



**Detection and Estimation Limits of Single
Electron Cyclotron Radiation with
Phased Array Antennas**

Dissertation zur Erlangung des Grades

DOKTOR DER NATURWISSENSCHAFTEN

am Fachbereich Physik, Mathematik und Informatik der
Johannes Gutenberg-Universität in Mainz

Florian Thomas

geboren am 23.05.1992 in Oberwesel

Mainz, den 11. Oktober 2024

Florian Thomas

Detection and Estimation Limits of Single Electron Cyclotron Radiation with Phased Array Antennas

Dissertation zur Erlangung des Grades DOKTOR DER NATURWISSENSCHAFTEN

1. Berichterstatter: [In der elektronischen Version aus Datenschutzgründen entfernt]

2. Berichterstatter: [In der elektronischen Version aus Datenschutzgründen entfernt]

Johannes Gutenberg Universität-Mainz

Datum der mündlichen Prüfung: 02. Juli 2025

Kurzfassung

Cyclotron Radiation Emission Spectroscopy (CRES) ist eine neue Technik zur Energiespektroskopie geladener Teilchen mit einer Auflösung im Sub-eV-Bereich. Mit CRES werden Teilchenenergien durch Messung der Frequenz der emittierten Zyklotronstrahlung in einem Magnetfeld bestimmt. Diese kann von einer Reihe von Radioantennen erfasst und in digitale Spannungen umgewandelt werden. Experimente, die große Volumina von Quellmaterial benötigen, um ausreichende statistische Daten zu sammeln, könnten potentiell Hunderte bis Tausende dieser Antennen erfordern, was pro Sekunde Hunderte Gigabyte an Rohdaten erzeugen könnte. Dies erfordert effiziente Lösungen zur Datenanalyse und -reduktion. In dieser Arbeit werden Algorithmen zur Detektion und Parameterschätzung dieser Signale in verrauschten Detektordaten diskutiert. Der Schwerpunkt liegt dabei auf der Entwicklung und Effizienzanalyse optimaler Lösungen unter der Annahme unbegrenzter Rechenressourcen, insbesondere durch Matched Filtering und Maximum-Likelihood-Schätzung. Diese optimalen Lösungen sind von entscheidender Bedeutung, da ihre Grenzen für die Detektionseffizienz und die Energieauflösung Schlüsselfaktoren sowohl für die Sensitivitätsschätzung als auch für die Optimierung der experimentellen Designs sind. Zur Unterstützung dieser Arbeit wurde ein neues Simulationstool entwickelt, das durch Fortschritte bei der physikalischen Modellierung von CRES-Signalen eine 500-fache Beschleunigung im Vergleich zur Basissimulationssoftware erreicht. Dieses Simulationstool wurde zusammen mit der Implementierung der optimalen Lösungen eingesetzt, um die Detektionseffizienz und die Energieauflösung eines idealisierten Beispielaufbaus für eine Neutrinomassenmessung mit Tritium-Elektronenspektroskopie im Project 8 Experiment zu bewerten. Dies ist das erste Mal, dass eine Untersuchung der theoretischen Grenzen eines konkreten Experimentaufbaus auf der Grundlage eines großen Antennen Arrays durchgeführt werden konnte, was erst durch die verbesserte Simulationslaufzeit ermöglicht wurde. Eine erste Bewertung deutet darauf hin, dass die Ergebnisse mit den Anforderungen übereinstimmen, um das Ziel von Project 8 zu erreichen, für Neutrinomassen mit $m_\beta > 0.04 \text{ eV}$ empfindlich zu sein.

Abstract

CRES is a new technique for energy spectroscopy of charged particles with sub-eV resolution. With CRES, particle energies are determined by measuring the frequency of the emitted cyclotron radiation in a magnetic field, which can be detected by an array of radio antennas and converted into digital voltages. Experiments that require large volumes of source material for gathering sufficient statistics may potentially need hundreds to thousands of these antennas, which can generate hundreds of gigabytes of raw data per second. This necessitates efficient data analysis and reduction solutions. In this thesis, algorithms for detection and parameter estimation of these signals within detector noise are discussed. The focus is on the development and efficiency analysis of optimal solutions assuming unlimited computing resources, specifically through matched filtering and maximum likelihood estimation. These optimal solutions are critical because their limits on detection efficiency and energy resolution are key factors in both sensitivity estimation and the optimization of experimental designs. To support this work, a new simulation tool was developed, achieving a 500 times speedup compared to the baseline simulation software through advancements in the physical modeling of CRES signals. This simulation tool was utilized, alongside the implementation of the optimal solutions, to evaluate the detection efficiency and energy resolution of an idealized example setup for a neutrino mass measurement with tritium β -decay electron spectroscopy in the Project 8 experiment. This represents the first time a study of the theoretical limits of a concrete experimental design based on a large antenna array could be conducted, made possible by the improved simulation run time. An initial assessment suggests that the results are consistent with the requirements for achieving the project goal of being sensitive to neutrino masses with $m_\beta > 0.04$ eV.

Acknowledgement

[In der elektronischen Version aus Datenschutzgründen entfernt]

Author's Contributions

In this section, I want to clarify my personal contributions to the work presented in this thesis. Since I conducted my research as a member of the Project 8 collaboration, I cannot claim full credit for all the ideas, work, and results included in this thesis. While chapters 2 and 3 clearly are not my own work, as they provide background information to contextualize the research, the remaining chapters of the thesis body contain my own contributions, which are detailed as follows:

- **Chapter 4:** The development of the required theory for the antenna array signals is in most parts a compilation of existing literature results. By combining a number of different sources I built a cohesive theory that does not exist in literature in this form. I deduced new expressions like the explicit results for the drift motion and most importantly the approximation of the fundamental frequency of the radiation field with relativistic correction. The convenient mathematical framework to describe rotationally symmetric magnetic fields was suggested by [REDACTED].
- **Chapter 5:** The source code for the simulation tool CRESana was almost exclusively written by myself. The only exception are small bug fixes and additions of smaller features by [REDACTED]. For the development I took inspiration from the existing antenna array simulation in Locust. In particular the original model for the antenna response simulation, which was developed by [REDACTED], [REDACTED], [REDACTED], and [REDACTED], is included in CRESana except for the FIR filter. All critical changes in the modeling technique resulting in the performance gains were my own ideas and rely on the theory from chapter 4. Due to these changes, most of the simulation components are implemented with approaches that differ from those in Locust.
- **Chapter 6:** The beamforming algorithm was first implemented in Project 8 software by [REDACTED]. The Matched Filter concept was introduced to Project 8 by [REDACTED]. Most of the pioneering research on the applicability and limitations of both concepts for CRES was conducted collaboratively by [REDACTED], [REDACTED], and myself. The drift motion extension of the beamforming algorithm was developed by [REDACTED] and myself in parallel and independent of each other. Nevertheless, all presented results

were produced with code I wrote myself. The contributions that I can fully claim as my own are the mathematical proof that the beamforming algorithm is a matched filter variant, the introduction of the calculation of the detection efficiency without extensive monte carlo simulation, and the analysis of the computational cost for a real time application.

- **Chapter 7:** Most of the work in this chapter is based on established methods and standard software packages used in particle physics. My contributions are their first application to CRES with a suitable likelihood function that I formulated, as well as an investigation of the challenges for this application. The Cramér–Rao lower bound for the chirp model that I use as a comparison was derived by ██████████. I only corrected a mistake in the calculations in his thesis and also derived the Cramér–Rao lower bound for a slightly modified model based on his calculations.
- **Chapter 8:** The effective volume calculation used to compare results to Project 8 goals follows the method established by ██████████. Beyond that, I can take full credit for developing the method for analyzing the detection and estimation limits of an experiment design and for executing it on a specific example. For the setup of the examined example design I had help from ██████████ on various aspects of the antenna array, while the trap was implemented by ██████████. ██████████ first discovered the unacceptable correlation between energy and pitch angle in certain magnetic traps, which I confirmed as a general feature of harmonic field profiles. With the presented investigation of the different polynomial field profiles I established an essential trap design directive to mitigate this issue. The final trap utilized in my example was developed by ██████████, based on this directive.

Declaration on the use of AI tools

In this thesis, I have used ChatGPT as a writing assistant for proofreading and polishing of my manuscript.

Contents

Acknowledgement	vii
Author's Contributions	ix
List of Figures	xv
List of Tables	xxi
Acronyms	xxiii
1. Introduction	1
2. Neutrinos and Neutrino Mass	5
2.1. Neutrino Oscillation	5
2.2. Neutrino Mass Measurement	7
2.2.1. Kinematic Approach	7
2.2.2. The KATRIN Experiment	10
2.2.3. Other Approaches	13
3. The Project 8 Experiment	17
3.1. Cyclotron Radiation Emission Spectroscopy	18
3.1.1. CRES Detectors	18
3.1.2. Advantages of CRES	21
3.2. Project 8 Status	22
4. Theory of CRES Signals	25
4.1. Electron Motion	25
4.1.1. Relativistic Charged Particle Motion	25
4.1.2. Cyclotron Motion in Uniform Magnetic Fields	26
4.1.3. Motion in a Magnetic Trap	27
4.1.4. Drift Motions	30
4.2. Magnetic Fields with Rotational Symmetry	31
4.2.1. General Mathematical Description	31
4.2.2. Drift Motion in Rotationally Symmetric Fields	33

4.3.	Electromagnetic Waves from CRES Electrons	35
4.3.1.	General Solution	35
4.3.2.	Spectral and Spatial Power Distribution	36
4.3.3.	Polarization and Phase	42
5.	Simulated Data of an Antenna Array CRES Experiment	47
5.1.	Overview of Simulation Principles	47
5.2.	Kassiopeia	50
5.3.	Locust	51
5.4.	CRESana	52
5.4.1.	Electron Trajectory	52
5.4.2.	Electric Field at Antenna	61
5.4.3.	Antenna Response	64
5.4.4.	RF Receiver	69
5.4.5.	Validation	71
5.4.6.	Performance Benchmarks	74
5.4.7.	Future Developments	77
5.5.	Spectral Features of Simulated Signals	78
5.5.1.	Sidebands	79
5.5.2.	Constant Doppler Shifts	86
5.5.3.	Linear Frequency Chirp	87
6.	Detection of CRES Signals in Noise	89
6.1.	Basics of Signal Detection	91
6.2.	Digital Beamforming	93
6.2.1.	Mathematical Formulation	94
6.2.2.	The Test Statistic	97
6.2.3.	Extension for Drift Motion	99
6.2.4.	Axial Motion	104
6.3.	Matched Filter	105
6.3.1.	Multi-Channel Case	106
6.3.2.	Unknown Parameters	108
6.3.3.	Signal Model	113
6.3.4.	Signal to Noise Ratio	116
6.3.5.	Detection Performance	118
6.4.	Real Time Detection	123
6.4.1.	Algorithmic Choice	124
6.4.2.	Estimation of the Computing Requirements	125
6.4.3.	Potential Improvements of Computational Performance	128

6.5. Summary	129
7. Parameter Estimation of CRES Signals	131
7.1. Maximum Likelihood Estimation	131
7.1.1. Uncertainties	134
7.2. Determination of Estimation Limits	137
7.2.1. Analytic Approach: Cramér–Rao Lower Bound	138
7.2.2. Numeric Approach: Asimov Dataset	140
7.2.3. Validation of Gaussian Assumption for the Numeric Approach	142
7.2.4. Implementation of Numeric Approach for the Chirp Model . .	145
7.2.5. Results of the Chirp Model Analysis	151
7.3. Summary	156
8. Assessing the Limits of a Full Scale Detector: A Case Study	159
8.1. Experimental Setup	159
8.1.1. Antenna Array	159
8.1.2. Magnetic Trap	162
8.1.3. Effective Trap Volume	166
8.2. Energy Resolution	167
8.2.1. Likelihood Function and Relevant Parameters	167
8.2.2. Parameter Dependencies of the Energy Resolution	169
8.2.3. Result	171
8.3. Detection Efficiency	177
8.3.1. Parameter Dependencies	177
8.3.2. Result	177
8.4. Implications and Limitations of Results	182
8.4.1. Validation of Hesse Errors	182
8.4.2. Track Length Scaling and CRLB	183
8.4.3. Neutrino Mass Sensitivity	184
8.5. Summary	184
9. Conclusion and Outlook	187
Bibliography	191
A. Magnetic Fields	201
A.1. The Harmonic Trap	201
A.2. The Square Well Trap	204
A.3. The Bathtub Trap	205

B. Bugs in Locust	209
B.1. Applying Antenna Response to Undersampled Signals	209
B.2. Incorrect Field Results	211
C. Additional Proofs and Calculations	213
C.1. Calculation of the Orthogonal Gradient for the Drift Motion	213
C.2. Proofs for Matched Filter SNR	215
C.3. Calculation of Effective Trap Volume	216
C.4. CRLB of a Chirp with Dependent Chirp Rate	218
D. Additional Plots	221
D.1. Chirp Model Likelihood Scans	221
D.2. Comparison of High and Low Gain Antennas	224
D.3. Full Scale Detector Likelihood Scans	225
Lebenslauf	231

List of Figures

2.1.	Energy spectrum of electrons produced in β -decay of molecular tritium.	8
2.2.	Mass scale of the observable m_β as a function of the sum of the neutrino mass eigenstates m_i .	9
2.3.	Schematic of tritium β spectroscopy with a MAC-E filter in the KATRIN experiment.	11
2.4.	Illustration of a molecular tritium decay	13
2.5.	Feynman diagram for neutrinoless double β -decay.	14
3.1.	Mass scale of the observable m_β as a function of the sum of the neutrino mass eigenstates m_i with limits.	17
3.2.	Schematic of a CRES experiment with a phased array antenna as detector.	19
3.3.	Schematic of a CRES experiment with a cavity as detector.	21
4.1.	Magnetic mirror trap with field lines.	28
4.2.	Visualization of the retarded time.	36
4.3.	Angular coordinates of the observer's orientation relative to the magnetic field direction and cyclotron orbit plane.	37
4.4.	Power spectra of electrons in a 1 T magnetic field with pitch angle $\alpha = 90^\circ$ observed with $\theta = 90^\circ$.	38
4.5.	Angular distribution of the total radiated power of electrons undergoing cyclotron motion.	39
4.6.	Power spectra of electrons in a 1 T magnetic field with pitch angle $\alpha = 90^\circ$.	40
4.7.	Angular distribution of the power radiated in the fundamental frequency.	42
4.8.	Observed polarization ellipses in the ϕ - θ -plane for different θ	45
4.9.	Phase of CRES radiation fields.	46
5.1.	Electron track simulation steps.	48
5.2.	Chart of multiple concurrent tracks and events.	49
5.3.	3D electron motion	54
5.4.	Plot of magnetic field lines in a trapping field.	55
5.5.	Construction of the axial solution.	57
5.6.	Approach for calculating the axial motion on the first quarter period.	58
5.7.	Visualization of retarded time calculation.	61
5.8.	Simulated frequency response of the Project 8 antenna.	66

5.9.	CREsana simulated gain patterns of a single element of the Project 8 antenna.	67
5.10.	E- and H-plane of antennas in the cylindrical array.	68
5.11.	Simulated gain patterns of the Project 8 antenna.	69
5.12.	Diagram of a direct conversion receiver which produces in-phase and quadrature (I/Q) data with a frequency mixer.	70
5.13.	Comparison of trajectories of CREsana and Kassiopeia	72
5.14.	Comparison of the cyclotron frequency in Kassiopeia and CREsana. . .	73
5.15.	Simulated antenna power depending on the angle θ	74
5.16.	Benchmark results for peak memory and runtime depending on the number of antenna channels for different radii and pitch angles with a fixed trajectory length and number of samples.	75
5.17.	Benchmark results for peak memory and runtime depending on the number of time samples (i.e. depending on trajectory length τ) for different radii and pitch angles with a fixed number of antenna channels.	76
5.18.	Spectra of electron signals with $E_{\text{kin}} = 18.6 \text{ keV}$ in a harmonic trap taken with the Project 8 prototype antenna.	79
5.19.	Spectra of AM and FM signals.	82
5.20.	Spectra for $\alpha_0 = 87.5^\circ$ and $E_{\text{kin}} = 18.6 \text{ keV}$ with the five-slot antenna in a harmonic and a square well trap.	84
5.21.	Spectra of an electron with and without drift motion.	86
5.22.	Time frequency spectrogram of a signal with sidebands.	87
6.1.	Noise power spectrum compared to signal power spectra.	90
6.2.	Antenna array generating data and signal detection.	91
6.3.	Probability density functions of $T(\mathbf{x})$	92
6.4.	Schematic for beamforming with a ring of antennas.	94
6.5.	Results of beamforming with a simulated $\alpha_0 = 90^\circ$ electron with different positions.	98
6.6.	Comparison of power spectra in beamforming.	99
6.7.	Beamforming result of an electron experiencing drift motion.	100
6.8.	Beamforming result with drift motion tracking.	102
6.9.	Comparison of Beamforming results with and without drift motion tracking.	103
6.10.	Power spectra of beamformed signals with varying pitch angles.	105
6.11.	Calculation of the test statistic with the multi-channel matched filter. . .	107
6.12.	ROC curves of the matched filter for different SNRs.	118
6.13.	ROC curves of the matched filter and quadrature matched filter from theory and MC experiments.	119

6.14.	ROC curves of template banks using the normalized quadrature matched filter and the generalized likelihood ratio test.	120
6.15.	Theoretical ROC curves of the quadrature matched filter template bank compared to MC results.	122
6.16.	Arithmetic bound and memory size bound depending on n_s and n_b	127
6.17.	Estimated number of required GPUs for an MF template bank.	128
7.1.	Calculation of the likelihood of CRESana data.	132
7.2.	Log-likelihood function of a 1-D estimation problem with Gaussian noise.	135
7.3.	Calculation of confidence regions in a 2-dimensional estimation problem.	136
7.4.	Comparison of Hesse and Minos errors.	137
7.5.	Determination of estimation limits with an Asimov data set compared to MC experiments.	142
7.6.	Demonstration of proposed Gaussian validation.	144
7.7.	Asimov log-likelihood of a chirp plotted for all coordinate planes.	146
7.8.	Comparison of different 1D likelihood functions of the chirp model.	147
7.9.	Asimov log-likelihood of a chirp with modified log-likelihood function.	148
7.10.	High resolution scans of likelihood functions for the chirp model along t_0 axis.	149
7.11.	Application of window function to the chirp model.	150
7.12.	Scan of modified asimov log-likelihood for the chirp model with high SNR along t_0 axis.	151
7.13.	Minuit results for Δf of a chirp with parameters corresponding to the detection of 1% of the electron's Larmor power in $B = 0.96$ T with $T = 5$ K noise.	153
7.14.	Minuit results for Δf of a chirp with parameters corresponding to the detection of 100% of the electron's Larmor power in $B = 0.96$ T with $T = 5$ K noise.	154
7.15.	Minuit results for Δf of a chirp with parameters corresponding to the detection of 100% of the electron's Larmor power in $B = 0.04$ T with $T = 5$ K noise.	155
8.1.	Likelihood scans in $E_{\text{kin}}-\alpha_0$ -plane with CRESana and polynomial magnetic fields	163
8.2.	Magnetic trap for the demo setup.	165
8.3.	Illustration of distances between an electron and antennas in a rotationally symmetric array, in x - y -plane.	169
8.4.	Energy resolution of a chirp for a $\alpha_0 = 90^\circ$ electron plotted as a function of the kinetic energy.	170

8.5.	Results of energy resolution analysis of the demo setup as a function of track length.	172
8.6.	2D scans of the likelihood function of the demo setup for $\alpha_0 = 90^\circ$, $\tau = 240 \mu\text{s}$	173
8.7.	2D scans of the likelihood function of the demo setup for $\alpha_0 = 90^\circ$, $\tau = 240 \mu\text{s}$	174
8.8.	Normalized SNR as a function of radius and pitch angle	178
8.9.	Schematic of a non-isotropic antenna array's sensitivity to radiation in the cross-sectional plane.	179
8.10.	Detection probability as a function of track length and false alarm probability of simulated electrons in the demo setup.	180
8.11.	Detection probability as a function of track length and background rate of simulated electrons in the demo setup.	181
A.1.	Field profile of a harmonic trap off axis.	201
A.2.	Field profile of single coil and harmonic approximation on axis.	202
A.3.	Field profile of single coil and harmonic approximation off axis.	203
A.4.	Field profile of a square well trap.	205
A.5.	Field profile of a bathtub trap on axis.	206
A.6.	Field profile of a bathtub trap off axis.	207
A.7.	Field profile of realistic trap with harmonic approximation.	208
B.1.	Filtering of the sampled field signal.	210
B.2.	Aliased harmonic inside a Locust spectrum.	211
B.3.	Comparison of angular power distribution and power spectrum of field solutions between Locust and theory.	212
D.1.	Asimov log-likelihood of a chirp with modified log-likelihood function $B = 0.96 \text{ T}$, $\tau = 100 \mu\text{s}$	221
D.2.	Asimov log-likelihood of a chirp with modified log-likelihood function $B = 0.96 \text{ T}$, $\tau = 1000 \mu\text{s}$	222
D.3.	Asimov log-likelihood of a chirp with modified log-likelihood function $B = 0.04 \text{ T}$, $\tau = 1000 \mu\text{s}$	223
D.4.	Comparison of energy resolution and detected power for high and low gain antennas.	224
D.5.	2D scans of the likelihood function of the demo setup for $\alpha_0 = 90^\circ$, $r = 0 \text{ m}$, $\tau = 671 \mu\text{s}$	225
D.6.	2D scans of the likelihood function of the demo setup for $\alpha_0 = 90^\circ$, $r = 1.09 \text{ m}$, $\tau = 671 \mu\text{s}$	226

D.7.	2D scans of the likelihood function of the demo setup for $\alpha_0 = 82.17^\circ$, $r = 0 \text{ m}$, $\tau = 671 \mu\text{s}$	226
D.8.	2D scans of the likelihood function of the demo setup for $\alpha_0 = 82.17^\circ$, $r = 1.09 \text{ m}$, $\tau = 671 \mu\text{s}$	227

List of Tables

2.1. Summary of neutrino mass limits.	16
8.1. Configuration parameters of the electromagnets that produce the trapping field	165

Acronyms

ADC analog-to-digital converter

AM amplitude modulation

BLAS Basic Linear Algebra Subprograms

CMB cosmic microwave background

CNN convolutional neural network

CPU central processing unit

CRES Cyclotron Radiation Emission Spectroscopy

CRLB Cramér–Rao lower bound

DBF digital beamforming

DFT discrete Fourier transform

EM electromagnetic

FIR finite impulse response

FM frequency modulation

FMA fused multiply-add

GLRT generalized likelihood ratio test

GPS Global Positioning System

GPU graphics processing unit

HFSS High Frequency Structure Simulator

I/Q in-phase and quadrature

IO inverted ordering

KATRIN Karlsruhe Tritium Neutrino

LO local oscillator

LW Liénard–Wiechert

MAC-E Magnetic Adiabatic Collimation with Electrostatic

MC Monte Carlo

MF matched filter

MLE maximum likelihood estimate

MSW Mikheyev–Smirnov–Wolfenstein

NO normal ordering

NP Neyman-Pearson

PDF probability density function

PMNS Pontecorvo–Maki–Nakagawa–Sakata

QMF quadrature matched filter

RAM random access memory

RF radio frequency

RMS root mean square

ROC receiver operating characteristic

SNO Sudbury Neutrino Observatory

SNR signal-to-noise ratio

TF transfer function

WGN white Gaussian noise

Introduction

1

“ *The only way of discovering the limits of the possible is to venture a little way past them into the impossible.*

— **Arthur C. Clarke**
Science fiction author

In researching the most fundamental physical processes of the universe, precision measurements are an indispensable tool for pushing the boundaries of knowledge into the realm of the unknown. They can help validate or challenge predictions of existing theories and uncover subtle new phenomena that may hint at new physics. The pursuit of precision often drives the development of new technologies and methods. One such method is Cyclotron Radiation Emission Spectroscopy (CRES) [1], a technique for energy spectroscopy of charged particles that provides the sub-eV resolution required by next-generation detectors. With CRES, particle energies are determined by measuring the frequency of their emitted cyclotron radiation in a magnetic field. This radiation can be detected, for example, by radio frequency (RF) antennas, and subsequently converted to digital signals, which are analyzed with digital signal processing techniques. The CRES technique was conceived for the proposed Project 8 experiment, which seeks to determine the neutrino mass through a very precise measurement of the β -decay spectrum of tritium.

Despite being the most abundant massive particles in the universe [2], neutrinos remain the only known elementary particles whose absolute mass is still a mystery. Unraveling this mystery is important due to the neutrinos' role in the universe's expansion history and structure formation, as well as the unresolved theoretical mechanism of their mass generation, which has broader implications for the understanding of particle physics [3]. With current experimental knowledge their mass is non-zero and must be about 6-7 orders of magnitude smaller than the mass of the electron, the next heavier elementary particle [4, 5, 6]. Therefore, attempts to measure the neutrino mass are perfect examples of extremely challenging precision experiments that require high statistical significance, excellent energy resolution and near-perfect understanding of the detector and its systematic effects.

Many of the most sensitive physics experiments on this cutting-edge level are notorious for producing vast amounts of data, demanding considerable computing resources as well as efficient data analysis and reduction techniques. Extreme cases are the experiments at the Large Hadron Collider like ATLAS, which generate data rates exceeding 10 TB s^{-1} [7], and the largest projected radio telescope in the world, the SKA Observatory, which is expected to produce more than 2 TB s^{-1} [8]. While these represent the upper limits of data generation, other experiments such as the upcoming DUNE experiment and the IceCube Neutrino Observatory also handle significant data rates, with 1 GB s^{-1} [9] and a more manageable 1 TB d^{-1} [10], respectively. CRES experiments have a similar tendency, typically generating 400 MB s^{-1} of raw data from a single antenna, depending on the requirements for the digitizer settings (namely, bit depth and sampling rate). CRES experiments that require large volumes of source material to acquire significant statistical data will demand an array of antennas with multiple independent channels. For example, a prototypical proof-of-concept experiment design considered by the Project 8 collaboration would utilize 60 antennas, producing a substantial 24 GB s^{-1} of raw data. As will be demonstrated in chapter 8, a full-scale experiment for a neutrino mass measurement could potentially require thousands of antennas with a corresponding increase in data rate.

While the exact amount of data produced by such a large antenna array will vary depending on experimental design choices, it is evident that such an experiment will require highly efficient computing solutions for analyzing the data. The primary data analysis tasks for CRES experiments, regardless of the specific physics application, are the detection of signals within the noisy digital data and the estimation of the particle energy from the observed signals. Due to the limitations in available data storage capacity, these two tasks may need to be executed in real-time before any raw data is written to disk. Signal detection is a decision problem, where the task is to determine whether a segment of data contains only noise or a combination of signal and noise. Ideally, this process can considerably reduce the data rate by discarding data that contain only noise. Once a signal is identified, the task becomes an optimization problem, where the goal is to find the best fit that explains the observed data, yielding an estimate for the signal parameters.

The effectiveness of signal detection and parameter estimation — defined as detection efficiency and energy resolution — both significantly influence the experiment's sensitivity. Given the significant cost of constructing and operating a large-scale CRES experiment with thousands of antennas, it is crucial to develop experiment designs guided by simulated data, optimizing for this experimental sensitivity. Therefore,

accurate estimates of an experiment's achievable detection efficiency and energy resolution are essential for guiding the design development.

So far, the CRES technique is lacking advanced data analysis approaches and methods for estimation of the detection efficiency and energy resolution of an experiment. Therefore, the goal of this dissertation is to further develop CRES as a technology in this area. Although the discussions throughout the thesis are frequently motivated by the Project 8 experiment and use it as an example, the presented methods are applicable beyond this specific application.

The primary objectives are twofold: to develop algorithms for the detection and parameter estimation of CRES signals from an antenna array detector, and to characterize their detection efficiency and energy resolution. Since the detection algorithms initially considered have demonstrated insufficient efficiency, the priority is to identify the optimal solutions that could be achieved with unlimited computing resources. The topic of computationally feasible solutions is only briefly addressed. To advance these primary goals, an essential prerequisite is the development of a detailed simulation model capable of efficiently generating large data sets for a CRES experiment involving thousands of antennas. This simulation model will be used to produce signals that can be analyzed with the detection and estimation approaches. The result of this dissertation will be the first cohesive framework for determining a CRES experiment's theoretical limits in detection efficiency and energy resolution, starting from experimental design, through data simulation, to data analysis.

This leads to the following thesis structure. Since the development of the CRES technique was motivated by the challenges of neutrino mass measurement, chapter 2 begins with an overview of the fundamentals of neutrino masses and the current state of measurement approaches. This is followed by a detailed discussion of the Project 8 experiment and CRES with an antenna array detector in chapter 3. In chapter 4, the mathematical details of the physical processes that define the expected detector signals are established, laying the groundwork for the development of the simulation model, as covered in chapter 5. Subsequently, chapter 6 addresses the algorithmic approaches for detecting these signals in noisy data, their computational feasibility and the derivation of the theoretical limits of the detection efficiency. Chapter 7 then explores the energy estimation from the detected signals and its theoretical limits, including a discussion of the expected challenges based on a simplified signal model. All of these aspects culminate in chapter 8, which presents a simulation-based analysis of the limits in detection efficiency and energy resolution for an example experimental design with a full-sized antenna array for a neutrino mass measurement. The thesis concludes with a summary and outlook in chapter 9.

Neutrinos and Neutrino Mass

” *I have done a terrible thing, I have postulated a particle that cannot be detected.*

— **Wolfgang Pauli**

Pioneer of quantum mechanics

Neutrinos, often dubbed "ghost particles" due to their elusive nature, are arguably the most fascinating elementary particles. First proposed by Wolfgang Pauli in 1930 as a solution to the continuous electron energy spectrum observed in β -decay [11], their existence was confirmed with the experimental detection of the electron neutrino in the 1950s [12, 13]. In the Standard Model of particle physics, neutrinos are electrically neutral leptons that participate only in the weak interaction. As a consequence of this, they rarely interact with other particles, allowing them to pass through most matter unhindered and making their detection extremely challenging. While the Standard Model initially predicted neutrinos to be massless, the discovery of neutrino oscillations [14, 15] proves that neutrinos possess nonzero masses. Therefore, the study of neutrinos and their masses provides valuable insights into physics beyond the Standard Model and sheds light on many open questions in modern elementary particle physics. This chapter briefly summarizes the essential backgrounds of neutrino masses and their experimental determination.

2.1 Neutrino Oscillation

Neutrinos interact with matter in one of three flavors — ν_e , ν_μ , or ν_τ — each paired with its corresponding charged lepton in charged current weak interactions. In the theory of neutrino mixing and oscillations these flavor eigenstates are linear combinations of the so-called mass eigenstates ν_1 , ν_2 , and ν_3 such that [16]

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \mathbf{U} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (2.1)$$

In eq. (2.1) U denotes the unitary 3×3 Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix, which is most commonly parameterized by three angles θ_{12} , θ_{23} , θ_{13} and a complex phase δ_{CP} .

The time evolution of the mass eigenstates $|\nu_j\rangle$ is described by [16]

$$|\nu_j(t)\rangle = |\nu_j(0)\rangle e^{-i \frac{m_j^2 c^4 t}{2E\hbar}}, \quad (2.2)$$

where c is the speed of light, \hbar the reduced Planck constant, E the particle energy, and m_j the mass of the ν_j eigenstate. The dependence on m_j implies that the mass eigenstates do not propagate coherently if their masses are not identical. Consequently, the composition of an initial flavor state $|\nu_\alpha(t)\rangle = \sum_j U_{\alpha j} |\nu_j(t)\rangle$ (with $\alpha = e, \mu, \tau$) changes periodically over time, allowing for the observation of a different flavor at a later time. This periodic flavor conversion is known as *neutrino oscillation*.

The probabilities for measuring a specific flavor after a propagation time t can be derived with the time evolution of the mass eigenstates (eq. (2.2)) and the PMNS matrix. In a simplified model of just two neutrino flavors¹ α and β , the probability that a neutrino of original flavor α is observed as flavor β after a travel distance $L = t \cdot c$ can be found to be [16]

$$P_{\alpha \rightarrow \beta} = \sin^2 \left(\frac{\Delta m_{21}^2 c^4 L}{4E \hbar c} \right) \sin^2 (2\Theta_m). \quad (2.3)$$

In eq. (2.3) $\Delta m_{21}^2 = m_2^2 - m_1^2$ is the difference of squared masses of the two neutrino mass eigenstates, and Θ_m is the angle characterizing the 2×2 transformation matrix in this model. Notably, if $\Delta m_{21} = 0$ or $\Theta_m = 0$, it follows that $P_{\alpha \rightarrow \beta} = 0$. Thus, the observation of a transition from one flavor to another requires that the PMNS matrix is not the identity matrix, meaning that the mass eigenstates are not identical to the flavor eigenstates, and that their masses are different.

Experiments that measure the flux of neutrinos of a specific flavor can determine these transition probabilities if the initial flux and flavor at the neutrino source are known. As a result, these experiments are sensitive to the mixing angles θ_{ij} of the PMNS matrix and the differences of squared masses Δm_{ij}^2 . This led to the experimental discovery of neutrino oscillations in the Super-Kamiokande detector [14] and the Sudbury Neutrino Observatory (SNO) [15], which proves that m_1 , m_2 , and m_3 must be different and thus at least two of them must be non-zero.

¹A similar more general formula for the three flavor case can be found in [16] but for the discussion of the essential features the two flavor case is sufficient.

An important question is the ordering of the three neutrino masses, as the naturally assumed order $m_1 < m_2 < m_3$ may not be correct due to the arbitrary naming of mass eigenstates as ν_1, ν_2, ν_3 . Hence, this ordering must be determined experimentally by measuring the mass differences, including their signs. Unfortunately, the sign is known only for Δm_{21}^2 , which is measured by experiments detecting neutrinos originating from the Sun's core. As these neutrinos propagate outward, their interaction with solar matter modifies the oscillation compared to the vacuum case, known as the Mikheyev–Smirnov–Wolfenstein (MSW) effect, which allows the sign to be resolved [17]. In contrast, $|\Delta m_{32}^2|$ is measured by experiments detecting Earth-based neutrinos from the atmosphere or particle accelerators. Determining its sign is more challenging and has not yet been accomplished because the effects of matter on neutrino propagation are small on terrestrial scales [17]. With $|\Delta m_{32}^2| \gg \Delta m_{21}^2$ the unknown sign results in two possible ordering relations: the normal ordering (NO) with $m_1 < m_2 \ll m_3$, or the inverted ordering (IO) with $m_3 \ll m_1 < m_2$.

Unfortunately, neutrino oscillation experiments only provide the differences Δm_{ij}^2 , not the absolute masses. Therefore, other experimental approaches are needed to determine the absolute masses. Nevertheless, the measurements of the mass differences can set lower limits for the absolute masses under the assumption that the lightest neutrino has a mass of zero. The most recent results from combined global data, as reported in [4, 5], establish lower limits of $\sum m_i > 58.7$ meV for the NO and $\sum m_i > 99.2$ meV for the IO.

2.2 Neutrino Mass Measurement

With the lower limits provided by the oscillation experiments, the heaviest neutrino mass is at least 0.05 eV, which is about 7 orders of magnitude smaller than the mass of the electron. Dealing with such small scales, a measurement of the absolute mass still remains challenging. To this day all measurements could only set upper limits for the mass as the results are indistinguishable from zero within the margin of error.

2.2.1 Kinematic Approach

One strategy for a direct measurement of the mass is the study of the kinematic properties of weak decays such as β -decays. For instance, β^- decays of an isotope

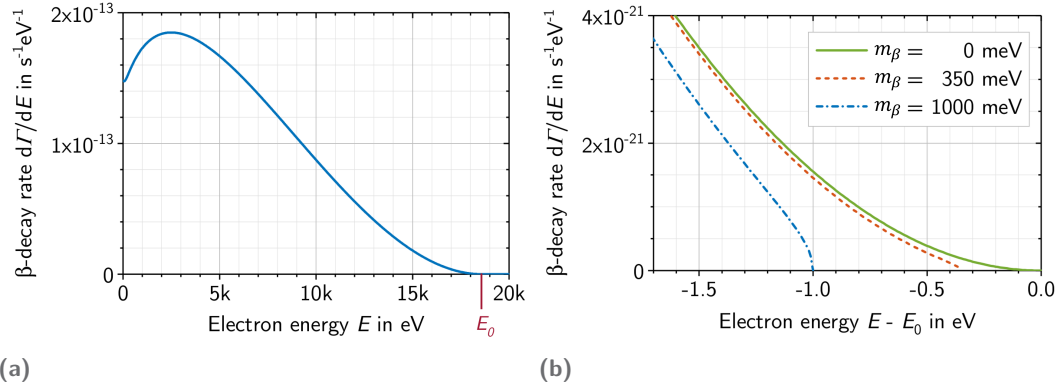
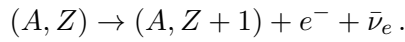


Fig. 2.1.: Theoretical energy spectrum of electrons produced in β -decay of molecular tritium. (a) shows the full range of energies from 0 to the endpoint energy E_0 . (b) shows a close-up of 1.7 eV below the endpoint and the impact of different values of m_β on the shape and the maximum energy of the spectrum. From [18].

with mass number A and charge number Z occur via the emission of an electron and an electron antineutrino:



As Fermi first discussed in his early work on β -decay theory [19], the neutrino mass can be inferred from the energy spectrum of the emitted electron based on conservation of energy and momentum.

The total energy available in the decay process is limited by the rest energy of the parent atom $m_p c^2$. The energy that is released is given by the Q -value, which is defined as the difference in rest energy of the parent atom and the decay products. With the daughter atom mass m_d , electron mass m_e , and effective neutrino mass m_β , the Q -value is

$$Q = (m_p - m_d - m_e - m_\beta) \cdot c^2.$$

This excess energy is available as kinetic energy and is shared between the three decay products². The electron mass is known and the mass difference between parent and daughter can be measured independently with mass spectrometry methods [20]. The endpoint energy is defined as the maximum possible kinetic energy of the electron under the assumption that $m_\beta = 0$:

$$E_0 = (m_p - m_d - m_e) \cdot c^2 - E_{\text{rec}}.$$

²This conceptual discussion assumes a simple scenario of non-molecular parent and daughter in the ground state without energy transitioning to nuclear, atomic or molecular excitation.

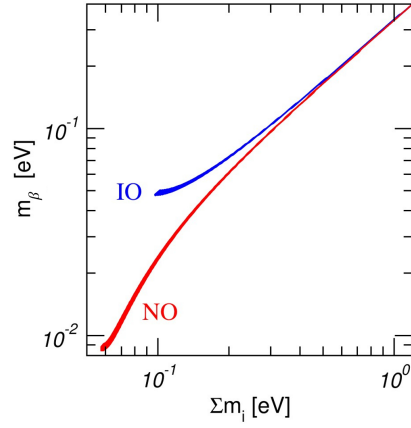


Fig. 2.2.: Mass scale of the observable m_β as a function of the sum of the neutrino mass eigenstates m_i . The two lines show the possible range of values of the observable mass under the NO and IO. From NuFit v5.2 [4] [5].

Due to the daughter nucleus being much heavier than the electron, its recoil energy E_{rec} is tiny compared to Q , with negligible variation near the endpoint. It can be calculated analytically by applying conservation of momentum [21].

As a result, m_β can be deduced from a precision measurement of the electron spectrum near E_0 , as the neutrino mass reduces the maximum electron energy by $m_\beta c^2$ and modifies the spectral shape. Figure 2.1 depicts the theoretical electron spectrum for β -decay of molecular tritium ${}^3\text{H}_2$, derived with Fermi's golden rule [21, 18]. The effect of the spectral distortion under different values of m_β is illustrated in fig. 2.1b. Therefore, a fit of the theoretical spectrum to the data yields m_β .

The observable m_β of such an experiment is the effective mass of the electron neutrino defined as weighted sum of its mass eigenstate content:

$$m_\beta^2 = \sum_{i=1}^3 |U_{ei}|^2 m_i^2. \quad (2.4)$$

In eq. (2.4) U_{ei} is the element of the PMNS matrix corresponding to the amplitude of mass eigenstate i in the electron flavor state. Due to the good knowledge of the mixing angles θ_{ij} and the squared mass differences Δm_{ij}^2 from oscillation experiments, a measurement of m_β can determine the absolute masses of the neutrinos m_1 , m_2 , and m_3 . The relation between m_β and $\sum m_i$ is depicted in fig. 2.2.

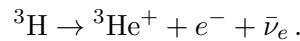
The lower bounds of $\sum m_i$ translate to lower bounds of $m_\beta > 0.05$ eV under the IO and $m_\beta > 0.009$ eV under the NO, while direct measurements have already set upper

limits well below 1 eV [6]. These scales present a significant experimental challenge as they are almost negligible compared to the endpoint energies. For example, in the case of tritium the endpoint energy is $E_0 = (18\,574.00 \pm 0.07) \text{ eV}$ [18]. Measuring such a small deviation from the endpoint requires precision spectroscopy with excellent energy resolution. Additionally, fig. 2.1b illustrates that only a minuscule fraction of the total decays occur within the last 1 eV of the spectrum — specifically, in the case of tritium only one in $\sim 10^{13}$ [22]. This inherent rarity of relevant decays necessitates a very high source activity and good background suppression.

2.2.2 The KATRIN Experiment

Experimental Concept

The current best limit of $m_\beta < 0.45 \text{ eV}$ [6] was obtained by the ongoing Karlsruhe Tritium Neutrino (KATRIN) experiment [23, 24]. KATRIN measures the neutrino mass using the kinematic approach outlined above by studying the β -decay of tritium:



Tritium has a long history in neutrino mass experiments, due to a number of advantages. It has the simplest nuclear structure among all radioactive isotopes, making it ideal for the theoretical calculations of the interaction processes [20]. Furthermore, its high specific activity and relatively low Q -value³ are advantageous for achieving high statistical sensitivity. The latter is due to the fact that the number of decays in the endpoint region scales as $\left(\frac{m_\beta}{Q}\right)^3$ [20].

The KATRIN experiment records an integral electron spectrum employing a spectrometer of the Magnetic Adiabatic Collimation with Electrostatic (MAC-E) filter [25, 26] type. A schematic of the experimental approach in KATRIN is depicted in fig. 2.3. Tritium gas decays in the source volume on the left, the decay electrons are then magnetically guided to the spectrometer under conservation of energy. The spectrometer acts as a high-pass filter, allowing only electrons to pass with energies above the cutoff energy E_c . The detector on the right counts all arriving electrons. By scanning different values of E_c , this approach determines the integral of the tritium spectrum from fig. 2.1.

