\mathrm{ch}})}. \tag{4.38}$$

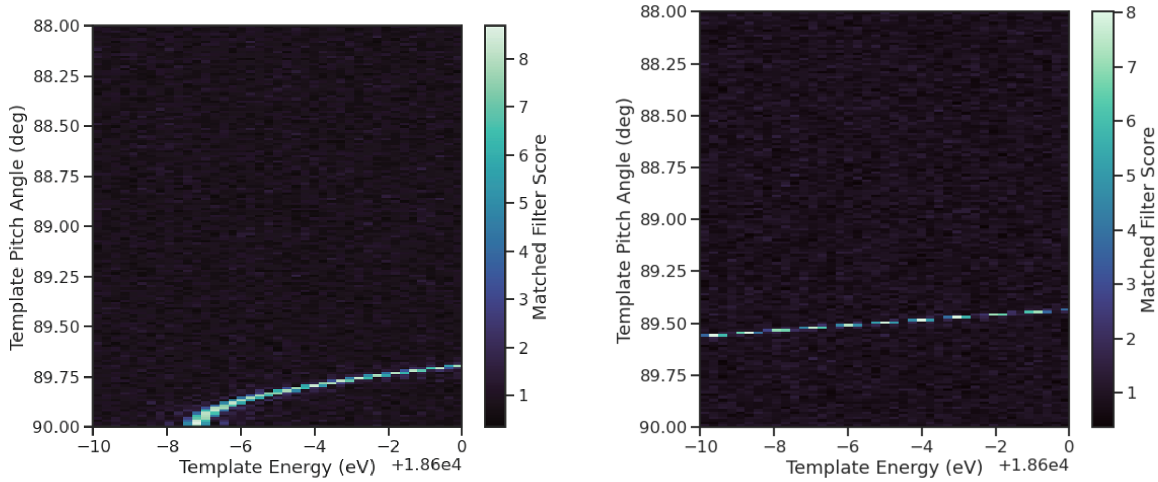
If we limit ourselves to a number of estimated drift frequencies of $O(1)$ then we see that the estimated computational cost reduction of the hybrid approach is of $O(N_{\mathrm{ch}})$. This is quite a large reduction considering that the FSCD antenna array contains sixty antennas in the baseline design.

The main drawback of the hybrid approach is that the limited number of allowed drift frequency guesses can lead to detection efficiency loss due to phase mismatch. The degree of phase error from an incorrect drift frequency is proportional to the length of the array data vector used by the signal detection algorithm. For signals with lengths equal to the baseline FSCD Fourier analysis window of 8192 samples, typical phase errors from using an average versus the exact ∇B-drift frequency are on the order of a few percent in terms of the signal energy. This has a relatively small impact on the overall detection efficiency, however, future experiments with antenna array CRES will want to balance optimizations such as these during the design phase to keep experiment costs to a minimum while still achieving scientific goals.

Kinetic Energy and Pitch Angle Degeneracy

More accurate modeling of a matched filter requires that we consider the effects of mismatched signals and template, since this more accurately reflects the real-world usage of a matched filter where many incorrect templates are convolved with the data until the matching template is found. One way to study this is to use the grid of simulated signals to compute the matched filter scores between mismatched signals and templates and evaluate the matched filter scores under this scenario. What one finds when performing this analysis is that templates for kinetic energies and pitch angles that do not match the underlying signal can have matched filter scores that are indistinguishable from the matched filter score of the correct template (see Figure 4.21 and Figure 4.21).

This degeneracy in matched filter score is the result of correlations between the kinetic energy of the electron and the pitch angle caused by changes in the average magnetic field experienced by an electron for different pitch angles. While in principle helpful for the purposes of signal detection these correlations are unacceptable since they greatly reduce the energy resolution of the experiment by causing electrons with specific kinetic energy to templates across a wide range of energies. It is important to emphasize that this



(a) (b)

Figure 4.21: Two example illustrations of the correlation between kinetic energy and pitch angle imparted by the shape of the FSCD magnetic trap. The correlations manifest themselves as degeneracies in the matched filter score where multiple matched filter templates have the same matched filter for a particular signal. These degeneracies are a sign that the magnetic trap must be redesigned in order to break the correlation between pitch angle and kinetic energy.

degeneracy cannot be fixed by implementing a different signal reconstruction algorithm. As revealed by the matched filter scores the shapes of the signals for different parameters are identical. Resolving this degeneracy between pitch angle and energy requires the design of a new magnetic trap with steeper walls so that the average magnetic field experienced by an electron is less dependent on pitch angle.

4.3.3 Machine Learning

Machine learning is a vast and rapidly developing field of research [72]. In this Section we shall attempt to provided a brief introduction to some of the concepts and techniques of machine learning that were applied to CRES signal detection rather than attempt a comprehensive overview.

Introduction to Machine Learning

Digitization of the FSCD antenna array generates large amounts of data that must be rapidly processed to enable real-time signal detection and reconstruction. While digital beamforming combined with a power threshold is relatively computationally inexpensive,

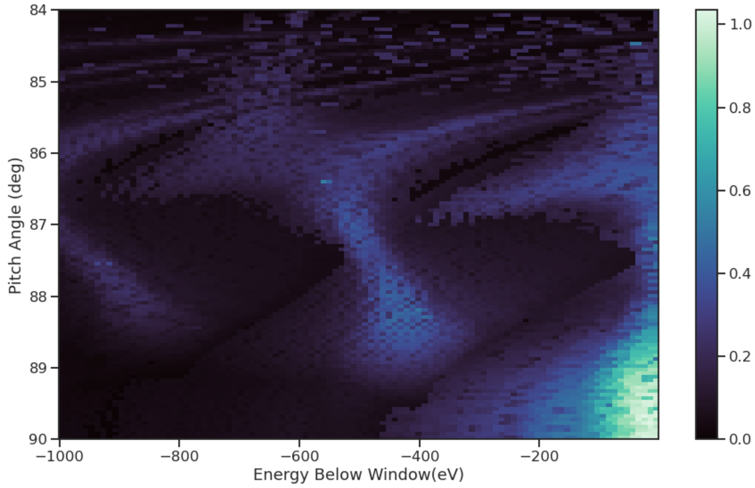


Figure 4.22: A visualization of the correlation between energy and pitch angle in the FSCD magnetic trap. The image is formed by computing the match of the best template from a grid consisting of pitch angles from 84 to 90 degrees in steps of 0.05 degrees, kinetic energies from 17574 to 18574 eV, located at 2 cm from the central axis, and simulated for a length of three FSCD time-slices. The signals used to compute the best matching template consisted of a grid from 84 to 90 degrees in steps of 0.05 degrees, kinetic energies from 18550 to 18575 eV in steps of 0.25 eV, located 2 cm from the central axis, and simulated for three FSCD time-slices. The colored regions of the plot show how well signals with lower energy can match those of higher energy for the FSCD magnetic trap, which is proportional to the achievable energy resolution of the FSCD design.

it is relatively ineffective at detecting CRES signal with small pitch angles, since it relies on a visible frequency peak above the noise. On the other hand, a matched filter is able to detect signals with a significantly larger range of parameters, however, the exhaustive search of matched filter templates can be computationally expensive. Machine learning based triggering algorithms have been used successfully in many different high-energy physics experiments [73] and recent developments have shown success in the detection of gravitational wave signals using machine learning techniques [74, 75] in place of the more traditional matched filtering method. This motivates the exploration of machine learning as a potential CRES signal detection algorithm.

There are several different approaches to machine learning, but the one most important to our discussion here is the supervised learning approach. In supervised machine learning one uses a differentiable model or function that is designed to map the input data to the appropriate label [72]. The data is represented as a multidimensional matrix of floating point values such as an image or a time-series, and the label is generally a class name such as signal or noise for classification problems or a continuous value like kinetic energy in the case of regression problems.

In supervised learning the model is trained to map from the data to the correct label by evaluating the output of the model using a training dataset consisting of a set of paired data and labels. To evaluate the difference between the model output and the correct label a loss function is used to quantify the error between the model prediction and the ground truth. For example, a common loss function in regression problems is the squared error loss function, which quantifies error using the squared difference between the model output and label.

Using the outputs of the loss function the next step in supervised learning is to compute the gradient of error with respect to the model parameters in a process called backpropogation. Using the model parameter gradients the last step in the supervised learning process is to perform an update of the parameter values in order to minimize the error in the model predictions across the whole dataset. This loop is performed many times while randomly shuffling the dataset until the error converges to a minimum value at which point the training procedure has finished. It is standard practice to monitor the training procedure by evaluating the performance of the model using a separate validation dataset that matches the statistical distribution of the training data and to check the performance of the model after training using yet another dataset called the test dataset. These practices help to guard against overtraining which is a concern for models with many parameters.

Convolutional Neural Networks

A popular class of machine learning models are neural networks. A neural network is essentially a function composed of a series of linear operations called layers which take a piece of data typically represented as a matrix, multiplies the elements of the data by a weight, and then sums these products to produce an output matrix. Neural networks composed of purely linear operations are unable to model complex non-linear behavior, therefore, non-linear activation functions are applied to the outputs of each of the layers to increase the ability of the neural network to model complex relationships between the data.

Neural networks are typically composed of at least three layers, but with the present capabilities of computer hardware they more often contain many more than this. The first layer in a neural network is called the input layer, because it takes the data objects as input, and the last layer in a neural network is known as the output layer. The output layer is trained by machine learning to map the data to a desired output using the supervised learning procedure described in Section 4.3.3. In between the input and the output layer are typically several hidden layers that receive inputs from and transmit outputs to other layers in the neural network model. The term deep neural network (DNN) refers to those neural networks that have at least one hidden layer, which have proven to be extremely powerful tools for pattern recognition and function approximation.

An important type of DNN are convolutional neural networks (CNN) that typically contain several layers which perform a convolution of the input with a set of filters. These convolution operations are typically accompanied by layers that attempt to down-sample the data along with the standard neural network activation functions. A standard CNN is composed of several convolutional layers at the beginning of the network and ends with a series of fully-connected neural network layers at the output. Intuitively, one can imagine that the convolutional layers are extracting features from the data that fully-connected layers use to perform the classification or regression task.

Deep Filtering for Signal Detection in the FSCD

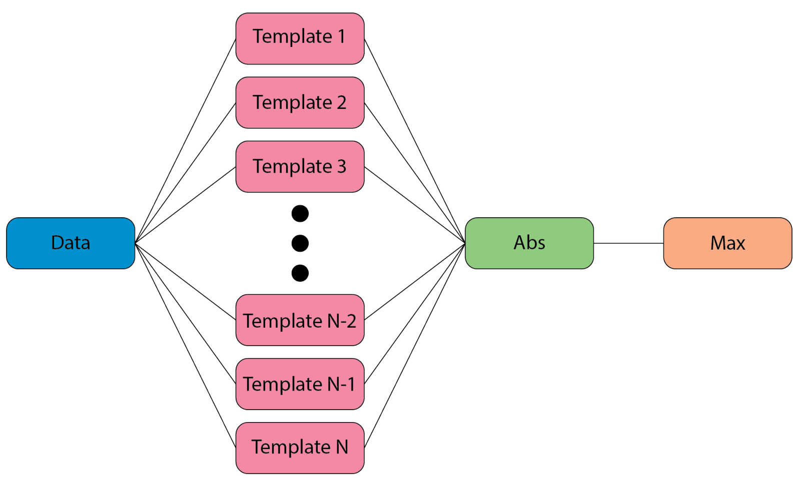


Figure 4.23: A representation of a matched filter template bank as a convolutional neural network. The network has a single layer composed of the templates, which act as convolutional filters. The activation of the neural network is an absolute value followed by a max operator.

CNNs have been extremely influential in the field of computer vision, particularly tasks such as image segmentation and classification, but have also been applied in numerous experimental physics contexts. Given the particular challenge posed by signal detection and reconstruction in the FSCD we are interested in exploring the potential of machine learning as an effective algorithm for real-time signal detection, since this application requires both high efficiency and fast evaluation.

In the machine learning paradigm signal detection is equivalent to a binary classification problem between the signal and noise data classes, and my investigation focuses specifically on the application of CNNs to signal detection in the FSCD, which is motivated by relatively recent demonstrations of CNNs achieving classification accuracies for gravitational wave time-series signals comparable to a matched filter template bank. In this framework it is possible to interpret the matched filter as a type of CNN composed of a single convolutional layer with the templates making up the layer filters (see Figure 4.23). Since this neural network has no hidden layers, it is not a DNN like we have been discussing so far, but we can attempt to construct a proper CNN that attempts to reproduce the classification performance of the matched filter network.

The name deep filtering refers to this scheme of replacing a matched filter template bank with a DNN. The reason why one might want to do this is that it may be possible to exploit redundancies and correlations between templates that may allow one to perform signal detection with similar accuracy but with fewer computations, which is important for real-time detection scenarios like the FSCD experiment. In Section 4.4 we perform a detailed comparison of the signal detection performance of a CNN to beamforming and a matched filter template bank.

Deep filtering is conceptually a simple technique. Similar to a matched filter template bank a large number of simulated CRES signals are generated and used to train a model to distinguish between signal and noise data (see Figure 4.24). In order to reduce the dimensionality of the input FSCD data a digital beamforming summation is applied to the raw time-series data generated by Locust to compress the 60-channel data to a single time-series. CRES signal have a sparse frequency representation and experiments training CNN's on time-series and frequency series data found that models trained on frequency spectrum data performed significantly better, therefore, an FFT is applied to the summed time-series before being normalized and fed to the classification model.

The data used to train the model consists of an equal proportion of signal and noise frequency spectra. Unique samples of WGN are generated and added to the signals during training time to avoid have to pre-generate and store large samples of noise data. The binary cross-entropy loss function combined with the ADAM optimizer proved effective at training the models to classify CRES data. A simple hyperparameter optimization was performed by manually tuning model, loss function, and optimizer parameters. The model and training loops was implemented in python using the PyTorch deep learning framework. Standard machine learning best practices were followed when training the

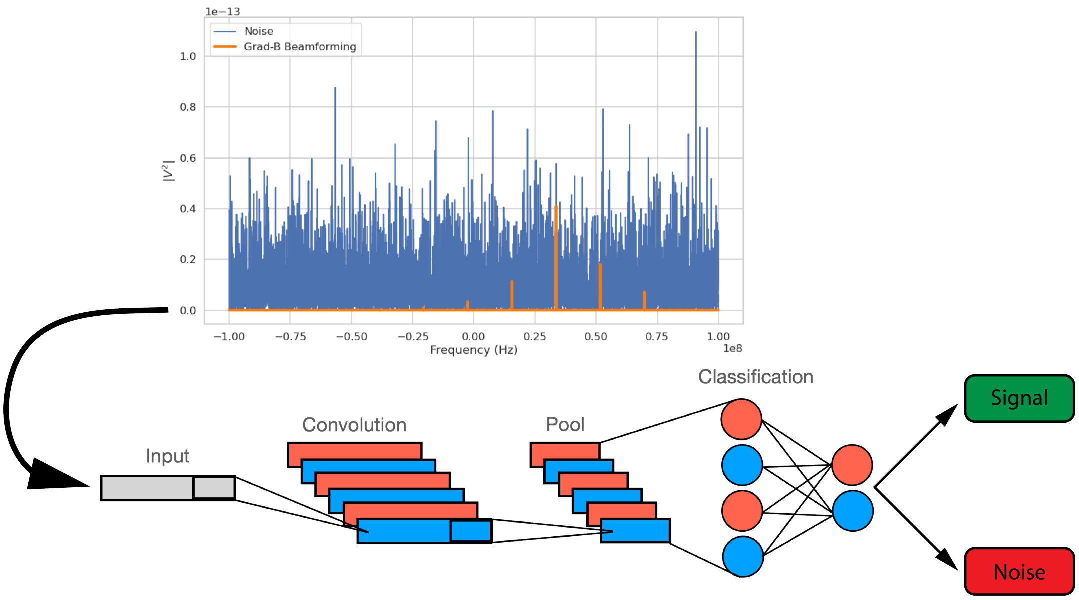


Figure 4.24: A graphical depiction of CRES signal detection using a CNN. A noisy segment of data is converted to a frequency series using digital beamforming and a FFT. The complex-valued frequency series is input into a trained CNN model that classifies the data as signal or noise using a decision threshold on the CNN output.

models, such as overtraining monitoring using a validation dataset. Models were trained until the training loss and accuracy converged and then evaluated using a separate test data set.

The classification results of the test dataset are used to quantify the relationship between the true positive rate and the false positive rate for the model. The true positive rate is analogous to detection efficiency and the false positive rate is a potential source of background in the detector. One can limit the rate of false positives using a sufficiently high threshold on the model output at the cost of a lower detection efficiency (see Figure 4.25 and Figure 4.26). As expected, the performance of the model at signal classification is negatively effected the noise power, which is quantified by the noise temperature.

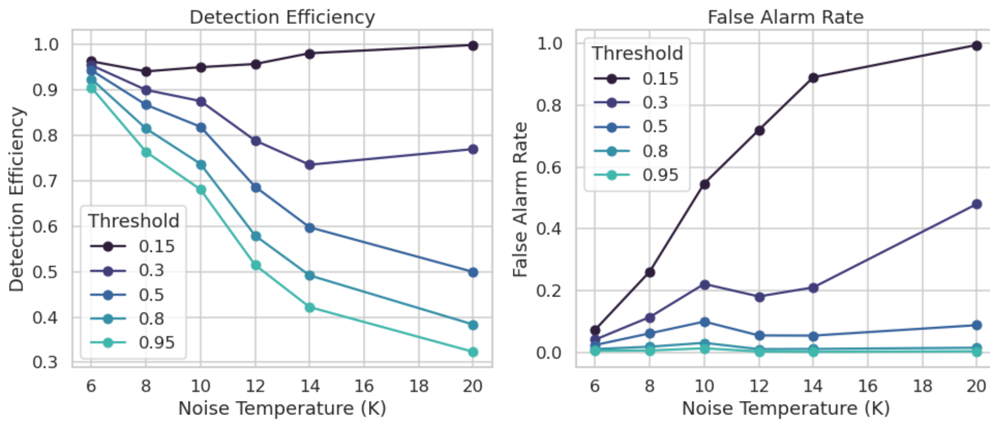


Figure 4.25: The detection efficiency and false alarm rate (false positive rate) as a function of the decision threshold for different values of the noise temperature. The model is trained to output a value close to one for data that contains a signal and outputs a value near zero when the data contains only noise. One sees that a lower decision threshold will have a high detection efficiency at the cost of a high rate of false alarms.

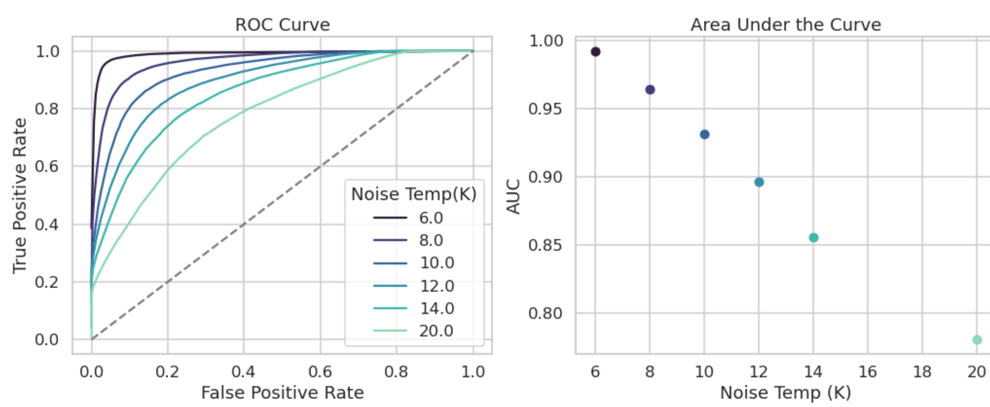


Figure 4.26: ROC curves for a CNN model classifying CRES signals. One can see that the area under the curve, which is a figure of merit that descibes the performance of the classifier, is roughly linearly dependent with the noise temperature.

4.4 Analysis of Signal Detection Algorithms for the Antenna Array Demonstrator

This section contains an early version of the manuscript for the triggering paper prepared for publication in JINST. In it I present a relatively detailed analysis of the signal detection performance of the three signal detection approaches discussed so far using a population of simulated CRES signals generated with Locust. The focus of the paper is on the performance of the signal detection algorithms for pitch angles below 88.5° where the beamforming power threshold begins to fail.

4.4.1 Introduction

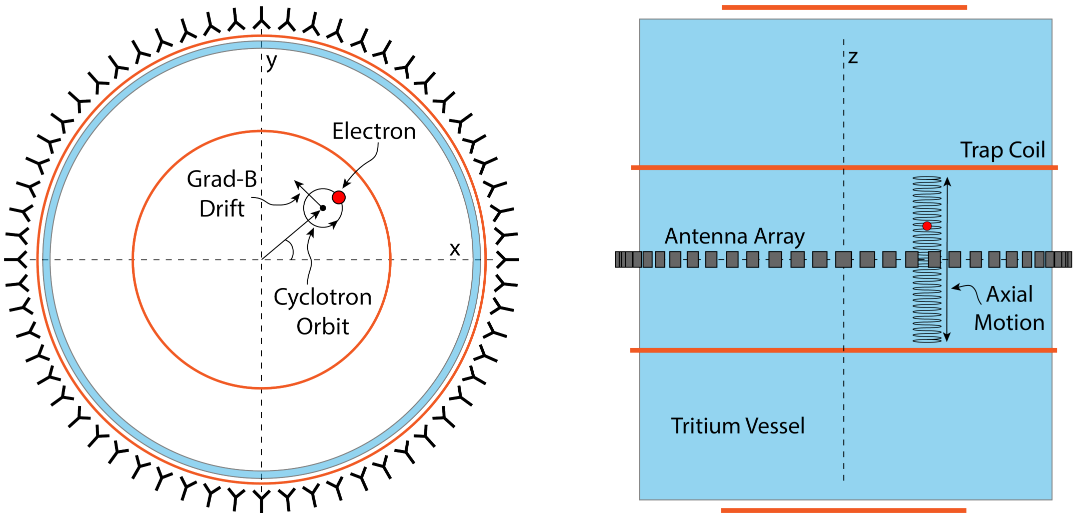
Cyclotron Radiation Emission Spectroscopy (CRES) is a technique for measuring the kinetic energies of charged particles by observing the frequency of the cyclotron radiation that is emitted as they travel through a magnetic field [38]. The Project 8 Collaboration is developing the CRES technique as a next-generation approach to tritium beta-decay endpoint spectroscopy for neutrino mass measurement. Recently, Project 8 has used CRES to perform the first ever tritium beta-decay energy spectrum and neutrino mass measurement [40, 41].

Previous CRES measurements have utilized relatively small volumes of gas that are directly integrated with a waveguide transmission line, which transmits the cyclotron radiation emitted by the trapped electrons to a cryogenic amplifier. While this technology has had demonstrable success, it is not a feasible option for scaling up to significantly larger measurement volumes. In particular, the goal of the Project 8 Collaboration is to use CRES combined with atomic tritium to measure the neutrino mass with a 40 meV sensitivity. Achieving this sensitivity goal will require a multi-cubic-meter scale measurement volume in order to obtain the required event statistics in the tritium beta-spectrum endpoint region; hence, there is a need for new techniques to enable large volume CRES measurements for future experiments.

One approach is to surround a large volume with an array of antennas that together collect the cyclotron radiation emitted by trapped electrons [39, 76]. A promising array design is an inward-facing uniform cylindrical array that surrounds the tritium containment volume. Increasing the size of the antenna array, by adding additional rings of antennas along vertical axis, allows one to grow the experimental volume until a sufficient amount of tritium gas can be observed by the array. A challenging aspect of this approach is that the total radiated power emitted by an electron near the tritium spectrum endpoint is on the order of 1 fW or less, which is then distributed between all the antennas in the array. Consequently, detecting the presence of a CRES signal and determining the electron's kinetic energy requires reconstructing the entire antenna array output over the course of the CRES event, posing a significant data acquisition and signal reconstruction challenge.

Project 8 has developed a triggering system to enable real-time identification of CRES events using an antenna array [77]. Previous measurements with the CRES technique have utilized a threshold on the frequency spectrum formed from a segment of CRES time-series data. This algorithm relies on the detection of a frequency peak above the thermal noise background, which limits the kinematic parameter space of detectable electrons. Due to the limitations of this power threshold, Project 8 has been investigating alternative signal identification approaches, including both matched filtering and machine learning based classifiers, to improve the detection efficiency of the experiment. In order to evaluate the relative gains in detection efficiency that come from utilizing these alternative algorithms, we develop analytical models for the power threshold and matched filter signal classifier performance applicable to an antenna array based CRES detector. In addition, we implement and test a basic convolutional neural network (CNN) as a first step towards the development of neural-network based classifiers for CRES measurements. These results allow us to compare the estimated detection efficiencies of each of these methods, which we weigh against the associated computational costs for real-time applications.

The outline of this paper is as follows. In Section 4.4.2 we give an overview of a prototypical antenna array CRES experiment, and describe the major steps involved in the proposed approach to real-time signal identification. In Section 4.4.3 we develop models for the power threshold and matched filter algorithms, and introduce the machine learning approach and CNN architecture. In Section 4.4.4 we describe our process for generating simulated CRES signal data and the details of training the CNN. Finally, in Section 4.4.5 we perform a comparison of the signal classification accuracy of the three approaches and discuss the relevant trade-offs in terms of detection efficiency and computational cost.



(a) Top-down view. (b) Side view.

Figure 4.27: An illustration of the conceptual design for an antenna array CRES tritium beta-decay spectrum measurement. The antenna array geometry consists of a 20 cm interior diameter with 60 independent antenna channels arranged evenly around the circumference. The nominal antenna design is sensitive to radiation in the frequency range of 25-26 GHz, which corresponds to the cyclotron frequency of electrons emitted near the tritum beta-spectrum endpoint in a 1 T magnetic field. The array is located at the center of the magnetic trap produced by a set of current-carrying coils. The nominal magnetic trap design is capable of trapping electrons up to 5 cm away from the central axis of the array and traps electrons within an approximately 6 cm long axial region centered on the antenna array.

4.4.2 Signal Detection with Antenna Array CRES

4.4.2.1 Antenna Array and DAQ System

In order to explore the potential of antenna array CRES for neutrino mass measurement, the Project 8 Collaboration has developed a conceptual design for a prototype antenna array CRES experiment [39, 76], called the Free-space CRES Demonstrator or FSCD, which could be used as a demonstration of the antenna array measurement technique (see Figure 4.27). The FSCD utilizes a single ring of antennas, which is the simplest form of a uniform cylindrical array configuration, to surround a radio-frequency (RF) transparent tritium gas vessel. A prototype version of this antenna array has been built and tested by the Project 8 collaboration to validate simulations of the array radiation pattern and beamforming algorithms [42]. In the FSCD the antenna array is positioned at the center of the magnetic trap formed by a set of electro-magnetic coils that are designed to produce a magnetic trap with flat central region and steep walls both radially and axially.

When a beta-decay electron is trapped its motion consists of three primary components. The component with the highest frequency is the cyclotron orbit whose frequency is determined by the size of the background magnetic field. The FSCD design assumes a background magnetic field value of approximately 0.96 T, which results in cyclotron frequencies for electrons with kinetic energies near the tritium beta-spectrum endpoint from 25 to 26 GHz. The component with the next highest frequency is the axial oscillation experienced by electrons with pitch angles of less than 90° [59]. The flat region of the FSCD magnetic trap extends approximately 3 cm above and below the antenna array plane causing electrons to move back and forth as they are reflected from the trap walls. Typical oscillation frequencies are on the order of 10's of MHz, which results in an oscillation period that is $O(10^3)$ smaller than the duration of a typical CRES event. Therefore, when reconstructing CRES events we treat the electron as occupying only an average axial position at the center of the magnetic trap, since we are not able to resolve the axial position as a function of time. The component of motion with the smallest frequency is ∇B-drift caused by radial field gradients in the trap, producing an orbit of the electron around the central axis of the trap with a frequency on the order of a few kHz, dependent on the pitch angle and the radial position of the electron.

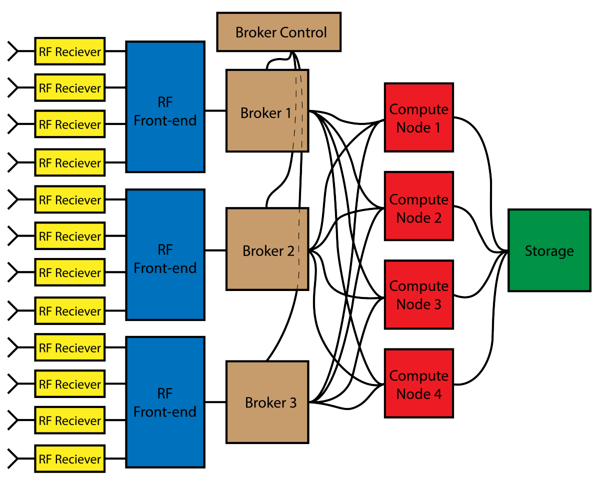


Figure 4.28: A high-level diagram of the DAQ system archtecture envisioned for the FSCD.

The data acquisition (DAQ) system digitizes the signals from the antenna array and combines thee data streams into a time-ordered matrix of array snapshots that can be used by the reconstruction algorithms. The FSCD DAQ system design [77] is divided into three layers 4.28. The first layer is the RF front-end, which includes the antenna array, the RF receiver boards, and the digitization electronics. The receiver board contains an amplifier, RF mixer, and bandpass filter to enable down-conversion, and is followed by the digitization electronics that sample the CRES signals at 200 MHz. In order to achieve an adequate signal-to-noise ratio to detect CRES events, the DAQ system for the antenna array demonstrator must have a total system noise temperature of ≈ 10 K, which we can achieve by using low-noise amplifiers and operating at cryogenic temperatures. After digitization, the array data must be reorganized from individual data streams sorted by channel into array snapshots sorted by time. In order to solve this data transfer and networking problem the second layer of the DAQ system consists of a set of broker computer nodes that reorganize the array data into time-ordered chunks. This approach allows us accommodate different data transfer requirements by scaling the number of broker nodes in this layer accordingly. Next, the broker layer distributes these chunks of array data to the final layer of the DAQ system, which consists of a set of identical reconstruction nodes that perform the calculations required for CRES reconstruction. Similar to the broker layer, the number of reconstruction nodes can be increased or decreased depending on the amount of computer power required for real-time CRES reconstruction.

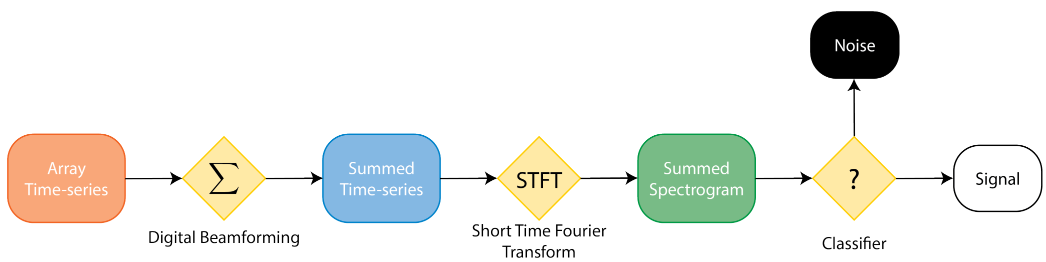


Figure 4.29: A block diagram illustration of the real-time triggering algorithm proposed for antenna array CRES reconstruction.

The design of the FSCD DAQ system is intended to enable a significant portion of the CRES event reconstruction to occur in real-time. The motivation for this comes from the fact that the FSCD antenna array generates approximately 1 exabyte of raw data per year of operation. Therefore, in order to reduce the data-storage requirements, it is ideal to perform at least some of the CRES event reconstruction in real-time so that it is possible to save a reduced form of the data for offline analysis. The first step of the real-time reconstruction would be a real-time signal detection algorithm, which is the focus of this paper. Our approach consists of three main operations performed on the time-series data blocks including digital beamforming, a short time Fourier transform

(STFT), and a binary classification algorithm to distinguish between signal and noise data (see Figure 4.29).

4.4.2.2 Real-time Signal Detection

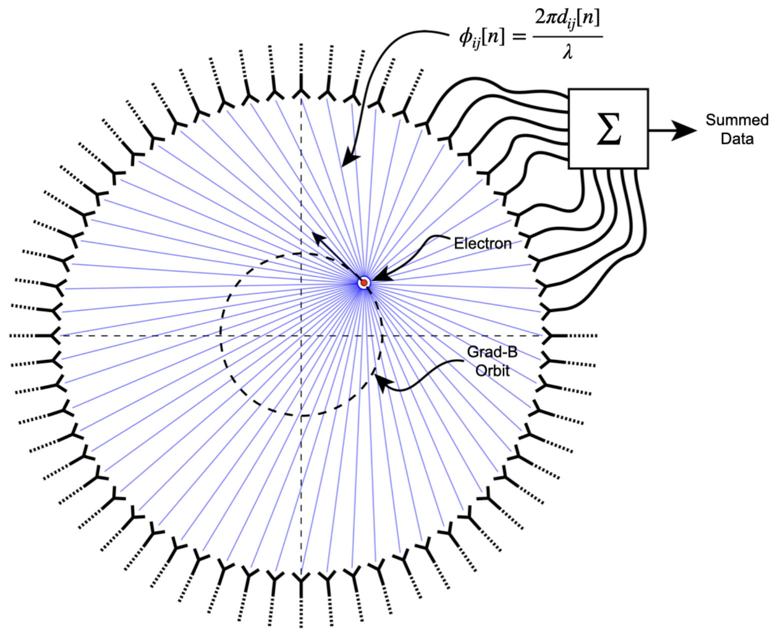


Figure 4.30: An illustration of the digital beamforming procedure. The blue lines indicate the various distances from the beamforming position to the antenna. In the situation depicted the actual position of the electron matches the beamforming position, so we should expect constructive interference when the phase shifted signals are summed. To prevent the electron's ∇B-motion from moving the electron off of the beamforming position, the beamforming phase include a time-dependence to follow the trajectory of the electron in the magnetic trap.

The first step in the real-time detection algorithm is digital beamforming, which is a phased summation of the signals received by individual antennas in the array (see Figure 5.21). The phase shifts correspond to the path length differences between a spatial position and each individual antenna such that, when there is an electron located at the beamforming position, all the signals received by the array constructively interfere. Since we do not know ahead of time where an electron will be produced in the detector, we define a grid of beamforming positions that cover the entire region where electrons can be trapped and perform a phased summation for each of these points for every time-step in the array data block. As we saw in Section 4.4.2.1, the axial oscillation of the electrons prevents us from resolving it's position along the Z-axis of the trap, therefore our beamforming grid need only cover the possible positions of the electron in the two-dimensional plane defined by the antenna array.

The equation defining digital beamforming can be expressed as

$$\mathbf{y}[n] = \mathbf{\Phi}^T[n]\mathbf{x}[n], \tag{4.39}$$

where $\mathbf{x}[n]$ is array snapshot vector at the sampled time n, $\mathbf{\Phi}[n]$ is the matrix of beamforming phase shifts, and $\mathbf{y}[n]$ is summed output vector that contains the voltages for each of the summed channels that correspond to a particular beamforming position. The elements of the beamforming phase shift matrix can be expressed as a weighted complex exponential

$$\mathbf{\Phi}_{ij}[n] = A_{ij}[n] \exp\left(2\pi i \phi_{ij}[n]\right), \tag{4.40}$$

where the indices i and j label the beamforming and antenna positions respectively. The weight A_{ij} accounts for the relative power increase for antennas that are closer to the position of the electron, and ϕ_{ij} is the total beamforming phase shift for the j-th antenna at the i-th beamforming position.

The beamforming phase shift is a sum of two terms

$$\phi_{ij}[n] = \frac{2\pi d_{ij}[n]}{\lambda} + \theta_{ij}[n], \tag{4.41}$$

where the first term is the phase shift originating from the path length difference ($d_{ij}[n]$) between the beamforming and antenna positions, which are represented by the vectors (r_j, θ_j) and $(r_i, \theta_i[n])$, and the second term is the angular separation ($\theta_{ij}[n]$) of the two positions. The angular separation enters into the beamforming phase due to an effect caused by the circular orbit of the electron that produces radiation whose phase is linearly dependent on the relative azimuthal position of the antenna [78,79]. The time-dependence of the beamforming phases is intended to correct for the effects of ∇B-drift, which cause the guiding centers of electrons to orbit the center of the magnetic trap. By including a linear time-dependence in the azimuthal beamforming position,

$$\theta_i[n] = \omega_{\nabla B} t[n] + \theta_{i,0}, \tag{4.42}$$

where $\omega_{\nabla B}$ is the azimuthal grad-B drift frequency, $t[n]$ is the time vector and, $\theta_{i,0}$ is the starting azimuthal position, we can configure the beamforming phases to effectively track the XY-position of the guiding center over the event duration. Predicting accurate values of $\omega_{\nabla B}$ for a specific trap and set of kinematic parameters will be done by simulations, which are performed using the Kassiopeia software package [58] by Project 8.

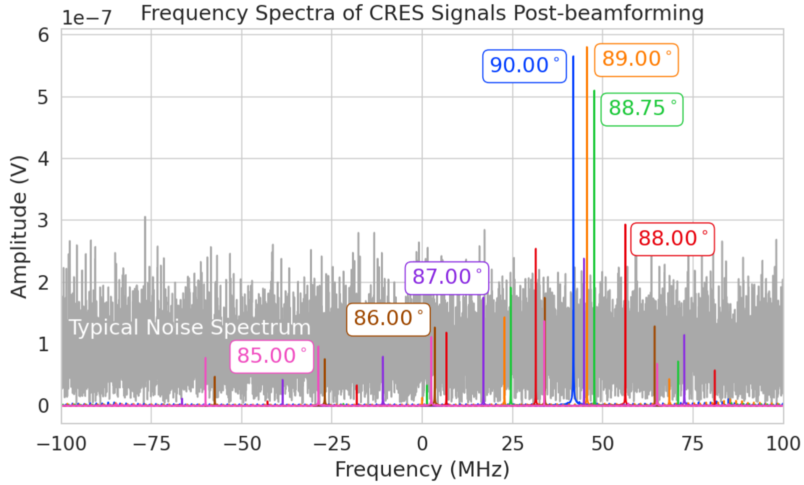


Figure 4.31: Frequency spectra of simulated CRES signals post-beamforming. The signal of a 90° electron consists of a single frequency component that is easy to detect with a power threshold on the frequency spectrum. This power threshold is still effective for signals with relatively large pitch angles such as 89.0° and 88.75°, which are composed of a main carrier and a few small sidebands. Signals with smaller pitch angles, below about 88.5°, tend to be dominated by sidebands such that no single frequency component can be reliably distinguished from the noise with a power threshold.

After digital beamforming, we apply a short-time Fourier transform (STFT) to the summed time-series to obtain the frequency spectrum representation of the signals (see Figure 4.31). From the detection perspective, the frequency representation of the CRES data is advantageous compared to the time domain, because the frequency spectra of CRES signals are well-approximated by a frequency and amplitude modulated sinusoidal whose carrier frequency increases as a linear chirp. The modulation is caused by the axial oscillation of the electron in the magnetic trap and produce frequency spectra that are well-described by a small number of frequency components. The linear chirp is caused by the energy loss due to cyclotron radiation, which results in a relatively slow increase in the frequency components of the CRES signal over time. During the standard Fourier analysis window for the FSCD of 40.96 μsec, we expect a typical CRES signal to increase in frequency by approximately 15 kHz, which is smaller than the frequency bin width given the 200 MHz sample rate. Therefore when considering a single frequency spectrum it is justifiable to neglect the effects of the linear frequency chirp.

In the cases where the electron's pitch angle is $\gtrsim 88.5°$, the majority of the signal power is contained in a single frequency component, with the remaining signal power contained in a small number of sidebands proportional to the electron's axial modulation (see Figure 4.31). In these cases detection is relatively straight-forward by implementing a power threshold on the STFT, since the amplitude of the main signal peak is distinct from the thermal noise spectrum. However, as the pitch angle of the electron is decreased below 88.5°, the modulation index of the signal increases causing the maximum amplitude of the frequency spectrum to be comparable to typical noise fluctuations. At this point, the power threshold trigger is no longer able to distinguish between signal and noise leading to a reduction in detection efficiency. The neutrino mass sensitivity of the FSCD is directly linked to the overall detection efficiency. And, because the distribution of electron pitch angles is effectively uniformly distributed across the range of pitch angles that can be trapped, the overall detection efficiency is directly influenced by the range of pitch angles that have detectable signals. Therefore, utilizing a signal detection algorithm that can more effectively identify signals with pitch angles less than 88.5° will improve both detection efficiency and ultimately the neutrino mass sensitivity of the FSCD and other CRES experiments.

