University of California
Santa Barbara

Cyclotron Radiation Emission Spectroscopy: first demonstration and performance benchmarks from the Project 8 experiment

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Physics

by

Benjamin Hines LaRoque

Committee in charge:

Professor Benjamin Monreal, Chair
Professor David Berenstein
Professor Harry Nelson

March 2017

The Dissertation of Benjamin Hines LaRoque is approved.

Professor David Berenstein

Professor Harry Nelson

Professor Benjamin Monreal, Committee Chair

January 2017

Cyclotron Radiation Emission Spectroscopy: first demonstration and performance benchmarks from the Project 8 experiment

by

Benjamin Hines LaRoque

In memory of my grandparents, Jean and Cecil Turner and Naomi and Walter LaRoque.

Acknowledgements

As with any large project, the number of people who have helped me along the way is a long one. Project 8 may still be a relatively small collaboration, and without the contributions of every person involved, we would not made nearly so much progress. In particular, working with and recieving guidance from Jared, Gray, and Noah from day one, through the successful operation of Phase I was incredibly valuable.

I would also like to thank my friends and family for their unyielding support. My parents, Mark and Roseann, have always supported and encouraged me to follow my passion. My partner Kimberly is on the same Ph.D. journey, and her support has helped me maintain focus when times were discouraging.

I'd also like to express my gratitude to the administrative support staff of the University of California at Santa Barbara Physics department and both CENPA and Physics at the University of Washington; the work they do has allowed me to focus on doing physics in the phase of administrative challeneges.

Curriculum Vitæ

Benjamin Hines LaRoque

Education

2017 Ph.D. in Physics (this work), University of California, Santa Barbara.

2013 M.A. in Physics, University of California, Santa Barbara.

2010 B.S. in Physics, University of North Carolina, Chapel Hill.

Publications

1. D. M. Asner et al. (Project 8 Collaboration), "Single-Electron Detection and Spectroscopy via Relativistic Cyclotron Radiation", *Phys. Rev. Lett.* **114**, 162501 (20 April 2015).
2. S. MacMullin, M. Boswell, M. Devlin, S. R. Elliott, N. Fotiades, V. E. Guiseppe, R. Henning, T. Kawano, B. H. LaRoque, R. O. Nelson, and J. M. O'Donnell, "Neutron-induced *gamma*-ray production cross sections for the first excited-state transitions in Ne-20 and Ne-22", *Phys. Rev. C* **86**, 067601 (21 December 2012).
3. S. MacMullin, M. Boswell, M. Devlin, S. R. Elliott, N. Fotiades, V. E. Guiseppe, R. Henning, T. Kawano, B. H. LaRoque, R. O. Nelson, and J. M. O'Donnell, "Partial γ-ray production cross sections for (n,xnγ) reactions in natural argon at 1-30 MeV," *Phys. Rev. C* **85**, 064614 (21 June 2012).
4. S. R. Elliott, B. H. LaRoque, V. M. Gehman, M. F. Kidd, M. Chen, "An Improved Limit on Pauli-Exclusion-Principle Forbidden Atomic Transitions", *Found. Phys.* **42** 1015 (16 March 2012).
5. S. R. Elliott, V. E. Guiseppe, B. H. LaRoque, R. A. Johnson, and S. G. Mashnik, "Fast-neutron activation of long- lived isotopes in enriched Ge", *Phys. Rev. C* **82**, 054610 (16 June 2010).

Presentations

1. *Results from the Project 8 phase-1 cyclotron radiation emission spectroscopy detector* "Neutrino 2016," London, United Kington. July 2016 **(contributed poster)**
2. *Design and performance of the Project 8 phase-1 detector.* "Determination of the absolute electron (anti)-neutrino mass," European Centre for Theoretical Studies in Nuclear Physics and Related Areas, Trento Italy. April 2016 **(invited talk)**
3. *Status of the Project 8 Experiment.* "IPA 2015: Cosmic Neutrinos, What Next?" Wisconsin IceCube Particle Astrophysics Center, Madison Wisconsin. May 2015. **(invited talk)**
4. *Investigation of an active radio frequency technique for electron spectroscopy* "New Directions in Neutrino Physics" Aspen Center for Physics, Aspen Colorado. February 2013. **(contributed poster)**

5. *Empirical determination of effective germanium detector efficiency for use in sample assays* "3rd Joint Meeting of the APS Division of Nuclear Physics and the Physical Society of Japan" Waikoloa Hawaii. October 2009. **(contributed poster)**

Research Experience

Graduate Student Researcher June 2011 - present
Monreal Group, Department of Physics, University of California at Santa Barbara

- Designed and implemented software framework for monitoring and operating the experiment which is modular and distributed
- Participated in the design and commissioning of the vacuum, magnet, and data acquisition subsystems and maintained systems during operation. (publication 1)
- Coordinate the efforts of other students and postdocs, including assignment of tasks and integration into the production system

Post Baccalaureate Research Assistant May 2010 - June 2011
Weak Interactions Team, Neutron Science & Technology Group, Los Alamos National Laboratory

- Collaborated with an engineer to design, then built and operated an apparatus to search for violations of the Pauli-Exclusion principle in atomic transitions. (publication 4)
- Provided operational support and participated in data analysis for measurements of neutron scattering cross sections for nobel gases relevant to dark matter detection experiments. (publications 2, 3)
- Conducted quality assurance testing of BEGe detectors following the proceedure established during summer 2009 research experience.

Undergraduate Research Assistant August 2009 - May 2010
Experimental Nuclear and Particle Astrophysics group, Department of Physics and Astronomy, University of North Carolina at Chapel Hill

- Developed and implemented a noise performance test plan for digital signal filters using python, C++, and ROOT programming languages.
- Worked with scientists at Los Alamos National Laboratory and engineers from Agilent Technologies to repair and troubleshoot a compactPCI digitizer system.

ST2 in Undergraduate Student Program June-August 2009
Weak Interactions Team, neutron Science & Technology Group, Los Alamos National Laboratory

- Standardized the report format for contract specification compliance of Cenberra BEGe detector crystals
- Developed an empirical method for nonlinear detector calibration of custom germanium detector (publication 5)

Undergraduate Research Assistant Summers 2007, 2008, Fall 2008
Coastal Circulation and Transport Lab, Department of Marine Sciences, University of North Carolina at Chapel Hill

- Ran large scale tidal simulations using highly parallel Fortran software on research computing cluster
- Created original Matlab and Fortran code to process and analyze simulation results and compare to empirical data

Abstract

Cyclotron Radiation Emission Spectroscopy: first demonstration and performance benchmarks from the Project 8 experiment

by

Benjamin Hines LaRoque

The Project 8 collaboration is taking a phased approach to developing an experimental search for the absolute neutrino mass scale, based on a novel technique, Cyclotron Radiation Emission Spectroscopy. The first phase was a demonstration of this new spectroscopy technique using a well understood source of narrow conversion electron lines, ^{83m}Kr, as a proof of principle. Results from the first successful operation of the detector are presented, demonstrating the viability of the approach. The strong conversion electron lines near 17.8, 30.4, and 32 keV were observed with full width at half maximum between 140 eV and 15 eV depending on the choice of trapping configuration used.

Various upgrades were made to the detector prior to its being operated with the specific goal of determining a performance baseline for planning future phases. These included alternative trapping configurations, with which the observed full width at half maximum has been improved to 3.6 eV. Evaluation of the event reconstruction and data quality are presented based on this data collection period. Areas where improvements will be required for phase II, when the approach will be used for the first time to measure a electrons from a continuous spectrum, are identified.

Contents

Chapter 1

Introduction to neutrino mass measurement

Neutrinos have been been part of the development of our understanding of nuclear and particle physics since they were first proposed in the early 20th century. That history is an interesting story which could, and has, filled many volumes well beyond the scope of a single dissertation. Therefore, here I will not even attempt to do justice to that history, but will instead point out some highlights that motivate our modern experimental efforts towards understanding neutrino masses. For a more complete historical treatment in a less technical context, I recommend *Are There Really Neutrinos?* [1]; a more complete and technical treatment is available in the collection of review articles *Current Aspects of Neutrino Physics* [2], though obviously it is becoming a bit dated. The review articles and parameter values in the *Review of Particle Physics* [3] are, of course, regularly updated.

1.1 Our understanding of the neutrino

By the end of the 1920s, it was well established that the β decay spectrum is continuous, but with only two known particles in the final state, conservation of energy and momentum predicts a discrete spectrum. In order to avoid abandoning those

conservation laws, Wolfgang Pauli postulated the existence of an undetected third particle in the final state, which would balance the observed products and explain the observed spectral shape [4]. Over the following decades, various experiments substantiated this hypothesis, culminating with the work of Reines and Cowan, which made the first direct detections of neutrinos, observing the inverse β decay process

$$\bar{\nu} + p^+ \rightarrow n + e^+. \tag{1.1}$$

Note that here the incident particle is actually an anti-neutrino; at the time these experimental results were released, that distinction was not commonly made, and the results used the term neutrino.

The understanding of neutrinos continued to develop in conjunction with the rest of nuclear and particle physics through the twentieth century. As originally formulated, the Standard Model of particle physics includes neutrinos which are massless and with three distinguishable eigenstates under the weak interaction, each paired with one of the charged leptons (electron, muon, and tau). The solar neutrino problem, raised by the unexpected results of the Homestake experiment, led directly the search for, and discovery of, neutrino oscillations. A critical implication of the observation of oscillations is that neutrinos must have distinct mass eigenstates which are not simultaneous with the eigenstates under the weak interaction.

The relationship between the two bases is given by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, which can be expanded in terms of empirically accessible

mixing parameters as

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.2}$$

where s_{li} and c_{li} are shorthand for $\sin(\theta_{li})$ and $\cos(\theta_{li})$. Here l indexes the flavor eigenstates (corresponding to electron, muon, and tau neutrinos, but sometimes labeled 1, 2, and 3 respectively) and i indexes the mass eigenstates (conventionally labeled 1, 2, and 3). In addition to the three mixing angles (θ_{12}, θ_{13}, and θ_{23}), there are three phases (δ, α_1, and α_2); taken together, these angles and phases fully define the matrix. The Dirac phase, δ, allows interactions which violate Charge-Parity (CP) symmetry. The Majorana phases, α_1 and α_2, are only present in the case where the neutrino is its own antiparticle, implying that total lepton number is not a conserved quantity. The existence of neutrino oscillations means that individual lepton numbers, associated with each of the weak states, are not conserved.

There have been many experimental efforts seeking to explore the various components of the PMNS matrix. In general these efforts are sensitive to the difference between a pair of mass states as well as a specific subset of mixing angles and phases. A summary of these efforts, including very recent results measuring θ_{13} can be found in chapter 14 of the *Review of Particle Physics* ([3]). Efforts to measure the CP-violating Dirac phases are ongoing, as well as more efforts to more precisely measure the mixing angles and to determine the sign of the larger mass difference. In table 1.1 I capture the current empirical understanding of the various mixing parameters (values are taken from table 14.1 of [3]).

There are a few important observations to be made when reviewing table 1.1. First, while the sign of Δm^2_{21} is known, that of the larger mass difference is not;

Parameter		Best fit value	Allowed interval at 3σ
Δm^2_{21}		$7.37{*}10^{-5} eV^2$	(6.93 - 7.97)$*10^{-5} eV^2$
$\lvert\Delta m^2\rvert$		$2.50{*}10^{-3} eV^2$	(2.37 - 2.63)$*10^{-3} eV^2$
–	$(\Delta m^2 < 0)$	$2.46{*}10^{-3} eV^2$	(2.33 - 2.60)$*10^{-3} eV^2$
$\sin^2(\theta_{12})$		0.297	0.250 – 0.354
$\sin^2(\theta_{23})$		0.437	0.379 – 0.616
–	$(\Delta m^2 < 0)$	0.569	0.383 – 0.637
$\sin^2(\theta_{13})$		0.0214	0.0185 – 0.0246
–	$(\Delta m^2 < 0)$	0.0218	0.0186 – 0.0248
δ/π		1.35	*no exclusion at 3σ*
–	$(\Delta m^2 < 0)$	1.32	*no exclusion at 3σ*

Table 1.1: Neutrino mixing parameters as listed in [3]. Following that source, the parameter $|\Delta m^2|$ is defined as $|\Delta m^2| = m_3^2 - \frac{1}{2}(m_2^2 + m_1^2)$ and $\Delta m^2_{ij} = m_i^2 - m_j^2$. For parameters that depend on the mass ordering, the first entry always corresponds to the $\Delta m^2 > 0$ case and the second to the $\Delta m^2 < 0$ case (the latter case is indicated). Note that at 3σ, all values of δ are currently allowed.

that is, it is not known if the mass states are ordered $m_1 < m_2 < m_3$ (normal ordering) or $m_3 < m_1 < m_2$ (inverted ordering). This has an impact when evaluating many parameters from experimental data, which is why each has a best fit value and allowed interval listed under each scenario. Second, oscillation experiments are not sensitive to the magnitude of the masses, but instead provide a measurement only of the differences between the masses. Absolute measurements must come from another class of measurement. The rest of this section will summarize the main experimental avenues for investigating the absolute neutrino mass scale.

1.1.1 Observational cosmology

A very sensitive probe of the neutrino mass scale comes from observational cosmology. Neutrinos produced in the early universe behaved as a source of hot dark matter, which impacted the formation of large scale structures. While neutrinos alone do not account for the apparent density of dark matter, their impact must still be included in any model which tries to explain the evolution of the universe. Observations of

the Cosmic Microwave Background (CMB), and other cosmology measurements, can therefore be used to either estimate or constrain the neutrino mass scale, under the assumption of the cosmology model being considered.

For these measurements, the parameter extracted is the sum of neutrino masses $m_{\mathrm{tot}} = \sum m_i$. The Plank collaboration's most recently released limit, which combines CMB spectra with a number of other observations, yields an upper bound of $m_{\mathrm{tot}} < 0.17\,\mathrm{eV}$ at 95% confidence level [5]. This limit is derived under the assumption of the ΛCDM model and involves the simultaneous extraction of many parameters. While the statistical methods used are surely robust, it is important to note that any change to the underlying model can impact surely impact the results for individual parameters. If the neutrino mass scale were determined by some other experimental means, that information could be used as an external constraint and the statistical power of the combined cosmological data sets could be applied to other parameters not terrestrially accessible.

1.1.2 Searches for neutrinoless double β decay

Double β decay is a process whereby a nucleus, which is energetically forbidden from undergoing normal β decay, undergoes a single decay process where it emits a pair of β particles, simultaneously converting a pair of neutrons both to protons in an energetically allowed daughter. Such a process is allowed under the standard model if each electron is emitted with an associated antineutrino, a process which has been observed. If neutrinos are Majorana particles, then they are their own anti-particles, and it is possible that the antineutrino associated with one electron can be absorbed as a neutrino at the vertex where the second electron is produced, resulting in a final state with only the daughter nucleus and the two electrons. Since there are no neutrinos present to share the released energy, the signature is a narrow excess of

events at the endpoint of the decay, just above the end of the continuous spectrum from two neutrino decay. The rate at which such a process occurs is sensitive to the effective mass for the exchanged particle, $m_{\beta\beta}$, given by

$$m_{\beta\beta}^2 = \left| \sum m_i U_{e,i}^2 \right|^2 . \tag{1.3}$$

It is more common for experiments setting bounds from null results to express them as lower bounds on the half-life of the decay.

Interpreting the results from experimental searches requires estimates of the relevant nuclear matrix elements, which are not not currently well known. As as result, it can be complicated to compare results from experiments which utilizing different source isotopes, and determining bounds on neutrino mass parameters have similar difficulties. Depending on the choice of nuclear matrix elements, combined experimental results in ^{136}Xe give upper bounds on $m_{\beta\beta}$ between 0.11 eV and 0.25 eV, while combined experimental results in ^{76}Ge range from 0.21 eV to 0.44 eV [6].

In addition to the uncertainty from the nuclear matrix elements, these limits are dependent on the assumption that neutrinos are Majorana particles. If neutrinos are purely Dirac, then the process would cannot occur, regardless of the neutrino mass scale, and these limits would not be meaningful. Further, the situation can be reversed. If the neutrino mass scale were known from another experiment, then a non-observation of neutrinoless double β decay at the rate predicted by the mass value could rule out the Majorana nature of the neutrino.

1.1.3 Spectroscopy of single β nuclear transitions

Electron capture and β decay are related nuclear processes whereby a lower energy nuclear configuration is achieved by converting one nucleon (proton or neutron) to

the other type. Being general, one can consider four distinct processes,

$$\mathrm{p}^+ + \mathrm{e}^- \rightarrow \mathrm{n} + \nu_e \text{ (electron capture)} \tag{1.4}$$

$$\mathrm{n} + \mathrm{e}^+ \rightarrow \mathrm{p}^+ + \bar{\nu_e} \text{ (positron capture)} \tag{1.5}$$

$$\mathrm{n} \rightarrow \mathrm{p}^+ + \bar{\nu_e} + \mathrm{e}^- \ (\beta^- \text{ decay}) \tag{1.6}$$

$$\mathrm{p}^+ \rightarrow \mathrm{n} + \nu_e + \mathrm{e}^+ \ (\beta^+ \text{ decay}), \tag{1.7}$$

which convert back and forth between protons and neutrons. Of these four, positron capture is not observed because atoms have only electrons, not positrons, available for capture. In principle, high precision spectroscopy of any of these processes will be sensitive to the non-zero neutrino mass scale through a distortion of the spectral shape near the reaction's endpoint, since the rest mass of the neutrino limits the amount of energy available for the kinetic energy of the daughter particles.

These processes are sensitive to the effective mass term m_β, given by

$$m_\beta^2 = \sum |U_{ei}|^2 \, m_i^2. \tag{1.8}$$

More specifically, the energy spectrum is given by Fermi's golden rule as

$$\frac{\mathrm{d}N}{\mathrm{d}E} = G_F^2 \frac{m_e^5 c^4}{2\pi^3 \hbar^7} \cos^2(\theta_C) \, |M|^2 \, F(Z,E) p_e (T + m_e c^2)(E_0 - T) \sqrt{(E_0 - T)^2 - m_\beta^2 c^4}, \tag{1.9}$$

where G_F, m_e, and θ_C are the weak coupling constant, electron rest mass, and Cabibo angle; T, and p_e are the kinetic energy and momentum of the electron; M and E_0 are the nuclear matrix element and endpoint energy (under the assumption that $m_\nu = 0$) for the decay; and $F(Z,E)$ is a phase space correction factor accounting for the electric interaction between the β and the nucleus.

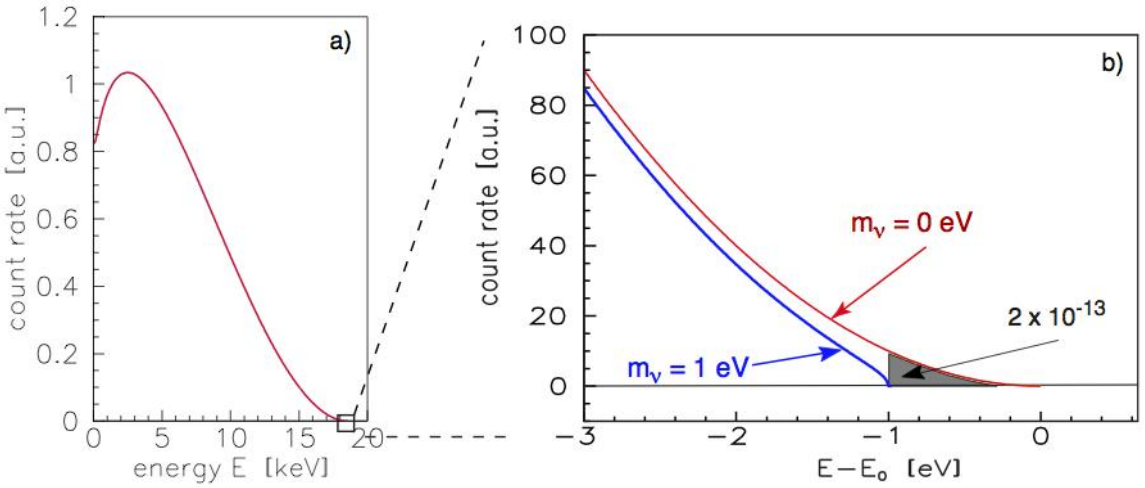


Figure 1.1: β^- energy spectrum for tritium decays. On the left is the full spectrum, the neutrino mass contribution is not visible at this scale. On the right is a zoom of the last 3 eV of the spectrum, shown for the assumption that $m_\beta = 0$eV in red, and $m_\beta = 1$eV in blue. This figure is a reproduction of figure 6 in the KATRIN design report [7] (note that in the annotation in this figure the term m_β in this text is labeled as m_ν). As is labeled, in the (unphysical) $m_\beta = 0$ eV case, only 2 decays in 10^13 populate the 1 eV nearest to the endpoint.

Inspecting equation 1.9, it is clear that only the final term ($\sqrt{(E_0 - T)^2 - m_\beta^2 c^4}$) depends on the effective neutrino mass, defined in equation 1.8. The true spectrum has a distortion at $E_0 - m_i$ for each of the mass states, m_i, as the decay phase space for each neutrino mass state becomes available. Because the difference of the masses of the three mass eigenstates are much smaller than the resolution of any current experiment, these three distortions are not visible. Further, for electron kinetic energies far from E_0 the impact of neutrino mass becomes negligible. For this reason, only the portion of the spectrum nearest the endpoint is relevant to neutrino mass scale determinations. In figure 1.1 the last few eV of the tritium spectrum are shown, demonstrating the slight distortion to the spectrum which results from the non-zero value of the neutrino mass. Because such a small fraction of total decays populate the region of interest, it is necessary to have a relatively large source activity and a detector system which is able to cope with the high rate of lower energy electrons.

Experiment	Mass limit or value (95% confidence) [eV]	Year
Troitsk	< 2.12 (Bayesian) <2.05 (Feldman & Cousins)	2011
Mainz	< 2.3	2005
Mainz	< 2.2	2000
Troitsk	< 2.5	1999
Zürich	< 11.7	1992
Los Alamos	< 9.3	1991
INS, Tokyo	< 13.1	1991
ITEP, Moscow	26^{+6}_{-5}	1987

Table 1.2: Historical neutrino mass limits derived from tritium β decay. All values are taken from [10] except for the two most recent results which come from [11] and [12]. The most recent results from the Troitsk experiment includes limits determined based on both a Bayesian statistical approach, as well as the ubiquitous approach proposed by Feldman and Cousins so both are listed here.

In practice, the specific details of the decay of a particular parent nucleus inform which isotopes are most empirically advantageous. The higher the total energy released in a decay, the smaller the fraction of events which populate the region near the endpoint, and therefore the larger the total source activity required to observe the same number of events within the region of interest. Furthermore, the detector technology being employed determines the source phase and may impose chemical property requirements (sources are often pure fluids or incorporated into crystal structures). Tritium β decay has historically been the process of choice for neutrino mass experiments, though there are currently several experimental efforts focused on electron capture of ^{163}Ho [8, 9]; the rest of this document will focus on tritium only. A summary of the limits on m_β from various tritium-based experiments is presented in table 1.2.