The working principle of the MAC-E filter is as follows: When tritium decays in the source electrons are produced with random, isotropic directions. A strong

³See the endpoint energy given before. Assuming negligible neutrino mass the Q -value differs from the endpoint only by the nuclear recoil. For tritium $E_{\text{rec}} = 1.72 \text{ eV}$ [18].

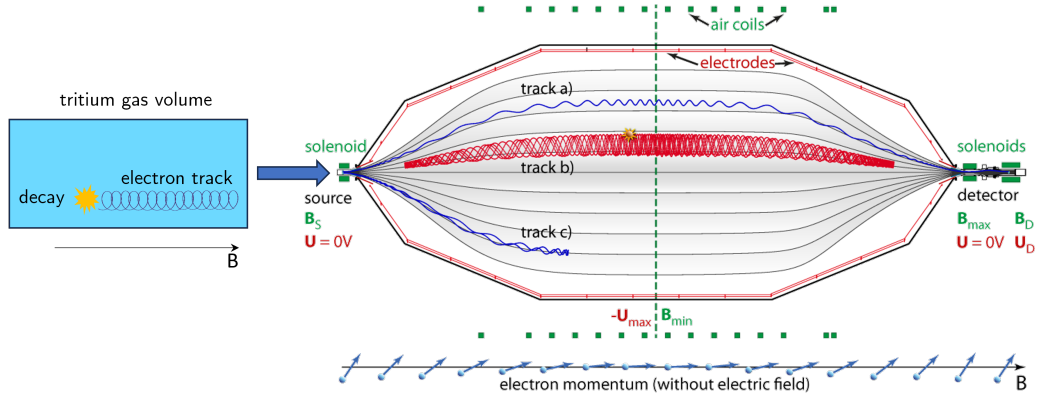


Fig. 2.3.: Schematic of tritium β spectroscopy with a MAC-E filter in the KATRIN experiment. From left to right: tritium source, MAC-E filter, electron detector. The tritium source is radially enlarged and not true to scale. Additional components are omitted for clarity. Adapted from [27].

magnetic field with strength B_S is applied, forcing the electrons into a helical motion that follows the magnetic field lines, guiding them to the spectrometer. Inside the spectrometer, which is an ultra-high vacuum vessel with a diameter of 9.8 m, the magnetic field strength decreases by about four orders of magnitude to B_{\min} , causing the field lines to spread out radially due to the conservation of magnetic flux. The electrons follow these field lines, and in the process, their momentum vectors become aligned with the magnetic field direction⁴, converting their rotational energy into longitudinal kinetic energy (as visualized in fig. 2.3 below the MAC-E filter). With the electron momenta aligned, the high-pass filter is achieved by applying a high voltage U_{\max} to create an electrostatic potential between the source and the magnetic field minimum. Only electrons with longitudinal kinetic energies $E_{\text{kin}} > E_c = qU_{\max}$, where q is the electron charge, can successfully overcome this retarding potential and reach the counting detector at the end.

The energy resolution of the spectrometer is determined by the amount of energy that remains in the circular motion of electrons as they cross the maximum of the electrostatic potential, and it is proportional to the ratio $\frac{B_{\min}}{B_S}$. KATRIN was designed for $\Delta E = 0.93$ eV, which was later adjusted to $\Delta E = 2.77$ eV in standard operating conditions [24]. Although this resolution may seem poor compared to the expected sub-eV scale of the neutrino mass, in an integral measurement, the limited energy resolution of the spectrometer is compensated by a precise knowledge of the transmission function [24]. The transmission function, which provides the probability of transmission for an electron of a given energy, can be characterized using a precisely calibrated source of electrons with well-known properties.

⁴The mathematical details underlying this process are elaborated in section 4.1.

Sensitivity Limitations

KATRIN is designed to reach a sensitivity of $m_\beta > 0.2 \text{ eV}$ [24]. While this represents a significant improvement over previous limits from direct measurements, results from cosmology already favor smaller neutrino masses (see section 2.2.3) and an improvement of another order of magnitude would be required to probe at least the entire phase space under the IO. However, the MAC-E filter approach, which has been developed through several generations of predecessor experiments in Mainz [28] and Troitsk [29], has likely reached its full potential with KATRIN. As will be discussed below, significant further advancements beyond KATRIN's design sensitivity are unlikely due to practical constraints.

Achieving an order of magnitude improvement in the limit on m_β requires a 10000 times increase in the tritium decay rate, as the limit scales with $r^{-\frac{1}{4}}$, where r is the experiment's event rate in the last 1 eV of the tritium spectrum [20]. The tritium decay rate could be increased by either raising the source's density or expanding its volume. However, as the tritium decays, the electrons must still reach the spectrometer before any energy measurements can take place (see fig. 2.3). Increasing the density or length also raises the likelihood that electrons will lose energy through collisions with tritium gas molecules along their path, which skews the results. KATRIN already operates at the very limit of gas density and source length compatible with sufficiently low collision probabilities, allowing only for radial scaling of the source [24, 30]. Due to the conservation of magnetic flux, however, the diameter of the spectrometer scales proportionally with the source diameter [23]. Therefore, radially scaling up the source volume by a factor of 10000, would require a spectrometer with a diameter of 1 km, all other things being equal. It is unlikely that an ultra-high vacuum vessel of this size can be constructed.

In addition to the unfavorable scaling of the spectrometer, all existing tritium beta decay experiments are subject to a systematic effect, which places a fundamental lower limit on the sensitivity at $m_\beta > 0.1 \text{ eV}$ [30]. In its natural form tritium only appears as the molecule T_2 . After one of the tritium atoms decays the resulting molecule $^3\text{HeT}^+$ can end up in one of many rotational or vibrational states, which take some of the decay energy [22, 20] (see fig. 2.4 for an illustration). This effect adds a systematic uncertainty for the endpoint energy of the electron. To overcome this lower limit, experiments that operate with atomic tritium must be designed. An experiment capable of producing atomic tritium needs to prevent the T_2 recombination by trapping the atoms.

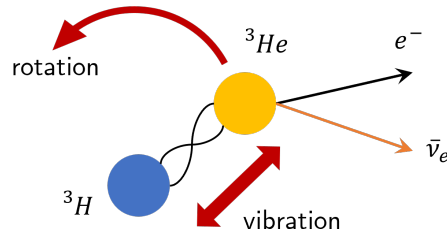


Fig. 2.4.: Illustration of a molecular tritium decay. Initially two tritium atoms are bound as a molecule. If one of them decays into ${}^3\text{He}$ some of that decay energy will go into vibrations and rotations of the bound molecule.

2.2.3 Other Approaches

Besides the kinematic measurement with a MAC-E filter in tritium β -decay, several other approaches have been pursued to investigate the neutrino mass. This section provides a high-level overview of these alternative approaches.

Cosmology

One possibility is to constrain the masses through observations of cosmic structures, which are influenced by the neutrino masses in numerous ways. A significant factor is the presence of massive relic neutrinos from the early universe, which contribute a non-negligible amount to the total mass of the universe. This additional mass affects the expansion history of the universe and the gravitational formation of large-scale structures, such as galaxy clusters [31]. In addition, the neutrino mass has an effect on the anisotropies observed in the cosmic microwave background (CMB) in a non-trivial way. Most notably, the mass of the relic neutrinos influences the gravitational lensing of the CMB photons by galactic structures [31].

Through these and other less significant cosmological observables, the complex cosmological models are sensitive to the sum of the three masses $\sum m_i$. With data from telescopes that map the cosmic structures in high resolution, such as the Planck space telescope [32], cosmological models were able to set tight limits of $\sum m_i < 0.12 \text{ eV}$ at a 95% confidence level [33]. With the next generation of experiments a sensitivity of $\sum m_i > 0.02 \text{ eV}$ is expected [20]. As indicated by fig. 2.2, this will be sufficient to cover the entire range of the NO, making a measurement of nonzero neutrino masses highly likely in the near future.

Nevertheless, there are significant concerns with results from cosmological observations, as they are highly dependent on the underlying models that could prove to be

incomplete. Furthermore, most of the fitting parameters in these models are inaccessible in laboratory measurements. An independent measurement of the neutrino mass would remove a degree of freedom, thereby helping to better constrain those parameters that cannot be measured directly [20].

Neutrinoless Double β -decay

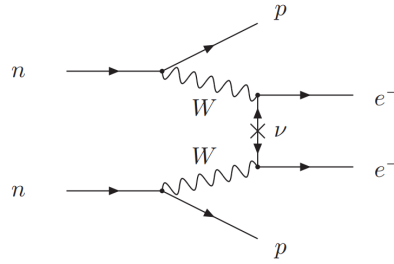


Fig. 2.5.: Feynman diagram for neutrinoless double β -decay. From [2].

Neutrinoless double β -decay ($0\nu\beta\beta$) is a hypothetical nuclear decay that requires neutrinos and their anti-particles to be identical (Majorana particles) [2]. Figure 2.5 depicts the Feynman diagram of this process. It is a decay that simultaneously emits two electrons without emission of any neutrinos:

$$(A, Z) \rightarrow (A, Z + 2) + e^- + e^- .$$

Experiments searching for $0\nu\beta\beta$ utilize isotopes where regular β -decay is forbidden but that are known to undergo ordinary double β -decay with the emission of two anti-neutrinos. The mass of the two neutrinos reduces the available kinetic energy for the two electrons. Thus, the observation of an additional peak above the endpoint of the continuous electron spectrum would indicate double β -decay without the emission of neutrinos [34].

The observable that is measured in these experiments is the half-life of the decay, which depends on the effective Majorana mass given by [2]

$$m_{\beta\beta} = \left| \sum_{i=1}^3 \mathbf{U}_{ei}^2 m_i \right| ,$$

where \mathbf{U} is the PMNS matrix for Majorana neutrinos. Limits on $m_{\beta\beta}$ strongly depend on the assumed model for the calculation of matrix elements that describe the nuclear decay process. The current best mass limit of $m_{\beta\beta} < 0.028\text{--}0.122$ eV at 90% confidence level was set with the KamLAND detector [35], where the reported

values are the minimum and maximum obtained over all commonly used models for the matrix elements. According to the conversion from [20], this corresponds to $m_\beta < 0.028\text{--}0.4\text{ eV}$. In converting to m_β , the already broad range of limits is further widened because the lightest neutrino mass is not unambiguously determined by $m_{\beta\beta}$, due to the unknown, so-called Majorana phases [20]. Nonetheless, despite the broad range of possible limits, they are already tighter than those obtained from β -decay. Ultimately, however, any current limit set by $0\nu\beta\beta$ experiments will be meaningless if neutrinos are Dirac particles after all.

Time of Flight Measurement

Another technique is the measurement of the time of flight of neutrinos produced in the short burst of a Type II Supernova explosion [20]. A neutrino produced in a Supernova at a distance D and with energy E arrives at a detector after a time

$$t = \frac{D}{v} = \frac{D}{c\sqrt{1 - \frac{m^2c^4}{E^2}}}. \quad (2.5)$$

In eq. (2.5) m represents the mass of all three neutrinos in the degenerate regime⁵. While the absolute time t between emission and detection is unknown, the time delay $\Delta t(E)$ of the arrival of lower energy neutrinos can be easily identified. As a result, a precise measurement of the arrival times and energies of the neutrinos can determine the mass m . This method assumes that energy dependencies of the emission time are accurately included in the emission model used in the analysis.

A first analysis of this kind was performed with data from Supernova SN 1987A, yielding a limit of $m < 5.7\text{ eV}$ (95% confidence) [36]. The upcoming JUNO [37], DUNE [38], and Hyper-Kamiokande [39] experiments are expected to achieve improved sensitivities of $m > 0.5\text{--}1\text{ eV}$ for Supernovae at a distance of $D \approx 20\text{ kpc}$. A major drawback for this kind of measurement is the need for a nearby Supernova to occur in the first place.

Other Kinematic Measurements

All of the methods presented above are either model dependent, require unconfirmed neutrino properties, or require rare events. In contrast, direct kinematic measurements such as the tritium β spectroscopy rely only on the conservation of energy and

⁵The degenerate regime refers to cases where the lightest neutrino is much heavier than the measured mass differences, and thus $m = m_1 \approx m_2 \approx m_3$.

momentum. Kinematic measurements are not limited to tritium and generally work with any weak decay since they produce neutrinos. The most promising alternative is a calorimetric measurement with Holmium as proposed in [40]. The isotope ^{163}Ho decays into ^{163}Dy via electron capture:

$$p + e^- \rightarrow n + \nu_e.$$

The energy released in this decay is shared between the neutrino, X-rays or Auger electrons emitted after de-excitation of the daughter atom, and the recoil of the daughter nucleus. Like in tritium β -decay the neutrino mass limits the energy available to the other decay products. Thus, a precise measurement of the endpoint region of the decay spectrum provides information about m_β .

Modern Holmium experiments, led by the ECHo [41] and HOLMES [42] collaborations, employ low temperature microcalorimeters embedded with ^{163}Ho to absorb the entire decay energy. In their first published results the ECHo collaboration set the current best Holmium based limit with $m_\beta < 150$ eV at 95% confidence level [43]. This result is statistics limited with just four days of data taking and a small amount of Holmium source material.

Finally, the Project 8 experiment [30] is another tritium-based experiment that applies a new technique for the measurement of the electron energies. Given its importance to the overall context of this thesis, it will be discussed separately in the next chapter.

A summary of the current limits from the different methods is presented in table 2.1, including both the results in their original experimental observables and the approximate equivalents in m_β .

Tab. 2.1.: Summary of neutrino mass limits.

Method	Observable	Value [eV]	m_β [eV]	Reference
Neutrino oscillations	$\sum m_i$	>0.0587 (NO)	>0.009	[4, 5]
		>0.0992 (IO)	>0.05	
Cosmological	$\sum m_i$	<0.12	<0.032 (NO)	[33]
			<0.053 (IO)	
$0\nu\beta\beta$	$m_{\beta\beta}$	<0.028–0.122	<0.028–0.4	[35]
Supernova time of flight	m	<5.7	<5.7	[36]
Tritium β -decay	m_β		<0.45	[6]
Holmium electron capture	m_β		<150	[43]

The Project 8 Experiment

” Any sufficiently advanced technology is indistinguishable from magic.

— Arthur C. Clarke
Science fiction author

The Project 8 Experiment is an endeavor for a next generation experiment that measures the absolute neutrino mass [30]. Its goal is to go beyond KATRIN’s sensitivity limit of $m_\beta > 0.2 \text{ eV}$ [23] to a limit of $m_\beta > 0.04 \text{ eV}$ which would be sufficient to probe the entire phase space of neutrino masses under the inverted ordering as can be seen in fig. 3.1. Like KATRIN, Project 8 tries to measure the neutrino mass through a kinematic measurement with tritium β -decay, as discussed in section 2.2.1.

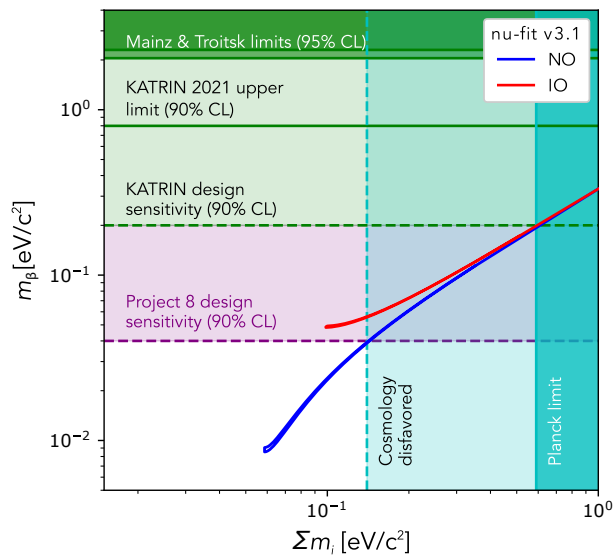


Fig. 3.1.: Mass scale of the observable m_β as a function of the sum of the neutrino mass eigenstates m_i under NO and IO. The horizontal and vertical lines add current limits from cosmology and direct measurements alongside the projected sensitivity limits of KATRIN and Project 8 (The most recent KATRIN limit is not indicated). Image by courtesy of A. Lindman with data from [44, 45, 20, 46].

However, instead of employing a MAC-E filter, Project 8 has pioneered an entirely new approach to electron energy spectroscopy called Cyclotron Radiation Emission Spectroscopy (CRES), which overcomes the limitations of the measurement with a MAC-E filter. In addition, Project 8 must use atomic tritium to reach a sensitivity of $m_\beta > 0.04$ eV due to the molecular sensitivity limit of $m_\beta > 0.1$ eV discussed in section 2.2.2.

In this chapter the Project 8 experiment is discussed on a conceptual level. This includes an explanation of CRES, its advantages for tritium spectroscopy and its detector principles as well as an overview of the current experimental status.

3.1 Cyclotron Radiation Emission Spectroscopy

With CRES the decay electrons are placed in a magnetic field, prompting them to emit cyclotron radiation. For relativistic electrons, the frequency of the radiation is given by [47]

$$\omega_c = \frac{eB}{\gamma m_0} = \frac{eB}{m_0 + E_{\text{kin}}/c^2}, \quad (3.1)$$

where e is the electron's charge, B is the magnetic field strength, γ is the electron's relativistic gamma factor, m_0 is the electron's rest mass, E_{kin} its kinetic energy and c is the speed of light. Thus, by detecting the emitted radiation and measuring its frequency, the kinetic energy can be inferred for a given magnetic field value. In contrast to the integral tritium spectrum measured by the KATRIN experiment, this approach is a differential measurement since the spectrum can be constructed by measuring individual electron energies and counting them.

3.1.1 CRES Detectors

To construct a suitable detector that is sensitive to the electron's cyclotron radiation, one has to take into account the frequency of that radiation. For a neutrino mass measurement with tritium beta spectroscopy E_{kin} in eq. (3.1) can be constrained to a narrow window of a few eV near the tritium endpoint at $E_0 = (18\,574.00 \pm 0.07)$ eV [18], as only these energies are sensitive to the neutrino mass. Thus, the range of possible frequencies is determined by the choice of the magnetic field B . With strong superconducting magnets fields as high as 12 T are conceivable (High-Luminosity Large Hadron Collider [48]) resulting in an upper frequency limit of 325 GHz. On the other hand, using fields below 0.01 T would be impractical as

that results in wavelengths longer than 1 m, leading to large experiment sizes due to the required distance between an electron and a radiation detector. Therefore, CRES experiments are effectively restricted to utilizing radio frequencies (RFs) in the range 250 MHz–325 GHz. Radio frequencies are utilized in various technical applications in the fields of radar and telecommunication. Hence, conventional RF devices such as waveguides, antennas, or resonant cavities can be used to collect the cyclotron radiation in a CRES experiment.

Antenna Array

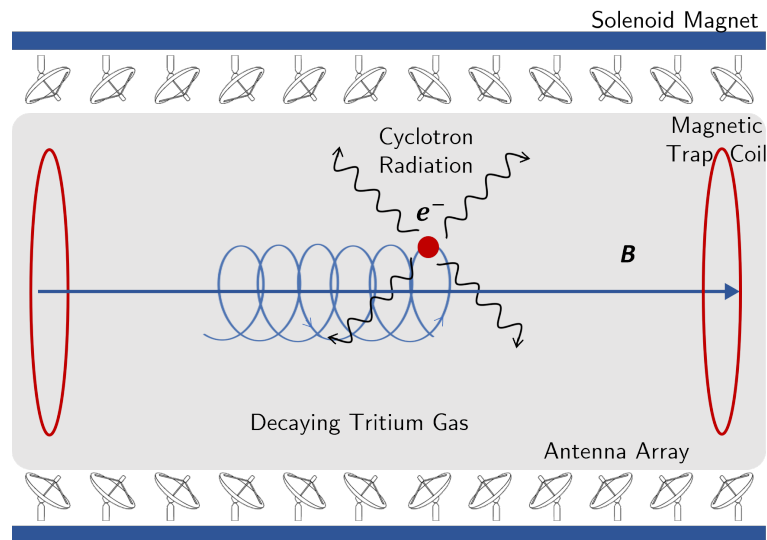


Fig. 3.2.: Schematic of a CRES experiment with a phased array antenna as detector. Tritium decays inside the field of a solenoid magnet. The β -decay electrons are trapped with additional magnetic trapping coils and their cyclotron radiation is detected by the antenna array. Adapted from [49].

Figure 3.2 depicts a schematic overview of a CRES experiment using antennas as detection devices. Tritium gas is filled into a volume inside a uniform magnetic field, which is provided by a solenoid magnet. The tritium decays inside the magnetic field, which forces the decay electrons into a helical motion. Because the electrons can travel significant distances that are much longer than the detector, another magnetic field is added that confines the parallel motion via the magnetic mirror effect (see section 4.1.3). This magnetic trapping field is formed by adding one or more electromagnetic coils, which introduce local maxima or minima to the uniform background field magnitude. The electron trapping is vital since the signal-to-noise ratio (SNR) and frequency resolution scale with the observation time. Both scaling relations will be discussed in more detail in chapters 6 and 7.

The electron's continuous centripetal acceleration on the helical path induces the cyclotron radiation, which can propagate freely inside the detector volume. In order to detect this radiation, an array of RF antennas is placed in the far field of the radiating electron. A detector as depicted in fig. 3.2 would be rotationally symmetric, hence the antenna array has to cover the entire cylinder circumference with each antenna facing inward to the cylinder's symmetry axis.

This antenna array would be implemented as a phased array antenna [50]. With a phased array antenna it is possible to focus the entire array to a single direction by appropriately phase shifting and combining the signals of all antennas. The signal combination can be done either passively or actively. Passive arrays use adjustable analog phase shifters at each antenna element and a single receiver for the combined analog signal of all antenna elements. Active arrays on the other hand, provide greater flexibility by using dedicated receivers for each antenna element, enabling the use of multiple simultaneous frequencies and focus directions [51].

The most flexible version of an active array is a digital beamforming (DBF) phased array where all antenna elements are equipped with digital receivers and the signal combining can be done entirely in software with the digitized signals [52, 53]. Since electron locations are a priori unknown the DBF array is the preferred choice for a CRES experiment. However, as a trade-off between computational cost and detection accuracy, hybrid versions are conceivable in which subarrays use analog beamforming prior to digitization with a shared analog receiver [54, 55].

Cavity Resonator

Figure 3.3 shows the schematic for a CRES experiment with a resonant cavity as detector in place of the antenna array. The setup elements like the solenoid for the background field, tritium gas cell and magnetic trapping coils are the same as before and they serve the same purpose as outlined in the antenna array discussion.

In the case of a cavity detector the whole detection volume is inside of a hollow conductive metal structure. The electron radiation does not propagate freely; instead it excites standing waves inside the cavity which resonate depending on the frequency and cavity dimensions [56]. With this approach it appears easier to collect more of the radiation using a small number of digital receivers. However, the size of this experiment is restricted by the cavity size, which cannot be chosen independently of the frequency considering that the physical dimensions must be on the scale of the wavelength to create standing waves [47, 57]. Thus cavity experiments that are compatible with the need for statistical sensitivity through large volumes, can only

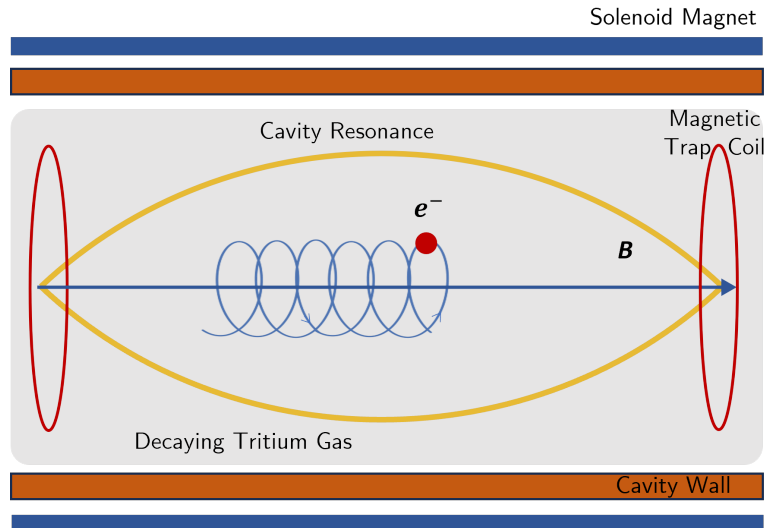


Fig. 3.3.: Schematic of a CRES experiment with a cavity as detector. Tritium decays inside the field of a solenoid magnet. The β -decay electrons are trapped with additional magnetic trapping coils and their cyclotron radiation excites standing waves that resonate with modes of the cavity.

be implemented with frequencies $\lesssim 1$ GHz. Furthermore, the estimation of electron positions becomes more challenging in a cavity, while this is straightforward with an antenna array, as will be demonstrated in section 6.2.

3.1.2 Advantages of CRES

The introduction of the CRES technique has a number of advantages that help increasing the neutrino mass sensitivity reach of tritium spectroscopy experiments.

Considering that CRES detectors are able to measure all electron energies simultaneously, they produce a differential tritium spectrum. However, only the electrons close to the endpoint of the spectrum add information for the neutrino mass. Keeping the detector signal clean of the overwhelming amount of low energy electrons in the spectrum can be easily accomplished by applying low-pass frequency filters. The filtered differential spectrum offers a significant advantage over an integrated one by reducing background counts and avoiding systematic effects introduced by instability of the source or spectrometer [30].

With this differential spectroscopy, each electron needs to be resolved individually with high precision, placing greater emphasis on sub-eV energy resolution than in

KATRIN's integral measurement. Fortunately, because CRES is a frequency measurement, its precision scales excellently with the signal observation time¹ (see also section 7.2.1). The observation time is limited, however, as it may end prematurely by an interaction of the electron with tritium gas or the electron leaving the trap, and ultimately, when the electron loses too much energy through its emitted radiation. Sensitivity studies conducted by the Project 8 collaboration [58] indicate that an energy resolution of $\Delta E = 0.115$ eV is required to achieve the experimental goals². Whether this required resolution can be attained in a concrete experimental setup is one of the major questions that will be addressed in this thesis.

The most significant advantages of CRES are introduced by the fact that the tritium gas source is transparent to the cyclotron radiation [30] and therefore the decay electrons can be observed and detected right at the decay position (see figs. 3.2 and 3.3). This is fundamentally impossible to do for any experiment of the MAC-E filter type since the electrons are required to cross the retarding potential outside of the source, preventing any meaningful improvements in their statistical sensitivity beyond KATRIN (see section 2.2.2). In a CRES experiment on the other hand, electrons can be observed from the moment of their creation, and without the need for a sizable main spectrometer, gains in radial source volume are attainable.

3.2 Project 8 Status

CRES and the Project 8 experiment were initially proposed in 2009 by Monreal and Formaggio [1]. While the renowned KATRIN experiment is pushing an already established technology for measuring the neutrino mass (see Mainz [28] and Troitsk [29] experiments) to its limits, Project 8 utilizes a relatively young measurement technique. Thus, Project 8's path to a 0.04 eV sensitivity experiment is a long ongoing research and development program, which is split in four phases, each involving smaller scale demonstrator experiments [59].

Phase I served as a first demonstration of CRES itself [60] in which the cyclotron radiation of single electrons was observed for the very first time in 2014. The electrons were produced by a gaseous source of the meta-stable radioactive Krypton isotope $^{83\text{m}}\text{Kr}$, which decays via internal conversion. The conversion electrons appear with a discrete energy spectrum that features several narrow spectral lines at

¹This is easily understood if one considers the bin-width of a discrete Fourier spectrum which can be reduced by including more time domain samples.

²This value is not directly comparable to the $\Delta E = 2.77$ eV resolution of KATRIN's MAC-E filter, given the integral spectrum and the role of the transmission function discussed in section 2.2.2.

energies of 17.8–32.0 keV [61]. The proximity of its spectral lines to the endpoint of the tritium energy spectrum near 18.6 keV, has firmly established $^{83\text{m}}\text{Kr}$ as a calibration source in neutrino mass experiments reliant on tritium [62, 63, 64].

In Phase II the first electron spectrum of molecular tritium β -decay was recorded over a period of 82 days, wrapping up in 2020. The data were analyzed using both a Bayesian and a frequentist approach, which yielded limits of $m_\beta < 155 \text{ eV}/c^2$ and $m_\beta < 152 \text{ eV}/c^2$ [65, 66]. Background events due to RF noise were limited to less than one with 90% probability. In addition, calibration data were taken with $^{83\text{m}}\text{Kr}$ in order to characterize the efficiency and detector response. From this data a good energy resolution of $(1.66 \pm 0.19) \text{ eV}$ (FWHM) was measured for the setup. Furthermore, the systematic effects of the measurement were thoroughly investigated and are well understood using both the $^{83\text{m}}\text{Kr}$ data and simulated data.

However, the result of Phase II is not competitive with the world-leading limit of $m_\beta < 0.45 \text{ eV}$ set by KATRIN [6]. The measurement was mostly limited by the statistical uncertainty due to a small number of recorded events. Both Phase I and Phase II were conducted with only a cm^3 -scale detection volume: the gas was confined to a small cell inside of a waveguide — a hollow conductive metal structure that transports RF electromagnetic (EM) waves similar to optical light guides. Therefore, the goal of Phase III is to demonstrate the scalability of the CRES technique to m^3 -scale volumes. This reduces the statistical uncertainty by increasing the number of observable decays while still increasing the energy resolution. For CRES in large volumes, the two new detector technologies outlined in section 3.1.1 have been considered. The initial antenna array concept was extensively studied in simulations, while in a more recent development a substantial effort has been put into studies of the potential of the cavity concept.

In addition to confirming the scalability of CRES, it is essential to also demonstrate the feasibility of an atomic tritium experiment in Phase III. An atomic tritium beamline would consist of a high flux hot atomic tritium source [67] and several cooling and purification stages. The beamline would have to provide a 100 m^3 magnetic atom trap³ with $10^{17} \text{ atoms}/\text{m}^3$ after taking into account all losses in the beamline and the trap. The magnetic trapping of tritium atoms imposes a stringent requirement on the atom temperature of 1 mK [30]. Furthermore, a purity of only one molecule in 10^4 atoms is necessary [30]. This is because a simultaneous measurement of the atomic and molecular tritium spectra is undesirable, as the endpoint of

³The atom trap is not to be confused with the electron trap discussed in section 3.1.1. Both are necessary at the same time and have different requirements.

the molecular tritium spectrum is ~ 8 eV higher than the atomic one [22], which introduces molecular backgrounds beyond the endpoint to the atomic spectrum. Crucially, the atom trap design needs to be compatible with the use of the CRES detector which has to measure the energy of the decay electrons from the atomic tritium inside the atom trap.

Finally, if the goals of Phase III can be reached, Phase IV would consist of a fully scaled-up atomic tritium experiment that can achieve the 0.04 eV sensitivity within a few years of runtime.

The work presented in this thesis was done in the context of Phase III and it contributes to the simulation studies of the antenna array concept.

Theory of CRES Signals

” *A theory is not complete until you can explain it to the person in the street.*

— **David Hilbert**

Founder of Modern Abstract Algebra

In chapter 3 CRES was established as a new approach to electron spectroscopy based on a relatively simple core idea: one measures the frequency of an electron’s cyclotron radiation which is used to calculate its energy with eq. (3.1). The frequency in question is the frequency at which the emitted electric and magnetic fields oscillate. These oscillating fields generate oscillating voltages in the antennas, which is the signal measured by the detector. In reality this signal is more complicated than the simple picture of a sinusoidal voltage as it is a combination of the electron’s motion in a magnetic trap, its characteristic radiation fields, and the response of an RF antenna to the fields. Therefore, the goal of this chapter is the discussion of the theoretical groundwork of the physics that describe CRES electrons. It is mostly an aggregation of results from classical electrodynamics with some additional derivations of specific required equations. The presented theory is used as foundation for a simulation model of the experimental data of electrons that are observed by an antenna array CRES detector which follows in chapter 5.

4.1 Electron Motion

4.1.1 Relativistic Charged Particle Motion

The movement of charged particles in an electromagnetic (EM) field is dictated by the Lorentz force. For electrons at high energies a description of the motion in relativistic form is necessary. The relativistic momentum vector of a particle with rest mass m_0 is defined as

$$\mathbf{p} = \gamma(v)m_0\mathbf{v} ,$$

where \mathbf{v} is the velocity vector of the particle. $\gamma(v)$ is the relativistic Lorentz factor with

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}},$$

where $\beta = \frac{v}{c}$ is the ratio of the absolute velocity v to the speed of light c . With the Lorentz factor the total energy E , the kinetic energy E_{kin} and the rest mass are connected via

$$E = m_0c^2 + E_{\text{kin}} = \gamma m_0c^2. \quad (4.1)$$

For a tritium CRES experiment with $E_{\text{kin}} = 18.6$ keV at the endpoint one can find with eq. (4.1) that $\gamma \approx 1.036$ or $\beta \approx 0.26$. While this is not highly relativistic, the relativistic treatment is still relevant for precision spectroscopy.

Using the momentum vector, the relativistic Lorentz law yields the equation of motion for an electron

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E} + e\mathbf{v} \times \mathbf{B}, \quad (4.2)$$

where e is the elementary charge, \mathbf{E} is the electric field vector and \mathbf{B} is the magnetic field vector.

4.1.2 Cyclotron Motion in Uniform Magnetic Fields

The work done by a force along a path C is given by $W = \int_C \frac{d\mathbf{p}}{dt} \cdot d\mathbf{r}$, and therefore the work done by the Lorentz force on the electron is

$$W = \int_C e\mathbf{E} \cdot d\mathbf{r} + \underbrace{\int_C e(\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{r}}_{=0}.$$

The work done by the magnetic field is zero since $\mathbf{v} \perp d\mathbf{r}$, and hence only electric fields contribute to the work done. Therefore, achieving an accurate measurement of the kinetic energy, with zero work done on the electron, necessitates the absence of electric fields in a CRES experiment. As a result, $\mathbf{E} = 0$ in eq. (4.2) due to these experimental design requirements.

Using only a static uniform magnetic field \mathbf{B} , the energy can be assumed constant and thus $\gamma(v) = \text{const.}$ In that case the solution of eq. (4.2) can be found to be a

helical motion with a certain frequency, pitch¹, and radius [47]. The frequency is called the *cyclotron frequency* and is given by

$$\omega_c = \frac{eB}{\gamma m_0} = \frac{eB}{m_0 + E_{\text{kin}}/c^2}. \quad (4.3)$$

As evident from the vector ω_c , the axis of rotation aligns with the magnetic field direction.

The radius and pitch of the helix both depend on the direction of the electron's velocity v relative to B . For any circular motion with frequency ω the radius is given by $r = \frac{v_{\perp}}{\omega}$ where v_{\perp} is the tangential speed, which in this case is the velocity component perpendicular to B . The pitch of the helix is the distance traveled along the axis in a complete turn, i.e. $h = \frac{v_{\parallel} 2\pi}{\omega}$ where v_{\parallel} is the velocity component parallel to B . By defining the electron's *pitch angle* α as the angle between the velocity and the magnetic field vectors the two velocities are $v_{\perp} = |v| \sin(\alpha)$ and $v_{\parallel} = |v| \cos(\alpha)$. With that the helix' radius, which is called the *gyroradius*, is

$$R_g = \frac{m_0 \gamma |v| \sin(\alpha)}{eB}. \quad (4.4)$$

A special case can be found for a pitch angle of $\alpha = 90^\circ$ when the electron has no parallel velocity and thus the helix' pitch is $h = 0$. This turns the helical motion into a plain circular motion in a plane that is perpendicular to B . However, a deviation from this special case of 1° already results in a traveled distance along the helix axis of more than 1000 m in just 1 ms. It should be of no question that it would be absolutely infeasible to build a detector that is able to observe the electron for this full distance. Because the energy resolution of CRES and the SNR improve with increased observation time (see chapters 6 and 7), this axial motion is what necessitates the electron trap that was mentioned in chapter 3.

4.1.3 Motion in a Magnetic Trap

Due to the requirement of the conservation of the electron's kinetic energy, the trap needs to be purely magnetic. This is accomplished with a magnetic field $B = |B|$ that has a local minimum B_0 along the direction of v_{\parallel} .

A field of this type is called a "magnetic mirror", since electrons are reflected back when entering the regions of higher magnetic fields [47, 68]. The working principle is visualized in fig. 4.1, which depicts a helical electron trajectory starting at $z = 0$

¹The pitch of a helix denotes the height of one complete turn, measured along the helix axis.

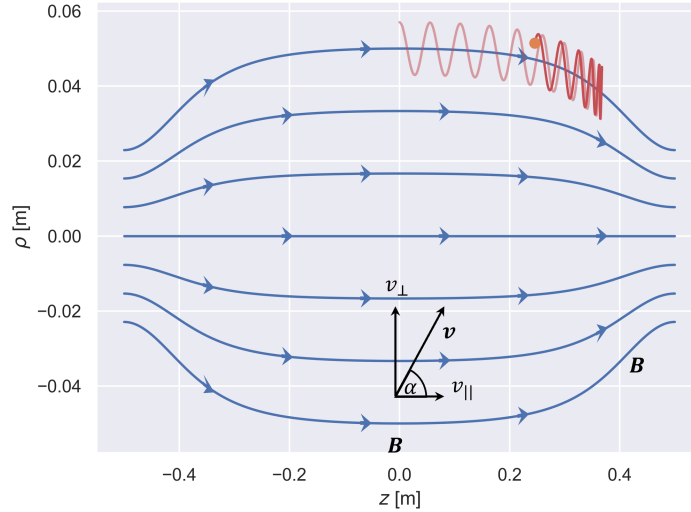


Fig. 4.1.: Magnetic field lines (blue) and an electron trajectory (red) in a magnetic mirror trap. In the right part the electron turns around on its path due to the increase in magnetic field line density.

with some velocity v_{\parallel} parallel to the local magnetic field \mathbf{B} . The direction of the acceleration $e\mathbf{v} \times \mathbf{B}$ (eq. (4.2)) is by definition orthogonal to the direction of v_{\parallel} at this point. The position $z = 0$ is the position of the minimum B_0 , which is indicated by the low field line density. In the high field regions at both ends the density of field lines increases which causes the magnetic field lines to bend and therefore the local magnetic field vector \mathbf{B} changes its direction. As a result the direction of v_{\parallel} is no longer aligned with \mathbf{B} in the high field region and thus $e\mathbf{v} \times \mathbf{B}$ has a component that is antiparallel to v_{\parallel} slowing down this part of the motion². If the field line density gets high enough the electron eventually stops and turns around.

For the mathematical treatment of the motion in a magnetic mirror it is assumed that the magnetic field has rotational symmetry around the z -axis, which is reasonable given the cylindrical detector geometry that is presented in section 3.1.1. Furthermore it is assumed that \mathbf{B} mainly points in z direction and only slowly varies along z . This is true for CRES experiments as well, since the magnetic field consists of the strong solenoid that provides a very uniform background field of approximately 1 T and an added much smaller trapping field of about 1 mT.

Under the assumption of a slowly varying field, the electron motion is called *adiabatic* and v_{\perp}^2/B becomes an invariant of motion [47]. Considering that $|\mathbf{v}| = \text{const}$, this

²Because the kinetic energy is conserved in this kind of field this does not slow the absolute velocity. Any loss in v_{\parallel} of course increases v_{\perp} .

results in the following relation that determines how the instantaneous pitch angle α changes if the magnetic field B changes during the motion:

$$\frac{\sin^2(\alpha)}{B} = \frac{\sin^2(\alpha_0)}{B_0}, \quad (4.5)$$

where α_0 is the pitch angle at the minimum B_0 . A magnetic mirror trap confines electrons that reach $\alpha = 90^\circ$ (i.e. $v_{\parallel} = 0$) under eq. (4.5) at some point in the magnetic field. Electrons reaching the maximum field B_{max} with $\alpha < 90^\circ$ still escape, which means that this concept of trapping is inherently imperfect. With B_{max} the condition for trapping follows from eq. (4.5)

$$\alpha_0 \geq \arcsin \left(\sqrt{\frac{B_0}{B_{max}}} \right).$$

Thus, high trapping efficiencies can be accomplished if $B_{max} \gg B_0$.

The introduction of a trap to confine the axial motion of the electron results in a periodic axial motion along the z axis with frequency ω_a . The corresponding equation of axial motion is [47]

$$\left(\frac{dz}{dt} \right)^2 = v_0^2 - \sin^2(\alpha_0) v_0^2 \frac{B(z)}{B_0},$$

where $v_0 = |v|$ is the absolute value of the velocity. The solution is found by separation of variables with subsequent integration of both sides. Using eq. (4.5) to substitute $\frac{\sin^2(\alpha_0)}{B_0} = \frac{1}{B(z_{max})}$, where z_{max} is the z position that maximizes B , the result is

$$t(z) = \frac{\sqrt{B(z_{max})}}{v_0} \int_{z(0)}^z \frac{dz'}{\sqrt{B(z_{max}) - B(z')}}. \quad (4.6)$$

This result yields the time t that the electron takes in the given magnetic field profile to move from some initial position $z(0)$ to some final position z . Taking the inverse of the result for a particular $B(z)$ yields the axial motion $z(t) = t^{-1}(z) - z(0)$. The result is not periodic per se since eq. (4.6) is only defined for $0 < z < z_{max}$ (it cannot distinguish between the same z values that are reached in different axial periods), and therefore the periodicity has to be forced onto the result $z(t)$. Unfortunately, the integral in eq. (4.6) only has a closed-form solution if $B(z)$ is a polynomial with order 2 at most³ and thus in most cases numeric integration is required.

³When allowing elliptic integrals as solutions, the integral also has a solution for 4th order even polynomials.

4.1.4 Drift Motions

Apart from the cyclotron and axial motions, the electron undergoes gradual drifts resulting from small spatial field variations on top of the gradient along the z direction. In a purely magnetic trap, the primary influence arises from a gradient of B within the plane of the cyclotron orbit, causing variations of the field experienced by the electron during a single revolution. Consequently, a drift emerges that is perpendicular to both the magnetic field and its gradient. The drift velocity associated with this *grad-B* motion is given by [47]

$$\mathbf{v}_{\nabla} = \frac{v_{\perp}^2}{2B^2\omega_c} \mathbf{B} \times \nabla_{\perp} B, \quad (4.7)$$

where $\nabla_{\perp} B$ is the gradient of B in the plane orthogonal to \mathbf{B} .

Finally, if the field lines are curved with a curvature radius $R \gg R_g$, the motion along the curved lines acts as an additional centrifugal force on the electron. This force adds a second drift motion with a velocity [47]

$$\mathbf{v}_c = \frac{v_{\parallel}^2}{\omega_c B^3} \mathbf{B} \times (\mathbf{B} \cdot \nabla) \mathbf{B}. \quad (4.8)$$

In the case of the magnetic mirror trap which requires a gradient along the z axis this curvature is present due to the small radial field components (see fig. 4.1), although it contributes on a much smaller scale to the drift motions than the gradients [69].