Modeling the detection performance of alternative signal detection algorithms for the FSCD requires that we pose the signal detection problem in a consistent manner. The approach we take is to perform a binary hypothesis test on the frequency spectra generated by the STFT. Mathematically, this is expressed as,

$$\mathcal{H}_0 : y[n] = \nu[n] \tag{4.43}$$

$$\mathcal{H}_1 : y[n] = x[n] + \nu[n]. \tag{4.44}$$

Where under hypothesis $\mathcal{H}_0$, the vector representing the frequency spectrum ($y[n]$) is composed of pure white Gaussian noise (WGN) represented by $\nu[n]$, and under hypothesis $\mathcal{H}_1$ the frequency spectrum is composed of a CRES signal ($x[n]$) with added WGN. The dominant source of noise in a FSCD-like experiment is expected to be thermal Nyquist-Johnson noise, which is well approximated by a WGN distribution. In order to decide between these two hypotheses we follow the standard Neyman-Pearson approach by performing a log-likelihood ratio test between the probability distributions of the signal classifier output under $\mathcal{H}_1$ and $\mathcal{H}_0$ [69]. The output of the log-likelihood ratio test is called the test statistic, which is used to assign the data as belonging to the noise ($\mathcal{H}_0$) or signal ($\mathcal{H}_1$) classes by setting a decision threshold on the value of the test statistic.

In practice, we select the decision threshold by finding the value of the test statistic that guarantees an acceptable rate of false positives and then attempt to maximize the signal detection probability under that fixed false positive rate. Because the signal classifier will be used to evaluate the spectra of $O(10^2)$ beamforming positions every 40.96 μsec, we will require the signal classifiers to operate with decision thresholds that provide false positive rates significantly smaller than 1%. This reduces the burden placed on later stages of the CRES reconstruction chain to reject these false positives and decreases the overall likelihood of reconstructing a false event. Below, we calculate the probability distributions that allow us characterize how different detection algorithms will perform for CRES signals in an FSCD experiment.

4.4.3 Signal Detection Algorithms

4.4.3.1 Power Threshold

The power threshold detection algorithm uses the maximum amplitude of the frequency spectra as the detection test statistic. To model the performance of this approach, consider first the case where the signal is pure WGN. For a single bin in the frequency spectrum, the probability that the amplitude falls below a specific threshold value is given by the Rayleigh cumulative distribution function (CDF),

$$\mathrm{Ray}(x;\tau) = 1 - \exp\left(-|x|^2/\tau\right), \tag{4.45}$$

where the complex amplitude of the frequency bin is x, and τ is the WGN variance. Because the noise samples for each frequency bin are independent and identically distributed (IID), the probability that every bin in the frequency spectrum falls below the threshold is the joint CDF formed by the product of each individual frequency bin CDF,

$$F_0(x;\tau,N_{\mathrm{bin}}) = \mathrm{Ray}(x;\tau)^{N_{\mathrm{bin}}}. \tag{4.46}$$

The PDF for the power threshold classifier can then be obtained by differentiating the CDF.

The probability distribution for the power threshold classifier under $\mathcal{H}_1$ is formed in a similar way, but the frequency bins that contain signal must be treated separately. For a frequency bin that contains both signal and noise we can describe the probability that the amplitude of the bin will fall below our threshold using the Rician CDF,

$$\mathrm{Rice}(x;\tau,\nu) = 1 - Q_1\left(\frac{|\nu|}{\sqrt{2\tau}}, \frac{|x|}{\sqrt{2\tau}}\right), \tag{4.47}$$

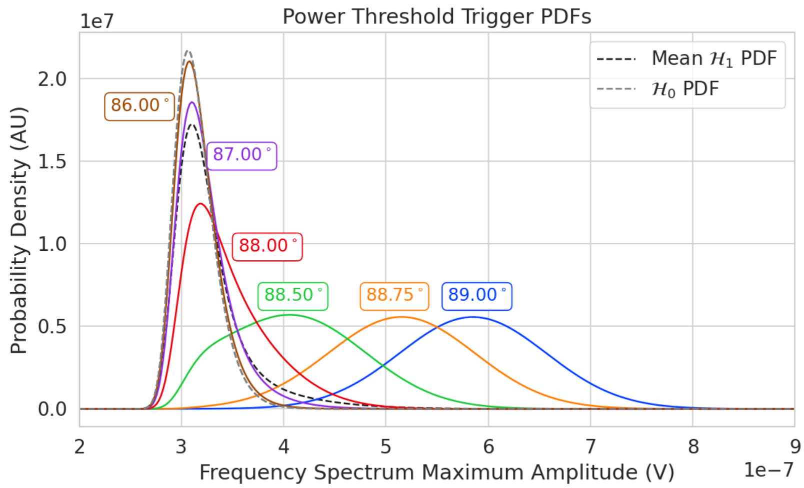


Figure 4.32: PDFs of the power threshold test statistic for CRES signals with various pitch angles as well as the PDF for the noise-only signal case. The average PDF computed for pitch angles ranging from 85.5 to 88.5° is also shown. As the pitch angle is decreased the signal PDF converges towards the noise PDF which indicates that the power threshold trigger is unable to distiguish between small pitch angle signals and noise.

where the parameter $|\nu|$ defines the noise-free amplitude of the signal and Q_1 is the Marcum Q-function. This time the CDF that describes the probability that the entire spectrum falls below the decision threshold is the product of both signal and noise CDFs,

$$F_1(x;\tau,\nu,N_{\text{bin}},N_s) = \text{Ray}(x;\tau)^{N_{\text{bin}}-N_s} \prod_{k=0}^{N_s} \text{Rice}(x;\tau,\nu_k). \tag{4.48}$$

The first half of Equation 4.48 is the contribution from the bins in the frequency spectrum that contain only noise, and the second half is the product of the Rician CDFs for the frequency bins that contain signal peaks with a noise-free amplitude of $|\nu_k|$. In Figure 4.32 we show plots of example PDFs under $\mathcal{H}_1$ and $\mathcal{H}_0$.

4.4.3.2 Matched Filtering

The shape of a CRES signal is completely determined by the initial conditions of the electron as it is emitted from beta-decay, which implies that it is possible to apply matched filtering as a signal detection algorithm. With a matched filter one uses the shape of the known signal, which is called a template, to filter the incoming data by computing the convolution between the signal and the data [69]. For cases where the signal is buried in WGN, the matched filter is the optimal detector in that it achieves the maximum probability of a true detection for a fixed false positive rate. Since CRES signals have an unknown shape but are deterministic, we can apply a matched filter by using simulations to generate a large number of signal templates called a template bank, which spans the parameter space of possible signals. Then at detection time, we use the template bank to identify signals by performing the matched filter convolution for each template in an exhaustive search.

As we saw from the frequency spectra in Figure 4.31, CRES signals are highly periodic in nature. In such cases, it is advantageous to utilize the convolution theorem to replace the matched filter convolution with an inner product in the frequency-domain. With the convolution theorem, the matched filter test statistic that describes the detection of a signal buried in WGN using a matched filter template bank is given by

$$\mathcal{T} = \max_{\mathbf{h}} \left| \sum_{n=0}^{N_{\text{bin}}} h^{\dagger}[n] y[n] \right|, \tag{4.49}$$

where $h^{\dagger}[n]$ is the complex conjugate of the signal template. For the case when our template bank consists of only a single template it is possible to derive an exact analytical form for the PDF describing the matched filter test statistic. First, we derive PDF under the signal hypothesis, where the equation describing the matched filter test statistic, also known as the matched filter score, becomes

$$\mathcal{T} = \left| \sum_{n=0}^{N_{\text{bin}}} h^{\dagger}[n] y[n] \right|. \tag{4.50}$$

Each noisy frequency bin represented by $y[n]$ is the sum between value of the signal at that bin and complex WGN, which means that $y[n]$ is itself Gaussian distributed. Therefore, the value of the inner product between the template and the data is also a complex Gaussian variable; and, since the matched filter score is the magnitude of this inner product, it must follow a Rician distribution.

We can derive the equation for the Rician PDF by expressing the matched filter template $\mathbf{h}$ in terms of the corresponding simulated signal, which we write as $\mathbf{x}_h$ to distinguish from the signal in the data. Using the standard normalization and assuming uncorrelated WGN, the matched filter templates can be written as

$$\mathbf{h} = \frac{\mathbf{x}_h}{\sqrt{\tau |\mathbf{x}_h|^2}} \tag{4.51}$$

where τ is the noise variance. Inserting this into Equation 4.49 and expressing the data as a sum between a signal and a WGN vector yields,

$$\mathcal{T} = \frac{1}{\sqrt{\tau|\mathbf{x}_h|^2}} \left| \sum_{n=1}^{N_{\text{bin}}} x_h[n] \left(x[n] + \nu[n] \right) \right|. \tag{4.52}$$

Next, we transform the expression by isolating the randomly distributed components giving

$$\mathcal{T} = \frac{\left| \sum_{n=1}^{N_{\text{bin}}} x_h[n] x[n] \right|}{\sqrt{\tau|\mathbf{x}_h|^2}} + \frac{1}{\sqrt{\tau|\mathbf{x}_h|^2}} \left| \sum_{n=1}^{N_{\text{bin}}} x_h[n] \nu[n] \right|. \tag{4.53}$$

The first term of 4.53 can be simplified by using the Cauchy-Schawrz inequality to express the magnitude of the inner product in terms of the magnitudes of the signal and template as well as an orthogonality constant which we call "match" (Γ). Using this we obtain,

$$\mathcal{T} = |\mathbf{h}||\mathbf{x}|\Gamma + \frac{1}{\sqrt{\tau|\mathbf{x}_h|^2}} \left| \sum_{n=1}^{N_{\text{bin}}} x_h[n] \nu[n] \right|. \tag{4.54}$$

The second term is a sum of Gaussian distributed variables, which we should expect also follows a Gaussian distribution. Each of the samples $\nu[n]$ is described by

$$\nu[n] \sim \mathcal{N}(0, \tau), \tag{4.55}$$

where $\mathcal{N}(0, \tau)$ is a complex Gaussian distribution with zero mean and variance τ. Therefore,

$$\frac{x_h[n]}{\sqrt{\tau|\mathbf{x}_h|^2}} \nu[n] \sim \mathcal{N}\left(0, \frac{x_h[n]^2}{|\mathbf{x}_h|^2}\right), \tag{4.56}$$

$$\sum_{n=1}^{N_{\text{bin}}} \frac{x_h[n]}{\sqrt{\tau|\mathbf{x}_h|^2}} \nu[n] \sim \mathcal{N}\left(0, \frac{\sum_{n=1}^{N_{\text{bin}}} x_h[n]^2}{|\mathbf{x}_h|^2}\right) = \mathcal{N}(0, 1), \tag{4.57}$$

$$|\mathbf{h}||\mathbf{x}|\Gamma + \sum_{n=1}^{N_{\text{bin}}} \frac{x_h[n]}{\sqrt{\tau|\mathbf{x}_h|^2}} \nu[n] \sim \mathcal{N}(|\mathbf{h}||\mathbf{x}|\Gamma, 1). \tag{4.58}$$

We see that $\mathcal{T}$ is magnitude of a complex variable with mean $|\mathbf{h}||\mathbf{x}|\Gamma$ and variance one. In order to simply the expression a bit further, we define the quantity $\mathcal{T}_{\text{ideal}} = |\mathbf{h}||\mathbf{x}|\Gamma$, which we call the ideal matched filter score, because it represents the value of the matched filter inner product that we would expect if no noise was present in the signal. We can write the matched filter test statistic as the magnitude of a two-dimensional vector in the complex plane

$$\mathcal{T} = \left| \left(\mathcal{T}_{\text{ideal}} + n_r, n_i \right) \right|, \tag{4.59}$$

where n_r and n_i are the real and imaginary components of the noise each with variance 1/2, which is modeled by a Rician distribution with shape factor $\mathcal{T}_{\text{ideal}}$. Therefore, the probability distribution of the matched filter test statistic is given by,

$$P_1(x; \mathcal{T}_{\text{ideal}}) = 2x \exp\left(-\left(x^2 + \mathcal{T}_{\text{ideal}}^2\right)\right) I_0\left(2x\mathcal{T}_{\text{ideal}}\right), \tag{4.60}$$

where I_0 is the zeroth-order modified Bessel function.

The shape of the matched filter score distribution is controlled by the parameter $\mathcal{T}_{\text{ideal}}$, which is effectively the value of the matched filter score if the data contained no noise. Without noise, the data vector reduces to the signal, $\mathbf{x}$, in which case Equation 4.50 becomes the magnitude of an inner product between two vectors. We can write the magnitude of an inner product in terms of the lengths of the individual vectors and a constant that describes the degree of orthogonality between them. Applying this to Equation 4.50, we obtain

$$\mathcal{T}_{\text{ideal}} = \left| \mathbf{h}^\dagger \cdot \mathbf{x} \right| = \left| \mathbf{h} \right| \left| \mathbf{x} \right| \Gamma \tag{4.61}$$

where Γ describes the orthogonality between $\mathbf{h}$ and $\mathbf{x}$. From the point of view of matched filtering, we can interpret Γ as describing how well the template matches the underlying signal in the data.

The matched filter score PDF under the noise hypothesis can be readily obtained from Equation 4.60 by setting the value of $\mathcal{T}_{\text{ideal}}$ to zero, since the data contains no signal in the noise case. Doing this, we obtain the Rayleigh distribution that describes the matched filter score under $\mathcal{H}_0$,

$$P_0(x) = 2x \exp\left(-x^2\right). \tag{4.62}$$

Equations 4.60 and 4.62 describe the behavior of the matched filter test statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$ for a single template. However, defining a PDF that describes the matched filter test statistic in the case of multiple templates is in general a mathematically intractable problem, since there is no guarantee of orthogonality between matched filter templates. This leads to correlations between the matched filter scores of different templates because only one sample of noise is used to compute the matched filter scores of the template bank. In order to proceed, we need to make the simplifying assumption that we can treat the matched filter scores as IID variables, which allows to ignore correlations between templates. The overall effect of this will be an underestimate of the performance of the matched filter, since we are under counting the number of templates that could contribute a detectable score.

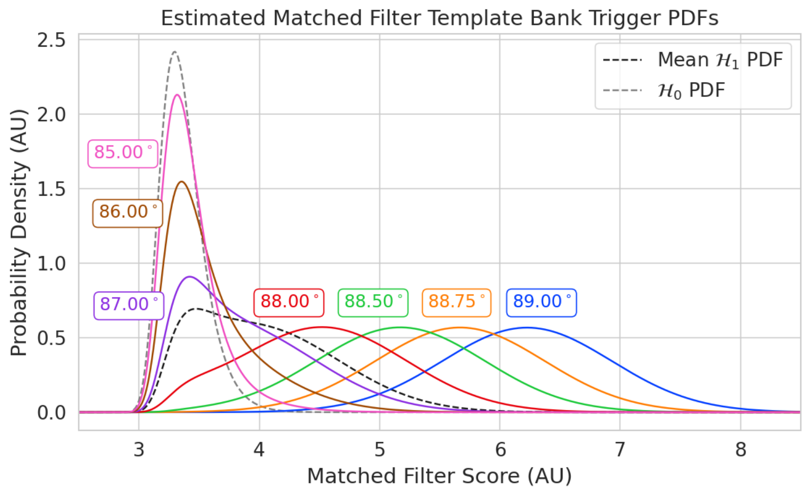


Figure 4.33: Plots of the estimated PDFs for the matched filter template bank test statistic for CRES signals with various pitch angles as well as the estimated PDF for the noise only signal case. We assume an estimated number of templates of 10^5 and perfect match between signal and template i.e. $\Gamma_{\text{best}} = 1$. The mean PDF includes signals ranging from $85.5 - 88.5°$ in pitch angle. There is a much larger distinction between the signal PDFs at small pitch angle compared to the power threshold indicating a higher detection efficiency for these signals.

For $\mathcal{H}_0$ we model the probability that the matched filter score falls below our threshold using the CDF obtained by integrating Equation 4.62. Because we are assuming that the matched filter scores using different templates are independent, the probability that the matched filter score for all templates falls below a threshold value is the joint CDF formed by multiplying the CDF for each template. Under $\mathcal{H}_0$ this is

$$F_0(x) = \left(1 - e^{-x^2}\right)^{N_t}, \tag{4.63}$$

where x is the matched filter score threshold and N_t is the number of templates. We should expect that the distribution describing the matched filter template bank maximum score depends on N_t, because with more templates there is a greater chance of a random match between the template and data.

For $\mathcal{H}_1$, we start by denoting the CDF of the best matching template as $F_{\text{best}}(x; \mathcal{T}_{\text{best}})$, and treat the matched filter scores for all other templates as negligible ($\mathcal{T}_{\text{ideal}} \approx 0$). Then we form the joint CDF by combining the distributions for all templates used during detection. Since we are exhaustively checking the matched filter scores, the number of templates checked will be a randomly distributed variable that ranges from zero to the total number of available templates. If we assume that signals are uniformly distributed across the parameter space spanned by the template bank then on average we check $(N_t - 1)/2 \approx N_t/2$ templates for each inference. Therefore, the estimated CDF under $\mathcal{H}_1$ is

$$F_1(x; \mathcal{T}_{\text{best}}) = F_{\text{best}}(x; \mathcal{T}_{\text{best}}) \left(1 - e^{-x^2}\right)^{N_t/2}. \tag{4.64}$$

In Figure 4.33 we show plots of the estimated matched filter template bank classifier PDFs under both $\mathcal{H}_0$ and $\mathcal{H}_1$.

4.4.3.3 Machine Learning

In this paper we focus on Convolutional Neural Networks (CNN) as an example of a machine learning based signal classifier. CNNs are constructed using a series of convolutional layers, each composed of a set of filters that are convolved with the input data. The individual convolutional filters can be viewed as matched filter templates that are learned from a set of simulated data rather than being directly generated. This opens the possibility of finding a more efficient representation of the matched filter templates during the training process that can potentially reduce computational cost at inference time while still offering good classification performance.

The machine learning approach is distinct from both the power threshold and matched filtering in that we do not attempt to manually engineer a test statistic that is computed from the data for classification. Instead, we attempt calculate the test statistic by constructing a differentiable function that maps the complex frequency series generated by the STFT to a binary classification as either signal or noise. The test statistic for the machine learning classifier can be expressed as

$$\mathcal{T} = G(\mathbf{y}; \mathbf{\Omega}) \tag{4.65}$$

where $\mathbf{y}$ is the noisy data vector and $G(\mathbf{y}; \mathbf{\Omega})$ is the machine learning model parameterized by the weights $\mathbf{\Omega}$. By using supervised learning on a labeled set of training signals, we can modify the function parameters to learn the mapping from the data to the likelihood of $\mathbf{y}$ belonging to either $\mathcal{H}_1$ or $\mathcal{H}_0$.

The CNN architecture used for this work is summarized by Table 4.1. No strategic

Table 4.1: A summary of the CNN model layers and parameters. The output of each 1D-Convolution and Fully Connected layer is passed through a LeakyReLU activation function and re-normalized using batch normalization before being passed to the next layer in the model. The output of the final Fully Connected layer in the model is left without activation so that the model outputs can be directly passed to the Binary Cross-entropy loss function used during training.

Layer	Type	Input Channels	Output Channels	Parameters
1	1D-Convolution	2	15	$(N_{\text{kernel}} = 4,\ N_{\text{stride}} = 1)$
2	Maximum Pooling	15	15	$(N_{\text{kernel}} = 4,\ N_{\text{stride}} = 4)$
3	1D-Convolution	15	20	$(N_{\text{kernel}} = 4,\ N_{\text{stride}} = 1)$
4	Maximum Pooling	20	20	$(N_{\text{kernel}} = 4,\ N_{\text{stride}} = 4)$
5	1D-Convolution	20	25	$(N_{\text{kernel}} = 4,\ N_{\text{stride}} = 1)$
6	Maximum Pooling	25	25	$(N_{\text{kernel}} = 4,\ N_{\text{stride}} = 4)$
7	Fully Connected	3200	512	NA
8	Fully Connected	512	64	NA
9	Fully Connected	64	2	NA

hyper-parameter optimization approach was implemented beyond the manual testing of different CNN architecture variations, so this particular model is best viewed as a proof-of-concept rather than a rigorously optimized design. Numerous model variations were tested, some with significantly more layers and convolutions filters per layer, as well as others that were even smaller than the architecture in Table 4.1. Ultimately, the model architecture choice was driven by the motivation to find the minimal model whose classification performance was still comparable to the larger CNN's tested, because of the importance of minimizing computational cost in real-time applications. It is possible that more sophisticated machine learning models could improve upon the classification results achieved here, but we leave this investigation for future work.

4.4.4 Methods

4.4.4.1 Data Generation

To test the triggering performance of the classifiers, simulated CRES signals were generated using the Locust simulations package [60, 78] developed by the Project 8 collaboration. Locust uses the separately developed Kassiopeia package to calculate the magnetic fields produced by a user defined set of current carrying coils along with any specified background magnetic fields, resulting in a magnetic trap. Next, Kassiopeia calculates the trajectory of an electron in this magnetic field starting from a set of user specified initial conditions. The Locust software then uses the electron trajectories from Kassiopeia to calculate the resulting electromagnetic fields using the Liénard-Wiechert equations, and determine the voltages generated in the antenna array with the antenna transfer function. Locust then simulates the down-conversion, filtering, and digitization steps resulting in the simulated CRES signals for an electron.

The shape of the received CRES signal is determined by the initial kinematic parameters, including the starting position of the electron, the starting kinetic energy of the electron, and the pitch angle. For the studies performed here we constrain ourselves to a single initial electron position located at $(x, y, z) = (5, 0, 0)$ mm, and using this starting position we generate two datasets by varying the initial kinetic energy and the starting pitch angle. The first dataset consists of a two-dimensional square grid of kinetic energy and pitch angle spanning an energy range from 18575-18580 eV with a spacing of 0.1 eV, and pitch angles from 85.5-88.5° with a spacing of 0.001°, resulting in 153051 signals with a unique energy-pitch angle combination. This dataset is intended to represent a matched filter template bank. The second dataset was generated by randomly sampling uniform probability distributions covering the same parameter space to produce approximately 50000 signals randomly parameterized in energy and pitch angle. This dataset provides the training and test data for the machine learning approach, and acts as a representative sample of signals to evaluate the performance of the matched filter template bank.

Each signal was simulated for a duration of 40.96 μs, which is equivalent to 8192 samples at the FSCD digitization rate, and begins at time $t = 0$ s for all simulations. This duration represents a single frequency spectrum generated by the STFT. The output of the Locust simulation is a matrix of array snapshots with size given by the number of channels times the event length ($N_{\text{ch}} \times N_{\text{sample}}$), which we pre-process using the digital beamforming summation and STFT described in Section 4.4.2.2. The ∇B-drift correction uses the exact value of $\omega_{\nabla B}$, obtained from the Kassiopeia simulation of that electron. In practice, an average value for $\omega_{\nabla B}$ could be used, because there is limited variation in drift frequency across this parameter space.

4.4.4.2 Template Number and Match Estimation

The estimated PDF for the matched filter template bank depends on the score of the best matching template or equivalently the match of the best template (Γ_{best}) as well as the number of templates. One expects that with a higher number of templates the average value of Γ_{best} will increase, however, there is a point of diminishing returns at which more templates will not significantly increase match, but will still increase the likelihood of false positives. Therefore, it is desirable to use the minimum number of templates that provide an acceptable mean value of Γ_{best}.

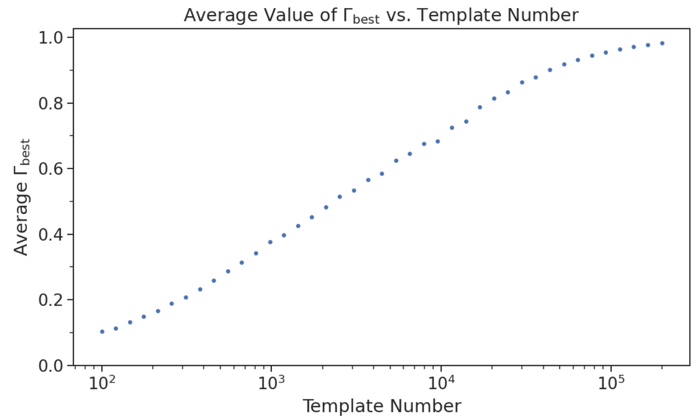


Figure 4.34: The mean match of the matched filter template bank to a test set of randomly parameterized signals as a function of the number or density of templates. The parameter space includes pitch angles from $85.5 - 88.5°$ and energies from $18575 - 18580$ eV.

To quantify the relationship between match and template number, we calculated the mean match of the random dataset to a selection of templates obtained from the regularly spaced dataset. The results are shown in Figure 4.34, where we find that the average value of Γ_{best} is an exponential function of the number of templates. From this plot we select the desired value of mean match at which we would like to evaluate the matched filter PDF and can infer the required number of templates.

4.4.4.3 CNN Training and Data Augmentation

To prepare the data for training the model, we split the random dataset in half to create distinct training and test datasets. Additionally, a randomly selected 20% of the training data is isolated for use as a validation set during the training loop. The size of the training, validation, and test datasets are then tripled by appending two additional copies of the data to increase the sample size of the dataset after data augmentation. The data is loaded with no noise, which is added to each data batch during the training phase by generating a new noise sample from a complex WGN distribution. In order to ensure an even split between signal and noise data we append to the noise-free signals an equal number of empty signals composed of all zeros. Therefore, as the data is randomly shuffled during training, on average an equal number of empty signals will be included with the training signals. After adding the sample of WGN to the data batch, the empty signals represent the noise-only data that the model must distinguish from signal data.

As the training signals are loaded we apply a unique random phase shift as the first form of data augmentation. Since the data is generated using the same initial axial position and cyclotron orbit phase, the randomization is an attempt to prevent overtraining on these features. During each training epoch the data is randomly shuffled and split into batches of 2500 signals. Each batch of signals is then circularly shifted by a random number of frequency bins to simulate a kinetic energy shift from -75 to 20 eV to simulate a training dataset with a larger energy range. Next, a sample of complex WGN, consistent with the expected 10 K Nyquist-Johnson noise expected for the FSCD, is generated and added to the signal, which prevents overtraining on noise features. As a final step, the data is renormalized by the standard deviation of the noise so that the range of values in the data is close to $[-1, 1]$, which helps ensure well-behaved back-propagation.

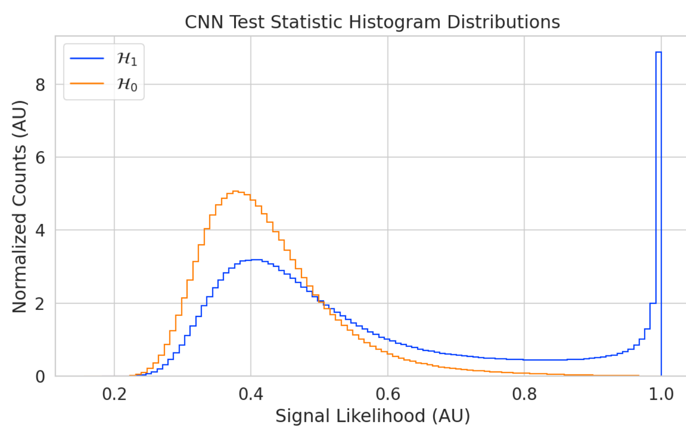


Figure 4.35: Histograms of the trained CNN model output from the test dataset. The blue histogram shows the model outputs for signal data. The oddly shaped peak near the end is the result of the softmax function mapping the long tail of the raw output distribution to the range $[0, 1]$.

The Binary Cross-entropy loss function is used to compute the loss for each batch of data and the model weights are updated using the ADAM optimizer with a learning rate of 5×10^{-3}. After each training epoch, the loss and classification accuracy of the validation dataset are computed to monitor for overtraining. It was noticed that the relatively high noise power and the fact that a new sample of noise was used for each batch together provided a strong form of regularization, since no evidence of over-training was observed even after several thousand epochs. Typically, the loss and classification accuracy of the model converged after a few hundred training epochs, but the training loop was extended to 3000 epochs to attempt to achieve the best possible performance. The training procedure generally took about 24 hrs using a single NVIDIA V100 GPU [80].

After training the model, we use it to classifying the test dataset and generate histograms of the model outputs for both classes of data. The data augmentation procedure for the evaluation of the test data mirrors the training procedure without the validation split. Since a random circular shift and a new sample of WGN is added to each batch, the testing evaluation loop is run for 100 epochs to get a representative sample of noise and circular shifts. The model outputs for each batch are passed through a softmax activation and then combined into histograms, which we show in Figure 4.35.

4.4.5 Results and Discussion

4.4.5.1 Trigger Classification Performance

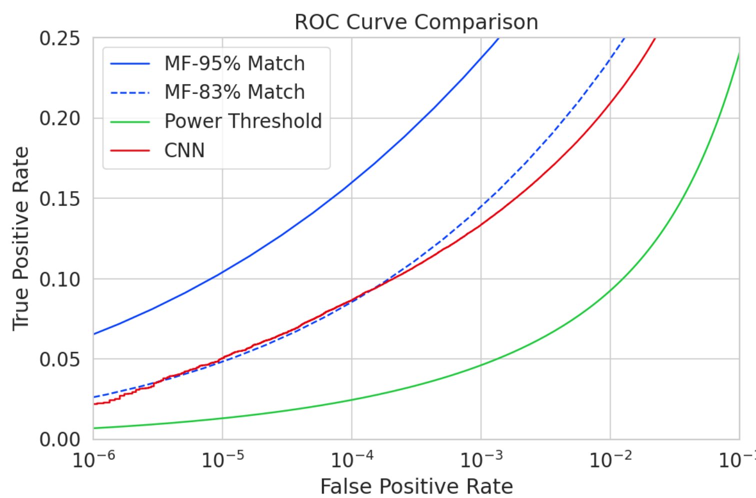


Figure 4.36: ROC curves describing the detection efficiency or true positive rates for the three signal classification algorithms examined in this paper.

Using the matched filter and power threshold CDFs, along with the classification results from the CNN, we compare detection performance by computing receiver operating characteristic (ROC) curves. Specifically, we compare the detection performance averaged over the full signal parameter space in order to get a measure of the overall detection efficiency achieved by each algorithm. For the power threshold and matched filter algorithms, we obtain the mean ROC curve by taking the average over all signals in the regularly spaced dataset. In the case of the matched filter, we examine two cases using different numbers of templates, which have different values of mean match. The ROC curve describing the CNN is obtained by forming a histogram of the network outputs for each class of signal and from this computing the estimated CDFs and ROC curve. In Figure 4.36, we show the ROC curves obtained for each of the detection algorithms, visualized in terms of true positive rate and false positive rate.

The true positive rate of a signal classifier is equivalent to its detection efficiency, and we see that for the population of signals with pitch angles $< 88.5°$ the power threshold has a consistently lower detection efficiency than the CNN and the matched filter. This result could have been predicted from the visualization of signal spectra in Figure 4.31, where we see that there is no way to distinguish between a noise peak and a signal peak with high confidence at small pitch angles. The CNN offers a significant and consistent increase in detection efficiency over the power threshold approach, with the relative improvement in detection efficiency increasing as the false positive rate decreases. If we compare the CNN to the matched filter, we see that the performance of the tested network is roughly equivalent to a matched filter detector with an average match of about 83%, which uses approximately 20000 matched filter templates. The overall best detection efficiency is achieved by the matched filter classifier if a large enough template bank is used. We show in the plot the ROC curve for a matched filter template bank with 95% average match, which is achieved with approximately 100000 templates. Since the matched filter is known to be statistically optimal for detecting a known signal in WGN, it is somewhat expected that this algorithm has the highest detection efficiency.

A potentially impactful difference between the matched filter and CNN algorithms is that the CNN relies upon convolutions as its fundamental calculation mechanism, whereas our implementation of a matched filter utilizes an inner product. Since convolution is a translation invariant operation, the detection performance of CNN can be extended to a wider range of CRES event kinetic energies with less cost than the matched filter, a feature that we exploited during the CNN training by including circular translations of the CRES frequency spectra in the training loop. Increasing the range of kinetic energies detectable by a matched filter requires a proportional increase in the number of templates, which directly translates into increased computational and hardware costs. From a practical perspective, the detection algorithm is always limited by the available computational hardware, so estimating the relative costs is a key factor in determining their feasibility. Below we perform a more detailed analysis of the relative costs of each of the detection algorithms.

4.4.5.2 Computational Cost and Hardware Requirements

In the process of investigating triggering approaches for an antenna array CRES experiment, we have uncovered a strong tension between detection efficiency and computational resources. To relate the computational cost estimates to actual costs, we compare the theoretical amount of computer hardware required to implement the signal classifiers for real-time detection in an FSCD experiment. To do this we shall utilize order of magnitude estimates of the theoretical peak performance values for currently available Graphics Processing Units (GPUs) as a metric. This approach will underestimate the amount of required hardware, since it is unlikely that any CRES detection algorithm could reach the theoretical peak performance of the hardware.

Of the three detection algorithms tested, the power threshold classifier is the least expensive. It requires that we check whether the amplitude of each frequency bin in the STFT is below or above our decision threshold. The STFT combined with digital beamforming produces $N_{\mathrm{bin}}N_{\mathrm{b}}$ frequency bins that must be checked every $N_{\mathrm{bin}}/f_{\mathrm{s}}$ seconds. This requires approximately $O(10^{10})$ FLOPS to check in real-time. Current generations of GPUs have peak theoretical performances in the range of $O(10^{13}) - O(10^{14})$ FLOPS [81], dependent on the required floating-point precision of the computation. Therefore, the entire computational needs of a real-time triggering system using a power threshold classifier, including digital beamforming and generation of the STFT, could be met by a single high-end GPU or a small number of less powerful GPUs. Since triggering is only one step of the full real-time signal reconstruction approach, limiting the computational cost of this stage is ideal. However, we have seen that the power threshold classifier does not provided sufficient detection efficiency across the entire range of possible signals, which is the primary motivation for exploring more complicated triggering solutions.

As discussed, the computational cost of the matched filter approach requires counting the number of templates that must be checked for each frequency spectra produced by the STFT. Computing the matched filter scores requires $O(N_{\mathrm{b}}N_{\mathrm{t}}N_{\mathrm{bin}})$ operations, since for each of the N_{b} beamforming positions we must multiply N_{t} templates with a data vector that has length N_{bin}. The time within which we must perform this calculation is equal to $N_{\mathrm{bin}}/f_{\mathrm{s}}$ to keep up with the data generation rate. To cover the 5 eV kinetic energy range spanned by the template bank, we saw that 10^4 to 10^5 templates are required in order to match or exceed the detection efficiency of the CNN. If the number of templates scales linearly with then kinetic energy range of interest as expected, then we would require 10^5 to 10^6 matched filter templates with this more realistic range of energies. Considering this, the estimated computational cost of the matched filter is between

$O(10^{15})$ to $O(10^{16})$ FLOPS, which is $O(10^2)$ to $O(10^3)$ high-end GPUs.

Lastly, we have the CNN classifier. To estimate the computational cost we simply sum the number of convolutions and matrix multiplications specified by the network architecture shown in Table 4.1. Each convolutional layer consists of $N_{\text{in}}N_{\text{out}}N_{\text{kernel}}L_{\text{input}}$ floating-point operations, where N_{in} is the number of input channels, N_{out} is the number of output channels, N_{kernel} is the size of the convolutional kernel, and L_{input} is the length of the input vector, and the fully connected layers each contribute $N_{\text{in}}N_{\text{out}}$ operations. Summing all the neural network layers we estimate that the CNN would require $O(10^6)$ floating point operations for each frequency spectra; therefore, the total computation cost of the CNN trigger is this cost times the number of beamforming positions per the data acquisition time, which is $O(10^{13})$ FLOPS or $O(10^0)$ GPUs.

Compared with the matched filter approach the CNN requires $O(100)$ to $O(1000)$ fewer GPUs to implement, dependent on the exact number of templates used in the template bank. The 100 eV kinetic energy range is motivated by the application of these detection algorithms to an FSCD-like neutrino mass measurement experiment. However, if a significantly larger range of kinetic energies is required, a CNN may be the preferred detection approach despite the lower average detection efficiency due to computational cost considerations. The low estimated computational cost of the CNN is directly related to the small network size.

Additional experiments with larger CNNs, generated by increasing the depth and width of the neural network, and we observed that these changes provided minimal ($\lesssim 1\%$) improvement in the classification accuracy of the model. A potential reason for this could be the sparse nature of the signals in the frequency domain and the low SNR which makes for a challenging dataset to learn from. Future work could investigate modifications to the neural network architecture such as sparse convolutions, which may improve the classification accuracy of the model or further reduce the computational costs of this approach. Alternatively, more complicated CNN architectures such as a ResNet [82] or VGG model [83] may provide improved classification performance over a basic CNN. An additional promising area of investigation are recurrent neural networks, which may be able to exploit the time-ordered features of the STFT for more accurate signal detection if the electron signals last for multiple Fourier transform windows.

Our estimate of the computational cost of the matched filter is somewhat naive if we notice that the majority of the values that make up a CRES frequency spectra are zero (see Figure 4.31). Therefore, the majority of operations in the matched filter inner product are unnecessary, and we could instead evaluate the matched filter inner product using only the $\lesssim 10$ frequency peaks that make up CRES signal. This optimization reduces the number of operations required to check each template by a factor of $O(100)$ to $O(1000)$, which brings the estimated computational cost of the matched filter in line with the CNN. Although this level of sparsity results in a multiplication with very low arithmetic complexity, the resulting sparse matched filter algorithm is still likely to be constrained by memory access speed rather than compute speed. Ultimately, the comparison of the relative computational and hardware costs between the matched filter and CNN will depend on the efficiency of the software implementation and hardware support for neural network and sparse matrix calculations.

4.4.6 Conclusion

Increasing the detection efficiency and overall event rate of the CRES technique represents a key developmental path towards new scientific results and broader applications of the CRES technique. It is what motivates both the antenna array detection approach and the development of real-time signal reconstruction algorithms. We have demonstrated that significant gains in the detection efficiency of the CRES technique are achievable by utilizing triggering algorithms that account for the specific shape of CRES signals in the detector. These algorithms emphasize the need for accurate and fast methods for CRES simulation, since they directly contribute to the success of matched filter methods by providing a way to generate expected signal templates and also serve as a source of training data for machine learning approaches.

The improvements in detection efficiency offered by these alternative approaches to triggering are crucial to the success of efforts to develop scalable technologies for CRES measurement, since they provide a significant increase in the detectable parameter space of CRES events, which allows for a better utilization of the larger detection volume. While we have focused on the real-time detection of CRES signals from antenna arrays, these same signal classifiers could be used in CRES experiments utilizing a different detector technologies, since the same principles of signal detection will apply. For example, previous CRES measurements by the Project 8 collaboration that utilized a waveguide gas cell, could have improved their detection efficiency by employing a matched filter or neural network classifier to identify trapped electrons with pitch angles that are too small to be detected by the power threshold approach. Furthermore, alternative CRES detector technologies such as resonant cavities [39] could also see similar improvements in detection efficiency, which is of crucial importance to future efforts by the Project 8 collaboration to utilize CRES to measure the neutrino mass.

Chapter 5

Antenna and Antenna Measurement System Development for the Project 8 Experiment

5.1 Introduction

The FSCD and antenna array CRES represent an innovative approach to beta-decay spectroscopy. While much can be learned from simulations about the systematics of CRES with antenna arrays, laboratory measurements and demonstrations provide critical inputs to sensitivity and simulation models as well as provide a means for calibration and commissioning of the experiment. Therefore, a robust program of antenna and antenna measurement hardware development is important to the success of the FSCD and the development of antenna array CRES more broadly.