1.2 Modern tritium β decay experiments

Tritium β decay spectroscopy has been the leading method of making direct neutrino mass measurements for several decades. The most recent experimental efforts, the now completed Troitsk and Mainz experiments, as well as the upcoming KATRIN experiment, utilize Magnetic Adiabatic Collimation combined with an Electrostatic (MAC-E) filter technology. More complete reviews of these experiment can be found in several of the citations, especially [13, 7, 10]. The rest of this section will focus on this approach. An alternative experimental approach being pursued by the Project 8 collaboration is discussed in chapter 2.

1.2.1 Basic operation of a magnetic collimation and electrostatic filter spectrometer

In a MAC-E filter experiment, the tritium source is located within a superconducting magnet which produces a strong local field. The β particles released in decay undergo cyclotron motion perpendicular to the magnetic field, and constant drift parallel to the field. The magnetic moment for an electron undergoing cyclotron motion is given by

$$\mu = \frac{T_\perp}{B}, \tag{1.10}$$

where $T_\perp$ is the energy of the motion perpendicular to the magnetic field (B), and μ is constant for adiabatic motion.

In the main spectrometer, the magnetic field magnitude is slowly reduced by many orders of magnitude relative to the source region. Keeping the magnetic moment constant in a lower field region means that the energy in motion perpendicular to the field must also reduce by the same factor. Since the motion is adiabatic, however, the

electron's energy is not lost, so it must be translated into linear motion parallel to the magnetic field. The very dense magnetic field lines in the source region will diverge as the field strength is reduced, requiring that the spectrometer have a cross-section which is correspondingly much larger than that of the source. A larger field difference results in more perfect collimation, but requires a larger diameter spectrometer.

In conjunction with the reduced magnetic field magnitude, an adjustable electric field applies an electrostatic barrier potential opposing the electron motion parallel to the magnetic field. Only those electrons with energy larger than the electrostatic potential will pass through the spectrometer, all others will have their motion stopped and will be ejected back in the direction of the source. A second superconducting magnet, on the downstream side of the spectrometer, increases the magnetic field and contracts the magnetic field lines and corresponding electron trajectories, allowing them to be directed to a solid state detector of a reasonable size.

1.2.2 Experimental reach of a MAC-E system

Experiments based on a MAC-E system must optimize several competing experimental considerations. These ultimately combine to determine the reach of such an experiment. The publications from the relevant collaborations, especially the KATRIN design report [7], provide a more complete evaluation, but a few important considerations are discussed here for the sake of context.

In order to increase statistical sensitivity of an experiment, the usable source activity must be increased. If an electron scatters while traveling out of the source region, its energy will be perturbed effecting the measurement. This limits the source size in the transport direction (a less dense source may be longer, but the total effective activity is not improved). It is possible to enlarge the source perpendicular to the direction of electron transport, but as was described above, the divergence

of magnetic field lines in the spectrometer section means that slight increases in source diameter require substantial increases in the scale of the spectrometer. The scale of the KATRIN experiment, which includes a spectrometer 10 m across and maintained at better than 10^{-11} mbar to avoid internal scattering, is already an impressive engineering feat, and one which is unlikely to be significantly scaled up.

Another major limitation comes from the use of tritium molecules (T2) as a source). There exists a spectrum of molecular rotational and vibrational states which can be excited by the energy released in a β decay. These states provide an unavoidable systematic limit for any experiment using T2 as a source. A very comprehensive review of the molecular final states is detailed in the work of Bodine, Parno, and Robertson [14].

The use of an atomic source would eliminate the limitation of the molecular final states but introduces the obvious technical challenge of designing and operating such a source. Furthermore, the decay of tritium in T2 has a slightly higher endpoint, due to the molecular binding energy, which means that even a very small molecular contamination will have a major impact on any experiment attempting to operate with an atomic source.

Finally, MAC-E experiments operate by counting the total number of events with energy above the barrier potential; stepping the barrier allows the spectral shape to be determined. This method means that the energy of each detected electron is not reconstructed, and so the full statistical power of the events collected is not utilized. Also, the approach is dependent on the system being stable during the barrier potential scan cycle. Any instability in the source strength or spectrometer parameters could introduce a distortion to the observed spectral shape, impacting the neutrino mass determination or limit.

One major advantage of the MAC-E filter is that the electrostatic barrier naturally

solves the challenge of dealing with the larger number of lower energy decays discussed in section 1.1.3. While the overwhelming majority of decays produce electrons with energies well below the region of interest, all of those electrons will be reflected by the spectrometer. Thus the detector and data acquisition systems do not need to be able to accommodate the larger number of events which do not contribute to the measurement.

As was listed in table 1.2, the current best limits on m_β from tritium β decay are between 2.05 eV and 2.3 eV, depending on the experiment and statistical interpretation used. Both the Mainz and Troitsk experiments used MAC-E filter spectrometers, though they employed different source technologies. The KATRIN experiment will also do so, and is currently undergoing final commissioning, with production data collection expected to begin in early 2017. Assuming a null result, KATRIN is projected to have a sensitivity of 0.2 eV at the end of operations, with statistical and systematic uncertainties contributing roughly equally at that point. This order of magnitude improvement in sensitivity is a significant gain. Nevertheless, there remains quite some distance to go before the lower bound is reached. Further, the most recent cosmology bounds, discussed above, prefer a value below even the reach of KATRIN. It is therefore advisable to simultaneously pursue alternative direct measurement technologies which can be extended even further.

Chapter 2

The Project 8 program: Cyclotron Radiation Emission Spectroscopy with an atomic source

The Project 8 collaboration formed shortly after the CRES technique was proposed in 2009 [15] and has since established a phased experimental program for investigating the neutrino mass, centered on the technique. The first phase, which has formally concluded and off of which this document this is based, was focused on demonstration of CRES as a viable method of performing spectroscopy. The second phase, which is the main focus of current efforts within the collaboration, will adapt CRES for measurement of the continuous β spectrum from tritium; it will establish a performance baseline and quantify the gains which must be made for subsequent phases. In the third phase, scalability will be demonstrated with a goal of results comparable to the best current limits of $\sim$2 eV [12, 11]. Finally, phase four will integrate an atomic tritium source, allowing neutrino mass sensitivity beyond the irreducible systematic from the molecular final states present in a conventional source.

2.1 Fundamentals of Cyclotron Radiation Emission Spectroscopy

In this section I will review CRES as a technique. To begin, I will cover the core concepts upon which the technique is based, as well as fundamental scales and relations. After, I will specifically address ways in ways in which the performance of CRES are expected to differ meaningfully from the traditional MAC-E approach.

2.1.1 Electron motion

The core concept of CRES is that any charged particle in a magnetic field will move in a trajectory which circles the magnetic field lines (cyclotron motion), and that the acceleration of this circular motion results in the emission of electromagnetic radiation with precisely the cyclotron frequency. We start by considering an electron moving in a plane perpendicular to the local magnetic field, we can write the central force due to the magnetic field from the Lorentz force law as

$$\vec{F} = q\vec{v} \times \vec{B}. \tag{2.1}$$

Noting from basic mechanics that the central force can also be written in terms of the perpendicular velocity and radius of the motion we have

$$evB = \frac{mv^2}{r} \tag{2.2}$$

or rearranging

$$\frac{v}{r} = \frac{eB}{m}. \tag{2.3}$$

The left side of 2.3 is the cyclotron frequency $\omega_c = 2\pi f_{\text{cyclotron}}$. Here the mass is the relativistic mass, rewriting in terms of the rest mass and combining we have

$$f_{\text{cyclotron}} = \frac{1}{2\pi}\frac{eB}{\gamma m_e} = \frac{1}{2\pi}\frac{eB}{T + m_e}, \tag{2.4}$$

where γ is the usual Lorentz factor, e and m_e are the charge and rest mass of the electron, and T is the electron's kinetic energy. The cyclotron frequency is shown in figure 2.1 over a range of energies, and with particular energies of interest indicated. For completeness, we can invert this to express the parameter of interest (energy) in terms the observable (frequency) as

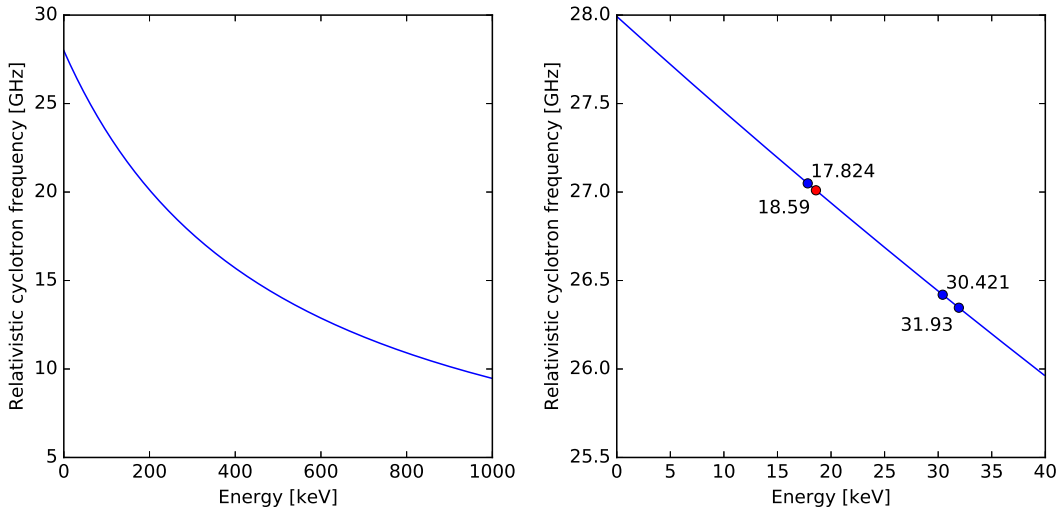
$$T = \frac{eB}{2\pi f_{\text{cyclotron}}} - m_e. \tag{2.5}$$

It is useful to pause here to note that there are only two non-constant parameters, the cyclotron frequency and the magnetic field magnitude. Also, if there is a velocity component parallel to the magnetic field, it does not contribute to this motion (i.e. vanishes from the cross product), but would introduce a $\sin(\theta)$ term to all velocity magnitudes, where theta is defined as the angle between the electron's velocity and the magnetic field and is hereafter referred to as the pitch angle. Note, however, that the relativistic mass correction is based only on the total energy, regardless of pitch angle.

The power radiated by a particle was calculated generally by Liénard, and is reproduced by Jackson [16] in equation 14.26 of that text as

$$P = \frac{2}{3}\frac{e^2}{c}\gamma^6 \left[\left(\dot{\vec{\beta}}\right)^2 - \left(\vec{\beta} \times \dot{\vec{\beta}}\right)^2 \right]. \tag{2.6}$$

Solving for the motion in our particular case gives the result shown in equation 14.31



(a) (b)

Figure 2.1: Plot of the cyclotron frequency as a function of electron energy (equation 2.4) for the case where B =1 T. On the left shows the range from 0 MeV to 1 MeV, while the right zooms in on the region of interest, which covers all conversion electrons from ^{83m}Kr and the tritium endpoint. The point in red indicates the tritium endpoint at 18.59 keV, while the points in blue mark select conversion electron lines from ^{83m}Kr.

of the Jackson text, or in terms of magnetic field, electron's Lorentz factor, pitch angle, and constants we have

$$P = \frac{e^4}{6\pi\epsilon_0 m_e^2 c} B^2 \sin^2(\theta)(\gamma^2 - 1). \tag{2.7}$$

In a reasonable experiment, it is important to be sensitive to a range of pitch angles, since the accepted solid angle is an efficiency factor which directly impacts the required source intensity. Electrons with a velocity component parallel to the magnetic field, in a perfectly uniform magnetic field, would travel with fixed velocity in that direction and soon pass beyond the sensitive region of any antenna system. Since magnetic fields do no work, a non-uniform magnetic field which is able to confine electrons will not disturb their energy. For CRES, we utilize the same concept as in a MAC-E system, with the opposite goal; rather than transport electrons to a region of lower magnetic field, increasing the fraction of their momentum parallel to the field, a region of high field reduces the momentum parallel to the field. If there is a finite difference in field magnitude, there will be a corresponding range a pitch angles which have their motion completely reversed, trapping the electrons in the direction parallel to the magnetic field and establishing a periodic axial motion.

Establishing the axial field gradients results in a corresponding radial gradient. For any electron not perfectly centered on the symmetry axis of the trap, the magnetic field magnitude will differ slightly at different locations in the cyclotron orbit. This slight differences causes the center of the cyclotron motion to slowly orbit the symmetry axis of the trapping field. A schematic of these motions is shown in figure 2.2. An extensive treatment of the motions of trapped electrons is presented in the work of Brown and Gabrielse [17], though it is important to note that there a Penning trap is used and it is the axial motion which is being measured.

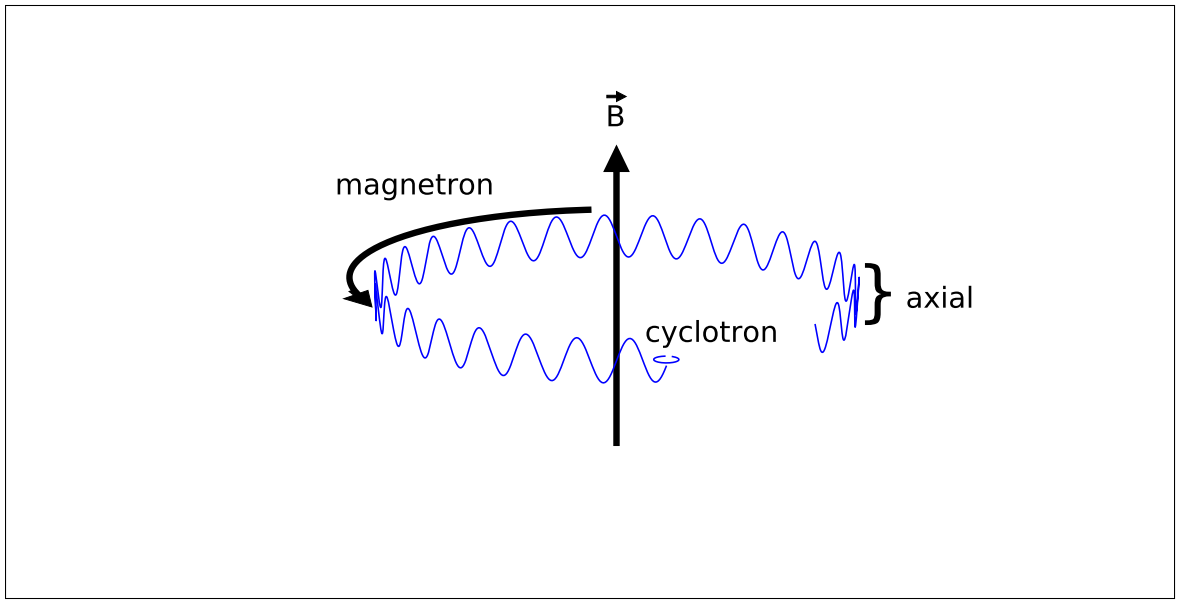


Figure 2.2: Schematic of the natural motions of an electron confined in a magnetic trap. The cyclotron motion is the smallest and is represented here with a single circle, the center of which follows the other motions described. The axial motion is an oscillation in the axial direction and has the same extent as the size of the trap. The magnetron motion is a slow orbit about the trap's axis of symmetry at fixed radius. The scales of the motions are exaggerated here to be visible and are not not meant to provide relative comparisons.

2.1.2 Using CRES for spectroscopy

There are numerous anticipated advantages to CRES in spectroscopic applications. The most motivating characteristic, compared to traditional techniques used for β decay, is the transparency of the source gas to electromagnetic radiation at the cyclotron frequency. This has the important implication that the detection region can be the same as the decay region. Because it is not necessary for electrons to make it out of the source region before their first scatter, the source requirements do not add constraints for how the source is enlarged. It is important to note, however, that magnet or receiver design considerations will still restrict in what ways the source can be effectively enlarged. Furthermore, modern MAC-E systems have nearly 2π acceptance of decay electrons (nearly all electrons emitted in the direction of the spectrometer as opposed to away from it), whereas thus far the CRES acceptance has not yet been demonstrated beyond the few percent level.

Another nice feature of CRES is that it reconstructs the energy of every electron detected, extracting maximal statistical power from events of interest. Further, the method is sensitive to the entire spectrum simultaneously. One major potential source of systematic error in a stepped experiment would be any instability in the system, especially the source intensity, on a timescale similar to the scanning rate. While stability is obviously always desirable, instabilities in a CRES system's source intensity are more likely to result in a change to the overall normalization, rather than a shape distortion.

Because the CRES measurement is made primarily in the frequency domain, there are a number of considerations which are different from those of common particle and nuclear physics measurements. In particular, the uncertainty in energy is not an intrinsic property of the detector response, but is instead primarily a function of the observation time, which determines the width of frequency bins when performing a

Fourier transform. For example, a standard discrete Fourier transform on a time series of 40 μs corresponds to a frequency bin width of 26 kHz; for a signal at 26 GHz that corresponds to a precision of one part in 10^6. This also sets an upper bound on the residual gas pressure in the source, since it is important that the scattering time be large compared to the required observation time.

For experiments which use a MAC-E spectrometer, lower energy electrons, which form the overwhelming majority of the decays, are rejected well before the detector system. In CRES, where the source region and the detector region are the same, there is no such rejection, and so the detector must be able to cope with the large number of signals from electrons outside of the region of interest. Fortunately, this should be achievable using a simple low-pass filter. One of the demonstration goals for phase II of Project 8 will be to demonstrate that this procedure is sufficiently effective.

2.2 An atomic source

While the advances of CRES provide an exciting new approach to spectroscopy, those advances are moot if working with a molecular tritium source. Ultimately, the rotational and vibrational final states of the daughter molecules set an irreducible systematic uncertainty. To overcome this limit, we are investigating ways to produce a high purity source of atomic tritium [18].

If an atomic source is able to be constructed, there are a number of possibilities as to what would become the limiting systematic. The magnetic field in equation 2.4 is the average magnetic field experienced by a particular electron. Imperfect knowledge of the field, whether due to non-uniformity, field drift, or limitations of the available field measurement precision, will set a limit on the achievable resolution. Similarly, the presence of noise, in conjunction with frequency change due to energy

lost to cyclotron radiation will limit how precisely we are able to determine the starting frequency of any individual track. Making a strong statement about what is achievable is difficult because it depends on the details of the analysis procedure used, and this will remain a major topic of research an development throughout all phases of Project 8. Fourier analysis makes a precise statement about bin widths in relation to observation duration, but given that CRES signals are not periodic, but instead chirp, the starting assumption of that analysis is invalid. Continuing to study what is achievable based on the actual noise levels present and specific techniques in use will inform the achievable neutrino mass sensitivity.

Chapter 3

Hardware subsystems of the Project 8 prototype detector

As with any physics experiment of even modest complexity, the Project 8 prototype is really the marriage of many interacting subsystems. Here I will present each subsystem as realized in phase 1 of Project 8, with an attempted focus on design constraints which come from either technical goals or integration with other systems.

3.1 The electron source

The electron source comprises both the radioactive source itself and the source delivery vacuum. For all stages of the experiment, this system will be responsible for providing enough source material to achieve a sufficiently high rate of events, while limiting the residual gas pressure so that signal electrons have a sufficiently long mean free path to be detected and precisely measured.

For the first phase of Project 8, ^{83m}Kr has numerous advantages as a source isotope. First, the metastable state's decays have been well studied [19, 20] and produces a number of internal conversion election lines including one at 17.8 keV which is near the tritium endpoint of 18.59 keV [21] as well as other lines useful for confirming a reliable calibration. Because the lines are discrete and relatively narrow,

it is possible to study the apparent resolution of the technique as well as making simultaneous measurements of signal (in the line's peak) and background (anywhere else). Finally, because of krypton is inert and the half life of ^{83m}Kr is less than two hours, it is much safer and logistically easier to work with than tritium.

The source itself is made of rubidium-83 which has been adsorbed onto zeolite beads, stored in a small vacuum vessel. The source is prepared by professional radiochemists according to a procedure established by the KATRIN collaboration [22]. When the rubidium decays, the krypton daughter atom then diffuses out of the zeolite and through the vacuum system.

The source gas is contained and transported by a custom vacuum system which is schematically shown in figure 3.1. The system is constructed using exclusively bakeable components, including all-metal valves and only all-metal conflat and VCR fittings. A turbomolecular pump provides initial vacuum and can be used to clear the system or while baking. During operation, the turbo pump is valved off and the base pressure is maintained by a pair of non-evaporable getter pumps, which adsorb chemically active species but have no impact on the residual pressure of krypton because it is chemically inactive.