The direction and magnitude of both drift motions depend on the specific magnetic field. In section 4.2 the differential vector operators from eqs. (4.7) and (4.8) are calculated for a general rotationally symmetric magnetic field which simplifies the calculation of the drift motions in practice.

As a whole the particle motion that was presented here is the so-called *guiding center* approximation. The lowest order of this approximation is the simple helix that spirals around the magnetic field lines and as higher order effects the drift motions are added. In this approximation it is sufficient to know the position of the guiding center which is the approximate dynamic position of the cyclotron motion's center of rotation. The actual particle position can be reconstructed at any time from the guiding center position using the gyration radius and the cyclotron frequency. A visualization of the kind of motion performed by the electron under this approximation is shown in fig. 5.3 but the concrete manifestation of the motion depends on the magnetic field. For this reason, a discussion of the mathematical basis of the magnetic field required for a simulation model follows in section 4.2.

4.2 Magnetic Fields with Rotational Symmetry

4.2.1 General Mathematical Description

In classical electrodynamics a complete mathematical description of the generation and the dynamics of electric and magnetic fields is provided by the four Maxwell equations. The two equations that are relevant for magnetic fields that are constant in time are [47] Gauss's law for magnetism

$$\nabla \cdot \mathbf{B} = 0 \quad (4.9)$$

and Ampère's extended circuit law

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{j} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right),$$

in which \mathbf{j} is an arbitrary current density and μ_0 and ϵ_0 are the magnetic and electric constants of vacuum. Since a CRES experiment requires that no electric fields are present Ampère's law becomes

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}. \quad (4.10)$$

Equations (4.9) and (4.10) provide the starting point for the calculation of magnetic fields that are generated by an arbitrary current distribution \mathbf{j} , e.g. an electromagnet which is built as a set of current-carrying wires. In practice, the calculation of \mathbf{B} from these equations can be quite challenging, especially for complicated current distributions, and is often done with numerical methods. However, it is possible to mathematically derive the fields that can possibly exist inside of a CRES detector by locally restricting the equations to the inner detector volume. The cylindrical inner volume can only be filled by tritium gas and thus it is a current-free space with $\mathbf{j} = 0$. Using this with Ampère's law will yield the valid mathematical description for any magnetic field that can exist inside the detector, without committing to the details of how the field is generated. The field has to be generated from a current density \mathbf{j} outside of the detector, where the results obtained here will not be valid.

In a current free space eq. (4.10) becomes $\nabla \times \mathbf{B} = 0$ and thus a scalar potential Φ exists with $\mathbf{B} = \nabla\Phi$. Using eq. (4.9) then yields the Laplace equation

$$\Delta\Phi = 0,$$

which has a well-known general solution in spherical coordinates $\Phi(r, \theta, \phi)$ based on the spherical harmonics [47]. In the case of the CRES detector with rotational symmetry there is no ϕ dependence and the potential takes a very simple form

$$\Phi(r, \theta) = \sum_{l=0}^{\infty} a_l r^l P_l(\cos(\theta)), \quad (4.11)$$

where $P_l(x)$ are the Legendre Polynomials and a_l are an infinite number of arbitrary parameters that can shape the potential.

The magnetic field can then be calculated as the gradient of the potential $\mathbf{B} = \nabla\Phi$. In cylindrical coordinates with $r = \sqrt{\rho^2 + z^2}$ and $\theta = \arccos\left(\frac{z}{\sqrt{\rho^2 + z^2}}\right)$ the results are

$$B_\rho(\rho, z) = \sum_{l=1}^{\infty} a_l (\rho^2 + z^2)^{l/2-1} \left(l P_l\left(\frac{z}{\sqrt{z^2 + \rho^2}}\right) \rho - \frac{z\rho}{\sqrt{\rho^2 + z^2}} \frac{\partial P_l(x)}{\partial x} \Big|_{\frac{z}{\sqrt{z^2 + \rho^2}}} \right), \quad (4.12)$$

$$B_\phi(\rho, z) = 0,$$

$$B_z(\rho, z) = \sum_{l=1}^{\infty} a_l (\rho^2 + z^2)^{l/2-1} \left(l P_l\left(\frac{z}{\sqrt{z^2 + \rho^2}}\right) z + \frac{\rho^2}{\sqrt{\rho^2 + z^2}} \frac{\partial P_l(x)}{\partial x} \Big|_{\frac{z}{\sqrt{z^2 + \rho^2}}} \right). \quad (4.13)$$

An important takeaway of these results is that due to the rotational symmetry $B_\phi = 0$, whereas $B_\rho(\rho, z)$ and $B_z(\rho, z)$ are another set of infinite series that are defined by the choice of a_l in eq. (4.11). It is important to point out that the choice of these coefficients entirely determines \mathbf{B} . Therefore, any choice of B_z that satisfies the solution always defines B_ρ as well and vice versa. On the symmetry axis with $\rho = 0$ the result of eq. (4.13) can be greatly simplified to

$$B_z(0, z) = \sum_{l=1}^{\infty} a_l z^{l-1} = \sum_{l=0}^{\infty} a'_l z^l. \quad (4.14)$$

This result means that any function $f(z)$ that can be written as a power series as in eq. (4.14) defines a magnetic field that is allowed for the constraints of rotational symmetry and current $\mathbf{j} = 0$. The choice of coefficients in $B_z(0, z)$ then determines the full vector $\mathbf{B}(\rho, z)$.

This is convenient because it allows for the study of simple analytic magnetic fields by using for example low order polynomials for $B_z(0, z)$. One such field uses $a'_l = 0$, for all l except for $l = 0$ and $l = 2$, which yields a profile $B_z(0, z)$ that is quadratic in z . This *harmonic trap* is one of few traps for which one can find a closed-form

solution for the axial motion with eq. (4.6). The mathematical details of this kind of trap and the motion in it are described in appendix A.1.

Two more traps that are often discussed in the Project 8 collaboration are the square well trap and the "bathtub" trap. A square well trap is one with $\mathbf{B}(\rho, z) = B_0 \hat{e}_z$ for $|z| < L/2$ with some length L and infinitely high walls outside of that. The infinitely sharp walls cannot be exactly reached with eq. (4.14) and therefore this kind of trap breaks the underlying laws of physics ⁴. Nevertheless, studying it in simulations is still valuable for establishing design principles for magnetic fields for CRES which will be seen in section 8.1.2, or even validation as the expected electron behavior is easy to comprehend. The mathematical details of the field and the motion of the square well trap are discussed in appendix A.2.

The "bathtub" trap on the other hand serves as a compromise, which combines the constant portion of the square well with harmonic walls in order to remove the infinities. However, in appendix A.3 it is demonstrated that this is unfortunately insufficient to fix the unphysical square well trap and introduces new issues that make it incompatible with Maxwell's equations. Therefore, it is not recommended to use it for the simulation of CRES.

4.2.2 Drift Motion in Rotationally Symmetric Fields

The second part of this section focuses on the calculation of the drift motions for the specific case of a magnetic field with rotational symmetry. These require the orthogonal gradient $\nabla_{\perp} B$, and the convective operator $(\mathbf{B} \nabla) \mathbf{B}$.

Appendix C.1 demonstrates that for a rotationally symmetric magnetic field the orthogonal gradient is $\nabla_{\perp} B = D_{\hat{e}_{\theta}} B(\rho, z) \hat{e}_{\theta}$. This means it points in the direction of \hat{e}_{θ} , which is one of the spherical basis vectors of \mathbf{B} , and its magnitude is given by the result of applying the differential operator $D_{\hat{e}_{\theta}}$ to the field magnitude $B(\rho, z) = |\mathbf{B}(\rho, z)| = \sqrt{B_{\rho}(\rho, z)^2 + B_z(\rho, z)^2}$. This differential operator is defined as:

$$D_{\hat{e}_{\theta}} B(\rho, z) = \frac{1}{B(\rho, z)^2} \left(B_z(\rho, z)^2 \frac{\partial B_z(\rho, z)}{\partial \rho} - B_{\rho}(\rho, z)^2 \frac{\partial B_{\rho}(\rho, z)}{\partial z} + B_z(\rho, z) B_{\rho}(\rho, z) \left(\frac{\partial B_{\rho}(\rho, z)}{\partial \rho} - \frac{\partial B_z(\rho, z)}{\partial z} \right) \right). \quad (4.15)$$

⁴If required, this trap can at least be approximated with eq. (4.14) by using a high order even polynomial.

Substituting this result and $\mathbf{B} = B(\rho, z)\hat{\mathbf{e}}_r$ into eq. (4.7), simplifies the gradient drift velocity to

$$\mathbf{v}_\nabla = \frac{v_\perp^2}{2B^2\omega_c} B (D_{\hat{\mathbf{e}}_\theta} B(\rho, z)) \hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\theta = \frac{v_\perp^2}{2B\omega_c} (D_{\hat{\mathbf{e}}_\theta} B(\rho, z)) \hat{\mathbf{e}}_\phi, \quad (4.16)$$

which makes use of the other two spherical basis vectors $\hat{\mathbf{e}}_r, \hat{\mathbf{e}}_\phi$.

The curvature drift velocity, given in eq. (4.8), uses the operator $(\mathbf{B}\nabla)\mathbf{B}$ which is called the convective operator. In cylindrical coordinates with $B_\phi = 0$ it is [70]

$$(\mathbf{B}\nabla)\mathbf{B} = \left(B_\rho \frac{\partial B_\rho}{\partial \rho} + B_z \frac{\partial B_\rho}{\partial z} \right) \hat{\mathbf{e}}_\rho + \left(B_\rho \frac{\partial B_z}{\partial \rho} + B_z \frac{\partial B_z}{\partial z} \right) \hat{\mathbf{e}}_z.$$

As a consequence, explicitly expanding the cross product from eq. (4.8) and making use of the orthogonality of the cylindrical basis vectors yields:

$$\begin{aligned} \mathbf{B} \times (\mathbf{B}\nabla)\mathbf{B} &= (B_\rho \hat{\mathbf{e}}_\rho + B_z \hat{\mathbf{e}}_z) \times (\mathbf{B}\nabla)\mathbf{B} \\ &= \left(B_\rho \frac{\partial B_\rho}{\partial \rho} + B_z \frac{\partial B_\rho}{\partial z} \right) (B_\rho \hat{\mathbf{e}}_\rho \times \hat{\mathbf{e}}_\rho + B_z \hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_\rho) \\ &\quad + \left(B_\rho \frac{\partial B_z}{\partial \rho} + B_z \frac{\partial B_z}{\partial z} \right) (B_\rho \hat{\mathbf{e}}_\rho \times \hat{\mathbf{e}}_z + B_z \hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_z) \\ &= \underbrace{\left(B_z \left(B_\rho \frac{\partial B_\rho}{\partial \rho} + B_z \frac{\partial B_\rho}{\partial z} \right) - B_\rho \left(B_\rho \frac{\partial B_z}{\partial \rho} + B_z \frac{\partial B_z}{\partial z} \right) \right)}_{C_B} \hat{\mathbf{e}}_\phi. \end{aligned} \quad (4.17)$$

By substituting this result into eq. (4.8) and combining it with eq. (4.16), the total drift velocity from both effects is given by

$$\mathbf{v}_D = \left(\frac{v_\perp^2}{2B\omega_c} (D_{\hat{\mathbf{e}}_\theta} B(\rho, z)) + \frac{v_\parallel^2}{\omega_c B^3} C_B \right) \hat{\mathbf{e}}_\phi. \quad (4.18)$$

This means that the drift velocity of a rotationally symmetric field is always pointed in the direction $\hat{\mathbf{e}}_\phi$. In other words, it causes a simple circular motion of the guiding center on a circle with radius ρ around the symmetry axis of the trap. The absolute value v_D is calculated by evaluating the scalar values of eqs. (4.15) and (4.17) for the field and applying these in eq. (4.18).

4.3 Electromagnetic Waves from CRES Electrons

4.3.1 General Solution

The phenomenon of electromagnetic waves is mathematically described by the inhomogeneous electromagnetic wave equations

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{E}_r}{\partial t^2} - \nabla^2 \mathbf{E}_r = - \left(\frac{1}{\epsilon_0} \nabla \rho + \mu_0 \frac{\partial \mathbf{j}}{\partial t} \right), \quad (4.19)$$

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{B}_r}{\partial t^2} - \nabla^2 \mathbf{B}_r = \mu_0 \nabla \times \mathbf{j}, \quad (4.20)$$

which directly follow from the full set of the four Maxwell equations. The right hand side of these equations describe the sources of the fields, while the left hand side describe the propagating waves. These summarize that EM waves are generated by gradients of charge density ρ , time variations in current density \mathbf{j} or circulations of current density \mathbf{j} .

A charged particle that moves along a path $\mathbf{r}_s(t)$ is characterized by the charge and current densities

$$\begin{aligned} \rho(\mathbf{r}, t) &= q \delta^3(\mathbf{r} - \mathbf{r}_s(t)), \\ \mathbf{j}(\mathbf{r}, t) &= q \mathbf{v}(t) \delta^3(\mathbf{r} - \mathbf{r}_s(t)), \end{aligned}$$

where q denotes the particle's charge and δ^3 the 3 dimensional Dirac delta function. With these it is apparent that any charged particle that is accelerated in any form ($\frac{d\mathbf{v}}{dt} \neq 0$) radiates energy through the emission of EM waves due to the time varying current density.

Solving eqs. (4.19) and (4.20) with the trajectory of an electron with charge e for the radiated fields $\mathbf{E}_r(\mathbf{r}, t)$, $\mathbf{B}_r(\mathbf{r}, t)$ is not straightforward and is often done through the introduction of the Liénard–Wiechert (LW) potentials [47]. These potentials produce the relativistically correct electric and magnetic fields at any time t and position $\mathbf{r} = R \hat{\mathbf{n}}$ that result from the emission of the wave by an electron in arbitrary motion:

$$\mathbf{E}_r(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \left(\underbrace{\frac{e(\hat{\mathbf{n}} - \boldsymbol{\beta})}{\gamma^2 (1 - \hat{\mathbf{n}} \cdot \boldsymbol{\beta})^3 R^2}}_{\text{static}} + \underbrace{\frac{e\hat{\mathbf{n}} \times ((\hat{\mathbf{n}} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}})}{c(1 - \hat{\mathbf{n}} \cdot \boldsymbol{\beta})^3 R}}_{\text{radiation}} \right)_{t_r}, \quad (4.21)$$

$$\mathbf{B}_r(\mathbf{r}, t) = \frac{1}{c} \hat{\mathbf{n}}(t_r) \times \mathbf{E}_r(\mathbf{r}, t). \quad (4.22)$$

In eq. (4.21) $\beta = \frac{v(t)}{c}$ is the vectorial form of the relativistic β factor and $\dot{\beta} = \frac{d\beta}{dt}$ denotes its derivative. The first term which only depends on the velocity is the static component of the electric field and is insignificant at large distances due to the $\frac{1}{R^2}$ dependence. The radiation is the result of the second term since it also depends on the acceleration and it is the dominating field contribution at distances that are large compared to the wavelength of the EM wave.

The subscript t_r in eq. (4.21) denotes that all properties of motion are evaluated at the *retarded time*. Evaluation at t_r instead of t takes into account the propagation time of the EM wave from the source point $\mathbf{r}_s(t_r)$ to \mathbf{r} . Hence, the retarded time is defined via the implicit equation

$$c|t - t_r| = |\mathbf{r} - \mathbf{r}_s(t_r)| . \quad (4.23)$$

The relationship between observer time t and retarded time t_r is visualized in fig. 4.2. Due to the finite speed of light the electron has continued its motion to a new position by the time the wave reaches the observer. Therefore, an evaluation of all properties at t would yield an incorrect result.

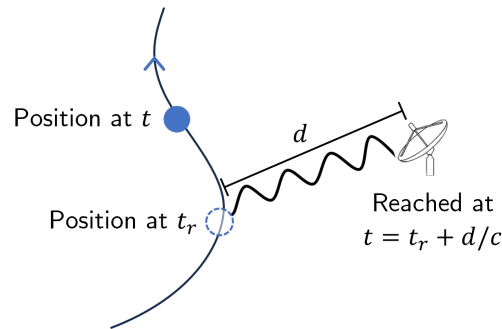


Fig. 4.2.: Visualization of the retarded time. Radiation reaching an observer at time t has been emitted at the retarded time t_r while the electron keeps moving.

Unfortunately, there is no exact closed-form expression for eq. (4.21) for describing an electron in helical motion due to an external magnetic field $\mathbf{B}(\rho, z)$. Nevertheless, in the next sections, features of the solution established in the literature are combined with a new approximation to develop an expression for the radiated electric field $\mathbf{E}_r(\mathbf{r}, t)$.

4.3.2 Spectral and Spatial Power Distribution

While it is not straightforward to write down a general closed-form expression for the time evolution of the wave in eq. (4.21), it is possible to calculate the emitted

power and its spatial and frequency distributions for free space radiation by taking time averages over a single cyclotron orbit. The total radiated power of the electron for example can easily be calculated with the relativistic Larmor formula [47] as

$$P = \frac{1}{4\pi\epsilon_0} \frac{2e^2\omega_0^2}{3c} \frac{\beta^2 \sin^2(\alpha)}{1 - \beta^2}, \quad (4.24)$$

where ω_0 is the non-relativistic cyclotron frequency $\omega_0 = \frac{eB}{m_0}$. With eq. (4.24) one finds that in a magnetic field of 1 T a tritium endpoint electron only radiates a power of $\sim 10^{-15}$ W. On this scale the detection of the emitted waves, which will be discussed in detail in chapter 6, can be extremely challenging due to a very low expected SNR.

Nevertheless, while the total power is informative for the energy loss rate $P = \frac{dE_{\text{kin}}}{dt}$ and estimations of the SNR, it does not yet help in understanding how the fields oscillate. In the simplest approximation the radiation fields would be treated as oscillating with the same frequency ω_c as the cyclotron orbit motion from eq. (4.3) like the electron itself. This plain model implies a spectrum with a single frequency that describes a sine wave. Yet, this simple picture is only accurate in the non-relativistic limit with $\beta \rightarrow 0$. In a general solution the spectrum gets more complicated for the majority of cases depending on the electron's β factor and pitch angle α . Additionally, the field's oscillation changes depending on how it is observed. For now the discussion is limited to a co-moving observer that keeps a constant position relative to the electron's guiding center. The case of an observer at rest requires modifications due to the Doppler effect and is discussed later in section 5.4.2 in order to keep the focus on the electron alone.

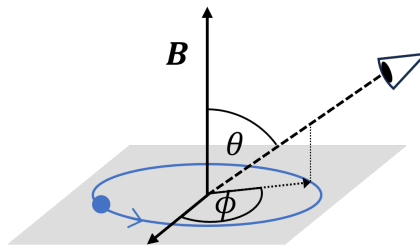
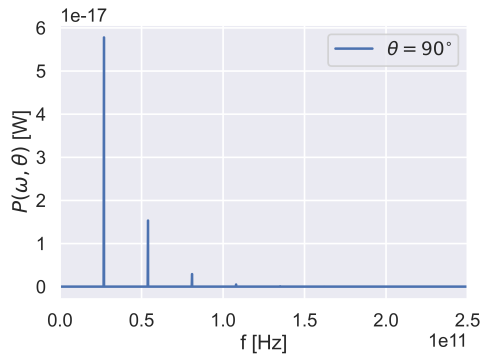


Fig. 4.3.: Angular coordinates of the observer's orientation relative to the magnetic field direction and cyclotron orbit plane.

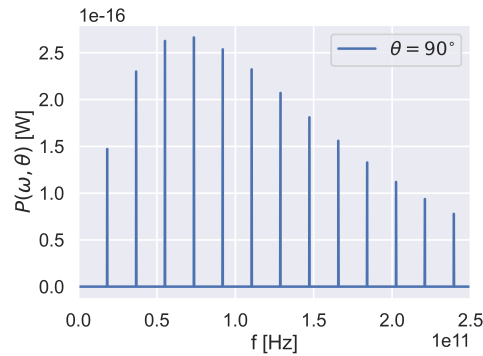
The orientation of the observer is characterized by two angles θ and ϕ , which are depicted in fig. 4.3. The angle θ describes the direction of the observer relative to the magnetic field B , which is always the normal vector of the cyclotron orbit's

plane. The angle ϕ is the azimuth angle in the plane of the cyclotron orbit relative to an arbitrary reference direction⁵.

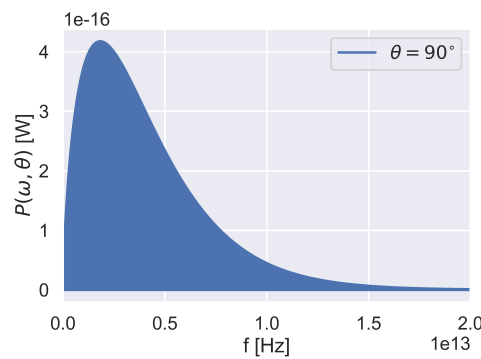
Power Spectrum



(a) Tritium endpoint electron $\beta \approx 0.26$.



(b) High energy electron $\beta = 0.75$.



(c) Highly relativistic electron $\beta = 0.99$.

Fig. 4.4.: Power spectra of electrons in a 1 T magnetic field with pitch angle $\alpha = 90^\circ$ observed with $\theta = 90^\circ$. In (a) and (b) the fundamental frequency ω_c is the leftmost peak and the additional peaks are higher harmonics. In (c) the distinct frequency peaks transition into a continuum of frequencies.

Instead of just a single frequency, the actual spectrum has additional frequency components at integer multiples of the cyclotron frequency ω_c . The manifestation of these higher harmonics depends on the electron's β factor and the angle θ . In general, the highest number of harmonics is visible at $\theta = 90^\circ$. As can be seen in figs. 4.4a and 4.4b for that case the spectrum can vary from few additional peaks at low β to sophisticated spectra with many peaks in the higher energy cases, meaning that the wave is composed of many frequencies of variable proportion.

⁵Due to the symmetry the choice of the reference direction is irrelevant for the results.

In the extreme cases of highly relativistic electrons the spectrum transitions from discrete frequencies to a continuum of frequencies as can be seen in fig. 4.4c. This variable spectral shape is certainly difficult to describe and thus it is no surprise that no closed form solution for the instantaneous field of eq. (4.21) exists.

Spatial Distribution

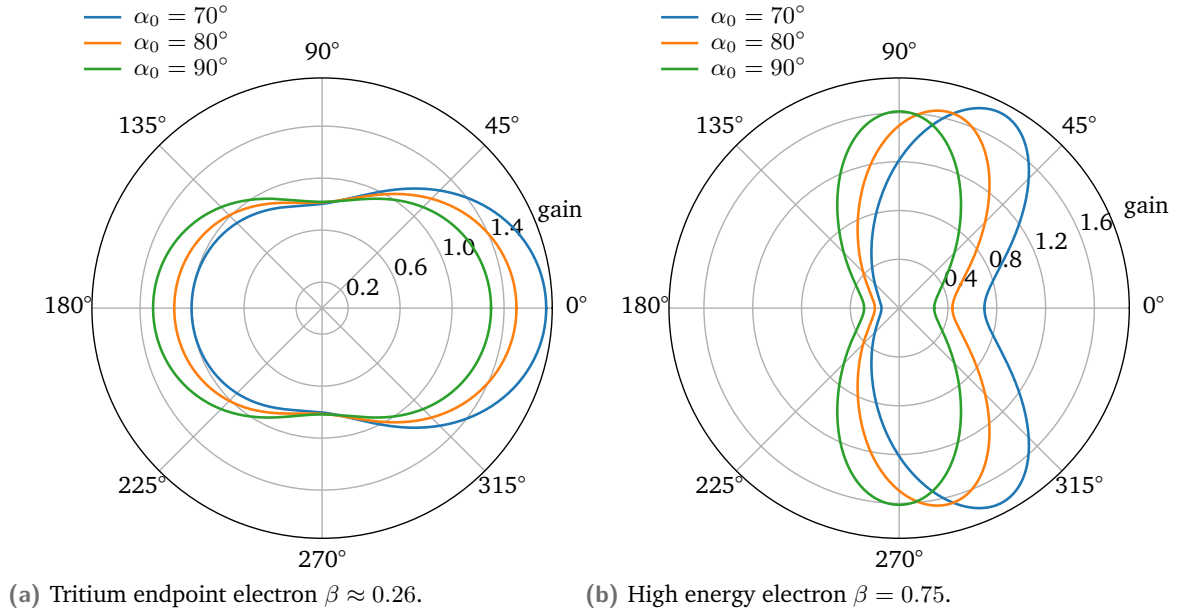


Fig. 4.5.: Angular distribution of the total radiated power of electrons undergoing cyclotron motion. The power is normalized to the total radiated power across all angles such that it is represented as linear antenna gain. The polar plots depict the gain as it depends on θ while it is symmetric around the azimuth ϕ . The distribution changes its shape depending on the electron's relativistic β and its pitch angle α .

The electron radiates its power from eq. (4.24) non-isotropically and the spatial distribution of the radiated power $\frac{dP}{d\Omega}(\theta)$ from all frequencies is given as a long closed-form expression in [71]⁶. It only depends on the angle θ between \hat{n} and \mathbf{B} and not on ϕ , i.e. it is rotationally symmetric around the magnetic field line. The dependence on θ dramatically changes in shape with the electron's energy E_{kin} and pitch angle α . In the interesting case of tritium beta-decay electrons close to the endpoint with $\beta \approx 0.26$ the radiated power has a slight preference for directions parallel to the external magnetic field (see fig. 4.5a) as opposed to highly relativistic cases where the radiated power has a strong preference for directions orthogonal

⁶The derived expression only assumes axial and cyclotron motion and thus neglects effects from drift motions. This is a reasonable approximation, as the kinetic energy due to the drift velocity only amounts to a tiny fraction of the total.

to \mathbf{B} (see fig. 4.5b). For the case of pitch angles $\alpha = 90^\circ$ the spatial distribution is also symmetric around $\theta = 90^\circ$. This symmetry breaks for the general case of $\alpha \neq 90^\circ$ where an excess of power is observed in the direction of axial motion. For the high energy case with $\beta = 0.75$ (fig. 4.5b), the pattern appears almost rotated by 90° with the power radiated mostly in the plane of the cyclotron orbit. As the energy increases further, this radiation beam becomes progressively narrower in that plane.

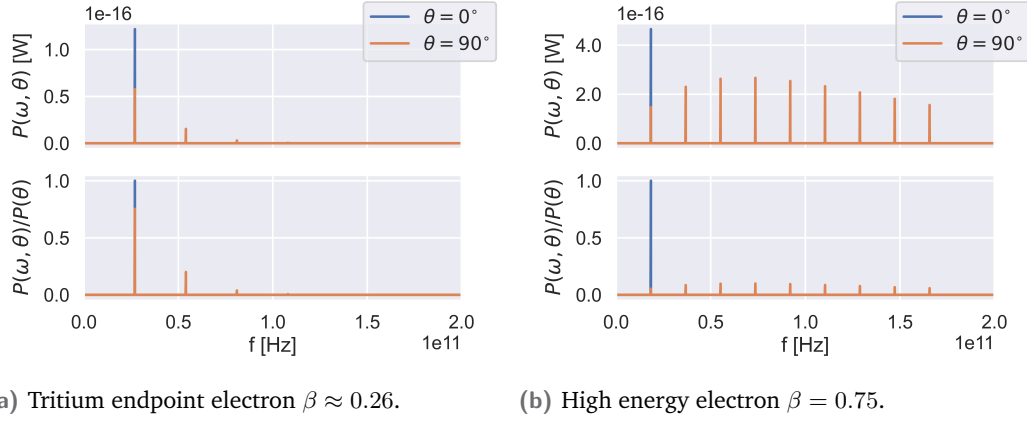


Fig. 4.6.: Power spectra of electrons in a 1 T magnetic field with pitch angle $\alpha = 90^\circ$. For an observer at angle $\theta = 0^\circ$ to the magnetic field all power is in a single peak at the cyclotron frequency while it spreads out to harmonics of the cyclotron frequency at other observation angles. The bottom plots show the power spectra relative to the total power radiated into the observer direction.

Considering both the frequency content and the anisotropy, the power frequency spectrum of the emitted wave is

$$\frac{\partial^2 P}{\partial \Omega \partial \omega}(\theta, \omega) = \sum_{n=1}^{\infty} \frac{dP_n}{d\Omega}(\theta) \delta(\omega - n\omega_c), \quad (4.25)$$

where $\frac{dP_n}{d\Omega}(\theta)$ is the angular power distribution for the n -th harmonic, which is given in a long analytic expression in [72]. Figure 4.6 shows spectra according to eq. (4.25) under different angles θ for two electron energies. Additionally, the respective bottom plots show relative power spectra by normalizing all peaks to the total power for the depicted angle θ . This helps putting into perspective how much the power spreads out into the harmonics in the different cases. From the analytic spectra one can find that at $\theta = 0^\circ$ for all energies and pitch angles all the power is actually radiated with the fundamental frequency. The harmonics start to gradually appear with increasing θ which reaches its peak at $\theta = 90^\circ$. Still, for tritium endpoint electrons the fundamental frequency is the most powerful under all

angles θ and it only drops to 75% of the total power in that direction θ . In contrast, for the high energy case (fig. 4.6b) it shows that only a small fraction of the power $P(\theta)$ is radiated by the fundamental frequency and the power is generally spread out to a wide bandwidth of frequencies.

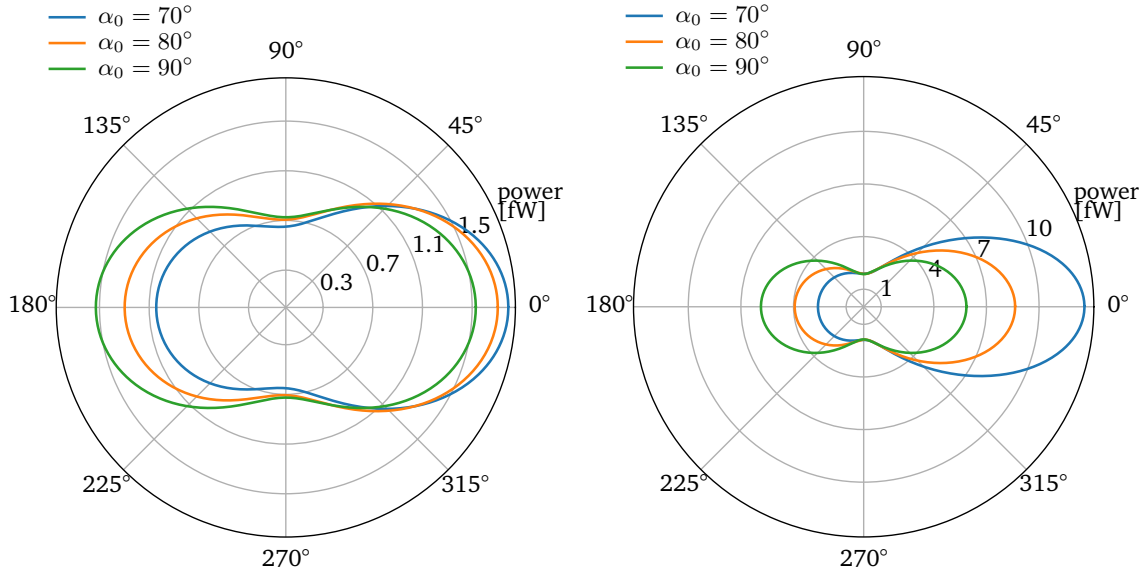
These results have some implications for the feasibility of CRES due to the resulting bandwidth. In both cases of fig. 4.6 the fundamental frequencies are near 20 GHz and thus an RF receiver that is sensitive to more than a single harmonic needs a bandwidth of $>40\text{--}50$ GHz. This pushes the requirement for the sampling rate close to at least 100 GHz, resulting in data rates $> 100 \text{ GB s}^{-1}$ for each receiver⁷. This data rate is a technical challenge on its own but the same is true for the design of an RF receiver of this bandwidth as higher bandwidths are generally associated with lower antenna gains and increased physical dimensions [50]. For this reason, RF receivers of a CRES detector should be designed such that they are only sensitive to one of the harmonics of the electron spectrum. This is supported by section 5.5 where it will be demonstrated that possible electron spectra of a single harmonic after applying the detector response only have a very narrow bandwidth.

Using only the fundamental frequency results in a bearable power loss of 25% for the tritium endpoint electron. while for high energy cases most power would be lost as each individual harmonic only carries a small fraction of the total. Nevertheless, overall this is compensated by the fact that the high energy electron's total radiated power, as calculated by eq. (4.24), is ≈ 20 times higher. Consequently, there is still a slightly higher power in the fundamental frequency compared to its counterpart in the lower energy case (see the upper plots in fig. 4.6), making the detection appear similarly challenging for both cases.

By using several receivers that are sensitive to different harmonics of the high energy spectrum it even sounds plausible to detect higher fractions of the total power and thus improving the efficiency, although at the expense of higher data rates and overall cost. Nevertheless, with the focus on tritium spectroscopy the further exploration of such ideas is outside the scope of this thesis. Furthermore, at the time of writing the only existing CRES experiment working at high energies with $\beta \approx 0.99$ [73] uses waveguide technology for the detection where the presented results of free space radiation for the spectral and spatial power distribution do not apply.

A somewhat unexpected result is the high power at the fundamental frequency for $\theta = 0^\circ$ in fig. 4.6b, which at first seemingly contradicts the narrow beam at $\theta = 90^\circ$ that is seen in fig. 4.5b. However, a similar polar plot of the power radiated by only the fundamental frequency in fig. 4.7 reveals a consistent image with a radiation

⁷Using only 8 bit samples.



(a) Tritium endpoint electron $\beta \approx 0.26$.

(b) High energy electron $\beta = 0.75$.

Fig. 4.7.: Angular distribution of the power radiated in the fundamental frequency. The polar plots depict the gain as it depends on θ while it is symmetric around the azimuth ϕ . The distribution changes its shape depending on the electron's relativistic β and its pitch angle α .

pattern that is surprisingly similar to the one of the low energy case. This provides evidence that the elementary characteristics of the radiation field's fundamental frequency persist even at higher energies. Given the bandwidth limitations and all prior evidence supporting the sole use of the fundamental frequency in high-energy cases, and considering its significantly higher relative power at the tritium endpoint, this frequency will be treated as the only component of the radiation field that is relevant to observe.

4.3.3 Polarization and Phase

With the decision to neglect the higher harmonics of the radiation field, the wave is effectively approximated as a sine wave with frequency ω_c . However, its explicit vectorial form $\mathbf{E}_r(\mathbf{r}, t)$ with its polarization and phase are still unknown. The complete mathematical form can be approached by a low energy Taylor expansion of eq. (4.21). For small β the radiation component of eq. (4.21) is to first order

$$\mathbf{E}_r(\mathbf{r}, t) \approx \frac{e}{4\pi\epsilon_0 c} \left(\frac{1}{R} \hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \dot{\boldsymbol{\beta}}) \right)_{t_r}. \quad (4.26)$$

Similar to the calculation of the axial motion, it is assumed that the field only slowly varies along the axial trajectory and thus the motion is locally approximately helical with the helix axis aligned with the direction of \mathbf{B} . Therefore, the calculation assumes a coordinate system for \mathbf{E}_r that has its z axis aligned with the direction $\hat{\mathbf{B}}$. Hence, the result will be given relative to the local magnetic field direction, which is not always exactly aligned with the symmetry axis z from section 4.2.

These assumptions result in

$$\dot{\boldsymbol{\beta}} = \frac{1}{c} \begin{pmatrix} \omega_c^2 R_g \cos(\omega_c \cdot t + \varphi_0) \\ -\omega_c^2 R_g \sin(\omega_c \cdot t + \varphi_0) \\ \dot{v}_{\parallel} \end{pmatrix},$$

where v_{\parallel} is the velocity parallel to $\hat{\mathbf{B}}$ and φ_0 is the initial phase into the helix. In order to simplify the following derivation it is preferable to use $\dot{v}_{\parallel} \approx 0$ under the assumption that it is much smaller than the x and y amplitudes⁸. Using spherical coordinates such that

$$\mathbf{r} = R\hat{\mathbf{n}} = R \begin{pmatrix} \sin(\theta) \cos(\phi) \\ \sin(\theta) \sin(\phi) \\ \cos(\theta) \end{pmatrix}$$

and substituting $\hat{\mathbf{n}}$ and $\dot{\boldsymbol{\beta}}$ into eq. (4.26) results in

$$\mathbf{E}_r(\mathbf{r}, t) = \left(\frac{e\omega_c^2 R_g}{4\pi\epsilon_0 c^2 R} \begin{pmatrix} -(\cos(\theta))^2 \cos(\varphi_0 + t\omega_c) + \sin(\phi) \sin(\theta)^2 \sin(\phi + \varphi_0 + t\omega_c) \\ \cos(\theta)^2 \sin(\varphi_0 + t\omega_c) + \cos(\phi) \sin(\theta)^2 \sin(\phi + \varphi_0 + t\omega_c) \\ \frac{1}{2} \sin(2\theta) \cos(\phi + \varphi_0 + t\omega_c) \end{pmatrix} \right)_{t_r}.$$

This result is more accessible in the spherical coordinate basis by applying the coordinate transformation $\hat{\mathbf{x}}' = \hat{\mathbf{e}}_{\phi}$, $\hat{\mathbf{y}}' = \hat{\mathbf{e}}_{\theta}$ and $\hat{\mathbf{z}}' = -\hat{\mathbf{e}}_r = -\hat{\mathbf{n}}$. With the corresponding transformation matrix $\mathbf{M} = (\hat{\mathbf{x}}', \hat{\mathbf{y}}', \hat{\mathbf{z}}')^{-1}$ this yields

$$\mathbf{E}'(\mathbf{r}, t) = \mathbf{M} \cdot \mathbf{E}_r(\mathbf{r}, t) = \frac{e}{4\pi\epsilon_0 c^2} \left(\frac{\omega_c^2 R_g}{R} \begin{pmatrix} \sin(t\omega_c + \varphi_0 + \phi) \\ -\cos(\theta) \cos(t\omega_c + \varphi_0 + \phi) \\ 0 \end{pmatrix} \right)_{t_r}. \quad (4.27)$$

⁸Assuming $E_{\text{kin}} \approx \text{const}$ on short time scales and with the assumption of the slowly varying field, the ratio of the amplitudes of $\dot{\boldsymbol{\beta}}$ is $\omega_c^2 R_g (|\dot{v}_{\parallel}|)^{-1} = v_0 \sin(\alpha) \omega_c (v_0 \sin(\alpha) \dot{\alpha})^{-1} = \omega_c \dot{\alpha}^{-1}$. In the magnetic traps under consideration the maximum change in pitch angle from the minimum to the maximum field is $\Delta\alpha = 5^\circ$ on timescales of at least $\Delta t = 0.01 \mu\text{s}$. Thus, $\dot{\alpha} = \frac{\Delta\alpha}{\Delta t}$ is on the order of 10^7 s^{-1} (This is a roughly estimated average over one axial period and it could be much higher on very short segments), while on the other hand ω_c is on the order of 10^{11} s^{-1} . As a result, $|\dot{v}_{\parallel}| \ll \omega_c^2 R_g$.

That means for any direction $\hat{\mathbf{n}}$ the radiation field is a 2 dimensional sine wave in the plane orthogonal to $\hat{\mathbf{n}}$ with a phase shift of $-\pi/2$ between the two components.

The result obtained here is a full mathematical description including polarization and phase of the radiation field of the fundamental frequency in the low energy limit. Yet, it does not contain the relativistic effects for the total power (eq. (4.24)) and its spectral and spatial distributions from the previous section. For a complete model that includes the relativistic effects a new amplitude $A(\mathbf{r}, t)$ is introduced in eq. (4.27) such that

$$\mathbf{E}_{rel}(\mathbf{r}, t) = \left(A(\mathbf{r}, t) \begin{pmatrix} \sin(t\omega_c + \varphi_0 + \phi) \\ -\cos(\theta) \cos(t\omega_c + \varphi_0 + \phi) \\ 0 \end{pmatrix} \right)_{t_r}. \quad (4.28)$$

The amplitude $A(\mathbf{r}, t) \sim \sqrt{P(\mathbf{r}, t)}$ is calculated with the direction dependent power of the fundamental frequency from the previous section and serves as a relativistic correction to eq. (4.27). Together with the use of the relativistic cyclotron frequency ω_c from eq. (4.3) this results in a full relativistic description of the fundamental frequency component of the radiation field.

Polarization

The result in eq. (4.28) can easily be recognized as the general form of elliptical polarization in the ϕ - θ plane of the observer. This is due to the factor $\cos(\theta)$ scaling down the amplitude of the E_θ -component with higher polar angles, while the amplitude of the E_ϕ component remains constant. The polarization ellipses are characterized by their axial ratio which is the ratio between the length of the major and minor axes. In this case it is $AR = \frac{1}{\cos(\theta)}$.

Figure 4.8 shows a variety of cases for the polarization ellipse under different observer angles θ . For the case where $\theta = 0^\circ$ and $\hat{\mathbf{n}}$ is parallel to \mathbf{B} , the ratio is $AR = 1$, indicating perfect circular polarization in the ϕ - θ -plane. In this configuration the ϕ - θ -plane is identical to the original x - y -plane. For the case where $\theta = 90^\circ$ the ratio is $AR = \infty$, corresponding to linear polarization along the azimuthal axis ϕ . Angles in between are elliptic in general with $AR > 1$. Finally, a comparison of these results to the schematic in fig. 4.3 is helpful for the geometric interpretation of the polarization: For an observer viewing the electron's circular motion under an angle θ , this circle will appear as an ellipse, corresponding to the polarization ellipses in fig. 4.8.

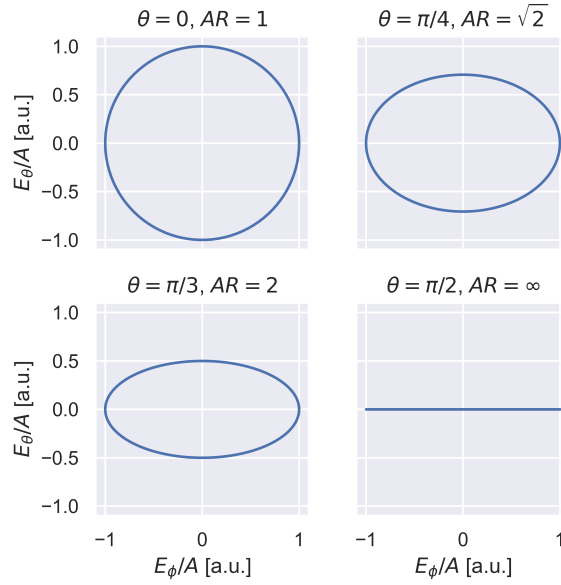


Fig. 4.8.: Observed polarization ellipses in the ϕ - θ -plane for different polar angles θ . In each plot title the axial ratio AR of the polarization ellipse is indicated. Field components are normalized to A .