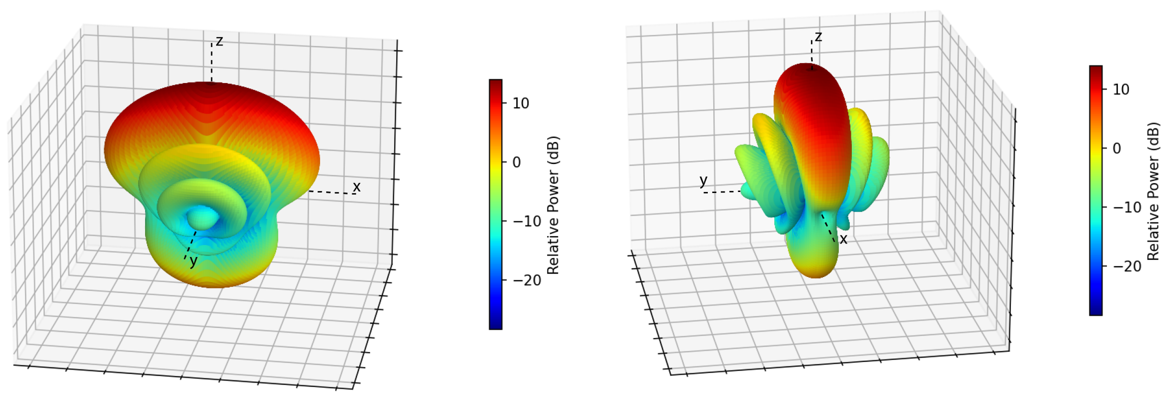
In this chapter we summarize the development of an antenna measurement system at Penn State to implement and test the techniques of antenna array CRES on the bench-top, in order to support the efforts of the Project 8 collaboration. In Section 5.2 we provide an introduction to some fundamental parameters and concepts related to antenna measurements as well as an overview of the Penn State antenna measurement system hardware. In Section 5.3 we include the manuscript of a paper [79] which details the design and characterization of a specialized antenna developed to mimic the electric fields emitted by an electron in a CRES experiment. This antenna, called the Synthetic Cyclotron Antenna (SYNCA), is intended as a calibration tool for antenna arrays developed for CRES measurements. Lastly, in Section 5.5 we summarize a set of prototype FSCD antenna array measurements with the SYNCA [42], which we use to validate the simulated performance of the antenna array and estimate systematic errors associated with the antenna array.

5.2 Antenna Measurements for CRES experiments

5.2.1 Antenna Parameters

Antenna characterization measurements are intended to validate simulations of the antenna array performance, which ultimately informs the neutrino mass sensitivity of the experiment. In this section, I shall summarize a few fundamental concepts relating to antennas and antenna measurement, before introducing how Project 8 uses antenna measurement for the development of antenna array CRES.

5.2.1.1 Radiation Patterns



(a) (b)

Figure 5.1: An example radiation pattern generated using HFSS simulations. The color and radial distance of the surface from the origin indicate the relative magnitude of radiation power emitted by the antenna in that direction. The primary goal of most antenna measurements is typically to measure the antenna pattern, which is used to derive many useful antenna performance parameters.

Antennas are conductive structures designed to carry alternating electric currents in order to transmit energy in the form of electro-magnetic (EM) waves [63]. Perhaps the most fundamental way to characterize an antenna, is to map out the radiated power density as a function of position, which is called the radiation pattern (see Figure 5.1). We find the radiation power density by calculating the time-averaged Poynting vector for all positions surrounding the antenna, which in equation form is

$$\mathbf{W}(x, y, z) = \langle \mathbf{E}(x, y, z, t) \times \mathbf{H}^*(x, y, z, t) \rangle_t \,, \tag{5.1}$$

where $\mathbf{E}(x, y, z, t)$ and $\mathbf{H}(x, y, z, t)$ are the time-dependent electric and magnetic fields produced by the antenna [47]. The radiation power density has units of $\mathrm{W/m^2}$ and is more typically called the energy flux density in physics applications, since it is a measure of the amount of energy passing through a unit area over time.

Because the radiation power density is a measure of power per unit area, its value in a particular direction will depend on the distance from the antenna at which we are measuring. This is undesirable for practical applications A related quantity, which is distance independent, is the energy flux per unit solid angle or radiation intensity, which is computed directly from the radition power density by multiplying by the squared distance from the antenna. Specifically,

$$U = r^2 W(x, y, z), \tag{5.2}$$

where r is the distance from the antenna to the field measurement point. The radiation intensity is typically defined in regions where the Poynting vector consists only of a radial component where it is safe to treat as a scalar quantity.

5.2.1.2 Directivity and Gain

Since the radiation intensity is a measure of average power per unit solid angle, it is independent of distance and more useful as feature for antenna measurement. However, most antenna measurements are performed in terms of the directly related directivity and gain quantities. Directivity is defined as the ratio between the radiation intensity at particular point on the radiation pattern to the average radiation intensity computed over all solid angles [63]. The equation that relates the radiation intensity to directivity is

$$D = \frac{U}{U_0} = \frac{4\pi U}{P_{\mathrm{rad}}}, \tag{5.3}$$

where U_0 is the average radiation intensity over all solid angles, which simply the total radiated power (P_{rad}) divided by 4π. Closely related to directivity is concept of gain, which accounts for energy losses that occur inside then antenna when attempting to transmit or receive a signal. The antenna gain is given by

$$G = \frac{4\pi U}{P_{\mathrm{in}}}, \tag{5.4}$$

where P_{in} is the total power delivered to the antenna. Gain can be thought of as the ratio of the antenna's radiation intensity to that of a hypothetical isotropic, lossless radiator.

The maximum values of gain and directivity exhibited by the main lobe of the antenna pattern as well as the ratio between the gain of the main lobe and any side-lobes are important figures of merit used to evaluate antenna designs.

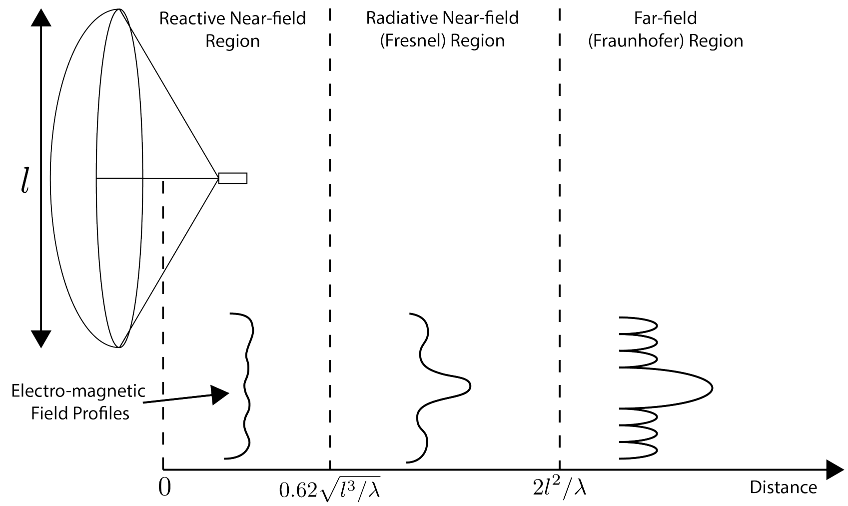


Figure 5.2: An illustration of the three field regions important for the analysis of an antenna system. Very close to the antenna the electric fields are primarily reactive so there is no radiation. If a receiving antenna were placed in this region most of the energy would be reflected back to the transmitter. Outside of the reactive near-field is the radiative near field. At these distances the antenna does radiate, but the radiation pattern is not well-defined since it changes based on the distance of the receiving antenna. It is only in the far-field region where the radiation pattern becomes constant as a function of distance, which is where the majority of antenna engineering is assumed to take place. The antenna arrays developed by Project 8 for CRES measurements operate in the radiative near-field due to the importance of limiting power loss from free-space propagation, which complicates the design of the antenna system.

5.2.1.3 Far-field and Near-field

Radiation patterns are only well-defined in regions where the shape of the radiation pattern is independent of distance. The region where this approximation is valid is called the "far-field", and in this region we can treat the EM fields from the antenna as spherical plane waves. A rule of thumb for antennas is that the far-field approximation is valid when the condition

$$R > \frac{2l^2}{\lambda} \tag{5.5}$$

is met. In this expression, R is the distance from the antenna, l is the largest characteristic dimension of the antenna, and λ is the wavelength of the radiation (see Figure 5.2).

The region very close to the antenna is called the reactive near-field, because in this region the reactive component of the EM field is dominant. Unlike radiative electric fields, the reactive electric and magnetic fields are out of phase from each other by 90°, since they are the result of electrostatic and magnetostatic effects coming from the self-capacitance and self-inductance of the antenna. The reactive fields are unable to transfer energy a significant distance from the antenna and are thus completely negligible for most antenna applications. The limit of the reactive near-field for an electrically-large antenna is typically taken to be

$$R < 0.62\sqrt{l^3/\lambda}. \tag{5.6}$$

The unique application of antennas by Project 8 is somewhat limited by reactive near-field effects in the form of a maximum radial position for electrons inside the uniform cylindrical antenna array. If electrons are too close to the edge of the array than reactive near-field effects leads to a large reduction in the received power and consequently detection efficiency. This leads to a significant volume inside of the antenna array that is unsuitable for CRES lowering the volumetric efficiency of the antenna array CRES technique relative to a cavity experiment.

In between the reactive near-field and the far-field is the radiative near-field region. In this region the fields are primarily radiative, however we are still too close to the antenna for the spherical plane wave approximation to apply. Therefore, interference effects between EM waves emitted from different points on the antenna occur causing the shape of the radiation pattern to change as a function of distance from the antenna. If we evaluate the far-field distance limit for the FSCD one finds an estimated far-field distance of 43 cm, which is a factor of four larger than the radius of the antenna array designed for the experiment. Consequently, we expect near-field effects to influence the performance of the antenna array highlighting the importance of calibration and characterization measurements.

5.2.1.4 Polarization

The polarization of an EM wave defines the spatial orientation of the electric field oscillations in the plane perpendicular to the direction of the propagation, and is defined in terms of orthogonal polarization components. In our application, one analyzes the properties of radiation propagating along the radial ($\hat{r}$) direction away from the antenna, which implies that the electric fields can be described as a linear combination of orthogonal

polarization components
$$\mathbf{E}_{\text{tot}} = E_x\hat{x} + E_y\hat{y} + E_z\hat{z}, \tag{5.7}$$
in Cartesian coordinates, or
$$\mathbf{E}_{\text{tot}} = E_\theta\hat{\theta} + E_\phi\hat{\phi}, \tag{5.8}$$
in spherical coordinates.

In general, one defines partial radiation patterns, directivities, and gains so that the performance of the antenna for the desired polarization can be analyzed. The radiation pattern defined in terms of partial patterns is

$$U_{\text{tot}} = U_\phi + U_\theta, \tag{5.9}$$

where U_ϕ and U_θ are the radiation intensities in a particular direction for the respective polarization components. Similarly, a quantity such as gain can be written in terms of partial gains,
$$G_{\text{tot}} = G_\phi + G_\theta = \frac{2\pi U_\phi}{P_{\text{in}}} + \frac{2\pi U_\theta}{P_{\text{in}}}. \tag{5.10}$$

If we view an electron performing a circular orbit in the XY-plane from the side, that is, along the X or Y axes, then we would observe the electron to be performing a linear oscillation perpendicular to the viewing axis. From this intuitive picture, we can predict that the primary polarization of electric fields from CRES events to be linearly polarized in the $\hat{\phi}$ direction when viewed with an antenna positioned in the XY-plane.

5.2.1.5 Antenna Factor and Effective Aperture

A useful way to characterize the performance of an antenna is to measure the electric field magnitude required to produce a signal with an amplitude of one volt in the antenna terminals. This ratio between the magnitude of the incoming electric field and the magnitude of the signal produced by the antenna is called the antenna factor, which is written as
$$A_{\text{F}} = \frac{|\mathbf{E}_{\text{in}}|}{V_{\text{ant}}}, \tag{5.11}$$
where A_{F} is the antenna factor, E_{in} is the incoming electric field, and V_{ant} is the magnitude of the voltage produced by the antenna.

The antenna factor can be expressed in terms of the antenna's gain through a related quantity called the effective aperture. The effective aperture defines for a given incident radiation power density (W/m^2) the power that is received by the antenna. Therefore, the effective aperture gives the equivalent area of the antenna,

$$A_{\text{eff}} = \frac{P_{\text{rec}}}{P_{\text{in}}} = \frac{\lambda^2}{4\pi}G, \tag{5.12}$$

where the received power is P_r and the total incoming power is P_{in}.

If we express the incident power in terms of the magnitude of the Poynting vector, then

$$|\mathbf{S}_{\text{in}}| = |\mathbf{E}_{\text{in}}|^2/\eta_0, \tag{5.13}$$

where η_0 is the impedance of free-space, which relates the magnitudes of the electric and magnetic fields in a vacuum, and is defined by

$$\eta_0 = \frac{|\mathbf{E}|}{|\mathbf{H}|} = \sqrt{\frac{\epsilon_0}{\mu_0}}. \tag{5.14}$$

The total received power by the antenna can therefore be expressed as

$$P_{\text{rec}} = |\mathbf{S}_{\text{in}}|A_{\text{eff}} = |\mathbf{S}_{\text{in}}|\frac{\lambda^2}{4\pi}G = \frac{|\mathbf{E}_{\text{in}}|^2\lambda^2 G}{4\pi\eta_0}. \tag{5.15}$$

To relate this to the antenna factor recall that we can relate the voltage produced by the antenna to the received power with

$$P_{\text{rec}} = \frac{V_{\text{ant}}^2}{Z} = \frac{|\mathbf{E}_{\text{in}}|^2}{A_{\text{F}}^2 Z}, \tag{5.16}$$

where Z is the system impedance. Setting Equations 5.15 and 5.16 equal to each other, we obtain the following expression for antenna factor in terms of gain

$$A_{\text{F}} = \sqrt{\frac{4\pi\eta_0}{ZG\lambda^2}} = \frac{9.73}{\lambda\sqrt{G}}. \tag{5.17}$$

The second expression in Equation 5.17 is obtained by evaluating the constant terms assuming a system impedance of 50 Ω.

We have gone through the effort of expressing the antenna factor in terms of gain to highlight that the majority of antenna parameters that we care to measure for a CRES experiment can be obtained from the radiation or gain pattern of the antenna. The antenna factor is a particularly important parameter for CRES measurements due to it's relevance to antenna array simulations with the Locust software [60, 78]. Specifically, Locust simulates the trajectory of an electron in a magnetic trap by running the Kassiopeia software package [58] and then uses the Liénard-Wiechert equations [61,62] to calculate the electric fields that are incident on the antenna.

To compute the response of the antenna to the electric field, Locust relies upon linear time-invariant system theory, which computes the response of the antenna (i.e. the voltage time series generated by the antenna) using a convolution between the electric field time-series and the antenna impulse response. This approach is necessary for correctly modeling the antenna response to the electric field due to the broadband and non-stationary nature of the electric fields from CRES events. Since antenna measurements take place under steady-state conditions, parameters such as the radiation pattern, gain, and antenna factor are defined in the frequency domain. However, by performing an inverse Fourier transform on the antenna factor we can obtain the antenna impulse response, which allows us to simulate CRES events in the antenna array demonstrator experiment.

5.2.2 Antenna Measurement Fundamentals

5.2.2.1 Friis Transmission Equation

The antenna factor, sometimes called the antenna transfer function, is used to model how the antenna will respond to electric fields emitted from a CRES event. Therefore, being able to measure the antenna transfer function of the antenna array is a key step in the commissioning and calibration phases of an antenna array CRES experiment. A common approach to antenna characterization is to perform a two antenna transmit-receive measurement where an antenna with a known gain is used to characterize the unknown gain of the antenna under test (see Figure 5.3).

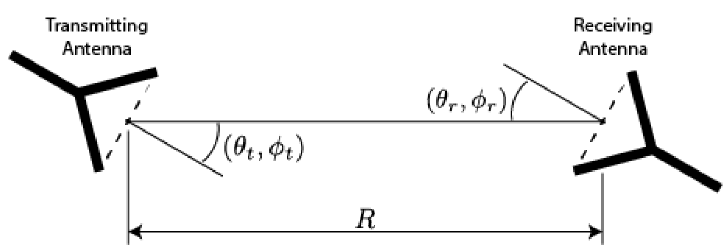


Figure 5.3: An illustration of the Friis measurement technique commonly used for antenna characterization measurements.

To analyze this two antenna setup we seek to calculate the amount of power from the transmitting antenna that we will detect with the receiving antenna. Using our understanding of antenna gain, we can calculate the power density transmitted by an antenna in a direction (θ_t, ϕ_t) at frequency f and distance R, which is given by

$$w_t = \frac{P_t}{4\pi R^2} G_t(\theta_t, \phi_t, f). \tag{5.18}$$

Here, P_t is the total power delivered to the transmitting antenna and $G_t(\theta_t, \phi_t, f)$ is the value of the transmitting antenna gain. The power density is the power per unit area, so to calculate the total power delivered to the receiving antenna we multiply the transmitted power density by the effective area of the receiving antenna,

$$P_r = w_t A_{eff,r} = P_t \frac{G_t(\theta_t, \phi_t, f) G_r(\theta_r, \phi_r, f) c^2}{(4\pi R f)^2}, \tag{5.19}$$

where $G_r(\theta_r, \phi_r, f)$ is the gain of the receiving antenna. Equation 5.19 is called the Friis transmission equation [84], which is of fundamental importance for antenna measurements, since it allows one to measure the gain of an unknown antenna by measuring the power received from an antenna with a known gain pattern. Alternatively, if no antenna with a known gain pattern is available, two identical antennas with unknown gain patterns can be used.

5.2.2.2 S-Parameters and Network Analyzers

Instead of directly measuring the power received by the antenna under test, it is more common to measure the ratio of the received power to the transmitted power,

$$\frac{P_r}{P_t} = \frac{G_t(\theta_t, \phi_t, f) G_r(\theta_r, \phi_r, f) c^2}{(4\pi R f)^2}. \tag{5.20}$$

This power ratio can be easily measured using a vector network analyzer (VNA), which automates a significant fraction of the measurement process. Network analyzers are used to measure the scattering or S-parameters of a multi-port RF device [85], which describes how waves are scattered between the device ports. The antenna measurements we have been considering can be modeled as a two-port microwave device that we can characterize by measuring how incident voltage waves are transmitted or reflected (see Figure 5.4). We can write the scattered waves (V_1^- and V_2^-) in terms of the incident (V_1^+ and V_2^+) waves using the scattering matrix

$$\begin{pmatrix} V_1^- \\ V_2^- \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} V_1^+ \\ V_2^+ \end{pmatrix}, \tag{5.21}$$

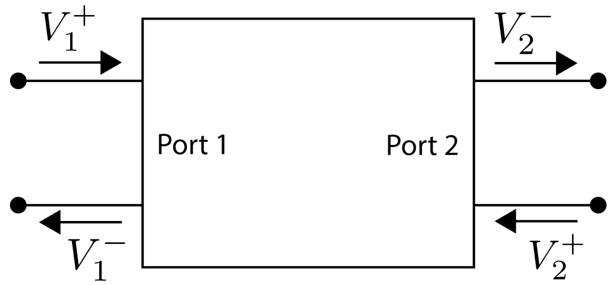


Figure 5.4: Illustration of a two-port S-parameter measurement setup. S-parameters characterize how incoming waves of voltage or power scatter of of the RF device under test. This allows you to measure important properties of the device. In particular, we can use this framework to model a two antenna radiation pattern measurement, which we can then automate using a VNA.

where the elements of the matrix are the device S-parameters. It is assumed that, when exciting the device from a particular port, that all other ports in the network are terminated at the system impedance. This ensures that the incident waves from other ports in the network are zero. Therefore, the S-parameters are the ratios between the scattered and incident waves,

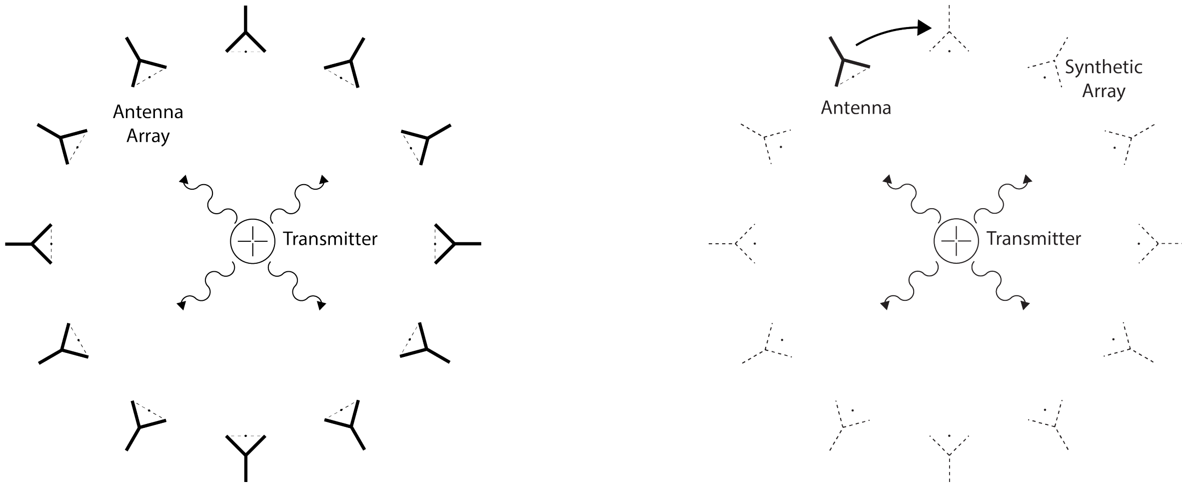
$$S_{ij} = \frac{V_i^-}{V_j^+}. \tag{5.22}$$

Alternatively, S-parameters can be defined as the ratio of the scattered and incident power, which is proportional to the ratio of the squared voltage waves. Returning to our antenna measurement setup, we see that measuring the ratio of the received to the transmitted power is equivalent to measuring the ratio of power being scattered from port 1 to port 2 in a RF network. Therefore, measuring an antenna's gain can be accomplished quite easily, by using a VNA to perform a two port S_{21} measurement.

5.2.2.3 Antenna Array Commissioning and Calibration Measurements

Up to this point we have been discussing calibration and commissioning measurements as they apply to a single antenna. While these measurements play an important role in validating the radiation patterns of the individual array elements, the ultimate goal is to use a phased array of these antennas. Therefore, we must also consider antenna measurement techniques that apply to the whole array system.

By measuring the gain of each individual array element we can predict the features of the signals received during a CRES event using the antenna factor (see Section 5.2.1.5). However, unpredictable changes to the antenna performance can be introduced by the



(a) (b)

Figure 5.5: Two measurement approaches to characterizing an antenna array for CRES measurements. The full-array approach (a) requires a complete antenna array with all the associated hardware. The synthetic array approach (b) utilizes a single antenna and a set of rotation/translation stages to reposition the transmitter or the recieing antenna to synthesize the signals that would be received by the full-array. This approach reduces the cost and complexity of array measurements. A down-side of the synthetic array approach is that multi-channel effects such as reflections cannot be measured. Utilizing both the full-array and the synthetic array is a powerful way to quantify the impact of errors from the multi-channel array.

incorporation of the antennas into the circular array geometry, therefore, we employ both individual antenna and full-array measurements in the commissioning of the FSCD to account for these effects.

There are two main approaches to array measurements that could be used for characterization and calibration (see Figure 5.5). One approach is to construct the complete array and use a omni-directional transmitting antenna to measure the power received by each channel in the antenna array. In Section 5.3 we describe the development of an omni-directional transmitter that also mimics the radiation phase characteristics of a CRES event, which is useful because the entire array can be tested without repositioning. Alternatively, a full antenna array can by synthesized by repeatedly moving and measuring a single array element. This approach is ideal for identifying if different channels in the antenna array are affecting each other through multi-path interference by comparing the measurement results of the synthetic array to the real array.

5.2.3 The Penn State Antenna Measurement System

The development of antenna array based CRES requires the capability to test and calibrate different antenna array designs to validate the performance of the as-built antenna array before and during the experiment. With these aims in mind we developed an antenna measurement system at Penn State specifically designed to mimic the characteristics of the antenna experiment designed for demonstration of the antenna array CRES technique by the Project 8 collaboration.

The Penn State antenna measurement system utilizes a two antenna measurement configuration with a stationary reference antenna and a test antenna mounted on a set of motorized translation and rotation stages (see Figure 5.6). The antenna measurement system can be operated in two distinct modes, one focused on the characterization of the radiation patterns of prototype antennas and the other focused on the validation of data-acquisition (DAQ) and CRES signal reconstruction techniques to bridge the gap between real measurements and simulation. In both measurement configurations it is critical to isolate the antennas from the environment so that multi-path reflections do not negatively influence the measurement results. For this reason we surround the measurement volume with microwave absorber foam (AEMI AEC-1.5) specifically designed to attenuate microwave radiation near the 26 GHz measurement range of the system.

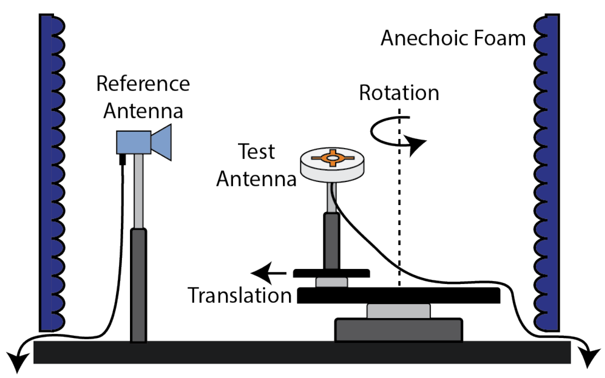
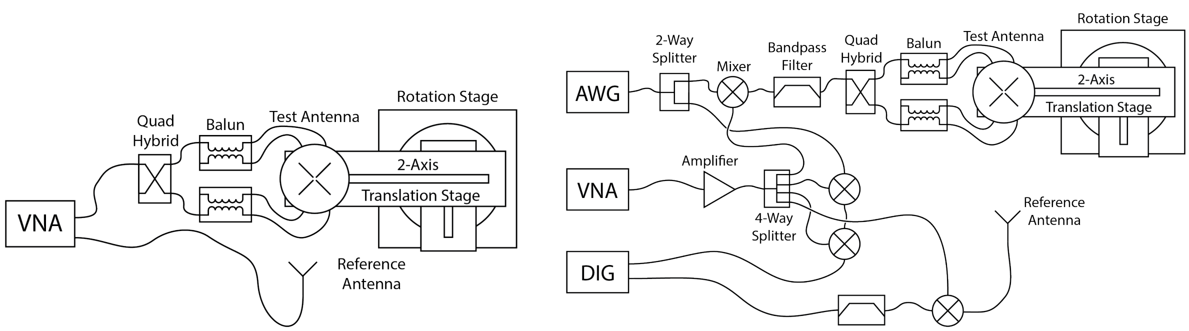


Figure 5.6: Illustration of the antenna measurement system developed for the Project 8 Collaboration. The reference and test antennas can be connected to different data acquisition configurations depending on the measurement goals The reference antenna is typically a standard horn antenna and the test antenna is mounted on a set of translation stages for positioning. Automated translation stages allows for relatively painless data-taking enabling synthetic antenna array measurements using only a single receiving antenna. Anechoic form designed to mitigate RF reflections surrounds the setup.

In the first measurement configuration the reference antenna is typically a well-characterized horn antenna as pictured, since horn antennas have well-known and stable radiation patterns making them ideal as standard references. For characterization measurements, the test antenna represents the antenna-under-test whose pattern we wish to characterize. Mounting the test antenna on motorized rotation and translation stages allows us to automate the procedure significantly speeding up the radiation pattern measurement process.

In the second measurement configuration one is interested in recreating the conditions of an antenna array CRES experiment as it concerns the antenna array and DAQ system. In this case, the reference antenna is a prototype FSCD antenna, which will be used to construct the antenna array in the FSCD experiment, and the test antenna is a specially designed synthetic cyclotron antenna (SYNCA) as picture in Figure 5.6. The SYNCA is designed such that the radiation pattern mimics that of a CRES electron so that the signals received by the prototype CRES array antenna mimic what is expected for a real CRES experiment.



(a) (b)

Figure 5.7: Diagrams of two measurement system configurations. Configuration (a) utilizes a VNA and is more suited to antenna characterization. Configuration (b) utilizes an AWG and VNA as a signal generation system and digitizer to collect measurement data, which is more suited to simulating CRES measurements. The transmission chain utilizes a quadrature hybrid and a pair of baluns to drive the cross-dipole variant test antenna developed for synthetic CRES measurements.

In Figure 5.7 we show two high-level system diagrams of the Penn State antenna measurement system that depict the important system components and the connections between them. The two configurations of the measurement system utilize different hardware. For characterization and radiation pattern measurements, one prefers the configuration shown in Figure 5.7a. In this case a vector network analyzer (VNA) is used as both the transmission source and data acquisition system and it is relatively easy to calibrate over a wide range of frequencies. Whereas, if one is more interested in recreating what would take place in the FSCD experiment then the configuration shown in Figure 5.7b is preferable, since this system effectively mimics the receiver chain envisioned for the FSCD experiment.

The characterization configuration utilizes a network analyzer (Keysight N5222A) with two independent sources and four measurement ports as the primary measurement tool. A standard reference antenna is connected to one measurement port, and the test antenna is connected to a separate port. The typical reference antenna used for these studies is a Pasternack PF9851 horn antenna . In the measurement shown, the test antenna represents a SYNCA antenna, which requires a transmission chain consisting of quadrature hybrid coupler (Marki QH-0226) connected to two baluns (Marki BAL-0026) to generate feed signals with the appropriate phases. The VNA measures the radiation pattern by performing a transmission S-parameter measurement, which can be used with the knowledge of the reference antenna's radiation pattern to determine the radiation pattern of the test antenna (see Section 5.2.1).

The second configuration is more complicated and incorporates more hardware components in order to more closely mimic the DAQ system envisioned for the FSCD experiment. The basic approach is to produce CRES-like radiation and use an antenna combined with a realistic RF receiver chain to acquire the signals. On the transmit side, an arbitrary waveform generator (AWG, RIGOL DG5252) is used to generate a waveform that mimics a CRES signal at a baseband frequency up to 250 MHz. This frequency is then up-converted to the CRES signal frequency band of 25.8 to 26.0 GHz using a mixer (Marki MM1-0832L) and a bandpass filter (K&L Microwave 3C62-25900/T200-K/K) to reject unwanted mixing components outside out of the 200 MHz CRES signal band. The local oscillator signal for mixing is provided by one of the VNA sources configured to run in a continuous wave setting. On the receive side, a prototype antenna is used to detect the radiation emitted by the test antenna, which is down-converted and filtered using the same mixer and bandpass filter as the transmission chain. Lastly, data acquisition is performed using a 14-bit ADC sampling at 500 MSa/s (CAEN DT530) to digitize the down-converted signals.

In order to distribute the LO to all mixers a 4-way power splitter (MiniCircuits ZC4PD-18263-S+) along with an amplifier (Marki APM-6848) is used to drive the four mixers used in the measurement system. A limitation of using the VNA as an LO source is that there is no control of the LO phase when a measurement is triggered by the control script, which leads to a random phase offset between acquisitions. This makes it impossible to perform synthetic array measurements, which require strict control over the starting phase of the transmitted signal. In order to monitor the random phase of the LO, a 2-way power splitter (MiniCircuits Z99SC-62-S+) is used to split the signal from the AWG between the transmission path and a LO monitoring path. The LO monitoring path consists of an up-conversion and down conversion using two mixers connected by a coaxial cable, and monitors the relative phase of the LO using a channel on the digitizer to sample this path. A phase shift in the LO will lead to a proportional phase shift in the mixed signal, which is measured and removed from the received signals.

The test antenna is mounted on a set of motorized stages, which are identical for both measurement configurations. A rotational stage (ThorLabs PRMTZ8) is used as the base layer with additional translation stages mounted on top of this. The rotational stage is ideal for measuring a complete azimuthal scan of the test antenna's radiation pattern as well as for moving a SYNCA antenna in circular motion to recreate the symmetry of the FSCD antenna array. On top of the rotational stage we mount two linear translation stages (ThorLabs MTS50-Z8 and MTS25-Z8) in a cross-wise manner so that the test antenna can be moved along two perpendicular axes. Using the linear stages in combination with the rotational stage allows one to fine-tune the positioning of the test antenna so that it can be perfectly aligned with the central axis of the array. A LabView script was developed to automate the measurement of a full 360° radiation pattern and control the measurement electronics. Data from these acquisitions is stored on university provided cloud storage.

5.3 Development of a Synthetic Cyclotron Antenna (SYNCA) for Antenna Array Calibration

This section is the manuscript of the publication [79] detailing the development of a Synthetic Cyclotron Antenna (SYNCA) for antenna array characterization measurements by the Project 8 collaboration.

5.3.1 Introduction

Neutrinos are the most abundant standard model fermions in our universe, but due to weak interaction cross-sections with other particles, neutrinos are particularly difficult to study. Consequently, many fundamental properties of neutrinos are still unknown including the absolute scale of the neutrino mass [28]. Direct, kinematic measurements of the neutrino mass are particularly valuable due to their model independent nature [35].

To date the most sensitive direct neutrino mass measurements have been performed by the KATRIN collaboration [86], which measures the molecular tritium β-decay spectrum to infer the neutrino mass. Current data from neutrino oscillation measurements [28] allow for neutrino masses significantly smaller than the design sensitivity of the KATRIN experiment; therefore, there is a need for new technologies for performing direct neutrino mass measurements to probe lower neutrino masses.

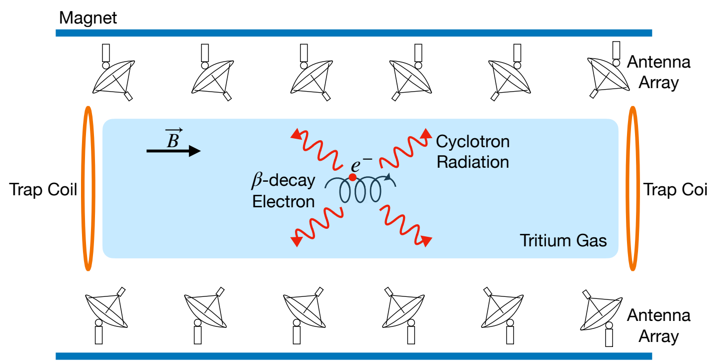


Figure 5.8: A sketch of an antenna array large-volume CRES experiment. Electrons from β-decays are confined in a magnetic field using a set of trap coils. The cyclotron radiation produced by the motion of the trapped electrons can be detected by a surrounding antenna array to determine the electron energies. Measuring the energies of many electrons produces a β-decay spectrum.

The Project 8 collaboration is developing new methods for neutrino mass measurement based on Cyclotron Radiation Emission Spectroscopy (CRES) [54, 87–89], with the goal of measuring the absolute scale of the neutrino mass with a 40 meV/c^2 sensitivity [**?**, 35]. This sensitivity goal will require the development of two separate technical capabilities. First is the development of an atomic tritium source, which avoids significant spectral broadening due to molecular final states [53]. Second is the technology for performing CRES in a multi-cubic-meter experimental volume with high combined detection and reconstruction efficiency, which is required in order to obtain sufficient event statistics near the tritium spectrum endpoint.

One approach for a large-volume CRES experiment is to use an array of antennas, which surrounds a volume of tritium gas, to detect the cyclotron radiation produced by the β-decay electrons when they are trapped in a background magnetic field using a set of magnetic trapping coils (see Figure 5.8). Project 8 has developed a conceptual experiment design to study the feasibility of this approach. The design consists of a single circular array of antennas with a radius of 10 cm and 60 independent channels positioned around the center of the magnetic trap. The motivation behind this antenna array design is to first develop an understanding of the antenna array approach to CRES with a small scale experiment before attempting to scale the technique to large volumes by using multiple antenna rings to construct the full cylindrical array. The development of the antenna array approach to CRES has largely proceeded through simulations using the Locust software package [78, 90], which is used to model the fields emitted by CRES events and predict the signals received by the surrounding antenna array. To validate these simulations, a dedicated test stand is being constructed to perform characterization measurements of the prototype antenna array developed by Project 8 (see Figure 5.9) and benchmark signal reconstruction methods using a specially designed transmitting calibration probe antenna.

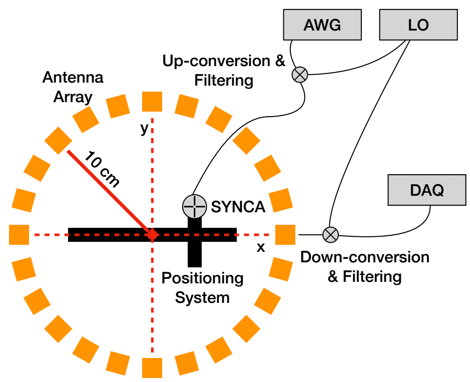


Figure 5.9: A schematic of the antenna array test stand. The circular antenna array has a radius of 10 cm with 60 independent channels (limited number shown for clarity). The test stand includes an arbitrary waveform generator (AWG), local oscillator (LO), and data acquisition (DAQ) hardware. Finally, a specialized Synthetic Cyclotron Antenna (SYNCA) is used to inject signals to test the antenna array.

We call this probe antenna the Synthetic Cyclotron Antenna or SYNCA. The SYNCA is a novel antenna design that mimics the cyclotron radiation generated by individual charged particles trapped in a magnetic field, which will be used in the antenna test stand to perform characterization measurements, simulation validation, and reconstruction benchmarking. This paper provides an overview of the design, construction, and characterization measurements of the SYNCA performed in preparation for its usage as a transmitting calibration probe.

In Section 5.3.2 we provide a description of the cyclotron radiation field characteristics that we recreate with the SYNCA. In Section 5.3.3 we give an overview of the simulations performed to develop an antenna design that mimics the characteristics of cyclotron radiation. In Section 5.3.4 we outline characterization measurements to validate that the fields generated by the SYNCA match simulation, and finally in Section 5.3.5 we demonstrate an application of the SYNCA to test phased array reconstruction techniques on the bench-top.

5.3.2 Cyclotron Radiation Phenomenology

To understand the cyclotron radiation phenomenology that the SYNCA should mimic, we consider a charged particle moving at relativistic speed in the presence of an external magnetic field (see Figure 5.10). In the special case we shall examine, the entirety of the electron's momentum is directed perpendicular to the magnetic field; therefore, the trajectory of the electron is confined to the cyclotron orbit plane. Because the momentum vector is oriented perpendicular to the magnetic field, electrons with these trajectories are said to have pitch angles of 90°.

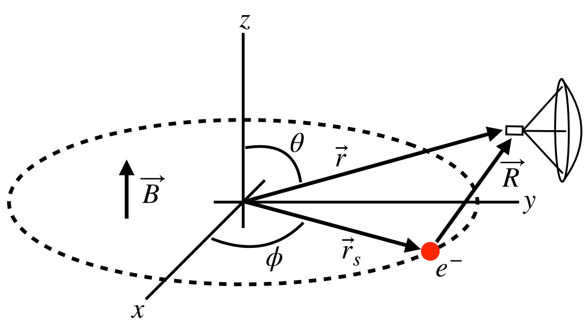


Figure 5.10: An electron (red dot) performing cyclotron motion in the x-y plane. The resulting cyclotron radiation is observed by an antenna located at the field point of interest.

The cyclotron radiation fields generated by this circular trajectory are those which we aim to reproduce with the SYNCA. We can describe the electromagnetic (EM) fields using the Liénard-Wiechert equations [47, 78], which in non-covariant form express the electric field as

$$\vec{E} = e\left[\frac{\hat{n}-\vec{\beta}}{\gamma^2(1-\vec{\beta}\cdot\hat{n})^3|\vec{R}|^2}\right]_{t_\mathrm{r}} + \frac{e}{c}\left[\frac{\hat{n}\times[(\hat{n}-\vec{\beta})\times\dot{\vec{\beta}}]}{(1-\vec{\beta}\cdot\hat{n})^3|\vec{R}|}\right]_{t_\mathrm{r}}, \tag{5.23}$$

where e is the particle's charge, $\hat{n} = (\vec{r}-\vec{r_s})/|\vec{r}-\vec{r_s}|$ is the unit vector pointing from the electron to the field measurement point, $\vec{\beta} = \dot{\vec{r_s}}/c$ is the velocity of the particle divided by the speed of light, and γ is the relativistic Lorentz factor. The equation is meant to be evaluated at the retarded time as indicated by $t_\mathrm{r} = t - |\vec{R}|/c$, which accounts for the time delay due to the finite speed of light between the point where the field was emitted and the point where the field is detected.