Prior to the first successful operation of the prototype, it was unclear if a lack of signal was a result of poor source density in the detection region or a problem in some other subsystem. This led to a design that included both an inlet and an outlet from the detection region, with both a Passivated Implanted Planar Silicon (PIPS) electron detector and Residual Gas Analyzer (RGA) on the downstream side. Once signals were seen via CRES, and as a result of a leak which was unable to be isolated and fixed, the downstream source section was eliminated and is no longer used. Prior to its elimination, the two-sided system was used to measure temperature dependence of krypton transport (for details, see chapter 4 and specifically figure 4.5

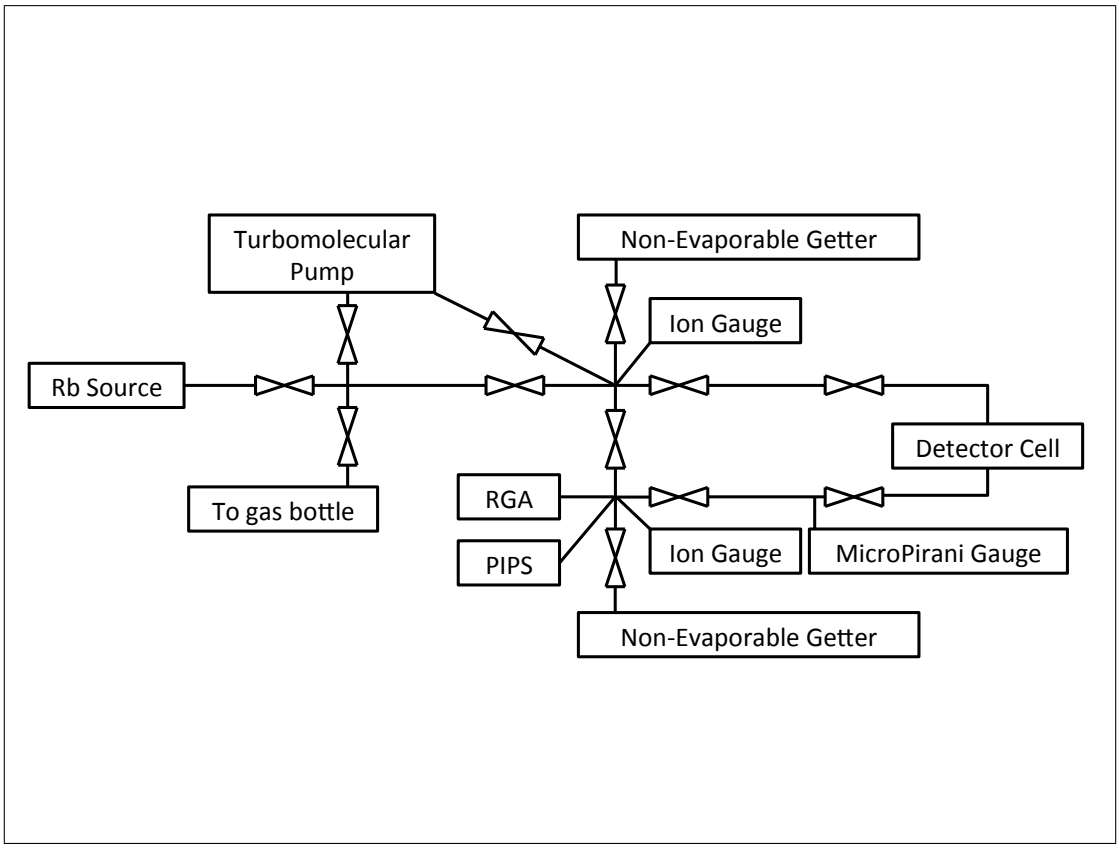


Figure 3.1: Schematic of the source delivery vacuum system for Project 8 phase I. On the upper left is the source itself, where ^{83}Rb which has been adsorbed onto zeolite beads is contained in a stainless steel vial enclosed in a lead shield. The source is plumbed both directly to the turbomolecular pump and to the rest of the system. The valving is such that it is possible to pump and/or isolate the source. Both the source and the turbomolecular pump connect to a six-way cross which also connects to a non-evaporable getter pump, ion gauge, and the detector itself. A second six-way cross, with its own non-evaporable getter and ion gauge, is connected to the other side of the detector volume. The second cross also contains a Passivated Implanted Planar Silicon (PIPS) detector and a Residual Gas Analyzer (RGA) and is connected to the first cross via a gate valve. The gate valve allows the turbo pump to effectively pump the entire system when open, or when closed forces any source gas to pass through the cell before reaching the RGA and PIPS, allowing the content of the cell to be monitored.

from the dissertation of Jared Kofron [23]). The results clearly showed that surface temperatures less than 100 K result in significant freezing of the krypton gas. This provides a lower bound on the acceptable temperature for anything with physical contact with the krypton source gas, including the waveguide cell discussed further in the RF section 3.4. Similar measurements also confirm that operating an ion gauge produces active pumping of krypton gas. However, with a sufficiently strong source the event rate can still be substantial even with the ion gauge operating and turning it off would leave no means by which to monitor the stability of the system pressure.

3.2 Magnetic Field

As was discussed in section 2.1.1, in a CRES system the ambient magnetic field determines the motion of electrons, causing the cyclotron motion and providing radial confinement. In the first phase of Project 8, there are two sets of coils used to generate and shape the magnetic field. A commercial nuclear magnetic resonance spectroscopy magnet consists of a primary superconducting coil, which provides the roughly 1 T background field, and nine trim coils which can be used to fine tune the field shape within the region of interest. In our magnet only eight of the trim coils are operational and their currents were selected to try and produce a uniform magnetic field in the trapping region.

Our earliest attempts to demonstrate CRES were conducted in a cold-bore magnet, where the insert system was in thermal contact with the magnet's cryogenic system. This was a problem because the magnet could only be operated for roughly a week without being refilled, and that only if the insert were installed prior to operation and not removed. Moving to a warm-bore magnet allowed the magnet itself to be commissioned and studied independently from the insert, and allowed the insert

to be commissioned and operated over an extended period of time.

In addition to the NMR system, three normal-conducting coils are wrapped directly around the waveguide cell and are used to produce local field perturbations. There are two current configurations which have been used commonly by Project 8. When a single coil is used to locally reduce the magnetic filed, creating a small region which retains electrons produced within it, it is referred to as a harmonic trap configuration. Alternately, in the configuration referred to as a bathtub trap, a pair of coils can be used to locally increase the magnetic field, confining electrons in the lower field region between the two coils.

Prior to commissioning of the waveguide insert, a standard NMR probe was used to map the magnetic field magnitude within the superconducting magnet. This was done to guide field shimming, but is necessarily modified by the trapping coils and the presense of insert system. To provide further diagnostics, a hall probe is installed on the outside of the insert, and a sample of 2,2-diphenyl-1-picrylhydrazyl (DPPH) is placed inside the waveguide for use as a standard source for electron spin resonance measurements. Because the insert is able to be shifted vertically, the DPPH sample can be raised into the position typically occupied by the trapping region. Despite these handles, the exact magnetic field experienced by an individual electron depends on its initial position, energy, and pitch angle. Our ability to properly account for the effective magnetic field will likely be a primary source of reducible systematic uncertainty for all phases of Project 8.

3.3 Cryogenics

The cryogenic requirements for the first phase of Project 8 are, compared to other systems, relatively straightforward. The low-noise amplifiers in the analog signal path

have a lower noise floor when operated at lower temperature and so it is beneficial to keep them as cold and thermally stable as possible. The same is true of any resistive component in the signal path prior to the first amplifiers. Because the amplifiers provide over 50 dB of gain across the bandwidth of interest, both the signal and noise present at that point are amplified to a level where further contributions, even from room temperature components, are negligible. However, as discussed previously, the krypton source will freeze out around 100 K, setting a competing constraint.

To deal with these requirements the system is cooled using a single stage cryocooler. The amplifiers are thermally connected directly to the cryocooler and allowed to reach their base temperature (typically between 30 K and 40 K). The waveguide cell is connected to the cryocooler via a separate thermal path, which is considerably longer because it is physically further away. A resistive heater on the thermal path to the waveguide cell allows an additional heat load to be introduced and is controlled via a software PID loop to maintain the desired temperature at the cell.

Insulation for the cryogenic system is maintained by an insulation vacuum with primarily polymer gaskets. There are custom components which mate to the bore of the NMR magnet, including one with a sliding seal, which allows the axial position of the insert to be adjusted without breaking the vacuum. As part of the upgrades prior to successful operation, the assembly area, which contains the cryogenic amplifiers, was moved from six-inch conflat hardware into a standard ISO-200 6-way cross. While this change may seem inconsequential, the increased internal volume made the assembly much easier and the ability to open and then reseal connections without disassembly to replace a gasket was critical to simplifying the commissioning process.

3.4 Analog signal path

The preceding sections describe various systems required to deliver a source of electrons to a small region where the ambient magnetic field is such that they are both confined, and produce cyclotron radiation of sufficient power to be detected. In order to turn that into a measurement, we must then collect the radiated power and record enough information to reconstruct the event. Here I will focus only on the custom analog part of that process, the instrument's internal analog and all digital aspects will be discussed with the data acquisition in chapter 5.

A schematic of the analog signal path is shown in figure 3.2. Following the path of the signal, an electron produces power at its cyclotron frequency which couples to the propagating modes of the WR-42 waveguide in the cell region. Half of that power travels down the waveguide and is reflected by a short before following the upward propagating signal. A long transport section gets the signal out of the high-field and physically constricted space of the NMR magnet's bore to a region with more space for components. A ninety degree twist of the WR-42 is required for mechanical assembly and a WR-42 to WR-28 transition allows the signal to be adapted to the input of a pair of low noise amplifiers (Low Noise Factory model LNF-LNC22_40WA [24]) which each provide frequency-dependent gain of 25 dB to 30 dB across the their operational bandwidth. The pair of amplifiers provide over 50 dB of amplification, and boosts both the signal and noise floor power to significantly above the noise contribution of the downstream components which operate at room temperature, effectively fixing the signal to noise ratio. The use of these amplifiers is another major hardware upgrade which led directly to successful operation. The cryogenic amplifiers used in earlier attempts exhibited unstable gain without other symptoms, making them difficult to operator and monitor.

After the second low noise amplifier, the signal is passed into a coaxial cable and traditional connectorized components are able to be used. There are two heterodyne mixing stages, the first of which uses a fixed local oscillator frequency of 24.2 GHz to mix the signals down to less than 2 GHz. At this frequency, losses in coaxial cabling are not significant and both components and instruments are much more affordable. After the first mixing stage the signal is split, with one branch entering a second mixing stage and the other connected to a real-time spectrum analyzer (RSA) which has its own internal mixing stage and serves as a triggered digitization system. The second mixing stage uses a tunable external local oscillator which allows selection of the particular frequency band that will be passed to the baseband streaming digitization system.

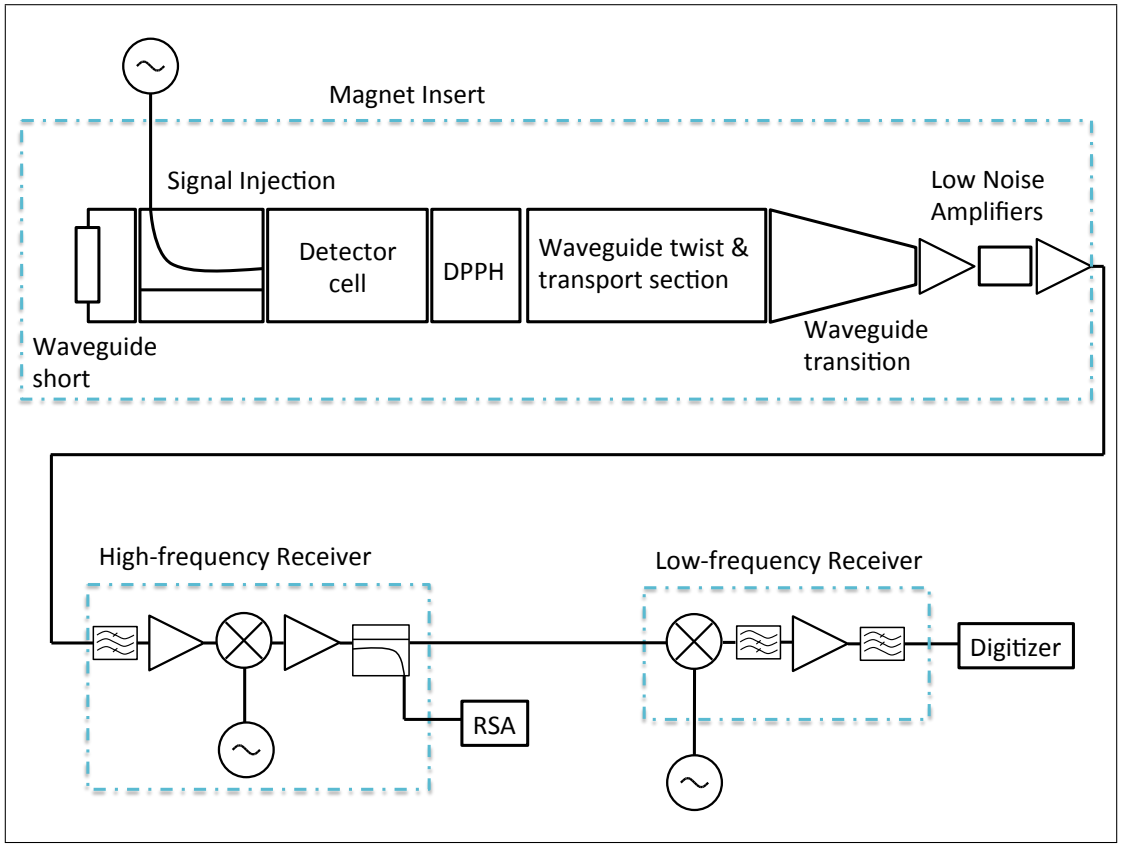


Figure 3.2: Schematic of the custom radio frequency signal path components. The signal path is broken down into three sections, indicated by broken blue boxes. The Magnet Insert includes everything which is inside the magnet and cryogenic insulation vacuum and is constructed primarily of custom WR-42 waveguide components. Below the trapping cell, a coupler allows signals to be injected into the system for diagnostics and a short reflects down-going signals back towards the receiver. Above the cell is a waveguide section with DPPH sample, then transport waveguide which brings the signal out of the center of the magnet to the cryogenic amplifiers. The High-frequency Receiver receives the signal after it is fed out of the insulation vacuum; it mixes all input signals down by a fixed 24.2 GHz. From the high-frequency stage, the signal is either passed directly to the RSA or it is mixed down further in the Low-frequency Receiver, which uses a variable local oscillator to select the signal band which is presented to the digitizer.

Chapter 4

Dripline: a protocol for hardware control and operational coordination

For any scientific endeavor it is important to be able to monitor and control experimental conditions. While there do exist many commercial products capable of providing a centralized slow control, monitoring, and orchestration system, none provided the full set of features and flexibility desired for Project 8. As a result, we developed the dripline protocol, which defines interactions and standard behaviors upon which a control system is developed. The decision to implement a network-based interface first ensures that it is possible to implement new components in new languages while maintaining system compatibility. The primary implementation of dripline was done in the python programming language, which provides the majority of the control systems used for ongoing stages of the experiment. In this chapter I will detail the major design features and decisions that led to the version of dripline used in phase I.

4.1 Starting with a protocol

One of the primary goals when developing dripline was that components (by component, here, I mean the software providing user control and access to a particular instrument or device) be deployable as independent executables. There are several benefits from this, but the two largest are that if there is a problem in one component which causes a crash or necessitates a shutdown, others able to continue uninterrupted, and that new components are able to be developed and deployed without any need to disrupt the already running system components. Further, new components can be distributed across servers and implemented in different languages. These are critical in the early phases of Project 8, because development and commissioning are much longer than periods of stable operation. Dripline achieves this goal by defining a communication protocol rather than being primarily a particular program. In this way, any program which conforms to the dripline standard is a valid component and can be incorporated, regardless of operating system, programming language, or any other particular detail. Dripline has currently been implemented in python, golang, and C++, with python being the most feature rich and the version I'll focus on here whenever not being fully generic.

The design of dripline was heavily informed by REpresentational State Transfer (REST) architecture. In particular, dripline defines a uniform interface where resources do not store the state of other objects and any details of implementation are opaquely hidden behind the interface, in a layered fashion. Because of this REST-based structure, it is possible to make changes or upgrades to the implementation of individual components in a way which is transparent to other components.

Finally, dripline itself defines a set of valid interactions, but lives on top of Advanced Message Queuing Protocol (AMQP), an open standard for sending and re-

ceiving messages [25]. An AMQP server is responsible for routing messages between components and dripline defines the structure and content of those messages. In principle, the standards defined in dripline could be be combined with any message encoding format and delivery system. In practice we use exclusively the rabbitMQ (see [26]) implementation of AMQP to deliver messages, and json (see [27]) to encode them. The decision to use rabbitMQ is motivated by its availability from package mangers on most *nix systems, demonstrated stability, and availability of libraries in a wide range of programming languages. Encoding with json was selected because of the easy availability of encoding/decoding libraries in most modern languages, and because it is reasonably human readable, easing debugging procedures.

4.2 Fundamental structure of the dripline protocol

At its heart, dripline defines a notion of a valid message. The restrictions on these messages in turn defines a set of three valid interactions: a request, which instructs a specific actor to send a reply; a reply sent in response to a request; and an alert sent to any program which chooses to listen. Every message has an integer-valued field *msgtype* which indicates which sort of message it is (3, 2, and 4 respectively); these are each treated in the following subsections. The available fields within a message are summarized in table 4.1.

All information transmitted is part of the message itself, with the exception that a message is sent with a routing key as defined by the AMQP standard. AMQP specifies that a routing key is a period delimited collection of strings, we further specify that the first element always specifies a *destination* or *target*, which determines which service, or services in the case of an Alert, are to receive the message. All elements after the first period is referred to here as a *routing-key specifier* (RKS) and is used

Field	Type	Required by	Values
msgtype	integer	all	2 (Reply), 3 (Request), 4 (Alert)
msgop	integer	Requests	0 (Set), 1 (Get), 7 (send), 8 (run), 9 (command)
timestamp	string	all	Following RFC3339 format [28]
lockout_key	string	Requests	16 hexidecimal digits
sender_info.package	string	all	Software package used to send the message
sender_info.exe	string	all	Full path of the executable used to send the message
sender_info.hostname	string	all	Name of the host computer where the message originated
sender_info.username	string	all	Name of system user, on the originating host, owning the process which sent the message
payload	any	none	The content of the message
retcode	integer	Reply	Execution/exception status, see table 4.2
return_msg	string	Reply	Human-readable explanation of the return code

Table 4.1: Summary of fields defined for compliant dripline messages. The message itself is a dictionary, and the fields are the keys in the top level dictionary.

by the receiving service to determine how the rest of the message is processed. The most obvious example of this is that the RKS can be used in either a "Get" or "Set" request to specify the attribute of the target which should be returned or modified.

4.2.1 Alert Messages

Alert messages are sent on a dedicated AMQP exchange, also named alerts, and are notable in that they are used for one-directional information flow. Any process which generates an alert message has no way of knowing if there are any other services which will receive it or what they might do with the information contained. The AMQP routing key of an alert message follows a set of conventions which allow services to bind on all messages of a particular category and to parse information

from the RKS. There are currently two established routing key prefixes in use.

First, routing keys of the form "sensor_value.*sensor_name*", where *sensor_name* is the string name of some sensor or other object which has values which are of interest to track over time. The payload of such a message is expected to contain at least one of the keys "value_raw" or "memo" and may optionally include "value_cal" as well. Currently the main use of these messages is providing data to a service which stores the values into a postgreSQL database and providing current state information to the input of control loops (like the software PID temperature controller) and a system health monitor.

The other class of alerts have routing keys of the form "status_message.*severity*.*origin*" and provide information about the current status of a particular object or service. These are used to provide information to on-call users, with the current implementation based on the web collaboration service Slack [29]. For the current implementation, the values for *severity* match the available Slack channels ("#p8_notices", "#p8_alerts", and "#p8_critical") and *origin* names the program that created the status, though the dripline standard does not specify which channels exist. The payload for such messages is a string, and a dedicated service processes these messages and posts them to their respective Slack channels, attempting to use the proper Slack bot token corresponding to the value of *origin* in the RKS.

4.2.2 Request Messages

A request message is one which causes its recipient to send back a reply type message. The message has a required field *msgop*, which indicates the operation type which is being requested. Upon receiving a request, an endpoint will try to take the requested action and then, regardless of the result or any exceptions raised, send a reply (the reply generally indicates the result and if there was a problem as discussed

in section 4.2.3).

The valid types of a request are Set, Get, Send, Run, and Command (encoded with integers 0, 1, 7, 8, and 9 respectively), each of which has a few very specific properties as follows:

1. Set: Assign a value, provided in the message payload, to either the endpoint itself (if no RKS was provided), or to a particular attribute of the endpoint (if routing key has an RKS).

2. Get: Return the value of the endpoint (or if an attribute is specified in the RKS, the value of that attribute). There is no payload content.

3. Send: The endpoint relays the payload verbatim via its internal send method. This is commonly used with *Provider* classes for debugging because it provides a way to manually specify the exact command to send to an instrument.

4. Run: This is conceptually similar to a high-level trigger. The expectation is that the target starts some long-lived process but will return quickly, indicating that it has started doing whatever run means in the particular case. This is generally used for data acquisition systems, but could in principle be expanded.

5. Command: Generically call available methods of the service not covered by the other operations. The method is named in the RKS. The payload may contain any arguments (positional or keyword) for the method. In principle all other operations could be achieved via this one but are separated due to their specific and common usage (and for convenience in command line utilities).

4.2.3 Reply Messages

A reply message is sent by a service in response to receiving a request. The payload of the message is used to include any direct response to the request which was made (often taking the form of raw and calibrated values of some requested parameter). A reply message also contains two status fields, *retcode* which contains an integer value indicating either successful processing of the request, or a particular class of error has occurred. The *return_msg* is a string and may contains human-readable clarifying information about any return code indicating that there was a problem. Currently defined return codes are summarized in table 4.2.

It is expected that any object submitting a request can and should check the value of the *retcode* of the reply message to determine if there has been an error and possibly apply logic for how to proceed. It is incorrect usage to place this information into, and parse it from, a payload (which is specifically for results in a format consistent with successful operation). In some cases, there will be an error, or more commonly a warning, coexisting with a valid payload; this distinction in storing the information enables consistent behavior when dealing with this case.

4.3 Fundamental behaviors of the dripline protocol

In addition to the structural standards described above, dripline also defines some core behaviors that all components must implement. These basic behaviors are important for ensuring uniformity and allowing complex systems with many services to coordinate effectively. The application-specific behaviors of each service are then built on top of this minimal set of actions.

Return Code	Description
0	Success
1	No action taken warning
2-99	Unassigned, non-error warnings
100	Generic AMQP-related error
101	AMQP connection error
102	AMQP routing key error
103-199	Unallocated AMQP errors
200	Generic hardware-related error
201	Hardware connection error
202	Hardware no response error
203-299	Unallocated hardware errors
300	Generic dripline error
301	No message encoding error
302	Decoding failed error
303	Payload-related error
304	Value error
305	Timeout
306	Method not supported
307	Access denied
308	Invalid key
309	Deprecated feature
310-399	Unallocated dripline errors
400	Generic database error
401-998	Unallocated
999	Unhandled core-language or dependency exceptions

Table 4.2: Valid error codes for reply messages under the dripline standard. Any value less than 100 indicates success, though warning information may still be passed. Larger values are grouped by categories, in blocks of 100, to allow organization as new errors are added. The errors which are divisible by 100 are intended to cover fairly generic classes of problems, with intermediate values being more specific cases.

4.3.1 Resource locking

All services must implement a lockout system to restrict access for certain types of requests (specifically those that make a state change to the service). The lockout is intended to prevent unexpected system-state changes from occurring during a measurement and to prevent multiple users or processes from interfering with each other. This engineering control is intended to compliment administrative controls which address the same interference issue, and should provide useful feedback to users; it is not intended to provide security against unauthorized access.

The key used for lockout is sixteen bytes long. When included in a message it is stored as the value of the *lockout_key* field of a request message, encoded as a string of thirty-two hexidecimal characters. For messages, the format should either be continuous, or in four hyphen separated groups with eight, four, four, and sixteen characters respectively.

There are various conditions under which a service will respond to requests. If the service is not currently locked, it will respond to all incoming request. If the service is locked, but receives the appropriate key, it will respond to the request. Finally, if the service is locked it will always respond to certain requests, including those which have no impact on state and a few special cases. In particular, it is always possible to send an OP_GET request to either an endpoint or one of its attributes since this action is read only. The ping command is also passive and will receive a response regardless of lockout status. A special case is made for the set_condition command, which does change the system state, but which is intended as an override by design.