Phase

In addition to the $\pi/2$ phase shift between the E_ϕ and E_θ components and the initial helix phase φ_0 , both components have an additional shift ϕ . This phase ϕ is the azimuthal angle ϕ of the observer (see fig. 4.3) and it is defined with respect to an arbitrary direction that is orthogonal to \hat{B} . Therefore, this result only determines a relative phase shift with respect to the reference direction instead of an absolute one. That means any two observers at the same polar elevation θ and at the same distance to the electron will observe radiation with a phase shift of $\Delta\phi$ equal to their azimuthal angular distance. Figure 4.9 shows normalized instantaneous amplitudes of \mathbf{E}_r in the x - y -plane of an electron at $\rho = 0$. Regions with matching colors are indicative of identical phases. The characteristic spiral pattern formed by these regions is a result of the azimuth phase shift ϕ .

The phase of the field at a position (ρ, ϕ) in the x - y -plane evaluated at the retarded time $t_r = t - \frac{\rho}{c}$ is

$$\varphi = \omega_c \left(t - \frac{\rho}{c} \right) + \varphi_0 + \phi.$$

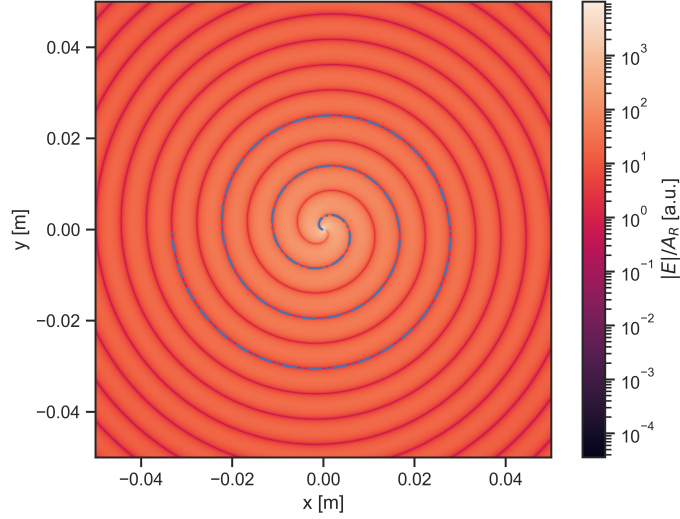


Fig. 4.9.: Instantaneous field amplitudes at time t in the x - y -plane normalized to $A_R = A \cdot R$. Matching colors share the same phase. The pattern is recognized as an Archimedean spiral, which is overlaid as dashed blue line. Without the log-scale the amplitudes drop off quickly, due to the $\frac{1}{R}$ scaling.

As a result, for positions with equal phase δ , the following relationship between ρ and ϕ holds:

$$\rho = \frac{c}{\omega_c} \left(\phi + \underbrace{\omega_c t + \varphi_0 - \delta}_{\text{constant for all } (\rho, \phi)} \right). \quad (4.29)$$

This result can be recognized as an Archimedean spiral, which is defined by

$$\rho = b \cdot \phi_A,$$

which relates the polar angle ϕ_A , to the radial distance ρ of each point of the spiral's curve, where the parameter b sets the distance between spiral loops. From eq. (4.29) one finds that $b = \frac{c}{\omega_c}$ and $\phi_A = \phi + \text{const}$, where the constant part adds an angular offset to the spiral. The corresponding Archimedean spiral is overlaid as the dashed blue line in fig. 4.9.

Simulated Data of an Antenna Array CRES Experiment

” *Creativity is just connecting things.*

— **Steve Jobs**
Co-Founder of Apple Inc.

Simulated data are an indispensable ingredient for research in modern fundamental physics. They play a critical role for steering the design process of new experiments, which can often be multi-million Euro investments, and help assessing the capability of reaching experimental goals and evaluating the performance of data analysis algorithms. Furthermore, even in the presence of experimental data, simulated data still play a pivotal role in gaining the desired insights about a physical theory from the experimental data. The data are the product of a mathematical model for the simulated system with its underlying physical processes that ideally provides an accurate description of reality within the bounds of its modeling assumptions. However, such modeling accuracy can obviously not be guaranteed for the complexity of a real experiment. This is why a thorough validation and continuous challenging of the model are imperative.

This chapter discusses how data taken by a CRES experiment can be simulated. It briefly covers the implementation of a simulation model with the software packages Kassiopeia and Locust and their shortcomings both in computational efficiency and modeling accuracy. These are addressed by another implementation with the newly developed software package CRESana, which is one of the main contributions of this thesis.

5.1 Overview of Simulation Principles

A full simulation of a CRES experiment yields the data measured by a detector over a time t_m for a defined detector setup and a given source of electrons. The data of each individual antenna are a time series of a time varying digital voltage, which

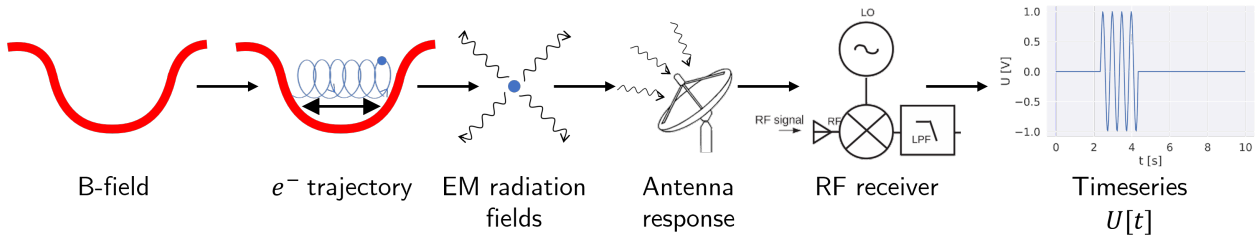


Fig. 5.1.: Electron track simulation steps.

is obtained by sampling at a fixed rate f_s . The samples are taken as in-phase and quadrature (I/Q) data, which is an RF sampling technique that samples a real signal and its 90° phase-shift with two separate analog-to-digital converters (ADCs) and combines them into complex samples¹. Therefore, the data are a complex $n_a \times n_s$ matrix \mathbf{U} , where n_a denotes the number of antennas and $n_s = t_m \cdot f_s$ is the number of samples per antenna.

Figure 5.1 shows the essential steps for the simulation of electron data. The first step in the simulation is the generation of a 3-dimensional magnetic field based on the distribution of electrical currents of the detector geometry. With the magnetic field provided, the next step is to compute the electron trajectory, which is fully determined by its initial parameters θ . The trajectory encompasses the time development of the electron's position and momentum, which are obtained by solving the equations of motion. This is followed by the calculation of the EM fields of the emitted radiation, which are defined by eqs. (4.21) and (4.22) for any position r . It is sufficient to calculate the fields only at the positions of the antenna elements in the array. After that the characteristic antenna response is calculated, which simulates the voltages induced in the antenna circuit by the input EM fields. The last step is the simulation of the RF receiver, which includes amplifiers, frequency mixing, low-pass filtering and digitization. The steps in the diagram following the electron trajectory are done for each antenna element independent of each other and produce a time series $U[t]$ for each antenna.

For the main application in Project 8 the source of electrons would be a tritium gas inside the detector volume at some given density. In a stochastic process the source produces decay electrons at positions inside the detector volume. When an electron *event* occurs the electron moves through the gas and can randomly scatter off of gas molecules, which results in a sudden change of kinetic energy and pitch angle. This

¹This has advantages since it always preserves instantaneous amplitude and phase of the signal and it allows for distinguishing positive and negative frequencies.

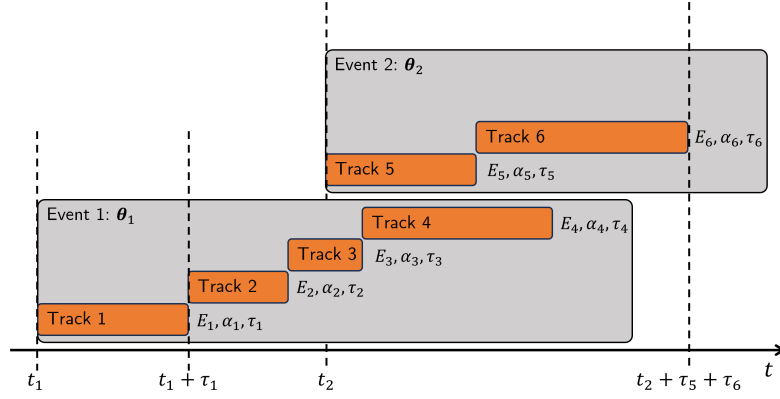


Fig. 5.2.: Chart depicting the concept of multiple concurrent tracks and events. An event occurs at t_1 with initial parameters θ_1 . After τ_1 the electron scatters which starts a new track with length τ_2 and changes energy and pitch angle to E_2, α_2 . Two more scatters occur before the first event ends. A second event appears at t_2 with different initial parameters θ_2 . Its two tracks partly overlap in time with the first event.

can occur multiple times and the event ends when the detector is no longer sensitive to the electron energy, or when the electron leaves the trap.

The periods of an event between scatters are called electron *tracks*. An electron track can be characterized by eight parameters: they are the electron's initial position and momentum vectors (x, y, z) and (p_x, p_y, p_z) , the start time t_0 and the track duration τ which is the time it takes before the electron scatters. An equivalent but more convenient representation of these parameters is $\theta = (E_{\text{kin}}, \alpha_0, \rho, \phi, \varphi_{\text{axial}}, \varphi_{\text{cycl}}, t_0, \tau)$, where α_0 is the pitch angle at the minimum of the magnetic field², ρ the radial position, ϕ the azimuthal angle of the position, φ_{axial} the initial phase of the axial motion and φ_{cycl} the initial phase of the cyclotron motion.

For the initial track all parameters except for τ are random variables arising from the stochastic production process of the source, while τ is a random variable that depends on the electron's mean free path for all tracks. The energy and pitch angle in subsequent tracks are random variables that depend on their previous values and the scattering mechanism, typically with $E_{\text{kin}i+1} < E_{\text{kin}i}$. Figure 5.2 depicts the concept of tracks and events along the time axis.

In this thesis the focus is on the simulation of deterministic electron tracks for given parameters θ , since this is the part that is specific to CRES while the stochastic event generation process in tritium is known from other tritium spectroscopy experi-

²Throughout this thesis α_0 will be called simply "the electron's pitch angle" for better readability although it actually always means "the electron's pitch angle at the minimum of the magnetic field". It should not be confused with the instantaneous pitch angle $\alpha(t)$.

ments [74, 75]. Combining the track simulation with a Monte Carlo event generator would be straightforward: the event generator draws tracks from the distributions that are appropriate for the gas which yields inputs θ_i to be processed by the track simulator. All simulated tracks are then combined by addition of the time series of output voltages. Assuming that electrons do not interact with each other it is also possible to have multiple overlapping events generated with this approach.

5.2 Kassiopeia

Kassiopeia is a particle tracking software package written in C++ and developed by the KATRIN collaboration [74, 76]. It can calculate the electric and magnetic fields of charge and current distributions in complex geometries, produce different types of particles and track them through the geometry. The particle production can be configured in many ways to enable the generation of particle properties from static inputs or by drawing from probability distributions. Available distributions include those relevant to tritium beta decay. Particles are tracked by solving an equation of motion, which can be configured by the user. During the particle tracking interactions with the medium like scattering off the tritium gas and interactions with surfaces can be configured as well.

In an electron track simulation for Project 8, Kassiopeia is responsible for the first two steps in fig. 5.1. The geometry only consists of one or more current carrying electromagnetic coils, which produce the magnetic field for the simulation in addition to a background field. For track simulations with given parameters θ including track length τ , an electron is produced with fixed properties and the gas interactions are disabled since scattering ends a track. In the baseline configuration the trajectory uses an equation of motion that tracks only the guiding center of the cyclotron motion with the addition of drift motions (see section 4.1), which is solved with an 8th-order Runge-Kutta integrator. But in general Kassiopeia also supports the exact equations of motion without guiding center approximation and other integration algorithms. For more details about possible ways to configure Kassiopeia readers should consult the Kassiopeia documentation [77].

5.3 Locust

Locust, developed by the Project 8 collaboration, is another software package written in C++ for the primary purpose of simulating CRES experiments [78, 79]. It models the response of an antenna and the full RF receiver chain to time-varying electromagnetic fields. The simulation of the receiver chain also includes simulation of an ADC, thus generating digital data with limited bit-depth in the same data file format as data from a laboratory experiment. The electromagnetic input signal can be configured to a variety of internal sources including simple test signals and external ones.

For the use case of an electron track simulation Locust implements the last four steps in fig. 5.1 but it needs a CRES electron for the signal source. This is accomplished by compiling Locust with Kassiopeia as a submodule. The combined Kassiopeia-Locust package runs with two concurrent threads for both subpackages, which enables Locust to query the electron trajectory at runtime. The electromagnetic fields are calculated in Locust by numerically solving eqs. (4.21) and (4.22) for the trajectory from Kassiopeia. The trajectory and field samples are both temporarily stored in ringbuffers and the threads and their access to the buffers are properly coordinated such that trajectory and RF sampling are synchronized.

Unfortunately, working with Locust for this thesis has uncovered two major bugs in the simulation of an antenna array setup which have not been fixed at the time of writing. The first one is introduced from an incorrect scheme for applying the antenna response to undersampled signals with the goal of saving computation time. The second one was found in the solution of eqs. (4.21) and (4.22) which produces incorrect results. For more details on these bugs see appendix B.

Aside from these bugs the main shortcoming of using Kassiopeia-Locust is the computation time of electron simulations for an antenna array. For a short $\sim 40\mu\text{s}$ track in a moderately sized array of 60 antennas a runtime of ~ 90 min was found. The main culprits for the long computation time are the numerical solving of eq. (4.21) for the radiation fields, and the simulation of the antenna response, which involves applying digital filters with convolution operations on a densely sampled time series. Both need to be done per antenna. It is estimated that both the track length and the array size will need to be a factor 10-100 higher with a proportional increase in computation time. The resulting runtime is unacceptable considering that intended use cases for the simulation range from application in optimization routines to generating large datasets for training machine learning algorithms.

5.4 CRESana

CRESana [80] was developed as an alternative to the Kassiopeia-Locust implementation of a CRES simulation with the goal of improving the performance by taking a different simulation approach. The CRESana approach employs analytic results, certain approximations and exploits symmetries and periodicity in order to reduce the computational cost of the numeric simulation. In CRESana, all the steps of CRES simulation from fig. 5.1 are integrated into a single package in order to simulate the I/Q voltage time series. Although implemented purely in Python the heavy use of the NumPy [81] and SciPy [82] libraries and the more efficient simulation approach allow for a significant reduction in computation time: CRESana is able to simulate the same setup which takes ~ 90 min in Kassiopeia-Locust in (10.5 ± 0.4) s — a speedup of ~ 500 .

An advantage of the Python implementation and the reliance on NumPy arrays as the main data structure is the effortless interfacing of the simulated data to the numerous libraries for optimization and machine learning that are available in the Python ecosystem and very widely used by physicists. In addition to the improved runtime this also greatly reduces IO overhead times for work with these Python libraries since the data are not written to a file first, which is mandatory with Kassiopeia-Locust.

The following sections go over the details of how CRESana combines electron motion, the electric fields and the antenna response in order to form the recorded spectrum based on the theory discussed in chapter 4. The section structure follows the diagram in fig. 5.1.

5.4.1 Electron Trajectory

The first two steps in fig. 5.1 are the magnetic field and the electron trajectory simulation. The simulation of the electron trajectory $\mathbf{T}(t) = (\mathbf{r}_s(t), \mathbf{B}(t), E_{\text{kin}}(t), \alpha(t))$ yields the time evolution of all parameters that are relevant to the calculation of the emitted radiation. This includes the position $\mathbf{r}_s(t)$, the magnetic field vector $\mathbf{B}(t)$, the kinetic energy $E_{\text{kin}}(t)$ and the pitch angle $\alpha(t)$.

The trajectory is determined by the three-dimensional magnetic field $\mathbf{B}(\rho, z)$, which is restricted in CRESana to be both rotationally symmetric around the z -axis and symmetric with respect to the plane $z = 0$. Assuming adiabatic motion for electrons in these fields (see section 4.1) the simulation of the trajectory uses the guiding

center approximation. In the lowest order of approximation the motion is a helix spiraling around the magnetic field lines. Therefore, instead of tracking the exact electron position, it is sufficient to track the position of the guiding center $\mathbf{r}_s(t)$ as the local approximate location of the helix axis. If required, the exact position could be reconstructed from that using the gyration radius with eq. (4.4).

Still, the motion is only an exact static helix in an entirely uniform magnetic field. With the higher orders of the approximation, the guiding center slowly moves in a circle around the symmetry axis of the magnetic field (i.e. its azimuth position ϕ changes) due to the drift motions introduced in section 4.1 [47]. Finally, in order to achieve the magnetic mirror effect the field lines have to converge along the z -axis and thus the radial position ρ of the field line followed by the guiding center eventually decreases at higher z -positions (see fig. 5.4). Accordingly, the full time evolution of the guiding center is

$$\mathbf{r}_s(t) = \begin{pmatrix} \rho(z(t)) \cos(\phi(t) + \phi_0) \\ \rho(z(t)) \sin(\phi(t) + \phi_0) \\ z(t) \end{pmatrix}, \quad (5.1)$$

where ϕ_0 determines the initial azimuth position of the guiding center. Figure 5.3 visualizes the kind of motion that is described by eq. (5.1).

With the guiding center position known, the remaining trajectory parameters are easy to calculate: the radial and z position determine the magnetic field $\mathbf{B}(t) = \mathbf{B}(\rho(z(t)), z(t))$, the magnetic field determines the pitch angle $\alpha(t) = \alpha(\mathbf{B}(t))$ with eq. (4.5) and the pitch angle and magnetic field determine the kinetic energy $E_{\text{kin}}(t) = E_{\text{kin}}(\mathbf{B}(t), \alpha(t))$.

Therefore, CRESana handles the trajectory simulation in a few calculation steps in a specific order where each step requires the results of the step before:

1. A magnetic field $\mathbf{B}(\rho, z)$ based on the setup
2. The field line for this magnetic field based on the initial guiding center position (ρ_0, z_0) which yields $\rho(z)$
3. The axial position $z(t)$ and the radial position $\rho(t) = \rho(z(t))$
4. The instantaneous pitch angle, magnetic field and kinetic energy
5. The azimuth position $\phi(t)$
6. The full guiding center position $\mathbf{r}_s(t)$

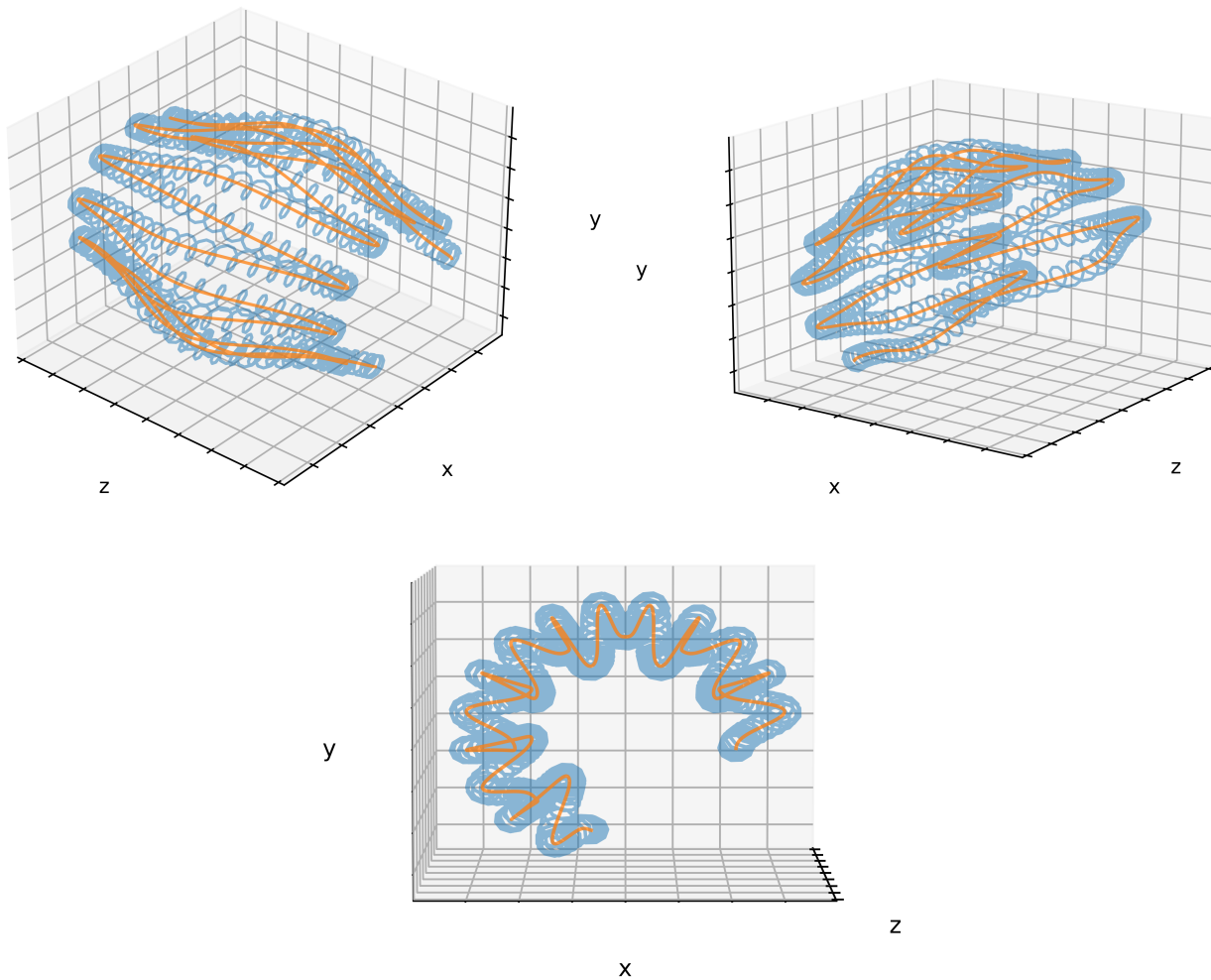


Fig. 5.3.: 3D electron motion with an orange line representing the guiding center and a blue line indicating the electron's actual position. An exaggerated synthetic motion is depicted aimed at improving visualization of key features rather than replicating the exact physical simulation.

Magnetic Field and Field Line

The standard simulation setup uses a field that is generated by a number of coaxial electromagnetic coils — all centered around the z -axis — and a background field $B_0\hat{z}$. The setup specifies the coils' locations z_i along the z -axis, the radii R_i and the currents I_i . The electromagnets are approximated as simple current loops for which analytic solutions with elliptic integrals can be found for the field $\mathbf{B}_{z_i R_i I_i}(\rho, z)$ and all its derivatives in [83]. Since magnetic fields satisfy the superposition principle, CRESana evaluates the field locally for any given position ρ, z as $\mathbf{B}(\rho, z) = \sum_i \mathbf{B}_{z_i R_i I_i}(\rho, z) + B_0\hat{z}$. Analogously all derivatives that are required for the drift velocity (eq. (4.18)) are evaluated via superposition. Figure 5.5a shows the absolute

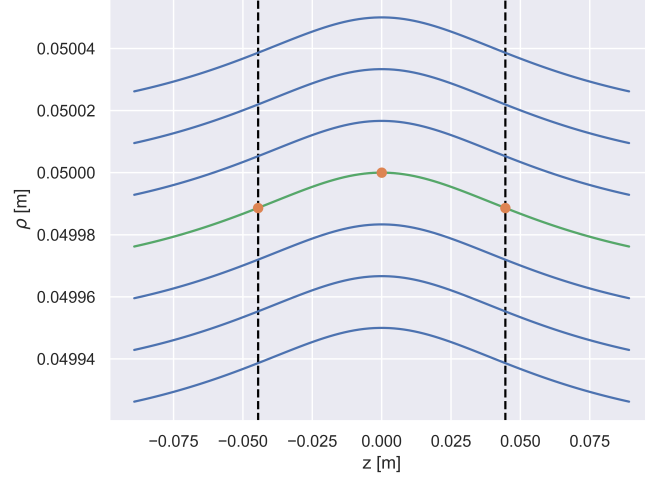


Fig. 5.4.: Plot of magnetic field lines in a trapping field. An electron (orange dots) starting at $(\rho_0, z_0) = (0.05, 0)$ m would follow the green field line to its maximal z extents which is marked by the dashed lines. For this reason the electron's radial position depends on its z position.

value of the magnetic field of a single trapping coil for a fixed radial position ρ simulated by this approach in CRESana.

The second option for a setup in CRESana is a field that is given by a polynomial function $B_z(z)$ for the z component of the field vector on the cylinder axis $\rho = 0$. This analytically determines the full vector $\mathbf{B}(\rho, z)$ and the derivatives at any location via the Legendre polynomial solution (see section 4.2) and thus CRESana can evaluate them locally.

The field line $\mathbf{r}_{\text{FL}}(s) = (\rho(s), z(s))$ is defined via the differential equation [84, 74]

$$\frac{d\mathbf{r}_{\text{FL}}(s)}{ds} = \frac{\mathbf{B}(\rho(s), z(s))}{|\mathbf{B}(\rho(s), z(s))|},$$

where s is an arbitrary parameter. For the solution a simple Euler integration scheme is used:

$$\begin{aligned} \rho_{i+1} &= \rho_i + \frac{B_\rho(\rho_i, z_i)}{|\mathbf{B}(\rho_i, z_i)|} \Delta s \\ z_{i+1} &= z_i + \frac{B_z(\rho_i, z_i)}{|\mathbf{B}(\rho_i, z_i)|} \Delta s. \end{aligned}$$

The values (ρ_i, z_i) are iteratively calculated starting with the initial guiding center position (ρ_0, z_0) until $z_i > z_{\text{stop}}$.

Field lines calculated by this approach are depicted in fig. 5.4. The step size Δs and the stopping value z_{stop} are implemented as configuration parameters of the trap

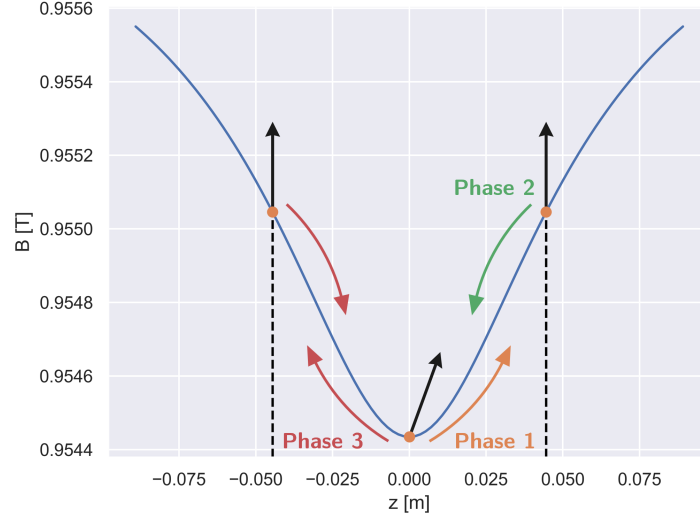
setup and in general both need to be adapted to the physical dimensions of the setup. Δs can trade-off accuracy and computation time and should be small compared to z_{stop} which needs to be set high enough that it is greater than the highest z position any electron of interest can reach in the trap. The resulting values (ρ_i, z_i) are used to form a function $\rho(z)$ with a B-spline interpolation from the SciPy library [82, 85].

Axial Motion

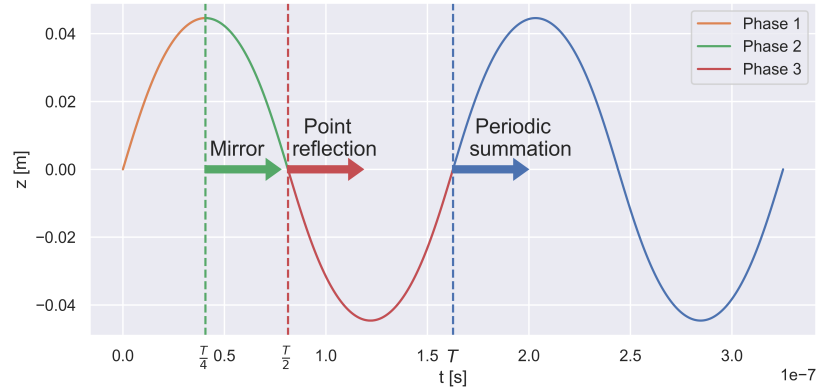
The trajectory of axial motion is found by solving the integral in eq. (4.6) along the electron's magnetic field line using the field magnitude $B(z) = |\mathbf{B}(\rho(z), z)|$. The resulting motion in the magnetic field is schematically outlined in fig. 5.5a with its three phases. Closed-form solutions of the integral only exist if $B(z)$ is polynomial with order ≤ 2 and have been implemented for the harmonic trap and the square well trap (see appendices A.1 and A.2). For all other cases the integration is carried out numerically with an adaptive quadrature method via SciPy's interface to the FORTRAN library QUADPACK [82, 86].

The first step towards the solution of the axial trajectory is the calculation of the upper integration bound z_{max} , which is the maximal z position that is reached by the simulated electron. It is defined by the root of the integrand's denominator and found by numerically solving the equation $B(z_{\text{max}}) - B(z') = 0$ for z' using an implementation of the secant method in SciPy [82]. $B(z_{\text{max}})$ is determined by the electron's pitch angle α_0 and the magnetic field at the minimum $B(0)$ and can be calculated without prior knowledge of z_{max} as $B(z_{\text{max}}) = \frac{B(0)}{\sin^2(\alpha_0)}$. This follows from eq. (4.5) (adiabatic invariance) using the fact that the electron reaches an instantaneous pitch angle of $\alpha = 90^\circ$ at $B(z_{\text{max}})$.

After the calculation of z_{max} , the simulation can proceed to eq. (4.6). This equation is a function $t(z)$, which returns the time it takes for the electron to move from $z(0)$, its z position at $t = 0$, to an arbitrary position $z \leq z_{\text{max}}$. By setting the start point as $z(0) = 0$, $t(z_{\text{max}})$ becomes the time it takes for the electron from the minimum of the magnetic field to the maximum it can reach with its pitch angle, which corresponds to phase one in fig. 5.5a. Once the electron reaches z_{max} , it comes to a halt with an instantaneous pitch angle of $\alpha = 90^\circ$, i.e. all its kinetic energy is momentarily in the cyclotron motion only. In the second phase the electron turns around and falls back into the magnetic minimum, repeating the same motion from z_{max} to 0 in reverse. From there it repeats the first two phases in negative z direction due to the symmetry of the magnetic field and finally it reaches $z = 0$ again after completing a full axial period.



(a) Plot of the absolute magnetic field of a trapping coil along the cylinder axis simulated in CRESana. In addition, the axial motion of an electron (orange dots) is schematically outlined. The black arrows indicate the electron's momentum direction relative to the z -axis and the dashed lines mark the maximum z -extent $\pm z_{max}$.



(b) Construction of full axial solution. The initial solution (orange) is valid for $t < \frac{T}{4}$ and is extended to the full solution by two reflections and periodic summation. The three phases here correspond to the phases marked in (a). The blue part of the curve marks the solution that is valid for any time $t > T$.

Fig. 5.5.: Construction of the axial solution.

As a result, $t(z_{max}) = \frac{T}{4}$ is the first quarter of the axial period and it is sufficient to evaluate this part of the axial motion since it is similar in the other three phases. However, in order to resolve the motion between $t = 0$ and $t = \frac{T}{4}$ in detail it is mandatory to evaluate $t(z)$ for a number of z positions, each of which is a separate integration with a different upper bound z . This is done with N_z equidistant points

$z_i \in [0, z_{max}]$ yielding $t_i = t(z_i)$, where N_z is a configuration parameter of the simulation. This is visualized in fig. 5.6.

From the results a B-spline interpolation function is constructed with t as independent variable yielding $z_{T/4}(t)$, an approximation to the inverse of $t(z)$ on the interval $[0, T/4]$. For the full axial trajectory $z(t)$ with $t \in [0, \tau]$, the solution $z_{T/4}(t)$ is extended to the full axial period by exploiting the symmetry of the trap and subsequently the full track length by periodic summation. This process is shown in fig. 5.5b.

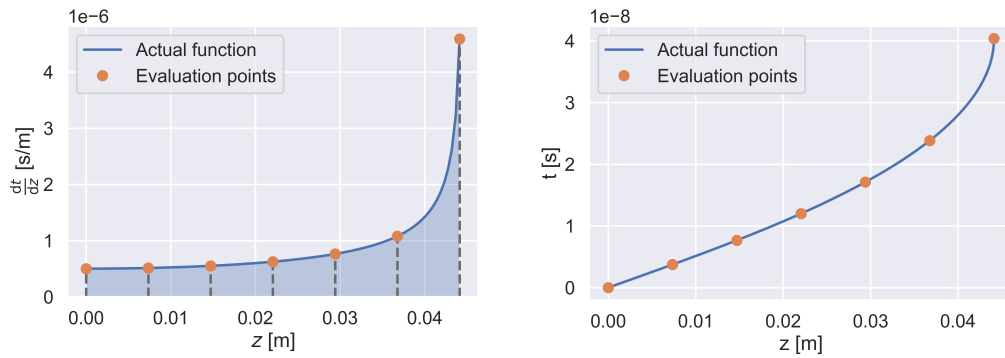


Fig. 5.6.: Approach for calculating the axial motion on the first quarter period. Equation (4.6) is evaluated for a number of z positions (right) and an interpolation function can yield the inverse $z(t)$. Each of the points on the right is the result of separately integrating the function on the left from 0 to the respective z position. At the rightmost point the function approaches a singularity.

Once the extended function $z(t)$ has been constructed, it can be evaluated for an arbitrary number of sample times along the full track without additional computational cost for integrations. The extra cost of the integration is fixed prior to the evaluation and can be adjusted with the simulation setup parameters. E.g. the number of z evaluation points N_z and thus the number of integrals can be reduced to accomplish faster but less accurate results.

The integration itself has proven to be rather challenging since its upper bound z_{max} is a singularity. Several different numeric integration schemes have been investigated. This includes the cumulative trapezoid rule from SciPy and a cumulative implementation of Simpson's rule, which both result in good computational performance but with an unacceptably high integration error close to the singularity. In addition, the so-called tanh-sinh quadrature method, which is well suited for integrals with singularities [87], has been assessed with an implementation from the mpmath Python package [88]. The integration quality turned out comparable to QUADPACK but with a significantly slower computation time.

In the final solution with QUADPACK, the leading contribution to the computation time of the trajectory solution is the magnetic field evaluation $B(z)$ in the integrand. The issue lies in the repeated evaluation of the integrand at multiple z -positions. These are dynamically selected by the adaptive integration routine from QUADPACK. Furthermore, since each point in fig. 5.6 corresponds to a separate call to the integration routine, the evaluated z -positions are not reused between different integrals. This leads to a large cumulative number of integrand evaluations.

However, the calculation of the magnetic field itself can be accelerated by using an interpolation function for $B(z)$. This interpolation function is constructed ahead of time by evaluating $B(z)$ along the field line at a number of positions $z_i \in [0, z_{max}]$. Depending on the number of positions, the evaluation of the interpolation function can be significantly faster than the calculation of the original $B(z)$ solution for the current loops. The number of positions is another setup parameter that can be used to trade off computation time and accuracy of this additional performance optimization. The use of this magnetic field interpolation is entirely optional.

Energy

The kinetic energy of the electron cannot be treated as constant since it emits cyclotron radiation that carries away power which is the derivative of the energy, i.e. $\frac{dE_{kin}}{dt} = -P(t)$. In this differential equation $P(t)$ is the total radiated relativistic power given by the Larmor equation eq. (4.24), which depends itself on $E_{kin}(t)$ due to the relativistic β factor. Using $\beta = \sqrt{1 - \left(\frac{m_0c^2}{m_0c^2 + E_{kin}}\right)^2}$ one finds the differential equation

$$\frac{dE_{kin}(t)}{dt} = -\frac{P_0(t)}{(m_0c^2)^2} \left(E_{kin}^2(t) + 2m_0c^2 E_{kin}(t) \right), \quad (5.2)$$

where $P_0(t) = P(t) \frac{1-\beta^2}{\beta^2}$ is the non-relativistic power which still depends on time due to the pitch angle and magnetic field.

The solution to this differential equation was found with Wolfram Mathematica [89]

$$E_{kin}(t) = \frac{2m_0c^2 E_{kin0}}{(m_0c^2 + E_{kin0}) e^{2E_L(t)/(m_0c^2)} - E_{kin0}} \quad (5.3)$$

where E_{kin0} is the initial kinetic energy and E_L is the integrated power

$$E_L(t) = \int_0^t P_0(t') dt'. \quad (5.4)$$

This integration is carried out numerically with the cumulative trapezoid rule from SciPy [82], which efficiently calculates the integral for a great number of time samples t_i by accumulating the approximate integrals of sub-intervals:

$$E_L(t_i) = \sum_{k=1}^i \int_{t_{k-1}}^{t_k} P_0(t') dt' = E_L(t_{i-1}) + \int_{t_{i-1}}^{t_i} P_0(t') dt' .$$

With the trapezoid approximation it is only required to evaluate the integrand a single time for each t_i , which can still be done at small computational cost. A faster approximate solution would be possible as well with the reasonable assumption that $P(t) \approx \text{const}$ on the relevant timescales which simplifies the solution to $E_{\text{kin}}(t) = E_{\text{kin}0} - P \cdot t$. However, CRESana still uses the exact solution of eq. (5.2) since its computational impact was found to be negligible.

A minor effect that is caused by the energy loss is a time dependence of the axial frequency. The solution of the axial motion in eq. (4.6) depends on the kinetic energy via the velocity v_0 . So far the presented simulation approach of the axial motion assumes that v_0 is constant, which is a valid assumption over a single axial period. However, for long simulation times the axial frequency should be slowly decreasing due to the energy loss. This was confirmed by comparison to a trajectory from Kassiopeia, which incorporates a term for energy loss from cyclotron radiation into the equations of motion. It is not possible to implement the energy loss properly into the axial motion with the approach of the periodic summation. Therefore, the effect was approximated by a simple linear correction: instead of calculating $t(z)$ with eq. (4.6) the simulation calculates the arch length $s(z) = t(z) \cdot v_0$ and the corresponding inverse interpolation function $z(s)$. This is then evaluated as $z_{T/4}(t) = z(v(t) \cdot t_{T/4}(t))$ where $t_{T/4}(t)$ is the time t converted to the corresponding time in the first phase of the axial motion (figs. 5.5a and 5.5b). For the conversion the axial period is rescaled with the current velocity as $T = s(z_{\text{max}})/v(t)$, assuming that only the maximum arch length is constant but the period length is not. This corresponds to a 1st order Taylor expansion that assumes $\frac{dv(t)}{dt} \approx 0$.

Azimuth Motion

The azimuth motion $\phi(t)$ from eq. (5.1) is caused by the drift of the guiding center. It can be calculated with the instantaneous total drift velocity v_D using eq. (4.18), which is not constant since it depends on the instantaneous pitch angle and energy and the local magnetic gradient and curvature at the current z -position. The

instantaneous drift phase is then the accumulated angular position as a result of that motion:

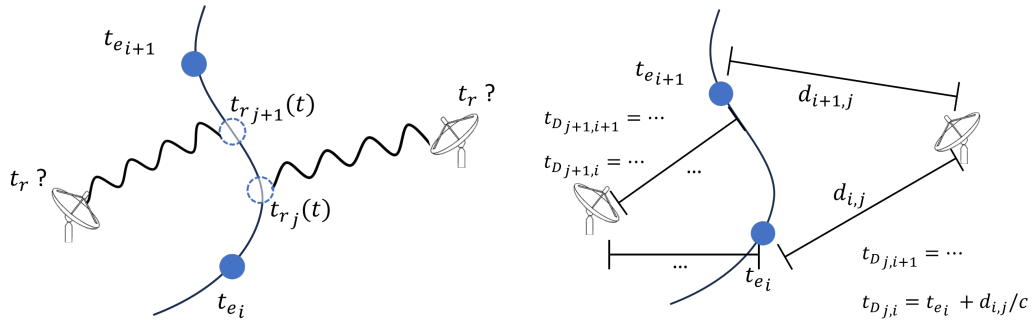
$$\phi(t) = \int_0^t \frac{v_D(z(t'))}{\rho(z(t'))} dt'. \quad (5.5)$$

This is the same kind of integration as eq. (5.4) and thus is calculated in the same way using the trapezoid rule.

5.4.2 Electric Field at Antenna

The electron trajectory simulation is followed by the calculation of the electric field at the antennas (see third panel in fig. 5.1). The three-dimensional electric field $\mathbf{E}_r(t) = \mathbf{E}_{rel}(\mathbf{r}, t)$ is calculated with the approximation of the fundamental frequency with relativistic correction. The vectorial form of this field is given in eq. (4.28) and it can be characterized by an instantaneous power $P_E(t)$ and phase $\varphi_E(t)$. The latter needs to take into account that the cyclotron frequency ω_c is time dependent due to the Doppler effect and magnetic field variations along the electron trajectory.

Retarded Time



(a) Retarded times of two antennas for same sample time visualized for sampled trajectory. Retarded times are in between trajectory sample times t_{e_i} . (b) Calculation of the distances between trajectory samples and antennas for the calculation of the delay times.

Fig. 5.7.: Visualization of retarded time calculation.

Since the electron is at motion the field at the antenna at time t given in eq. (4.28) is evaluated at the retarded time t_r , which is defined in eq. (4.23) where \mathbf{r} is the antenna position and \mathbf{r}_s is the electron's position. The goal of this calculation is visualized in fig. 5.7a: for each sample time t and each antenna j the corresponding

retarded time $t_{rj}(t)$ has to be determined. The exact times are (in most cases) different for each antenna and do not coincide with existing trajectory samples.

Unfortunately, eq. (4.23) has no closed-form solution for a helical trajectory. For the solution the electron trajectory is sampled at twice the sampling rate and the delay times to all antennas are calculated. The delay time is the time when radiation from a trajectory sample has propagated to an antenna at position \mathbf{r}_j , i.e. $t_{Dj}(t_e) = \frac{|\mathbf{r}_j - \mathbf{r}_s(t_e)|}{c} + t_e$. By interpolating the results of each antenna individually (again using B-spline interpolation in SciPy [82, 85]), the retarded time at an antenna's sampling time is evaluated as $t_{rj}(t) = t_{Dj}^{-1}$. This approach assumes that for each sample time there is only a single shortest path from the electron trajectory to the antenna and reflection effects are negligible.