We would like to simplify Equation 5.23 it at all possible. As a first step we analyze the relative magnitudes of the electric field polarization components. Consider an electron following a circular cyclotron orbit in a uniform magnetic field whose guiding center is positioned at the origin of the coordinate system. The equation of motion can be expressed as

$$\vec{r_s} = (r_c \cos\omega_c t_r)\hat{x} + (r_c \sin\omega_c t_r)\hat{y}. \tag{5.24}$$

For single antenna located along the y-axis at position $\vec{r} = r_a\hat{y}$ we are interested in the incident electric fields from the electron. The electric field is given by Equation 5.23, which we evaluate in the regime where $r_a \gg r_c$. This limit can justified by comparing the radius of the cyclotron orbit for an electron with the tritium beta-spectrum endpoint energy of 18.6 keV in a 1 T magnetic field to the typical ($r_a \simeq 100$ mm) radial position of the receiving antenna. We find that the cyclotron orbit has a radius of 0.46 mm which is approximately a factor of 200 smaller than the typical antenna radial position. In this regime we can make the approximation $\vec{R} \simeq r_a\hat{y}$ and the expression for the electric field at the antenna's position becomes

$$\vec{E} \;=\; \frac{e}{\gamma^2 r_a^2}\frac{\hat{x}(\frac{r_c\omega_c}{c}\sin\omega_c t_r) + \hat{y}(1-\frac{r_c\omega_c}{c}\cos\omega_c t_r)}{(1-\frac{r_c\omega_c}{c}\cos\omega_c t_r)^3} \;-\; \frac{e}{cr_a}\frac{\hat{x}(\frac{r_c^2\omega_c^3}{c^2} - \frac{r_c\omega_c^2}{c}\cos\omega_c t_r)}{(1-\frac{r_c\omega_c}{c}\cos\omega_c t_r)^3}. \tag{5.25}$$

Since the receiving antenna is part of a circular array of antennas, it is useful to rewrite Equation 5.25 in terms of the azimuthal ($\hat{\phi}$) and radial ($\hat{r}$) polarizations. Making use of the fact that for an antenna located at $R = r_a\hat{y}$ that $\hat{\phi} = -\hat{x}$ and $\hat{r} = \hat{y}$ we find

$$\vec{E} = \hat{\phi}E_\phi + \hat{r}E_r \tag{5.26}$$

$$E_\phi = \frac{e}{(1-\frac{r_c\omega_c}{c}\cos\omega_c t_r)^3}\left[-\frac{\frac{r_c\omega_c}{c}\sin\omega_c t_r}{\gamma^2 r_a^2} + \frac{\omega_c\left(\frac{r_c^2\omega_c^2}{c^2} - \frac{r_c\omega_c}{c}\cos\omega_c t_r\right)}{cr_a}\right] \tag{5.27}$$

$$E_r = \frac{e(1-\frac{r_c\omega_c}{c}\sin\omega_c t_r)}{\gamma^2 r_a^2(1-\frac{r_c\omega_c}{c}\cos\omega_c t_r)^3}. \tag{5.28}$$

For the purposes of designing a synthetic cyclotron radiation antenna we are interested in the dominant electric field polarization emitted by the electron. The antenna is being designed to mimic the cyclotron radiation produced by electrons with kinetic energies of approximately 18.6 keV in a 1 T magnetic field [53]. Since the relativistic beta factor for an electron with this kinetic energy is $|\vec{\beta}| \simeq \frac{1}{4}$, the approximations $\gamma \simeq 1$ and $\frac{r_c\omega_c}{c} \simeq \frac{1}{4}$ are justified. Inserting these expressions into the equations for the electric field components above simplifies the comparison of the magnitudes of the two components. Additionally, we compare the time-averaged magnitudes to evaluate the root mean squared electric field ratio. The time-averaged ratio of the radial and azimuthally polarizied electric fields with the above simplifications is given by

$$\frac{\langle|E_r|\rangle}{\langle|E_\phi|\rangle} = \frac{8-\sqrt{2}}{\left|1-\frac{r_a}{r_c}\frac{1-2\sqrt{2}}{8}\right|} \simeq \frac{r_c}{r_a}\frac{8(8-\sqrt{2})}{2\sqrt{2}-1} = 0.13, \tag{5.29}$$

where we have made use of the fact that for these magnetic fields and kinetic energies the cyclotron radius is much smaller than the radius of the antenna array.

From Equation 5.29 we see that the time-averaged azimuthal polarization is larger than the radial polarization by about a factor of 8, which makes it the dominant contribution to the electric fields at the position of the antenna. We must also consider the directivity of the receiving antenna which can have a gain that is disproportionately large for a specific polarization component. Because the E_ϕ component is dominant, the receiving antenna array is designed with an azimuthal polarization, which negates the voltages induced in the antenna from the radially polarized fields. Therefore, we conclude that for the purpose of designing the SYNCA antenna it is acceptable to approximate the electric fields from Equation 5.23 as purely azimuthally or ϕ-polarized. The simplified expression for the electric field received by an antenna becomes

$$\vec{E} = E_\phi\hat{\phi} = \frac{e\frac{r_c\omega_c}{c}}{4r_a r_c}\left[\frac{\frac{r_c\omega_c}{c} - \cos\omega_c t - \frac{4r_c}{r_a}\sin\omega_c t}{(1-\frac{r_c\omega_c}{c}\cos\omega_c t)^3}\right]_{t_r}\hat{\phi}, \tag{5.30}$$

where the radius of the cyclotron orbit is called r_c, the cyclotron frequency is called ω_c, and the radial position of the receiving antenna is called r_a. Equation 5.30 has been evaluated in the non-relativistic limit where $\gamma \simeq 1$, which is justified by the fact that $|\vec{\beta}| \simeq \frac{c}{4}$ for an electron with an 18.6 keV kinetic energy in a 1 T magnetic field.

This rather complicated expression can be simplified using Fourier analysis. Assuming a background magnetic field of 1 T and a kinetic energy of 18.6 keV we calculate numerically the electric field using Equation 5.30 and apply a discrete Fourier Transform to visualize the frequency spectrum (see Figure 5.11).

We observe that the azimuthally polarized electric field is periodic with a base cyclotron frequency of 25.898 GHz corresponding to the highest power frequency component in

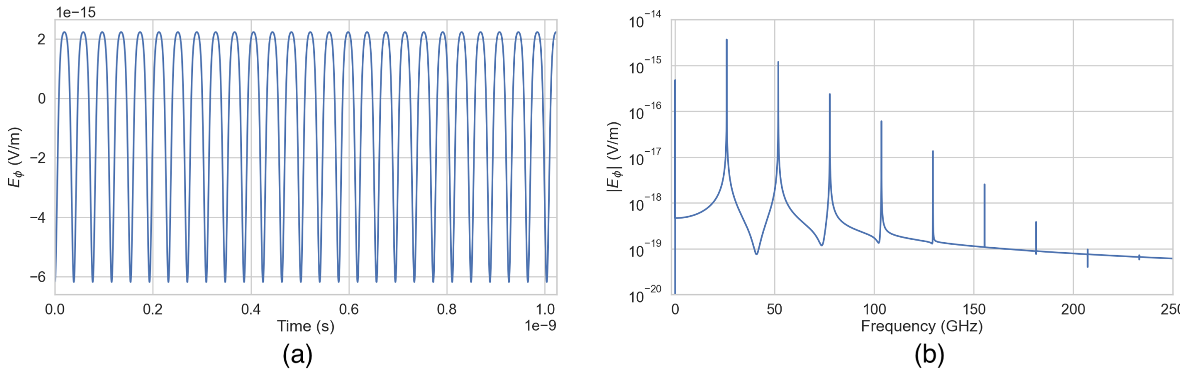


Figure 5.11: A plot of the numeric solution to Equation 5.31. The time-domain representation of the signal (a) is composed of a zero frequency term and a series of harmonics separated by the main cyclotron frequency as shown in the plot of the frequency spectrum (b). We can see that the relative amplitude of the harmonics beyond $k = 7$ are smaller than the main carrier by a factor of about 10^{-5} and are completely negligible.

Figure 5.11. The frequency spectrum reveals that the signal is composed of a constant term with zero frequency and a series of harmonics separated by 25.898 GHz. Therefore, we can represent the azimuthal electric fields from the electron as a linear combination of pure sinusoids with frequencies given by $\omega_k = k\omega_c$ ($k \in 0, 1, 2 ...$) and amplitudes extracted from the Fourier representation. Using this representation we can transform the equation for the azimuthally polarized electric fields in Equation 5.30 into

$$E_\phi = \frac{e\frac{r_c \omega_c}{c}}{4 r_a r_c} \sum_{k=0}^{7} A_k e^{i\omega_k t_r}, \tag{5.31}$$

where we have truncated the sum over harmonics at the 7th order for completeness. The amplitudes A_k are dimensionless complex numbers, which encode the relative powers of the harmonics as well as the starting overall phase of the cyclotron radiation. Because magnitude of the relative amplitudes exponentially decreases for higher harmonics, it is usually sufficient to consider only the terms up to $k = 4$ where the relative amplitude of the hamonics has decresed from the main carrier by a factor of approximately 100. However, for completeness we include harmonics up to 7th order in Equation 5.31. The range of frequencies to which the receiving antenna array in the antenna test stand is sensitive is defined by the antenna's transfer function. The receptive bandwidth for the antennas used in the test stand is a range of frequencies with a bandwidth on the order of a few GHz centered around the main cyclotron carrier frequency of 25.898 GHz. Therefore, the higher order harmonics as well as the zero frequency term can be ignored when considering only the signals that will be received by the antenna array.

Considering only the 1st order harmonic term from Equation 5.31, which represents the portion of the electric field that will be detected by the array, and evaluating this at the retarded time we obtain the following for the ϕ-polarized electric fields

$$E_\phi \propto \cos\left(\omega_c\left(t - |\vec{R}|/c\right) - \Delta\right), \tag{5.32}$$

where the arbitrary phase Δ is defined by $A_k = |A_k|e^{i\Delta}$. We are interested in the characteristics of the amplitude of the electric field as a function of the radial distance component ($|\vec{R}|$) of the retarded time. In particular, the maximum of E_ϕ occurs when the argument of the cosine function is equal $n\pi$ where $n \in \{0, \pm 2, \pm 4, ...\}$; however, the solutions where n is negative can be discarded since they represent unphysical negative overall phases. Applying this condition to Equation 5.32 gives a condition on the radial position of the maximum of E_ϕ

$$\omega_c(t - |\vec{R}|/c) - \Delta = n\pi, \tag{5.33a}$$

$$|\vec{R}| = \frac{c}{\omega_c}\left((\omega_c t - \Delta) - n\pi\right), \tag{5.33b}$$

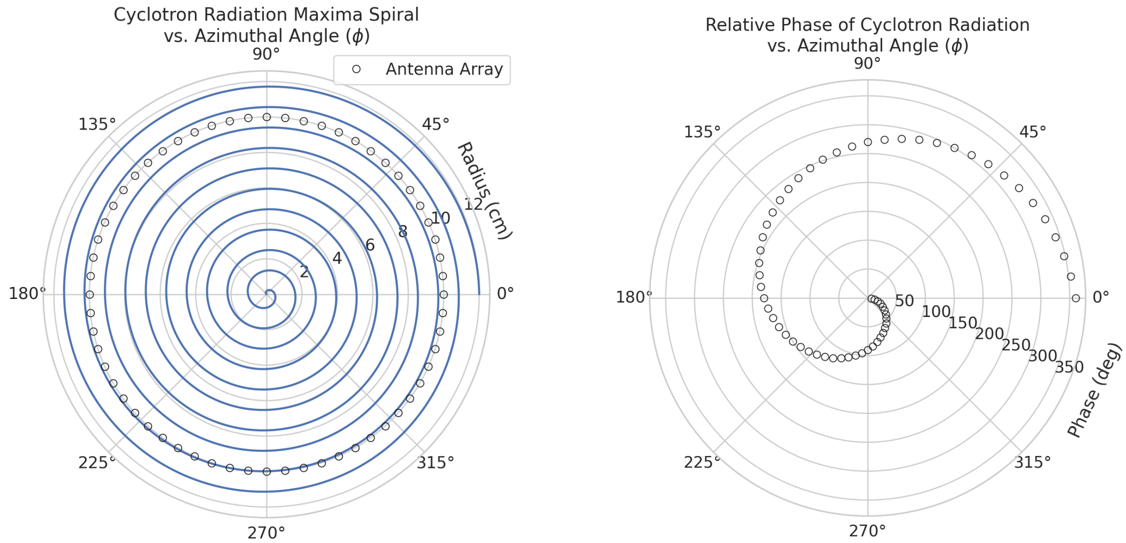
which is a function of time in the frame of the moving electron (t). Equation 5.33 can be further simplified by noticing that the azimuthal position of the electron ($\phi_e(t)$) as a function of time is defined by $\phi_e(t) = \omega_c t - \Delta$ which reduces Equation 5.33 to

$$|\vec{R}| = \frac{c}{\omega_c}(\phi_e(t) - n\pi). \tag{5.34}$$

Equation 5.34 represents an archimedian spiral which is formed when plotting the amplitude of E_ϕ in the x-y plane. The solution where $n = 0$ represents the leading edge of the radiation spiral which propagates outward from the electron at the speed of light. The additional solutions for $n > 0$ represent the persistent spiral at radii inside the leading edge of the radiated fields that have not yet been detected by the receiver at the current time. In Figure 5.12a we show the expected spiral pattern for the maxima of the cyclotron radiation.

In particular, we note that for the circular array geometry of the test stand, depicted as the series of circles in Figure 5.12a, each antenna receives a linearly polarized wave with a phase offset that corresponds to the azimuthal angle for that antenna element. Therefore, as we show in Figure 5.12b, when the relative phase of the received signal is plotted as a function of the recieving antenna's azimuthal position the result is also an

Archimedean spiral.



(a) (b)

Figure 5.12: The amplitude maxima of the cyclotron radiation form an Archimedean spiral as the radiation propagates outward from the cyclotron orbit center (a). A circular antenna array located at a fixed radius from the orbit center will receive electric fields with equal magnitude in each of its channels, but the phase of the electric field incident on each array channel will be linearly out of phase from its neighbor antennas by an amount equal to the angular separation of the two channels (b).

Based on these analytical calculations we can characterize the magnitude, polarization, and phase of the signals received by the antenna array using three criteria. These criteria are the basis of comparison for the radiation produced by the SYNCA and cyclotron radiation emitted by electrons and will be used to evaluate the performance of antenna designs. The criteria are:

1. Electric fields that are ϕ-polarized near $\theta = 90°$
2. Uniform time-averaged electric field magnitudes around the circumference of a circle centered on the antenna
3. Electric fields whose phase is equal to the azimuthal angle at the point of measurement plus a constant

The Locust simulation package [90] can be used to directly simulate the EM fields generated by electrons performing cyclotron motion to validate the analytical calculations. Locust simulates the EM fields by first calculating the trajectory of the electrons in the magnetic trap using the Kassiopeia software package [91]. The trajectory can then be used to solve for the EM fields using the Liénard-Wiechert equations directly with no approximations. The resulting electric field solutions drive a receiving antenna by convolving the time-domain fields with the finite-impulse response filter of the antenna or they can be examined directly to study the field characteristics that the SYNCA must reproduce. In the next section we compare the radiation field patterns for electrons simulated with Locust to patterns from a SYNCA antenna design.

5.3.3 SYNCA Simulations and Design

One potential SYNCA design is the crossed-dipole antenna [92]. A crossed-dipole antenna consists of two dipole antennas, one of which is rotated 90° with respect to the other, which are fed with signals that are out of phase from the opposite dipole by 90° (see Figure 5.13). This arrangement causes the signals fed to each arm of the dipole to be

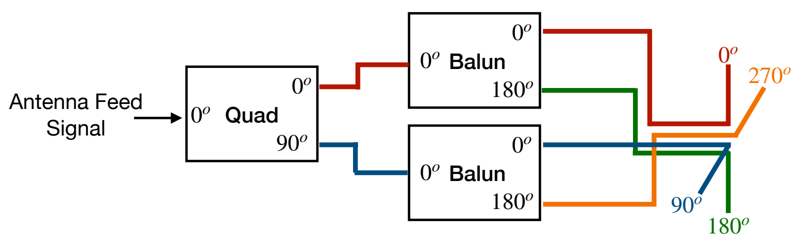


Figure 5.13: An idealized crossed-dipole antenna consists of two electric dipole antennas oriented perpendicular to each other and is fed with four signals with a quadrature phase relationship. An example antenna feed circuit is shown which is composed of a chained combination of a quadrature hybrid-coupler (Quad) and two baluns.

out of phase from each of the neighboring arms by 90°, which mirrors the spatial phase relationship of cyclotron radiation fields.

A potential drawback of this design is that standard crossed-dipole antennas do not radiate uniform electric fields near the $\theta = \pi/2$ plane. Typical crossed-dipole antennas use dipole arm lengths equal to $\lambda/4$ or larger [92], where λ is the wavelength at the desired operating frequency. Such large arm lengths cause the electric field magnitude to vary significantly around the circumference of the antenna. However, making the antenna electrically small by shrinking the arm length can improve the antenna pattern uniformity.

In general, the criterion for an electrically small antenna is that the largest dimension of the antenna (D) obey $D \lesssim \lambda/10$ [63]. In our application, we are attempting to mimic the cyclotron radiation emitted by electrons produced from tritium β-decay with energies near the spectrum endpoint. For a background magnetic field of 1 T, the corresponding cyclotron frequency of tritium endpoint electrons is approximately 26 GHz. Therefore, the electrically small condition would require that the largest dimension of the crossed-dipole antenna be smaller than 1.2 mm.

A crossed-dipole antenna with an overall size of 1.2 mm is challenging to fabricate due to the small dimensions of the dipole arms that, in practice, are fragile and unsuitable for use as a calibration probe. To mitigate some of the challenges with the fabrication of such a small antenna, a variant crossed-dipole antenna design using printed circuit board (PCB) technology (see Figure 5.14) was developed in partnership with an antenna prototyping company, Field Theory Consulting [1].

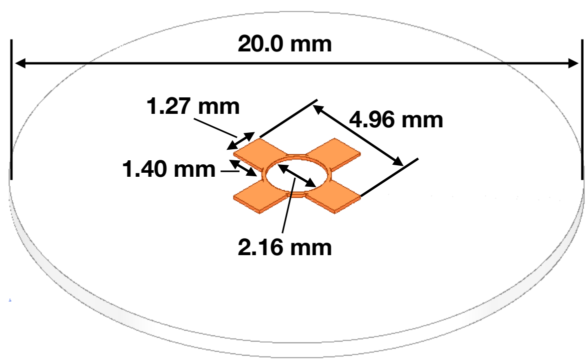


Figure 5.14: A model of the PCB crossed-dipole antenna with dimensions. The design has an inside diameter of 2.16 mm for the central circular trace, which is 0.13 mm wide. The dipole arms each have a width of 1.27 mm and protrude beyond the circular trace by 1.40 mm, which gives an overall width of 4.96 mm for the length of the antenna PCB trace from end-to-end. The overall size of the antenna is 20.0 mm the majority of which is the PCB dielectric material. This design was observed in simulation to maintain the field characteristics of the idealized crossed-dipole while being simpler to fabricate due to the increased size of the antenna.

The PCB crossed-dipole design uses four rectangular pads to represent the dipole arms, which are connected by a thin circular trace. The circular trace both adds mechanical stability to the antenna and improves the azimuthal uniformity of the electric fields compared to a more standard crossed-dipole geometry. Futhermore, the circular trace allows for a greater separation between dipole arms than standard crossed-dipoles, which

[1] https://fieldtheoryinc.com/

is required to accommodate the coaxial connections to each pad. The pads each contain a through-hole solder joint to connect coaxial transmission lines using hand soldering. The antenna PCB has no ground plane on the bottom layer as this was observed in simulation to significantly distort the radiation pattern in the plane of the PCB. The only ground planes present in the model are the outer conductors of the four coaxial transmission lines which feed the antenna. These are left unterminated on the bottom of the PCB dielectric material.

The antenna design development utilized a combination of Locust electron simulations and antenna simulations using ANSYS HFSS [64], a commercial finite-element electromagnetic simulation software. Two antenna designs were simulated: an idealized electrically small crossed-dipole antenna with an arm length of 0.40 mm and an arm separation of 0.05 mm, as well as a PCB crossed-dipole antenna with the dimensions shown in Figure 5.14. Plotting the magnitude of the electric fields generated by the antennas across a 10 cm square located in the same plane as the respective antennas reveals the expected cyclotron spiral pattern (see Figure 5.15) which closely matches the prediction for simulated electrons. The spiral pattern demonstrates that the electric fields have the appropriate phases to mimic cyclotron radiation, which fulfills SYNCA criterion 3 identified in Section 5.3.2.

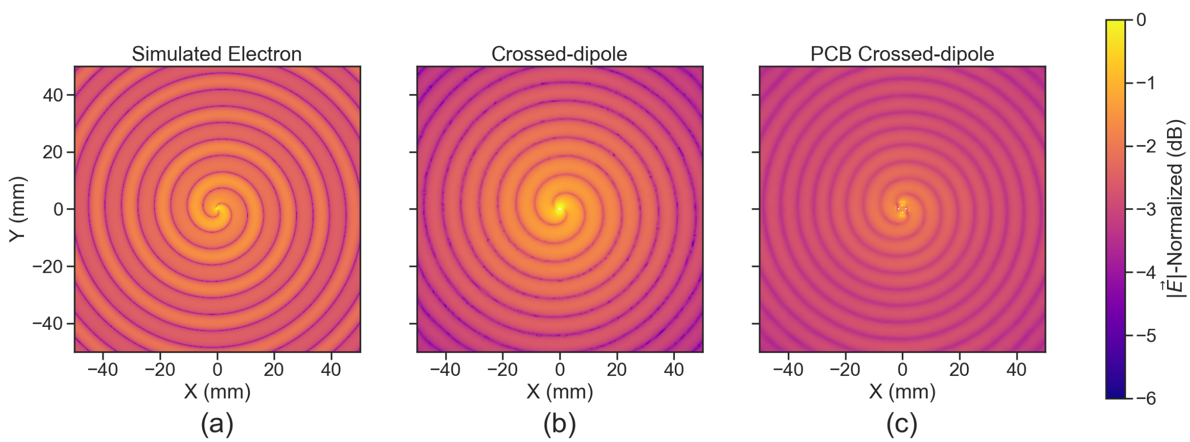


Figure 5.15: A comparison of the electric field magnitudes, normalized by the maximum value of the electric field in each simulation, plotted on a 10 cm square to visualize the Archimedean spirals formed by the electron (a), the crossed-dipole antenna (b), and a PCB crossed-dipole antenna (c). The matching patterns indicate that the electric fields have similar phase characteristics. These images were generated using Locust simulations for the electron and ANSYS HFSS for both antennas.

As we can see from Figure 5.16, the crossed-dipole antenna, which uses an idealized geometry, exhibits good agreement with simulation. The antenna has a maximum deviation from a simulated electron of approximately 0.5 dB in the total electric field, 1 dB for the ϕ-polarized electric field and 1 dB for the θ-polarized electric field.

In comparison, the pattern of the PCB crossed-dipole antenna, because the simulation incorporates the geometry of the coax transmission lines, exhibits some distortion from the idealized cross-dipole simulations. The vertically oriented ground planes of the coax lines introduce more θ-polarized electric fields than are observed for simulated electrons near $\theta = 90°$. The significant θ-polarized field minimum is still present but shifted to approximately $\theta = 65°$. The θ-polarized field deviations of the PCB crossed-dipole

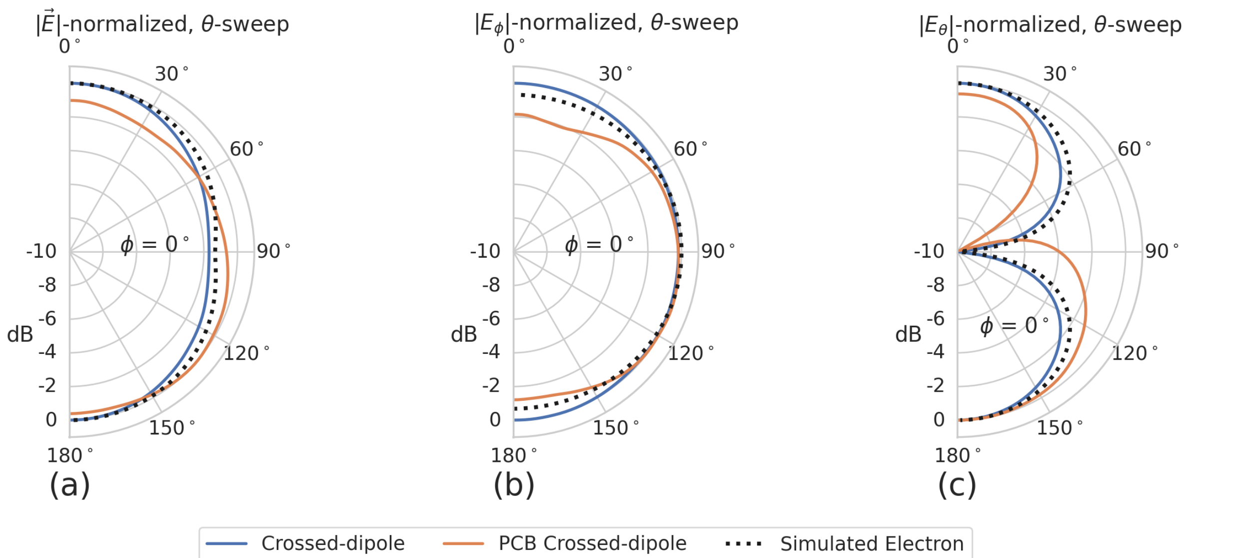


Figure 5.16: A comparison of the normalized electric field magnitudes for the ideal crossed-dipole, PCB crossed-dipole, and a simulated electron as a function of the polar angle (θ). (a) Shows the total electric field, (b) shows the ϕ-polarized electric field component, and (c) shows the θ-polarized electric field component. These images were generated using Locust simulations for the electron and ANSYS HFSS for both antennas.

antenna should not greatly impact the performance of the antenna because the receiving antenna array is primarily ϕ-polarized. Therefore deviations in the θ-polarized fields will be suppressed due to the polarization mismatch. More importantly, the ϕ-polarized electric field pattern generated by the PCB crossed-dipole closely matches simulated electrons across the polar angle range of $50° < \theta < 150°$. In this region the PCB crossed-dipole differs by less than 0.5 dB from simulated electrons. This range greatly exceeds the beamwidth of the receiving antenna array which is designed to be most sensitive to fields produced near $\theta = 90°$. Therefore, we conclude that the PCB crossed-dipole antenna generates a ϕ-polarized radiation pattern that fulfills SYNCA criterion 1 from Section 5.3.2.

The final SYNCA criterion is related to the uniformity of the electric fields when measured azimuthally around the antenna. As we saw for real electrons in Section 5.3.2 it is expected that the magnitude of the electric field be completely uniform as a function of the azimuthal angle due to the symmetry of the cyclotron orbit. In Figure 5.17 we plot the total electric field as a function of azimuthal angle for an electron, the crossed-dipole antenna, and the PCB crossed-dipole antenna. The crossed-dipole antenna exhibits

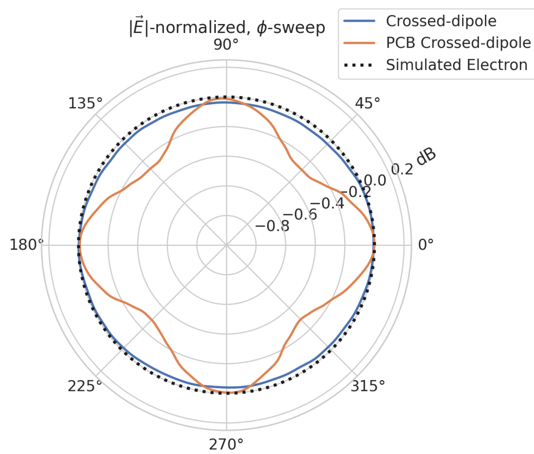
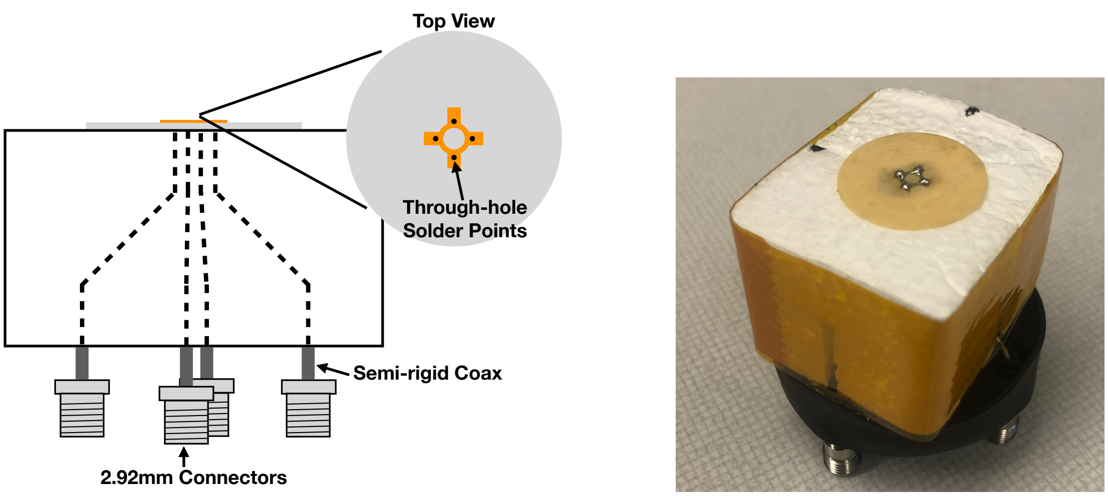


Figure 5.17: A comparison of the normalized electric field magnitudes for the crossed-dipole, PCB crossed-dipole, and a simulated electron as a function of the azimuthal angle (ϕ) evaluated at $\theta = 90°$. This image was generated using Locust simulations for the electron and ANSYS HFSS for both antennas.

perfect uniformity around the azimuthal angle, whereas the PCB crossed-dipole has a small periodic deviation with a maximum difference of 0.3 dB caused by the coaxial transmission lines below the PCB. Such a small deviation from uniformity is acceptable since it is smaller than the expected variation in uniformity caused by imperfections in the antenna fabrication process, which modifies the antenna shape in an uncontrolled manner by introducing solder blobs with a typical size of a few tenths of a millimeter on the dipole arms (see Figure 5.18). Additionally, the SYNCA will be separately calibrated to account for azimuthal differences in the electric field magnitude. Therefore we see from the simulated performance of the PCB crossed-dipole antenna that this antenna design meets all three of the SYNCA criteria.

5.3.4 Characterization of the SYNCA

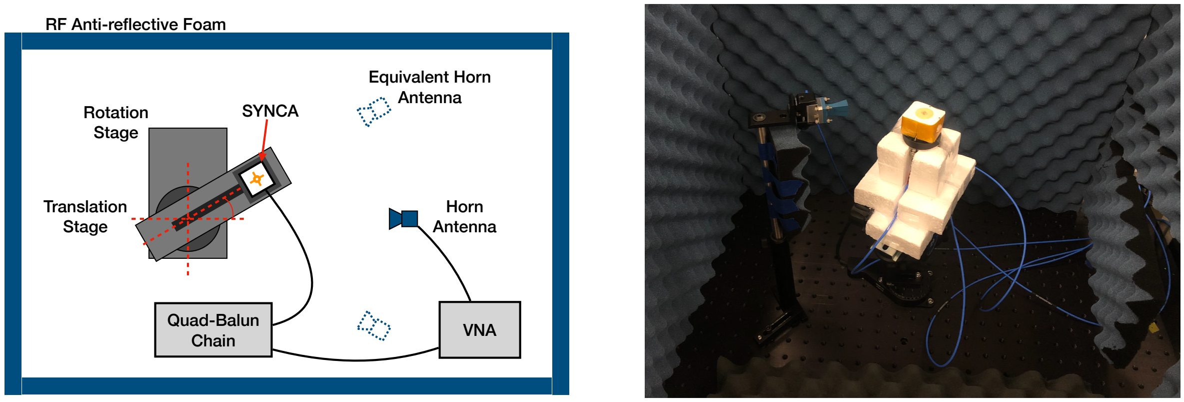
Two SYNCAs were manufactured using the PCB crossed-dipole design (see Figure 5.18). The antenna PCB (Matrix Circuit Board Materials, MEGTRON 6) is connected to four 2.92 mm coaxial connectors (Fairview Microwave, SC5843) using semi-rigid coax (Fairview Microwave, FMBC002), which also physically support the antenna PCB. The antenna PCB consists only of two layers which correspond to the copper antenna trace and the PCB dielectric. Each coax line is connected to the associated dipole arm using through-hole soldering and phase matched to ensure that the electrical length of each of the transmission lines is identical at the operating frequency. The antenna PCB is further reinforced using custom cut polystyrene foam blocks, which have an electrical permittivity nearly identical to air. A custom 3D printed mount is included at the base of the antenna to support the coax connectors and to provide a sturdy mounting base.



(a) (b)

Figure 5.18: (a) A cartoon schematic which highlights the routing of the semi-rigid coax transmission lines. (b) A photograph of a SYNCA constructed using the modified crossed-dipole PCB antenna design. Visible in the photograph of the SYNCA are four blobs of solder which are an artifact of the SYNCA's hand-soldered construction. These solder blobs are the most significant deviation from the SYNCA design shown in Figure 5.14 and are responsible for a significant fraction of the irregularities seen in the antenna pattern.

Characterization measurements were performed using a Vector Network Analyzer (VNA) to measure the electric field magnitude and phase radiated by the SYNCA to verify the radiation pattern (see Figure 5.19). The VNA is connected to the SYNCA



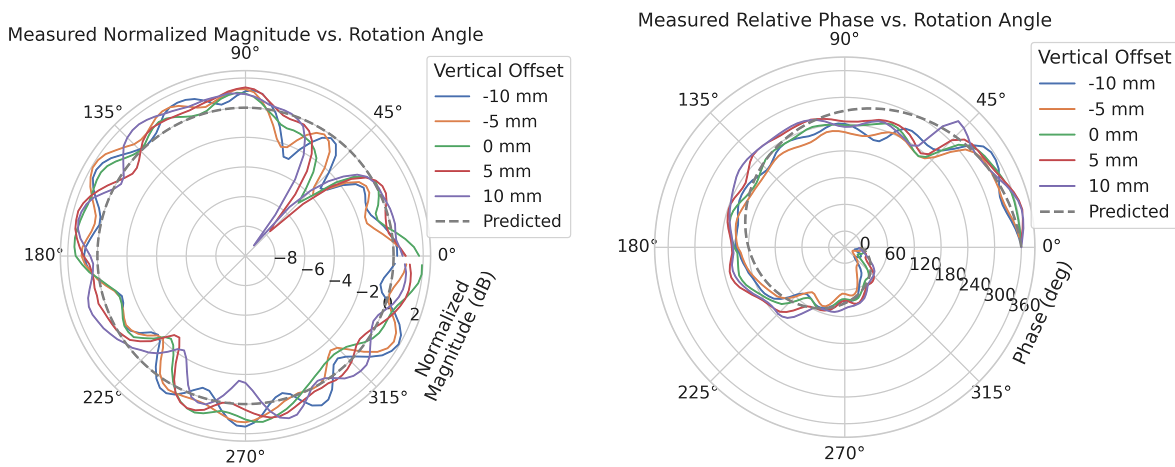
(a) (b)

Figure 5.19: A schematic of the VNA characterization measurements (a). This setup allows for antenna gain and phase measurements across a full 360° of azimuthal angles using a motorized rotation stage and control of the radial position of the SYNCA using a translation stage. A photo of the setup in the lab is shown in (b).

at one port through a hybrid-coupler whose outputs are connected to two baluns to generate the signals with the appropriate phases to feed the SYNCA (see Figure 5.13). The other port of the VNA is connected to a single reference horn antenna that serves as a field probe. To position the SYNCA, a combination of translation and rotation stages are used to characterize the antenna's fields across the entire radiation pattern circumference. This measurement scheme is equivalent to measuring the fields generated by the SYNCA using a full circular array of probe antennas.

The antenna measurement space is surrounded by RF anti-reflective foam to isolate the measurements from the lab environment (see Figure 5.19b) and remaining reflections are removed using the VNA's time-gating feature. The SYNCA is affixed to the stages by a custom RF transparent mount made of polystyrene foam. The coaxial cables deliver the antenna feed signals generated by the VNA to the SYNCA while still allowing unrestricted rotation. The horn antenna probe is nominally positioned in the plane formed by the antenna PCB ($\theta = 90°$ or $z = 0$ mm) at a distance of 10 cm from the SYNCA, to match the expected position of the antenna array relative to the SYNCA in the antenna array test stand. The horn antenna can be manually raised or lowered to different relative vertical positions to characterize the radiation pattern at different polar angles.

Several 360° scans were performed with probe vertical offsets of -10.0 mm, -5.0 mm, 0.0 mm, 5.0 mm, and 10.0 mm relative to the antenna PCB plane. These probe offsets cover a 2 cm wide vertical region centered on the SYNCA PCB, approximately equal to ±6 degrees of polar angle. The measurements show that the SYNCA is generating fields with nearly isotropic magnitude across the probed region. The standard deviation of the electric field magnitude measured around the antenna circumference is approximately 2.9 dB for a typical rotational scan. The presence of a significant pattern null is noted near 45° (see Figure 5.20), which we attribute to small imperfections in the antenna PCB that could be introduced from the hand soldered terminations connecting the coax cables to the antenna. There is no significant difference in the radiation pattern when measured across the 2 cm vertical range. The measured relative phases closely follow the expectation for an electron, being linear with the measurement rotation angle and forming the expected spiral pattern. Other than the small phase imperfections there is a slight sinusoidal bias to the phase data, which we determined is the result of a small ($\lesssim$ 1 mm) offset of the antenna's phase center from the rotation axis of the automated stages.



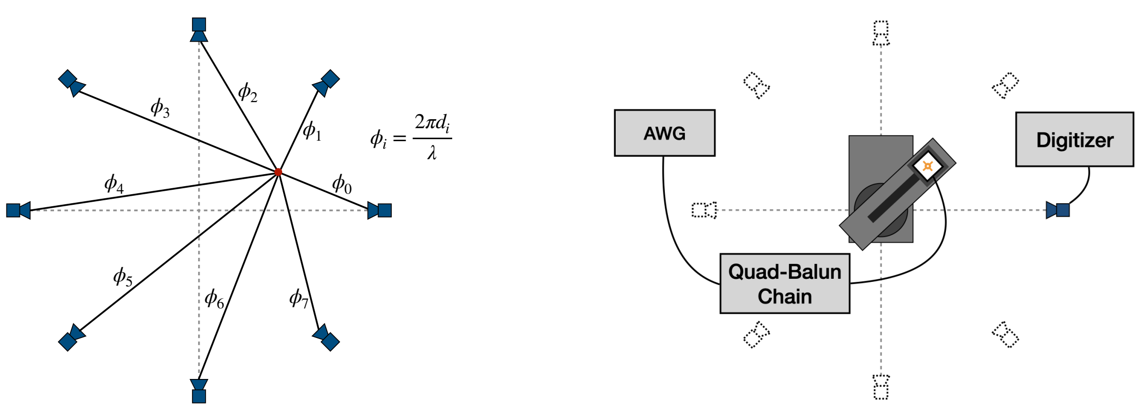
(a) (b)

Figure 5.20: Linear interpolations of the measured electric field magnitude (a) and phase (b). The data was acquired using a VNA at 120 points spaced by 3 degrees from 0 to 357 degrees of azimuthal angle. The different color lines indicate the vertical offset of the horn antenna relative to the SYNCA PCB and the dashed line shows the expected shape from electron simulations. No significant difference in the antenna pattern is observed for the measured vertical offsets.

The characterization measurements confirm the simulated performance of the SYNCA. As expected the fields generated by the antenna are nearly isotropic in magnitude, ϕ-polarized, and are linearly out of phase around the circumference of the antenna as predicted for cyclotron radiation in Section 5.3.2. Small imperfections in the magnitude and phase of the antenna are expected, particularly at the antenna's high operating frequency of 26 GHz where small geometric changes can have significant impacts on electrical properties. However, calibration through careful characterization measurements can be used to remove the majority of these pattern imperfections, including the relatively large pattern null near 45°, which will allow for the usage of the SYNCA as a test source for free-space CRES experiments utilizing antenna arrays. In the next section we use the VNA measurements obtained here as a calibration for signal reconstruction using digital beamforming.

5.3.5 Beamforming Measurements with the SYNCA

Digital beamforming is a standard technique for signal reconstruction using a phased array [93]. The SYNCA, since it exhibits the same cyclotron phases as a trapped electron, can be used to perform simulated CRES digital beamforming reconstruction experiments on the bench-top without the need for the magnet, cryogenics, and vacuum systems required by a full CRES experiment. The fields received by the individual elements of the antenna array will have phases dependent on the spatial position of the source relative to the antennas. Therefore, a simple summation of the received signals will fail to reconstruct the signal due to destructive interference between the individual channels in the array. However, applying a phase shift associated with the source's spatial position



(a) (b)

Figure 5.21: (a) A depiction of the relative phase differences for signals received by a circular antenna array from an isotropic source. The phases correspond to a unique spatial position. (b) A schematic of the setup used to perform digital beamforming.

removes phase differences and results in a constructive summation of the channel signals (see Figure 5.21). We can summarize the digital beamforming operation succinctly using the following equation

$$y[t_n] = \sum_{m=0}^{N-1} x_m[t_n] A_m e^{i\phi_m}, \tag{5.35}$$

where $y[t_n]$ represents the summed array signal at time t_n, $x_m[t_n]$ is the signal received by channel m at time t_n, ϕ_m is the phase shift applied to the signal received at channel m, and A_m is an amplitude weighting factor that accounts for the different signal power received by individual channels. By changing the digital beamforming phases, the point of constructive interference can be scanned across the sensitive region of the array to search for the location of a radiating source, which is identified as the point of maximum summed signal power above a specified threshold. The digital beamforming phases consist of two components,

$$\phi_m = 2\pi d_m/\lambda + \theta_m, \tag{5.36}$$

where d_m is the distance from the m-th array element to the source, and θ_m is the relative angle between the source position and the m-th antenna. The first component is the standard digital beamforming phase that corresponds to the spatial position of the source, and the second component is the cyclotron phase that corresponds to the relative azimuthal phase offset.