In order to lock a service, an OP_CMD request is sent with routing key specifier of lock (this can either be a broadcast or a direct request). The request to lock may specify a key in the message's *lockout_key* field, or may leave the field blank and the service will generate the key. Once locked, any request may include the *lockout_key*

field, with valid key value, to be processed despite the lock. Unlocking is simply a special case of an OP_CMD, but it also supports a forced mode, if the message payload includes a key *force* assigned the boolean value True, to allow the system to be unlocked if the key is lost or a guaranteed unlock of all services if sent to the *broadcast* target.

4.3.2 Broadcast commands

In most cases, dripline requests are addressed to a particular endpoint, which sends exactly one reply to the original requester. However, the standard also requires all services to respond to requests sent to the special name *broadcast*. In this case, the number and order of the replies is unknown, and the behavior is only used in a few special cases.

The ping command is intended for discovering which services are currently running, and also for confirming that a running service is able to properly respond to requests. Any service receiving a ping applies no logic and simply sends an empty reply to the requester.

The lock and unlock commands were described above, but at times it may be desirable to lock everything, rather than explicitly locking each resource. Care should be taken in this case, since if any resources are already locked, it could be easy to accidentally miss an error in replies other than the first. Similarly, doing a forced unlock will not only unlock the resources that were locked with a particular key, but any others which may have been locked by another user.

The set_condition command is an upcoming feature which service as a quick way to achieve a particular state across the entire experiment. It is being added specifically for the case where one service detects a problem that should halt data taking. It is implemented as a broadcast so that every service which could detect such a problem

does not need to know which other services are involved in taking data. It must be stressed that this method is intended to be heavy handed and is suitable for aborting an operation, but should be disfavored for procedures that care to ensure coordination between services or that every component has reached the desired state.

4.4 Abstractions of the python implementation of dripline

While defining dripline as a communication standard, rather than a library or program, provides incredible flexibility when extending it, that exercise does not result in any usable code. Dripline was first implemented in the python programming language (version 2.7), using the pika library to provide an API for AMQP interactions. Here I review the major abstractions which have been defined and how the major requirements of the dripline are applied. I also discuss some of the major design decisions and conventions which have been established.

Message The Message class abstracts the dripline message object in a python dictionary-derived subclass. The implementation supports encoding and decoding the json data used in the actual AMQP messages and provides straightforward access to the expected fields. The class also automatically enforces required fields and formatting.

Endpoint An Endpoint is any object which is able to receive a dripline Request Message. Endpoints must always have a unique name so that addressing is unique. The base class implements a core method for processing any request, including catching exceptions and converting them to a Reply Message with proper return code, but

this extending this behavior in derived classes is a standard way of implementing new control system features. Standard logic or actions taken by instruments are often each an Endpoint; the Provider and Service abstractions are also specific cases of Endpoint. It can be useful to think of an Endpoint as the abstraction which relates user interactions to instrument interactions. A user makes use of dripline requests to communicate with an Endpoint via its Service, and the Endpoint responds by interacting with the instrument via its Provider.

Provider The Provider abstraction is specific to the python implementation of dripline, and represents a translation or access layer (it is also a subclass of Endpoint). A Provider generally has some number of Endpoints associated with it and which make use of it. Providers are generally used to convert dripline messages and commands to communication with an instrument, over a network socket or GPIB bus; hence, they provide access to some resource. By abstracting this, generic Endpoint classes can be used for a variety of instruments which are programmable via a range of device interfaces. In many cases, a Provider can be thought of as representing a particular instrument or piece of hardware, while the associated Endpoints represent a specific action or interaction with the device.

Services Any continuously running program which is part of a dripline mesh is called a Service. A Service is itself an Endpoint since, for example, it must respond to ping commands, but it may contain any number of additional Endpoints which are not Services themselves (often they are Providers, which themselves have associated Endpoints). The Service is responsible for establishing queues and bindings with the AMQP broker, and for processing messages which are ultimately delivered to, and fulfilled by, those Endpoints.

Mesh A general term which refers to a particular AMQP broker and all of the Services bound to it. An experiment or project would normally run with a single mesh so that all components can communicate with each other and be have their operation coordinated. It is common for a developer to establish a local mesh while implementing new features or testing bug fixes to avoid any potential conflicts with the production system. The flexible structure of the dripline standard makes it possible to halt a Service on one mesh and start it on another without disrupting other Services (assuming they don't try to interact with it).

Chapter 5

Data acquisition and signal processing

The early stages of Project 8 have utilized several data acquisition systems and have tried many others. Fundamentally, these fall into two categories: streaming systems, which take data continuously once started, without any confirmation of signal content; and triggered systems, which make attempt to reduce the volume of raw data by only recording time series during which signals are present. The obvious trade-off being that a streaming system produces a high volume of data but a triggered system can miss events or otherwise be biased by the details of the trigger being used.

5.1 Streaming data acquisition for phase 1

In the initial successful running of the first phase of Project 8, data were collected using a Signatec PX1500 digitzer. Upon being processed through the full analog signal path (see section 3.4), signals were fed into the digitizer and sampled at 250 MS/s with an 8 bit analog to digital converter. The acquisition software was custom written by the collaboration and utilized a circular buffer of shared memory which allowed continuous collection of data without breaking phase coherence between samples. The software package, mantis, is no longer under active development or maintenance but is

freely available online [30]. The data were recorded in ten second chunks to prevent the resulting data files from becoming too large. Ten seconds of data collected as eight-bit samples at 250 MS/s comes to 2.5 GB per file.

The data collected using this system were saved using a custom data format named egg2. This format can be read and written using the monarch library, which is maintained by the collaboration [31].

The PX1500 became inoperable during phase 1 of Project 8 and other streaming digitizer systems were explored. All data taken since that time has utilized the RSA system discussed next. It was worth noting that while many manufacturers of high-speed digitizers claim to support linux, in most cases the linux drivers require extremely old kernel versions and hardware architectures, making them effectively unusable. The Windows versions of drivers were generally much more current, but developing software and achieving sufficient data transfer rates on a Windows host proved untenable.

5.2 Triggered data acquisition with the Real-time Spectrum Analyzer

Project 8 has used several different Tektronix Real-time Spectrum Analyzers (RSAs) for data collection. A Tektronix RSA6120A, owned by collaborators at PNNL, was used for initial studies, while multiple RSA5100B series units, both rented and eventually purchased, were used to collect part of the data in 2014, published in 2015 [32], as well as all triggered data since that time. The Textronix RSA utilizes both analog and digital techniques to process the input signals; these are treated separately in the following subsections and follow closely the description provided in the primer document "Fundamentals of Real-Time Spectrum Analysis" produced by Tektronix

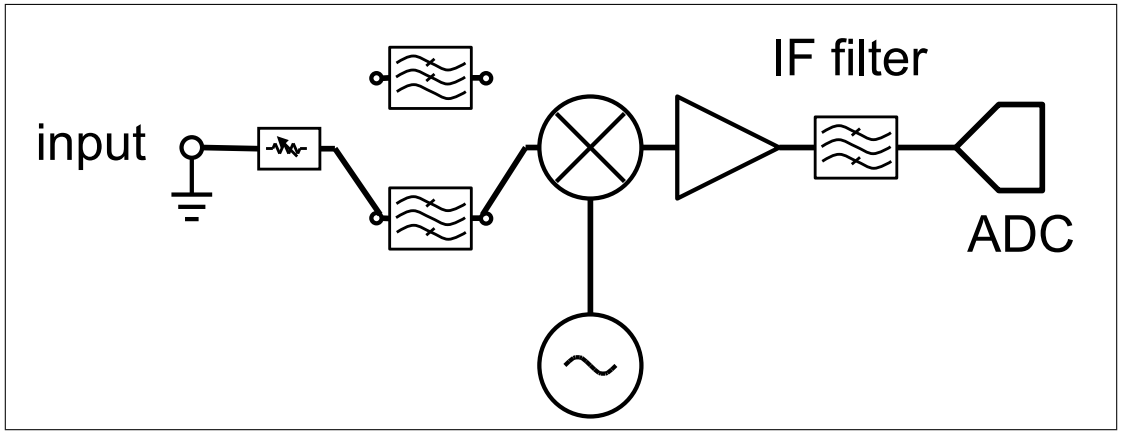


Figure 5.1: Simplified block diagram of the analog signal path of the Tektronix RSA. This figure is based on Figure 1-2,c in the Tektronix primer [33]

[33]. There is example worked out in some detail in appendix A.

5.2.1 Analog processing in the RSA

The analog signal path within the RSA is relatively straightforward, and looks very similar to the second mixing stage employed by the streaming acquisition system used by Project 8 (see figure 5.1 for a block diagram of the RSA's analog components and section 3.4 for a discussion of the custom mixing stages). The signal passes through a variable attenuator which, in conjunction with a downstream amplifier, allows the gain of the signal path to be configured based on the input power levels. Following the attenuator, a filtering stage suppresses power from outside the instrument's input bandwidth. A mixing stage with tunable local oscillator then mixes the input signal band of interest into the near-baseband region for which the digital receiver is sensitive. The output of the mixer is then amplified and passed through an IF filter which suppresses all power outside of the signal bandwidth. After the IF filter, the signals are passed to the analog to digital converter, where the digital processing discussed in the following section begins.

5.2.2 Digital processing in the RSA

The digital signal path in the RSA is an especially prominent source of confusion. At the end of the analog signal path, the signal is sampled by an ADC. According to the primer, the instrument's specifications for performance and stability exceed those of the analog components. This is achieved by applying various corrections to the digitizer output, the details of which are not provided.

It is generally not practical to mix an analog signal all the way to DC, especially when a wide bandwidth is desired. Instead, a commonly employed technique is physically mix the signal into a higher Nyquist zone, between half the sampling frequency and the sampling frequency, and allow the signal to be aliased by the digitization process. The discrete time samples are multiplied by both numerically generated sine and cosine signals at precisely the desired center frequency, allowing selection of the desired output signal bandwidth. The result of this digital mixing is two outputs, the coefficients of each of the sine and cosine functions, generally referred to as I and Q components (i.e. the in-phase and quadrature components).

Both the I and Q signal are even functions in frequency space, and can be decimated by two. An anti-aliasing filter is required prior to decimation to preserve signal fidelity. This is because decimation, like discrete sampling, can alias signals across the effective Nyquist frequency. The decimated I and Q samples, taken together, have the same number of data points as the original signal and contain the same information, but with the selected bandwidth now centered at zero frequency. Note that because the signal is centered at zero, it would also be possible to decimate by more than a factor of two, assuming that the bandwidth of interest is less than half of the original sampling frequency. This would allow the total volume of data saved and stored to be further reduced; the RSA's implementation does not do this, the roll-off of the internal filters are visible in the output bandwidth and all of the originally recorded

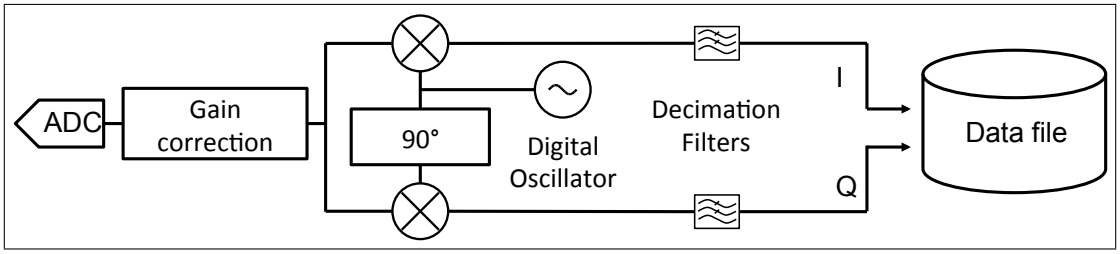


Figure 5.2: Simplified block diagram of the digital signal path of the Tektronix RSA. This figure is an approximation of Figure 2-4 in [33].

information is retained. A schematic block diagram of the digital signal processing is shown in figure 5.2.

For purposes of triggering, the RSA performs a discrete Fourier transform of the baseband IQ samples and compares this to a frequency-dependent threshold called the trigger mask. As is typical, if the configured trigger conditions are met, the instrument writes data, including some prescribed amount of buffered data from prior to the trigger event and some fixed total duration. While the trigger is evaluated using the FFT of fixed-width time intervals, the data stored can cover a much longer time period, and are recorded as the original I and Q time-domain values. This means that while triggering effects do dictate what is stored, the FFT details do not impact the quality of that data for use in other processing schemes.

5.2.3 Frequency-mask trigger

A powerful feature of the RSA is its capability to perform frequency-mask triggering. In figure 5.2 the schematic is simplified, showing only the path taken to disk storage. In actuality the output of the decimation filter is split, with one branch is passed to the input of a discrete Fourier transform, while the other is cached for recording only if a trigger occurs. The instrument is configured with a frequency-dependent threshold which can be arbitrarily defined. For all Project 8 applications, the threshold as configured as a fixed power offset above the RMS power level in each

bin. This is selected to account for the frequency-dependence of the receiver's gain. The FFT windows used are approximately 80 μs wide, corresponding to a frequency resolution of 12.5 kHz.

Much like a standard time-domain trigger compares each sample for a trigger condition, the RSA checks every frequency bin computed for each time window. There are available configurations for caching a fixed pre-trigger time and the total duration of the data stored. There are also configurations for a requirement that the trigger remain un-violated for some amount of time before it rearms, though the sophistication of these options is limited (this is part of the motivation for moving to the custom acquisition system discussed in the next section). It is important to stress that while the triggering is performed in the frequency domain, the RSA maintains an internal buffer of the digitally down-converted data in time domain. The values are continuous samples which maintain phase coherence of observed signals and are not distorted by the windowing used by the trigger.

5.3 Future hybrid system based on ROACH-2

The RSA is incredibly useful in phases one and two of Project 8, where the data are all single-channel, because it gives feedback in real time. Especially after major hardware changes, it is incredibly valuable to be able to confirm that signals are still present with immediate feedback, rather than having to wait hours or days to process data. There are two major failings, however, for this system. First, it is only usable if the signal-to-noise ratio in a single analog channel is large enough for it to trigger. Future phases of Project 8 will need to instrument a large physical volume, most likely using an array of antennas, with very little signal power in them individually. Second, the only available triggering schemes are those included in the proprietary

software. Moving forward we will want to develop triggering algorithms optimized to the known characteristics of a CRES signal.

The approach currently being pursued is based on the ROACH-2 platform. This architecture features a Xilinx Virtex-6 FPGA in a system which supports a wide range of interchangeable ADC input modules. We plan to develop a modular software system, which will enable a staged deployment and testing of new versions of components individually. Under the current scheme, the FPGA will perform an FFT on all input data, then ship both time and frequency domain data over a network to a processing computer. At first, the data will just be recorded in a streaming mode, essentially the same as the earliest data collect in 2014. Triggering algorithms can be tested against real data offline, and those with satisfactory performance implemented as available acquisition modes.

The first two acquisition modes desired are the simple streaming and a frequency-mask trigger that mimics the behavior of the RSA. Once those previously-used modes are demonstrated, various alternatives can be pursued. On promising example is a modified frequency-mask technique that uses a lower threshold, but requires not only that the trigger be violated for multiple time windows, but within a narrow bandwidth, consistent with a chirping electron signal. The reduced threshold of such a scheme should provide sensitivity to a wider range of electron events. Being able to operate in multiple modes with the same hardware will enable more thorough optimization of the trigger.

To move from the second to the third phase of Project 8, it will be necessary to perform digital beam forming calculations. A functional ROACH-based acquisition system could exchange the input ADC module for one with a higher channel count and otherwise proceed in its streaming mode, with all additional processing done offline or in CPU. It is also possible, and likely eventually necessary, for the phased

channel combinations to be calculated on the FPGA itself.

While the ROACH system is still being actively developed, it was selected based upon considerable time comparing this solution to other, more commercial, options. In particular, the possibility of a PXIe-based system from National Instruments was investigated and some hardware acquired. Such a system would have a similar architecture (ADC modules attached to FPGA systems running custom firmware), but using a commercial platform rather than an open one. The result of initial evaluations was that while the PXIe-based system has an advantage of commercial support, it is likely much more expensive and would tie us to development and operation based on labview, in a Windows environment, constraints which are seen as unacceptable. The ROACH-based approach is therefore being pursued as the primary solution.

Chapter 6

Results of the first successful operation of the Project 8 prototype

For the first phase of Project 8, the main goal was simply to demonstrate that CRES is possible. There were several initial designs which we attempted to operate, but which ultimately proved unsuccessful. A number of significant hardware changes were made at the end of 2013 and in the early months of 2014. These enabled our first successful operation in the summer of 2014, with results published in 2015 [32].

6.1 Major hardware changes

There were three major hardware changes which led directly to successful operation. While the relevant hardware subsystems have been discussed earlier (see chapter 3), it is worth highlighting each of those changes briefly here.

First, the cryogenic amplifiers were replaced. At the early stages, a pair of custom amplifiers were provided by the National Radio Astronomy Observatory for use. These were non-commercial amplifiers which we eventually determined were unstable during continuous operation. The symptoms of the failure mode were not obvious and so it is possible that otherwise functional conditions were achieved much earlier.

We then identified and purchased a commercially available amplifier from Low Noise Factory. These amplifiers have operated without any issue since they were purchased.

Second, the experiment was moved from an custom cold bore superconducting magnet to commercial warm bore magnet made for nuclear magnetic resonance measurements. In principle either system should have been usable, but the cold-bore system was only operable for a couple of weeks at a time whereas the warm-bore magnet is able to be operated continuously. This allowed us to make careful measurements of the magnetic field within. Furthermore, it meant that we could decouple commissioning the magnet itself from the challenge of commissioning the CRES insert and control software.

Finally, the insulation vacuum system, which was previously contained in six inch conflat hardware, was converted to ISO-200. The change significantly increased the amount of work space during construction. More important, because the system is based on reusable rubber gaskets, breaking a seal no longer requires disassembly to replace the gasket. This and the prior change may seem like minor matters of convenience, but it makes checking components a matter of roughly a day, rather than a week or more, and was probably necessary for the system to ever be successfully commissioned.

6.2 Data analysis framework

The Project 8 collaboration has developed a sophisticated framework for processing CRES data. The C++ package, named katydid, was developed from the ground up as a flexible and modular system and is freely available online [34]. Classes which implement procedures to be carried out on data are called processors, and they can be strung together however they are needed, assuming input and output types are

properly matched. The processors being used, their order, and the configuration of each, is specified in a configuration file which is processed at runtime. The obvious advantage to this organizational scheme is that it is often not necessary to recompile the code in order to change the logic. Furthermore, the modular nature allows individual components to be added or replaced without impacting other components.

Regardless of data acquisition system used, our analyses have always proceeded through the same few conceptual steps. First, the data are transformed from voltage values in the time domain, into power values in the frequency domain. The details of that transformation account for the differences between the data acquisition systems used, and specifically on the format of the stored data, after that stage all data should be on equal footing. Next, various techniques are used to identify time-frequency bins of interest. This generally includes a simple threshold, but may also including techniques based on the density or topology of high-power bins. Finally, the bins of interest are combined into more abstract objects such as tracks and events, which represent a portion or the entirety of an electron's signal respectively.

6.2.1 Katydid processors for the 2014 analysis

In this section I will review the katydid processors utilized for processing the data collected in 2014. These form the first stages for processing the 2015 data, which will be discussed further in the next chapter.

Reading data files

As discussed previously, Project 8 has made use of a number of different data acquisition systems, and each stores data in a very different format. At the time of this writing, there are three forms: early phase 1 streaming data are stored in what we call "egg2" format and include raw ADC values, encoded using protocol buffers

[35]; triggered data collected using the RSA are stored as IQ time series, which are a result of processing, stored in matlab format; and newer streaming data collected with the ROACH2 platform currently under development will use the newer "egg3" format, which encodes values using hdf5 [36].

The data reader is responsible for parsing any header information available from the datafile into a header object, which is passed along to the downstream processors. This header information can be used by downstream processors to configure how received data are interpreted.

Once the header is processed, the time series itself is passed to downstream processors in time windows of fixed duration. The length of the time windows, as well as any overlap between windows, is configurable, allowing sliding windows to be used if desired.

Conversion to power spectral density

The conversion of a time series to power spectral density proceeds in two steps, each in its own processor. A discrete Fourier transform changes the basis to frequency domain, followed by a unit conversion from the input ADC values to power and applying proper bin normalization.

Fourier transform: Time domain data, real or complex, are convert to frequency domain by use of a discrete Fourier transform as implemented in the ubiquitous FFTW library [37]. There is support for transforming both real and complex time series and the processor also takes care of computing frequency bin values and reordering the bins. After this processor, the technical differences between different acquisition systems should have been accounted for, and downstream processors can treat all data in the same way (though practical differences such as noise levels, data

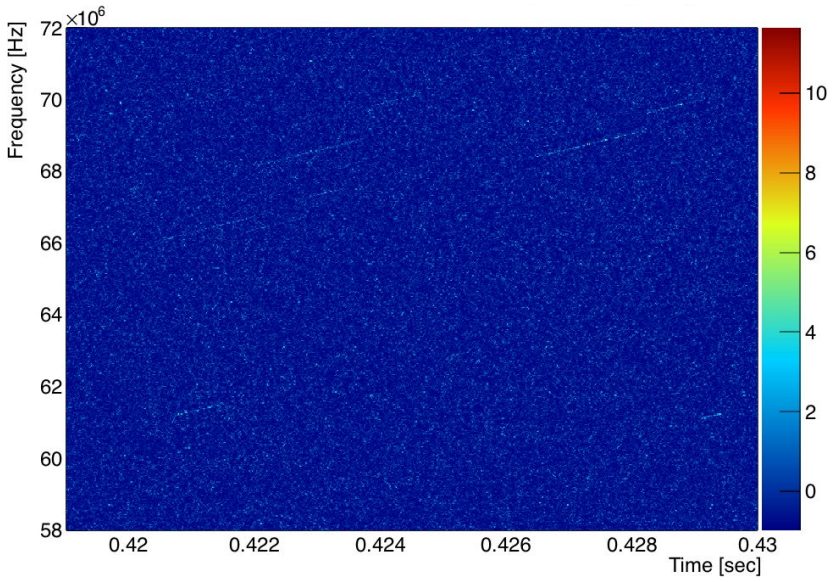


Figure 6.1: A representative spectrogram after power conversion. The color axis is power in units of decibels relative to noise per frequency bin.

precision, or optimal analysis parameters may still differ). The processor as implemented is able to return results represented either as real and imaginary coefficients or amplitude and phase for each frequency.

Power conversion Generically speaking, if the input to a Fourier transform has units, then the output will have units equal to those of the original values divided by the units of the new domain (usually Hz). The unit of interest is generally power, computed from the square of the amplitude for each frequency, divided by the characteristic impedance of the system (50 ohms for all cases here). The processor is able to return either the power spectrum or a power spectral density by properly accounting for the bin width. In figure 6.1, a typical spectrogram is shown. The x and y axes are time and frequency respectively, where the time binning is exactly the windowing used by the data reader, and the frequency binning is the result of a discrete FFT of fixed length.