Once the retarded time $t_{rj}(t)$ is known the trajectory parameters are used to calculate the cyclotron frequency $\omega_c(t_{rj}(t))$ with eq. (3.1), the larmor power $P_{la}(t_{rj}(t))$ with eq. (4.24) and the distance vector $\mathbf{d}(t_{rj}(t)) = \mathbf{r}_s(t_{rj}(t)) - \mathbf{r}_j$.

Power

In the next step, the incident field power at an antenna can be determined using Frii's transmission equation, which describes the relationship between the received and transmitted power of two antennas in each other's far field. The output power of the receiving antenna is given by [90]

$$P_r = G_r G_t P_t \left(\frac{c}{2\omega d} \right)^2 \quad (5.6)$$

where P_t is the transmitted power, G_t the gain of the transmitting antenna, G_r the gain of the receiving antenna, d their distance, and ω the frequency of the transmitted wave. By treating the electron as a transmitting antenna that transmits EM radiation with the Larmor power P_{la} and assuming that the electron and the receiving antennas are in each other's far field, eq. (5.6) can be applied to calculate the output power of the antennas in the array.

In this treatment, the receiver gain G_r in eq. (5.6) results from the antenna response function, which is discussed in the next section. Assuming $G_r = 1$ for now yields the instantaneous power of the field incident on the antenna:

$$P_{Ej}(t) = P_{la} G_e \left(\frac{c}{2\omega_c |\mathbf{d}|} \right)^2 \Big|_{t_{rj}(t)}, \quad (5.7)$$

where G_e denotes the electron's "transmitter gain", and all symbols are evaluated at the retarded time.

The electron gain G_e implements the angular power distribution from eq. (4.25), evaluated only at ω_c . As discussed in section 4.3.2, the field is reduced to its fundamental frequency, as the antenna is unable to detect the harmonics of the field (see the antenna's frequency response in fig. 5.8). In practice, this means that the anisotropy seen in fig. 4.5 is combined with a power damping factor to account for the fraction of larmor power that is radiated with the undetected harmonics as seen in fig. 4.6. Both these factors depend on the direction $\mathbf{d}(t_{r,j}(t))$ relative to the B-field.

Phase

The last step is the calculation of the instantaneous field phase at each antenna. It is given by the integral over all past instantaneous field frequencies. The frequency at the antenna is the Doppler-shifted cyclotron frequency. The calculation of the Doppler shift requires the relative velocity between electron and antenna, which can be calculated using the definition of the retarded time in eq. (4.23).

Using the fact that $t > t_r$ and introducing the instantaneous distance d as a shorthand for the right-hand side of that equation it can be written as

$$c(t - t_r(t)) = d(t_r(t)).$$

Differentiating both sides with respect to t yields the velocity:

$$\begin{aligned} \frac{d}{dt}c(t - t_r(t)) &= \frac{d}{dt}d(t_r(t)) \\ \Leftrightarrow c\left(1 - \frac{dt_r(t)}{dt}\right) &= \frac{dt_r(t)}{dt} \frac{d}{dt_r}d(t_r) \\ \Leftrightarrow v_e(t_r) &= \frac{c\left(1 - \frac{dt_r(t)}{dt}\right)}{\frac{dt_r(t)}{dt}}. \end{aligned}$$

Using this relative velocity yields for the Doppler-shifted frequency in the receiver's rest frame:

$$\omega(t) = \omega_c(t_r(t)) \frac{1}{1 + \frac{v_e(t_r(t))}{c}} = \omega_c(t_r(t)) \frac{dt_r(t)}{dt}.$$

Consequently, the phase at a receiver including Doppler shifts can be written as an integral over instantaneous cyclotron frequencies with a coordinate transformation

$$\varphi(t) = \int_0^t \omega(t') dt' = \int_0^t \omega_c(t_r(t')) \frac{dt_r(t')}{dt'} dt' = \int_{t_r(0)}^{t_r(t)} \omega_c(t'_r) dt'_r. \quad (5.8)$$

With $t_r(t) < t$ in general, the lower bound of this integral is $t_r(0) < 0$. However, negative retarded times indicate points in time before the electron appears, since the trajectory simulation uses strictly positive sample times in the laboratory frame. Therefore, CRESana calculates the instantaneous field phase at antenna j as

$$\varphi_{E_j}(t) = \int_0^{t_{r_j}(t)} \omega_c(t'_r) dt'_r + \varphi_{\text{cycl}} + \phi_A(t_{r_j}(t)). \quad (5.9)$$

The phase φ_{cycl} is the initial phase of the cyclotron motion and the additional phase shift ϕ_A adds the characteristic Archimedean spiral described in section 4.3, thus incorporating the relative phase between different antenna positions along the cyclotron orbit. The integration is done in the same way as the integrations for $E_{\text{kin}}(t)$ and $\phi(t)$ from eqs. (5.3) and (5.5). With the trapezoid approximation this only requires evaluations of the integrand $\omega_c(t')$ at all retarded times which has already been done earlier.

5.4.3 Antenna Response

The detector response function of one of the antennas converts the electric field of a radiation source into a voltage. This part of the simulation corresponds to the fourth panel in fig. 5.1. The result of this step is the voltage induced at an antenna, which is treated as an arbitrary modulated Cosine function

$$U_{\text{real}}(t) = A(t) \cos(\varphi(t)), \quad (5.10)$$

where $A(t)$ and $\varphi(t)$ are the instantaneous amplitude and phase of the modulated signal. With the antenna impedance Z the amplitude is a simple conversion of the instantaneous output antenna power via $A(t) = \sqrt{2P_{\text{out}}(t)Z}$. Therefore, the variables of interest are the instantaneous power and phase at the antenna output after applying the response function to the incident field's power P_E and phase φ_E from section 5.4.2. The response function depends on the frequency, the polarization, and the source direction of the radiation incident on the antenna, all of which can be treated separately with their respective effects on the frequency spectrum of the output voltage.

Polarization Mismatch

The antennas under consideration are only sensitive to linearly polarized radiation in a direction $\hat{\mathbf{p}}_a$, hence the instantaneous electric field $E(t)$ that drives the antenna is the component of the three-dimensional field vector $\mathbf{E}_r(t)$ that is parallel to the polarization direction with $E(t) = \hat{\mathbf{p}}_a \cdot \mathbf{E}_r(t)$. This polarization mismatch effect is implemented with a power loss factor M_{pol} for the antenna's output voltage spectrum. For a general radiation source with elliptic polarization the mismatch factor is $M_{pol} = A_x^2(\hat{\mathbf{x}} \cdot \hat{\mathbf{p}}_a)^2 + A_y^2(\hat{\mathbf{y}} \cdot \hat{\mathbf{p}}_a)^2$ where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are basis vectors aligned with the axes of the polarization ellipse and A_x and A_y are the amplitudes of the field in that respective direction. For an electron source radiating from a direction $\hat{\mathbf{d}}$ the polarization ellipses were derived in section 4.3. They are defined by $\hat{\mathbf{x}} = \hat{\mathbf{e}}_\phi$, with $A_x = 1$ and $\hat{\mathbf{y}} = \hat{\mathbf{e}}_\theta$ with $A_y = \cos(\theta)$, where the spherical coordinate system is defined such that θ is the angle enclosed by $\hat{\mathbf{e}}_r = -\hat{\mathbf{d}}$ and the magnetic field direction $\hat{\mathbf{B}}$. Therefore, knowing the direction of the radiation source, the polarization loss factor can easily be calculated.

Frequency Response

The frequency response is given by the antenna's transfer function (TF) defined as

$$H(\omega) = \frac{U(\omega)}{E(\omega)},$$

which relates the input electric field E to the output voltage U at frequency ω . This implies that the complex TF acts as a finite impulse response (FIR) filter on the input electric field signal $E(t)$, which is typically implemented by convolving the filter's impulse response with the input field.

For a concrete antenna design the Project 8 collaboration obtains the TF from an electromagnetic finite element simulation with the commercial package Ansys High Frequency Structure Simulator (HFSS) [91]. These simulations reveal that for antennas suited to this application, the bandwidth of $H(\omega)$ is much wider than the narrow bandwidth of interest for CRES in tritium beta spectroscopy and that it plateaus on these small scales for most frequencies.

This can be seen in the simulation results of Project 8's first manufactured prototype antenna in fig. 5.8. By placing the CRES spectrum (green area in the figure) in the plateau region, via the magnetic field, one can approximate $H(\omega) \approx H(\omega_0) = const$ where ω_0 is the central frequency of the CRES spectrum (orange dots in the figure).

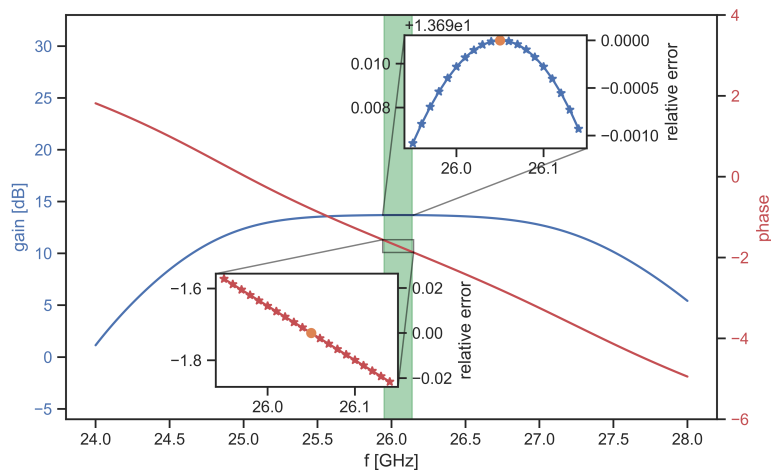


Fig. 5.8.: Simulated gain (blue) and phase (red) of the Project 8 antenna depending on frequency of the incoming radiation. Near the center of the antenna’s frequency band there is a plateau with almost constant gain. The green region marks a 200 MHz wide potential frequency band for CRES. The insets show zooms into that region each with a secondary y-axis showing the respective relative error if only the value at its center (orange) is used.

The inset plots show the relative error of this approximation where the phase errors are relative to 2π , thus this approximation introduces errors of $\lesssim 0.1\%$ for the gain and $\lesssim 2\%$ for the phase. For this reason, CRESana implements the frequency response as a constant gain $G_F = G(\omega_0)$ for the voltage power spectrum which is calculated from the HFSS simulated transfer function. The effect of the phase is ignored assuming that it is approximately constant.

With this simple approximation, CRESana can skip the computationally more expensive convolution of the electric field signal $E(t)$ and the FIR filter of the TF. This is a very powerful approximation, since the proper convolution requires the field $E(t)$ to be sampled at very high sampling rates (see appendix B for more details).

Directional Response

The antenna’s directivity $D(\hat{\mathbf{d}})$ describes its directional response in the form of a power damping factor depending on the direction $\hat{\mathbf{d}}$ of the radiation source in relation to the antenna normal vector. Figure 5.9 shows how $D(\hat{\mathbf{d}})$ changes the overall antenna gain of a single element of the Project 8 prototype antenna in its E and H-plane depending on the angle to the antenna normal at a frequency of 26 GHz.

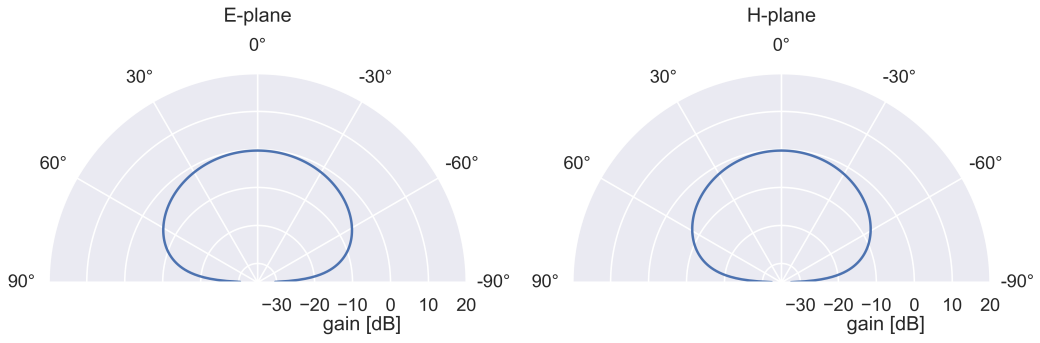


Fig. 5.9.: CRESana simulated gain patterns of a single element of the Project 8 antenna in the E and H planes at a frequency of 26 GHz. Distance of the line to the origin marks the gain D in the corresponding direction on a log-scale in dB. 0° marks the direction of the antenna's normal vector \hat{n} .

E and H-plane are terminology from antenna theory that describe the planes in which the electric and magnetic field of EM radiation emitted by the antenna propagate. By this definition, the normal \hat{n} and the polarization direction \hat{p}_a span the E-plane, whereas the normal and the cross-polarization direction $\hat{p}_c = \hat{n} \times \hat{p}_a$ span the H-plane. In a cylindrical antenna array for CRES the antennas are arranged such that their E-planes all are in the ρ - ϕ -plane and the H-planes all are in the ρ - z -plane in cylindrical coordinates. This is illustrated in fig. 5.10.

The directivity function depicted in fig. 5.9 is an analytic approximation that closely matches the simulated directivity of the prototype antenna in HFSS. Apart from this specific directivity function, the simulation in CRESana can also be configured with any arbitrary function $D: \mathbb{R}^3 \rightarrow [0, 1]$. For example, isotropic antennas are achieved with $D(\hat{d}) = 1$.

Element Combination

Finally, it is considered that antenna arrays may be build with passively combined sub-arrays as mentioned in section 3.1.1. For example, the Project 8 prototype antenna is a slot antenna [50], which acts as a linear array itself that combines five elements along the cylinder axis. With such sub-arrays, the analog voltages U_k from a subset of n elementary antennas are combined by summation with fixed phase relations φ_k , such that

$$U_{tot} = \sum_{k=0}^{n-1} U_k e^{i\varphi_k}. \quad (5.11)$$

In the case of the Project 8 antenna $n = 5$ and $\varphi_k = 0$ for all sub-elements.

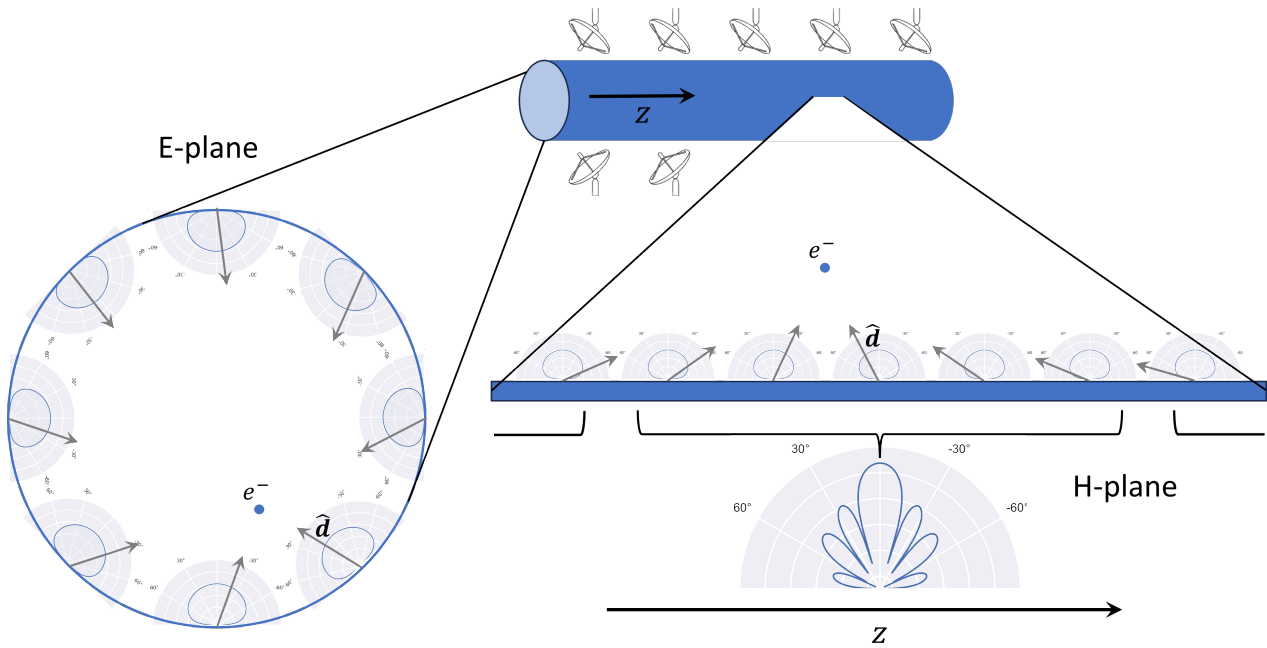


Fig. 5.10.: E- and H-plane of antennas in the cylindrical array. Circle on the left shows the cylinder cross section and the E-plane gain patterns. Right part shows a cylinder wall along the axis with the H-plane gain patterns. Along the axis, 5 antenna elements are combined to form a composite antenna. Bottom shows the H-plane pattern of the combined antenna. Instantaneous electron position e^- defines the direction vectors \hat{d} for the gain evaluation of each element.

Therefore, CRESana samples the voltages of each elementary antenna in a composite antenna with their respective directivity functions and then takes their sum with the respective phase relations. Figure 5.10 illustrates the placement of the elementary antennas and their E- and H-plane directivity gains in the full cylindrical array. Combining elements along the cylinder axis modifies the H-plane directivity of the resulting sub-array, which is depicted in the bottom right of fig. 5.10. The structure with multiple side-lobes is the result of interference of the sub-elements due to their displacement and the incoherent summation.

Figure 5.11 compares the measured gain pattern from simulations of the Project 8 prototype antenna in CRESana and the finite element simulation in HFSS. Using the analytic approximation for the directivity function of a single element, CRESana reproduces the HFSS results well except for angles greater than 70° in the E-plane. Since the antennas are arranged such that the E-plane is in the cross-sectional plane of the cylindrical array, these large angles correspond to directions that are nearly tangential to the cylinder. This results in a negligible fraction of the cylinder volume that is not accurately modeled (see fig. 5.10).

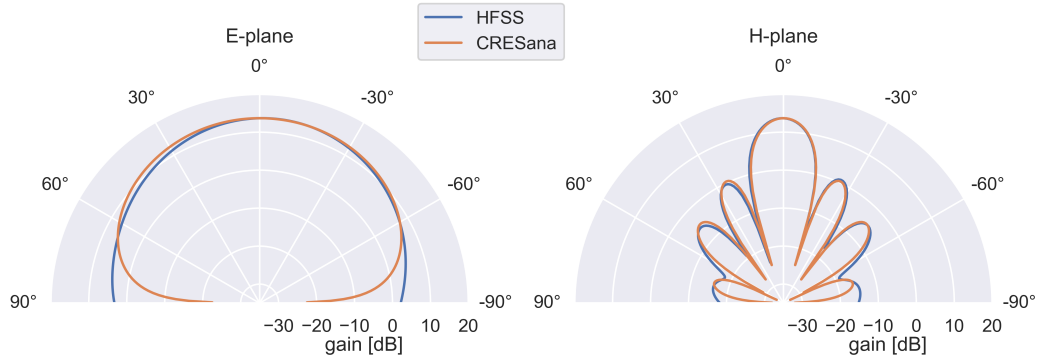


Fig. 5.11.: Simulated gain patterns of the Project 8 antenna in the E and H planes at a frequency of 26 GHz. Gain patterns from simulations of the antenna in HFSS are compared to the patterns produced by simulation in CRESana with the approach described in the text.

In summary the simulation of the antenna response calculates how the antenna polarization, directivity and the transfer function gain shape the instantaneous output power of a single antenna element:

$$P_{out}(t) = M_{pol}(t) \cdot D(\hat{\mathbf{d}}(t)) \cdot G(\omega_0) \cdot P_E(t), \quad (5.12)$$

depending on the input radiation power at the antenna P_E , the direction of the source $\hat{\mathbf{d}}$ and the polarization of the radiation. The directivity D and the polarization mismatch M_{pol} both need to be evaluated considering the electron's position at the retarded time. Because the frequency response is treated as constant, the antennas do not alter the frequency content and phases of the incident electric field $E(t)$. In the case of a composite antenna the resulting voltages of the individual elements in eq. (5.10) are then passively combined with eq. (5.11).

5.4.4 RF Receiver

The last component in the simulation, corresponding to the fifth panel in fig. 5.1, is the RF receiver, which converts the analog antenna voltages from eq. (5.10) to digital voltages. The former represent the real physical signals, while the latter are the measured data. Figure 5.12 shows the block diagram for a direct conversion receiver, a typical RF receiver for producing digital I/Q data. The incoming real-valued signal passes through an amplifier and is subsequently multiplied with a sinusoidal signal with frequency f_{LO} , which is produced by a local oscillator (LO) inside the receiver. In parallel a separate signal is generated by performing this frequency mixing process with a 90° phase-shifted version of the LO signal. The

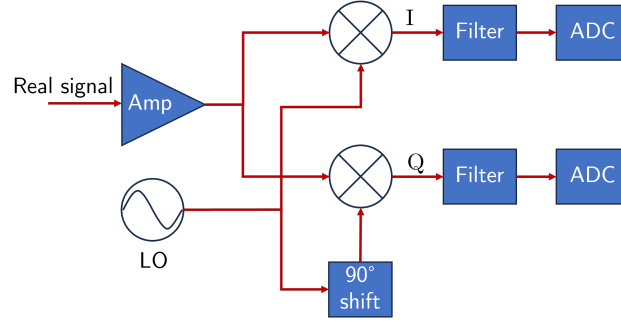


Fig. 5.12.: Diagram of a direct conversion receiver which produces I/Q data with a frequency mixer.

generation of this second identical but phase-shifted signal is the required step for sampling of the complex I/Q data.

Mathematically the operation is

$$U_{I/Q} = A_{LO} e^{i(\frac{\pi}{2} - 2\pi f_{LO} t)} U_{real}(t), \quad (5.13)$$

where A_{LO} is some arbitrary amplitude for the LO. Using the complex representation of the \cos function, eq. (5.13) can be written as:

$$U_{I/Q} = \frac{1}{2} A_{LO} A(t) \left(e^{i(\frac{\pi}{2} + \varphi(t) - 2\pi f_{LO} t)} + e^{i(\frac{\pi}{2} - (\varphi(t) + 2\pi f_{LO} t)} \right).$$

Thus, mixing with the LO produces a signal with two equal components with phases $\varphi(t) - 2\pi f_{LO} t$ and $-(\varphi(t) + 2\pi f_{LO} t)$.

By selecting f_{LO} close to the cyclotron frequency, the first component becomes a baseband signal while the second one shifts to a distant frequency band at $\approx -2f_{LO}$, which can easily be removed by the subsequent low-pass filter. This process greatly reduces the signal's bandwidth prior to digitization with the ADC in the last step. Conversely, the original signal of an electron in a 1 T magnetic field has a bandwidth of at least 26 GHz, which requires sampling rates $f_s > 52$ GHz according to the Nyquist–Shannon sampling theorem [92]. After down-conversion the sampling rate can be limited to the bandwidth needed to capture the full signal spectra at the relevant energies for tritium beta spectroscopy, which is expected to be < 200 MHz.

In the simulation, the whole RF receiver is implemented by assuming an idealized filter that removes the undesired frequency component entirely, avoiding aliasing, and thus, CRESana directly takes digital samples from

$$U_{I/Q}(t) = \frac{1}{2} A_{LO} A(t) e^{i(\frac{\pi}{2} + \varphi(t) - 2\pi f_{LO} t)}. \quad (5.14)$$

The instantaneous amplitude $A(t)$ and phase $\varphi(t)$ of the analog antenna voltage are the result of all the previous simulation steps from sections 5.4.1 to 5.4.3. However, their calculation is only triggered by the RF receiver, which requests an $A(t_j)$ and $\varphi(t_j)$ from each antenna at each sampling time step j . All intermediate results of the analog signals are only calculated at these requested time steps. Due to the limited bandwidth after down-conversion, eq. (5.14) can be sampled with a rate of just 200 MHz. This approach makes digitally sampling and processing of the intermediate analog signals at much higher sampling rates obsolete, and that is the main reason for the leap in performance in CRESana. This is only possible due to the use of the approximation of the fundamental frequency of the electric field, and the bypass of the full FIR filter in the antenna response.

As a final remark the noise of the RF receiver has to be addressed. The main source of noise for the recorded data is Nyquist-Johnson thermal noise. It can be approximated by a white Gaussian noise (WGN) distribution with variance

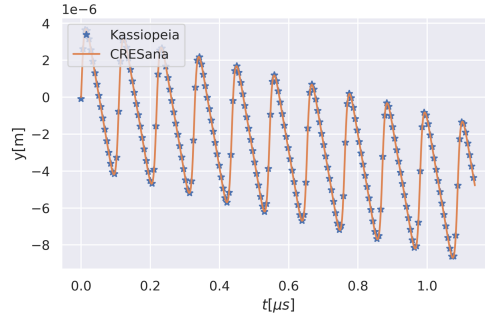
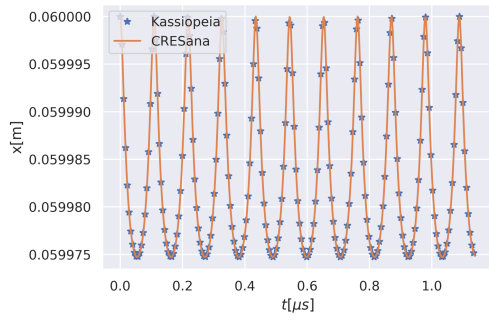
$$\sigma^2 = k_B T \Delta f Z, \quad (5.15)$$

where k_B is Boltzmann's constant, T is the system noise temperature, Δf is the bandwidth and Z is the system impedance [93, 94]. The system noise temperature is determined by the noise level of the first amplifier stage of the receiver.

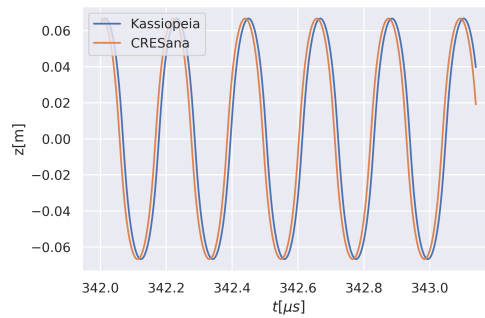
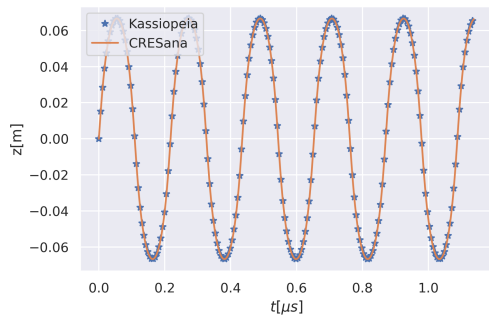
In a consequent mapping of reality this noise would be added prior to digitization onto the real signal that enters the receiver. However, sampling complex WGN is computationally much cheaper than the simulation of the deterministic electron track signal. Since this kind of noise is purely additive and the frequency mixer amplifies noise and signal likewise with the arbitrary A_{LO} , it is possible to add the noise only at the very end to the I/Q samples by setting $A_{LO} = 1$. In that way any simulated deterministic electron track can be combined with an arbitrary number of noise instances with minimum computational effort.

5.4.5 Validation

The results from CRESana were successfully validated in several ways. First of all, trajectory simulations were compared to Kassiopeia simulations of identical setups. One example of that is presented in fig. 5.13 which shows perfect agreement between both approaches in figs. 5.13a to 5.13c. Only fig. 5.13d shows that at a later time in the trajectory the axial motion is not perfectly aligned anymore. The reason for this is the electron's energy loss which slowly changes the axial frequency.



(a) Comparison of x coordinate of the trajectory. (b) Comparison of y coordinate of the trajectory.



(c) Comparison of z coordinate of the trajectory. (d) Comparison of z coordinate of the trajectory at a later time.

Fig. 5.13.: Comparison of trajectories of CRESana and Kassiopeia using the same magnetic field configuration and electron parameters for both. (a)–(c) show a perfect agreement of the trajectories, while (d) shows that for long simulation times the CRESana solution of z slowly goes out of phase which is due to the 1st order correction for the energy loss.

As mentioned in section 5.4.1 the initial integration in CRESana treats the energy as constant and the energy loss is only addressed with a subsequent linear correction. Kassiopeia on the other hand solves the equation of motion for each step with the instantaneous energy, which is more accurate. Nevertheless, this is an acceptable discrepancy considering the performance gained by the approximation.

In addition to the trajectory validation, Kassiopeia was used to validate the cyclotron frequency as can be seen in fig. 5.14. The plot also shows that the change of frequency over time is correct which means that the radiated power and the instantaneous energy in CRESana are in agreement with Kassiopeia.

Unfortunately, the validation of the antenna response simulation and the final signals are not as simple since there is no baseline simulation to compare to due to the unresolved issues in Locust (see appendix B). Without a baseline simulation, one can still find that the simulated spectra agree well with general expectations. For

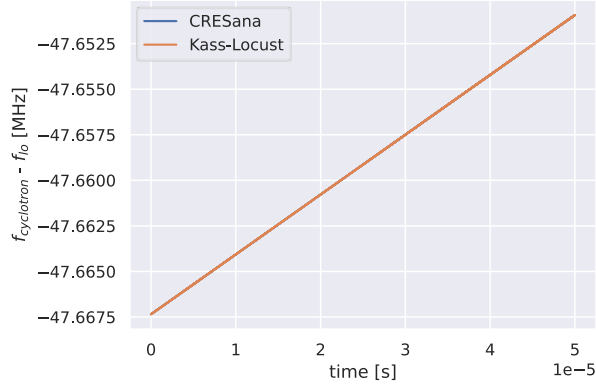


Fig. 5.14.: Comparison of the cyclotron frequency in Kassiopeia and CRESana. The blue line from CRESana is not visible as it perfectly aligns with the orange one from Kassiopeia.

example for cases with $\alpha_0 = 90^\circ$ the cyclotron frequency is located exactly where it is expected after down-conversion and also the power of the signal is consistent with the theory presented in chapter 4. The spectra of cases with $\alpha_0 < 90^\circ$ are more challenging to assess due to a complex set of features, although they generally align with qualitative expectations. A detailed discussion of the spectral features follows in section 5.5.

In addition to this qualitative assessment, the detected antenna power under different emission angles θ with different antenna polarizations was compared to theory which is presented in fig. 5.15. The angle θ is again the angle between observer direction and magnetic field. The red markers indicate the predicted power at the antenna using the electron's angular power distribution from the theory (see fig. 4.5a) and the free-space path loss. The, orange, blue and green data are the signal power measured from simulated data with a stationary electron in CRESana.

For each line the antenna is moved around the source at a fixed distance while always facing right to the source. The lines differ in the polarization direction used for the antenna, which are three right-handed base vectors \hat{x} , \hat{y} , \hat{z} such that \hat{z} is parallel to the magnetic field direction that fully decompose the radiation in any possible polarization. The total field power is the sum of the three orthogonally polarized field powers which is marked in purple and perfectly agrees with the prediction from the theory.

Furthermore, the three individual powers follow the expectation for the observation of elliptically polarized waves under the varying angle: at $\theta = 0$ an observer sees perfect circular polarization in the plane spanned by \hat{x} and \hat{y} , and consequently each of these two antennas detects half of the total power. At $\theta = 90^\circ$ an observers

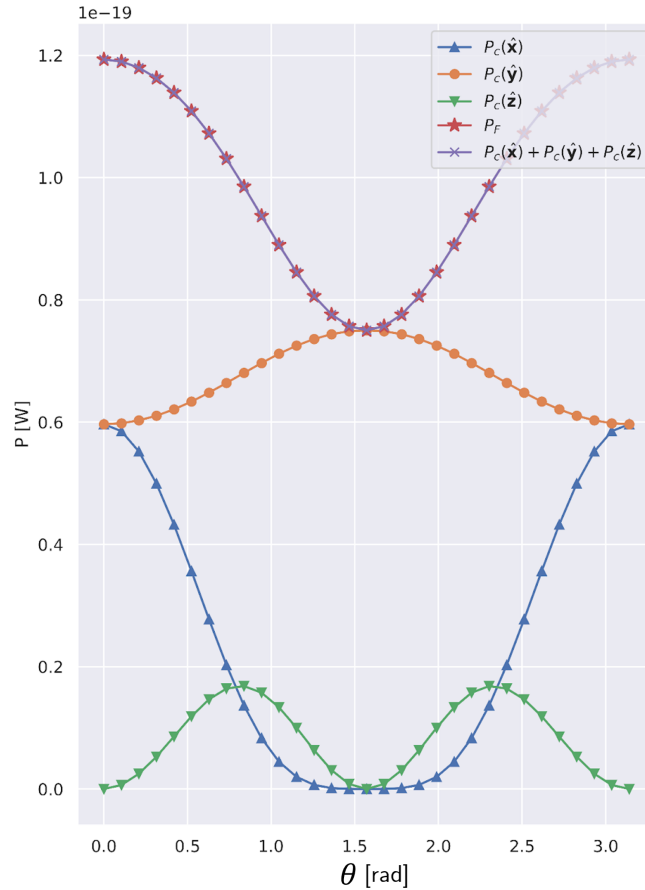


Fig. 5.15.: Simulated antenna power depending on the angle θ between observer and magnetic field. In the plot $P_c(\hat{x})$, $P_c(\hat{y})$ and $P_c(\hat{z})$ mark the powers of three antennas which are sensitive to linearly polarized waves in orthogonal directions. The sum of the three powers is compared to P_F , the electron's angular power distribution from theory.

sees perfect linear polarization which aligns with \hat{y} in the chosen coordinate system, and thus all the power is detected by the corresponding antenna and no power by the other two.

5.4.6 Performance Benchmarks

Finally, it is important to address the performance of CRESana, especially considering that it was primarily motivated by the long simulation times of Locust. As mentioned earlier, CRESana is ~ 500 times faster than a similar setup in Locust. Only this improvement makes the simulation of longer trajectories and larger antenna arrays feasible at all. Therefore, the goal of the benchmarks conducted here is to

demonstrate how the runtime and memory use of CRESana scale with the antenna array size and the number of samples. Extensive benchmarks of Locust for the sake of comparison were prohibitive due to the high requirement in central processing unit (CPU) hours.

For the benchmarks a fixed magnetic trap setup consisting of eight electromagnetic coils was used, which is the same that will be used for the results in chapter 8. The trap setup only influences the trajectory simulation and detailed benchmarks for different setups were not conducted. The trajectory simulation's runtime is independent of the antenna array size and its most expensive part is the integration, which is implemented such that it also does not scale with the simulation length. The integration only scales with the number of integration points and the complexity of magnetic field evaluations, i.e. the number of electromagnets in this case. The parameters for interpolation and number of evaluation points can be easily tuned without noticeable loss in simulation quality such that the constant cost of the integration is negligible compared to the sampling of voltages for long simulation times. This is demonstrated by the good agreement in fig. 5.13, which applied aggressive approximation settings yielding an integration time < 0.1 s.

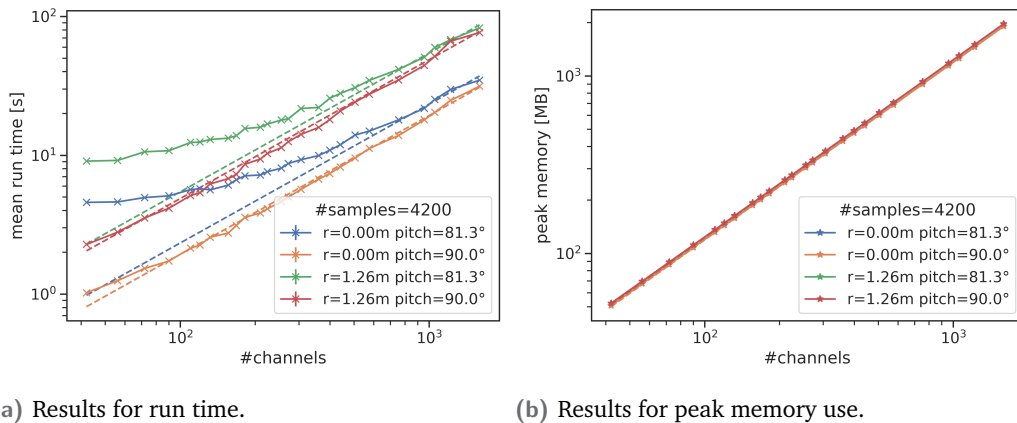
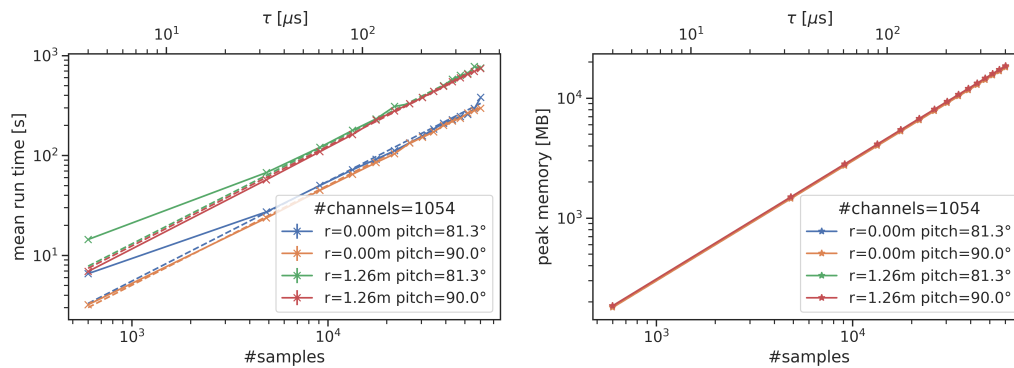


Fig. 5.16.: Benchmark results for peak memory and runtime depending on the number of antenna channels for different radii and pitch angles with a fixed trajectory length and number of samples.

Figure 5.16 displays the benchmark results depending on the number of antenna channels n_a for different electron parameters. In fig. 5.16a each plot point averages the total run time of 20 iterations. The two cases with pitch angle $\alpha_0 = 90^\circ$ scale linearly with the number of channels which would be expected here. The generally shorter run times for cases with $r = 0$ are caused by the calculation of the magnetic field, which is 2.4 times faster as a result of the equations greatly simplifying in that case.

For the cases of $\alpha_0 < 90^\circ$ the run time approaches linear scaling for high channel counts. Towards the lower end the run time approaches a plateau caused by the constant overhead of the integration in the trajectory simulation, which is skipped for the case of $\alpha_0 = 90^\circ$.

Figure 5.17a shows benchmark results as a function of the number of voltage samples n_s per channel, with a fixed number of channels. The general observations for all cases are similar to before: for high sample counts, the run times approach a linear scaling, as expected.



(a) Results for run time.

(b) Results for peak memory use.

Fig. 5.17.: Benchmark results for peak memory and runtime depending on the number of time samples (i.e. depending on trajectory length τ) for different radii and pitch angles with a fixed number of antenna channels.

Figures 5.16b and 5.17b show the benchmark results for the peak memory usage depending on number of channels and number of samples respectively. In both cases the memory usage scales perfectly linearly, which is expected considering that the whole simulation is implemented such that the output is generated as a single NumPy array by performing algebraic calculations with a series of intermediate arrays, which mostly have at least the same sizes and dimensions as the output array.

However, it is noticeable that the peak memory is about a factor 15 higher than the memory required for the output array, which is $n_a \times n_s \times 128$ bit for double precision complex numbers. This indicates that the produced intermediate arrays unnecessarily take up significant amounts of memory. Unfortunately, this was only recognized as an issue at a late stage of development, when long trajectories were simulated, whereas earlier typical simulation setups had negligible memory footprints. Yet, in the energy resolution studies in chapter 7 with the combination of required antenna array size and trajectory length, peak memory demands higher

than 20–60 GB were observed while only using a single CPU core. This memory consumption turned out non-ideal and even partly limiting for the work with the available compute cluster, which only has 5 GB of random access memory (RAM) per core.

5.4.7 Future Developments

To conclude the discussion of CRESana, this section briefly outlines potential future developments. First and foremost the memory issue needs to be resolved, as it currently limits the ability to simulate long tracks in large antenna arrays (e.g. ~ 1 ms with ~ 1000 channels). An apparent solution would be to split the calculation into smaller chunks. Instead of calculating the whole output array with NumPy in a single pass, one could iteratively calculate sub-arrays with a fixed number of samples. This comes at a small performance cost as it requires the undesired use of Python for-loops. Still, by choosing the sub-array size large enough, the actual impact should be small.

One complication for this approach is the use of integrals in several places of the voltage sampling. That means the correct calculation of sub-arrays depends on earlier results. This is particularly inconvenient for eq. (5.9), where the integration samples and bounds are different for each antenna and thus simple splits by sample time do not work. While not impossible, the required fix is not as straightforward to implement as splitting the calculation without dependencies and therefore it was not prioritized anymore for this thesis.

Additionally, despite the substantial speedup achieved so far, further improvements in CRESana's computational performance are desirable for its application in optimization problems with antenna arrays using thousands of channels. In sections 7.2.2 and 8.1.1 it will become evident that these problems are still limited by the simulation time and require certain computational shortcuts. Since CRESana's performance gains are primarily due to advanced physical modeling, its purely serial implementation in Python is somewhat prototypical and could be further optimized. A parallel implementation on either CPUs or graphics processing units (GPUs) of the most computationally intensive components is advisable, given that the simulation of many independent channels and time samples is highly parallel. Moreover, these components could benefit from optimized reimplementations in a compiled language like C++, with bindings to Python.

In addition to these technical improvements, the introduction of new features could be considered as well. As mentioned in section 3.1.1, resonant cavities are proposed

as an alternative detector concept for large CRES experiments. This naturally leads to the question of whether CRESana could be extended to support this concept as well. Unfortunately, many existing aspects of CRESana’s modeling approach cannot be applied to such a detector. The main issue lies in the solutions of the electron’s radiation fields discussed in section 4.3 and their sampling at the retarded time. These are only valid for free-space radiation and require alternative approaches when the electron couples to a cavity. Moreover, CRESana’s substantial gains in computational performance rely heavily on the assumption of an approximately plateauing, wide-bandwidth frequency response of the considered antennas. This allows it to skip the convolutions and bypass the numeric high-frequency sampling of the field solutions. With a resonant cavity this approximation is no longer valid, as the frequency response is a narrow resonance peak [47].

The only aspect of CRESana that can be straightforwardly reused for a cavity detector is its electron motion simulation, which serves as a replacement for the slower particle tracking in Kassiopeia. Nevertheless, there has been some concern within the Project 8 collaboration that the cavity’s reaction to the electron radiation might influence the electron’s trajectory. If this proves to be the case, the effect cannot be easily incorporated into CRESana’s periodicity-exploiting trajectory solution. Instead, it would likely require modifications to the differential equations and an iterative integration approach, similar to that used in Kassiopeia.