With a small modification to the hardware used to characterize the SYNCA (see Figure 5.19), we can perform a digital beamforming reconstruction of a synthetic CRES event. By replacing the VNA with an arbitrary waveform generator (AWG), the SYNCA can be used to generate cyclotron radiation with an arbitrary signal structure, which can then be detected by digitizing the signals received by the horn antenna. Rotational symmetry allows us to use the rotational stage of the positioning system to rotate the SYNCA to recreate the signals that would have been received by a complete circular array of antennas.

Using this setup, signals from a 60 channel circular array of equally spaced horn antennas were generated with the SYNCA positioned 10 mm off the central array axis, reconstructed using digital beamforming, and compared to Locust simulation (see Figure 5.22). When the cyclotron spiral phases are not used, which is equivalent to setting θ_m in Equation 5.36 to zero, the SYNCA's position is reconstructed as a relatively faint ring as predicted by simulation. However, when the appropriate cyclotron phases are used during the beamforming procedure, both the simulated electron and the SYNCA appear

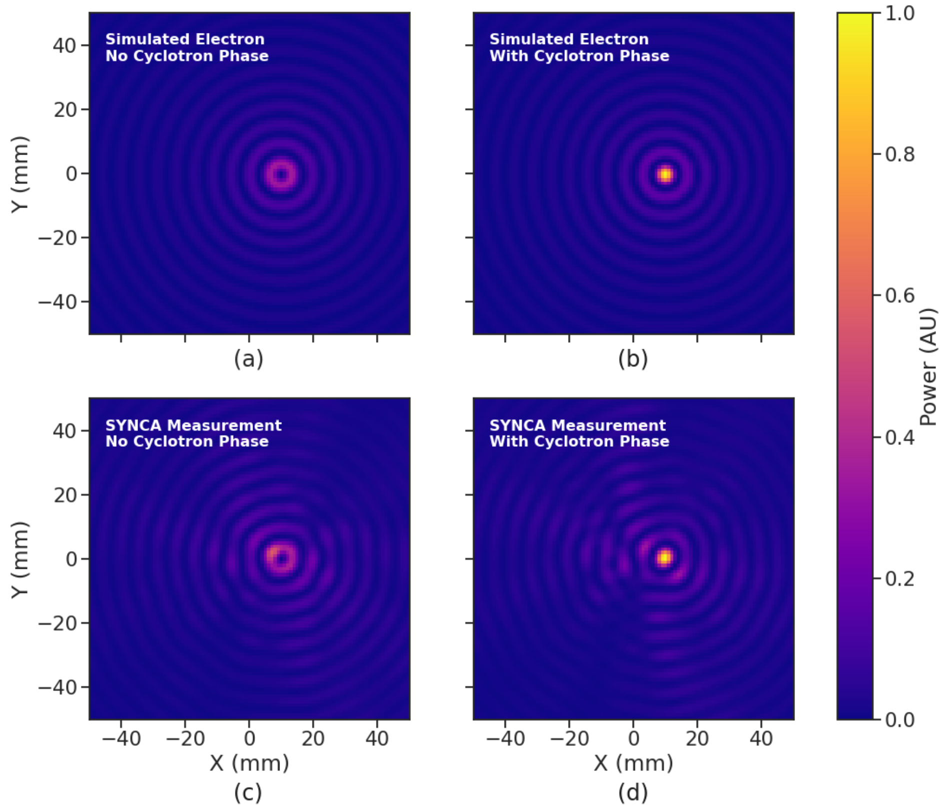
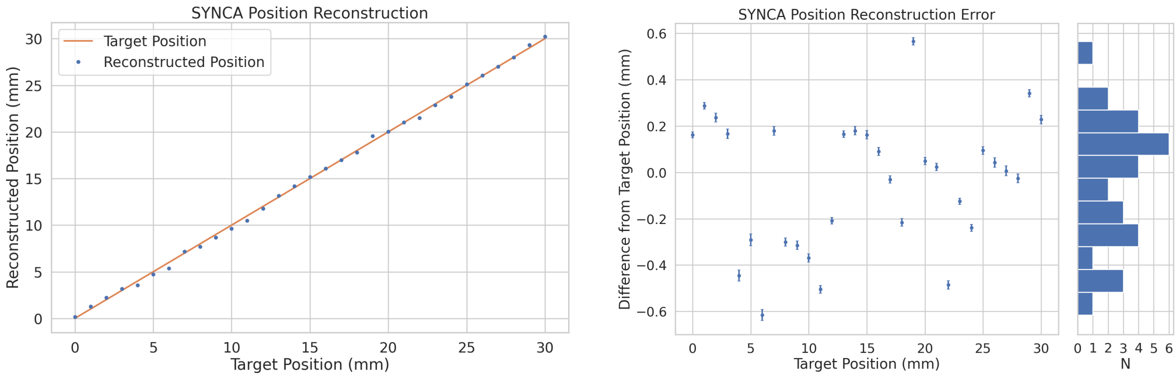


Figure 5.22: Digital beamforming maps generated using a simulated 60 channel array and electron simulated using the Locust package. (a) and (b) show the beamforming maps for simulated electrons without the cyclotron spiral phases and with the cyclotron spiral phases respectively. (c) and (d) show the beamforming maps produced from SYNCA measurements. We observe good agreement between simulated electrons and the SYNCA measurements.

as a single peak of high relative power corresponding to the source position. Therefore, we observe good agreement between the simulated and SYNCA reconstructions. While it may seem that for the case with no cyclotron phase corrections the ring reconstructs the position of the electron as effectively as beamforming with the cyclotron phase corrections, it is important to note that the simulations and measurements were generated without a realistic level of thermal noise. The larger maxima region and lower signal power, which occurs without the cyclotron phase corrections, significantly reduce the probability of detecting an electron in a realistic noise background.

To bound the beamforming capabilities of the synthetic array of horn antennas, we performed a series of beamforming reconstructions where the SYNCA was progressively moved off the central axis of the array (see Figure 5.23). To extract an estimate of the



(a) (b)

Figure 5.23: A plot of the SYNCA's reconstructed position using the synthesized horn-antenna array and digital beamforming. (a) Shows the reconstructed position of the SYNCA compared with the target position indicated by the positioning system readout. (b) Shows the reconstruction error, which is the difference between the target and reconstructed positions. The error bars in (b) are the uncertainty in the mean position of the 2D Gaussian used to fit the digital beamforming reconstruction peak obtained from the fit covariance matrix. The mean fit position uncertainty of 0.02 mm is an order of magnitude smaller than the typical reconstruction error of 0.3 mm obtained by calculating the standard deviation of the difference between the reconstructed and target position.

position of the SYNCA using the digital beamforming image we apply a 2-dimensional (2D) Gaussian fit to the image data and extract the estimated centroid value. We find that the synthetic horn antenna array reconstructs the position of the SYNCA with a 1σ-error of 0.3 mm with no apparent trend across the 30 mm measurement range. This reconstruction error is an order of magnitude larger than mean fit position uncertainty of 0.02 mm indicating that systematic effects related to the SYNCA positioning system could be contributing additional uncertainty to the measurements. Note that the current mean reconstruction error of 0.3 mm is a factor of 20 smaller than the full width at half maximum of the digital beamforming peak (6 mm), which could be interpreted as a naive estimate of the position reconstruction performance of this technique. Because these measurements are intended as a proof-of-principle demonstration, we do not investigate potential sources of systematic errors further; however, we expect that a similar and more thorough investigation will be performed using the Project 8 antenna array test stand, where typical reconstruction errors can be used to estimate the energy resolution limits of antenna array designs.

5.3.6 Conclusions

In this paper we have introduced the SYNCA, which is a novel antenna design that emits radiation that mimics the unique properties of the cyclotron radiation generated by charged particles moving in a magnetic field. The characterization measurements of the SYNCA validated the simulated performance of the PCB crossed-dipole antenna design. Additionally, the SYNCA was used to estimate the position reconstruction capabilities of a synthesized array of horn antennas and experimentally reproduced the simulated digital beamforming reconstruction of electrons.

While the SYNCA performs well, there exist discrepancies in the phase and magnitude of the radiation pattern compared to the simulated SYNCA design that are related to the small geometric differences in the soldered connections. Future design iterations that replace the soldered connections with a fully surface mount design could improve the radiation pattern at the cost of some complexity and expense. Furthermore, improving the design of the antenna PCB and mounting system would allow the antenna to be inserted into a cryogenic and vacuum environment where in-situ antenna measurement calibrations could be performed.

The discrepancies in the radiation pattern and phases exhibited by the as-built SYNCA should not greatly impact its performance as a calibration probe. Both magnitude and phase variations can be accounted by applying the SYNCA characterization measurements as a calibration to the data collected by the antenna array test stand. The separate calibration of the SYNCA radiation does not impact the primary goals for the antenna array test stand which are array calibration and signal reconstruction algorithm performance characterization, because it can be performed with standard reference horn antennas with well understood characteristics.

The SYNCA antenna technology advances the CRES technique by providing a mechanism to characterize free-space antenna arrays for CRES measurements without the need for a magnet and cryogenics system, which would be required for calibration using electron sources. Both the Project 8 collaboration as well as future collaborations which are developing antenna array based CRES experiments can make use of SYNCA antennas as an important component of their calibration and commissioning phases.

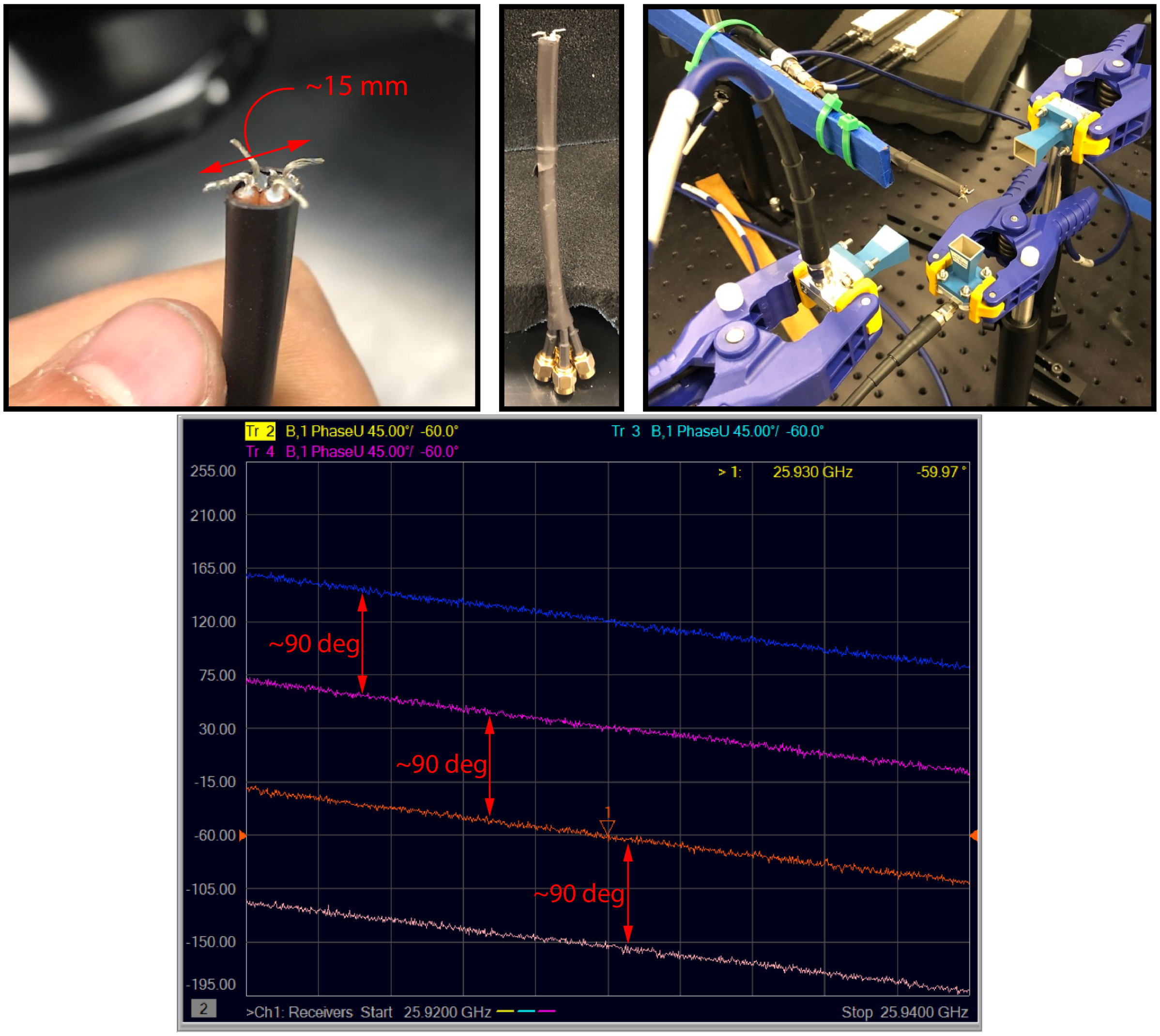


Figure 5.24: Images of an early prototype crossed-dipole antenna manufactured by hand and the first measurement setup. The antenna was constructed by hand using four stripped coaxial cables. The antenna was connected to one port of the VNA, and the remaining three ports on the VNA were connected to horn antenna arranged with 90 deg offsets around the crossed-dipole. The measured unwrapped S-parameter phases exhibit the desired relative phase behavior for a SYNCA. These early measurements were the first laboratory proof-of-principle for the crossed-dipole SYNCA.

5.4 SYNCA Development Discussion

A crossed-dipole antenna (see Figure 5.24) was identified early on as a candidate SYNCA design. The crossed-dipole is a circularly polarized antenna, consequently, the electric fields measured in the plane of the dipole antenna exhibit the same relative phase offsets as a 90° electron in a magnetic trap. This is explained in greater detail in Section 5.3.

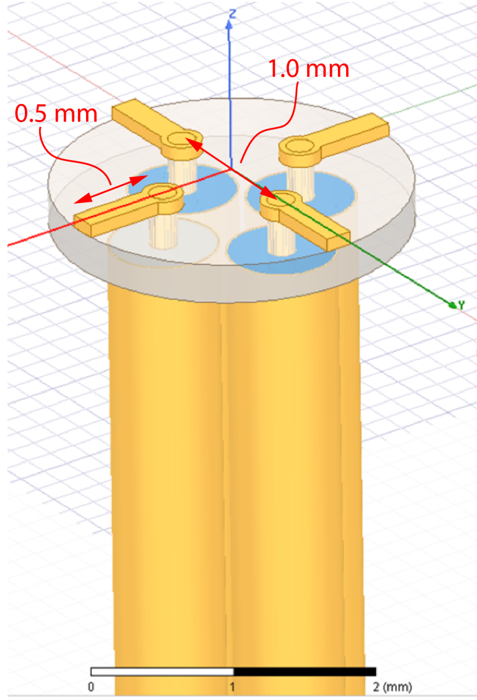


Figure 5.25: An early iteration of a crossed-dipole SYNCA antenna simulated in HFSS. The antenna is electrically small at 26 GHz, which requires dipole arms on the order of 1 mm long. This design is limited by the minimum achievable distance between the dipole arms caused by the available diameters of coaxial cables. The assumed termination scheme for the coaxial cables to the antenna is hand-soldering, which introduces random variation in the antenna pattern from the inevitable blobs of solder left on the surface of the PCB.

These phase offsets were measured with the first rudimentary crossed-dipole prototype manufactured from coaxial cables with the insulation and shield stripped away.

Because the SYNCA is ultimately a calibration tool, it is desireable that the antenna have a well-characterized and robust antenna pattern. Therefore, manufacturing a SYNCA using the stripped wire method shown in Figure 5.24 is infeasible. Studies of crossed-dipole antennas manufactured out of printed circuit boards were performed using HFSS to identify an antenna design that imitated an electron, while being more robust and simpler to manufacture (see Figure 5.25).

Identifying a design that was robust, manufacturable, and most importantly matched the electric fields of a trapped electron proved to be a non-trivial task. The primary factor driving the difficulty was the high operating frequency of the antenna (26 GHz) combined with the requirement that the antenna be electrically-small. An antenna that is electrically-small at 26 GHz has a largest dimension on the order of 1 mm, which poses significant manufacturability challenges given the limited available budget for SYCNA fabrication.

One of the key limitations with the small size requirements is the diameter of the coaxial cables needed to feed the crossed-dipole antenna. The smallest commonly available rigid coaxial cables available on the market have diameters of approximately 0.5 mm, which limited the spacing between dipole arms to a minimum of about 1 mm. The crossed-dipole antenna performs better as a SYNCA if the dipole arm separation is significantly less than the operating wavelength. Therefore, the high operating frequency ultimately limited how well the SYNCA could mimic an electron. If the desired cyclotron frequency was lowered by an order of magnitude to approximately 3 GHz a significantly higher quality SYNCA could be manufactured at lower cost.

The decision to use coaxial transmission lines terminated on the antenna PCB with a hand-soldered connection was driven primarily to limit the costs of SYNCA development and contributed to the observable variations in the SYNCA's gain and phase patterns. A second iteration of the SYNCA design that minimized hand-soldering by using surface-mount components could significantly reduce variations in the antenna pattern. The major drawback in the development of a surface-mount SYNCA is the cost, and given the transition to a cavity based design for Phase IV, such a design was never investigated.

5.5 FSCD Antenna Array Measurements with the SYNCA

5.5.1 Introduction

Using the SYNCA we can perform full-array measurements of prototype versions of the FSCD antenna array to test its performance with a realistic cyclotron radiation source (see Figure 5.26). The goal is to check how the measured power received by the array compares to FSCD simulations as a function of the radial and axial position of the SYNCA. These measurements are intended to validate the antenna research and development by Project 8, which has been driven primarily by simulations with Locust [60] and CRESana (see Section 4.2.3), and identify any discrepancies with these simulations tools. This knowledge will provide confidence in the simulations necessary for the analysis of the sensitivity of larger antenna array based CRES experiment designs to the neutrino mass.

As shown in Section 5.3, the SYNCA does have some radiation pattern imperfections that complicate the comparison between measurement and simulation data. One way to

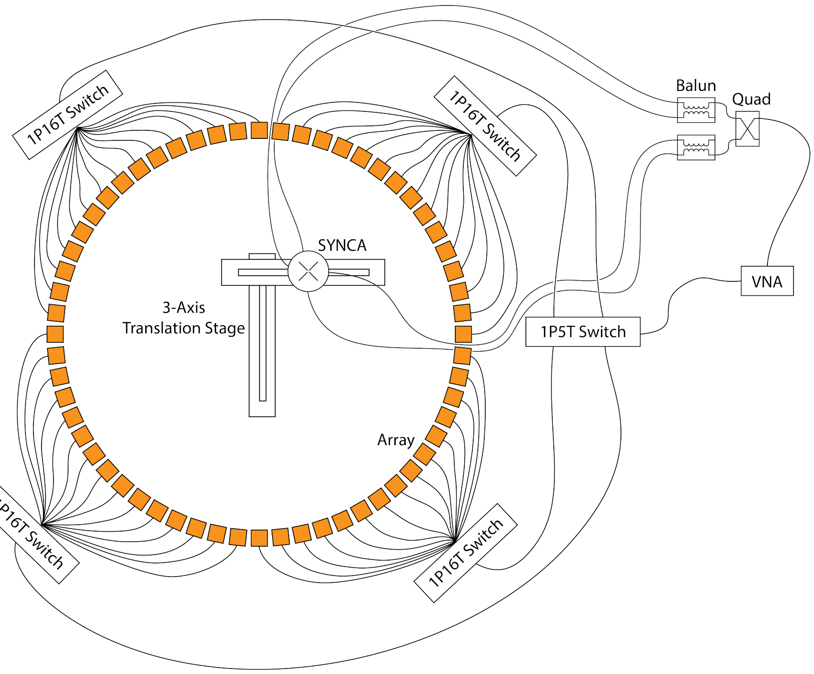


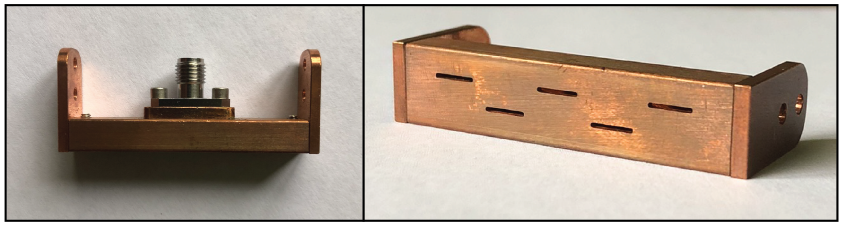
Figure 5.26: A diagram of the array measurement system used to test the prototype FSCD antenna array. A VNA is used as the primary measurement tool, which is connected to the array through a series of switches. The other port of the VNA connects to the SYNCA through the quad-balun chain used to provide the SYNCA feed signals. During measurements the SYNCA is positioned inside the center of the antenna array and translated to different radial and axial positions using a 3-axis manual translation stage setup.

disentangle some of the effects of these imperfections is to perform an additional set of measurements using a synthetic antenna array setup along with the SYNCA antenna. Since the synthetic array setup uses only a single array antenna, the data should be free of errors associated with individual antenna differences and multi-path interference, which are two error sources being tested with the full-array setup. By comparing the synthetic array data to the FSCD array data and to simulation data one can evaluate the significance of these effects relative to the errors introduced by SYNCA imperfections.

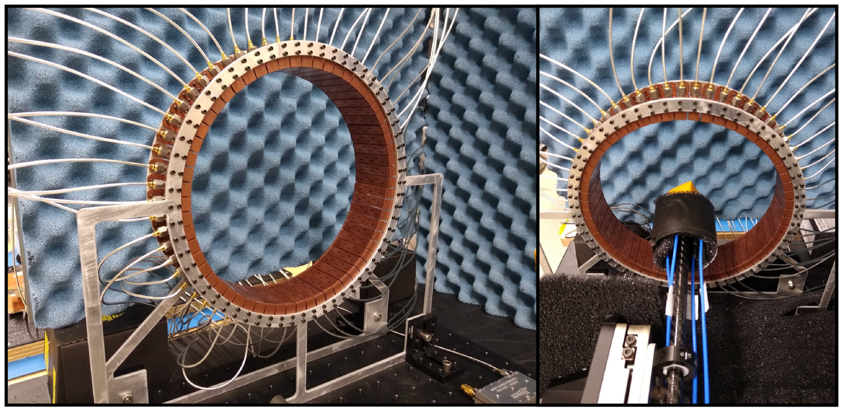
5.5.2 Measurement Setups

5.5.2.1 FSCD Array Setup

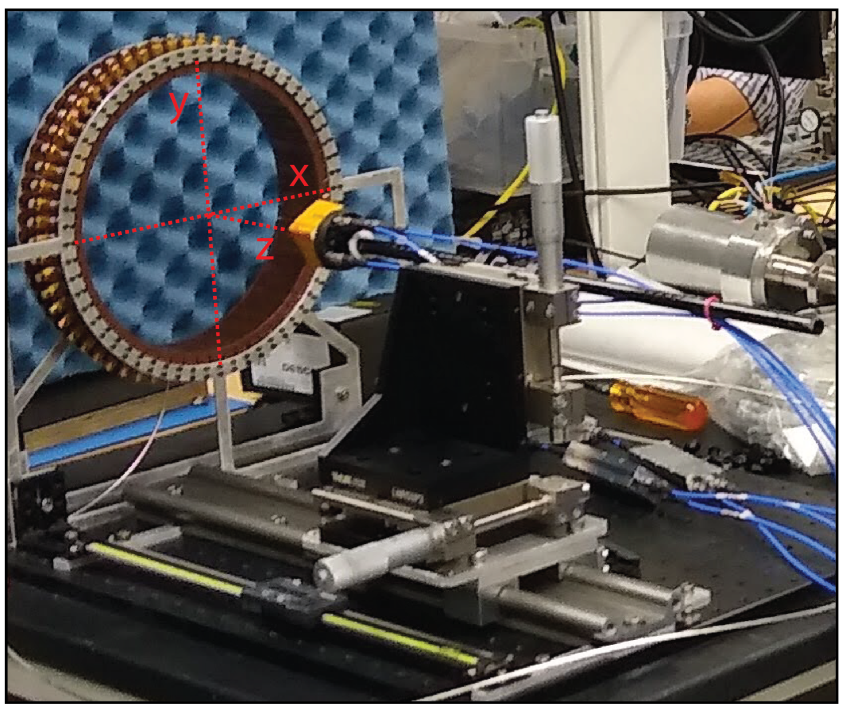
The antenna design that composes the array is the 5-slot waveguide antenna developed for the FSCD experiment (see Figure 5.27a). The antenna is 5 cm long and is constructed out of WR-34 waveguide with a 2.92 mm coax connector located at the center of the



(a)



(b)



(c)

Figure 5.27: Photos of the prototype FSCD antenna (a), the FSCD array and SYNCA (b), and the translation stages and coordinate system used to position the SYNCA (c).

antenna. Copper flanges located on both ends of the antenna are used to mount the antenna in the array support structure. The antennas are supported by two circular steel brackets that can be bolted to both ends of the waveguide to construct the circular array (see Figure 5.27b). The antenna array consists of sixty identical waveguide antennas with a radius of 10 cm. The array is mounted perpendicular to an optical breadboard surface using a pair of the steel brackets, which provide sufficient space for the coaxial cable connections and allows for easy positioning of the SYNCA antenna. The SYNCA is mounted on the end of a carbon fiber rod attached to a set of manual translation stages, which are used to move the SYNCA antenna to different positions inside the array (see Figure 5.27c). The stages allow for independent motion in three different axes and can position the SYNCA at radial distances up to 5 cm from the center.

Data acquisition is accomplished using a two-port VNA in combination with a series of microwave switches that allow the VNA to connect to each channel in the array . The first port of the VNA is connected to the quad-balun chain used to feed the SYNCA (see Section 5.3), and the second port of the VNA connects to a 1P5T microwave switch. The 1P5T switch is connected to four separate 1P16T switch boards that connect directly to the array. The data acquisition is controlled by a python script running on a lab computer, which is connected to the VNA and an Arduino board programmed to control the microwave switches. The script uses the switches to iteratively connect each of the antennas in the array to the VNA. The VNA is configured to load a specific calibration file for each antenna channel and performs the measurements of all available S-parameters. The separate calibration files is an attempt to remove phase and magnitude errors caused by different propagation through the RF switches. Array measurements were performed for the set of SYNCA positions consisting of radial (x-axis) positions from 0 to 50 mm in 5 mm steps and axial (z-axis) positions from 0 to 50 mm in 5 mm steps resulting in 121 array measurements. At each SYNCA position we measured the two-port S-parameter matrix using a linear frequency sweep from 25.1 to 26.1 GHz with 101 discrete frequencies.

5.5.2.2 Synthetic Array Setup

A photograph of the setup used to perform the synthetic array measurements is shown in Figure 5.28. One important difference between this setup and the FSCD array setup is that the synthetic array measurements were performed with a waveform generator and digitizer instead of a VNA. The electronics configuration is identical to the diagram in Figure 5.7b. Despite the differences, one is still able to compare the measured phases of the synthetic array and the relative magnitude of the power, since the digitized signal power is directly proportional to S21.

The arbitrary waveform generator in the setup is configured to produce a 64 MHz sine wave signal that is up-converted to 25.864 GHz using a mixer and the VNA source. This signal is passed through a bandpass filter and fed to the SYNCA quad-balun chain. A single FSCD antenna is positioned 10 cm from the SYNCA and aligned vertically so that the center of the 5-slot waveguide is in the plane of the SYNCA PCB (see Figure 5.28). This position corresponds to $z = 0$ in Figure 5.27c. The SYNCA is rotated

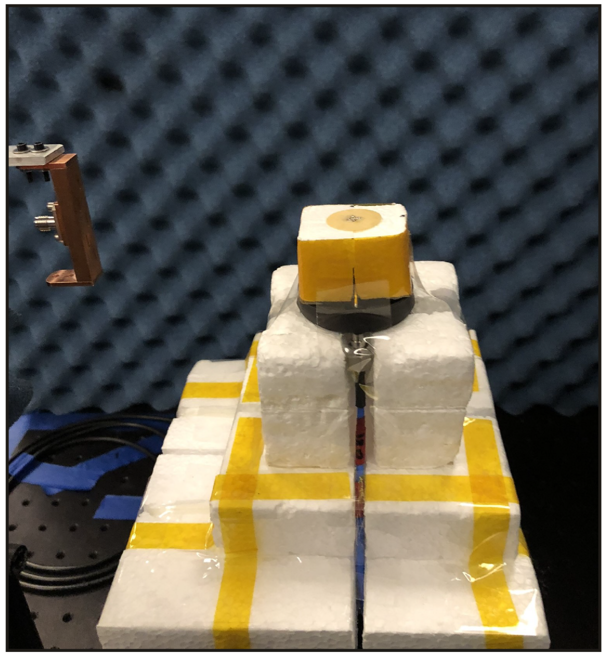


Figure 5.28: A photo of the FSCD antenna and the SYNCA in the synthetic array measurement setup at Penn State.

in three degree steps to synthesize an antenna array with 120 channels. This channel count is more than could physically fit in a 10 cm radius array, but there is no cost to over-sampling. Additionally, over-sampling allows for a check of the smoothness of the antenna array radiation pattern. The signals from the FSCD antenna are down-converted using the second mixer connected to the VNA source before being digitized at 250 MHz and saved to disk. Several synthetic array measurement scans were performed by using the linear translation stage to change the radial position of the SYNCA. In total eight scans were taken from 0 to 35 mm using a radial position step size of 5 mm.

5.5.3 Simulations, Analysis, and Results

The Locust and CRESana simulation packages utilize the antenna transfer functions to calculate the power that would be received by each antenna from a CRES electron. The equivalent quantity in the measurement setup is the S21 matrix element, which indicates the ratio of the power received by an antenna in the array to the amount of power delivered to the SYNCA. Therefore, the analysis focuses on comparing the relative magnitudes and phase of the S21 parameters measured by the VNA as a function of the array channel and the SYNCA position. Additionally, we apply a beamforming reconstruction to the S21 data to evaluate how the summed power and beamforming images change as a function of the position of the SYNCA.

5.5.3.1 Simulations

Simulations for the FSCD array measurements were performed using CRESana, which performs analytical calculations of the EM-fields produced by an electron at the position of the antennas. At each sampled time CRESana computes the electric field vector at the antenna positions, which is projected onto the antenna polarization axis to obtain the co-polar electric field. The magnitude of the co-polar electric field is then multiplied by a flat antenna transfer function to calculate the corresponding voltage signal. CRESana simulations exploit the flat transfer functions of the FSCD antennas, which allows the electric field to be multiplied by the antenna transfer function rather than performing the full FIR calculation. These calculations produce a voltage time-series for each of the antennas in the array that can be compared to the laboratory measurements.

CRESana was configured to simulate a 90° electron in a constant background magnetic field of ≈ 0.958 T with a kinetic energy of 18.6 keV. These parameters were chosen in order to mimic a CRES event near the tritium beta-decay spectrum endpoint in the FSCD experiment. The constant background magnetic field guarantees that the guiding center of the electron is stationary across the duration of the simulation which is consistent with the SYNCA in the laboratory measurements. Simulations were performed with the electron's guiding center at radial positions from 0 to 45 mm in steps of 1 mm and axial positions from 0 to 30 mm in steps of 1 mm. The simulations generated time series consisting of 8192 samples at 200 MHz for the sixty channel FSCD antenna array geometry.

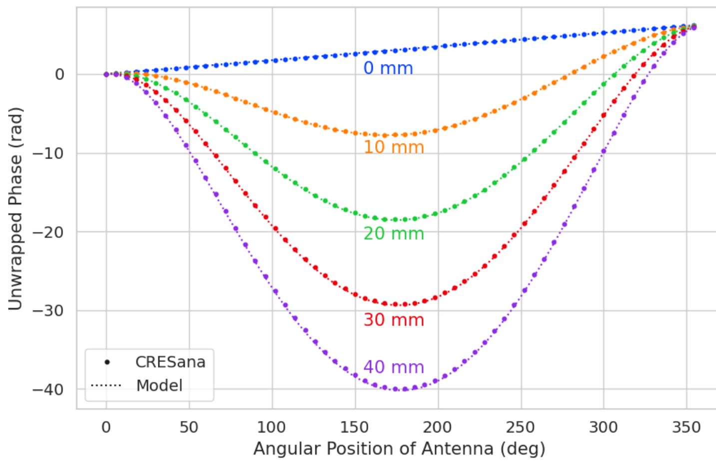


Figure 5.29: The unwrapped phases of signals received by the FSCD antenna array from an electron with a 90° pitch angle located in the plane of the antenna array. The data points indicated the phases extracted from simulation and the dashed lines show the model predictions.

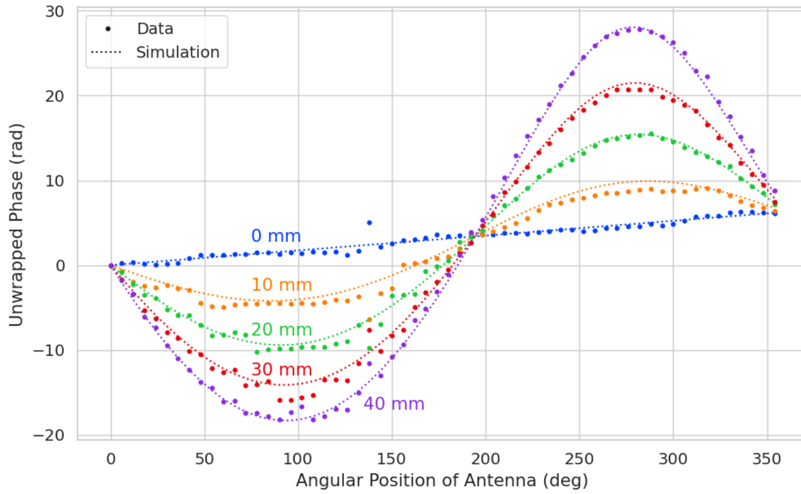
5.5.3.2 Phase Analysis

Correct modeling of the signal phases is fundamental to reconstruction for both beamforming and matched filter approaches. The beamforming reconstruction relies on a signal phase model developed from Locust simulations, which allows one to predict the relative signal phases for a specific magnetic trap and electron position. The equation for the model is

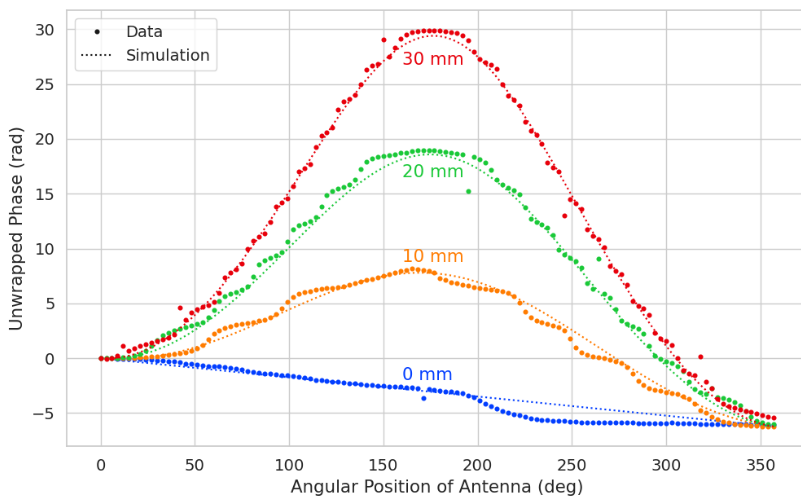
$$\phi_{ij}(t) = \frac{2\pi d_{ij}(t)}{\lambda} + \theta_{ij}(t), \tag{5.37}$$

where $d_{ij}(t)$ is distance between the assumed electron position and the antenna position, and $\theta_{ij}(t)$ is the angular separation between the electron and antenna positions. For details on the components of the phase model see Section 5.3.2. In Figure 5.29 we compare the phases predicted by Equation 5.37 to phases extracted from CRESana simulations of an electron located in the plane of the antenna array at a series of radial positions. One observes excellent agreement between the model and simulation.

The measured signal phases from the FSCD array and synthetic array are shown in Figures 5.30a and 5.30b compared to the signal phase model. The axial position of the SYNCA in both plots is $z = 0$ mm, such that the plane of the PCB is aligned with the center of the FSCD antenna. The data shown in Figure 5.30a corresponds to the S-parameters measured at 25.80 GHz which is the frequency closest to the one used in



(a)



(b)

Figure 5.30: Plots of the measured unwrapped phases from the FSCD array (a) and the synthetic array (b) compared to the model predictions for a series of radial positions. The different phases of the sinusoidal phase oscillations in the two plots reflects differences in the coordinate systems of the measurements.

the synthetic array setup. The different slope and sinusoidal phases exhibited by Figure 5.30a and 5.30b reflects differences in the coordinate system for each setup. In general, we see that the phase model predicts the large scale features of the phases quite well, but there are some small scale deviations or errors from the phase model that do not appear to be present in simulation.

A comparison of the phase errors, which are the difference between measurement and

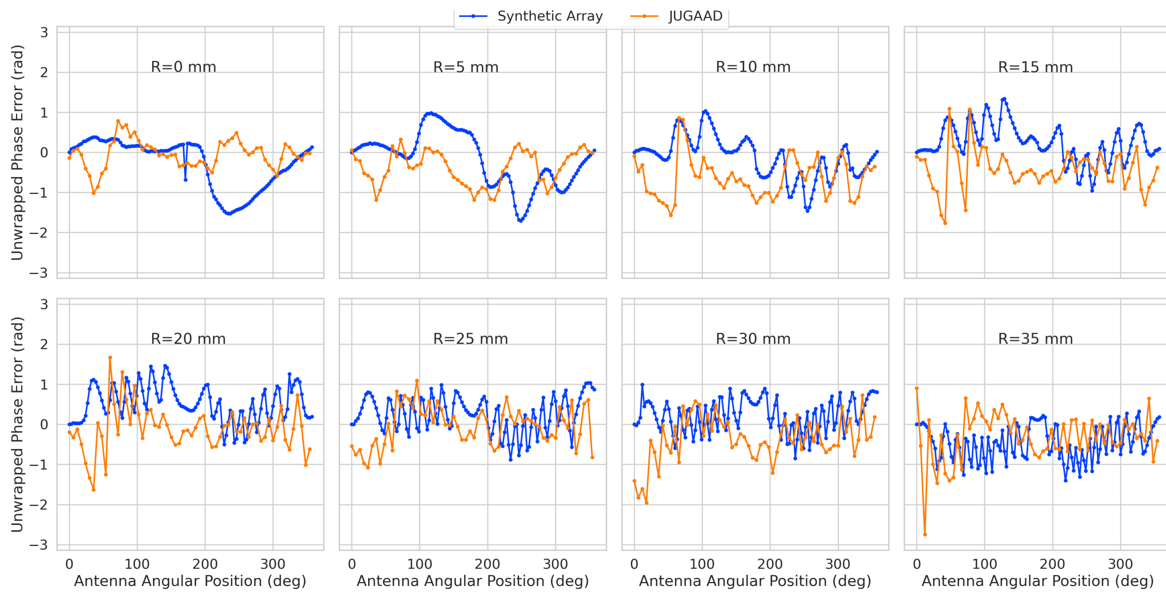
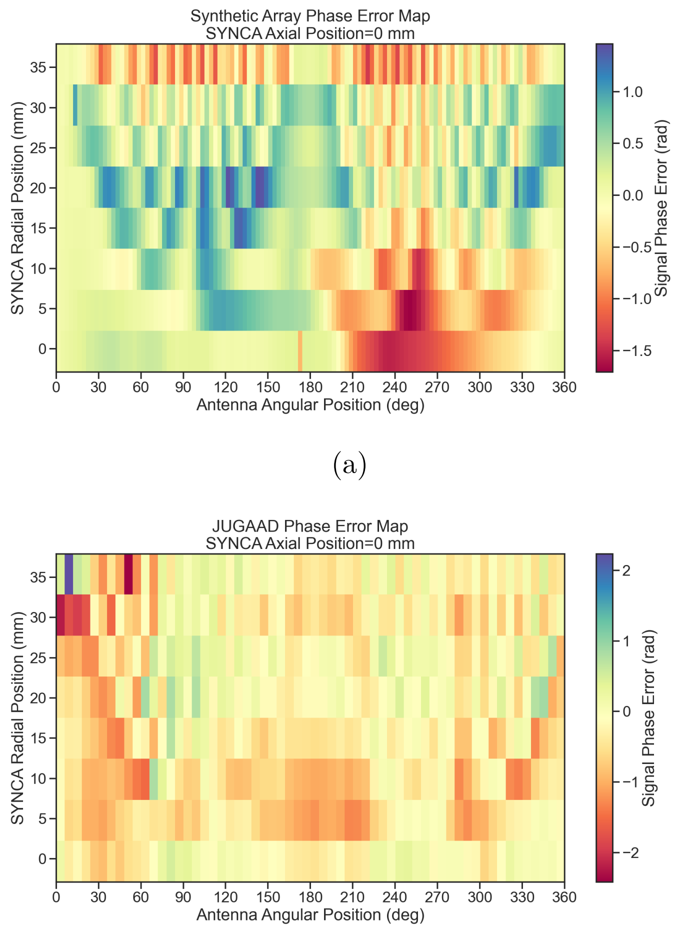


Figure 5.31: The phase errors between the measurement and model for the synthetic array (blue) and the FSCD array (orange) for a series of radial positions. The label JUGAAD refers to an alternative name for the FSCD array setup. As the SYNCA is translated off-axis phase errors with progressively higher oscillation frequency enter into the measurements.

model is shown in Figure 5.31. The FSCD array data is referred to as the JUGAAD data in the plot legend, which is an alternative name for the FSCD array setup.