Thresholding

The physical receiver exhibits a frequency dependent gain. Some of this is desired (such as the suppression of filters), but there are also contributions from the performance of realizable devices and interference effects due to reflections at component connection caused by imperfect impedance matching. While a flat response would be ideal, when looking for weak signals on top of noise, it is the signal-to-noise ratio which matters because the gain effects both signal and noise equally and drops out of the ratio. In order to leverage this when searching for excesses of power, the threshold used is defined to be frequency dependant (that is, a fixed offset from the noise floor), rather than a fixed power level.

Accumulator The data accumulator collects array-like data and stores it in a continuous average. It can be configured either as a cumulative average, or as a moving average (where older data are weighted less strongly relative to new data), and can either output values at a fixed interval, or upon receiving a signal that all data have been received. We use this to estimate the average power level, letting noise fluctuations average out.

Gain Variation For the actual analysis, we do not use the bin-by-bin values from the accumulator. Instead, the data are further averaged into a configured number of bins and approximated with a spline. This results in a very smooth estimator of the background which is not sensitive to fluctuations. How well this procedure reproduces the true average noise level depends on the amount of data used, the duty cycle with signals present, the degree of rebinning used, and the scale of structure present in the noise floor.

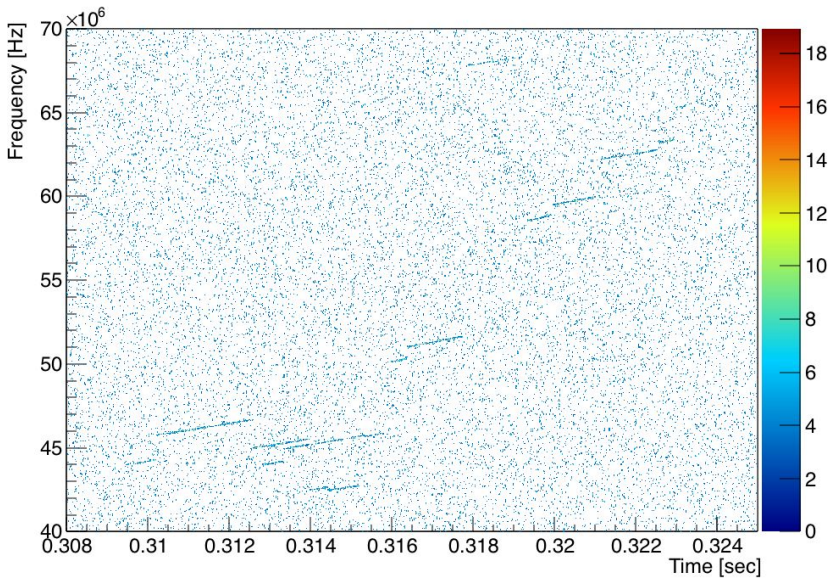


Figure 6.2: A spectrogram after the discriminator has removed those bins which do not exceed the specified threshold. By eye it is now relatively easy to identify tracks, though there is also still a relatively large number individual bins which have fluctuated above threshold.

Discriminator Once the gain variation has constructed an estimate of the noise floor, that is passed to the discriminator processor. This processor defines a threshold at some fixed power level above the noise floor. When a power spectrum (or power spectral density) for a particular time slice is passed in, each frequency bin is compared against this threshold. The processor then outputs, for each time slice, only those bins which exceed the threshold. In figure 6.2 is shown a typical spectrogram of data output from the discriminator. Note that bins below threshold are off scale and displayed as white.

6.2.2 Event reconstruction for the 2014 analysis

For the data collected in 2014, event reconstruction was conducted using a set of macros which have not been incorporated into the standard framework, and so

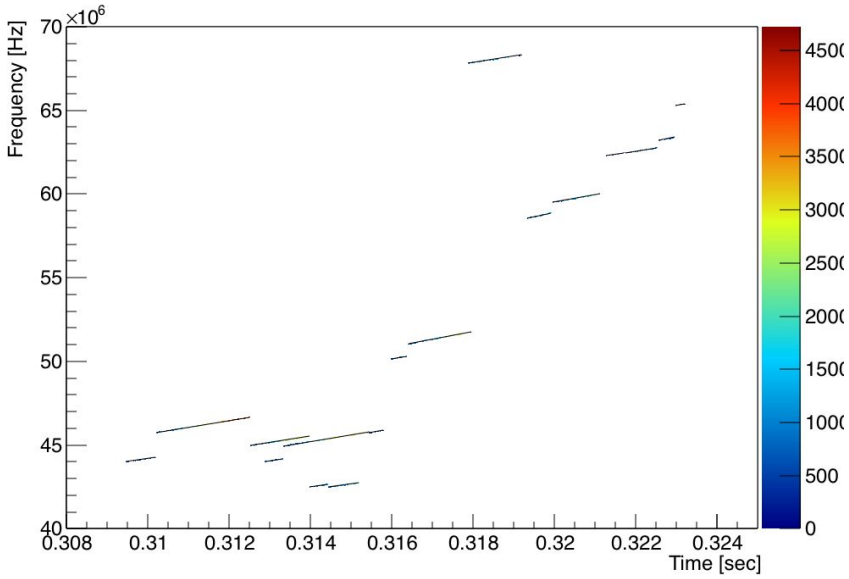


Figure 6.3: The high-power bins shown in figure 6.2 are passed through the track identification procedure; the identified lines are shown here.

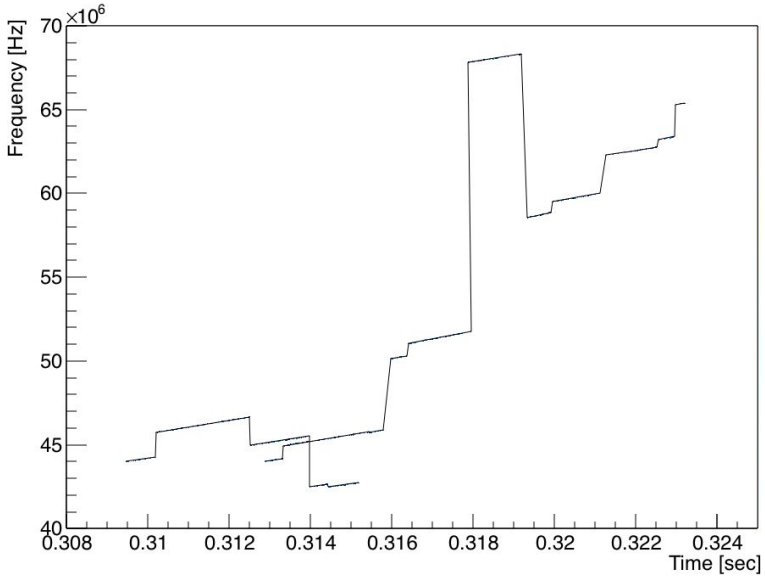


Figure 6.4: After identifying tracks shown in figure 6.3, they are combined head--to-tail into events. Note that here there are two events which partially overlap.

it is dealt with separately here. The procedure is based on using carefully chosen projections of the output of the katydid processing chain above.

As a first step, electron tracks were found by means of manually scanning the data and the average slope (i.e. the trajectory in time-frequency space of the excess of power) was estimated. The output of the discriminator is then transformed from the standard time-frequency domain to one rotated by precisely the track slope. This leaves the data in a new two-dimensional space, where both axes have units which are a combination of time and frequency, and where electron tracks show up as lines which extend in only one dimension. Integrating the data along the same axis in which tracks are extended results in very narrow peaks where tracks are located. For each bin in the projection which exceeds threshold, the perpendicular projection can be used to determine the start and end coordinates of the track. Once a track has been located, it can be converted back to the usual time-frequency domain and the starting frequency determined. Figure 6.3 shows a collection of tracks identified in this way. Tracks so located are then combined head-to-tail into event objects which include all the emissions from a single electron, including after scattering events (see 6.4 for an example of how tracks get combined). The starting frequency of the first track in each event then populates a histogram which is calibrated to energy.

6.3 Results from 2014

There were two primary sets of data collected in in 2014. The streaming data acquisition system was used to scan the majority of the bandwidth available to the system, including the highest energy conversion electron lines from the ^{83m}Kr source. This was done using a deep harmonic trapping configuration (800 mA current in the trapping coil), which increased the total event at a cost of degraded line width.

Later, data were collected using the RSA acquisition system and a weaker trapping field (400 mA in the trapping coil), which resulted in a lower event rate but narrower observed line width. These results were published by the collaboration in 2015 [32], with the results presented in the energy spectrum shown here in figure 6.5

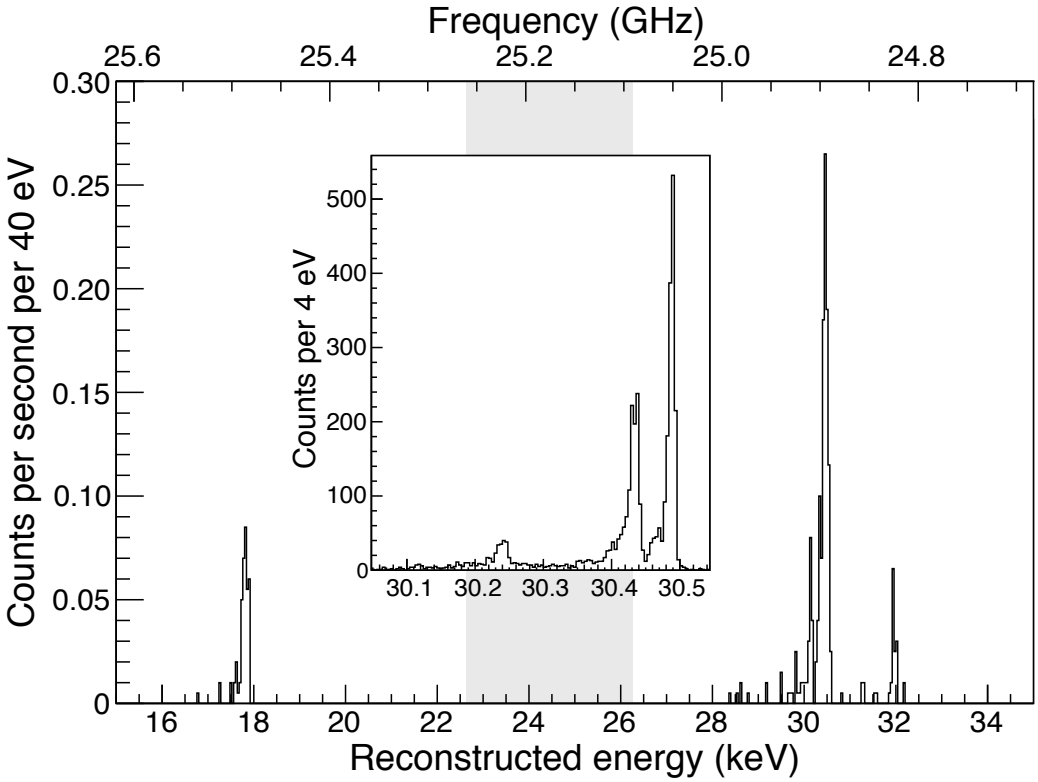


Figure 6.5: Full energy spectrum of data collected in 2014 and published in 2015. The data were collected in a sequence of partially overlapping frequency bands, shown together in the full figure; the grayed out regions represent the frequency band where no data were collected. Data in the main figure were collected using a single trapping coil with relatively large (800 mA) current, while the inset shows data collected with half the trapping current (400 mA) and the corresponding improvement in the observed line width. In the full spectrum, the primary feature near 30 keV has a FWHM of 140 eV, while the inset spectrum has a FWHM of 15 eV. This is a direct reproduction of the figure we originally published in 2015 [32].

Chapter 7

Investigation of sidebands and trigger diagnostics in continued phase I operation in 2015

After the successful demonstration of CRES described in chapter 6, we identified further goals for Project 8's first phase of operation and the transition into the second phase. In particular, having demonstrated that the technique works at all, it is valuable to quantify the performance achieved and identify the path to a system capable of making competitive neutrino mass measurements. That processes primarily manifests itself in the phased approach mentioned in chapter 2, where each stage has technical goals which demonstrate the various requirements of the ultimate experiment. However, especially with this first phase, it is also useful to do extra studies which help us understand exactly where we stand. In particular, we are interested in investigating the energy resolution, detection efficiency, and geometric scalability of the approach.

7.1 Hardware changes in advance of operation in 2015

Following successful operations, several hardware changes were identified to improve the performance of the phase 1 prototype. These changes were not major design changes or even particularly significant insights, but they are important to remember whenever comparing data taken during the different time frames.

The first piece of work is by far the most straightforward. Our cryocooler has regular maintenance cycle requirements for which we were overdue during initial operations. The symptom of this was that the base temperature was both warmer and less stable (including very sharp increases of many degrees) during operation. Standard maintenance, as expected, solved this problem. The temperature history as logged using the control system is shown in figure 7.1, showing the performance change. It is only worthy of mention here because there had been consternation over the origin of the temperature instability, and to emphasise the importance of doing one's due diligence.

At the end of the summer of 2014, the NMR magnet providing the background field quenched. This forced the end of data taking and provided the opportunity for the work described in this chapter. Prior to the quench, attempts to operate the prototype with the bathtub trapping geometry, described in 3.2, had all failed. As part of the re-energizing of the magnet, we performed a more thorough mapping and careful shimming of field. The expectation was that the main field may not be well aligned with the waveguide axis, and that the axial motion of the decay electrons brought them into contact with the walls when a trap configuration with longer axial extent was used. While the insert was removed, we also determined that several components (specifically electrical connectors and heater resistors) were made from

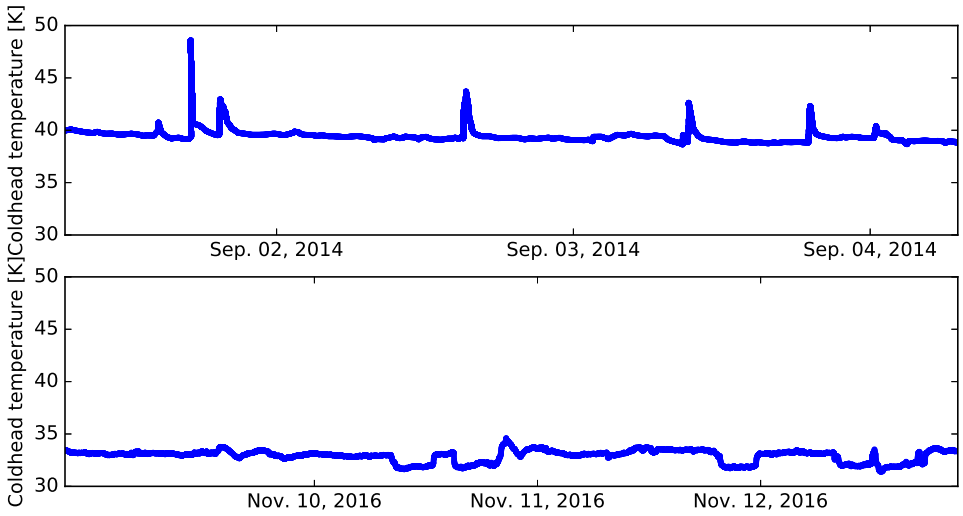


Figure 7.1: Time series data for the coldhead temperature prior to the refurbish and at the approximate time of this writing. In the upper figure, sudden jumps of several kelvin are visible over the course of several days; the detector system was nominally under consistent conditions over this time interval. The lower figure shows both a lowered ultimate temperature, and does not exhibit the discrete increases in temperature. There are smaller discrete changes in temperature which are a result in changing heat load due to changes in the current present in the trapping coils.

ferromagnetic components. We took the opportunity to replace or relocate these far from the trapping region.

Finally, the high-frequency analog signal chain was examined and triaged. During operation, the insert's waveguide connections had become loose, with visible gaps. We also found that over the various changes and iterations, some of the active and powered components in the signal chain had been assembled in such a way that the current draw slightly exceeded the specifications of the power supply. The power supplies were replaced so that everything would operate within the specified tolerance. While we have no specific evidence that any of this was causing a particular problem, the repairs would generally be expected to improve both the absolute performance and repeatability of the system's performance.

7.2 Expanded analysis processors

In addition to hardware changes, there were significant advances made to the way data are analyzed. Aided by experience looking at the data from 2014, we went on to develop more sophisticated analysis procedures. This included a number of new processors, described in the following subsections, which are run after those described in section 6.2.

7.2.1 The k-dimensional tree

The k-d tree is data structure optimized for certain types of relationship searches [38]. The details of the implementation and performance are subjects of computer science study; for our purposes it is simply taken as an established tool. In particular, given a large collection of points in a multi-dimensional space, it is able to quickly determine the set of points closest to any particular chosen point. The *create-kd-*

tree processor takes the time-frequency domain points which are above threshold and populates such a data structure. The data structure supports computing distances either as euclidean distance or rectilinear (also known as Manhattan or taxicab) distance. Further, the data structure expects all dimensions to have the same units. Since we are working in the time-frequency domain, all points are normalized to be dimensionless using a scaling factor which is configurable for each axis.

This processor marks a transition in the progression of the data processing. All prior steps involves sequential time chunks of data being processed and passed along. The k-dimensional tree, like the accumulator of the previous chapter, collects all of the data until it receives a signal that it has collected everything. Subsequent processors iterate over structures identified in the data (points, tracks, events, etc.) It is important understand how those structures are stored because that determines the order in which they are passed to subsequent processors; simple time ordering is not possible, for example, when objects have extended and unequal duration.

7.2.2 Consensus thresholding

Like the discriminator discussed in the previous chapter, the consensus thresholding processor seeks to efficiently reduce the total number of time-frequency points under consideration for inclusion in event reconstruction. This reduction is important because the downstream reconstruction procedures are relatively slow, and processing times can be improved substantially be reducing the number of elements being considered. The procedure leverages the k-dimensional tree populated in the previous processor to efficiently count the number of neighboring points within some radius. The radius and a minimum number of neighbors is configured and any points with fewer than the minimum number of neighbors are deemed to be noise.

7.2.3 Track identification via Density-Based Spacial Clustering of Applications with Noise

Density-Based Spacial Clustering of Applications with Noise (DBSCAN) is an algorithm for identifying grouped populations within a large multi-dimensional domain [39]. The technique is very similar to the consensus thresholding method described above. In both cases the units of each axis are normalized away and for each point those points within a unit hypersphere sphere are considered to be neighbors of the original point.

Rather than removing points with insufficiently many neighbors, DBSCAN creates a grouping, called a cluster, starting with the original point. All of that point's neighbors are added to the cluster, along with all of their respective neighbors. The process continues, adding new neighbors to the cluster, until the cluster contains all points reachable from the original point by stepping from point to point in steps smaller than the defined radius. Once a cluster is completed, the procedure continues selects another starting point and continues to build clusters until all points have been added to a cluster. For our analysis, any cluster with fewer than a configured minimum number of points is discarded.

7.2.4 Hough transform

The Hough transform is an algorithm for processing image-like data and extracting features, particularly line segments [40]. Each cluster identified by DBSCAN is analyzed using the Hough transform to identify the underlying line of excess power in time-frequency space. The cluster of points, along with the Hough result, are passed along as a candidate track object.

7.2.5 Track processing

The track processor starts with the results of the Hough transform, specifically the line segment constituting a track. In two stages, points which are further from the best-approximate line are removed and the line itself is recalculated. Once the final excess point removal is complete, relevant physics parameters for the track object are computed. These include the track's initial and final time and frequency, as well as the power and other parameters of interest.

7.2.6 Multi-peak event building

In the 2014 data there is no evidence that we have made any observations of sidebands. Our best explanation is that there are two factors leading to this. First, because all trapping was done using a harmonic configuration, the frequency of axial motion is expected to be large enough that when a main emission is visible at the cyclotron frequency, any sideband would fall outside of the receiver bandwidth. Second, the higher noise floor and reduced signal-to-noise ratio of the early data would generally make it less likely that a sideband would be visible at all, and especially unlikely that it would be sufficiently powerful that it triggers the acquisition system.

In the 2015 data, we see events with clearly visible sidebands (see figure 7.2, for example). This structure was predicted in the proposal of the CRES technique [15], so seeing it is expected, but it does create an edge-case for the early event construction algorithm. To solve this, I generalized the head-to-tail technique being used previously to cover the case where any given track may have parallel sidebands.

The procedure for this multi-peak event builder starts by taking the collection of found tracks and collecting any with the same start and stop time into a multi-peak track. To allow for slight fluctuations, a time interval is defined to set what counts as

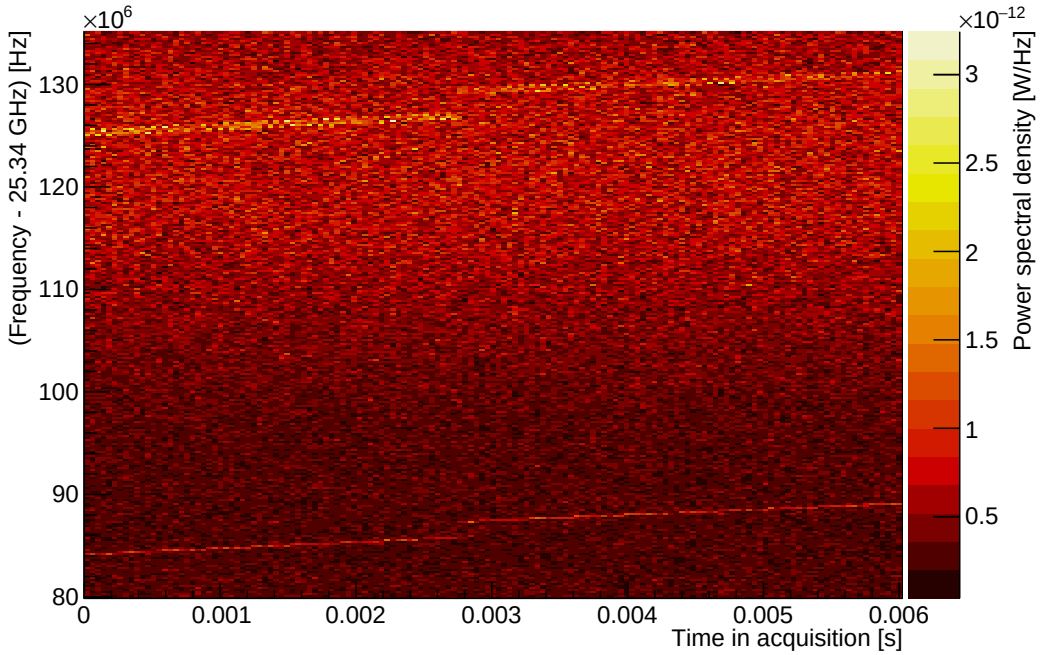


Figure 7.2: Spectrogram of an event with visible sideband. The vertical axis is presented in units of frequency as seen by the data acquisition, it has been mixed down in stages, first by 24.2 GHz, then by 1.14 GHz for a total of the 25.34 GHz indicated. The frequency axis has been rebinned by a factor of eight relative to the true frequency resolution to improve visibility. The central emission line begins at a time of zero and roughly 84 MHz (axis scale) then slowly increases in frequency until it scatters around 0.0028 s. A subsequent track continues from that time through the end of the time window shown. A parallel sideband is visible, beginning around 125 MHz at zero time and proceeding in the same fashion as the central line, including scattering at the same time.