Although there is no inherent major obstacle for implementing a cavity model from scratch into CRESana, it should be recalled that the main motivation for developing an alternative simulation tool were the long computation times in Locust. These are primarily driven by the high number of channels of an antenna array. A cavity experiment on the other hand, necessitates only a small number of channels, with the first prototype design using just a single one [95]. Thus, the simulation run times are expected to be more manageable for this detector concept. For this reason, there are no immediate plans for integrating the necessary functionality to model a cavity-based experiment into CRESana.

5.5 Spectral Features of Simulated Signals

In this section simulation results from CRESana are presented for the purpose of discussing the generic spectral features of CRES signals. The specific manifestation of spectral features is influenced by the detector setup consisting of magnetic trap and

antenna array and for any given setup the spectra depend strongly on the electron's pitch angle α_0 .

The spectral features will be discussed in power spectra of the simulated time series. The power spectrum of a time series \mathbf{x} with n_s samples is defined as

$$\mathcal{P}(\mathbf{x})_m = \frac{|\mathcal{F}(\mathbf{x})_m|^2}{n_s Z}, \quad (5.16)$$

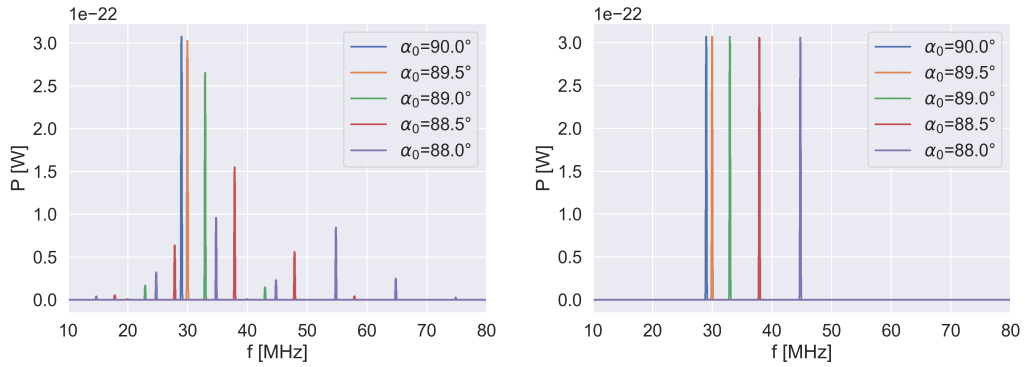
where Z denotes the impedance (typically $Z = 50 \Omega$ is used in signal processing) and $\mathcal{F}(\mathbf{x})_m$ represents the m -th frequency component of the discrete Fourier transform (DFT) of \mathbf{x} , defined as

$$\mathcal{F}(\mathbf{x})_m = \frac{1}{\sqrt{n_s}} \sum_{j=0}^{n_s-1} \mathbf{x}_j \cdot e^{-2\pi i \frac{jm}{n_s}}. \quad (5.17)$$

Using the normalization $\frac{1}{\sqrt{n_s}}$, eq. (5.17) defines a symmetric DFT.

5.5.1 Sidebands

Pitch Angle Dependence



(a) Spectra with all modulation effects.

(b) Spectra without modulation using the average cyclotron frequency.

Fig. 5.18.: Spectra of electron signal simulations in CRESana with $E_{\text{kin}} = 18.6 \text{ keV}$ in a harmonic trap taken with the Project 8 prototype antenna. A single antenna is used at a distance of 1 m. A *flattop* window function is applied for optimal amplitude accuracy.

Figure 5.18a shows spectra taken with the Project 8 prototype antenna in a trap with a harmonic profile along the rotation axis and the variation with the pitch angle.

Electrons with pitch angle $\alpha_0 = 90^\circ$ generate spectra with a single peak that corresponds to the downshifted cyclotron frequency. At different antenna positions (not depicted) this signal appears with different time delays, phase shifts and amplitudes according to the radiation's travel distance. The time delays and corresponding phase shifts are implemented in eq. (5.8) due to the upper integration limit which determines how much the phase has advanced and which changes with the travel time to the antenna. The amplitude change at different antenna positions is the result of the free-space path loss in eq. (5.12).

For electrons with $\alpha_0 < 90^\circ$ the spectra exhibit additional side peaks, which are characteristic features of frequency or amplitude modulated carrier waves that are common in radio communications [96]. These modulation effects arise from variations of the received cyclotron radiation caused by the periodic axial motion.

Modulation Sources

One source of frequency modulation (FM) are the local magnetic field changes as the electron moves through the axial field profile. Such changes of the local field change the instantaneous cyclotron frequency in eq. (5.8). This effect shifts the observed main peak position, placing it at the *average* cyclotron frequency. The average frequency

$$\bar{\omega}_c = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \omega_c(t) dt = \frac{1}{\Delta t} (\varphi_c(t_1) - \varphi_c(t_0)) ,$$

is defined by the difference in cyclotron phase at two reference points in the trajectory time. For this reason the average over an axial period is

$$\bar{\omega}_c = \frac{2\pi}{T_{ax}} \Delta\varphi_c = \omega_{ax} \Delta\varphi_c ,$$

where $\Delta\varphi_c$ is the difference of the cyclotron phase over one axial period. Since the cyclotron phase still evolves when the electron is out of view, this phase difference is detected, causing the observed average to change, even if the electron moves outside the sensitive detector volume during axial motion. The effect of the shifted main peak is emphasized in fig. 5.18b where all modulation effects were switched off in the simulation and the average cyclotron frequency was used.

Another source of FM are the Doppler-shifted frequencies due to the relative velocity to the antennas, which changes continuously when the electron moves through the trap. This shifts the perceived frequency periodically up and down.

Amplitude modulation (AM) is caused by two things: the simple change of the distance $|d|$ in eq. (5.7) and the change of the source direction relative to the antenna, which changes G_e and all the contributors to the antenna response in eq. (5.12). As a minor effect the Larmor power P_{la} also experiences small changes due the change in local magnetic field when the electron moves up the potential wall of the trap (see eq. (4.24)).

Theoretical Description of AM and FM

The spectra produced by AM and FM are well established in RF theory [96]. In the typical radio applications, the modulations are applied to a carrier wave $c(t) = A_{ca} \cos(\omega_{ca}t)$, where A_{ca} and ω_{ca} are the carrier amplitude and frequency. The simplest form of AM modulates the carrier's amplitude with a cosine $m(t) = 1 + m_A \cos(\omega_m t)$, where ω_m is the modulation frequency and m_A is the modulation index. It is defined via the modulation amplitude A_m as:

$$m_A = \frac{A_m}{A_{ca}}.$$

The spectrum of the amplitude modulated signal, given by

$$s(t) = m(t)A_{ca} \cos(\omega_{ca}t), \quad (5.18)$$

consists of the carrier frequency ω_{ca} , and two additional peaks at $\omega_{\pm} = \omega_c \pm \omega_m$, which are called sidebands. The amplitude of these sidebands is a fraction of the carrier amplitude that is determined by the modulation index as $A_{\pm} = \frac{m_A}{2} A_{ca}$. Figure 5.19a illustrates this expected behavior, showing spectra for eq. (5.18) with varying m_A .

For an FM signal, the spectrum is more complex. Again starting with the simplest form of modulating the carrier frequency with a cosine function such that $m(t) = 1 + \omega_{\Delta} \cos(\omega_m t)$, where ω_{Δ} is the maximum shift of the instantaneous frequency from the carrier frequency, the modulated carrier is

$$s(t) = A_c \cos\left(\int_0^t \omega_{ca} m(t') dt'\right) = A_c \cos\left(\omega_{ca}t + \frac{\omega_{\Delta}}{\omega_m} \sin(\omega_m t)\right). \quad (5.19)$$

The corresponding spectrum consists of the carrier frequency ω_{ca} and a multitude of sidebands at frequencies $\omega_n = \omega_{ca} + n\omega_m$ with $n \in \mathbb{Z}$.

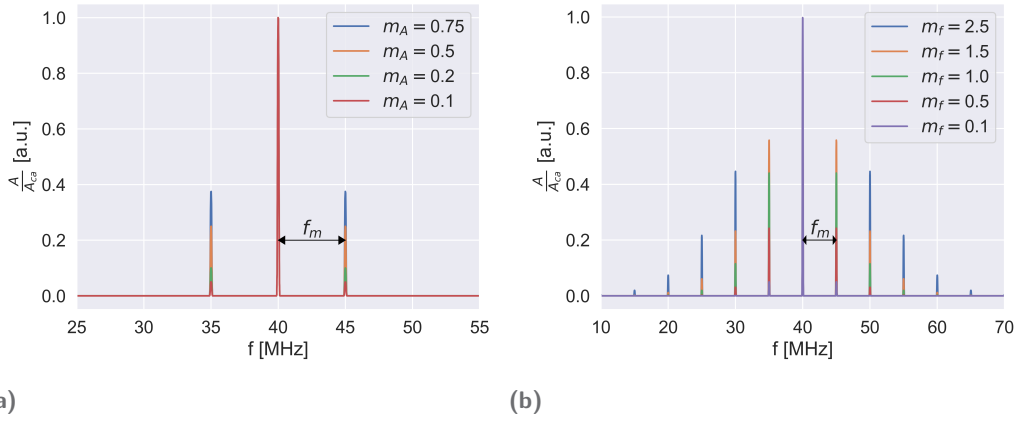


Fig. 5.19.: Spectra of AM (a) and FM (b) signals with varying modulation indices. Signals use a carrier frequency of $f_{ca} = 40$ MHz and modulation frequency of $f_m = 5$ MHz. Amplitudes are relative to the carrier amplitude A_{ca} .

The amplitudes of the sidebands can be calculated with Bessel functions of the first kind as $A_n = J_{|n|}(m_f)$. They depend on the modulation index, defined as

$$m_f = \frac{\omega \Delta}{\omega_m}. \quad (5.20)$$

Higher modulation indices increase the relative amplitudes in the higher order sidebands. For values of m_f up to 0.5, the sidebands have negligible power compared to the carrier. At $m_f = 1.5$, the sidebands begin to contain more power than the carrier, and at $m_f \approx 2.41$, all the power is in the sidebands, causing the carrier to disappear entirely. This behavior is demonstrated in fig. 5.19b, which shows spectra for eq. (5.19) with varying m_f . However, the disappearance of the carrier is not visible in this plot because the high-amplitude carrier in the $m_f = 0.1$ case conceals the other cases, but it is shown in fig. 5.20.

For CRES signals the carrier frequency corresponds to the average cyclotron frequency $\bar{\omega}_c$, and the modulation frequency is typically twice the axial frequency ω_{ax} , since the electron passes the antenna twice in a single axial period³. Therefore, the sidebands appear at frequencies

$$\omega_n = \bar{\omega}_c + 2n\omega_{ax}, \quad (5.21)$$

with $n \in \mathbb{Z}$. The two AM sidebands coincide with the FM sidebands at $n = 1$.

³This is not the case when the electron does not reach the antenna's z position at its maximal z extent.

For general periodic axial motion, the axial frequency ω_{ax} only describes the fundamental frequency of the motion. In addition, the motion can be decomposed in harmonics at integer multiples of the fundamental frequency $\omega_{axi} = i\omega_{ax}$. This complicates the spectrum of the modulated signal, as each modulating frequency component $\omega_{mi} = 2\omega_{axi}$ adds its own set of AM and FM sidebands. Fortunately, since each additional frequency is a multiple of $2\omega_{ax}$, they only add sidebands that overlap with the possible sideband positions of the fundamental frequency. Therefore, no matter how complicated the spectrum of the axial motion, all possible sideband positions are given by eq. (5.21).

Nevertheless, providing a complete theoretical description of the spectra, including the expected sideband amplitudes, is not straightforward. The amplitudes depend on the AM and FM modulation indices of each individual modulating frequency component ω_{axi} . Thus, despite a comprehensive theoretical understanding of the modulated signals, full simulations of the spectra are mandatory.

Observed simulation results (e.g. figs. 5.18 and 5.20) generally indicate that smaller pitch angles lead to higher modulation indices, thereby shifting the signal power to the higher-order sidebands. In most cases, smaller pitch angles also increase the axial frequency, which widens the distance between sidebands. However, the extent of this effect depends on the functional form of the magnetic field $B(z)$. With typical axial frequencies in the 10s of MHz, the signal spread across many sidebands for small pitch angles is the primary factor determining the bandwidth requirement of recording the CRES spectra.

Comparison of Modulation Effects in Different Setups

Figure 5.20 shows a comparison of spectra for one particular electron under different simulation conditions. The first column shows the spectra with all modulation effects disabled, the second enables only AM, the third enables only FM due to the local magnetic field changes, the fourth enables only FM due to the Doppler shifts, and the last enables all modulation effects. The first row shows the spectra for a harmonic trap (see appendix A.1), the second one for a square well trap (see appendix A.2).

Although both traps have similar magnetic field strengths and axial frequencies for this electron, their spectra differ due to the functional form of the magnetic trapping field. In the harmonic trap, local field variations shift the main peak to the average cyclotron frequency, as discussed before. In contrast, in the square well trap, the cyclotron frequency remains constant throughout the axial motion, due to the uniform background field and thus the main frequency is not shifted. This

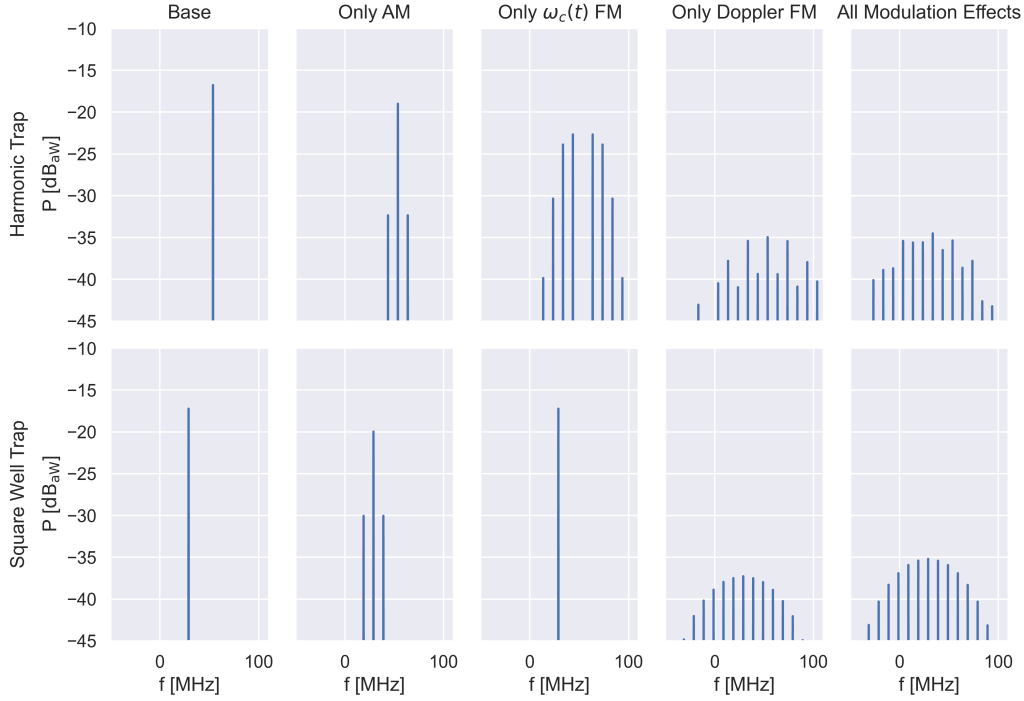


Fig. 5.20.: Spectra for $\alpha_0 = 87.5^\circ$ and $E_{\text{kin}} = 18.6 \text{ keV}$ in harmonic and square well trap taken with the Project 8 antenna at distance of 0.1 m. Powers in dB_{aW} , i.e. log-scale relative to $1 \text{ aW} = 10^{-18} \text{ W}$. In each column different modulation effects are enabled. *Flattop* window function is applied for optimal amplitude accuracy.

difference is evident in the first column. When AM is introduced, only two significant sidebands appear, as expected. This effect is more pronounced in the square well trap, which can be attributed to the electron's axial travel distance being about 70 % greater in that trap. The third column indicates a high FM modulation index with a disappearing carrier in the harmonic trap, whereas no sidebands are added in the square well trap, as the cyclotron frequency stays constant. Finally, in both traps, Doppler induced FM plays the largest role in shaping the final spectrum as seen in the fourth and fifth columns.

In fact, simulation results show that FM is the dominant effect for most setups, which can be explained by analyzing the modulation indices. For non-relativistic radial velocities v_r between electron and antenna, the observed Doppler shift is approximately $\Delta\omega_D = \frac{v_r}{c} \bar{\omega}_c$. Assuming that ω_Δ in the FM modulation index in eq. (5.20) is the maximum $\Delta\omega_D$ along the axial trajectory⁴ yields

$$m_f = \frac{\max(v_r)}{c} \frac{\bar{\omega}_c}{2\omega_{ax}}.$$

⁴This treatment is a simplification that is only exact for harmonic modulation.

Using trigonometric relations, the radial velocity can be expressed as

$$v_r = v_0 \cos(\alpha) \sin(\gamma),$$

where v_0 is the electron's velocity, α its instantaneous pitch angle, and γ is the angle between the antenna's line of sight to the electron and its normal vector, given by $-\hat{e}_\rho$ in cylindric coordinates. Noticeable power in FM sidebands requires that $m_f > 0.5$, which then results in

$$\gamma_D > \arcsin\left(\frac{1}{\beta \cos(\alpha_D)} \frac{\omega_{ax}}{\bar{\omega}_c}\right),$$

where γ_D and α_D are the two angles at $\max(v_r)$ and $\beta = \frac{v_0}{c}$.

Therefore, with axial frequencies of ≈ 10 MHz and a cyclotron frequency of ≈ 26 GHz, electrons at the tritium endpoint, with $\beta = 0.26$ that have their maximum Doppler shift with pitch angles of $\alpha_D = 88\text{--}89^\circ$, require $\gamma_D = 2.4\text{--}4.7^\circ$ to achieve $m_f = 0.5$. This corresponds to minimum axial travel distances of 4–8 % of the antenna array radius, which can easily be achieved in most antenna arrangements. However, this is a crude estimate that neglects the fact that the axial turning point z_{max} is only reached when $\alpha = 90^\circ$. This must necessarily occur beyond the position of maximum Doppler shift z_D since $\alpha_D < 90^\circ$ is required for observing Doppler shifts. A more rigorous determination of γ at z_{max} requires a specific $B(z)$, as the exact position of z_D relative to z_{max} depends on the form of $B(z)$. Nevertheless, this result demonstrates that noticeable FM due to the Doppler effect alone is generally possible even with short axial travel distances.

Conversely, in many cases, the amplitudes will not vary significantly over the same short axial scales. First, changes in the free space path loss are negligible for $\gamma = 2.4\text{--}4.7^\circ$, as it scales with $\sim 1/|d|^2$ (see eq. (5.7)), and $|d|$ can be considered constant for small angles γ . The antenna's directivity also does not vary strongly unless it has an extremely narrow beamwidth. For example, with the prototype antenna, these angles correspond to an amplitude variation of 3.5–13.5 %.

Achieving relative power in the first sideband with AM similar to that of FM at $m_f = 0.5$ requires a modulation index $m_A = 0.51$. This implies that the amplitude must vary by around 51 % from its average value. Momentarily achieving this variation due to the directivity of the prototype antenna requires the electron to reach at least $\gamma = 8.2^\circ$. However, achieving this variation from the average amplitude requires the electron to travel even further. As before, a more rigorous determination require a specific $B(z)$, as the trajectory influences the average amplitude. The

contribution of the free space path loss to AM can be neglected in this treatment, as it would require the electron to reach at least $\gamma = 48.5^\circ$.

In conclusion, this analysis suggests that significant AM requires considerably longer axial travel distances than FM. Given that the Doppler shift also increases with longer axial travel distances while ω_m decreases, it is likely that FM remains the dominant contribution at longer distances in many setups. Nonetheless, this relationship depends on the antenna's beamwidth and it could shift with higher-directivity antennas.

5.5.2 Constant Doppler Shifts

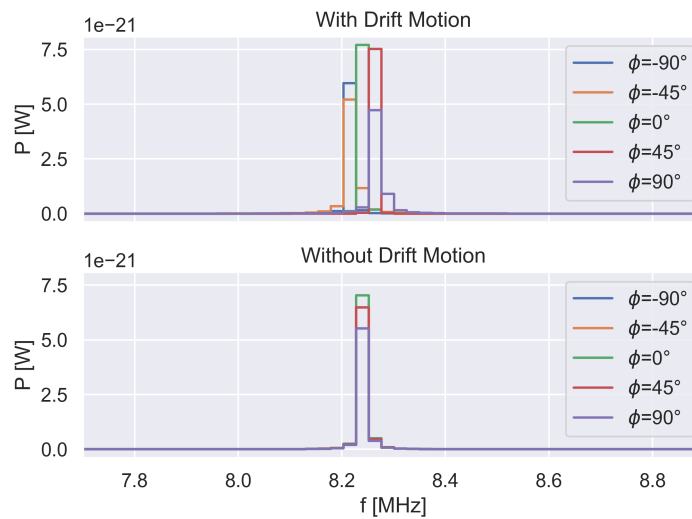


Fig. 5.21.: Spectra of an electron with $\alpha_0 = 90^\circ$ and $\rho > 0$ experiencing drift motion compared to spectra with drift motion disabled. The drift motion doppler-shifts the cyclotron frequency for different azimuthal antenna positions ϕ .

For electrons at radial positions $\rho > 0$ the drift motions add some minor additional features. While in theory the drift motions are another periodic motion in a different dimension they do not add visible AM and FM sidebands because the frequency of the drift motion is very small compared to the axial motion. However, a noticeable effect is a small variation in the cyclotron frequency for different azimuthal antenna positions ϕ (see fig. 5.21). This is caused by the different relative velocities between electron and antenna for different azimuthal positions which causes a change in the doppler-shifted frequency for different positions. While this effect might seem insignificant, if left unaccounted the incoherence of antennas causes a noticeable reduction in SNR of simple trigger algorithms (see section 6.2.3).

5.5.3 Linear Frequency Chirp

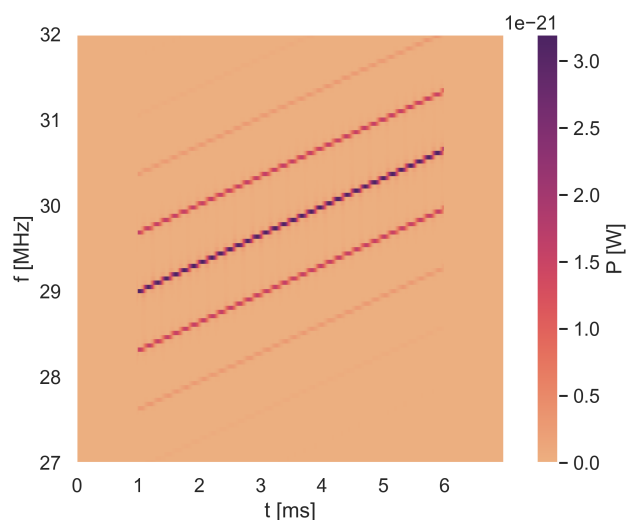


Fig. 5.22.: Time frequency spectrogram of a 5 ms long single antenna signal. Multiple sidebands are visible that shift in frequency over time due to the energy loss. A setup with small axial frequency was chosen such that all lines are well visible in a close-up.

The last feature of the signal spectra is caused by the energy loss of the electron due to the radiated power. For sufficiently long observation times the energy loss manifests itself as a linear frequency chirp, i.e. over time all frequencies shift to higher values in a linear fashion (see fig. 5.22). The frequency change rate (slope) will be mostly determined by the chosen magnetic field strength which influences the radiated power with some variations depending on pitch angle and electron energy (see eq. (4.24)).

Detection of CRES Signals in Noise

” *You can have data without information, but you cannot have information without data.*

— **Daniel Keys Moran**
Science fiction author

One of the greatest challenges of CRES with an antenna array is the detection of the signals described in section 5.5 under real measurement conditions. The power of the electron radiation is determined by eq. (4.24) and it depends on the magnetic field strength B . The Phase I and Phase II experiments of Project 8 (see section 3.2) operated with $B \approx 1$ T, which results in a radiated power on the order of 10^{-15} W. In future phases using atomic tritium an adequate atom trapping efficiency requires fields that are smaller than 0.1 T [30] with estimated electron powers on the order of 10^{-18} W. In both cases the signals are obviously very faint, making it challenging to distinguish them from background noise.

As mentioned in section 5.4.4, background noise is added to the signal in the form of thermal fluctuations in the receiver’s circuit. Thermal noise is approximately white Gaussian noise (WGN) [93, 94] and its power $P = k_B T \Delta f$ is determined by the bandwidth Δf and the system noise temperature T , which combines the noise levels of the antenna and all noise sources of the receiver up to the first amplification stage. With a bandwidth of 200 MHz, even lowering the system temperature to $T = 5$ K still yields a noise floor of $P = 1.38 \times 10^{-14}$ W, which is considerably higher than the signal power.

While this is a challenging scenario it is certainly not impossible to detect signals under these conditions. Similar conditions can be found for instance in the Global Positioning System (GPS). The signals transmitted by GPS satellites reach receivers on earth with similarly low powers of $\sim 10^{-18}$ W [97] and even use similar frequencies with 1.5 GHz [98]. Additionally, Phase I and II of Project 8 demonstrated that it is generally possible to detect the electron radiation despite the low levels of power [60, 65, 66].

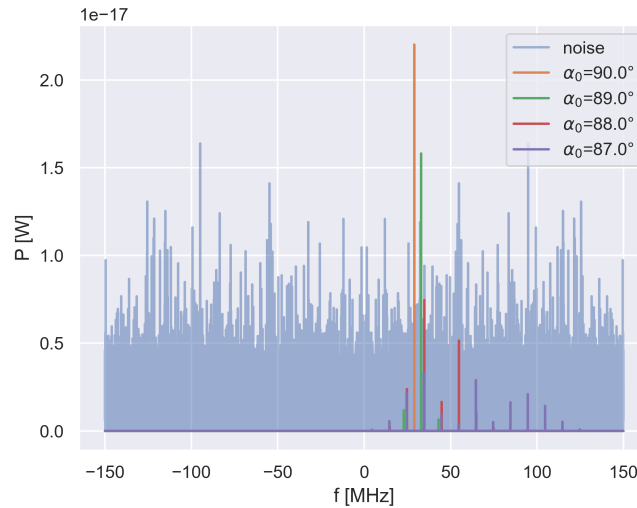


Fig. 6.1.: Power spectra obtained with a $42\ \mu\text{s}$ long Fourier Transform. WGN with $T = 5\ \text{K}$ is compared to signals of electrons with different pitch angles α_0 in a harmonic trap measured by a single antenna. For visualization, the signals are amplified to an SNR that is similar to Phase II. Actual signals from a single antenna are too weak for detection.

However, the main difference between the detection with an antenna array and the earlier phases of Project 8 is the signal power per receiver channel. In Phase II the electron's radiation was coupled to the transmitting mode of a waveguide and thus a high fraction of its radiated power reached the receiver. In the case of the antenna array the same emitted power is spread out to tens or potentially hundreds of antenna channels. Furthermore, the older experiments were only sensitive to electron pitch angles $> 89.3^\circ$ [66], which produce mostly simple spectra with a main peak and low-amplitude side-peaks.

In fig. 6.1, electron spectra from a single antenna are presented, where the signal powers have been artificially amplified to achieve an SNR that is comparable to that of Phase II in the $\alpha_0 = 90^\circ$ case. With a sufficiently long time window in the Fourier transform, a detection algorithm that searches for high-powered peaks in the spectrum works well for the cases with $\alpha_0 > 89^\circ$ [65]. Yet, with an increased pitch angle acceptance required in future phases, the spectra become more complex with many smaller peaks, rendering this approach ineffective. Therefore, the new challenges for signal detection going forward are the combination of weak signals from many channels and the identification of complicated spectra of electrons with $\alpha_0 < 89^\circ$ that have no single peak above the noise floor.

6.1 Basics of Signal Detection

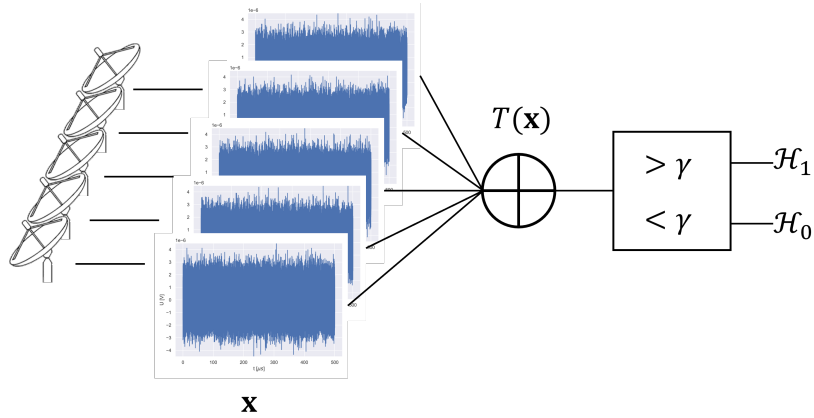


Fig. 6.2.: Noisy voltage time series generated by antennas are grouped into a data matrix \mathbf{x} . Signal detection is tasked with the decision between a noise and a signal hypothesis for the data \mathbf{x} .

The signal detection procedure in an antenna array is confronted with the situation depicted in fig. 6.2. A number of n_a antennas take data in the form of time-varying complex voltages. These time series are inherently noisy and may or may not contain a signal. From each antenna channel a fixed number of time samples n_s is grouped together into a complex $n_a \times n_s$ data matrix \mathbf{x} . In mathematical form, the general detection problem then is a statistical hypothesis test between the two hypotheses

$$\begin{aligned}\mathcal{H}_0: \mathbf{x} &= \mathbf{n} \\ \mathcal{H}_1: \mathbf{x} &= \mathbf{s} + \mathbf{n}.\end{aligned}$$

In the null hypothesis \mathcal{H}_0 the data matrix \mathbf{x} is purely noise which is denoted by the matrix \mathbf{n} . The alternative hypothesis \mathcal{H}_1 is the case where both a signal \mathbf{s} and noise are present in the data matrix. For the decision, the detector calculates a test statistic $T(\mathbf{x})$, which can be an arbitrary function of the data returning a scalar value. The signal detector then decides \mathcal{H}_1 if $T(\mathbf{x}) > \gamma$ for an arbitrary threshold γ [99].

The choice of γ is, in principle, arbitrary, but it affects the two important performance parameters of the detector and will be selected to optimize them. $T(\mathbf{x})$ is a continuous random variable due to the random noise in the data \mathbf{x} . Figure 6.3 shows exemplary probability density functions (PDFs) for this random variable under the two hypotheses, denoted as $p_{T(\mathbf{x})}(t; \mathcal{H}_0)$ and $p_{T(\mathbf{x})}(t; \mathcal{H}_1)$. The detection probability

$$P_D = P(\mathcal{H}_1; \mathcal{H}_1) = \int_{\gamma}^{\infty} p_{T(\mathbf{x})}(t; \mathcal{H}_1) dt$$

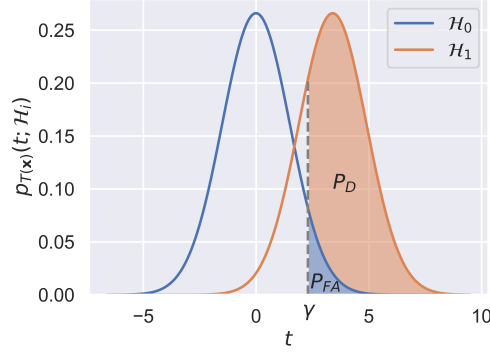


Fig. 6.3.: Exemplary probability density functions of a test statistic $T(\mathbf{x})$ under both hypotheses. The decision threshold γ defines the detection probability P_D and the false alarm probability P_{FA} .

is the probability that the detector correctly decides \mathcal{H}_1 and the probability of a false alarm

$$P_{FA} = P(\mathcal{H}_1; \mathcal{H}_0) = \int_{\gamma}^{\infty} p_{T(\mathbf{x})}(t; \mathcal{H}_0) dt$$

is the probability that the detector incorrectly decides \mathcal{H}_1 .

By varying the threshold γ one can choose a trade-off between a high detection probability and a low false alarm probability. In most cases one would have a strict requirement for an upper limit of $P_{FA} < a_0$ to limit the background rate. Since P_{FA} and P_D are the integrals of the PDFs for $T(\mathbf{x}) > \gamma$ it is possible to directly calculate the required γ that yields $P_{FA} = a_0$ and the P_D that results from this threshold (see fig. 6.3). Only in cases of complicated test statistics where the PDFs might not be known analytically or are difficult to model is it necessary to experimentally find the threshold that yields the desired false alarm probability. However, doing so while requiring that $P_{FA} \ll 0.01$ can easily result in a significant computational investment for the accumulation of sufficient statistical data. This would be the case for a neutrino mass measurement with CRES which demands a high background rejection with background rates below $4.7 \times 10^{-11} \text{ s}^{-1} \text{ eV}^{-1}$ [66]¹.

With the above considerations it is clear how to approach the choice of γ for a given test statistic $T(\mathbf{x})$, yet the choice of a good test statistic itself remains an open question and depends on the expected signals. The ideal test statistic is one that gives a good separation of the two PDFs which results in maximum P_D for a fixed P_{FA} . For a neutrino mass measurement the choice is critical for the experimental sensitivity, which directly depends on the P_D that is attainable for a fixed P_{FA} . In this

¹This number means that $P_{FA} < 4.7 \times 10^{-11}$ if the detector makes a statistically independent decision every second for an energy region of interest of 1 eV. For more details see section 8.3.2.

case the test statistic also needs to be able to deal with unknown signal parameters since the electron production from nuclear decays is a random process.

6.2 Digital Beamforming

The first test statistic under consideration is the result of a process called digital beamforming (DBF). Beamforming is a signal processing technique that is used to focus antenna arrays toward a desired direction [53, 100]. A signal that reaches an antenna array does not appear coherently in the individual elements due to their spatial displacements. Beamforming is used to achieve coherent signals in the desired direction by applying phase shifts or time delays and amplitude weighting to the channels. Digital refers to the fact that the data are first taken incoherently and digitized for each receiver and the beamforming process is subsequently applied by software to the digital data. This offers the advantage of being able to steer the array to "look" in different directions after the data have already been taken.

Beamforming has applications in many technical fields that use antenna arrays such as radar, sonar, seismology, radio astronomy, and communication [53, 100], primarily in the plane-wave approximation where the source distance is large compared to the antenna spacing. The application of DBF in this work is motivated by the decision to use a phased-array antenna and by the fact that electrons can appear at any position inside of the detection volume. However, the geometry of the used antenna array is slightly unusual for a DBF array. Typical DBF arrays would be linear or planar and facing in a single direction with a shared normal vector. Nevertheless, the cylindrical inward facing geometry has the advantage of enabling position pinpointing since the source is surrounded by observers.

Figure 6.4 schematically outlines the idea of DBF for a ring of antennas in the cross-sectional plane of the detector. An electron emitting EM waves with frequency ω_c from a position $\mathbf{p}_e = (x_e, y_e)$ in the plane generates signals that appear with phase shifts at the antennas. The phase shift $\varphi_{ei} = -\omega_c \frac{d_{ei}}{c}$ is determined by the distance d_{ei} between the point and the i -th antenna. The array can be focused onto the point by applying phase shifts of $-\varphi_{ei}$ to the measured data and taking the sum over all antennas. The phase shifted signals are now coherent and constructively interfere, which amplifies the signal after summation. Since electrons can appear at any position and with a range of frequencies, this procedure is repeated with probing positions \mathbf{p}_k and frequencies ω_l .

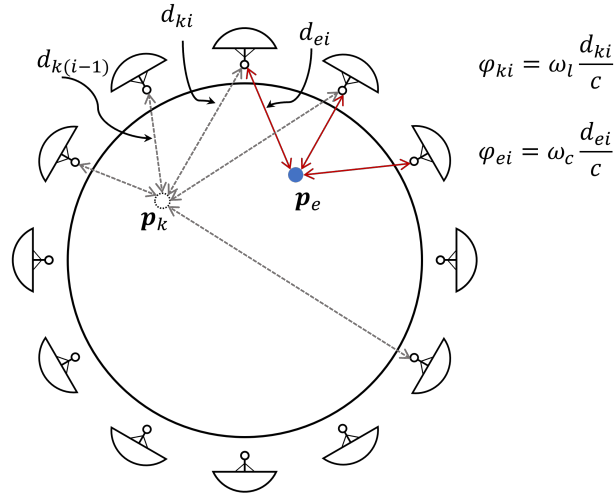


Fig. 6.4.: Schematic for beamforming with a ring of antennas. The picture depicts an electron at a position \mathbf{p}_e and a second probing position \mathbf{p}_k . The array can be focused onto a single point by applying phase shifts to each antenna that correspond to the respective distances to the point.

6.2.1 Mathematical Formulation

Time Domain

The mathematical description starts in the time domain. In the simplest approach it is assumed that the signal portion \mathbf{s} of the data matrix is a sinusoidal with constant frequency ω_c and constant amplitude A that is emitted by a stationary electron at \mathbf{p}_e . At an antenna with position \mathbf{p}_i this assumption results in the following signal

$$\mathbf{s}_{ij} = A_i \exp \left(i \left(t_j (\omega_c - \omega_{LO}) - \omega_c \frac{d_{ei}}{c} + \delta_{ei} \right) \right) . \quad (6.1)$$

In eq. (6.1) t_j denotes the sample time, and ω_{LO} is the frequency of the LO (see section 5.4.4). d_{ei} is the euclidean distance between antenna and electron and δ_{ei} is the Archimedean spiral shaped phase (see section 4.3.3) determined by the azimuth angle of the antenna measured in a coordinate system centered on the gyrating electron:

$$d_{ei}^2 = d_x^2 + d_y^2 = (x_i - x_e)^2 + (y_i - y_e)^2, \quad \delta_{ei} = \arctan 2(d_y, d_x) . \quad (6.2)$$

The goal of beamforming is, when applied to the data with the position of the electron \mathbf{p}_e , to produce a shifted signal \mathbf{s}'_{ij} for each antenna, with the antenna-dependent phases removed from eq. (6.1):

$$\mathbf{s}'_{ij} = A_i \exp(it_j(\omega_c - \omega_{LO})). \quad (6.3)$$

This defines the shift operation $S_{\mathbf{p}_k \omega_l}(\mathbf{x})$ for an arbitrary position \mathbf{p}_k and frequency ω_l that is applied to the data matrix \mathbf{x}

$$(S_{\mathbf{p}_k \omega_l}(\mathbf{x}))_{ij} = \mathbf{x}_{ij} \cdot \exp\left(-i\left(-\omega_l \frac{d_{ki}}{c} + \delta_{ki}\right)\right) = \mathbf{x}_{ij} \cdot e_{lki}. \quad (6.4)$$

By this definition $S_{\mathbf{p}_e \omega_c}(\mathbf{x})$ obviously results in eq. (6.3) for the signal part of \mathbf{x} . Taking the sum over the n_a antennas yields the beamforming result

$$BF_{\mathbf{p}_k \omega_l}(\mathbf{x}) = \sum_{i=0}^{n_a-1} S_{\mathbf{p}_k \omega_l}(\mathbf{x})_i = BF_{\mathbf{p}_k \omega_l}(\mathbf{s}) + BF_{\mathbf{p}_k \omega_l}(\mathbf{n}). \quad (6.5)$$

This result is a single time series of length n_s with an amplified signal amplitude of $\sum_i A_i$.

However, eq. (6.5) also acts on the noise part \mathbf{n} of the data, which produces $\mathbf{n}_{BF} = BF_{\mathbf{p}_k \omega_l}(\mathbf{n}) = \sum_i \mathbf{n}_i \cdot e_{lki}$. The assumption for the noise is again complex WGN, i.e. $\mathbf{n}_{ij} \sim \mathcal{CN}(0, \sigma^2)$, where \mathcal{CN} denotes the circularly-symmetric complex normal distribution. The noise is uncorrelated between antennas and time samples with thermal noise variance σ^2 according to eq. (5.15). Under these assumptions, \mathbf{n}_{BF} is a complex WGN random vector with $\mathbf{n}_{BFi} \sim \mathcal{CN}(0, \sigma^2 \sum_i |e_{lki}|^2)$. Since the phase shifting factor has unity length this means that $\mathbf{n}_{BFi} \sim \mathcal{CN}(0, n_a \sigma^2)$. As a result, the noise remains WGN after beamforming, but its variance increases by a factor of n_a .

Thus, it is convenient to introduce a normalization constant $\frac{1}{\sqrt{n_a}}$, such that eq. (6.5) becomes

$$BF_{\mathbf{p}_k \omega_l}(\mathbf{x}) = \frac{1}{\sqrt{n_a}} \sum_{i=0}^{n_a-1} S_{\mathbf{p}_k \omega_l}(\mathbf{x})_i. \quad (6.6)$$

Using this modification the noise is simply $\mathbf{n}_{BFi} \sim \mathcal{CN}(0, \sigma^2)$. After the normalization one finds that the signal part has an amplitude $\frac{1}{\sqrt{n_a}} \sum_i A_i = \sqrt{n_a} \bar{A}$, which is now clearly amplified due to the use of multiple antennas while the noise remains unchanged due to its incoherence between the antennas.

²If Z_i are random variables with $Z_i \sim \mathcal{CN}(0, \sigma_i^2)$ and $\alpha_i \in \mathbb{C}$ then $Z_s = \sum_i \alpha_i \cdot Z_i$ is a random variable with $Z_s \sim \mathcal{CN}(0, \sum_i |\alpha_i|^2 \sigma_i^2)$.