The phase error at $R = 0$ in Figure 5.31 forms a smooth curve, with the exception of an outlier data point caused by a bug in the data acquisition script. One can attribute the observed phase error at this position to imperfections in the antenna pattern of the SYNCA. As the SYNCA is moved away from $R = 0$ mm one observes that the phase error exhibits oscillations whose frequency increases as a function of the radial position of the SYNCA. These oscillations have the appearance of a diffraction pattern, which is particularly clear for the radii ≥ 15 mm, due to the bilateral symmetry of the phase error peaks around 180°.

One can observe a higher average variance in the phase errors measured for the FSCD array compared to the synthetic array. This is best seen by comparing the curves at $R \leq 15$ mm where the smooth synthetic array curves are distinct from the relatively noisy FSCD array errors. The extra noise in the FSCD array is most likely caused by differences in the radiation patterns of the antennas that make up the array as well as differences in the transmission lines through the switch network that introduce additional phase errors into the measurement. Since the synthetic array measurements use only a single antenna, these extra error terms are not present, which explains the relatively smoother phase error curves. Despite the extra phase errors in the FSCD array, it is still possible to observe a similar phase error oscillation effect as the SYNCA is moved away from $R = 0$ mm.



(a)

(b)

Figure 5.32: Two dimensional plots of the phase errors for the synthetic array (a) and the FSCD (JUGAAD) array (b). In both plots we observe evidence of a similar diffraction pattern with bilateral symmetry, but the FSCD array measurments have an additional phase error contribution from the different antennas and paths through the switch network.

The diffraction pattern exhibited by the phase error oscillations is more easily observed by plotting the phase errors in a two-dimensional map, which is done in Figures 5.32a and 5.32b. For the synthetic array data ones observes a relatively clear diffraction pattern that emerges as the SYNCA is moved radially. The bilateral symmetry of the diffraction patterns is due to the bilateral symmetry of the circular synthetic array around the translation axis of the SYNCA. A similar pattern is also visible in the FSCD array data, although, it is obscured by the additional phase error that results from the multi-channel array.

The physical origin of the phase error diffraction pattern is attributed to interference effects arising from path-length differences between the individual slots in the FSCD antenna and the SYNCA transmitter. Since we are operating in the radiative near-field of the FSCD antenna, the path length differences between the slots introduces a significant change in the summation of the signals that occurs inside the waveguide, which causes the radiation pattern of the antenna to change as a function of distance. Therefore, when the SYNCA is positioned off-axis the different path-lengths from the SYNCA to each antenna results in different radiation patterns leading to the observed diffraction pattern.

This near-field effect is not present in simulations, because in order to simplify the calculations we assume that the far-field approximation can be applied to the FSCD antennas. This means that the radiation pattern and antenna transfer functions are independent of the distance between the transmitter and the receiving antenna. In principle, we can account for these near-field effects with a more detailed simulation of the FSCD antennas either in CRESana or Locust, which would result in an additional term in the beamforming phase model. However, this would increase the computational intensity of the simulation software. In the next section we briefly discuss the impact of these near-field effects on the measured magnitude of the signals.

5.5.3.3 Magnitude Analysis

Exactly as for the signal phase, one can use simulations to construct a model that describes the magnitude of the signals received by each channel in the antenna array. By examining the results of simulations or by analyzing the Liénard-Wiechert equation one can show that radiation pattern from a 90° pitch angle electron in a magnetic field is omni-directional. Therefore the relative magnitudes of the signals received by each channel will be determined by the free-space power loss, which is proportional to the inverse distance between the assumed electron position and the antenna.

A consequence of this is that the signals produced in the array for electrons off the central axis will have larger amplitudes for the antennas closer to the electron compared to those which are further away. The amplitudes of the signals received by the array from an electron located at a series of radial positions are shown in Figure 5.33.

One expects to see a similar trend in the signal magnitudes in both the FSCD and

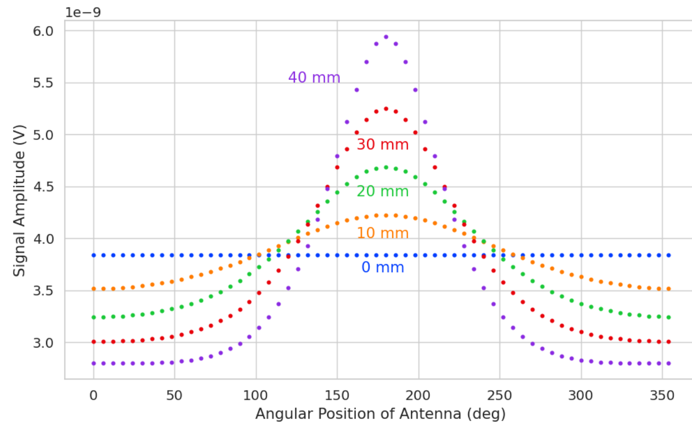


Figure 5.33: The amplitude of the signals from CRESana for the FSCD array from a 90° electron. As the electron is moved from $R = 0$ the signals begin to have unequal amplitudes depending on the distance from the electron to the antenna.

synthetic arrays. The normalized signal magnitudes extracted from the full and synthetic array setups for a series of radial SYNCA positions are shown in Figure 5.34. The data corresponds to a SYNCA axial position of $z = 0$ mm and at a frequency 25.86 GHz. One complication is that the radiation pattern of the SYNCA is not perfectly omni-directional, which causes the measured magnitudes at $R = 0$ mm to diverge from the perfectly flat behavior exhibited by electrons.

As the SYNCA is moved off-axis one observes a similar increase in the number of magnitude peaks in the synthetic array data that one would expect from a diffraction pattern, although this trend is not as stark compared to the phase data. Noticeably, there does not appear to be a set of channels with disproportionately larger amplitude at large R, which would be expected based on the trends from CRESana.

Comparing the magnitudes of the synthetic array to the FSCD array in Figure 5.34 we see that there is a similar amount of variability in the magnitudes at $R = 0$ mm, although there is potentially more small scale error in the magnitude curve caused by channel differences in the FSCD array. We observe a similar trend in the number of magnitude error peaks in the FSCD array data to the synthetic array data, which mirrors the diffraction effect observed in the phase data. The diffraction effect can be visualized more clearly by plotting a similar two-dimensional map of the magnitudes (see Figure 5.35).

The fact that one observes a similar diffraction pattern in the signal magnitudes as a function the SYNCA position reinforces the conclusions from the phase analysis that near-field effects are having a significant impact on the radiation pattern of the

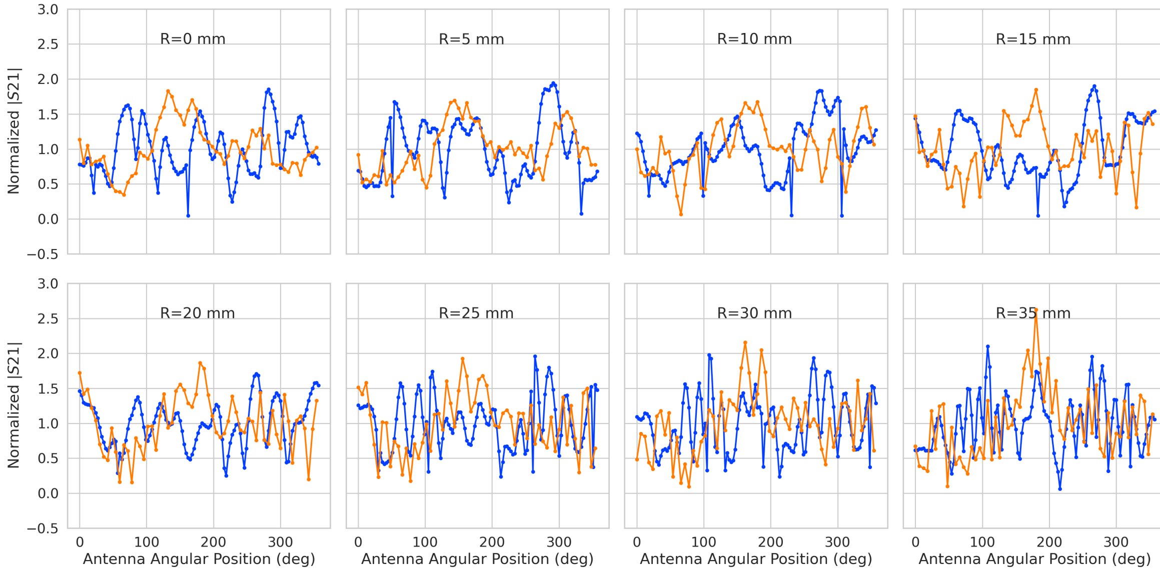
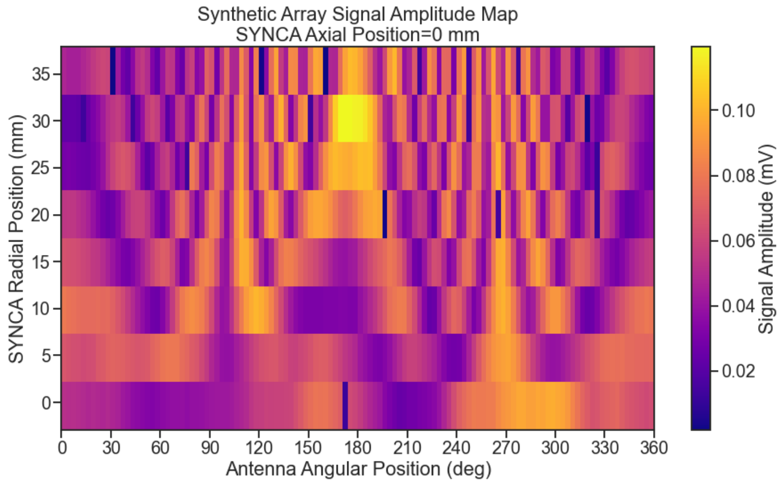


Figure 5.34: The normalized magnitudes of the S21 parameters measured in the FSCD (orange) and synthetic array (blue) setups. The dominant observed behavior as a function of radius is the increase in the number of magnitude peaks, which was noted in the phase error curves. There does not appear to be a strong change in the relative amplitude of a group of antennas as predicted by CRESana.

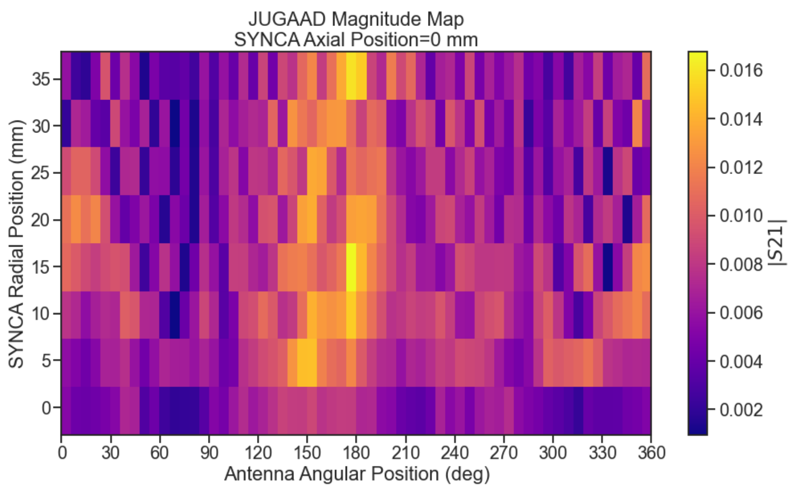
FSCD array. These near-field effects lead to changes in the magnitude and phase of the radiation pattern of the FSCD antenna as a function of distance. If left uncorrected these errors reduce detection efficiency by causing power loss in the beamforming or matched filter reconstruction due to phase mismatch. We explore the impact of these phase and magnitude errors on beamforming in the next section.

5.5.3.4 Beamforming Characterization

Errors in the signal magnitudes and phases lead to errors in signal reconstruction. For example, a matched filter reconstruction requires accurate knowledge of the signals in each channel to achieve optimal performance. Uncorrected errors leads to mismatches between the template and signal, which reduces detection efficiency and introduces uncertainty in the parameter estimation. In this section, we analyze the beamformed signal amplitude as a function of the position of the SYNCA to quantify the impact of the phase and magnitude errors on signal reconstruction. Because of the imperfections in the SYNCA source, it is inappropriate to directly compare the beamformed signal amplitude of the FSCD array or synthetic array. Such a comparison would not allow one to disentangle losses that occur because of the antenna array from those that occur



(a)

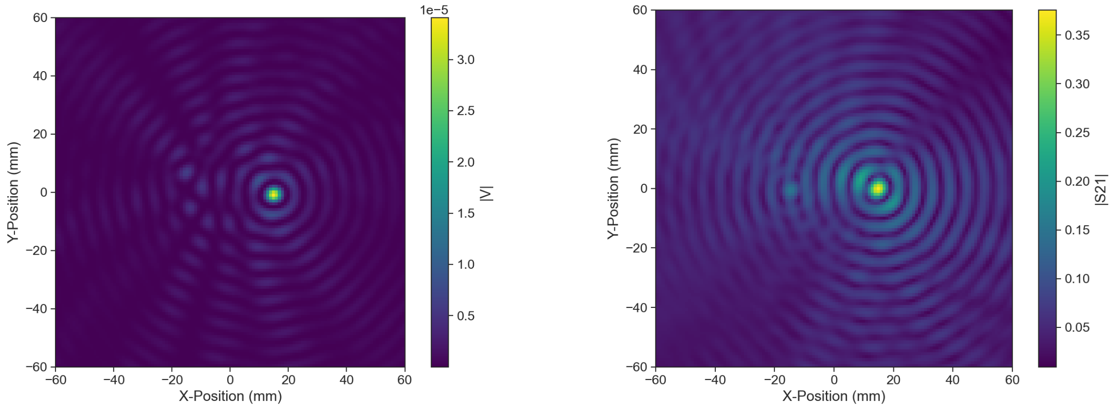


(b) The two-dimensional maps showing the diffractive pattern exhibited by the FSCD and synthetic array signal magnitudes.

Figure 5.35

because of the source. Therefore, we focus on comparing the beamforming of the FSCD array to the synthetic array.

The first method of comparison is to analyze the images generated by applying the beamforming reconstruction specified in Section 4.3.1 to the FSCD and synthetic array data (see Figure 5.36). The beamforming grid consisting of a square 121 $\times$ 121 grid spanning a range of -60-mm to 60 mm in the x and y dimensions. The beamforming images formed from the synthetic array produces a three-dimensional matrix where each grid position contains a summed time series. A single beamforming image is formed from



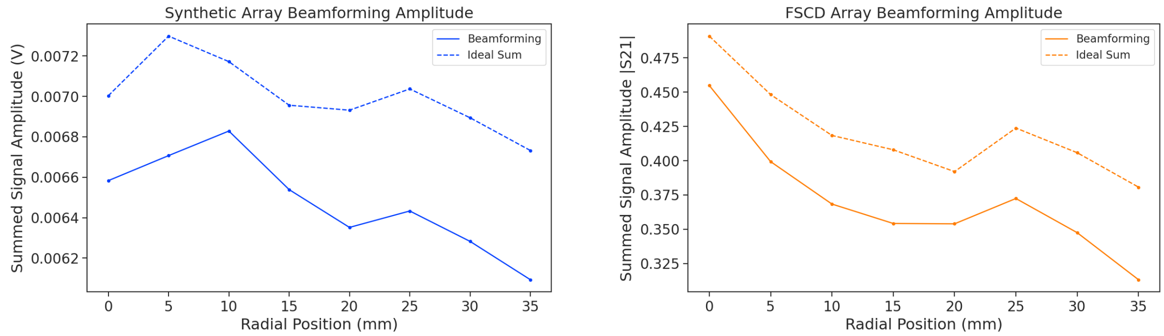
(a) (b)

Figure 5.36: Beamforming images from the synthetic array (a) and FSCD array (b) setups with the SYNCA positioned 15 mm off the central axis. In both images we see a clear maxima that corresponds to the true SYNCA position. However, in the FSCD array there is an additional faint peak located at the opposite position of the beamforming maximum. This additional peak is the mirror of the true peak and is the result of reflections between antennas in the FSCD array.

this data matrix by taking the mean over the time dimension. In the case of the FSCD array, the VNA generates frequency domain data such that each grid position contains a summed frequency series produced by the VNA sweep. For this data a single image is formed by averaging in the frequency domain.

There is a clear difference between the synthetic and FSCD array beamforming images, which is the additional faint beamforming maxima located directly opposite the maxima corresponding to the SYNCA position. The images in Figure 5.36 were generated with data collected at a SYNCA radial position of 15 mm, which agrees well with the observed beamforming maximum in both images. We observe that the faint beamforming peak is located directly opposite of the true beamforming maximum similar to a mirror image. Therefore, the origin of this additional feature appears to be reflections between the two sides of the circular antenna array that are not present for the synthetic array since only a single physical antenna is used.

From the beamforming images we extract the maximum amplitude, which we plot as a function of the radial position of the SYNCA (see Figure 5.37). The phase errors we observed in the FSCD and synthetic arrays leads to power loss at the beamforming stage due to phase mismatches between the signals at different channels. This power loss can be quantified by comparing the signal amplitude obtained from beamforming to the amplitude which would be obtained from an ideal summation. We perform the ideal



(a) (b)

Figure 5.37: A comparison of the maximum signal amplitude obtained by beamforming to the signal amplitude obtained with an ideal summation as a function of the radial position of the SYNCA. The amplitudes for the synthetic array are shown in (a) and the FSCD array are shown in (b). In both setups we observe that the signal amplitudes obtained from beamforming are smaller than the signal amplitude that could be attained with the ideal summation without phase mismatch.

summation by phase shifting each array channel to the same phase and then summing. The comparison between the beamforming and ideal sums is shown in Figure 5.37, where we observe that both the synthetic and FSCD arrays experience power losses from the beamforming summation.

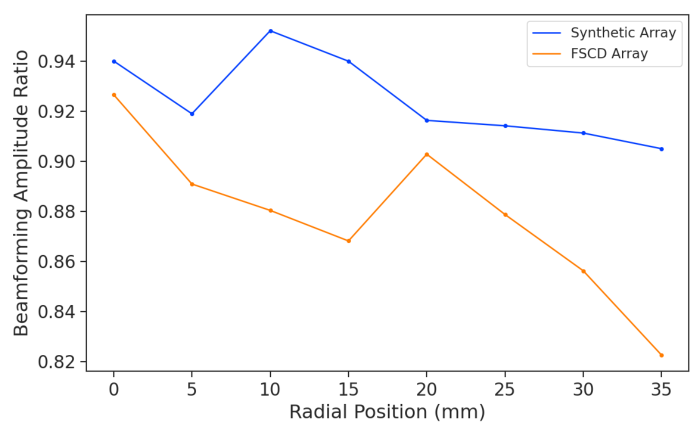


Figure 5.38: The ratio of the beamforming signal amplitude to the ideal signal amplitude for the FSCD and synthetic arrays. We see that the FSCD array has a larger power loss from phase error compare to the synthetic array which indicates that calibration errors associated with the multiple channels as well as reflections are impacting the signal reconstruction.

The beamforming power loss can be quantified using the ratio of the beamforming to ideal signal amplitudes. Computing this ratio as a function of SYNCA radial position radius for the FSCD and synthetic arrays we find that the FSCD array has a uniformly smaller beamforming amplitude ratio, which means that the FSCD array has a larger beamforming power loss (see Figure 5.38). The primary contributions to the beamforming power loss in the synthetic array are phase errors from the SYNCA and phase errors from the FSCD antenna near-field. Both of these phase errors contribute to beamforming losses in the FSCD array, but there are clearly additional phase errors in the FSCD array measurements contributing to the smaller ratio. Two potential error sources include phase differences in the different antenna channels that could not be corrected by calibration as well as reflections between antennas in the array. The total effect of these additional phase errors is to reduce the beamforming amplitude ratio by about 5% from the beamforming ratio of the synthetic array. Therefore, we estimate that if no effort is made to correct these phase errors in an FSCD-like experiment, then we expect approximately a 10% total signal amplitude loss from a beamforming signal reconstruction.

5.5.4 Conclusions

The estimated power loss of a beamforming reconstruction obtained from this analysis provides valuable inputs to sensitivity calculations of a FSCD-like antenna array experiment to measure the neutrino mass, since it helps to bound systematic uncertainties from the antenna array and reconstruction pipeline. This power loss lowers the estimated detection efficiency of the experiment since some of the signal power is lost due to improper combining between channels and also increases the uncertainty in the electron's kinetic energy by contributing to errors in the estimation of the electron's cyclotron frequency.

If these reconstruction losses prove unacceptable there are steps that can be taken to mitigate their effects. Some examples include the development of a more accurate antenna simulation approach that can reproduce the observed near-field interference patterns of the FSCD antennas and the implementation of a calibration approach that allows for the relative phase delays of the array to be measured without changing or disconnecting the antenna array configuration.

Chapter 6

Development of Resonant Cavities for Large Volume CRES Measurements

6.1 Introduction

The cavity approach was originally an alternative CRES measurement technology under consideration by the Project 8 collaboration for the Phase IV experiment. After pursuing an antenna array based CRES demonstrator design for several years, the increasing costs and complexity of the antenna arrays led to a reconsideration of the baseline technology for the ultimate CRES experiment planned by Project 8. Currently, a cavity based CRES experiment is the preferred technology choice for future experiments by the Project 8 collaboration including the Phase IV experiment.

In this chapter I provide a brief summary of resonant cavities and sketch out the key features of a cavity based CRES experiment. In Section 6.2 I provide a brief introduction to cylindrical resonant cavities and the solutions for the electromagnetic fields in the cavity volume.

In Section 6.3 I describe the main components of a cavity based CRES experiment, including the background and trap magnets, cavity geometry and design, and cavity coupling considerations. I also discuss some relevant trade-offs between an antenna array and cavity CRES experiment, and highlight some reasons for the transition of Project 8 to the development of a cavity based experiment.

Finally, in Sections 6.4 and 6.5, I present the design and development of an open mode-filtered cavity that could be used in a cavity based CRES experiment with atomic tritium. The results of the cavity simulations are confirmed by laboratory measurements of a proof-of-principle prototype that demonstrates key features of the design.

6.2 Cylindrical Resonant Cavities

Resonant cavities are sealed conductive containers, which allows us to describe the electromagnetic (EM) fields contained in the cavity volume as a superposition of resonant modes [85]. The field shapes of the resonant modes are determined by Maxwell's equations and the boundary conditions enforced by the cavity geometry. Of interest to Project 8 for CRES measurements are cylindrical cavities due to their ease of construction and integration with atom and electron trapping magnets.

6.2.1 General Field Solutions

Consider a long segment of conducting material with a cylindrical cross-section (see Figure 6.1). A geometry such as this can be used as a waveguide transmission line to transfer EM energy from point to point, or, if conducting shorts are inserted on both ends of the cylinder, the waveguide becomes a resonant cavity.

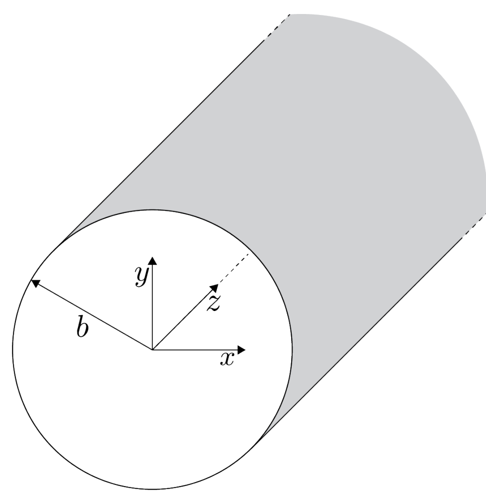


Figure 6.1: Geometry of a cylindrical waveguide with radius b.

The fields allowed inside a cylindrical cavity are determined by the boundary conditions of the cylindrical geometry. The general approach to solving the fields begins by assuming solutions to Maxwell's equations of the form

$$\boldsymbol{E}(x, y, z) = (\boldsymbol{e}(x, y) + \hat{z}e_z(x, y))e^{-i\beta z}, \tag{6.1}$$

$$\boldsymbol{H}(x, y, z) = (\boldsymbol{h}(x, y) + \hat{z}h_z(x, y))e^{-i\beta z}. \tag{6.2}$$

The solutions assume a harmonic time dependence of the form $e^{i\omega t}$ and propagation along the positive z-axis. The functions $\boldsymbol{e}(x, y)$ and $\boldsymbol{h}(x, y)$ represent the transverse $(\hat{x}, \hat{y})$ components of the electric and magnetic fields respectively, and $e_z(x, y)$, $h_z(x, y)$ represent the longitudinal components. The version of Maxwell's equations in the case where there are no source terms can be written as a pair of coupled differential equations,

$$\nabla \times \boldsymbol{E} = -i\omega\mu\boldsymbol{H}, \tag{6.3}$$

$$\nabla \times \boldsymbol{H} = i\omega\epsilon\boldsymbol{E}, \tag{6.4}$$

where ϵ and μ are the permittivity and permeability of the material inside the waveguide or cavity. Using the field solutions from Equations 6.1 and 6.2 one can solve for the transverse components of the fields in terms of the longitudinal fields. Because we are interested in cylindrical cavities it is advantageous to write the field solutions in cylindrical coordinates. After performing this transformation the set of four equations for the transverse field components are,

$$H_\rho = \frac{i}{k_c^2}\left(\frac{\omega\epsilon}{\rho}\frac{\partial E_z}{\partial\phi} - \beta\frac{\partial H_z}{\partial\rho}\right), \tag{6.5}$$

$$H_\phi = \frac{-i}{k_c^2}\left(\omega\epsilon\frac{\partial E_z}{\partial\rho} + \frac{\beta}{\rho}\frac{\partial H_z}{\partial\phi}\right), \tag{6.6}$$

$$E_\rho = \frac{-i}{k_c^2}\left(\beta\frac{\partial E_z}{\partial\rho} + \frac{\omega\mu}{\rho}\frac{\partial H_z}{\partial\phi}\right), \tag{6.7}$$

$$E_\phi = \frac{i}{k_c^2}\left(\frac{-\beta}{\rho}\frac{\partial E_z}{\partial\phi} + \omega\mu\frac{\partial H_z}{\partial\rho}\right), \tag{6.8}$$

where k_c is the cutoff wavenumber defined by $k_c^2 = k^2 - \beta^2$ with $k = \omega\sqrt{\mu\epsilon}$ being the wavenumber of the EM radiation.

This set of equations can be used to solve for a variety of different modes that can be obtained by setting conditions on E_z and H_z. For cylindrical cavities two types of modes are allowed, which correspond to solutions where $E_z = 0$ and $H_z = 0$ respectively.

6.2.2 TE and TM Modes

The TE family of modes corresponds to the case where $E_z = 0$. This implies that H_z is a solution to the Helmholtz wave equation

$$(\nabla^2 + k^2)H_z = 0. \tag{6.9}$$

For solutions of the form $H_z(\rho, \phi, z) = h_z(\rho, \phi)e^{-i\beta z}$, Equation 6.9 can be solved using the standard technique of separation of variables. Rather than reproduce the derivation here we shall simply quote the solutions for the transverse fields [85], which are

$$H_\rho = \frac{-i\beta}{k_{c_{nm}}}(A \sin n\phi + B \cos n\phi) J_n'(k_{c_{nm}}\rho)e^{-i\beta_{nm}z}, \tag{6.10}$$

$$H_\phi = \frac{-i\beta n}{k_{c_{nm}}^2 \rho}(A \cos n\phi - B \sin n\phi) J_n(k_{c_{nm}}\rho)e^{-i\beta_{nm}z}, \tag{6.11}$$

$$E_\rho = \frac{-i\omega\mu n}{k_{c_{nm}}^2 \rho}(A \cos n\phi - B \sin n\phi) J_n(k_{c_{nm}}\rho)e^{-i\beta_{nm}z}, \tag{6.12}$$

$$E_\phi = \frac{i\omega\mu}{k_{c_{nm}}}(A \sin n\phi + B \cos n\phi) J_n'(k_{c_{nm}}\rho)e^{-i\beta_{nm}z}. \tag{6.13}$$

One can observe that the solutions have a periodic dependence on ϕ, and radial profiles given by the Bessel functions of the first kind. The integer indices n and m arise from continuity conditions on the EM fields in the azimuthal and radial directions. For the TE modes $n \geq 0$ and $m \geq 1$. $k_{c_{nm}}$ is the cutoff wavenumber for the TE$_{nm}$ mode given by

$$k_{c_{nm}} = \frac{p'_{nm}}{b}, \tag{6.14}$$

where b is the radius of the cavity or waveguide and p'_{nm} is the m-th root of the derivative of the n-th order Bessel function (see Table 6.1).

Table 6.1: A table of the values of p'_{nm}.

n	p'_{n1}	p'_{n2}	p'_{n3}
0	3.832	7.016	10.174
1	1.841	5.331	8.536
2	3.054	6.706	9.970

The TM mode family corresponds to the case where $H_z = 0$, and $(\nabla^2 + k^2)E_z = 0$. Again, we assume solutions of the form $E_z(\rho, \phi, z) = e_z(\rho, \phi)e^{-i\beta z}$, for which the general form of the solutions is the same as for the TE modes. However, the different boundary conditions for the TM modes results in particular solutions with a different from, which we shall quote here without derivation. The transverse fields of the TM modes are given by

$$H_\rho = \frac{-i\omega\epsilon n}{k_{c_{nm}}^2 \rho}(A \cos n\phi - B \sin n\phi) J_n(k_{c_{nm}}\rho)e^{-i\beta_{nm}z}, \tag{6.15}$$

$$H_\phi = \frac{-i\omega\epsilon}{k_{c_{nm}}}(A\sin n\phi + B\cos n\phi)J_n'(k_{c_{nm}}\rho)e^{-i\beta_{nm}z} \tag{6.16}$$

$$E_\rho = \frac{-i\beta}{k_{c_{nm}}}(A\sin n\phi + B\cos n\phi)J_n'(k_{c_{nm}}\rho)e^{-i\beta_{nm}z}, \tag{6.17}$$

$$E_\phi = \frac{-i\beta n}{k_{c_{nm}}^2\rho}(A\cos n\phi - B\sin n\phi)J_n(k_{c_{nm}}\rho)e^{-i\beta_{nm}z}, \tag{6.18}$$

which one may notice are the same solutions as the TE modes with H and E flipped. The cutoff wavenumber for the TM modes is given by, $k_{c_{nm}} = p_{nm}/b$, where the values of p_{nm} correspond to the m-th zero of the n-th order Bessel function (see Table 6.2).

Table 6.2: A table of the values of p_{nm}.

n	p_{n1}	p_{n2}	p_{n3}
0	2.405	5.520	8.654
1	3.832	7.016	10.174
2	5.135	8.417	11.620

6.2.3 Resonant Frequencies of a Cylindrical Cavity

A cylindrical cavity is constructed by taking a section of cylindrical waveguide and shorting both ends with conductive material. This means that the electric fields inside a cylindrical cavity are exactly those we derived in Section 6.2.2 with the additional condition that the electric fields must go to zero at $z = 0$ and $z = L$ (see Figure 6.2).

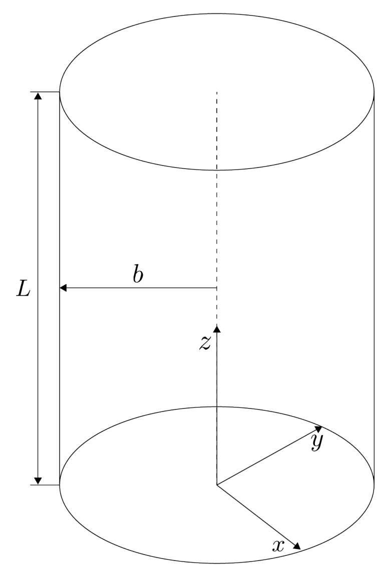


Figure 6.2: The geometry of a cylindrical cavity with length L and radius b.

The transverse electric field solutions for a cylindrical waveguide are of the form

$$\boldsymbol{E}(\rho,\phi,z) = \boldsymbol{e}(\rho,\phi)\left(A_+e^{-i\beta_{nm}z} + A_-e^{i\beta_{nm}z}\right), \tag{6.19}$$

where A_+ and A_- are arbitrary amplitudes of forward and backward propagating waves. In order to enforce that $\boldsymbol{E}$ is zero at both ends of the cavity we require that

$$\beta_{nm}L = 2\pi\ell, \tag{6.20}$$

where $\ell = 0, 1, 2, 3\ldots$. Using this constraint on the propagation constant we can solve for the resonant frequencies of the $\mathrm{TE}_{nm\ell}$ and the $\mathrm{TM}_{nm\ell}$ modes in a cylindrical cavity. For the TE modes the resonant frequencies are

$$f_{nm\ell} = \frac{c}{2\pi\sqrt{\mu_r\epsilon_r}}\sqrt{\left(\frac{p'_{nm}}{b}\right)^2 + \left(\frac{\ell\pi}{L}\right)^2}, \tag{6.21}$$

and the frequencies of the TM modes are

$$f_{nm\ell} = \frac{c}{2\pi\sqrt{\mu_r\epsilon_r}}\sqrt{\left(\frac{p_{nm}}{b}\right)^2 + \left(\frac{\ell\pi}{L}\right)^2}. \tag{6.22}$$

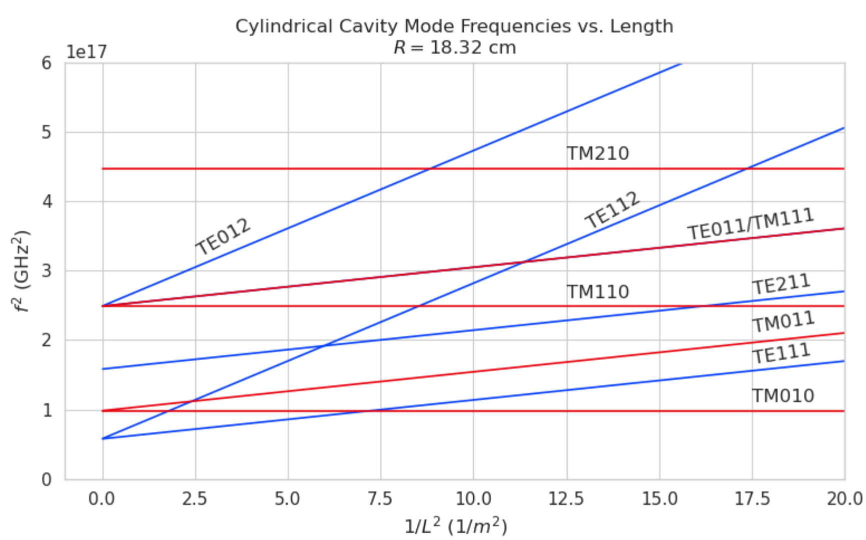


Figure 6.3: Relation of mode frequency to cavity length for a cylindrical cavity with a radius of 18.32 cm.

6.2.4 Cavity Q-factors

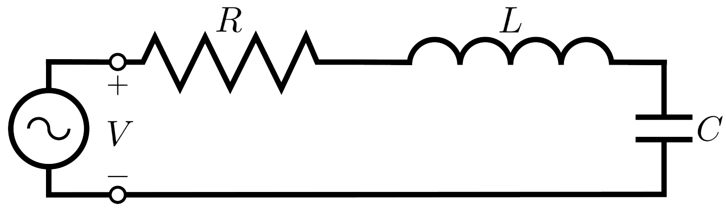


Figure 6.4: A series RLC circuit.

The resonant behavior of cylindrical cavities can be modeled as a series RLC circuit (see figure 6.4). The input impedance of the circuit can be obtained by applying Kirchhoff's laws to calculate the impedance of the equivalent circuit. For a series RLC circuit the input impedance is

$$Z_{\text{in}} = \left(\frac{1}{R} + \frac{1}{i\omega L} + i\omega C\right). \tag{6.23}$$

The resistance in the circuit represents all sources of loss in the cavity, which is primarily caused by the finite conductivity of the cavity walls. The inductor and capacitor represent the energy stored in the cavity in the form of electric and magnetic fields. If the circuit is being driven by an external power source we can write the input power in terms of the circuit input impedance and the source voltage

$$P_{\text{in}} = \frac{1}{2} Z_{\text{in}} |I|^2 = \frac{1}{2} |I|^2 \left(\frac{1}{R} + \frac{1}{i\omega L} + i\omega C\right). \tag{6.24}$$

The resistor introduces a loss into the system with a power given by

$$P_{\text{loss}} = \frac{1}{2} |I|^2 R, \tag{6.25}$$

and the capacitor and inductor store energies given by

$$W_e = \frac{1}{4} \frac{|I|^2}{\omega^2 C}, \tag{6.26}$$

$$W_m = \frac{1}{4} |I|^2 L, \tag{6.27}$$

respectively. Using these expressions we can write the input power and input impedance expressions in terms of the lost power and stored energy

$$P_{\text{in}} = P_{\text{loss}} + 2i\omega(W_m - W_e), \tag{6.28}$$

$$Z_{\text{in}} = \frac{P_{\text{loss}} + 2i\omega(W_m - W_e)}{\frac{1}{2}|I|^2}. \tag{6.29}$$

The condition for resonance in the RLC circuit is that the stored magnetic energy is equal to the stored electric energy (W_e=W_m). When this occurs $Z_{\text{in}} = R$, which is a purely real impedance, and $P_{in} = P_{loss}$. The resonant frequency of the circuit can be determined from the condition $W_e = W_m$ from which one finds that

$$\omega_0 = \frac{1}{\sqrt{LC}}. \tag{6.30}$$

An important performance parameter for any resonant system is the Q-factor, which quantifies the quality of the resonator as the ratio of the stored energy multiplied by the resonant frequency to the average energy lost per second. For the series RLC circuit, the Q-factor is given by the expression

$$Q_0 = \omega\frac{W_e + W_m}{P_{\text{loss}}} = \frac{1}{\omega_0 RC}, \tag{6.31}$$

from which one observes that as the resistance of the RLC circuit is decreased the quality factor of the resonator increases. From the perspective of cylindrical cavities this implies that as one decreases the resistance of the cavity walls it is expected that the Q-factor of the cavity should increase, which is indeed the case. In certain applications where a high Q is desireable it is possible to manufacture a cavity out of superconducting materials in order to minimize the power losses of the system.