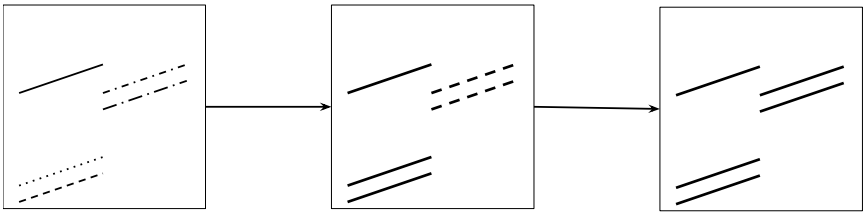


Figure 7.3: Simple illustration of the grouping progression used by the multi-peak event building processor. First, a set of tracks are received from whatever processor is being used to find and characterize tracks; the cartoon here has five tracks. Second, the tracks are grouped into sequences. All tracks within a sequence have the same start time and end time, but their frequencies will differ. Shown in the middle frame, the three tracks on the left are all grouped into a sequence and the two tracks on the right form a second sequence. Finally, if any sequence begins at the same time another ends, they are combined into a single event. In the right-most frame, the two sequences have been combined into a single event.

the same time. It is expected that this tolerance should be roughly the size of the time bin width, to allow for a case where the first bin is missed due to noise fluctuations. Edge cases, where parallel tracks have the same start or stop, but not both, are not treated, but they are detected and a boolean property set to indicate the detection of an unexpected event structure. Next, multi-peak tracks are combined into events using the same head-to-tail algorithm from before.

It is very important to note that the multi-peak event builder was implemented as a first pass at something which is capable of treating the case where parallel structures are present. It seeks to solve the problem of parallel tracks being treated as independent electron events, which results in sidebands populating an energy spectrum with unexpected extra peaks. No attempt is made to use the tracks which have been found for a more targeted search of previously unseen tracks. Similarly, there is

no judgement made about which tracks are sidebands and which are central emission lines. The hope is that this system will make it easier to study what sorts of event structures are present in the data, which will then guide the future development of a more sophisticated and robust event reconstruction.

7.3 Investigations of data from continued phase I operation in 2015

In this section, I present the results from several investigations of the data taken in early 2015. The software advancements of the previous section were developed as part of this process and a large portion of effort was spent validating their performance. Further, this time coincided with major changes to the way data for Project 8 are stored and accessed on the computing resources at PNNL. For that reason, it was necessary to work with a subset of the data, which could be stored and processed locally.

The following sections all make use of run 236, an extended data taking period centered on the pair of conversion electrons near 32 keV. This dataset comprises 10234 triggers of the RSA, which were processed using both the multi-peak track event builder described above, and the older event builder which uses DBSCAN to combine tracks into events.

7.3.1 Understanding the RSA trigger

As previously described (see section 5.2), the RSA's trigger uses a frequency-dependent but temporally single-bin threshold to detect events. This is a contrast to the event reconstruction used in the analysis, which requires multiple adjacent

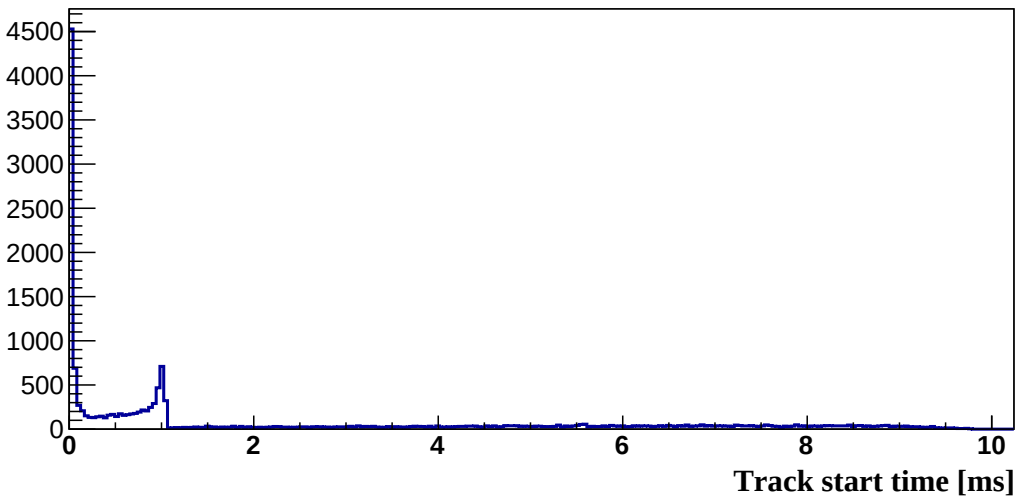


Figure 7.4: Distribution of reconstructed event start times for a collection of RSA triggers centered on the conversion electron lines near 32 keV. The peak at 1 ms is expected as this is is the pre-trigger time configured for the acquisition. The excess of events that start near 0 ms are what would be expected from either triggering when a second event (possibly below the trigger threshold) is present, or if an event lasts much longer than the acquisition time and is able to re-trigger the system. Shown are 10234 triggers of the RSA, which are reconstructed as 16746 distinct electron events by the analysis software.

high power bins in time-frequency domain to identify a track. The multiple-bin requirement allows the threshold to be lowered, since high-power fluctuations in several adjacent bins is far less probable than a single-bin fluctuation.

A study of the distribution of reconstructed event start times can tell us quite a bit about the trigger being used. In figure 7.4, the distribution of reconstructed event start times is show. For this operation, the acquisition was configured to be 10 ms long, with a pre-trigger of 1 ms.

The most obvious feature is the expected peak at the pre-trigger time of 1 ms. For those events with lower than average power, the RSA may trigger at some time after the track has already started; the reconstruction software finds the earlier start time due to the lower signal-to-noise ratio threshold. It is also possible that there are some events which have an initial track which is too low in power to trigger the RSA, but

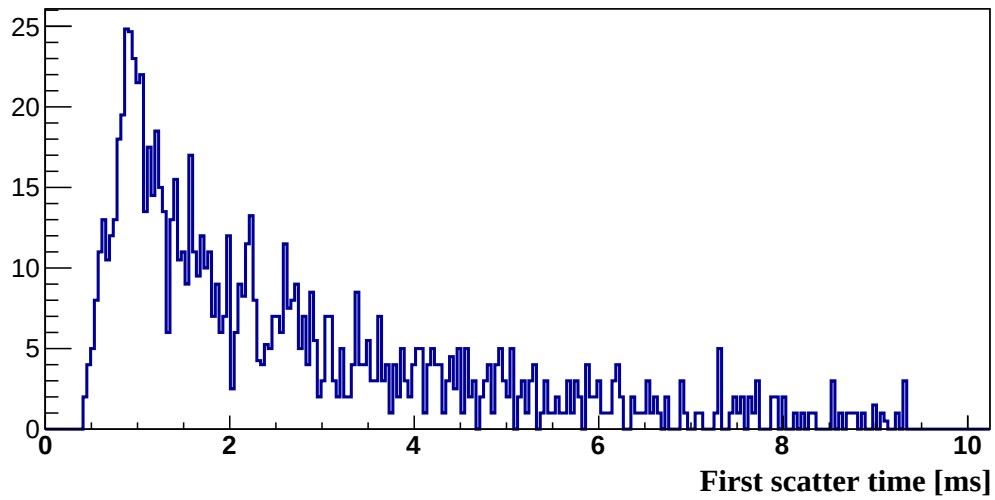


Figure 7.5: Distribution of the time of the first scatter for events with reconstructed start times between 0.2 ms and 1.8 ms. This cut selection was intended to skip the obvious features at 0 ms and 2 ms. There are 2400 events which start in this window, of which only 966 have a second track and thus are included in this histogram.

which produce a track with higher signal-to-noise ratio after scattering. This could be a result of an energy change that moves the electron's signal into a more favorable frequency region, or possibly a change in the electron's motion which results in more effective coupling of the radiated power to the sensitive modes of the waveguide.

Another population of events fill the time interval between the start of the file and the pre-trigger time. In figure 7.5 I plot the time of the first scatter for the subset of events which are reconstructed as starting between 0.2 ms and 0.8 ms. Qualitatively, it appears to be peaked near 0.2 ms, consistent with events with a first track invisible to the RSA. In the future studies will be made possible through the use of a streaming data acquisition system, which will clarify the situation and also facilitate development of a trigger scheme that does not exhibit the same behavior.

The other obvious peak in figure 7.4 is at 0 ms, the start of the file. These are presumably either tracks that the are completely missed by the RSA, but perhaps

after a scatter the new track triggers the RSA, or possibly tracks that are already present as soon as the RSA's trigger is rearmed. This population dies off quickly after the first couple of bins, indicating that tracks that actually start prior to the acquisition are found more-or-less immediately by the offline analysis.

Finally, there is a population of events which start after the pre-trigger time. In figure 7.4 there roughly 6952 events that start after after 1.2 ms. Some of these may represent coincident events, where a second electron is produced and detected during the acquisition window. Another possibility is that these tracks are actually related to the original event which caused the trigger. That would mean that either an intermediate track was not seen, possibly because the electron scattered to a configuration that is not visible, then scattered back into detectability, or a portion of a track was not found, breaking the event.

These observations serve to strongly underscore the insufficiency of trigger system employed by the RSA. Having a large fraction of stored data that likely missed the start of the event means that those triggers are not contributing statistics, but they do introduce dead time, since the trigger then has to rearm.

7.3.2 Energy reconstruction

One broad but important goal for project 8 moving forward is demonstration of robust and precise energy reconstruction for events. To investigate this, the pair of conversion electrons near 32 keV is especially interesting because those two lines are only separated by 7 eV. From the outset, the expectation was that the pair of lines would be distinguishable, and that improvements in event reconstruction would clear up spurious spectral features. Furthermore, understanding of expected features in the data, such as sidebands, should allow corrections which further improve resolution. It is important to stress that while for this discrete spectrum of electrons,

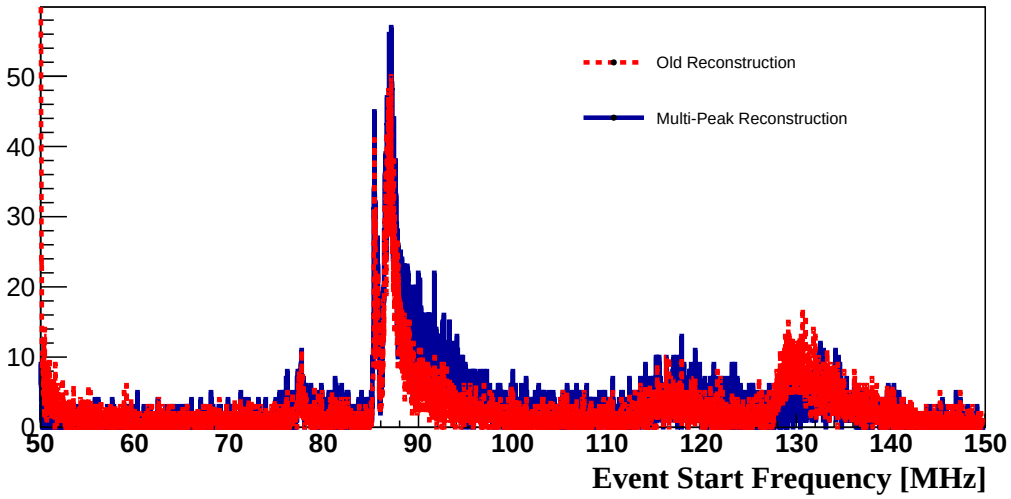
energy (or equivalently starting frequency) can be an interesting parameter for cuts or classification, it will not remain a usable parameter when evaluating tritium data because the spectrum is continuous.

The first thing considered was the multi-peak event reconstruction discussed previously. In figure 7.6, I show all events reconstructed from run 236 using both the older DBSCAN-based event reconstruction, and the new multi-peak technique. Even with only the original analysis, the full width at half maximum observed is 3.6 eV (where the conversion to energy is via equation 2.4 with the magnetic field magnitude taken to be 0.958 462 T. This is a significant improvement over the roughly 15 eV achieved previously and results from the use of a bathtub rather than harmonic trapping configuration.

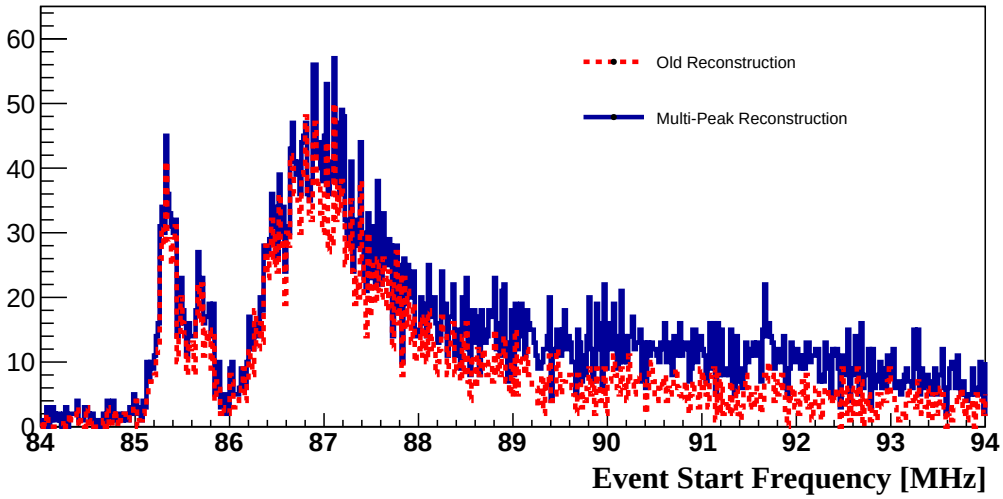
With the multi-peak reconstruction there is some reduction in the visible sideband feature, notably around 130 MHz, the technique also exhibits a significant increase in the broad continuum of unexpected events (specifically between 88 MHz and 95 MHz. Note that pair of peaks between 85 and 86 MHz are consistent with expectations for the true conversion lines; everything at higher frequencies, in both reconstruction schemes, is an indication of poor reconstruction.

The first place to look to clean this up is the event start-time distribution shown in figure 7.4. As was discussed in section 7.3.1, the distribution of starting times has two peaks, which naturally divide the event into four populations (i.e. each of the peaks, the region between them, and the region after the second peak). In figure 7.7 I plot the full spectrum, stacked in slices based on those populations. The start-time populations are clearly making different contributions to the reconstructed spectral features, which are interesting in the context of the possible origins of these populations suggested previously.

The most obvious population are the events which start very near to the configured



(a) Full acquisition bandwidth



(b) Zoom of central structure

Figure 7.6: Full spectrum of the 10234 trigger events from run 236. The older event reconstruction algorithm finds 13941 electron events, while the new method described in section 7.2.6 finds 16746. The upper plot shows the full 85 MHz of receiver bandwidth, centered at 100 MHz, while the lower plot has the same binning and zooms in on the structures of interest. The pair of peaks between 85 MHz and 86 MHz are consistent with the conversion electrons near 32 keV. The observed FWHM here is 3.6 eV, a marked improvement over the 15 eV shown in figure 6.5. The multi-peak event builder clearly suppresses some of the structures on the edges of the bandwidth, which are believed to be sidebands, but has a greater excess of unexpected events between 86 MHz and 100 MHz.

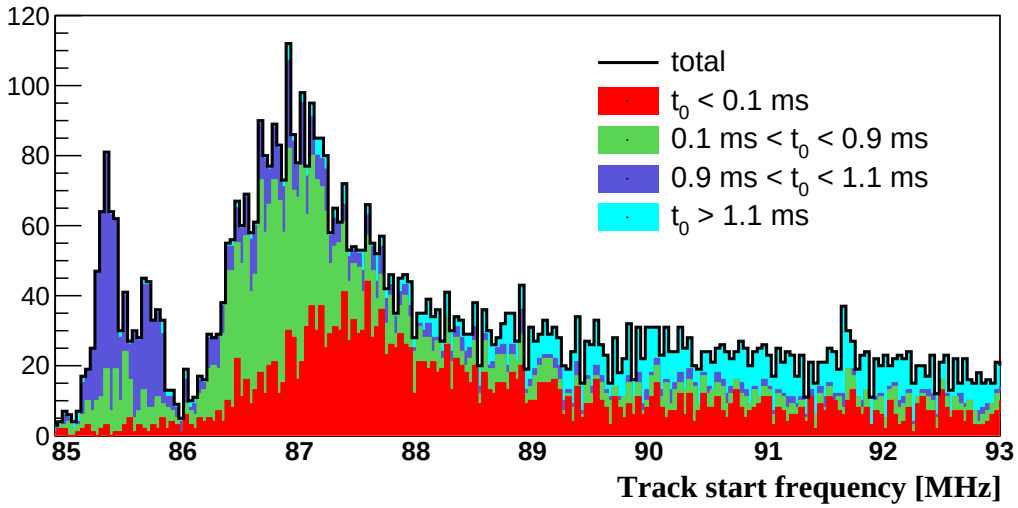


Figure 7.7: Starting frequency for events collected in the region near the 32 keV (run 236). Note that the four sub-populations are mutually exclusive and the histograms shown are stacked. The full spectrum, show in figure 7.6, includes all 16746 reconstructed events, though here I have zoomed in on the roughly 8 MHz band which contains the most interesting structure. The events are divided into four populations based on the start time of the event, with divisions intended to reflect the populations described in section 7.3.1 and visible in figure 7.4. It is clear that the pair of peaks between 85 MHz and 86 MHz contain mostly events with reconstructed start times very near the pre-trigger time of 1 ms, while other event start times primarily populate the broader distribution above 86 MHz.

pre-trigger. These events form the bulk of those in the peaks of interest. A small number show up in the continuum above 86 MHz, and the rate of this is very sensitive to the choice of 0.9 ms to separate those populations. As it is, there seems to be leakage in both directions across this boundary, so the exact choice is not critical for qualitative analysis of the results.

The next population to consider are those events which start near to the pre-trigger time (here defined to be between 0.1 ms and 0.9 ms). These events primarily populate the lowest frequency portion of the broad feature above 86 MHz, though as just discussed, some show up in the peaks of interest. Especially given that figure 7.5 indicates that these events tend to have subsequent scatters near the pre-trigger time, it is likely that these are events which start out emitting power which is missed, but which latter scatter to a configuration which triggers the data acquisition system. If the true initial condition of the electron event is missed, it would be expected that the subsequent detection would be at lower energy (higher frequency) as is seen in this population.

The population of events which start prior to 0.1 ms are similar to the previous population. That they do not contribute significantly to the peaks of interest is not surprising, since it is extremely likely that the true event start has been missed, leading to the same higher-frequency expectation as before.

Finally, there are those events which begin after 1.1 ms. These events tend to appear at higher frequencies than the prior two populations and obviously differ in that they come after the trigger rather than before it. For these, it seems plausible that they are actually related to the event which caused the original trigger, but which the event reconstruction failed to combine as scatters of the same electron. This would happen if an event consists of many scatters and an intermediate track were missed by the analysis, or if a single track were reconstructed as two with a

large enough gap between that they were not combined into the same event. If these events are largely actually electrons which have scattered at least twice, as would be required for a missed intermediate track, one would expect more energy loss before detection, consistent with the higher frequency than the other subpopulations just discussed.

7.3.3 Observation of events with sidebands

As mentioned earlier, in many events collected in 2015, such as the one shown in figure 7.2, sidebands are found in the data. These are expected features and understanding them in detail should provide an additional handle for achieving precise event reconstruction. In particular, the exact sideband separation is a result of the electron's axial motion, which provides a means of determining the electron's pitch angle. Therefore, a measurement of sideband separation should provide a means of correcting for systematic effects related to pitch angle.

In order to study the effect of corrections based on sidebands, it is important to find a population of events that can be well understood. The pair of conversion electrons near 32 keV are attractive because their natural widths are less than 2 eV and their separation is 7.7 eV [19], making it obvious if spectral resolution is better than that scale. In figure 7.8, I show a sequence of cuts being used to attempt to isolate a population of these events which have detected sidebands. However, the first cut, which eliminates those events which do not have detected sidebands also effectively eliminates the narrow lines of interest. The events which remain are those in the broad structure at higher frequency. This population was discussed in the previous section and consists primarily of events with start times which are inconsistent with the configured pre-trigger in the data acquisition system. While it may be possible to correct the energy reconstruction of these events based on information from the

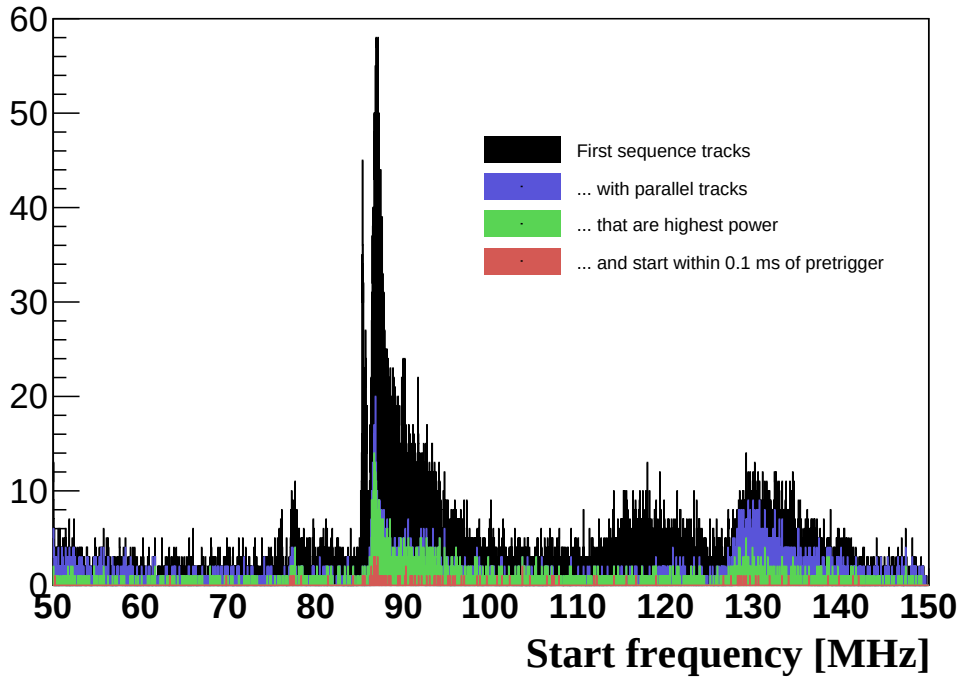


Figure 7.8: Investigation of cuts for isolating a population of events with sidebands. In black is the start frequency of all tracks which are part of the first sequence in an event. In purple, all events with only a single track have been eliminated. Note that this cut essentially eliminates the population of events between 85 MHz and 86 MHz which represent the actual events of interest. In green, only the track with the highest average power is shown. This significantly reduces the feature near 130 MHz, which is consistent with the expected location for an upper sideband. Finally, in red, events with start times which differ significantly from the pre-trigger (more than 0.1 ms away) have been removed. As might be expected based on figure 7.7, this cut leaves few events and little structure.

sideband separation, because they are not part of a narrow spectral structure there is not a good way to evaluate such a correction.