Frequency Domain

In addition to the previously derived result in the time domain, it is also beneficial to examine the outcome in the frequency domain. In this context, the notation $\mathcal{F}(\mathbf{x})_m$ represents the m -th frequency component of the DFT of the time series \mathbf{x} . Applying a DFT with window size n_s to eq. (6.6) produces

$$\begin{aligned}
 \mathcal{F}(BF_{p_k \omega_l}(\mathbf{x}))_m &= \frac{1}{\sqrt{n_s}} \sum_{j=0}^{n_s-1} BF_{p_k \omega_l}(\mathbf{x})_j \cdot e^{-2\pi i \frac{jm}{n_s}} \\
 &= \frac{1}{\sqrt{n_s n_a}} \sum_{j=0}^{n_s-1} \sum_{i=0}^{n_a-1} \mathbf{x}_{ij} \cdot e^{-i\left(-\omega_l \frac{d_{ki}}{c} + \delta_{ki}\right)} \cdot e^{-2\pi i \frac{jm}{n_s}} \\
 &= \frac{1}{\sqrt{n_s n_a}} \sum_{i=0}^{n_a-1} e^{-i\left(-\omega_l \frac{d_{ki}}{c} + \delta_{ki}\right)} \sum_{j=0}^{n_s-1} \mathbf{x}_{ij} \cdot e^{-2\pi i \frac{jm}{n_s}} \\
 &= \frac{1}{\sqrt{n_a}} \sum_{i=0}^{n_a-1} e^{-i\left(-\omega_l \frac{d_{ki}}{c} + \delta_{ki}\right)} \mathcal{F}(\mathbf{x}_i)_m. \tag{6.7}
 \end{aligned}$$

With the above result it is possible to add a modification that improves the way the unknown frequency ω_l is handled. So far in the derivation it is assumed arbitrary and eqs. (6.6) and (6.7) can be evaluated for any exact value of ω_l . However, the bandwidth B of possible frequencies of the data is limited by the sampling rate and the original frequencies of the emitted EM wave are converted into B by the LO. Therefore, by using $\omega_l = \omega_{LO} + \omega_0$ with $\omega_0 \in B$ all frequencies that can possibly cause the phase shift in the data are covered. Furthermore, the frequencies ω_m of the DFT, which are discrete values with $\omega_m \in B$, are an excellent candidate for ω_0 . Thus, with this modification of eq. (6.7) the final definition of frequency domain beamforming $BF_{p_k}(\mathbf{x})$ for a data array \mathbf{x} at a probe point p_k is

$$(BF_{p_k}(\mathbf{x}))_m = \frac{1}{\sqrt{n_a}} \sum_{i=0}^{n_a-1} e^{-i\left(-(\omega_m + \omega_{LO}) \frac{d_{ki}}{c} + \delta_{ki}\right)} \mathcal{F}(\mathbf{x}_i)_m. \tag{6.8}$$

The subtle change seems almost superficial, but it removes the additional dimension ω_l from eq. (6.7). In addition, it has the advantage that the result now uses a multitude of phase shifts, which are all exact for the corresponding frequency in the spectrum. Importantly, this is also the correct way for the case of a signal with more than one frequency, where the initial derivation starting from the time domain with a single assumed frequency is inferior³.

³It should be mentioned that in practice the phase shifts are precise enough with any frequency that is comparable to ω_c and using e.g. $\omega_l = \omega_{LO}$ fixed would be sufficient. Nonetheless, the correct treatment here is still important for section 6.3.3.

The result of frequency domain beamforming is a frequency spectrum of a signal that is the combination of all antennas for the assumed source position p_k . Hence, each probed position gets its own frequency spectrum assigned. Using eq. (6.8) to calculate the spectra of n_p probing points can have a computational advantage compared to evaluating eq. (6.6) and taking DFTs. The only part that depends on the evaluation point is the phase shift term while $\mathcal{F}(\mathbf{x}_i)$ is the same for all positions and can be calculated once in advance. Therefore, the frequency domain version requires n_a DFTs, whereas the time domain version requires n_p DFTs using a window size of n_s in both cases. Many cases require that $n_p \gg n_a$ (see section 6.2.2) and thus the frequency domain calculation would be at an advantage. The calculation of $\mathcal{F}(\mathbf{x}_i)$ is ideally executed with the symmetric DFT using the normalization $\frac{1}{\sqrt{n_s}}$, as is the case in eq. (6.7). With this normalization choice the noise again remains complex WGN with the same variance as before.

6.2.2 The Test Statistic

The outcome of beamforming is a frequency spectrum of size n_s , which should have a strong signal above the noise floor if an electron is present and if the evaluation takes place for its position. Due to the unknown position, it is mandatory to repeat the DBF procedure for many positions. If this is done on a $n_g \times n_g$ 2D grid G in the cross-sectional plane this produces a $n_s \times n_g \times n_g$ tensor \mathbf{x}_{BF} . Converting to power spectra using eq. (5.16), each frequency in the tensor corresponds to a spatial power map for that frequency, which can be visualized as an image. Figure 6.5 shows several of such images for different electron positions using a grid of size 101×101 . The images were taken at the frequency that contains the highest power peak across all grid points and frequencies and clearly show the electrons close to their true position.

Figure 6.6 compares the resultant spectra at different beamforming positions for an electron at the center. For $\alpha_0 = 90^\circ$, the spectra display a single peak. In the beamformed spectrum of the true position, the power of this peak is enhanced compared to the single-channel spectrum and clearly stands out above the noise floor. However, at offset positions, the power of this peak is reduced and may even drop below the single channel power due to destructive interference.

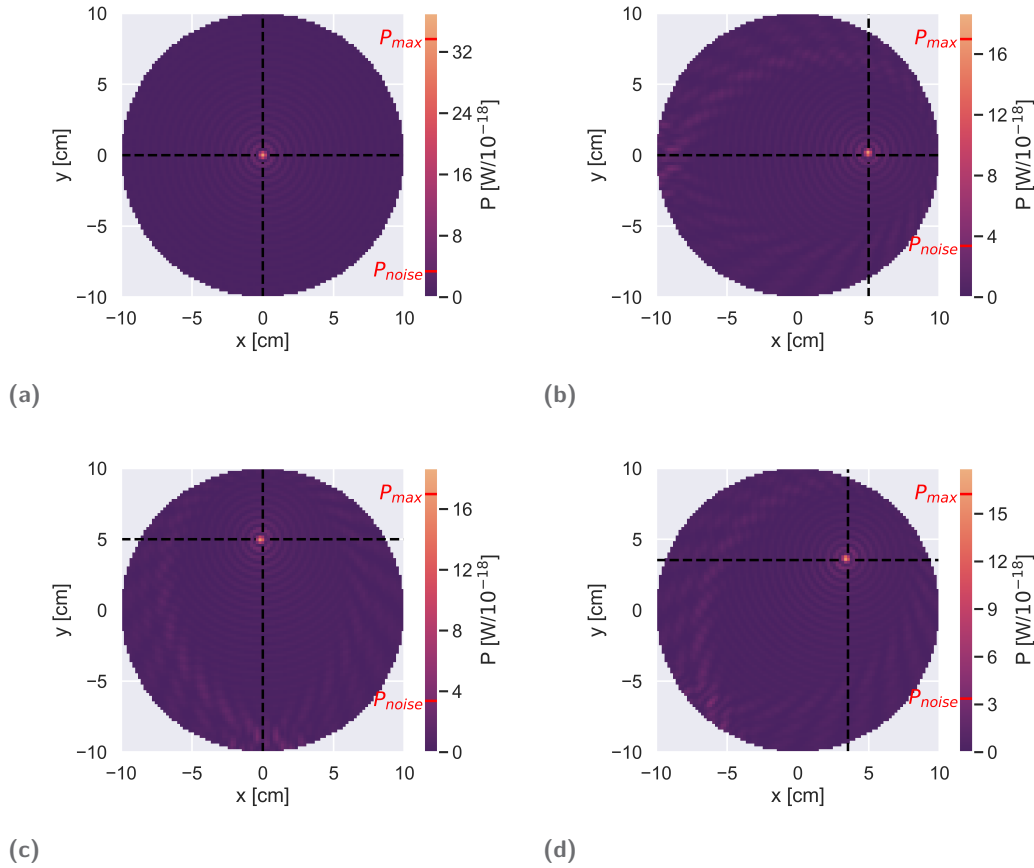


Fig. 6.5.: Results of beamforming with a simulated $\alpha_0 = 90^\circ$ electron in a harmonic trap with different positions in the plane $z = 0$ (without noise). An array with 60 of the Project 8 prototype antennas on a ring with radius 0.1 m in the plane $z = 0$ is used. Interference patterns are visible that are characteristic for beamforming with high powered peaks close to the true electron positions, marked by the intersection of the dashed lines. P_{max} marks the peak's power value, P_{noise} the expectation for 5 K noise⁴. See text for details on the image production.

Following these observations, an obvious test statistic to build on top of the described DBF procedure is a threshold on the beamformed spectra similar to Phase II (see also fig. 6.1 and [66]). The test statistic is

$$T(\mathbf{x}) = \max_{m \in M, \mathbf{p}_k \in G} \left[(BFF_{\mathbf{p}_k}(\mathbf{x}))_m \right], \quad (6.9)$$

where M denotes the frequencies of the DFT and G the $n_g \times n_g$ grid that is used for the DBF evaluation. Finally, this detector would decide \mathcal{H}_1 if $T(\mathbf{x}) > \gamma$ for a threshold γ that is appropriate for the desired P_{FA} .

⁴Conversion of the noise to a power spectrum with eq. (5.16) transforms the Gaussian distribution to an exponential distribution with mean $\frac{\sigma^2}{n_s R}$.

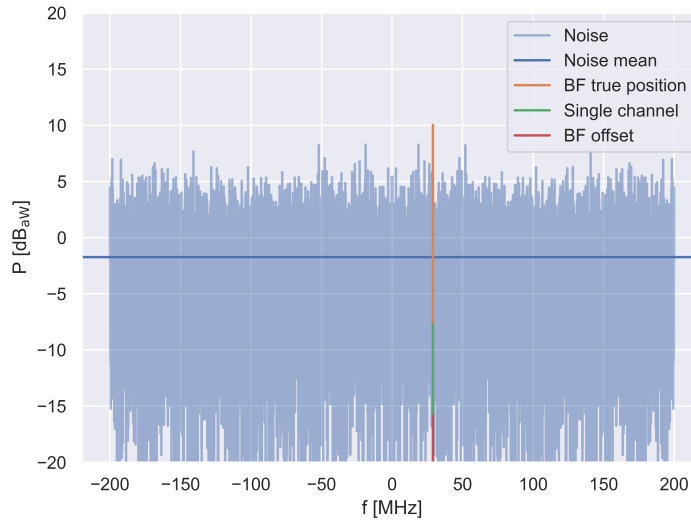


Fig. 6.6.: Power spectra for electrons in a harmonic trap with $\alpha_0 = 90^\circ$ at $\rho = 0$ using a $42 \mu\text{s}$ DFT. Comparison of beamformed spectrum at the electron's true position, beamformed spectrum 1 cm offset, single channel's spectrum, and a noise spectrum with $T = 5 \text{ K}$. Powers on logarithmic scale relative to $1 \text{ aW} = 10^{-18} \text{ W}$ in units dB_{aW} . *Hamming* window is applied to reduce spectral leakage.

This test statistic uses the assumption that the electron is static in the cross-sectional plane of the detector. However, depending on the magnetic field and the pitch angle the electron actually moves both in the plane due to the drift motion and out of the plane due to the axial motion.

6.2.3 Extension for Drift Motion

A case where drift motion in the plane becomes problematic is presented in fig. 6.7. The peak power found with DBF is roughly an order of magnitude smaller than in fig. 6.5. This power loss is induced by two distinct effects. As evident from the plot, the power is simply not found at a single location any more due to the electron moving away from the focus point. The second more subtle reason is the Doppler effect, which slightly shifts the observed frequency of an antenna. Due to the antenna arrangement this shift is not consistent for all antennas, which was demonstrated in fig. 5.21. As a consequence, the beamformed power is also not at a single frequency peak either, which can be seen in fig. 6.9a.

The lower power caused by these two effects significantly weakens the detector defined by the test statistic in eq. (6.9) for cases with strong drift motion. Cases like this will occur for electrons with any pitch angle in a magnetic field that has

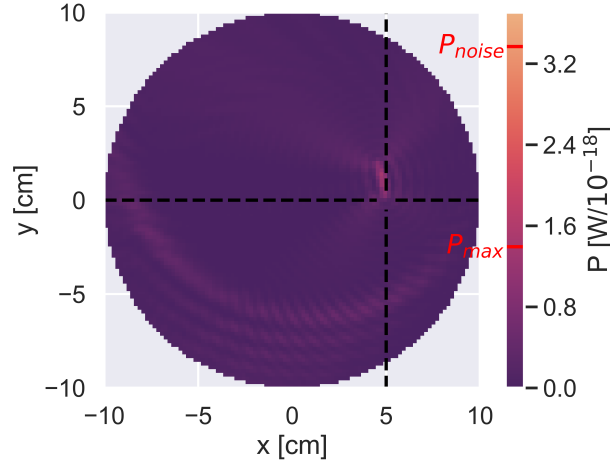


Fig. 6.7.: Beamforming result of an electron with $\alpha_0 = 90^\circ$ in a trap consisting of two EM coils that generate a considerable drift motion. The power of the electron signal spreads to several positions of smaller power below the noise level instead of a single high power peak.

a considerable orthogonal gradient or curvature, which define the drift velocity according to eq. (4.16).

In the harmonic trap that was used in fig. 6.5, this effect is negligibly small and might have gone unnoticed: upon closer examination, however, the peak exhibits a very slight shift from the true position. In the current trap, this effect is much more pronounced. Furthermore, the coil trap is a configuration that could be implemented in a laboratory setup as it is, unlike the harmonic trap that is purely analytic without a real world implementation. Hence, in the context of a real-world configuration, the occurrence of drift motions is a feature that must be anticipated.

The prospects of inadequate detection under these circumstances motivate the development of a modified beamforming procedure, designed to track the drift motion of electrons. It builds upon the base premise that the electron position is now a circular trajectory

$$\mathbf{p}_e(t) = \begin{pmatrix} \rho_e \cos(\omega_D t + \phi_e) \\ \rho_e \sin(\omega_D t + \phi_e) \end{pmatrix}, \quad (6.10)$$

where ω_D denotes a constant angular drift frequency, and (ρ_e, ϕ_e) denotes the initial electron position in polar coordinates. Assuming a constant frequency ω_c the assumed antenna signal is

$$\mathbf{s}_{ij} = A_i(t_r(t_j)) \exp [i(\omega_c t_r(t_j) - \omega_{LO} t_j + \delta_{ei}(t_r(t_j)))] , \quad (6.11)$$

which is the analog to eq. (6.1) of the static case. What sets this apart from the previous version are the introduction of time dependence for A and δ_e and the evaluations at the retarded time t_r to account for the electron motion⁵.

The retarded time is defined in eq. (4.23), which unfortunately has no closed-form solution for the trajectory of eq. (6.10). However, the drift motion still is an effect that takes place on a significantly longer time scale than the sampling time step Δt . Therefore, in a detector with a radius on the scale of centimeters to meters it holds that $v_D \Delta t \ll d(t_r)$.

This motivates a Taylor expansion of eq. (4.23):

$$c(t - t_r) = d(t_r) \approx d(t) + \left(\frac{dd(t_r)}{dt_r} \right)_t (t_r - t) + \frac{1}{2} \left(\frac{d^2 d(t_r)}{d(t_r)^2} \right)_t (t_r - t)^2 + \mathcal{O}(t_r^3). \quad (6.12)$$

By using expansion terms with orders 2 or less it is possible to find a solution of eq. (6.12) for t_r . E.g. with just the 0th order expansion one trivially finds that

$$t_r = t - \frac{d(t)}{c}. \quad (6.13)$$

In light of its simplicity, this is the preferred solution. More complex higher-order solutions can still be considered if inadequate results are observed for very high drift velocities. So far this solution has proven sufficient for common drift velocities.

Substituting eq. (6.13) into eq. (6.11) produces

$$\mathbf{s}_{ij} = A_i(t_j) \exp \left(i \left(t_j (\omega_c - \omega_{LO}) - \omega_c \frac{d_{ei}(t_j)}{c} + \delta_{ei}(t_j) \right) \right),$$

which has the exact same form as the static case due to the 0th order expansion, except for the dependencies on the sample time t_j which tracks the drift motion. The goal of DBF again is to produce shifted signals that remove the antenna dependent phases as in eq. (6.3), which is done analogous to the static case using eq. (6.4) with an arbitrary position \mathbf{p}_k and frequency ω_l . The position $\mathbf{p}_k = (\rho_k, \phi_k)$ now defines an assumed starting position, which evolves over time like \mathbf{p}_e in eq. (6.10).

Thus, DBF with tracking of the drift velocity for the data matrix \mathbf{x} is defined as the following procedure $BFD_{\mathbf{p}_k}(\mathbf{x})$ that produces the beamformed signal time series:

- For all antennas i and time samples j calculate $d_{ki}(t_j)$ and $\delta_{ki}(t_j)$ with eq. (6.2) assuming the instantaneous drift position $\mathbf{p}_k(t_j)$

⁵Strictly speaking, the evaluation at the retarded time is already incorporated in eq. (6.1) since $t_r(t_j) = t_j - \frac{d_{ei}}{c}$ for a static electron.

- Calculate eq. (6.4) with $d_{ki} = d_{ki}(t_j)$ and $\delta_{ki} = \delta_{ki}(t_j)$ for all i and j
- Take the sum over antennas i as in eq. (6.6)
- For simplicity use a fixed $\omega_l = \omega_{LO}$

Based on this procedure the new test statistic is

$$T(\mathbf{x}) = \max_{m \in M, \mathbf{p}_k \in G} \left[\mathcal{F} (BFD_{\mathbf{p}_k}(\mathbf{x}))_m \right].$$

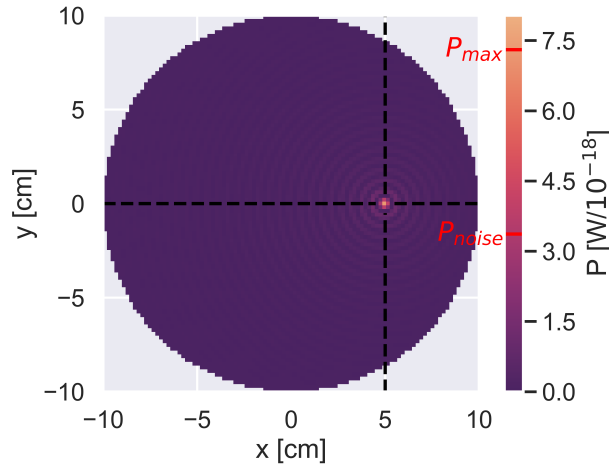


Fig. 6.8.: Beamforming result with drift motion tracking using the same simulation setup as in fig. 6.7.

Figures 6.8 and 6.9 present results with this new test statistic for a simulation setup identical to the one in fig. 6.7. In the beamforming image depicted in fig. 6.8 all the power is now concentrated at a single position, which exactly coincides with the true electron position.

Figure 6.9 compares results with drift motion tracking to regular beamforming results. The power loss for regular beamforming due to the Doppler shift is evident from fig. 6.9a where the frequency peak of the beamformed spectrum splits into two smaller peaks. With the drift motion tracking, however, the issue is resolved as the power spectrum indeed only has a single peak. Moreover, from fig. 6.9b it is evident that the use of standard DBF results in a significant reduction in detected power for all cases with $\rho > 0$. Once again, most of the power is recovered by tracking the drift motion.

The large variations with ρ in the plot are caused by binning effects. By varying ρ it can occur that the electron position falls between grid points and thus the power is

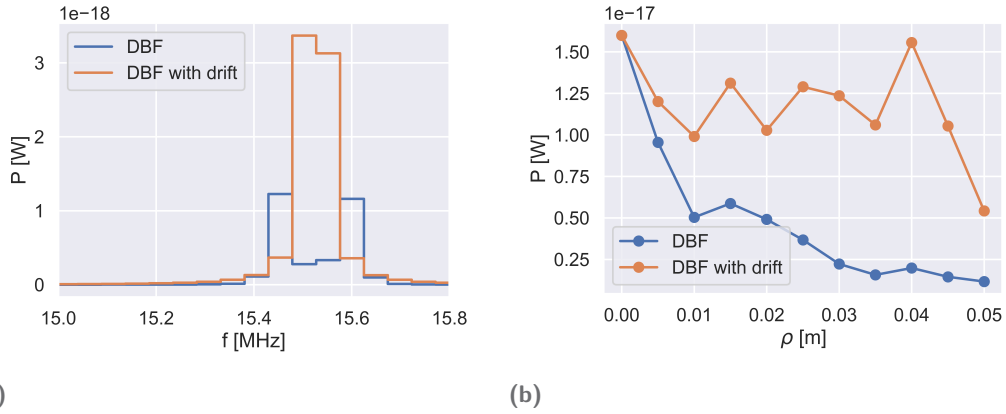


Fig. 6.9.: Beamforming results with and without drift motion tracking. (a) shows a close-up of the power spectrum for $\rho = 0.05$ m. With DBF the signal appears with two separate peaks due to Doppler shifted signals in the antennas, which is fixed by accounting for the drift motion. (b) depicts power as a function of radial position ρ , which increases by accounting for the drift.

shared between several positions, which is not accounted for with the simple test statistic. The same effect appears with the DFT bins since the electron frequency depends on ρ . An extreme case of this is seen in fig. 6.9a, which also explains the considerable drop of power at $\rho = 0.05$ m. This demonstrates that the success of the test statistic can be greatly diminished if the evaluation points in space and frequency are too sparse or poorly placed. The binning issues might be alleviated by applying an appropriate window function prior to the DFT. However, this must be chosen carefully, since window functions can also reduce the SNR.

A crucial differentiating factor from the first test statistic is the fact that the DFT must be the last step before the threshold. A similar simplification that allows the use of the DFT of the data matrix, as it was achieved with eq. (6.8), is not possible since the shifting terms depend on the sampling time t_j . Therefore, a DFT for every grid point is mandatory. Furthermore, with the phase shifts depending on t_j , they have to be recalculated for every time sample. Both these factors lead to an increased demand of computational resources compared to the standard DBF approach.

As a final remark, the drift frequency ω_D that is used for the tracking is in general a free parameter of the detection procedure. Strictly speaking, it should be found by optimization. However, knowing the magnetic field of a setup, an estimate can be calculated using the theoretic drift velocity (eq. (4.16)). Unfortunately, this estimate depends on the unknown pitch angle, which, up to this point in the discussion of DBF, has been assumed to be 90° .

6.2.4 Axial Motion

For $\alpha_0 < 90^\circ$ the electron additionally performs axial motion and thus it moves outside of the DBF plane. With the electron shifting away from the focus point, this presents a problem that is similar to what was encountered with the drift motion. However, the axial position of the electron oscillates with a frequency of several MHz, which is significantly faster than the drift motion. With the 200 MHz sampling rate this means that the electron completes many axial cycles over the time window of the DFT. Therefore, a pixel in the beamforming image now only represents the electron's polar coordinates (ρ_e, ϕ_e) instead of the actual guiding center position. If the array is focused onto that polar position with DBF, it still detects a signal due to the many electron passages.

Yet, as illustrated in fig. 6.10, the detection of that beamformed signal in the presence of noise is substantially more challenging. As demonstrated before in section 5.5, the smaller pitch angles produce modulated signals with sidebands, which cannot be detected in noise by the simple peak finding test statistic

Following the success of the drift motion tracking from the previous section, the obvious next step seems to be an expansion of the test statistic that includes tracking of the axial motion. Nevertheless, there are many counterarguments to this:

- This adds the pitch angle as another dimension of the test statistic.
- The axial trajectory has no simple universally applicable solution that can be tracked in the same way. The solution is always specific for a trap.
- The Taylor expansion solution to the retarded time of eq. (6.12) is no longer viable for the faster axial motion.
- The presented drift motion tracking becomes ineffective since the gradient and curvature of the field are not constant along the symmetry axis.

The resulting detection algorithm would be quite sophisticated and computationally demanding. Moreover, for this whole chapter it was assumed that the antennas are only arranged on a ring in the cross-sectional plane. In this arrangement the antennas are unable to continuously observe the electron in their field of view. Hence, consistent with the initial concept depicted in fig. 3.2, the axial motion mandates the incorporation of additional antennas along the symmetry axis in order to improve the detection performance. In the established DBF framework it is not clear yet how additional antennas should contribute to the test statistic. In the next

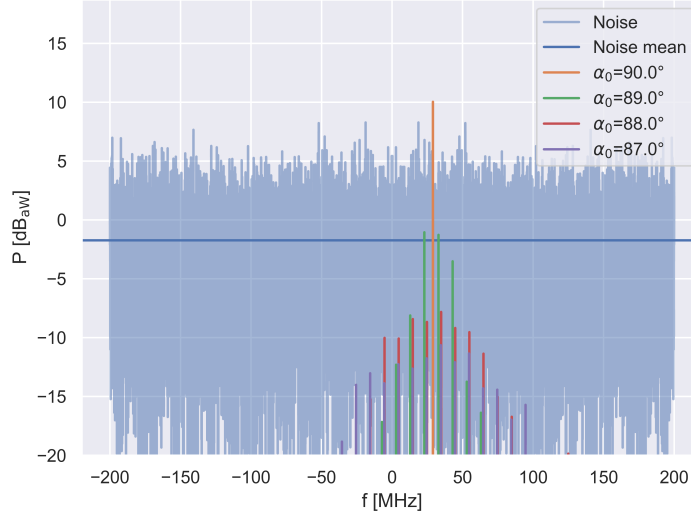


Fig. 6.10.: Power spectra of the beamformed signal at the electron's position for electrons in a harmonic trap with pitch angles α_0 at $\rho = 0$ obtained using a $42 \mu\text{s}$ DFT. Powers on log scale relative to $1 \text{ aW} = 10^{-18} \text{ W}$ in units dB_{aW} . Plot also shows noise power spectrum with $T = 5 \text{ K}$ for comparison. A *Hamming* window function is applied to reduce spectral leakage.

section a general detection framework is established that is valid for any signal structure and antenna arrangement.

6.3 Matched Filter

A general solution for the detection problem is provided in standard literature [99] with the Neyman-Pearson (NP) test. For a vector \mathbf{x} of observations the NP test relies on the likelihood ratio of the two hypotheses \mathcal{H}_1 and \mathcal{H}_0 defined in section 6.1, meaning that \mathcal{H}_1 is the signal plus noise hypothesis and \mathcal{H}_0 the noise-only hypothesis. As a result, the test statistic

$$T(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)}.$$

defines the detector, which decides \mathcal{H}_1 if $T(\mathbf{x}) > \gamma$. The Neyman-Pearson theorem then states that this detector maximizes the detection probability P_D for a given false alarm probability $P_{FA} = \alpha_0$ defined by

$$P_{FA} = \int_{\{\mathbf{x}|T(\mathbf{x})>\gamma\}} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x} = \alpha_0.$$

While this is a general solution to the detection problem, for a specific application it still necessitates models for the two PDFs $p(\mathbf{x}; \mathcal{H}_{0/1})$ that describe the random vector \mathbf{x} under the two hypotheses. For the detection of a known signal \mathbf{s} in general additive Gaussian noise the PDFs that determine \mathbf{x} are both multivariate complex normal distributions with a $n_s \times n_s$ covariance matrix \mathbf{C} . More precisely under \mathcal{H}_0 , $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{C})$ and under \mathcal{H}_1 , $\mathbf{x} \sim \mathcal{CN}(\mathbf{s}, \mathbf{C})$. Under these assumptions, the NP detector becomes the so-called matched filter (MF) with the test statistic [99]

$$T(\mathbf{x}) = \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{x} \right) . \quad (6.14)$$

\mathbf{s}^H denotes the conjugate transpose of the signal vector \mathbf{s} . Thus, the operation in the test statistic is the inner product of \mathbf{s} and $\mathbf{C}^{-1} \mathbf{x}$, i.e. a sum over all time samples.

The general covariance matrix \mathbf{C} includes the variances of each observation in the random vector \mathbf{x} and the correlations between them. In many practical applications with well-designed electronics, the noise-generating process can be assumed to be identical for each observation and observations independent of one another. In this case, $\mathbf{C} = \sigma^2 \mathbf{1}$, where σ^2 denotes the variance for each individual observation. Although this assumption is typically applied to CRES data, the discussions will retain the more general covariance matrix for now, as the results remain valid in non-ideal scenarios, provided the covariance matrix can be estimated from the data.

6.3.1 Multi-Channel Case

The generalization for the data matrix \mathbf{x} of an antenna array with multiple channels is straightforward by appropriately rearranging the data. The $n_a \times n_s$ data and signal matrices can both be written as flat vectors $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{s}}$ in row-major order, e.g. $\tilde{\mathbf{x}}_i = \mathbf{x}_{jl}$ with $j = \lfloor i/n_s \rfloor$ and $l = i \bmod n_s$. The corresponding noise covariance matrix $\tilde{\mathbf{C}}$ is a $(n_s n_a) \times (n_s n_a)$ matrix that contains the $n_s \times n_s$ covariance matrices \mathbf{C}_j of the isolated measurements of each channel as sub-matrices as well as potential correlations between channels. In general, the covariance matrices \mathbf{C}_j can also differ between channels.

With these rearrangements the solution to the detection problem is given by eq. (6.14) as well and thus the test statistic for array data with the most general covariance matrix is [99]

$$T(\mathbf{x}) = \text{Re} \left(\tilde{\mathbf{s}}^H \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{x}} \right) . \quad (6.15)$$

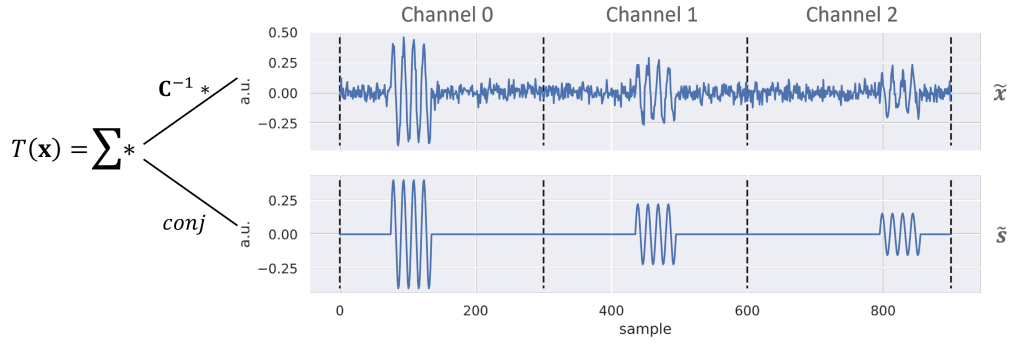


Fig. 6.11.: Calculation of the test statistic with the multi-channel MF. The signal \tilde{s} is an arbitrary example. For the data \tilde{x} simple uncorrelated white noise is added to the same signal. The noise variance is substantially smaller than in a real CRES detection scenario in order to better visualize how the test statistic filters out the signal from the data.

Although it has the same form as eq. (6.14), the inner product of the new test statistic is a sum over both channels and time samples due to the rearranged data. The calculation of this test statistic is depicted in fig. 6.11 for an example signal.

This general result can be simplified by assuming that correlations between channels are negligible, a condition known to be true for thermal noise if the antenna spacing is greater than a wavelength [101]. The result is a covariance matrix \tilde{C} that is mostly zero except for the sub-matrices C_j (meaning that it may still include correlations between time samples). With this covariance matrix one finds that eq. (6.15) simplifies to [99]

$$T(\mathbf{x}) = \sum_{j=0}^{n_a-1} T_j(\mathbf{x}_j) = \sum_{j=0}^{n_a-1} \text{Re} \left(\mathbf{s}_j^H \mathbf{C}_j^{-1} \mathbf{x}_j \right). \quad (6.16)$$

Hence, the test statistic for an antenna array is a simple sum of the test statistics T_j of all channels. As long as the antenna spacing is greater than the wavelength, eq. (6.16) is the optimal test statistic for array data independent of the antenna arrangement. This is an important finding because a major concern with large antenna arrays was a perceived ambiguity in incorporating more antennas into the beamforming detection framework.

Yet, for this optimal detector the exact knowledge of the expected signals in each channel is mandatory. Fortunately, although the signals can differ between the channels they are not entirely independent. Instead, all signals depend on a common source, which deterministically defines the full signal matrix \mathbf{s} . This is also illustrated with the example in fig. 6.11, in which all channels receive the same signal with different amplitudes and delay times. In a CRES detection the common source

would of course be an electron. Thus, the optimal detection is achieved with a suitable model for the signal matrix that is produced by an electron track (for a definition of electron tracks, see section 5.1) with its eight parameters $\boldsymbol{\theta} = (E_{\text{kin}}, \alpha_0, \rho, \phi, \varphi_{\text{axial}}, \varphi_{\text{cycl}}, t_0, \tau)$. Since the tracks are produced in random processes, it is impossible to know the track parameters for the signal matrix in practice. For that reason, the next section will focus on the case of detection using a model with unknown parameters for the signal matrix.

6.3.2 Unknown Parameters

A detector using the test statistic in eq. (6.15) should appropriately be termed a "clairvoyant" detector, as it achieves the best possible detection performance only if the true signal is perfectly known. However, in scenarios with incomplete knowledge of the signal, detectors often must rely on signal models with unknown parameters, leading to a degradation in detection performance compared to the ideal case. This degradation arises from the inherent uncertainty in estimating the unknown parameters, rather than from potential mismatches between the assumed and true signal, whether due to incorrect parameter values or errors in modeling.

The case in which the PDFs of both \mathcal{H}_1 and \mathcal{H}_0 can have unknown parameters is a more general composite hypothesis test. This can be approached by a generalized likelihood ratio test (GLRT), which evaluates the test statistic at the maximum likelihood estimate (MLE) of the unknown parameters under both hypotheses [99]. For the case of Gaussian noise with known covariance matrix \mathbf{C} , only the PDF under $\mathcal{H}_1: \mathbf{x} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{n}$ depends on unknown parameters, while $\mathcal{H}_0: \mathbf{x} = \mathbf{n}$ is parameter-free. Thus the test statistic is

$$T(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)},$$

where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ under \mathcal{H}_1 . The MLE is found by maximization, which yields the equivalent test statistic

$$T(\mathbf{x}) = \max_{\boldsymbol{\theta}} \frac{p(\mathbf{x}; \boldsymbol{\theta}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)}.$$

Using the PDFs of a complex Gaussian under \mathcal{H}_0 and \mathcal{H}_1 , the GLRT is

$$\max_{\boldsymbol{\theta}} \exp\left(-(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^H \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})) + \mathbf{x}^H \mathbf{C}^{-1} \mathbf{x}\right) > \gamma,$$

which is equivalent to the test

$$\max_{\boldsymbol{\theta}} \left(2 \operatorname{Re} \left(\mathbf{s}(\boldsymbol{\theta})^H \mathbf{C}^{-1} \mathbf{x} \right) - \mathbf{s}(\boldsymbol{\theta})^H \mathbf{C}^{-1} \mathbf{s}(\boldsymbol{\theta}) \right) > \gamma'. \quad (6.17)$$

This is the most general solution without further assumptions about the unknown parameters or the signal model. In the next part of this section, this general solution is adapted to account for the specific unknown electron parameters, resulting in specialized detection approaches.

Unknown Phase

In many situations it must be assumed that the signal arrives at the detector with an arbitrary phase that cannot be known in advance. This is relevant for an electron track because the initial phase of the cyclotron motion φ_{cycl} is unknown. Additionally, phase shifts can be imposed on the analog signal by the readout electronics prior to digitization. For example, even if the source signal is perfectly known including its phase, using eq. (6.14) can still result in poor detection performance if the instantaneous phase of the local oscillator is disregarded.

Assuming that the signal \mathbf{s} is known except for an unknown phase φ , the model is $\mathbf{s}_\varphi(\varphi) = e^{i\varphi} \mathbf{s}$. Inserting $\mathbf{s}_\varphi(\varphi)$ into eq. (6.17) results in the GLRT under Gaussian noise for this model:

$$\max_{\varphi} \left(2 \operatorname{Re} \left(\left(e^{i\varphi} \mathbf{s} \right)^H \mathbf{C}^{-1} \mathbf{x} \right) - \left(e^{i\varphi} \mathbf{s} \right)^H \mathbf{C}^{-1} e^{i\varphi} \mathbf{s} \right) > \gamma. \quad (6.18)$$

Using $(e^{i\varphi})^* e^{i\varphi} = 1$, where $(\cdot)^*$ denotes complex conjugation, simplifies eq. (6.18) to

$$\max_{\varphi} \left(2 \operatorname{Re} \left(e^{-i\varphi} \mathbf{s}^H \mathbf{C}^{-1} \mathbf{x} \right) \right) > \gamma + \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}. \quad (6.19)$$

The result of the inner product $\mathbf{s}^H \mathbf{C}^{-1} \mathbf{x}$ is a complex number z with $z = |z| e^{i\varphi_z}$. Thus, $\operatorname{Re} \left(e^{-i\varphi} z \right) = \operatorname{Re} \left(|z| e^{i(\varphi_z - \varphi)} \right) = |z| \cos(\varphi_z - \varphi)$. This expression is maximal for $\varphi = \varphi_z$, which produces $\max_{\varphi} |z| \cos(\varphi_z - \varphi) = |z|$. As a result, eq. (6.19) transforms into

$$\left| \mathbf{s}^H \mathbf{C}^{-1} \mathbf{x} \right| > \frac{1}{2} \left(\gamma + \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} \right) = \gamma'.$$

In most standard literature [99], detection with unknown phase is addressed using the equivalent test statistic

$$T(\mathbf{x}) = \left| \mathbf{s}^H \mathbf{C}^{-1} \mathbf{x} \right|^2. \quad (6.20)$$

This is referred to as a quadrature matched filter (QMF), and for unknown phases, this test statistic serves as a universal replacement for the matched filter test statistic from eq. (6.14).

Unknown Arrival Time

Another very common problem is an unknown arrival time. For the case of the electron detection this is reflected in the start time t_0 , which is the result of an unpredictable decay. For a signal s , known except for its delay, embedded in simple WGN (i.e. $\mathbf{C}^{-1} = \frac{1}{\sigma^2} \mathbf{1}$) the generalized likelihood ratio test yields the test statistic

$$T(\mathbf{x}) = \max_{n \in [0, n_x - n_\tau]} \operatorname{Re} \left(\sum_{m=n}^{n+n_\tau-1} \mathbf{x}_m \mathbf{s}_{m-n}^* \right). \quad (6.21)$$

Here, n and $n_\tau = \lfloor \tau f_s \rfloor$ denote the delay and signal length in discrete time samples and n_x the number of time samples in \mathbf{x} . The MLE of n yields an estimator for the start time $\hat{t}_0 = \frac{\hat{n}}{f_s}$, though its resolution is limited by the sampling interval. The proper non-discrete estimation of t_0 will be discussed in chapter 7, but for the signal detection alone this is not required.

Equation (6.21) has the form of a convolution of the data with a FIR filter \mathbf{h} . The discrete convolution without zero-padding with a filter of length n_τ can be written as

$$(\mathbf{x} * \mathbf{h})_n = \sum_{m=n}^{n+n_\tau-1} \mathbf{x}_m \mathbf{h}_{n_\tau-1+n-m}.$$

With this definition one can easily see that the test statistic is in fact a convolution with a filter that is the complex conjugate reversed signal. In other words, using $\mathbf{h}_n = \mathbf{s}_{n_\tau-1-n}^*$, eq. (6.21) can be written as

$$T(\mathbf{x}) = \max_{n \in [0, n_x - n_\tau]} \operatorname{Re} (\mathbf{x} * \mathbf{h})_n. \quad (6.22)$$

This implementation as a filter operation is the origin of the term *matched filter*, since it is a very common approach to apply a set of FIR filters to data in order to find matching patterns or signals in it. This is also closely related to classic pattern detection in images, in which the same is done with 2-dimensional filters.

In a practical implementation the test statistic is often not evaluated for all discrete delays n . In order to relax the computational requirements it is instead common to slide the filter with a stride over the data. Typically stride lengths with 50% or less overlap between the windows are employed. In the extreme case of no overlap, the

data time series x is split into time slices to which the matched filter test statistic (e.g. eq. (6.20)) is applied individually⁶.

Apart from the computational benefits, using non-overlapping time slices has the additional advantage of simplifying the theoretical treatment of the detection performance: the test statistic of eq. (6.22) is a maximum over correlated random variables since the same noise samples influence the results of multiple delays n . Without overlap, these random variables become independent and therefore, the detection probability in each time slice is independent of other time slices.

Track Length

A situation that is less common in typical signal detection problems, such as radar applications, is that of a signal with unknown length τ . Still, this problem is characteristic for CRES detection since electron tracks disappear in a random scattering event. Including the random track length τ as an unknown signal parameter into the sliding filter approach from above results in an additional optimization over a varying filter length n_τ . This method is impractical since the number of required inner product evaluations does not scale linearly with the number of data samples n_x .

A better approach would be to use a fixed filter length n_s . In this case, a signal would result in a sequence of time slices with high test statistic outputs. Due to the noise, the individual time slice results will fluctuate, and therefore, the detector ultimately needs to decide if a sequence of time slice results is caused by a consecutive track or not. A good choice for the arbitrary filter length could be the mean of the track length $n_s = \lfloor \langle \tau \rangle f_s \rfloor$.

While a proper treatment of the detection with a sequence of time slice results is beyond the scope of this thesis, at this point, existing research on this problem in the context of Project 8 is worth noting: In [102] an application of the Viterbi algorithm [103] is explored in order to decide if a sequence of observed time slice filter results is caused by an electron or not. The algorithm treats the data as a hidden Markov model, generating the observed test statistic outputs for each time slice, with the hidden state representing either \mathcal{H}_0 or \mathcal{H}_1 . Utilizing dynamic programming, the Viterbi algorithm determines the most likely sequence of hidden states corresponding to the observed outputs.

⁶This is analogous to a spectrogram produced by a short-time Fourier transform.

Full Set of Electron Parameters

For the remaining electron parameters, no special treatment can be applied. With the full set of electron parameters unknown, the optimal detector is given by eq. (6.17). By separating the static phase shifts φ from the remaining unknown parameters $\boldsymbol{\theta}$, the detector decides \mathcal{H}_1 if

$$\max_{\boldsymbol{\theta}} \max_{\varphi} \left(2 \operatorname{Re} \left(\left(e^{i\varphi} \mathbf{s}(\boldsymbol{\theta}) \right)^H \mathbf{C}^{-1} \mathbf{x} \right) - \left(e^{i\varphi} \mathbf{s}(\boldsymbol{\theta}) \right)^H \mathbf{C}^{-1} e^{i\varphi} \mathbf{s}(\boldsymbol{\theta}) \right) > \gamma.$$

As before, maximizing over φ results in the absolute value of the inner product. As a result, the generalized likelihood ratio test (GLRT) for electron detection in Gaussian noise decides \mathcal{H}_1 if

$$\max_{\boldsymbol{\theta}} \left(2 \left| \mathbf{s}(\boldsymbol{\theta})^H \mathbf{C}^{-1} \mathbf{x} \right| - \mathbf{s}(\boldsymbol{\theta})^H \mathbf{C}^{-1} \mathbf{s}(\boldsymbol{\theta}) \right) > \gamma. \quad (6.23)$$

However, in practical applications of this kind of detection scheme, e.g. in [104], one often finds a slightly different version of the test statistic of eq. (6.23) in which the energy term $\mathbf{s}(\boldsymbol{\theta})^H \mathbf{C}^{-1} \mathbf{s}(\boldsymbol{\theta})$ is used as a normalization instead of subtracting it. With the quadrature matched filter for the case of multiple channels and using $\tilde{\sim}$ as notation for flat data vectors, this results in the test statistic

$$T(\mathbf{x}) = \max_{\boldsymbol{\theta}} \frac{\left| \tilde{\mathbf{s}}(\boldsymbol{\theta})^H \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{x}} \right|^2}{\tilde{\mathbf{s}}(\boldsymbol{\theta})^H \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{s}}(\boldsymbol{\theta})}. \quad (6.24)$$

Although the corresponding test is not strictly equivalent⁷ to the actual GLRT, the performance is almost identical, as will be demonstrated in section 6.3.5. In contrast, ignoring the energy term altogether causes a preference of high energy signals and a disproportionate reduction in detection probability of signals with low energy. The normalization, which is constant for any choice of $\boldsymbol{\theta}$, can be integrated into the signal model $\tilde{\mathbf{s}}$. The advantage of using eq. (6.24) is a small reduction in computational cost for the evaluation of the test statistic, which is an important aspect that is further discussed in section 6.4.