The Q-factor of the resonator also determines with bandwidth (BW) of the system. A cavity with a high Q-factor will resonant with a smaller range of frequencies than a cavity with a low Q-factor. To see this we can examine the behavior of the RLC circuit when driven by frequencies near the resonance. For a frequency $\omega = \omega_0 + \Delta\omega$, where $\Delta\omega = \omega - \omega_0 \ll \omega_0$, we can write the input impedance as

$$Z_{\text{in}} = R + i\omega L\left(\frac{\omega^2 - \omega_0^2}{\omega^2}\right), \tag{6.32}$$

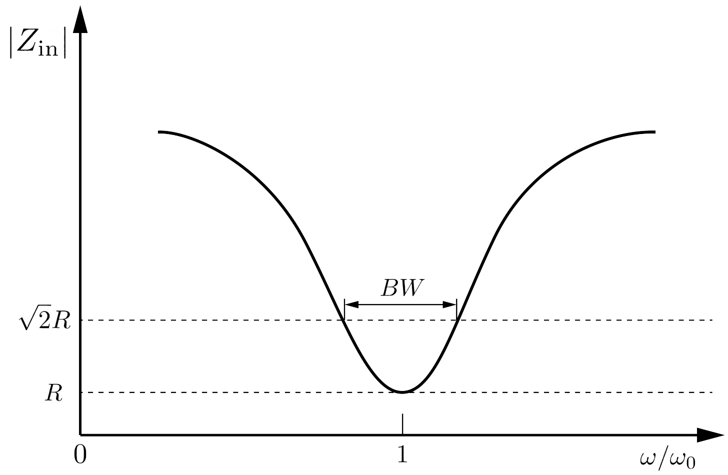


Figure 6.5: Illustration of the behavior of the input impedance of the series RLC circuit as a function of the driving frequency. The BW is proportion to the width of the resonance, which is inversely proportional to Q.

and by expanding $(\omega^2 - \omega_0^2)/\omega^2$ to first order in $\Delta\omega$, we obtain

$$Z_{\text{in}} \approx R + i\frac{2RQ_0\Delta\omega}{\omega_0}. \tag{6.33}$$

Therefore, the magnitude of the input impedance near the resonance is given by

$$|Z_{\text{in}}| = R\sqrt{1 + 4Q_0^2\frac{\Delta\omega^2}{\omega^2}}, \tag{6.34}$$

from which we observe that for the series RLC circuit the input impedance is minimized at the resonant frequency, which corresponds to the maximum input power (see Figure 6.5). The half-power BW is the range of frequencies over which the input power drops to half the input power on resonance. This occurs when $|Z_{\text{in}}| = \sqrt{2}R$, which corresponds to $\Delta\omega/\omega = \text{BW}/2$. Using Equation 6.34 one can find that

$$2R^2 = R^2(1 + Q_0^2\text{BW}^2), \tag{6.35}$$

which implies

$$\text{BW} = \frac{1}{Q_0} \tag{6.36}$$

It is important to emphasize that the Q-factor defined here, Q_0, is technically the unloaded Q. It reflects the quality of the cavity or resonant circuit without the influence of any external circuitry. In practice, however, a cavity is invariably coupled to an

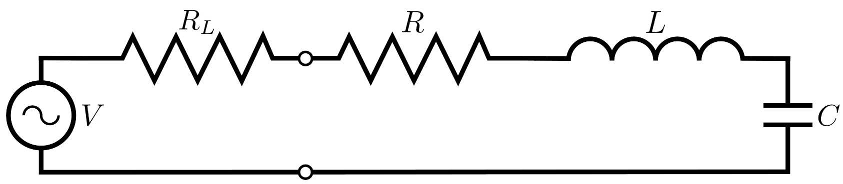


Figure 6.6: A series RLC circuit coupled to an external circuit with input impedance R_L.

external circuit to drive a cavity resonance or to measure the energy of a resonant mode. Coupling a cavity to an external circuit changes the Q by loading the equivalent cavity RLC circuit (see Figure 6.6). The Q-factor of the cavity when it is loaded by an external circuit is called the loaded Q, which is the quantity that one actually measures when exciting a resonance in the cavity. Using the series RLC circuit model one can see that the load resistor in Figure 6.6 will add in series with the resistor in the circuit for a total equivalent resistance of $R + R_L$. Therefore, the loaded Q is given by

$$Q_L = \frac{1}{\omega_0 (R + R_L) C}, \tag{6.37}$$

from which one observes that the loaded Q is always less than the intrinsic Q of the cavity.

The amount of coupling that is desireable depends on the specific application of the resonator. If one wants a resonator that is particular frequency selective than it makes sense to limit the amount of coupling to the cavity to maintain a small BW, alternatively, if a larger BW is need one can increase the cavity coupling by tuning the input impedance of the external circuit. The critical point, where maximum power is transferred between the cavity and the external circuit, occurs when the input impedance of the cavity matches the input impedance of the external transmission line. For the series RLC circuit on resonance, this matching condition corresponds to

$$Z_0 = Z_{\text{in}} = R, \tag{6.38}$$

where Z_0 is the impedance of the transmission line. The loaded Q at this critical point is, therefore,

$$Q_L = \frac{1}{2\omega_0 Z_0 C} = \frac{Q_0}{2}. \tag{6.39}$$

One can described the degree of coupling between the cavity and an external circuit by defining a coupling factor, g, such that,

$$g = \frac{Q_0}{Q_L} - 1. \tag{6.40}$$

When $g = 1$ then $Q_L = Q_0/2$, and the cavity is said to be critically coupled as we described. If $Q_L < Q_0/2$, then the cavity is undercoupled to the transmission line, corresponding to $g < 1$. Alternatively, if $Q_L > Q_0/2$, then $g > 1$, and the cavity is overcoupled to the transmission line. Various specialized circuits can be used to tune the input impedance of the external circuit as seen by the cavity to achieve a wide range of different coupling factors based on the desired application of the cavity.

6.3 The Cavity Approach to CRES

6.3.1 A Sketch of a Molecular Tritium Cavity CRES Experiment

Resonant cavities can be used to perform CRES measurements, and they represent the current preferred technology by the Project 8 collaboration. The basic approach to a neutrino mass measurement using a resonant cavity and molecular tritium beta-decay source is illustrated by Figure 6.7.

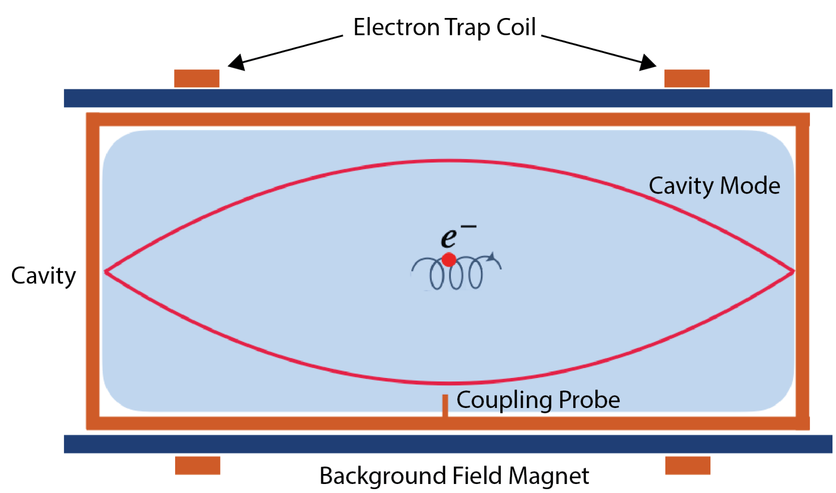


Figure 6.7: A cartoon depiction of a cavity CRES experiment. A metallic cavity filled with tritium gas is inserted into a uniform background magnetic field to perform CRES measurements. Electrons from beta-decays inside the cavity can be trapped and used to excite a resonant mode(s). By coupling to the cavity mode with a suitable probe one can measure the cyclotron frequency of the electron and perform CRES.

At the core of the experiment is a large resonant cavity filled with tritium gas. The filled cavity is then placed in a uniform magnetic field provided by a primary magnet that provides the background magnetic field. The value of the background magnetic field sets the range of cyclotron frequencies for electrons emitted near the tritium spectrum endpoint. When a beta-decay electron is produced in the cavity it is trapped using a set of magnetic pinch coils that keep electrons inside the cavity volume.

Electrons trapped inside the cavity do not radiate in the same way as electrons in free-space. Effectively, the same boundary conditions that were used to derive the resonant modes of a cylindrical cavity in Section 6.2 apply to the radiation of the electron as well. The coupling of an electron performing cyclotron motion in a cavity has been studied in detail for measurements of the electron's magnetic moment [94–96] If an electron is emitted with a kinetic energy that corresponds to a cyclotron frequency that matches a resonant frequency of the cavity, then energy radiated by the electron excites a corresponding resonance in the cavity. The strength of the electron's coupling to the cavity is given to first order by the dot product between the electrons trajectory and the electric field vector of the resonant mode. Additional effects, such as the Purcell enhancement [97], alter the emitted power from the free-space Larmor equation [48]. If an electron is moving with a cyclotron frequency that is far from any resonant modes in the cavity, then radiation from the electron is suppressed. One can interpret this somewhat surprising effect as the metallic walls of the cavity reflecting the radiated energy back to the electron.

Detecting an electron in the cavity is accomplished by coupling the cavity to an external transmission line that leads to an amplifier and RF receiver chain [98]. The coupling of the cavity resonance to the amplifier occurs through a coupling probe or aperture designed to read-out the excitation of the mode(s) excited by the electron. For CRES measurements, the placement of a wire antenna coupling probe inside the cavity volume leads to unacceptable losses of tritium atoms due to recombination to molecular tritium on the antenna surface, therefore, apertures are the preferred coupling method for cavity CRES experiments.

One of the attractive features of the CRES technique for neutrino mass measurement is the gain in statistics that comes from the differential nature of the tritium spectrum measurement. Initially, this seems incompatible with cavities, due to the narrow resonances of cavity modes giving relatively small bandwidth. However, by intentionally over-coupling to a single cavity mode one can achieve bandwidths of a few 10's of MHz (see Section 6.2), which is sufficient for a measurement of the tritium spectrum endpoint region.

6.3.2 Magnetic Field, Cavity Geometry, and Resonant Modes

Magnetic Field and Volume Scaling

For a CRES experiment, cylindrical cavities are a natural choice since they match the geometry of standard solenoid magnets, which are needed in order to produce the background magnetic field for CRES measurements. Furthermore, the cylindrical shape is compatible with a Halbach array, which is the leading choice of atom trapping technology for future atomic tritium experiments by the Project 8 collaboration. Cylindrical cavities also benefit from well-established machining practices that are able to achieve high geometric precision at large lengths scales. More exotic cavity designs are under-consideration and there are on-going efforts to investigate the potential advantages these may have over the standard cylindrical geometry.

As we saw in Section 6.2, the physical dimensions of the cavity are directly coupled to the resonant frequencies of the cavity. This dependency links the size of the cavity to the magnitude of the background magnetic field, because the magnetic field determines the cyclotron frequencies of trapped electrons. Specifically, as the size of the cavity is increased to accommodate larger volumes of tritium gas, the frequencies of the resonant modes decrease proportionally. This requires that the magnetic field also decrease in order to maintain coupling between electrons and the desired cavity mode.

The required cavity size is ultimately determined by the required statistics in the tritium spectrum endpoint region. Because the gas density must be kept below a certain level to ensure that electrons have sufficient time to radiate before scattering, larger volumes become the only way to achieve higher event statistics. To achieve the sensitivity goals of Phase III and IV cavity volumes on the order of several cubic-meters are required, which pushes one towards frequencies in the range of 100's of MHz.

Single-mode Cavity CRES

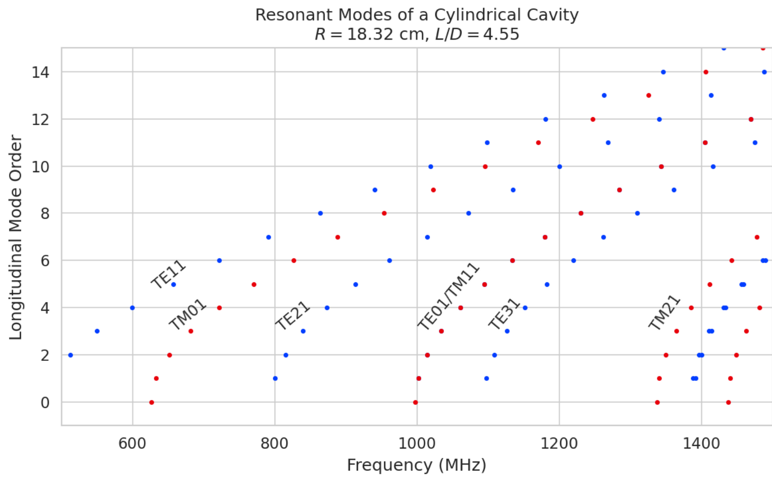
It is tempting to consider maintaining a high magnetic field, while still increasing the size of the cavity, in order to increase the radiated power from trapped electrons for better SNR. However, if one were to maintain the same magnetic field while increasing the size of the cavity, the electrons would begin to couple to higher order modes with more complicated transverse geometries. The danger with this approach is that a complicated mode structure could introduce systematic errors into the CRES signals. Example systematics include unpredicted mode hybridization or changes in the mode shapes from imperfections in the cavity construction, which would prevent reconstruction of the electron's starting kinetic energies with adequate resolution. For this reason, it is ideal to operate with magnetic fields that give cyclotron frequencies near the fundamental frequency of the cavity, where the mode structure is relatively simple (see Figure 6.8). In this frequency region it is possible to perform CRES by coupling to only a single resonant mode, however, it is currently an open question if a single mode measurement will provide enough information about an individual electron's position to reconstruct the full event. Regardless, developing a solid understanding of the CRES phenomenology when an electron is coupling to a single mode will be a necessary step towards a future multi-mode cavity experiment.

Considerations for Resonant Mode Selection

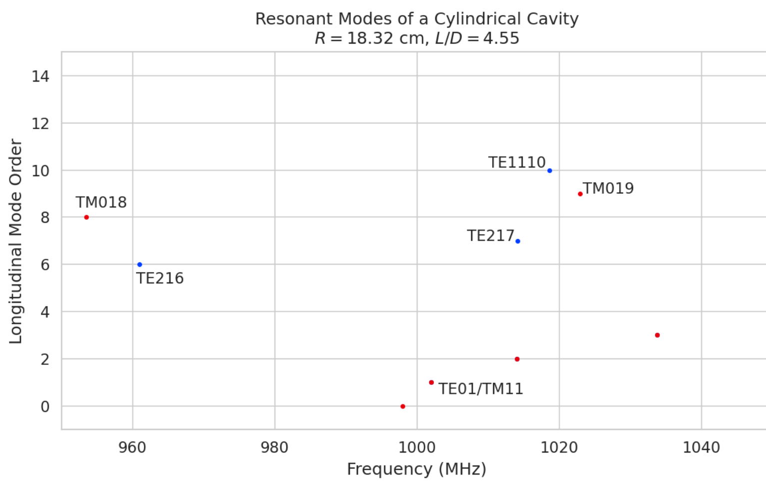
A single-mode cavity experiment begs the question, which resonant mode is best for CRES measurements? There is an immediate bias towards low order TE_{nm} and TM_{nm} modes due to the multi-mode considerations discussed above. Additionally, there is a preference towards modes with longitudinal index $\ell = 1$ with a single antinode along the vertical axis of the cylindrical cavity. The reason for this is that there is a phase change in the electric fields between antinodes that leads to modulation effects that destroy the carrier frequency signal information.

A second consideration for mode selection is the volumetric efficiency of the mode. Volumetric efficiency can be thought of as an integral over the volume of the cavity weighted by the relative amplitude of the mode. From the perspective of simply maximizing the volume useable for CRES measurements this integral would be as close to unity as possible. However, there is a requirement to reconstruct the position of the electrons inside the cavity volume so that the local magnetic fields can be used to convert the measured cyclotron frequency to a kinetic energy. With a single mode this necessarily requires a variable transverse mode amplitude, which lowers the volumetric efficiency, so that position of the electron in the cavity can be estimated from the average amplitude of the CRES signal. Longitudinal indices of $\ell = 1$ have an advantage in volumetric efficiency over higher order ℓ modes, since there are only two longitudinal nodes, one at each end of the cavity. Therefore, the average coupling strength of trapped electrons as they oscillate axially is higher for $\ell = 1$ modes.

The longitudinal variation in the mode strength is ultimately critical for achieving the energy resolution required for neutrino mass measurements. Correcting for the change in



(a)



(b)

Figure 6.8: Examples of the resonant mode frequencies of a cylindrical cavity. This cavity has a radius of 18.32 cm and a length to diameter ratio of 4.55.

the average magnetic fields experienced by electrons with different pitch angles requires that information on the axial motion of the electron be encoded into the CRES signal. The longitudinal variation in the mode amplitude leads to amplitude modulation of the CRES signal with a frequency proportional to the electron's pitch angle.

An additional factor for mode selection is the intrinsic or unloaded Q of the mode. In terms of SNR it is advantageous to use a mode with a very high Q_0, which is then highly overcoupled to achieve the necessary bandwidth to cover the tritium endpoint spectrum. This scheme leads to a decoupling of the physical cavity temperature from the effective noise temperature after the amplifier, which allows us to achieve adequate SNR without the requirement of cooling the entire cavity to single Kelvin temperatures.

An example of a resonant mode that exhibits these traits is the TE_{011} mode. At present the TE_{011} mode is the preferred resonance for a single-mode cavity CRES experiment by the Project 8 collaboration. TE_{011} is a low order mode located in a region relatively far from other cavity modes. Furthermore, the separation of the TE_{011} mode can be improved by various mode-filtering techniques discussed in Section 6.4.2 below. TE_{011} consists of a single longitudinal antinode that can provide pitch angle information in the form of amplitude modulation, and has an electric field with a radial profile given by the J_0' Bessel function allowing for radial position estimation. Lastly, the TE_{011} mode has a relatively high intrinsic Q compared to nearby modes, which helps with SNR. Unloaded Q's greater than 80000 are achievable for a 1 GHz TE_{011} resonance using a copper walled cavity.

6.3.3 Trade-offs Between the Antenna and Cavity Approaches

The choice between cavities and antennas for large-scale CRES measurements is not without trade-offs. Both the antenna array and cavity approaches are relatively immature techniques, at present there are no known obstacles that would prevent either approach from being used for a large scale neutrino mass experiment. The preference for cavities is largely driven by important practical considerations that could make a cavity based experiment significantly cheaper than an antenna experiment of similar size and scope. However, the switch to cavities also introduces new challenges less relevant to the antenna array, which must be solved in order for Project 8 to achieve its neutrino mass measurement goals.

One of the major relative drawbacks of the antenna array approach is the size and complexity of the data-acquisition system. A large-scale antenna array experiment requires $O(100)$ antennas independently digitized at rates of $O(10)$ to $O(100)$ MHz. Since there is insufficient information in a single antenna channel to detect or reconstruct the CRES signal, the entire array output must be processed during the signal reconstruction. Because data storage becomes an issue with these data volumes, there is a real-time signal reconstruction requirement that allows one to detect CRES signals buried in the thermal noise. As we discuss in Section 4.4, the computational cost of these real-time detection algorithms are potentially quite large for even a small scale antenna array experiment. However, the operating principle of a cavity experiment allows the CRES signal to be detected using only a single read-out channel digitized at rates of O(10) MHz, which reduces the cost of the data acquisition system by many orders of magnitude.

From an engineering perspective, the simple geometry and thin-walls of a cylindrical cavity are simpler to interface with the cryogenic and magnetic subsystems needed for a CRES experiment. Whereas, the antenna array requires careful design and engineering to accommodate the antenna array and receiver electronics in proximity to the trapping magnets. Additionally, due to near-field interference effects, the antenna array is unable to reconstruct CRES events within the reactive near-field distance of the antennas. Because atom trapping requirements require magnetic fields which correspond to cyclotron frequencies for endpoint electrons less than 1 GHz, the required stand-off distance leads to a significant loss in useable experiment volume, necessitating larger and more expensive magnets.

Another advantage to the cavity approach is the relatively compact sideband structure, which is a result of the low modulation index for cavity CRES signals. The axial motion in an antenna array experiment leads to frequency modulation and sidebands. The shape of the sideband structure is determined by the modulation index, $h = \frac{\Delta f}{f_a}$, where Δf is the size of the frequency deviation and f_a is the axial frequency. The large electron traps required for a cubic-meter-scale experiment leads to high modulation indices, which causes the signal spectrum to be made up of numerous low power sidebands that make reconstruction and detection challenging. This behavior was observed in simulations of the FSCD in which carrier power decreased with pitch angle due to the increase in modulation index (see Figure 4.31). For cavities, however, the modulation index remains near $h = 1$ even for very long magnetic traps due to the high phase velocity in cavities relative to the axial velocity of the electron. This results in an almost ideal spectrum shape that has a strong carrier frequency with a few sidebands whose relative amplitudes encode pitch angle information.

A downside of the cavity approach is the apparent difficulty of estimating the position of the electron using only the coupling of the electron to a single mode. The amplitude of the TE_{011} mode is completely independent of the azimuthal coordinate, therefore, position reconstruction using the TE_{011} mode is only able to estimate the radial position of the electron. This position degeneracy may lead to magnetic field uniformity requirements that are too challenging to meet due to mechanical uncertainties in cavity and magnet construction, as well as uncertainties caused by nuisance external magnetic fields such as the Earth's field and magnetic fields from building materials. A multi-mode cavity experiment may provide a way to extract more precise information on the position of the electron by analyzing the coupling of the electron to several modes that overlap in different ways.

6.4 Single-mode Resonant Cavity Design and Simulations

The single-mode cylindrical cavities envisioned for the Phase III and IV experiments must be carefully engineered in order to measure the neutrino mass with the desired sensitivity. In this section I summarize some simulation studies performed to analyze early design concepts for a single-mode cavity. The primary tool for these investigations was Ansys HFSS, which was also used for the development of the SYNCA antenna described in Section 5.3.

6.4.1 Open Cylindrical Cavities with Coaxial Terminations

Design Concept

A basic cavity design question relevant to Project 8's ultimate goal of an atomic tritium CRES experiment is how to build a cavity that can be efficiently filled with atomic tritium. To keep the rate of atom loss from recombination on surfaces it is ideal if the ends of the cylindrical cavity are as open as possible so that tritium atoms can flow inside unimpeded. Additionally, one of the primary calibration techniques planned for future CRES experiments involves CRES measurements using electrons injected from an electron gun source, which also requires an opening at the cavity end. Cylindrical cavities with open ends can be manufactured, however, the intrinsic Q-factors of these cavities are orders of magnitude less than their sealed counterparts, which reduces the signal-to-noise ratio when that cavity is used for CRES measurement.

Cylindrical cavities with mostly open ends that also exhibit Q values for the $\mathrm{TE}_{01\ell}$ modes similar to sealed cavities can be built by using coaxial endcaps to terminate the cavity. Cavities of this type have been manufactured for specialized applications related to the measurements of the dielectric constants of liquefied gasses (see Figure 6.9) [99,100]. This cavity design leaves the ends of the cavity wide open, but retains high Q-values for the $\mathrm{TE}_{01\ell}$ modes due to the coaxial endcap, which are designed to perfectly reflect the electric fields of $\mathrm{TE}_{01\ell}$ modes. Coupling to the $\mathrm{TE}_{01\ell}$ mode is achieved via an aperture located at the center of the cavity wall.

A cavity similar to Figure 6.9 is a candidate design for the future CRES experiments by Project 8, since it appears to elegantly solve many practical issues that arise when combining cavity CRES and atomic tritium. The coaxial endcaps leave significant regions of the cavity ends completely open, which allows for the entrance of atomic tritium as well as the pumping away of molecular tritium that has recombined on the cavity walls.

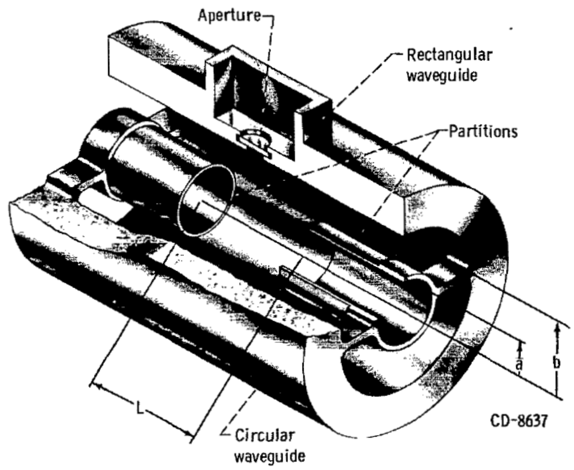


Figure 6.9: An image of an open cavity with coaxial terminations used for dielectric constant measurements. Figure from [100].

These open ends are achieved while preserving the high Q-values of the $TE_{01\ell}$ modes, which is important for extracting as much signal power from the electron as possible. In subsequent sections we shall analyze this cavity design in more detail, primarily by using HFSS simulations to analyze the resonant mode structure of this cavity geometry.

Coaxial Terminator Constraints

The reason that coaxial endcaps can be used to achieve high Q-values for the $TE_{01\ell}$ modes is that the electric fields for these modes are purely azimuthally polarized (see Equations 6.12 and 6.13). Therefore, the boundary conditions that require the electric field to go to zero at the cavity ends can be supplied using a coaxial partition of the correct radius (see Figure 6.10). Because the cylindrical shape enforced by the partition does not match the boundary conditions of other cavity modes, these terminations also significantly suppress the Q-factors of non-$TE_{01\ell}$ modes, which is potentially beneficial for a single-mode cavity CRES experiment.

The correct radius of the cylindrical partition is derived by setting up the boundary value problem in Figure 6.10, and analyzing the reflection and transmission coefficients for waves incident on the coaxial terminators. The basic problem is to identify the radius a where the reflection coefficient for the $TE_{01\ell}$ modes becomes equal to 1. One can show that if the coaxial partitions are made sufficiently long relative to the wavelength of the TE_{01} modes than perfect reflection can be achieved. This derivation is quite lengthy and complex and is presented in full in [99]. Here, we shall simply explain the resulting

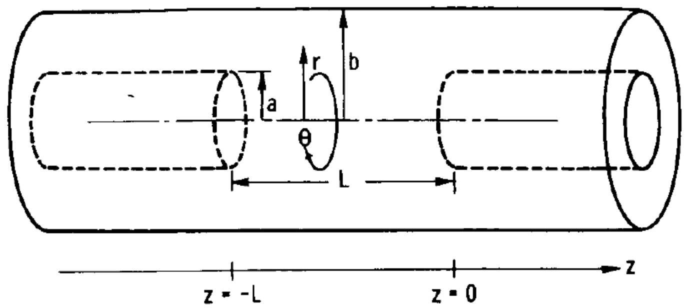


Figure 6.10: The simplified geometry of an open cavity with coaxial terminations. Figure from [99].

conditions on the partition radius for perfect reflection.

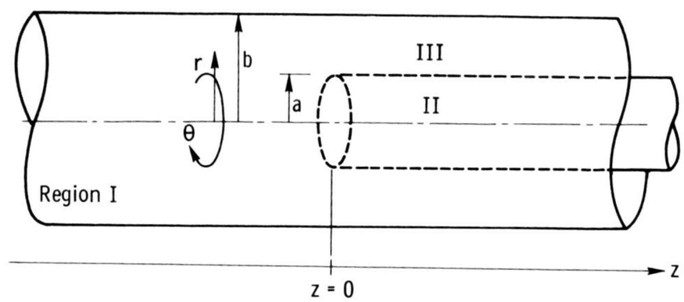


Figure 6.11: Electric field regions for the open cavity boundary value problem. Figure from [99].

The open cavity boundary value problem is solved by expressing the forms of the electric fields in the different regions of the cavity and requiring that the electric fields are continuous. There are effectively three distinct regions in the open cavity corresponding to the central cavity volume, the inner coaxial volume, and the outer coaxial volume (see Figure 6.11).

In Region I, the boundary conditions are those of a cylindrical waveguide, and we require that E_ϕ for the TE_{0m} modes go to zero at the cavity wall ($r = b$). This requires that $J'_{0m}(k_{c_{0m}} b) = 0$. We aim to solve for the radius a in the specific situation where the TE_{01} mode can propagate but all other TE_{0m} modes are below the cutoff frequency for the circular waveguide. This is equivalent to requiring

$$3.832 < k_{c_{0m}} b < 7.016, \tag{6.41}$$

where the numbers 3.832 and 7.016 correspond to the first and second zeros of the Bessel function (see Table 6.1).

In Region II the boundary conditions are those of a cylindrical waveguide, but with a smaller radius. The condition that $E_\phi = 0$ at the cylindrical partition radius is that $J'_{0m}(k_{c_{0m}}a) = 0$. To ensure perfect reflection, we want all modes in Region 1 of the cavity to be below the cutoff frequency of the circular waveguide formed by the inner volume of the coaxial terminator. Therefore, we consider the solutions where

$$k_{c_{0m}}a < 3.832. \tag{6.42}$$

Finally, in Region III the boundary condition are those of a coaxial waveguide. We need to guarantee that $E_\phi = 0$ at both $r = b$ and $r = a$, which involves finding the eigenvalues of the following equation

$$J'_0(k_{c_{0m}}a)Y'_0(k_{c_{0m}}b) - J'_0(k_{c_{0m}}b)Y'_0(k_{c_{0m}}a) = 0, \tag{6.43}$$

where Y'_0 the zeroth-order derivatives of the Bessel function of the second kind. The solutions to this equation depend on the value of the ratio b/a. The approximate solution is given by

$$\delta_n a \simeq \frac{n\pi}{b/a - 1}, \tag{6.44}$$

where δ_n are eigenvalues of Equation 6.43. Similar to Region II, we are interested in solutions for which the TE_{01} modes of Region I are below the cutoff of Region III. Therefore, we require that

$$k_{c_{0m}} < \delta_1. \tag{6.45}$$

In general, one has some freedom in specifying the value of b/a. A value typically used in practice is $b/a = 2.082$, which corresponds to positioning the radius of the cylindrical partition at the maxima of the TE_{01} electrical fields.

Using the constraints from the three field regions one can develop a coaxial terminator that acts as a virtual perfectly conducting surface for the TE_{01} modes. The only required inputs are the desired frequency of the TE_{011} mode and a choice for the value of b/a.

6.4.2 Mode Filtering

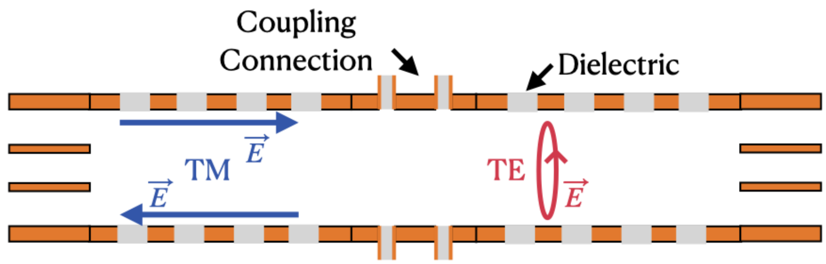
The general case of an electron coupling to a resonant cavity is complicated. This is because cavities contain an infinite number of resonant modes, which for higher order modes, have couplings to the electron with a complex spatial dependence. The danger is that improper modeling of the electron's coupling to the cavity can lead to systematic errors in the CRES measurements that prevent a high-resolution measurement of the electron's kinetic energy. This in part drives the preference for a single-mode cavity experiment that uses only the electron's coupling to the TE_{011} mode to perform CRES, assuming that sufficient information on the electron's position can be obtained with a single mode.

The TE_{011} mode is in a region where there are relatively few other modes to which the electron could couple(see Figure 6.8). However, one can see that the frequency of the TE_{011} is perfectly degenerate with the TM_{111} mode, which means that electrons will inevitably couple to both modes if they have the correct cyclotron frequency.

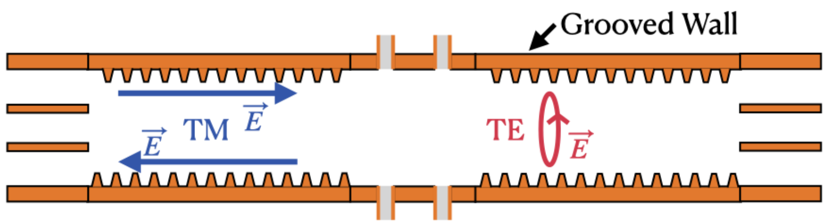
The magnitude of the impact of the electron coupling to both TE_{011} and TM_{111} is currently unknown. To first order an electron coupling to more both modes will lose more energy overtime, which can be measured by observing the frequency chirp rate of the signal. This effect may be small enough to be negligible or simple enough to model that the cavity can be treated as an effective single-mode cavity. Alternatively, the one could consider devising a coupling scheme that is sensitive to both the TE_{011} and the TM_{111} modes. By measuring the coupling of the electron to both modes more information on the position of the electron could be obtained, which could improve the position and energy resolution of the CRES measurements.

A different approach is the mode filtering approach, which seeks to obtain a single TE_{011} mode cavity using perturbations to the cavity walls that selectively impede the TM modes, while leaving the TE modes mostly unperturbed. The type of perturbations required can be determined by visualizing the surface currents induced in the cavity walls by each type of mode (see Figure 6.12). By definition, all TM have electric fields directed along the vertical axis of the cylindrical cavity, which means that perturbations that impede currents in this direction will modify TM resonances. On the other hand, the TE_{01} modes induce azimuthal currents in the cavity walls, therefore, it is possible to break the degeneracy between TE_{01} and TM_{11} using a cavity perturbation that imepedes axial currents, but does not affect the flow of azimuthal currents.

Figure 6.12 shows two cavity design concepts that achieve this selective current perturbation. The resistive approach inserts a series of thin dielectric rings into the walls of the cavity that introduces a resistive and capacitive impedance to the longitudinal currents, while leaving azimuthal current paths intact. Cavities of this type with high TE_{01} Q's have also been constructed by tightly wrapping a thin, dielectric coated wire around a mold to form the cavity wall. An alternative method is to introduce an inductive



(a)



(b)

Figure 6.12: Two mode filtering concepts to break the degeneracy of TE_{01} and TM_{11} modes. The resistive approach uses dielectric materials to impede currents that travel vertically along the cavity while leaving azimuthal currents unperturbed. An alternative approach is to impede the currents using grooves cut into the cavity wall, which achieve the same effect with an inductive impedance.

impedance by cutting grooves or a thread pattern on the inside wall of the cavity. For reasons of manufacturability and compatibility with tritium the grooved cavity approach is the preferred method for mode-filtered cavity construction by Project 8.

6.4.3 Simulations of Open, Mode-filtered Cavities

A candidate design for a single TE_{011} mode CRES experiment is a cavity that utilizes the coaxial terminations combined with a mode-filtering wall. The first step towards validating that a cavity that combines these two design features will operate as expected is a thorough simulation effort for which finite element method (FEM) simulation software is invaluable. The primary tool for electromagnetic FEM calculations inside Project 8 is Ansys HFSS, which has a robust and well-established eigenmode solver that can identify the resonant frequencies and associated Q-factors for given structure.

Four variations of a cavity design with a ~ 1 GHz TE_{011} resonance were implemented in HFSS (see Figure 6.13). The four designs include a standard cylindrical cavity, an open cavity with smooth walls, an open cavity with resistive walls, and an open cavity with grooved walls. The relevant design parameters are summarized in Table 6.3. All

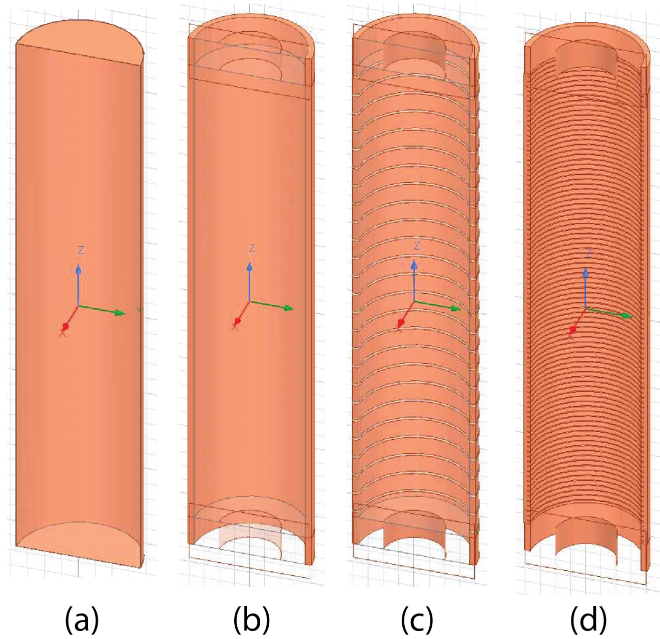


Figure 6.13: Four cavity design variations. (a) is a standard sealed cylindrical cavity, (b) is an open cavity with smooth walls, (c) is an open cavity with resistive walls, and (d) is an open cavity with grooved walls. The main cavity and coaxial terminator parameter are identical for all four cavities.

cavities were simulated using copper walls and filled with a vacuum dielectric. The identities of the resonant modes found by HFSS were validated by visual inspection of the electric and magnetic field patterns and by comparison to analytical calculations of the mode frequencies.

Table 6.3: A table of cavity design parameters used for HFSS simulations.

Name	Qty.	Unit	Description
D_{cav}	326.4	mm	Cavity diameter
L_{cav}	1668.0	mm	Cavity length
D_{term}	200.2	mm	Inner diameter of coaxial terminator
L_{term}	100.0	mm	Terminator length
l_{die}	8.3	mm	Dielectric spacer thickness
Δl_{die}	66.7	mm	Distance between dielectric spacers
l_{groove}	3.0	mm	Groove height
d_{groove}	9.0	mm	Groove depth
Δl_{groove}	18.3	mm	Distance between grooves

The results of the HFSS simulations validate our predictions of the resonant behavior of an open, mode-filtered cavity developed in the preceding sections (see Figure 6.14) One

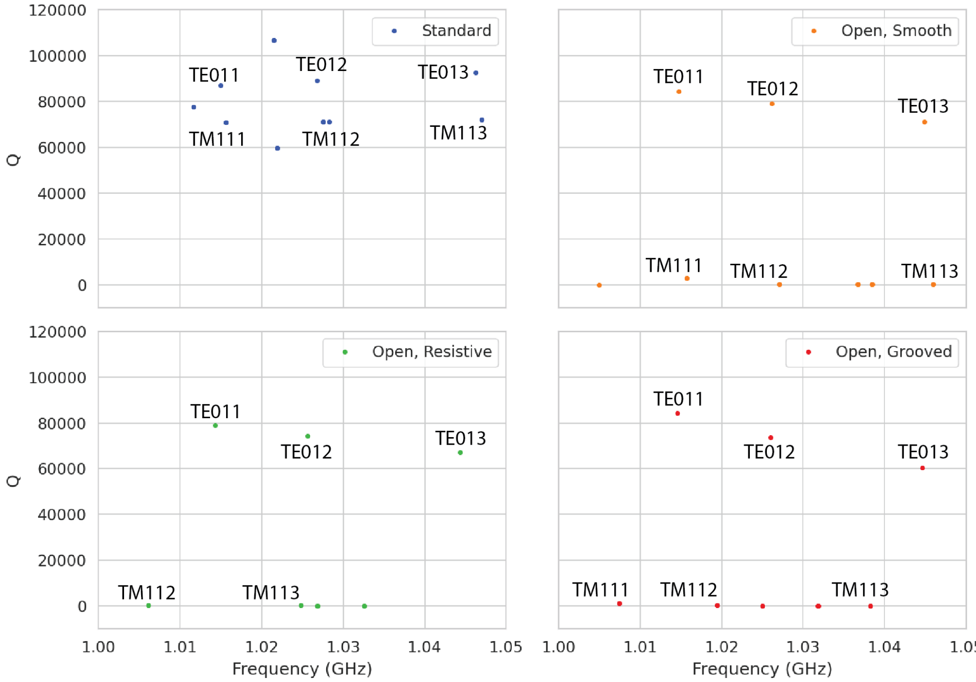


Figure 6.14: The frequencies and Q-factors of the resonant modes identified by HFSS for the cavity variations shown in Figure 6.13. The fully-sealed cavity with smooth walls has several high-Q modes near the TE_{011} resonance. Introducing the open-termination preserves the Q-factors of the $TE_{01\ell}$ modes and suppresses the Q-factors of the modes whose boundary conditions to not match the cylindrical partition. Both the resistive and grooved wall perturbations shift the resonant frequencies of the TM modes away from the TE_{011} mode. By properly tuning the geometry of the grooves or the resistive spacers several MHz of frequency separation can be achieved.

can see that for a standard cavity the TE_{01} and the TM_{11} are degenerate in frequency with relatively high Q-factors. The open-ended cavity preserves the high Q-factors of the TE_{01} modes, while the other modes, since their boundary conditions do not match the coaxial geometry, have their Q-factors suppressed. One can see that the effect of the resistive and inductive mode-filtering schemes is to effectively shift the resonant frequencies of the TM_{11} modes below those of the associated TE_{01} modes, which breaks the degeneracy. Optimization of the dielectric spacer or groove parameters can ensure that the TE_{011} mode is isolated from other modes by $O(10)$ MHz, which provides sufficient bandwidth for a measurement of the tritium spectrum endpoint.