Despite the limitations of the events with detected sidebands, there can still be some lessons learned from them. If we compare the purple and green histograms in figure 7.8, we see that selecting the highest power track within a sequence tends to retain tracks near the lines of interest (in the region between 85 MHz and 90 MHz) while suppressing those tracks near 130 MHz, which are consistent with sidebands. This indicates that track power may be a useful discriminator, on component of a discriminator, between sidebands and tracks at the central cyclotron frequency.

To investigate the power selection, it is useful to also look at the distribution of average detected track power (note here that the average is taken along the track, i.e. the track duration is factored out). This is shown in figure 7.9 for the same populations as are shown in figure 7.8. Here we see that there are clearly two populations, though they are not especially well separated. This indicates that average track power may be a useful tool for characterizing tracks as either central cyclotron frequency emission or sidebands, but that further discrimination will also be required.

Figure 7.10 shows the distribution of observed sideband separation for those events which have at least one detected sideband in the first sequence. There are features at symmetric positions relative to 0 MHz at absolute separations of roughly 85 MHz, 45 MHz, and less than 1 MHz. The features near ± 85 MHz are most likely a result of events where the track with the highest power turns out to be in the sideband. The largest populations are separated by just over 40 MHz and are what is expected for the case of a central line with a sideband on one or both sides. The two narrow lines near 0 MHz are a result of individual track objects being reconstructed as multiple nearby tracks rather than a single object.

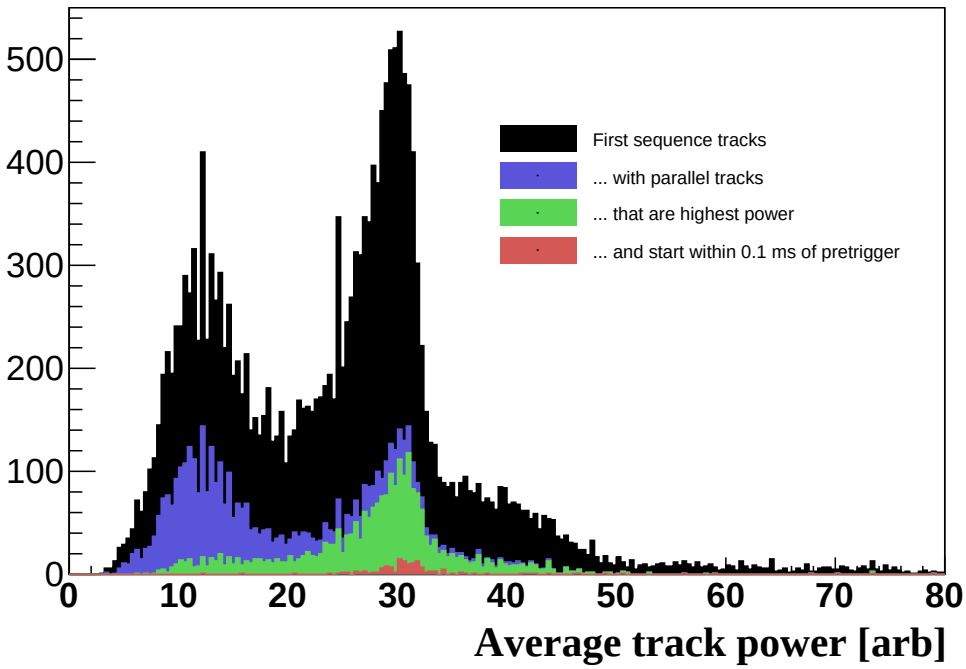


Figure 7.9: Distribution of average detected power along tracks. Here the power units are proportional to power delivered to the RSA, but because those units are after the complex receiver gain, it is not very instructive to compute physical units (note, however, that the units are linear, proportional to Watts, not dB). All tracks which are part of the first sequence of an event are shown in black. In purple, tracks are eliminated if they are the only track in their sequence. In green, only the track with the highest average power is shown for each sequence. Finally, the red includes only those events which start within 0.1 ms of the prescribed pre-trigger. These colors and populations match those shown in figure 7.8.

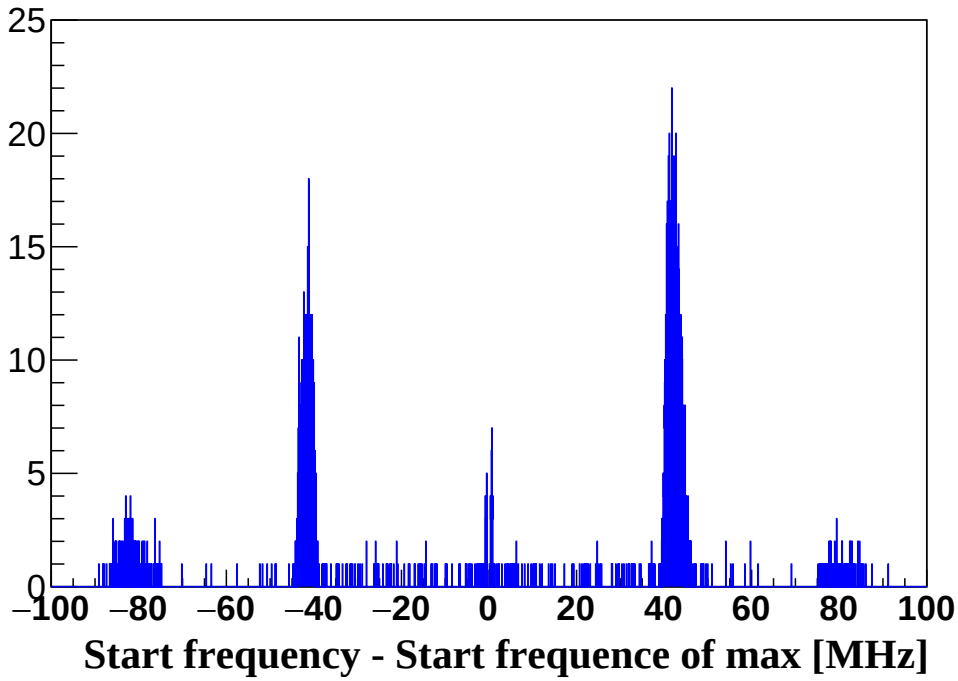


Figure 7.10: Histogram of the frequency separation of tracks in the first sequence of events where that sequence has multiple tracks. The track with the highest average detected power has been assumed to be the central cyclotron emission and all other tracks in the sequence are measured relative to it.

7.3.4 Ongoing investigations

The results presented thus far in this section stress the need for more sophisticated means for looking and understanding the properties of tracks and events once they are identified. Efforts in this direct are ongoing and recently collaborators have made exciting advances (for example, see [41, 42, 43]), which are summarized here.

Once an event has been identified, it is possible to go back to the time-frequency domain data and explore in further detail. In [41], the output of the discriminator processor is considered for each identified track. Following a similar projection procedure to that used on the 2014 data (described in section 6.2.2), bins above threshold are counted along a path parallel to the track, for the duration of the track, for each frequency bin. The result is a one dimensional histogram generally peaked at the starting frequency of the identified track. From that distribution, a number of potentially interesting parameters are calculated, including the first four moments (mean, variance, skewness, and kurtosis) and a Gaussian fit. From hand-scanning, it seems that sideband tracks (identified as such by their starting frequency) tend to have more spread than tracks at the cyclotron frequency Baseed on that, a power ratio, defined as total of the distribution within three standard deviations of the center of the Gaussian fit to the total of the distribution outside of that interval, can be computed for each event.

For the electrons near 30.4 keV (data run 177), similar to the events near 32 keV discussed in previous sections, there appear to be features corresponding to the central cyclotron frequency, as well as sidebands. For this work, a frequency window of roughly 40 MHz was used to label some tracks as central and all others were labeled as sidebands. With the identified tracks broken into populations of sidebands and central lines, the parameters of interest can be explored; the results for the power ratio just discussed are shown in figure 7.11. It is interesting to see that the two

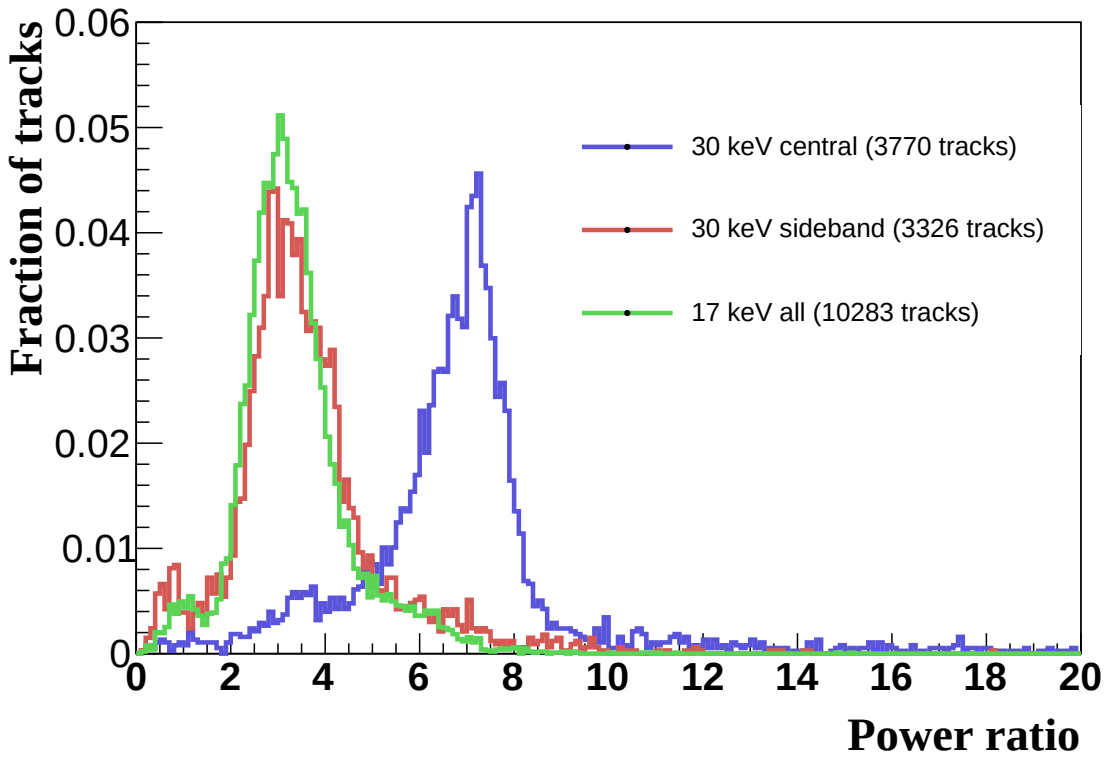


Figure 7.11: Ratio of the projected power near a track to that further away, for spectrograms which have been projected along the track (details of the projection are discussed further in the text). In purple and red are tracks from events near 30.4 keV (run 177), where the lines are defined to be the central cyclotron emission (3770 tracks) or sidebands (3326 tracks) based on their starting frequency. In green is the same ratio for all (10283) tracks collected near 17.8 keV (run 221). Because they have such different statistical content, all three histograms has been normalized to one.

populations of tracks are indeed qualitatively separated in this parameter. There is some overlap in the two distributions, but this serves to motivate both efforts to find parameters with better discrimination or techniques for considering multiple parameters in characterization, and also efforts to generate cleaner training populations of events, since some of the leakage is almost certainly from the way data tracks were originally classified.

A long-term point of investigation has been that the visible electron lines do

not seem to perfectly follow the expected energy-frequency relation of equation 2.4. Recently, the analytic trap calculations in [43] determined that prior estimates of the expected axial frequency were too large by a factor of two based on the experimental design, a confusing observation since the observed sidebands seemed to agree with the earlier prediction. Following up, it was realized that the presence of the short at the end of the waveguide would, at the frequencies corresponding to the lines of interest, suppress either the 1st order sideband, or the emission at the cyclotron frequency and the 2nd order sideband (all higher order sidebands should be outside of the receiver's bandwidth). This means that for the lines near 30 keV, what had been interpreted as 1st order sidebands are actually 2nd order, explaining their position. Similarly, for the pair of lines observed near 17.8 keV (where only one conversion line is expected), both of the observed lines are actually 1st order sidebands.

Following this line of reasoning, the entire collection of tracks from the 17.8 keV events are included in figure 7.11. It is clear from that figure that all of these tracks form a single population in the power ratio domain, and that that population is consistent with the higher energy events which had been identified as being sidebands. This is also consistent with the surprising observation that the trigger rate for 17.8 keV events was much lower than expected during data collection. This effect was previously attributed to the frequency-dependent response of the receiver, a known issue which will need to be treated carefully in future phases, but which was accepted in phase I where the goals are primarily demonstration. The observed electron events are shown in figure 7.12, along with the frequency to energy calibration based on equation 2.4 using a separately determined value for the magnetic field. That determination was made using only the electrons near 30 keV in a harmonic trapping configuration with variable trap depth. It is clear that both observed lines near 17 keV are far from the calibration, but that their midpoint (where the central

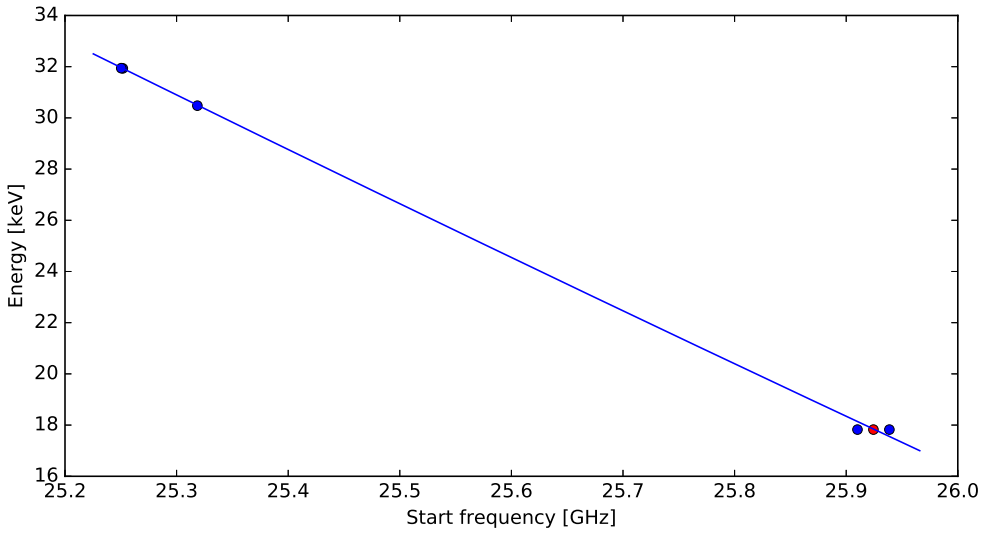


Figure 7.12: Scatter of the observed electron features, here shown with the observed frequency on the x-axis and the expected energies on the y-axis. The blue curve is the calibration of equation 2.4, where B was previously determined to be 0.958 462 T and is not a fit to any of the data shown. The points in blue are the observed events and the single point in red is the midpoint of the two observed features above 25.9 GHz (these two are those observed in the region of interest near the 17.8 keV conversion line. That their midpoint matches the energy calibration is consistent with the hypothesis that these points are both sidebands and any feature at the actual cyclotron frequency has been missed.

line would be if these are first order sidebands) is consistent with the calibration.

7.3.5 Outlook

In the proceeding section, I presented the results of a number of preliminary analyses of the data collected in the first part of 2015. The initial aim of this work, motivated by the first observation of sidebands in individual events, was to upgrade the data analysis to account for these new features, and then to explore their use in processing and interpreting the results. Unfortunately, as was shown in figure 7.6, an extremely large fraction of the events collected are not reconstructed as part of the conversion electron lines of interest. The most plausible explanation for this is

that the configuration of the triggered data acquisition system was not adequately able to capture many events. Nevertheless, distinct populations of electron events were able to be identified through the use of newly implemented analysis procedures. It will be very interesting to compare these results to those obtained by performing the same analysis on data collected with the streaming acquisition system currently being commissioned. Furthermore, recent efforts are showing great promise towards a complete understanding of the electron events as observed using CRES.

Chapter 8

Conclusions and Prospects

The Project 8 collaboration has made significant progress since CRES was first proposed in 2009 [15]. Five years after that initial proposal, we had constructed and attempted to operate a series of prototypes, culminating in the phase I detector, which was operated beginning in 2014 and used to collect data publish as the first demonstration of the technique in 2015 [32]. Since then efforts have been made to improve the performance of the technique at this demonstration scale, while the final design work is completed for the upcoming phase II operation, expected to begin in early 2017. Design and planning efforts for subsequent phases has also already begun and will ramp up as phase II transitions from the comissioning to the operations phase.

8.1 Major efforts

The path to the phase III and IV design goals will require upgrades to nearly all aspects of the experiment. For the demonstration phase of the project, many of the performance criteria have been qualitative judgements of what is good enough. To move forward, these need to become quantitative, with achieved performance determined and requirements identified. While the figure of merrit of most interest is

projected sensitivity to m_β, there are others which are easier to isolate to a particular subsystem, or measure at a particular stage of research and development. Without offering particular solutions, I will take the following short sections to outline some particular challenges where substantial efforts will clearly be required.

A few parameters are particularly critical to understanding the path from demonstration to neutrino mass measurement. Energy resolution is perhaps the most obvious figure of merrit as translates directly to achievable sensitivity. While the demonstrated performance is not adequate, it is already approaching the level of conventional experiments and it is reasonable to assume that further avancements in data processing will be possible. Reconstruction efficiency incorporates many related concepts including physical trapping efficiency, trigger efficiency, and reconstructablity of events. This is a major concern and the ability to be sensitive to a large fraction of source decays will likely be a major factor in determining the viability of phase IV.

8.1.1 Scalable and robust data acquisition

One of primary conclusions of chapter 7 is that the existing data acquisition system is inadequate for our needs moving forward. The RSA has been an incredibly useful diagnostic instrument, allowing us to make qualitative judgements of the experiment's health without having to wait for data transfers or processing steps. On the other hand, the available trigger configurations are limited, and it is clear even with the existing analysis procedures that important portions of electron events can be missed. Furthermore, it offers no means for combining or processing multiple channels, as will be required beginning in phase III.

In the nearest term, a new streaming acquisition system is being developed (see section 5.3). This system will allow continuous collection of data over a timescale of seconds, providing two significant benefits. First, the chances that the beginning

or end of any detected electron event will extend beyond the available time interval will be greatly reduced. This should reduce the fraction of detected events which are unable to be faithfully reconstructed, making the data collected more useful for performance studies. Second, we will have a sample of the exact data which will be available to a custom online triggering system. Proposed triggering algorithms can be run directly on this data and benchmarked for processing time, resource requirements, and efficiency.

Looking to the future this system will need to address several new challenges. Starting with phase III it will be necessary to collect signals spread across a number of channels. This is required both to provide coverage of a much larger source volume needed to achieve a sufficiently high exposure, and to facilitate the separation of concurrent events. The acquisition system will need to be able to coherently combine channels so that the signal to noise ratio is large enough for triggering. Further, a triggering scheme must be developed such that the total volume of stored data remains manageable, while having a much higher fidelity than the existing scheme. One advantage of the custom system currently envisioned is that we have complete control over all stages of the signal processing. The RSA's available algorithms are designed to work for a wide range of applications and are therefore generic and conceptually simple. We can, instead, implement more complex logic which is specific to the signals we expect, much as is done in the existing off-line analysis.

For the demonstration phase of Project 8, it was sufficient to operate in any mode where events were able to be identified. Moving forward it will be critical that the achieved resolution of energy reconstruction is measured. Furthermore, it will be critical to quantify the efficiency of the acquisition and analysis system as a function of various parameters, including pitch angle, position within the source volume, and energy. Any non-uniformity to those efficiencies is a potential source of systematic

errors, and the absolute value of the efficiency determines the actual scale of the experiment.

8.1.2 High performance computing

Even with a well designed trigger, the Project 8 acquisition system is expected to produce large amounts of data during operation. For the experiment to operate continuously for extended periods of time will require significantly re-thinking the computing infrastructure.

Under the current scheme, data are collected in the laboratory and first stored on a small number of local storage drives. The data are then transferred via 10 Gb network connection to the computing cluster at PNNL. Both the network connection and local storage elements are obvious points of potential failure and would be expected to have multiple issues when operating for extended periods of months. This is not a large problem through phase II, since the total operation time is not expected to be very long, and if something fails we can simply correct it and absorb the time required for repairs. For phase III and beyond, we will need to expand the local storage so that interruptions to transfer do not immediately cause substantial backup problems and force halting data collection. Ideally, the system would also be robust against the loss of individual drives and provide a reasonable means for replacing drives without interruption. A rolling set of storage elements, for example, could allow periodic service intervals on each without interrupting data collection.

Furthermore, the computational requirements for both signal combination from multiple channels, and the subsequent triggering logic, will be substantial. Achieving this will likely require dedicated hardware, either in the form of FPGAs, GPU compute servers, traditional CPU compute resources, or a combination of all of these. Given that the bandwidth for transferring data out of the lab cannot be scaled up

significantly, it will become critical that as much processing and data reduction as possible be done on site and in real time. Developing these tools can obviously begin on recorded data, but commissioning and operation will require both the physical resources and software framework to be in place.

The DIRAC framework [44] has been identified as a system for implementing our downstream requirements for data transfer, storage, and cataloging. That system has support for automatic redundancy and integrates with grid computing for data processing, but developing the internal expertise to work efficiently within it will require yet another significant investment of time.

8.1.3 Event reconstruction and analysis

Venturing into the more subjective, there will likely need to be a change to way analyzing the data is approached. In the early phases, the analysis procedures have developed organically and originally were driven by the assumption that it would be difficult to find events at all. As has been seen in the more recent data taking, the bigger problem is actually that events are found, but they can have lots of structure which needs to be identified.

I believe it will be very useful to consider the data processing in a number of stages, which require different procedures and optimizations. For example, the trigger, either online or as the first stage of off-line data process, needs to very reliably and quickly find events, but it is not critical that it capture the details of an event precisely. If we focused on procedures which can very effectively isolate time windows of interest, then we can use much more computationally intensive procedures to extract the details from just those intervals. Similarly, once we have identified some portion of an event, we can use that as an input for determining the full details.

As a concrete example, it is likely that simple FFTs with larger frequency win-

dows will be useful for identifying that an event is present. Once that is identified, however, it may be possible to use a different change of basis, such as a fractional Fourier transform or wavelet transform, to more precisely determine the starting time and frequency of an event. Similarly, if the parameters of an individual track have been found, that could be used to constrain the search region when looking for weak sidebands.

It will be important to continue to develop all stages of data processing, especially in the relatively near term. By improving the procedures used earlier in the signal processing, it will be much easier to understand the later stages, such as converting event details into a reconstructed energy or studying spectroscopic features. These advances will be required soon because understanding factors such as detection efficiency for sidebands and the efficacy of sideband corrections will have a major impact on the achievable solid angle acceptance in the source. Unless the acceptance is sufficiently large, any volume scalability advantages are insufficient for the technique to exceed the capabilities of conventional techniques.

8.1.4 Hardware and design

For each phase of Project 8, there will clearly be major engineering and design efforts required. It will always be necessary to develop source delivery, RF instrumentation, and magnet systems, each increasingly stringent requirements and they must all be mutually compatible. While these efforts are incredibly challenging, they tend to receive much more attention and there is already a number of people working on the early stages of each.

8.2 Outlook

The efforts of the Project 8 collaboration remain a promising means to study the mass of the neutrino. The CRES technique offers an entirely different approach from traditional technologies and we have only scratched the surface of understanding both the potential and the limitations of the method. Moving forward it will become critical that we quantify performance in terms of parameters such as efficiency, resolution, and computation burden, and that all of these are carefully connected to both achievable sensitivity and cost so that a viable final stage can be developed.

Through the previous chapter, and previous section of this chapter, I've presented numerous limitations of our current capabilities but this should not be interpreted as an indictment. The technique is a new one, and the size and quality of the available data is increasing rapidly. As new problems are discovered, new tools are developed to compensate, and a greater understanding of the subtleties of CRES is achieved. Furthermore, while many challeneges remain, there is currently a research path forward on each, which is limited most by the number of people available to work on them.

Appendix A

Digital signal processing example

In this appendix I treat in detail, the signal processing steps of the RSA for a simplified example. These procedures have proven a consistent source of confusion within Project 8, and so I will attempt to clarify in a way that is intuitive for physicists. The parameter values here are chosen to be relevant to the RSA as used by Project 8 so that it is easy to contextualize, these are summarized in table A.1. Also included will be code snippets in the python programming language used to compute the signal state at various stages and shown in the included figures. It is worth noting that the data acquisition system based on ROACH-2, still under development, will utilize similar data processing steps.

A.1 Analog mixing

The analog signal processing within the RSA is functionally equivalent to our custom mixing stages (see sections 3.4 and 5.2.1) and so it does not bare going through in detail. There is no noise included in this example and all gains are taken to be flat. In figure A.1 the signal is shown, in the frequency domain, at the RSA's input

Parameter	**Description**
signal	A sequence of five continuous-wave tones frequencies 1760, 1770, 1780, 1790, 1800 MHz with amplitudes 0.1, 0.08, 0.06, 0.04, 0.02
input frequency band	85 MHz wide: [1737.5, 1822.5]
analog mixing frequency	1550 MHz
ADC sampling rate	300 MS/s
record duration	100 μs
digital mixing frequency	230 MHz

Table A.1: A summary of the values being used for this particular example. They were selected to be representative values for the Project 8 RSA configuration.

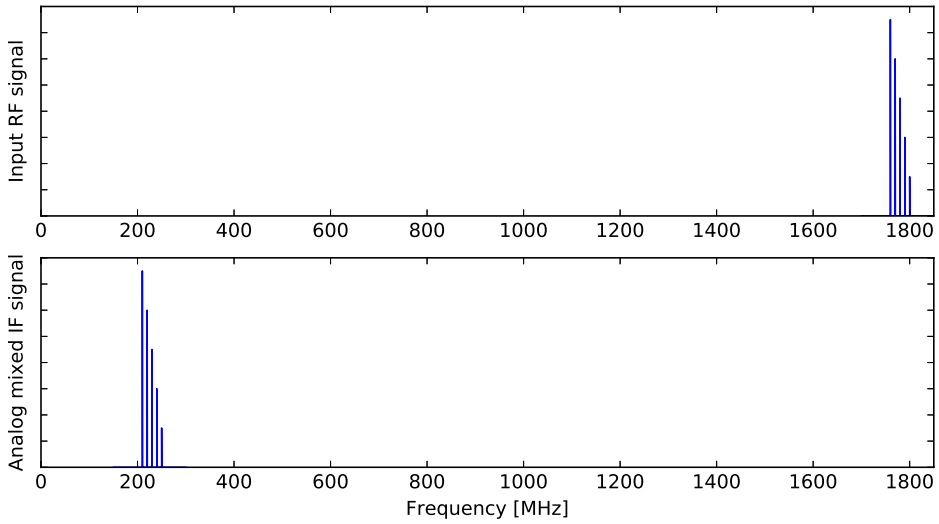


Figure A.1: The test signal at the analog input (above) and after analog mixing (below). In the actual signal processing chain both signals are still analog, though here the signal is represented in the discrete frequency domain. Note also that noise has not been included and the gain has been taken to be unity for all frequencies. The physical system obviously would have noise power and each component contributes a frequency-dependent gain, for which the RSA attempts to correct.

and after internal analog mixing into the input band of the digital stream.

The signal is computed as the sum of five sine waves with frequencies from 210 MHz to 250 MHz in 10 MHz steps and amplitudes which decrease with increasing frequency. In the actual signal processing, the signal is analog at this point, though obviously here it is necessary to use a discrete time representation. Populating the intermediate frequency values in the time domain is demonstrated in the following code snippet.

```
import numpy as np

sampling_rate = 300e6
times = np.arange(0, 100e-6, 1./sampling_rate)
frequency = np.linspace(210e6, 250e6, 5)
power = np.linspace(0.1, 0, len(frequency), endpoint=False)
signal = np.zeros(times.size)

for f,p in zip(frequencies, powers):
    signal += p * np.sin(2 * np.pi * f * t)
```

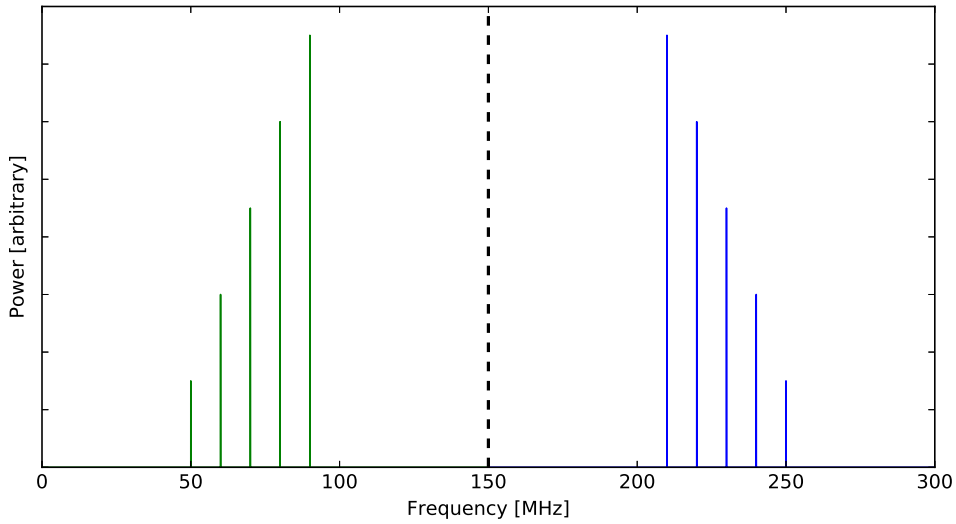


Figure A.2: The example signal after it has been discretely sampled in time domain, displayed in frequency domain using a discrete Fourier transform. The signal in blue is the input to the digitizer, note that all five tones are between the Nyquist frequency (150 MHz) and the sampling frequency (300 MHz), while in green is the result of aliasing, which shows up as a reflection across the Nyquist frequency.

A.2 Sampling and aliasing

The first subtlety comes when digitizing the signal. The ADC samples at 300 MS/s, but the analog mixing stage has placed the signal of interest not into the first Nyquist zone (between zero and half of the sampling frequency), but instead into the second Nyquist zone (between half of the sampling frequency and the full sampling frequency). The process of discretely sampling the data aliases any signal present in one Nyquist zone into all others, with the image in even and odd zones being reflections across the zone boundaries. As described earlier (see 5.2.2), this is done because it is not possible to design a reliable system that mixes all the way to baseband and has a wide output bandwidth. This procedure follows Nyquist theorem, which tells us only a bandwidth of half the sampling frequency can be faithfully represented in the data; here we are making a choice only of which band that is applied to. In this example, a sampling frequency of 300 MS/s means there are 150 MHz of bandwidth. Figure A.2 shows the input signal to the ADC and the aliased version from sampling, along with the Nyquist zone boundary.

A.3 Digital mixing

After the signal has been digitized, the practical constraints of analog mixing do not apply, and the precise analysis band of interest can be mixed down to be centered

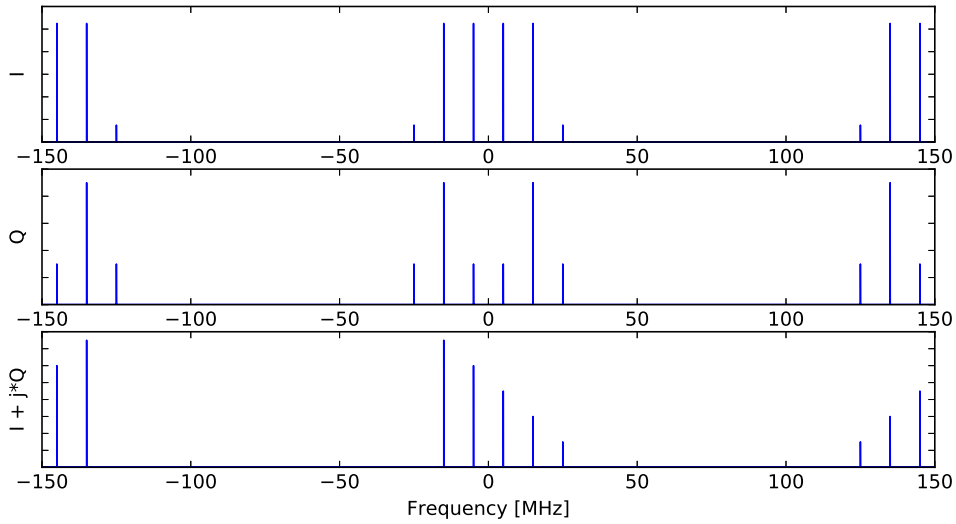


Figure A.3: Frequency domain representation of the I (top), Q (middle), and complex $\mathrm{I} + i\mathrm{Q}$ (bottom) time series. Note that both I and Q are even, and that both the region outside $\pm\frac{1}{4}$ of the sampling is a reflection of the data within that region. The complete complex data faithfully represents the original signal in that same region about the origin and has the same reflections in the region outside of $\pm\frac{1}{4}$ of the sampling frequency.

at zero frequency. The samples from the ADC are multiplied by two numerically synthesized signals, a sine and a cosine, each with frequency equal to the center of the bandwidth of interest in the digitizer's input. In this example, the bandwidth of interest, from 187.5 to 272.5 MHz, is centered at 230 MHz.

The result of multiplication with the generated sine and cosine are conventionally referred to as I and Q, and in this example are computed per the following code snippet. These can be considered as the real and imaginary parts of a complex time series. It is worth stressing that the signal values, generated mixing signals, and resulting I and Q values, are all computed sequentially in the time domain. The frequency domain representation of each is shown in figure A.3.

```
f_lo = 230e6
I = signal * np.sin(2 * np.pi * f_lo * t)
Q = signal * np.cos(2 * np.pi * f_lo * t)
IQ = I + 1j * Q
```

A.4 Downsampling

The I and Q values contain all of the information from the original data, but everything is stored redundantly and the size is twice that of the original input.

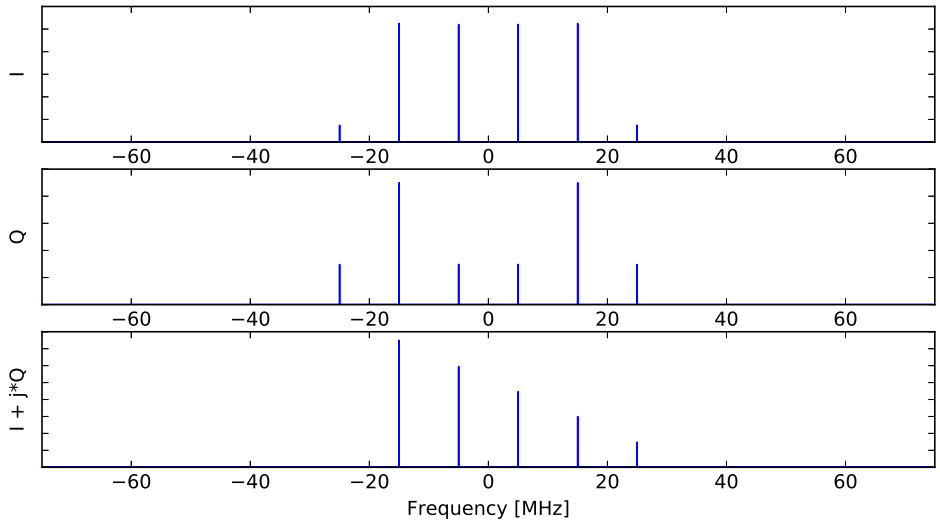


Figure A.4: I (top), Q (middle), and $\mathrm{I} + i\mathrm{Q}$ (bottom) represented in the frequency domain, after filtering and decimation. Note that the bandwidth has been reduced by half, but that the content in the central region is exactly the same as in figure A.3.

Considered in the frequency domain, each of I and Q contains useful information between negative and positive one quarter of the sampling frequency, with the rest of the bandwidth containing an aliased image of the same data. This is, therefore, still 150 MHz of faithfully stored bandwidth, consistent with the Nyquist theorem. Because all of the redundant information is stored in higher magnitude frequencies, reducing the effective sampling rate by decimating the signal by a factor of two will remove only the redundant information. As with sampling the analog signal, however, this process will alias signals across the Nyquist zone boundary of each time series. Having the signal bandwidth centered at zero allows this to be removed using a time domain low-pass filter. This reduces the size of the data being stored back to the same as the original digitizer output. For this example, I use the decimate function provided in the signal submodule of the scipy package, which includes an 8th order Chebyshev type 1 filter (see the scipy documentation [45]). The post-decimation representation of I, Q, and the complex combination, in frequency domain, is shown in figure A.4.

```
from scipy.signal import decimate
Id = decimate(I, 2, zero_phase=True)
Qd = decimate(Q, 2, zero_phase=True)
IQd = decimate(IQ, 2, zero_phase=True)
```

It is worth stressing once more that all digital processing steps are completed sequentially in the time domain. Further, the only processing step which requires using

more than one time step at the same time is the low-pass filter prior to decimation, which can be computed using a sliding window. This is important because it means that computationally intensive discrete Fourier transforms are not required, and there is no windowing which must be matched in subsequent or offline analysis.

Bibliography

[1] Franklin, A. *Are There Realy Neutrinos? An Evidential History* (Westview Press, 2004).

[2] Caldwell, D. O. (ed.) *Current Aspects of Neutrino Physics* (Springer Berlin Heidelberg, 2001).

[3] Patrignani, C. *et al.* Review of particle physics. *Chin. Phys.* **C40** (2016).

[4] Brown, L. M. The idea of the neutrino. *Physics Today* **31**, 23–28 (1978). URL `http://physicstoday.scitation.org/doi/abs/10.1063/1.2995181`.

[5] Planck Collaboration *et al.* Planck 2015 results - xiii. cosmological parameters. *AA* **594**, A13 (2016). URL `http://dx.doi.org/10.1051/0004-6361/201525830`.

[6] Bhupal Dev, P. S., Goswami, S., Mitra, M. & Rodejohann, W. Constraining neutrino mass from neutrinoless double beta decay. *Phys. Rev. D* **88**, 091301 (2013). URL `http://link.aps.org/doi/10.1103/PhysRevD.88.091301`.

[7] Angrik, J. *et al.* KATRIN design report 2004. `https://www.katrin.kit.edu/publikationen/DesignReport2004-12Jan2005.pdf` (2005).

[8] Gastaldo, L. *et al.* The electron capture ^{163}Ho experiment ECHo. *Journal of Low Temperature Physics* **176**, 876–884 (2014). URL `http://dx.doi.org/10.1007/s10909-014-1187-4`.

[9] Alpert, B. *et al.* HOLMES: The electron capture decay of ^{163}Ho to measure the electron neutrino mass with sub-ev sensitivity. *The European Physical Journal C* **75**, 112 (2015). URL `http://dx.doi.org/10.1140/epjc/s10052-015-3329-5`.

[10] Wilkerson, J. F. & Robertson, R. G. H. *Direct Measurements of Neutrino Mass*, 39–64 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2001). URL `http://dx.doi.org/10.1007/978-3-662-04597-8_3`.

[11] Kraus, Ch *et al.* Final results from phase II of the Mainz neutrino mass searchin tritium β decay. *Eur. Phys. J. C* **40**, 447–468 (2005). URL `http://dx.doi.org/10.1140/epjc/s2005-02139-7`.

[12] Aseev, V. N. *et al.* Upper limit on the electron antineutrino mass from the Troitsk experiment. *Phys. Rev. D* **84**, 112003 (2011). URL http://link.aps.org/doi/10.1103/PhysRevD.84.112003.

[13] Otten, E. W. & Weinheimer, C. Neutrino mass limit from tritium decay. *Reports on Progress in Physics* **71**, 086201 (2008). URL http://stacks.iop.org/0034-4885/71/i=8/a=086201.

[14] Bodine, L. I., Parno, D. S. & Robertson, R. G. H. Assessment of molecular effects on neutrino mass measurements from tritium β decay. *Phys. Rev. C* **91**, 035505 (2015). URL http://link.aps.org/doi/10.1103/PhysRevC.91.035505.

[15] Monreal, B. & Formaggio, J. A. Relativistic cyclotron radiation detection of tritium decay electrons as a new technique for measuring the neutrino mass. *Phys. Rev. D* **80**, 051301 (2009). URL http://link.aps.org/doi/10.1103/PhysRevD.80.051301.

[16] Jackson, J. D. *Classical Electrodynamcis* (Wiley, New York, NY, 1999), 3rd edn.

[17] Brown, L. S. & Gabrielse, G. Geonium theory: Physics of a single electron or ion in a penning trap. *Rev. Mod. Phys.* **58**, 233–311 (1986). URL http://link.aps.org/doi/10.1103/RevModPhys.58.233.

[18] Clark, B. M. *Magnetic Trapping of Atomic Tritium for Neutrino Mass Measurement*. B.S. thesis, California Institute of Technology (2014). URL http://resolver.caltech.edu/CaltechTHESIS:07252014-082021105.

[19] Picard, A. *et al.* Precision measurement of the conversion electron spectrum of ^{83m}Kr with a solenoid retarding spectrometer. *Zeitschrift für Physik A Hadrons and Nuclei* **342**, 71–78 (1992). URL http://dx.doi.org/10.1007/BF01294491.

[20] Vénos, D., Dragoun, O., Špalek, A. & Vobecký, M. Precise energy of the weak 32-kev gamma transition observed in 83mkr decay. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **560**, 352–359 (2006). URL http://www.sciencedirect.com/science/article/pii/S0168900205026252.

[21] Purcell, J. & Sheu, C. Nuclear data sheets for a = 3. *Nuclear Data Sheets* **130**, 1–20 (2015). URL http://www.sciencedirect.com/science/article/pii/S0090375215000551.

[22] Vénos, D., Špalek, A., Lebeda, O. & Fišer, M. Kr radioactive source based on rb trapped in cation-exchange paper or in zeolite. *Applied Radiation and Isotopes* **63**, 323–327 (2005). URL http://www.sciencedirect.com/science/article/pii/S0969804305001260.

[23] Kofron, J. *A novel method for electron energy measurement: Cyclotron Radiation Emission Spectroscopy.* PhD dissertation, University of Washington (2015). URL https://digital.lib.washington.edu/researchworks/handle/1773/33235.

[24] Low Noise Factory. LNF-LNC22_40WA s/n 010: 22-40 ghz cryogenic low noise amplifier (2014). Product data sheet.

[25] ISO/IEC 19464: Information technology - advanced message queuing protocol (amqp) v1.0 specification (2014). URL http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=64955.

[26] Pivotal Software. The rabbitmq homepage. https://www.rabbitmq.com.

[27] Ecma-404: The json data interchange format (2013). URL http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf.

[28] Klyne, G. & Newman, C. Date and time on the internet: Timestamps. https://www.ietf.org/rfc/rfc3339.txt (2002).

[29] Slack Technologies, Inc. Getting started with custom integrations. https://api.slack.com/custom-integrations.

[30] The Project 8 Collaboration. Mantis source repository. https://www.github.com/project8/mantis.

[31] The Project 8 Collaboration. Monarch source repository. https://www.github.com/project8/monarch.

[32] Asner, D. M. *et al.* Single-electron detection and spectroscopy via relativistic cyclotron radiation. *Phys. Rev. Lett.* **114**, 162501 (2015). URL http://link.aps.org/doi/10.1103/PhysRevLett.114.162501.

[33] Tektronix. Fundamentals of real-time spectrum analysis. info.tek.com/www-fundamentals-of-real-time-spectrum-analysis.html (2009).

[34] The Project 8 Collaboration. Mantis source repository. https://www.github.com/project8/katydid.

[35] Protocol buffers. https://developers.google.com/protocol-buffers/.

[36] The HDF Group. The hdf5 home page. https://support.hdfgroup.org/HDF5/.

[37] Frigo, M. & Johnson, S. G. The design and implementation of fftw3. *Proceedings of the IEEE* **93**, 216–231 (2005). URL http://ieeexplore.ieee.org/document/1386650.

[38] Bentley, J. L. Multidimensional binary search trees used for associative searching. *Commun. ACM* **18**, 509–517 (1975). URL http://doi.acm.org/10.1145/361002.361007.

[39] Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231 (Portland, OR, 1996). URL http://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf.

[40] Hough, P. V. C. Method and means for recognizing complex patterns (1962). US Patent 3069654 A.

[41] Zayas, E. Update on sideband processing analysis. https://basecamp.com/1780990/projects/338746/messages/64151630?enlarge=259644636#attachment_259644636 (2016). Project 8 internal report.

[42] Saldaña, L. Cataloging project 8 events: Phase 1 bathtub data. https://basecamp.com/1780990/projects/338746/messages/63972963?enlarge=258723243#attachment_258723243 (2016). Project 8 internal report.

[43] Ashtari Esfahani, A. Bathtub trap power spectrum. https://basecamp.com/1780990/projects/338746/messages/64359836?enlarge=260800910#attachment_260800910 (2016). Project 8 internal report.

[44] Tsaregorodtsev, A. *et al.* Dirac: a community grid solution. *Journal of Physics: Conference Series* **119**, 062048 (2008). URL http://stacks.iop.org/1742-6596/119/i=6/a=062048.

[45] The Scipy community. scipy.signal.decimate. https://docs.scipy.org/doc/scipy-0.16.1/reference/generated/scipy.signal.decimate.html.