Further optimization of the cavity design requires a more detailed cavity simulation that includes the cavity coupling mechanism as well as other geometry modifications required for integration into the magnetic and tritium gas subsystems. Perhaps more important is the development of the capability to simulate the interaction of electrons with the cavity so that simulated CRES signals can be generated using cavities designed for CRES measurements. Simulated CRES signals can then be used to estimate the neutrino mass sensitivity of the experiment, which allows for the optimization of the cavity design towards the configuration that provides the best measurement of the neutrino mass.

6.5 Single-mode Resonant Cavity Measurements

Measurement test stands play an important role in the research and development process that cannot be replaced by simulations. For example, constructing a prototype CRES cavity forces one to consider important practical issues such as manufacturability and machine tolerances that may require modifications to the design. Furthermore, by comparing laboratory measurements of a real cavity to simulations, one can quantify the impact of imperfections and real-life measurement systematics, which allows for more accurate sensitivity estimates of the experiment. Lastly, the development of these prototypes helps to build the necessary experience and expertise within the collaboration required for more complicated experiments to succeed.

In this spirit a prototype cavity was constructed to demonstrate the open, mode-filtered cavity concept explored in the previous sections. The primary goal of the measurements was to validate that an open, mode-filtered cavity suppressed the TM_{11} modes as predicted by HFSS simulations.

6.5.1 Cavities and Setup

Two rudimentary, cavities were constructed using segments of copper pipe available from McMaster-Carr (see Figure 6.15). The design consists of copper pipes of two diameters. The larger diameter pipe forms the main cavity wall and the smaller diameter pipe is used to create a coaxial termination. The diameter of the outer pipe was chosen to produce a TE_{011} resonance of approximately 6 GHz, while the diameter of the smaller pipe was selected based on the open termination criteria introduced in Section 6.4.1. The approximate diameters and lengths of the copper pipe are summarized in Table 6.4.

Coupling to the cavity was achieved using a hand-formable segment of coaxial cable stripped at one end to form a loop antenna. This was inserted into a small hole located at the center of the main cavity wall. The coaxial terminators were supported inside the

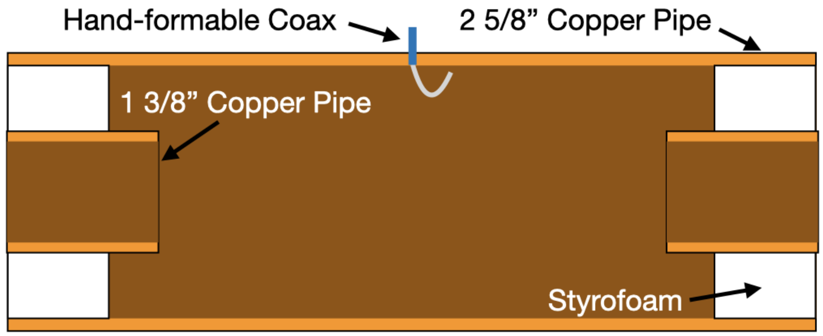


Figure 6.15: A cartoon depicting the design of the open-ended cavity prototype designed to operate at approximately 6 GHz. The main cavity wall was composed of a single copper pipe. A mode-filtered version of this cavity was constructed by

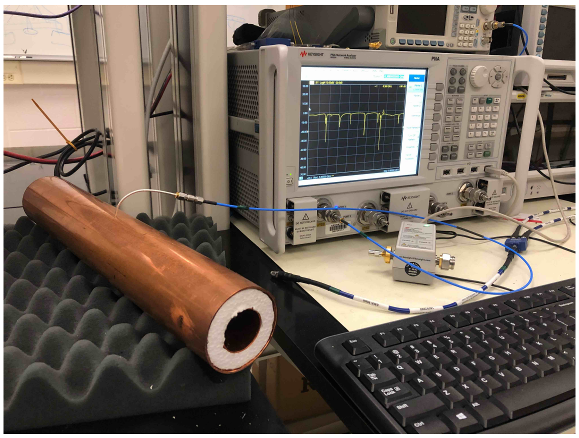
main cavity by carving a spacer from polystyrene foam (styrofoam) so that they could be easily inserted into the cavity and repositioned. The dielectric constant of styrofoam is quite close to air at microwave frequencies so this is expected to have minimal impact on the resonant properties of the cavity.

Table 6.4: A table of parameters describing the cavity prototypes. Certain values such as the cavity length and the distace between dielectric spacers are approximate due to variation in the machining of the copper. In particular, the filtered cavity was constructed from conducting copper segments that varied in size from 1.50" to 1.85".

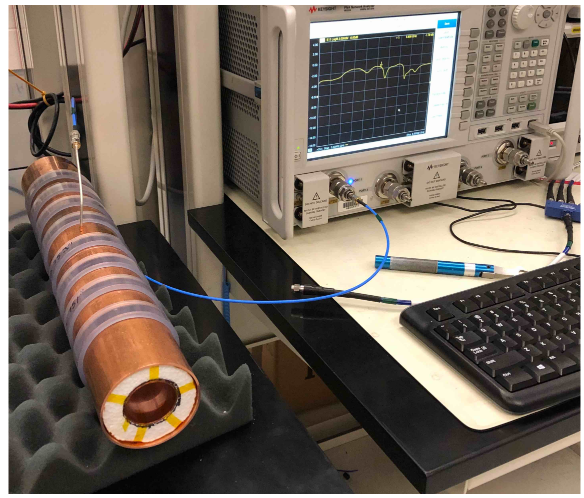
Name	Qty.	Unit	Description
D_{cav}	2.625	in	Cavity diameter
L_{cav}	≈ 13	in	Cavity length
D_{term}	1.375	in	Inner diameter of coaxial terminator
L_{term}	1.575	in	Terminator length
l_{die}	0.75	in	Dielectric spacer thickness
Δl_{die}	$\approx$ 1.50 to 1.85	in	Distance between dielectric spacers

The actual length of the cavity is given by the distance between the inner edges of the coaxial terminations. The length of the outer section of pipe that forms the main wall of the cavity is approximately 16" in length which leads to a cavity length of $\approx$ 13" when both terminators are inserted in the cavity. Because the terminators were not rigidly mounted this distance is only approximate, however, the uncertain length of the cavity will not prevent us from validating the open cavity design.

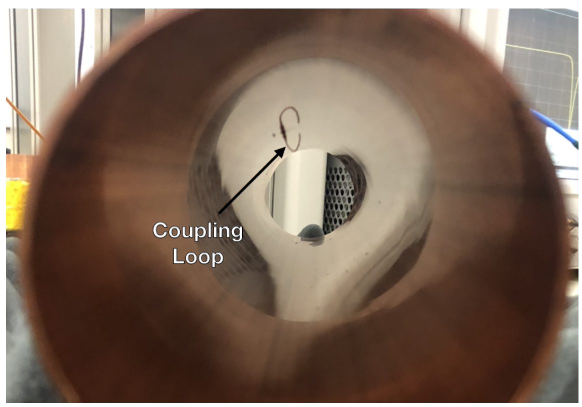
Along with the smooth-walled open cavity a resistively mode-filtered cavity was constructed by creating dielectric spacers out of segments of clear PVC pipe (see Figure 6.16). The spacers were machined such that the conductive segments of the cavity would be separated by 0.75" when the cavity was fully assembled. Due to variations in the lengths of the copper segments that make up the cavity wall the distance between spacers has significant variation with average value of about 1.7". Eight total spacers were used to build the cavity, which when assembled was approximately 16" in total length similar to the non-filtered cavity.



(a)



(b)



(c)

Figure 6.16: Images depicting the measurement of the filtered and non-filtered open cavities using the VNA. The coupling loop in the figure is shown in the TE orientation.

Measurements of both cavities were performed using a VNA connected to the cavity coupling probe (see Figure 6.16). By measuring the return loss over a range of frequencies one can measure the frequencies and relative Q-factors of the resonant modes in the cavity. Due to the opposite polarity of the electric fields for the TE and TM modes, the loop coupling probe must be rotated 90° to change the polarity of the loop antenna. When the antenna is oriented such that the loop opening faces the ends of the cavity, it couples primarily to the TE modes which have magnetic fields directed along the long axis of the cavity (see Figure 6.16). If the coupling loop is turned by 90° from where it is shown in the image then it will couple to the TM modes which have azimuthally directed magnetic fields. In this way both the TE and TM resonances can be measured independently.

6.5.2 Results and Discussion

The primary analysis for the prototype cavities involved a simple visualization of the return loss as measured by the VNA and a comparison between the filtered and non-filtered variations. Since the resonances measured by the VNA are not labeled, there is an uncertainty about the true identities of the modes measured by the VNA. To resolve this I performed a simulation of the simplest possible cavity that could be created from the prototype components, which is a fully open cavity created by removing the coaxial inserts. The fully-open cavity with the as-built dimensions was simulated in HFSS to get estimates on the positions of the TE_{011} and TM_{111} modes (see Figure 6.17).

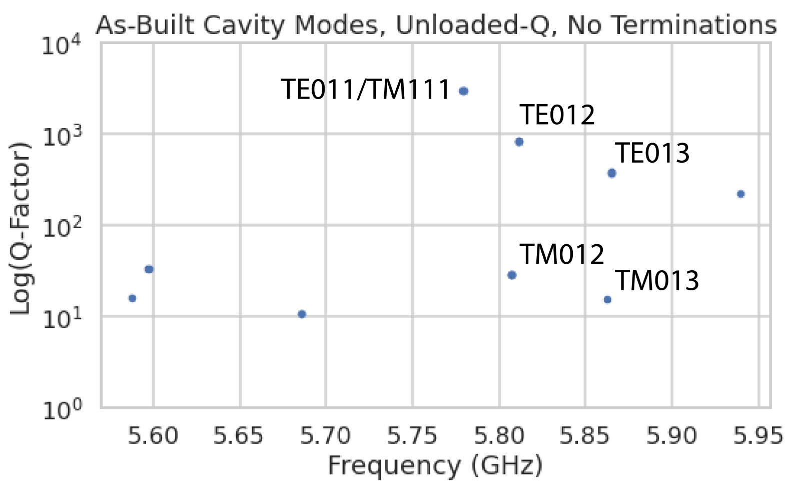
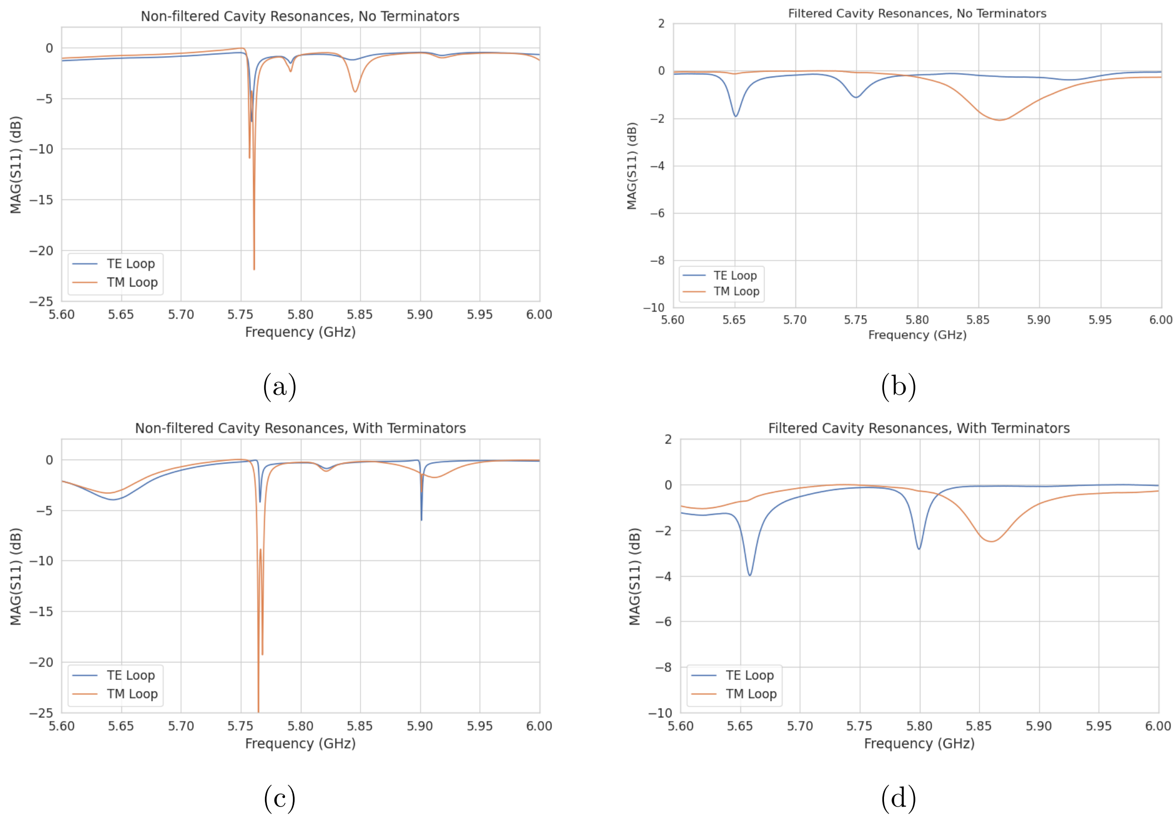


Figure 6.17: HFSS simulation results for a the as-built cavity with the coaxial terminators removed. The TE_{011}/TM_{111} frequency is approximately 5.78 GHz.

Simulation of the fully open cavity shows that the TE_{011}/TM_{111} modes have a frequency of approximately 5.78 GHz in the fully open cavity. If the frequency of this mode is compared to the measurments of the fitered and non-filtered cavities with the terminators removed one can easily identify the TE_{011} mode at approximately 5.75 GHz (see Figure 6.18).

Both variations of the non-filtered cavities one sees that the TE_{011} mode is degenerate in frequency with what appears to be a doublet of TM modes located at the TM_{111} frequency position. This doublet is actually the TM_{111} mode, which has two polarizations with opposite polarizations. Because the pipe used to construct the cavity is not perfectly round, the frequency degeneracy between the two polarizations is broken resulting in the doublet peaks.



(a) (b) (c) (d)

Figure 6.18: Measurements of the filtered and non-filtered prototype cavities acquired with the VNA.

The S-parameter plot for the filtered cavity without terminators has an isolated TE resonance at 5.65 GHz, associated with the TE_{011} mode. The frequency of this mode is lower than the non-filtered cavity due to a difference in the overall lengths of the cavities. An obvious difference between the filtered and non-filtered cavities is that there is no TM_{111} doublet at the TE_{011} frequency. This is what one would expect if the mode-filtering was suppressing the TM modes. There appears to be a noticeable difference in the Q of the TE_{011} resonance between non-filtered and filtered variations as indicated by the increased resonance depth for the filtered cavity. Overall, the Q-factors of the filtered cavity appear significantly smaller than the non-filtered cavity due to the increase in resonance width. This is likely caused by the relatively large widths of the dielectric spacers, which are partially impeding the TE modes.

One can see from these cavity measurements that, in principle, resistive mode-filtering can be used to separate the TE_{011} resonance from the degenerate TM_{111} modes in combination with the open cavity endcaps. This finding agrees with the expectations from HFSS, which should provide confidence that the eigenmode solver is correctly modeling the behavior of the cavity. Although I did not perform a similar study using a cavity with grooved walls it is expected that the resonant mode structure would be similar to the cavity studied here.

While this prototype cavity is a good first step, several deficiencies prevent this setup from providing more than qualitative information to the design of cavities for CRES. This includes the rudimentary approach to cavity coupling using a stripped coax antenna and the inability to map the field density in the cavity volume. Improvements in these areas are required so that measurements from a real cavity can provide useful information to cavity CRES simulations that will ultimately inform neutrino mass sensitivity estimates.

Future work with prototype cavities must include an improved cavity coupling scheme, which is robust and compatible with atomic tritium. Since the cavity will ultimately be filled with atomic tritium, a coupling antenna cannot be used due to the losses of atomic tritium caused by recombination on the antenna surfaces. Possible non-invasive coupling schemes include aperture coupling, where the cavity is coupled to an external waveguide structure through an aperture, or a split-ring coupling approach, where the center segment of the cylindrical cavity wall is replaced an isolated conductive ring with a small vertical slit. The aperture coupling approach is a standard coupling scheme [85] used in a wide range of applications, but at low frequencies the size of the external waveguide conflicts with design of the atom trapping magnet and cryogenics system. The split-ring approach could potentially be coupled to a small coaxial transmission line which is more compatible with the rest of the experiment design. A challenge is achieving adequate coupling through impedance tuning, which is a focus of current research.

The robustness of the coupling mechanism is relevant due to the difficulty in modeling its effect on the cavity modes. Small changes in geometry can have a large influence on the coupling and hence the performance of the cavity, therefore, correctly modeling the cavity coupling is critical for accurate CRES simulations. Coupling schemes that rely on connections to coaxial lines are potentially at a disadvantage in this regard due to the affect of soldering imperfections or unintended bends in the coax on the coupling. Future work will identify a coupling scheme for the cavity compatible with the neutrino mass goals of Project 8.

Imperfections in the geometry of a real cavity will necessarily distort the resonant modes away from simulation predictions. This will change the coupling of an electron to the cavity and thus change the expected signal structure. Ultimately, this effect will limit the achievable energy resolution of the experiment unless the differences between simulation and a real cavity can be sufficiently characterized and calibrated. One possible approach to this is to utilize a "bead puller" system [101] to strategically perturb the cavity by moving a conductive bead through the cavity volume. The small perturbation caused by the bead affects the phase of the cavity resonances proportional to the total magnitude of the electric field at that position, so by moving the bead through the cavity volume the total electric field can be mapped and compared to simulation. This information can provide bounds on the relative perturbations to the cavity mode structure from real-life imperfections compared to the idealized cavity in HFSS.

Chapter 7

Conclusion and Future Prospects

In this dissertation we have discussed research and development efforts towards the development of a scalable CRES measurement technology that can be used to build a CRES experiment at cubic-meter scales with sensitivity to neutrino masses of 40 meV. The primary contributions of my dissertation are the development and analysis of signal reconstruction algorithms for an antenna array based CRES experiment [102], which leads to estimates of the neutrino mass sensitivity; the development of a synthetic cyclotron radiation antenna (SYNCA) [79], which allowed for laboratory validation of antenna array CRES simulation models [42]; and the development of an open-ended cavity design compatible with atomic tritium for a cavity based CRES experiment. A measurable impact of this work is the transition of the Project 8 collaboration's experimental plan from an antenna array based approach to a cavity based approach, where my work played a key role in demonstrating the significantly higher cost and complexity of the antenna array experiment.

The transition from antenna arrays to cavities requires a new set of demonstrator experiments to make incremental progress towards a 40 meV measurement of the neutrino mass. At the time of writing, the near-term plan of Project 8 is to design and construct a small-scale cavity CRES experiment utilizing the 1 T magnet installed in the UW-Seattle. This cavity is designed to have a TE011 resonance with a frequency of about 26 GHz with a length-to-diameter ratio that mimics the larger cavities intended for the pilot-scale and Phase IV experiments. The goal of this experiment is to demonstrate cavity CRES as well as validate models of CRES systematics using electrons from ^{83m}Kr and an electron gun. Though the primary goal is demonstration, near-term physics measurements are available in the form of high-resolution measurements of the ^{83m}Kr conversion spectrum of interest to the KATRIN collaboration.

Furthermore, Project 8 is currently constructing a low-frequency CRES setup located at Yale University to better understand the principles of cavity based CRES at lower magnetic fields. The Low, UHF Cavity Krypton Experiment at Yale (LUCKEY) is a 1.5 GHz cavity CRES experiment the will use conversion electrons from ^{83m}Kr to perform CRES measurements at the lowest frequencies ever attempted with the technique. LUCKEY will validate frequency scaling models developed by Project 8 and will pave the way for the future Low-Frequency Apparatus (LFA), which will be a larger, 1 GHz cavity CRES experiment that includes a molecular tritium source. The target for the LFA is a measurement of the neutrino mass with a sensitivity of approximately 0.2 eV, which will build towards the atomic pilot-scale CRES experiment.

In parallel to the development of cavity CRES is the development of the atomic tritium source. Recent demonstrations of the production of atomic hydrogen are excellent steps towards the atomic tritium production needed for the pilot-scale experiment. One area of future study includes the development of a more detailed unstanding of the efficiency of atomic hydrogen production. Near-term plans include the development of a magnetic, evaporatively cooled beamline, as well as the prototyping of a Halbach array atoms trap. Nearly all of the components of the atomic tritium system will require demonstration before the complete system can be built. The long-term goal of the atomic tritium work is to construct a full atomic tritium prototype that demonstrates the production, cooling, trapping, and recycling of tritium at the rates needed for the pilot-scale experiment.

More broadly, the long-term goal of the Project 8 collaboration is to fully develop both the atomic tritium and cavity CRES technologies so that both can be combined in a pilot-scale CRES experiment. It is envisioned that this process will take approximately 10 years for both atomic tritium and cavity CRES. After these developments comes the pilot-scale experiment which will be the first CRES experiment that simultaneously demonstrates all the required technologies for Phase IV. Scaling to Phase IV with cavity CRES will require the construction of multiple copies (approximately 10) of the pilot-scale experiment to obtain sufficient statistics for 40 meV sensitivity.

Development of the CRES experimental technique by Project 8 has led to new experiments utilizing the CRES technique for basic physics research, such as the ^{6}He-CRES collaboration [103], and has also found applications as a new approach to x-ray spectroscopy [104]. Recently, a new experimental effort called CRESDA has begun in the UK to develop new quantum technologies applied to CRES measurements for the neutrino mass [105]. This flourishing of new experimental efforts based on the CRES technique is likely to continue as Project 8 continues to develop the technique towards its neutrino mass measurement goal.

Bibliography

[1] Das, A. and T. Ferbel (2003) *Introduction to Nuclear and Particle Physics*, World Scientific.

[2] Curie, P. (1898) "Sur une nouvelle substance fortement radioactive, contenue dans la pechblende," *Comptes rendus de 1'Academie des Sciences, Paris*, **127**, pp. 1215–1217.

[3] Rutherford, E. (1899) "Uranium radiation and the electrical conduction produced by it," *Philos. Mag.*, **47**, p. 109.
URL `https://cds.cern.ch/record/261896`

[4] Gerward, L. (1999) "Paul Villard and his Discovery of Gamma Rays," *Physics in Perspective*, **1**, pp. 367–383.

[5] Becquerel, H. (1900) "Deviation du rayonnement du radium dans un champ électrique," *C.R.*, **130**, pp. 809–815.

[6] D.Sc., E. R. M. and F. S. B.A. (1902) "XLI. The cause and nature of radioactivity.—Part I," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **4**(21), pp. 370–396, `https://doi.org/10.1080/14786440209462856`.
URL `https://doi.org/10.1080/14786440209462856`

[7] Chadwick, J. (1914) *Intensitätsverteilung im magnetischen Spektrum der ß-Strahlen von Radium B+C*, Druck von Friedr. Vieweg und Sohn.
URL `https://books.google.com/books?id=8s8UmQEACAAJ`

[8] Brown, L. M. (1978) "The idea of the neutrino," *Physics Today*, **31**(9).

[9] Pauli, W. (1978) "Dear radioactive ladies and gentlemen," *Phys. Today*, **31N9**, p. 27.

[10] Amaldi, E. (1984) "From the discovery of the neutron to the discovery of nuclear fission," *Physics Reports*, **111**(1), pp. 1–331.
URL `https://www.sciencedirect.com/science/article/pii/037015738490214X`

[11] Chadwick, J. (1932) "The existence of a neutron," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **136**(830), pp. 692–708, https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1932.0112.
URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1932.0112

[12] Wilson, F. L. (1968) "Fermi's Theory of Beta Decay," *American Journal of Physics*, **36**(12), pp. 1150–1160, https://pubs.aip.org/aapt/ajp/article-pdf/36/12/1150/11891052/1150_1_online.pdf.
URL https://doi.org/10.1119/1.1974382

[13] Crane, H. R. and J. Halpern (1938) "New Experimental Evidence for the Existence of a Neutrino," *Phys. Rev.*, **53**, pp. 789–794.
URL https://link.aps.org/doi/10.1103/PhysRev.53.789

[14] Cowan, C. L., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire (1956) "Detection of the Free Neutrino: a Confirmation," *Science*, **124**(3212), pp. 103–104, https://www.science.org/doi/pdf/10.1126/science.124.3212.103.
URL https://www.science.org/doi/abs/10.1126/science.124.3212.103

[15] Anicin, I. V. (2005) "The Neutrino - Its Past, Present and Future," *arXiv e-prints*, physics/0503172, physics/0503172.

[16] Davis, R. (2003) "Nobel Lecture: A half-century with solar neutrinos," *Rev. Mod. Phys.*, **75**, pp. 985–994.

[17] McDonald, A. B. et al. (2002) "Direct evidence for neutrino flavor transformation from neutral-current interactions in SNO," *AIP Conf. Proc.*, **646**(1), pp. 43–58.

[18] Fukuda, Y. et al. (1998) "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.*, **81**, pp. 1562–1567, hep-ex/9807003.

[19] Giganti, C., S. Lavignac, and M. Zito (2018) "Neutrino oscillations: The rise of the PMNS paradigm," *Prog. Part. Nucl. Phys.*, **98**, pp. 1–54, 1710.00715.

[20] An, F. et al. (2016) "Neutrino Physics with JUNO," *J. Phys. G*, **43**(3), p. 030401, 1507.05613.

[21] Bian, J. et al. (2022) "Hyper-Kamiokande Experiment: A Snowmass White Paper," in *Snowmass 2021*, 2203.02029.

[22] Abi, B. et al. (2020) "Long-baseline neutrino oscillation physics potential of the DUNE experiment," *Eur. Phys. J. C*, **80**(10), p. 978, 2006.16043.

[23] Workman, R. L. et al. (2022) "Review of Particle Physics," *PTEP*, **2022**, p. 083C01.

[24] Zuber, K. (2020) *Neutrino Physics*, CRC Press.

[25] Xing, Z. Z. and S. Zhou (2011) *Neutrinos in Particle Physics, Astronomy and Cosmology*, Springer.

[26] Mohapatra, R. N. et al. (2007) "Theory of neutrinos: A White paper," *Rept. Prog. Phys.*, **70**, pp. 1757–1867, `hep-ph/0510213`.

[27] de Gouvêa, A. (2016) "Neutrino Mass Models," *Ann. Rev. Nucl. Part. Sci.*, **66**, pp. 197–217.

[28] Workman, R. L. and Others (2022) "Review of Particle Physics," *PTEP*, **2022**, p. 083C01.

[29] Wong, Y. Y. Y. (2011) "Neutrino mass in cosmology: status and prospects," *Ann. Rev. Nucl. Part. Sci.*, **61**, pp. 69–98, `1111.1436`.

[30] Ade, P. A. R. et al. (2016) "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.*, **594**, p. A13, `1502.01589`.

[31] Carlstrom, J. et al. (2019) "CMB-S4," in *Bulletin of the American Astronomical Society*, vol. 51, p. 209, `1908.01062`.

[32] Elliott, S. R. and P. Vogel (2002) "Double beta decay," *Ann. Rev. Nucl. Part. Sci.*, **52**, pp. 115–151, `hep-ph/0202264`.

[33] Dolinski, M. J., A. W. P. Poon, and W. Rodejohann (2019) "Neutrinoless Double-Beta Decay: Status and Prospects," *Ann. Rev. Nucl. Part. Sci.*, **69**, pp. 219–251, `1902.04097`.

[34] Agostini, M., G. Benato, and J. Detwiler (2017) "Discovery probability of next-generation neutrinoless double- β decay experiments," *Phys. Rev. D*, **96**(5), p. 053001, `1705.02996`.

[35] Formaggio, J. A., A. L. C. de Gouvêa, and R. G. H. Robertson (2021) "Direct measurements of neutrino mass," *Physics Reports*, **914**, pp. 1–54, direct measurements of neutrino mass.
URL `https://www.sciencedirect.com/science/article/pii/S0370157321000636`

[36] Aker, M. et al. (2022) "New Constraint on the Local Relic Neutrino Background Overdensity with the First KATRIN Data Runs," *Phys. Rev. Lett.*, **129**(1), p. 011806, `2202.04587`.

[37] Formaggio, J. A., A. L. C. de Gouvêa, and R. G. H. Robertson (2021) "Direct Measurements of Neutrino Mass," *Phys. Rept.*, **914**, pp. 1–54, `2102.00594`.

[38] Monreal, B. and J. A. Formaggio (2009) "Relativistic cyclotron radiation detection of tritium decay electrons as a new technique for measuring the neutrino mass," *Phys. Rev. D*, **80**, p. 051301.
URL `https://link.aps.org/doi/10.1103/PhysRevD.80.051301`

[39] Esfahani, A. A. et al. (2022) "The Project 8 Neutrino Mass Experiment," `2203.07349`.

[40] ——— (2023) "Cyclotron Radiation Emission Spectroscopy of Electrons from Tritium Beta Decay and $^{83\mathrm{m}}$Kr Internal Conversion," `2303.12055`.

[41] ——— (2023) "Tritium Beta Spectrum and Neutrino Mass Limit from Cyclotron Radiation Emission Spectroscopy," `2212.05048`.

[42] "Antenna Arrays for Physics Measurements with Large-scale CRES Detectors," *In preparation*.

[43] Nicholson, T. et al. (2015) "Systematic evaluation of an atomic clock at 2 x 10-18 total uncertainty," *Nature Communications*, **6**(6896).

[44] Hänsch, T. W. (2006) "Nobel Lecture: Passion for precision," *Rev. Mod. Phys.*, **78**, pp. 1297–1309.
URL `https://link.aps.org/doi/10.1103/RevModPhys.78.1297`

[45] Gabor, D. (1946) "Theory of communication. Part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-part III: radio and communication engineering*, **93**(26), pp. 429–441.

[46] Esfahani Ashtari, A. et al. (2022) "Viterbi decoding of CRES signals in Project 8," *New J. Phys.*, **24**(5), p. 053013, `2112.05265`.

[47] Jackson, J. D. (1999) *Classical electrodynamics*, 3rd ed., Wiley, New York, NY.
URL `http://cdsweb.cern.ch/record/490457`

[48] F.R.S., J. L. D. (1897) "LXIII. On the theory of the magnetic influence on spectra; and on the radiation from moving ions," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **44**(271), pp. 503–512, `https://doi.org/10.1080/14786449708621095`.
URL `https://doi.org/10.1080/14786449708621095`

[49] Esfahani, A. A. et al. (2022) "The Project 8 Neutrino Mass Experiment," in *2022 Snowmass Summer Study*, `2203.07349`.

[50] Kraus, C. et al. (2005) "Final results from phase II of the Mainz neutrino mass search in tritium beta decay," *Eur. Phys. J. C*, **40**, pp. 447–468, `hep-ex/0412056`.

[51] Aseev, V. N. et al. (2011) "An upper limit on electron antineutrino mass from Troitsk experiment," *Phys. Rev. D*, **84**, p. 112003, `1108.5034`.

[52] Saenz, A., S. Jonsell, and P. Froelich (2000) "Improved Molecular Final-State Distribution of HeT+ for the β-Decay Process of T2," *Phys. Rev. Lett.*, **84**(2), p. 242.

[53] Bodine, L. I., D. S. Parno, and R. G. H. Robertson (2015) "Assessment of molecular effects on neutrino mass measurements from tritium β decay," *Phys. Rev. C*, **91**(3), p. 035505, 1502.03497.

[54] Asner, D. M. et al. (2015) "Single electron detection and spectroscopy via relativistic cyclotron radiation," *Phys. Rev. Lett.*, **114**(16), p. 162501, 1408.5362.

[55] Vénos, D., J. Sentkerestiová, O. Dragoun, M. Slezák, M. Ryšavý, and A. Špalek (2018) "Properties of 83mKr conversion electrons and their use in the KATRIN experiment," *JINST*, **13**(02), p. T02012.

[56] Lagendijk, A., I. F. Silvera, and B. J. Verhaar (1986) "Spin exchange and dipolar relaxation rates in atomic hydrogen: Lifetimes in magnetic traps," *Phys. Rev. B*, **33**, pp. 626–628.
URL https://link.aps.org/doi/10.1103/PhysRevB.33.626

[57] Silvera, I. and J. Walraven (1986) "Spin-polarized atomic hydrogen," *Prog. Low Temp. Phys.*, **X**, pp. 139–370.

[58] Furse, D. et al. (2017) "Kassiopeia: a modern, extensible C++ particle tracking package," *New Journal of Physics*, **19**(5), p. 053012.
URL https://doi.org/10.1088/1367-2630/aa6950

[59] Esfahani, A. A. et al. (2019) "Electron radiated power in cyclotron radiation emission spectroscopy experiments," *Phys. Rev. C*, **99**, p. 055501.
URL https://link.aps.org/doi/10.1103/PhysRevC.99.055501

[60] Ashtari Esfahani, A. et al. (2019) "Locust: C++ software for simulation of RF detection," *New J. Phys.*, **21**, p. 113051, 1907.11124.

[61] Wiechert, E. (1901) "Elektrodynamische Elementargesetze," .
URL https://doi.org/10.1002/andp.19013090403

[62] Liéard, A. (1898) "Champ électrique et Magnétique," *Léclairage électrique*, **16**(27-29).

[63] Balanis, C. (2015) *Antenna Theory: Analysis and Design*, Wiley.
URL https://books.google.com/books?id=PTFcCwAAQBAJ

[64] https://www.ansys.com/products/electronics/ansys-hfss.

[65] https://en.wikipedia.org/wiki/Linear_time-invariant_system.

[66] Nyquist, H. (1928) "Certain Topics in Telegraph Transmission Theory," *Transactions of the American Institute of Electrical Engineers*, **47**(2), pp. 617–644.

[67] Brun, R. and F. Rademakers (1997) "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth. A*, **389**, pp. 81–86.

[68] Ashtari Esfahani, A. et al. (2021) "Bayesian analysis of a future β decay experiment's sensitivity to neutrino mass scale and ordering," *Phys. Rev. C*, **103**, p. 065501.
URL `https://link.aps.org/doi/10.1103/PhysRevC.103.065501`

[69] Kay, S. (1998) *Fundamentals of Statistical Signal Processing: Detection Theory, Volume II*, Pearson.

[70] Neyman, J. and E. Pearson (1933) "On the problem of the the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc. Lond. A*, **231**.

[71] Stumpf, M. (2018) *Electromagnetic Reciprocity in Antenna Theory*, Wiley.

[72] Bishop, C. (2016) *Pattern Recognition and Machine Learning*, Springer.

[73] Plehn, T., A. Butter, B. Dillon, and C. Krause (2022) "Modern Machine Learning for LHC Physicists," `2211.01421`.

[74] George, D. and E. A. Huerta (2018) "Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data," *Phys. Lett. B*, **778**, pp. 64–70, `1711.03121`.

[75] Gabbard, H., C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith (2022) "Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy," *Nature Phys.*, **18**(1), pp. 112–117, `1909.06296`.

[76] Reimann, R. (2022) "Project 8: R&D for a next-generation neutrino mass experiment," *PoS*, **PANIC2021**, p. 283.

[77] LaRoque, B. (2020) "Zero-deadtime processing in beta spectroscopy for measurement of the non-zero neutrino mass," *EPJ Web Conf.*, **245**, p. 01014.

[78] Buzinsky, N. (2021) *Statistical Signal Processing and Detector Optimization in Project 8*, Ph.D. thesis, Massachusetts Institute of Technology.

[79] Esfahani, A. A. et al. (2023) "SYNCA: A Synthetic Cyclotron Antenna for the Project 8 Collaboration," *Journal of Instrumentation*, **18**(01), p. P01034.
URL `https://dx.doi.org/10.1088/1748-0221/18/01/P01034`

[80] `https://www.nvidia.com/en-us/data-center/v100/`.

[81] `https://www.nvidia.com/en-us/data-center/h100/`.

[82] He, K., X. Zhang, S. Ren, and J. Sun (2016) "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

[83] Simonyan, K. and A. Zisserman (2015) "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.).
URL `http://arxiv.org/abs/1409.1556`

[84] Friis, H. (1946) "A Note on a Simple Transmission Formula," *Proceedings of the IRE*, **34**(5), pp. 254–256.

[85] Pozar, D. M. (2005) *Microwave engineering; 3rd ed.*, Wiley, Hoboken, NJ.
URL `https://cds.cern.ch/record/882338`

[86] Aker, M. et al. (2022) "Direct neutrino-mass measurement with sub-electronvolt sensitivity," *Nature Physics*, **18**(2), pp. 160–166.
URL `https://doi.org/10.1038/s41567-021-01463-1`

[87] Monreal, B. and J. A. Formaggio (2009) "Relativistic Cyclotron Radiation Detection of Tritium Decay Electrons as a New Technique for Measuring the Neutrino Mass," *Phys. Rev. D*, **80**, p. 051301, `0904.2860`.

[88] Ashtari Esfahani, A. et al. (2017) "Determining the neutrino mass with cyclotron radiation emission spectroscopy—Project 8," *J. Phys. G*, **44**(5), p. 054004, `1703.02037`.

[89] Esfahani, A. A. et al. (2022) "The Project 8 Neutrino Mass Experiment," in *2022 Snowmass Summer Study*, `2203.07349`.

[90] Ashtari Esfahani, A. et al. (2019) "Locust: C++ software for simulation of RF detection," *New Journal of Physics*, **21**(11), p. 113051.
URL `https://doi.org/10.1088/1367-2630/ab550d`

[91] Furse, D. et al. (2017) "Kassiopeia: a modern, extensible C++ particle tracking package," *New Journal of Physics*, **19**(5), p. 053012.
URL `https://doi.org/10.1088/1367-2630/aa6950`

[92] Balanis, C. (2011) *Modern Antenna Handbook*, Wiley.
URL `https://books.google.com/books?id=UYpV8L8GNCwC`

[93] Wirth, W. (2001) *Radar Techniques Using Array Antennas*, Institution of Engineering and Technology.
URL `https://books.google.com/books?id=ALht42gkzLsC`

[94] Brown, L. S., G. Gabrielse, K. Helmerson, and J. Tan (1985) "Cyclotron motion in a microwave cavity: Lifetime and frequency shifts," *Phys. Rev. A*, **32**, pp. 3204–3218.
URL `https://link.aps.org/doi/10.1103/PhysRevA.32.3204`

[95] Hanneke, D., S. Fogwell Hoogerheide, and G. Gabrielse (2011) "Cavity control of a single-electron quantum cyclotron: Measuring the electron magnetic moment," *Phys. Rev. A*, **83**, p. 052122.
URL `https://link.aps.org/doi/10.1103/PhysRevA.83.052122`

[96] Hanneke, D. A. (2007) *Cavity control in a single-electron quantum cyclotron: an improved measurement of the electron magnetic moment*, Ph.D. thesis, Harvard U.

[97] Purcell, E. (1946) "Spontaneous Emission Probabilities at Radio Frequencies," *Phys. Rev.*, **69**, pp. 674–674.
URL `https://link.aps.org/doi/10.1103/PhysRev.69.674`

[98] Moskowitz, B. E. and J. Rogers (1988) "ANALYSIS OF A MICROWAVE CAVITY DETECTOR COUPLED TO A NOISY AMPLIFIER," *Nucl. Instrum. Meth. A*, **264**, pp. 445–452.

[99] Wenger, N. (1967) "Resonant Frequency of Open-Ended Cylindrical Cavity," *IEEE Transactions on Microwave Theory and Techniques*, **15**(6), pp. 334–340.

[100] Wenger, N. C. and J. Smetana (1972) "Hydrogen Density Measurements Using an Open-Ended Microwave Cavity," *IEEE Transactions on Instrumentation and Measurement*, **21**(2), pp. 105–114.

[101] Rapidis, N. M. (2018) "Application of the Bead Perturbation Technique to a Study of a Tunable 5 GHz Annular Cavity," *Springer Proc. Phys.*, **211**, pp. 45–51, `1708.04276`.

[102] "Real-time Signal Detection for Cyclotron Radiation Emission Spectroscopy Measurements using Antenna Arrays," *In preparation.*

[103] Byron, W. et al. (2022) "First observation of cyclotron radiation from MeV-scale e^{pm} following nuclear beta decay," `2209.02870`.

[104] Kazkaz, K. and N. Woollett (2021) "Using Cyclotron Radiation Emission for Ultra-high Resolution X-Ray Spectroscopy," *New J. Phys.*, **23**(3), p. 033043, `1911.05869`.

[105] Canning, J. A. L., F. F. Deppisch, and W. Pei (2023) "Sensitivity of future tritium decay experiments to New Physics," *JHEP*, **03**, p. 144, `2212.06106`.