⁷This might seem unexpected given that eqs. (6.18) and (6.20) were shown to be equivalent by scaling and shifting the threshold γ with the data-independent terms. However, the same transformations do not apply to eq. (6.23), as the data-independent terms are not constants within the max operation.

For WGN with $\tilde{\mathbf{C}} = \sigma^2 \mathbf{1}$, the test statistic becomes

$$T(\mathbf{x}) = \max_{\boldsymbol{\theta}} \frac{1}{\sigma^2 \tilde{\mathbf{s}}(\boldsymbol{\theta})^H \tilde{\mathbf{s}}(\boldsymbol{\theta})} \left| \sum_{i=0}^{n_a-1} \sum_{j=0}^{n_s-1} \mathbf{x}_{ij} \mathbf{s}_{ij}(\boldsymbol{\theta})^* \right|^2, \quad (6.25)$$

which serves as the baseline test statistic for the detection of CRES electrons with multiple antenna channels utilizing a matched filter. In this test statistic, $\mathbf{s}_i(\boldsymbol{\theta})$ has a fixed track length of n_s samples and assumes $t_0 = 0$.

To accommodate the unknown start time and signal length, eq. (6.25) will be evaluated applying the strategy of non-overlapping time slices as outlined earlier. This means that \mathbf{x}_i is only an n_s samples long slice of the whole data matrix. Consequently, the parameter vector is $\boldsymbol{\theta} = (E_{\text{kin}}, \alpha_0, \rho, \phi, \varphi_{\text{axial}})$, and it no longer includes the static phase φ_{cycl} , start time t_0 and track length τ .

6.3.3 Signal Model

The previous sections have established the MF as a universal framework for electron detection with optimal detection performance. Yet, it all depends on the use of a well matching model for the signal matrix \mathbf{s} , which is the topic of this section.

Sinusoidal Model

The simplest signal model is once again a sinusoidal model with fixed frequency and amplitude for the electron's radiation. A suitable model for the sinusoidal observation of an antenna at position \mathbf{p}_i was already given as the starting point of the derivation of the beamforming algorithm in eq. (6.1). The model is parameterized by the down-converted frequency ω_m and an electron position \mathbf{p}_k as

$$\mathbf{s}_{ij}(\omega_m, \mathbf{p}_k) = A_i \exp \left(i \left(t_j \omega_m - (\omega_m + \omega_{LO}) \frac{d_i(\mathbf{p}_k)}{c} + \delta_i(\mathbf{p}_k) \right) \right), \quad (6.26)$$

with the corresponding normalization

$$\tilde{\mathbf{s}}(\omega_m, \mathbf{p}_k)^H \tilde{\mathbf{s}}(\omega_m, \mathbf{p}_k) = \sum_{i=0}^{n_a-1} \sum_{j=0}^{n_s-1} A_i^2 = n_s \sum_{i=0}^{n_a-1} A_i^2.$$

The matched filter detector for this model is then constructed by inserting the model into the test statistic of eq. (6.25):

$$T(\mathbf{x}) = \max_{\omega_m, \mathbf{p}_k} \frac{1}{\sigma^2 n_s \sum_{i=0}^{n_a-1} A_i^2} \left| \sum_{i=0}^{n_a-1} \sum_{j=0}^{n_s-1} \mathbf{x}_{ij} A_i e^{-i \left(t_j \omega_m - (\omega_m + \omega_{LO}) \frac{d_i(\mathbf{p}_k)}{c} + \delta_i(\mathbf{p}_k) \right)} \right|^2. \quad (6.27)$$

By using $t_j = \Delta t j$ and restricting the frequency evaluation to discrete values $\omega_m = \Delta \omega m$ with $\Delta \omega = \frac{2\pi}{\Delta t n_s}$, eq. (6.27) can be rearranged to

$$T(\mathbf{x}) = \max_{m, \mathbf{p}_k} \frac{1}{\sigma^2 \sum_{i=0}^{n_a-1} A_i^2} \left| \sum_{i=0}^{n_a-1} A_i e^{-i \left(-(\omega_m + \omega_{LO}) \frac{d_i(\mathbf{p}_k)}{c} + \delta_i(\mathbf{p}_k) \right)} \underbrace{\frac{1}{\sqrt{n_s}} \sum_{j=0}^{n_s-1} \mathbf{x}_{ij} e^{-i \frac{2\pi j m}{n_s}}}_{DFT} \right|^2.$$

In the above equation one can now identify the n_s -point DFT of the data. If the maximization is also restricted to a fixed set or grid of points G , then the test statistic is almost identical to the test statistic of the beamforming algorithm (see eqs. (6.8) and (6.9)):

$$T(\mathbf{x}) = \max_{m \in M, \mathbf{p}_k \in G} \frac{1}{\sigma^2 \sum_{i=0}^{n_a-1} A_i^2} \left| \sum_{i=0}^{n_a-1} A_i e^{-i \left(-(\omega_m + \omega_{LO}) \frac{d_i(\mathbf{p}_k)}{c} + \delta_i(\mathbf{p}_k) \right)} \mathcal{F}(\mathbf{x}_i)_m \right|^2.$$

The main difference at this point is due to the amplitudes, since dropping the variance σ^2 and the square leads to an equivalent detector as a result of the thresholding.

By assuming a signal model with unity amplitude for all channels, the detector is entirely equivalent to beamforming, including the normalization constant $(\sum_{i=0}^{n_a-1} A_i^2)^{-1} = \frac{1}{n_a}$. On the other hand, this result demonstrates that a more general approach to beamforming would actually include amplitude weighting per channel. Without an elaborate model for the amplitude in each channel one option could be a constant amplitude weighted by distance $A_i = \frac{A}{d_i(\mathbf{p}_k)}$, which is motivated by Frii's transmission equation (see eq. (5.7)).

In conclusion, the derivation above proves that the beamforming algorithm from section 6.2 is in fact a matched filter for the detection of a sinusoidal signal with multiple channels at different spatial positions. This finding is not too surprising, given that the optimal detector for a sinusoid in a single channel is the DFT [99]. Thus, beamforming represents a straightforward generalization of this case for multiple channels with a single source.

Nevertheless, this is an important insight that caused a crucial paradigm shift for electron detection with an antenna array at the time of its discovery for several

reasons. First, the results in section 5.5 indicate that, neglecting energy loss, a sinusoidal signal emitted by a source moving on a circle around the symmetry axis (drift motion) is a good approximate model for an electron with $\alpha_0 = 90^\circ$. That means beamforming already results in close to optimal detection performance for 90° electrons and attempts to improve the performance for these electrons with alternative detection algorithms are futile. In addition, this result means that beamforming is an incomplete matched filter for the other electrons. Therefore, this also suggests that the most promising path forward is indeed an improved signal model instead of searches for entirely new detection algorithms or secondary algorithms after beamforming.

Template Bank

The construction of an improved closed-form signal model on the basis of eq. (6.26) is not a trivial task. It could be approached by either using a generic modulated signal — employing a Fourier series with adaptive coefficients $c(\omega)$ that allows for a signal with multiple frequencies — or by tracking the axial motion, similar to the drift motion. As discussed in section 6.2.4, tracking the axial motion is no straightforward calculation anymore due to the complexity of the axial trajectory and the retarded time calculation. In fact, initial efforts to establish a compact closed-form model with axial motion lead to the development of CRESana (see section 5.4), which ultimately evolved into a comprehensive CRES signal simulation tool. Moreover, the diverse set of signal spectra observed in simulations with CRESana, suggests that a universal Fourier series model may not be feasible.

In the absence of a compact closed-form signal model that can be directly implemented into the test statistic, the matched filter can be realized with a *template bank* instead. In this approach, eq. (6.25) is evaluated with a set of pre-calculated waveforms. Matched filter template banks are known from their utilization in gravitational wave detection as part of the data analysis of LIGO [105, 104]. The main challenge for a template bank is a good placement of templates in the parameter space. The template bank needs to cover the interesting regions of the parameter space well in order to ensure good detectability of electrons, while keeping it to a size that is manageable with limited computing resources. This is balanced through a metric that quantifies the SNR loss inflicted by the mismatch of the chosen templates with the signals.

For an application in Project 8 electron detection a large template bank can be generated with the available simulation tools, preferably with CRESana. An initial

study for its application was conducted with waveforms generated by Locust. In this study, it was found that, in order to achieve an average of 95% of the theoretically possible SNR, a template bank size of approximately 10^5 is required [106]. This study was conducted for a dataset that covers only a small 2D sub-region of the possible parameter space with pitch angles of $85.5\text{--}88.5^\circ$ and energies of $18\,575\text{--}18\,580$ eV. The exact result is of course setup dependent and might be compromised by the unresolved issues that were found in Locust in the meantime.

6.3.4 Signal to Noise Ratio

The relevant metric that characterizes the detection performance of all the test statistics is the SNR of the matched filter. Defining the SNR of a random variable X as

$$SNR(X) = \frac{\langle X \rangle^2}{\text{Var}(X)},$$

where $\langle X \rangle$ denotes the expectation value of X and $\text{Var}(X)$ its variance, the SNR of the matched filter is found with $SNR(T(\mathbf{x}; \mathcal{H}_1))$ using eq. (6.14). The required expectation value and variance are

$$\langle T(\mathbf{x}; \mathcal{H}_1) \rangle = \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}$$

$$\text{Var}(T(\mathbf{x}; \mathcal{H}_1)) = \frac{1}{2} \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}.$$

The proof for these results can be found in appendix C.2. Consequently, the SNR of the matched filter is

$$SNR_{MF} = SNR(T(\mathbf{x}; \mathcal{H}_1)) = 2\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}. \quad (6.28)$$

This represents the inherent SNR of the signal, which is the maximum output SNR achievable with any FIR filter applied to the signal. This maximum can only be attained when utilizing the matched filter.

With the assumption of approximately white noise (Nyquist-Johnson thermal noise), in which case the covariance matrix is $\mathbf{C} = \sigma^2 \mathbf{1}$ with $\sigma^2 = k_B T \Delta f R$, the result can be further simplified to

$$SNR_{MF} = \frac{2}{\sigma^2} \mathbf{s}^H \mathbf{s} = \frac{2P_T R n_\tau}{\sigma^2} = \frac{2P_T n_\tau}{k_B T \Delta f}. \quad (6.29)$$

In the equation above P_T denotes the total root mean square (RMS) power of the signal. For a signal \mathbf{s} with n_τ samples and a resistance R the RMS power is defined

as $P_{RMS} = \frac{1}{R n_\tau} \sum_i \mathbf{s}_i^* \mathbf{s}_i$. For the case of a signal $\tilde{\mathbf{s}}$ with n_a antenna channels, the power P_T is the sum of the RMS power of all channels:

$$P_T = \frac{\tilde{\mathbf{s}}^H \tilde{\mathbf{s}}}{R n_\tau} = \frac{1}{R n_\tau} \sum_{j=0}^{n_a-1} \sum_{i=0}^{n_\tau-1} \tilde{\mathbf{s}}_{n_\tau \cdot j+i}^* \tilde{\mathbf{s}}_{n_\tau \cdot j+i} = \sum_{j=0}^{n_a-1} P_{RMSj}.$$

Finally, substituting the signal duration $\tau = n_\tau \Delta t = \frac{n_\tau}{f_s} = \frac{n_\tau}{\Delta f}$ into eq. (6.29) yields:

$$SNR_{MF} = \frac{2\tau P_T}{k_B T}. \quad (6.30)$$

This result allows for a straightforward assessment of the inherent SNR of an electron track in a given experimental setup by assuming a noise temperature and numerically determining the RMS power of a simulated \mathbf{s} from CRESana. Moreover, eq. (6.30) can be used for simple SNR estimations by assuming that the RMS power is constant⁸ with $P_T = \epsilon P$. Here P is the electron's emitted Larmor power and $\epsilon \in [0, 1]$ is an assumed power efficiency of the antenna array for the trajectory. An upper bound for the SNR can be set with $\epsilon = 1$.

Equation (6.30) shows that the inherent SNR only depends on the received signal power P_T , the signal duration τ and the noise temperature T . Notably, it is independent of the signal's shape and distribution between antennas. In addition, it does not depend on the bandwidth Δf but it is implicitly assumed that the signal was recorded with its entire bandwidth and that the filter and signal duration match.

However, for CRES signals the more common case involves applying the filter to non-overlapping time slices of a fixed length of the data time series \mathbf{x} as mentioned in section 6.3.2. For WGN it is possible to calculate how the SNR that determines the detection performance in a single time slice, changes in this scenario. This is done by assuming that the actual signal is contained in the time series \mathbf{s} with a length of n_x and n_τ non-zero samples, while the used filter $\bar{\mathbf{s}}$ differs only in the number $n_{\bar{\tau}}$ of non-zero samples and is zero-padded to match the length n_x of the data.

Analogous to the calculation in appendix C.2 one finds that the corresponding test statistic has the properties $\langle T(\mathbf{x}; \mathcal{H}_1)_{\bar{\mathbf{s}}} \rangle = \frac{1}{\sigma^2} \bar{\mathbf{s}}^H \mathbf{s}$ and $\text{Var}(T(\mathbf{x}; \mathcal{H}_1)_{\bar{\mathbf{s}}}) = \frac{1}{2\sigma^2} \bar{\mathbf{s}}^H \bar{\mathbf{s}}$. This yields for the SNR defined in eq. (6.28):

$$SNR_{\bar{\mathbf{s}}} = \frac{2}{\sigma^2} \frac{(P_T R \min(n_\tau, n_{\bar{\tau}}))^2}{P_T R n_{\bar{\tau}}} = SNR'_{MF} \frac{\min(n_\tau, n_{\bar{\tau}})^2}{n_{\bar{\tau}}}. \quad (6.31)$$

⁸Fluctuations of the emitted and received electron power are negligible if the RMS power is taken over many axial cycles. The slow reduction of emitted power due to the energy loss can be disregarded except for very long τ .

In eq. (6.31) it is assumed that P_T is approximately constant over the full signal length. Hence, $SNR'_{MF} = \frac{SNR_{MF}}{n_\tau}$ is used as the matched filter SNR per time bin. The result indicates that for a fixed time slice length $n_s = n_\tau$ the SNR grows with n_τ^2 until it reaches a constant value when $n_\tau \geq n_s$. The full SNR can be retrieved by combining the results of all time slices.

6.3.5 Detection Performance

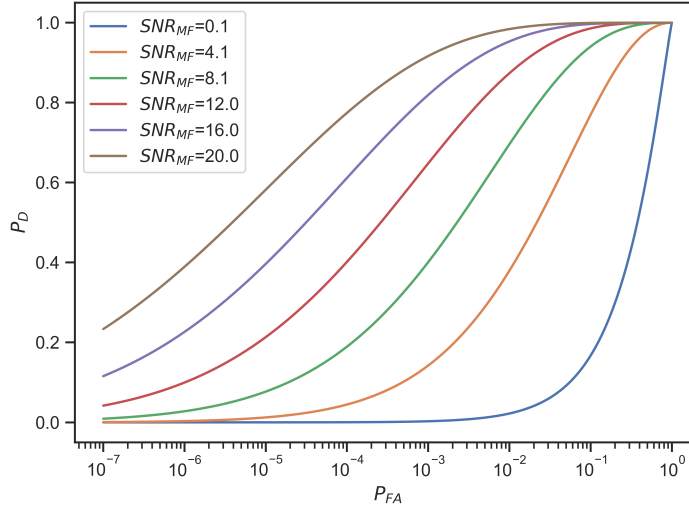


Fig. 6.12.: ROC curves of the matched filter for different SNRs.

The standard method for evaluating the performance of a detector is to use receiver operating characteristic (ROC) curves. The ROC curve plots $P_D(P_{FA})$ and thus one can read off the reachable detection probability for the acceptable probability of false alarm. The detection probability P_D of the matched filter with the test statistic of eq. (6.14) can be shown to be [99]

$$P_D(P_{FA}) = S\left(S^{-1}(P_{FA}) - \sqrt{SNR_{MF}}\right), \quad (6.32)$$

where S is the survival function of the standard normal distribution. As depicted in fig. 6.12, ROC curves from eq. (6.32) are fully determined by the inherent SNR of the signal. Therefore, given an electron track the theoretical ROC curve is found by simulating the noise-less signal in CRESana and subsequently calculating SNR_{MF} using eq. (6.30).

However, employing eq. (6.14) as a detector assumes a true matched filter without unknown parameters. Hence, eq. (6.32) only establishes an upper bound on the

detection performance of any realizable detector, which is necessarily smaller in practice. For a CRES signal the phase will be unknown, which necessitates the use of the quadrature matched filter with eq. (6.20). Its detection probability is [99]

$$P_D(P_{FA}) = S_{\chi_2^2(SNR_{MF})}(-2 \log(P_{FA})) , \quad (6.33)$$

where $S_{\chi_2^2(\lambda)}$ is the survival function of the non-central χ^2 distribution with 2 degrees of freedom and non-centrality parameter λ . The parameter λ is given by SNR_{MF} , which is the same as for the true matched filter.

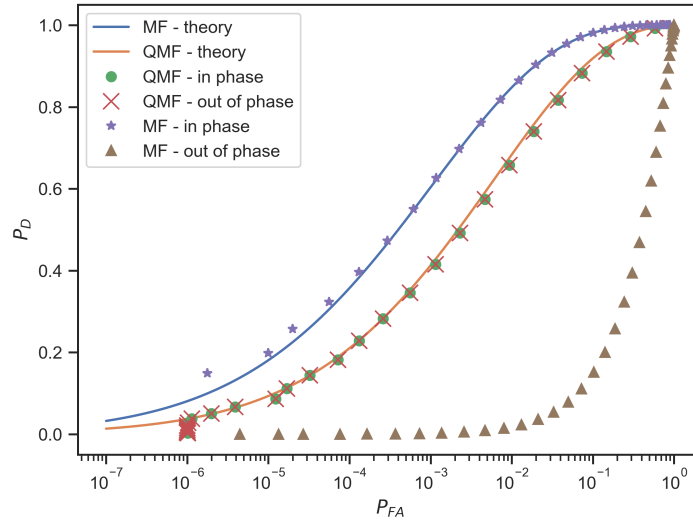


Fig. 6.13.: Theoretical ROC curves (solid lines) of the matched filter (MF) and quadrature matched filter (QMF) compared to results from 10^6 MC experiments (data points). See text for details about the generation of the MC data points.

In fig. 6.13 ROC curves of the quadrature matched filter and the true matched filter are compared to each other both using the theoretical predictions and results from 10^6 Monte Carlo (MC) experiments. CRESana was utilized to generate an antenna array signal s of a $\alpha_0 = 88^\circ$ electron from which $SNR_{MF} = 11$ was found as input for the theory curves according to eqs. (6.32) and (6.33). For the MC experiments WGN was added to s and the results of the test statistics were calculated with eqs. (6.14) and (6.20) in order to generate the ROC curves. In addition, a second set of MC experiments was conducted with a 1.5 rad phase shifted signal s' , while adhering to s for the detection.

The results demonstrate that the use of the quadrature matched filter yields identical detection performance for both cases. In contrast, the matched filter displays slightly better performance when matching the original signal but is inadequate for detecting

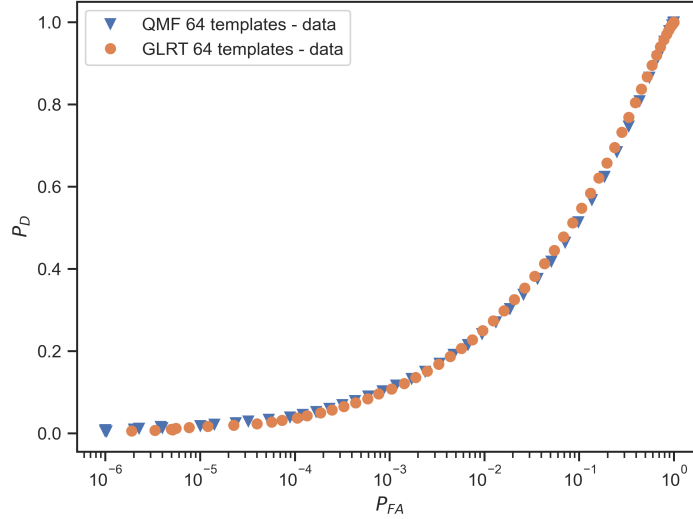


Fig. 6.14.: Experimental ROC curves of template banks using the normalized quadrature matched filter (QMF) and the generalized likelihood ratio test (GLRT) with $n_t = 64$. The data points were obtained through MC experiments. The true signal that is used in the MC experiments is present in the template bank.

the shifted signal. This reduction in detection performance of the matched filter depends on the phase shift.

Finally, the general detection case with all electron parameters unknown is solved with the quadrature matched filter template bank, which employs the test statistic of eq. (6.25). First, it still needs to be proven that this test statistic has a performance that is similar to the actual solution provided by the generalized likelihood ratio test (see eq. (6.23)). This is again done through MC experiments similar to fig. 6.13. Figure 6.14 shows that the MC experiments provide very similar ROC curves for both test statistics.

For the derivation of the theoretical ROC curve of the template bank, it is assumed that the template bank contains the true electron signal and that it is built with n_t nearly orthogonal signals, meaning that for all θ_i, θ_j with $i \neq j$ in the template bank

$$\frac{\tilde{\mathbf{s}}(\theta_i)^H \tilde{\mathbf{s}}(\theta_j)}{\sqrt{\tilde{\mathbf{s}}(\theta_i)^H \tilde{\mathbf{s}}(\theta_i)} \sqrt{\tilde{\mathbf{s}}(\theta_j)^H \tilde{\mathbf{s}}(\theta_j)}} \rightarrow 0. \quad (6.34)$$

This is motivated by the fact that the spectra feature a small number of narrow peaks that are distinct for the parameter combinations (see section 5.5). The assumption in eq. (6.34) then means that the overlap of peaks between different combinations in the template bank is small.

Under these assumptions eq. (6.33) already describes the probability of detecting the true signal using the correct template. Additionally, due to the assumed orthogonality, the probability of detection with an incorrect template is small. Consequently, eq. (6.33) also represents the detection probability of the signal when the full template bank is utilized in the test statistic.

However, P_{FA} denotes the probability of a false alarm from only the true signal template. The test statistic results of all templates are uncorrelated random variables, given the assumption of orthogonality of the templates. Consequently, false alarms from one template are independent of false alarms originating from any other template. The probability of false alarm from any template is then given by

$$P'_{FA} = 1 - \prod_{i=0}^{n_t-1} (1 - P_{FAi}),$$

with P_{FAi} representing the probability of false alarm of the i -th template. Assuming that a common detection threshold for all templates exists such that $P_{FAi} = P_{FA}$ (this should be reasonable with the normalization of eq. (6.25)) gives

$$P'_{FA} = 1 - (1 - P_{FA})^{n_t}. \quad (6.35)$$

Solving for P_{FA} produces

$$P_{FA} = 1 - (1 - P'_{FA})^{\frac{1}{n_t}}.$$

Finally, substituting this into eq. (6.33) yields the detection probability of the template bank

$$P_D(P_{FA}) = S_{\chi^2_2(SNR_{MF})} \left(-2 \log \left(1 - (1 - P_{FA})^{\frac{1}{n_t}} \right) \right), \quad (6.36)$$

where P_{FA} now denotes the probability of false alarm from all n_t templates.

In practice, the template bank will be too densely populated to fulfill the orthogonality criterion in order to minimize the risk of missing signals that may fall between templates. Thus, eq. (6.34) can only be applied for pairs of templates i, j that are sufficiently distant from one another in parameter space. While a precise distance metric is not defined here, it can be assumed, without loss of generality, that the majority of template pairs are adequately separated to justify the orthogonality assumption. However, for a small number of templates that are close to the true template, the detection probabilities will be non-zero and correlated with each other. Similarly, false alarm probabilities of all pairs of templates that are close to each

other, are correlated. Both complicate the exact mathematical treatment. As a result, eq. (6.36) must be treated as an approximation assuming that the detection probability of the true template is the dominant contributor and that the P_{FAi} of each individual template are small.

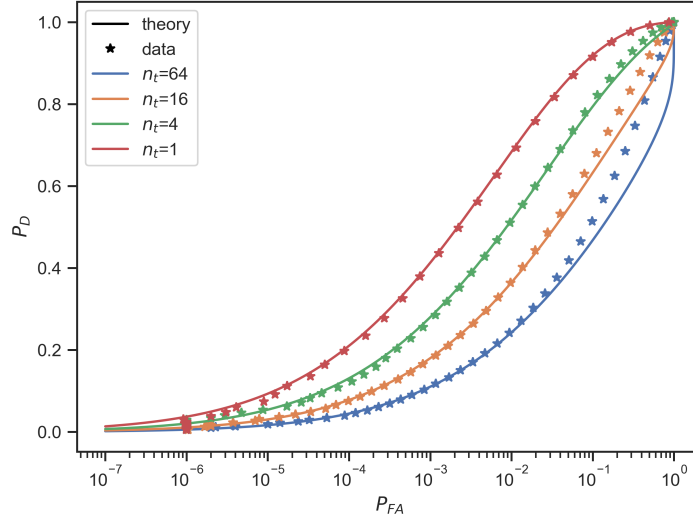


Fig. 6.15.: Theoretical ROC curves of the quadrature matched filter template bank (solid lines) compared to results from 10^6 MC experiments (data points) for different template bank sizes. The true signal that is used in the MC experiments is present in the template bank, which consists of CRESana signal simulations on a regular square grid in E_{kin} and α_0 .

The accuracy of the approximation is tested using MC experiments, from which ROC curves are constructed in the same manner as those in fig. 6.13. The MC experiments are set up such that detection is recorded if any template exceeds the threshold, not necessarily the correct one. Figure 6.15 depicts the results compared to the theoretical curves according to eq. (6.36) for a variety of template numbers. Good agreement is observed at small P_{FA} in all cases. However, some discrepancies appear at high P_{FA} and high n_t . These results are within expectations, considering that orthogonality was not enforced during the template bank construction. Thus, in the regime of high P_{FA} (low thresholds) and with many templates, the probability of any template other than the correct one exceeding the threshold, along with the contributions of correlations to the probabilities, become significant.

Nevertheless, given that the goal of CRES is highly sensitive measurements, such as neutrino mass determination, the relevant region of the ROC curve is at $P_{FA} < 0.01$, where the agreement is good. This suggests that eq. (6.36) is a suitable approximate solution. Yet, it must be noted that this is only the case if the exact signal is part of the template bank and a mismatch diminishes the reachable SNR.

6.4 Real Time Detection

One of the major challenges in an antenna array based CRES experiment are the expected data rates. The data rate of a single I/Q receiver can easily be estimated with $D = 2b f_s$, where b represents the bit depth of the ADC, $2b$ is the memory size of a complex sample, and f_s is the sampling rate. The sampling rate is determined by the required bandwidth Δf , given by $f_s = \Delta f$ (in the case of complex I/Q samples), as prescribed by the sampling theorem [92]. The baseline assumption is a required bandwidth of 200 MHz. Running the ADCs with minimal bit depths of $b = 8$ bit the resulting data rate is $D = 400 \text{ MB s}^{-1}$ per digital receiver. Even the smallest prototype designs under consideration for Phase III utilize 60 independent receivers. Therefore, operating an antenna array for CRES produces at least 24 GB s^{-1} of data in a minimal configuration. However, for a full-scale Phase IV experiment, the number of independent receivers is estimated to be in the thousands (see chapter 8). Thus, assuming no further design optimization that could help reduce the data rate⁹, the generated data can easily exceed 1 TB s^{-1} , particularly if better sampling depths are required.

These data rates quickly exhaust typical data storage capacities available at the time of writing, making data rate reduction critical. This reduction is achieved by recording only data that presumably contain signals and discarding data that presumably contain uninteresting noise. Consequently, this approach necessitates a detection algorithm that runs in real-time and can serve as a trigger for storing the data.

The effectiveness of this data reduction depends on both the average event length and the trigger rate. The trigger rate is determined by the true event rate and the configuration of the detection algorithm, which yields a detection probability and probability of false alarm. All of these parameters in question directly influence the achievable neutrino mass sensitivity. Hence, a meaningful analysis of the reachable data reduction first requires an experimental setup and a configuration of the detection algorithm that are both optimized for neutrino mass sensitivity. Instead, the main topic of this section is therefore an analysis of the real time capabilities of the template bank approach and the computing requirements associated with it.

⁹An optimized design might feature smaller sampling rates by minimizing the signal bandwidth, require fewer high-gain antennas, and reduce the number of digital channels by combining analog signals before digitization.

6.4.1 Algorithmic Choice

The slightly different test statistics of the matched filters presented in section 6.3 all have in common that they are a variant of an inner product that can be interpreted as matrix-vector or matrix-matrix multiplications. For instance, the template bank test statistic in eq. (6.25) can be formulated as the following product of a matrix and a vector

$$T(\mathbf{x}) = \max_k \left| \sum_{j=0}^{n_a n_s - 1} \tilde{\mathbf{S}}_{kj} \tilde{\mathbf{x}}_j \right|^2, \quad (6.37)$$

where \sim denotes flat vectors of the data and signal matrices (see section 6.3.1) and the signal component is a pre-computed template matrix including normalization: $\tilde{\mathbf{S}}_{kj} = \tilde{\mathbf{s}}_j(\boldsymbol{\theta}_k)^* \left(\sigma^2 \tilde{\mathbf{s}}(\boldsymbol{\theta}_k)^H \tilde{\mathbf{s}}(\boldsymbol{\theta}_k) \right)^{-\frac{1}{2}}$. This matrix has the dimensions $n_t \times n_a n_s$ with n_t denoting the total size of the template bank¹⁰.

If data are taken continuously, it is obviously impossible, to indefinitely accumulate the n_s samples for the evaluation of the test statistic in a single pass due to memory limitations. For that reason, the evaluation has to proceed in batches of fixed size n_s . For a time T_c of continuous data production the number of batches n_B is defined via the relation

$$T_c f_s = n_s n_B. \quad (6.38)$$

With the execution in batches eq. (6.37) can be rearranged as

$$T(\tilde{\mathbf{X}})_l = \max_k \mathbf{R}_{kl} = \max_k \left| \sum_{j=0}^{n_a n_s - 1} \tilde{\mathbf{S}}_{kj} \tilde{\mathbf{X}}_{jl} \right|^2.$$

Thus, the test statistic calculates the matrix product of $\tilde{\mathbf{S}} \in \mathbb{C}^{n_t \times n_a n_s}$ and $\tilde{\mathbf{X}} \in \mathbb{C}^{n_a n_s \times n_B}$, producing a matrix $\mathbf{R} \in \mathbb{R}^{n_t \times n_B}$.

The formulation as a matrix-matrix product has a number of appealing advantages. First, it is a highly parallel operation, which suggests the execution on a multi-core computing platform. It is a very well established algorithm that often features low-level implementations with hardware-specific optimizations for the platforms it runs on with easy access via the Basic Linear Algebra Subprograms (BLAS) [107] standard. BLAS implementations exist for most high-performance platforms, e.g. Intel MKL [108] for CPUs or cuBLAS [109] for GPUs. Matrix multiplications also

¹⁰To clarify: All matrices here are, in fact, high dimensional tensors. For example, the template bank can be up to 8D when considering the entire parameter space of CRES. Each template is then itself a multi-channel $n_s \times n_a$ matrix. For the algorithmic treatment, these higher dimensions are effectively disregarded, and all can be flattened out in the way that is most convenient for computation.

serve as fundamental building blocks for Deep Learning, a field which profits immensely from their implementation on GPUs [110, 111]. Finally, the more basic matrix-vector products like eq. (6.37) are always limited by memory access times because all data are accessed exactly once. Matrix-matrix products, on the other hand, utilize certain segments multiple times and can thus be limited by either memory or arithmetic operations, depending on the dimensions of the matrices. Given that memory access times are typically considerably slower than the arithmetic operations on modern computing platforms (see e.g. [112]), it is preferable to operate within the arithmetic limit.

With all these considerations, it is a reasonable assumption that the template bank should be implemented on a multi-GPU server. Each GPU then evaluates the matrix product with the full set of n_t templates, processing n_b batches at a time, where $n_b < n_B$, such that the total number of batches n_B is distributed equally among the GPUs. Therefore, the number of GPUs must be sufficient to ensure that the computations keep up with the incoming data.

6.4.2 Estimation of the Computing Requirements

The objective of this section is to estimate the required number of GPUs for a real time implementation of the template bank approach, inspired by the methodology outlined in [113]. For the estimation it will be assumed that Nvidia's most recent flagship server GPU H100 will be used with its ability to efficiently process 16 bit floats on tensor cores. For the initial real-time detection, accuracy is not critical, making the use of 16 bit floats adequate. In this case the H100 GPU has a peak computing performance of¹¹ $P_{FP16} = 989.4 \text{ TFLOP s}^{-1}$, a peak memory access rate of $M_a = 3.35 \text{ TB s}^{-1}$ and a total memory size of $M = 80 \text{ GB}$ [112].

The evaluation of the full matrix \mathbf{R} requires $n_B \cdot n_t \cdot n_a \cdot n_s$ multiplications and additions of complex numbers for the matrix product. Since the product $z_1 \cdot z_2 = x_1x_2 - y_1y_2 + i(y_1x_2 + x_1y_2)$ for $z_i = (x_i + iy_i) \in \mathbb{C}$, this corresponds to a total of $4n_B \cdot n_t \cdot n_a \cdot n_s$ multiplications and additions of real numbers. On the hardware level these would be handled by fused multiply-add (FMA) instructions, which each reflect 2 floating point operations. After the matrix product, each element in the matrix \mathbf{R} is produced by taking the norm-squared. Given that $|z|^2 = x^2 + y^2$, this operation

¹¹Inputs and outputs are in 16 bit but the accumulation uses 32 bit.

requires $3n_B \cdot n_t$ floating point operations for the entire matrix \mathbf{R} . Therefore, the operation count of the calculation is

$$OP = n_B n_t (8n_a n_s + 3).$$

Using eq. (6.38) and assuming that the contribution of the norm-squared can be neglected since n_a and n_s are large, produces

$$OP = 8n_t n_a T_c f_s. \quad (6.39)$$

Assuming that the calculation is not limited by memory accesses, the operation count processed by N_{GPU} GPUs in a time T_c is $N_{GPU} P_{FP16} T_c$. Comparing this with eq. (6.39) yields

$$N_{GPU} = \frac{8n_t n_a f_s}{P_{FP16}}. \quad (6.40)$$

Whether or not a matrix-matrix product is limited by memory access or arithmetic operations depends on the sizes of the matrices relative to each other. These sizes are solely decided by the chosen batch size n_s . Possible values for n_s are restricted by the available amount of memory on a single GPU. Each value of the complex input matrices is comprised of 2 16 bit floats, hence 4 B of memory are required per value. The real output matrix on the other hand requires 2 B per value. Thus, the restriction for n_s set by the available memory is given by

$$4n_b n_a n_s + 4n_t n_a n_s + 2n_t n_b \leq M. \quad (6.41)$$

A second restriction for n_s is found by requiring that the calculation is not limited by memory access times. This can be examined by calculating the arithmetic intensity, which is the ratio of the operation count to the total memory accessed:

$$\begin{aligned} AI &= \frac{8n_b n_t n_a n_s}{4n_b n_a n_s + 4n_t n_a n_s + 2n_t n_b} \\ &= \frac{4n_t n_a}{n_a + n_t \left(\frac{n_a}{n_b} + \frac{1}{2n_s} \right)}. \end{aligned}$$

The calculation is not limited by memory accesses if the arithmetic intensity is greater than the ratio of peak computing performance to peak memory access rate:

$$AI \geq \frac{P_{FP16}}{M_a}. \quad (6.42)$$

This kind of treatment neglects the role of the on-chip memory caches, which can be safely ignored when the calculation is limited by arithmetic operations.

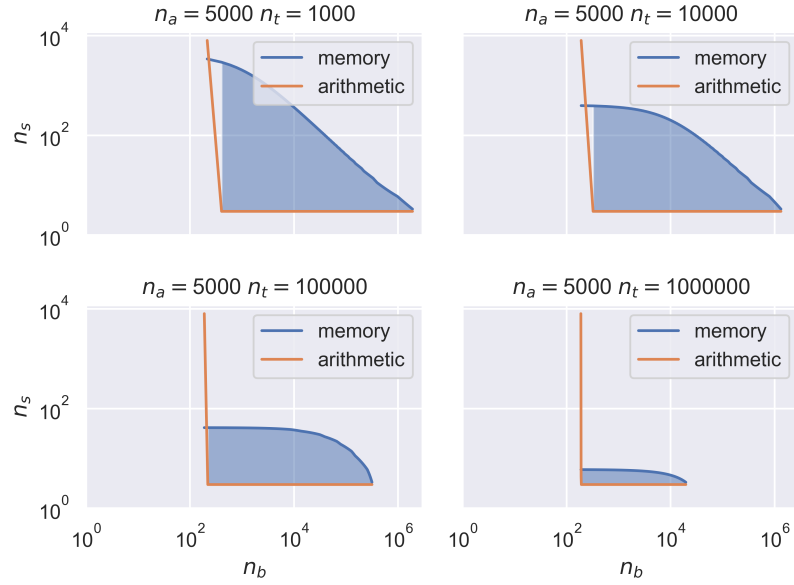


Fig. 6.16.: Arithmetic bound and memory size bound depending on n_s and n_b . The orange line shows the boundary provided by eq. (6.42). The blue line indicates the memory limit given by eq. (6.41). Any combination of n_s and n_b in the shaded area is both in the faster arithmetic limit and fits onto the GPU memory.

Figure 6.16 shows how the conditions eqs. (6.41) and (6.42) restrict the choice of n_s and n_b for given n_a and n_t . As long as the blue area is non-zero, it is generally possible to pick the batch size n_s such that the matrix multiplication is feasible. The plots demonstrate that with 5000 antennas and up to 1000000 templates, it is appropriate to use eq. (6.40) as an estimate for the number of GPUs.

Figure 6.17 shows results of the calculation using eq. (6.40), varying both the number of templates and the antenna array size. In addition, the dashed black line represents the equivalent number of H100 GPUs of the computing performance of Frontier, the current fastest supercomputer on the TOP500 list [114] reaching $1.194 \text{ EFLOP s}^{-1}$ (E=Exa) for 64 bit operations.

These results show that small scale prototype experiments with ~ 100 antennas and template bank sizes up to 10^6 might be viable with a medium scale GPU cluster, but the computational demands required to meet Phase IV design requirements appear unattainable at this stage. One should keep in mind, however, that a potential Phase IV experiment is still many years in the future and gains in available computing power are expected. On the other hand, it should be emphasized that this result is only a basic estimate that disregards many practical aspects of a real implementation. For instance, this basic treatment assumes all data is already available for calculations

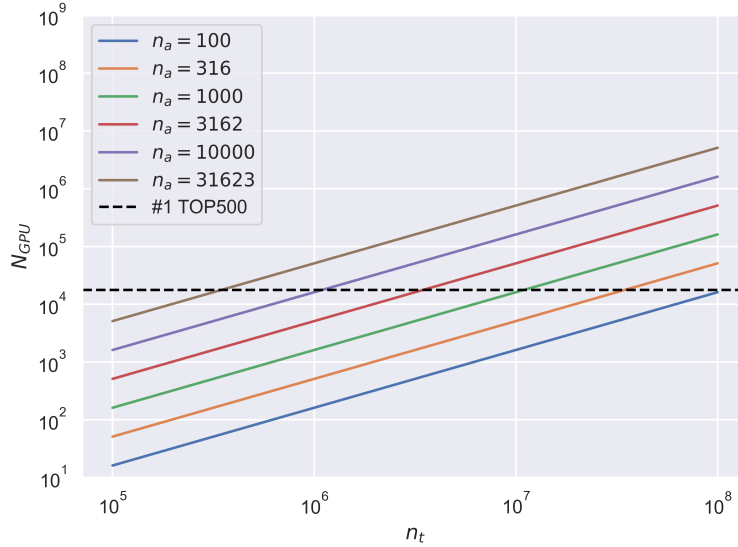


Fig. 6.17.: Estimated number of required Nvidia H100 GPUs for an MF template bank as a function of the number of templates n_t . The estimate is plotted for different assumed number of antennas n_a . The dashed black line represents the equivalent of the current most powerful supercomputer in the TOP500 [114] in number of H100 GPUs.

on the GPUs and entirely neglects data transfers to and from the GPUs. As always, a realistic assessment can only be obtained through actual benchmarks.

6.4.3 Potential Improvements of Computational Performance

There may be ways to reduce the computing requirements for a real time detection algorithm, which will be discussed briefly in this section.

First, a significant efficiency improvement of the template bank evaluation may be possible by using Parseval's theorem [115] in order to calculate the inner products in the frequency domain instead of the time domain. As shown in section 5.5, the signal spectra feature only few discrete peaks. This implies that the template matrix \tilde{S} would largely consist of entries close to zero. Therefore, significant performance gains could be possible through the use of sparse matrix algebra. Nevertheless, this potential performance gain entirely depends on the signal structure obtained for an experimental design. This gain is purely theoretical at this point and needs to be studied in detail.

A critical parameter for the computational cost of the template bank is its size, which depends on the tolerable SNR loss from not exactly matching the true signal with

a discrete set of templates. An early indication for the required size is provided in [106], which found a lower limit of 10^5 . However, this value likely varies with the experiment design. Additionally, the study covers only a subset of the full parameter space with templates, and it remains unclear how much the size increases as more dimensions are included.

Moreover, the tolerable SNR loss of 5% was chosen arbitrarily, whereas in a real implementation, it must be determined by optimizing the detection probability and probability of false alarm for a neutrino mass sensitivity goal. With optimized settings, the acceptable SNR loss could be higher, leading to a reduction in template bank size. For example, allowing a 30% SNR loss in [106] reduces the size of the template bank already by an order of magnitude. In conclusion, future work must establish a formal procedure for selecting a grid within the 8-dimensional parameter space of electron tracks that strikes the right balance between detection performance and computational cost. A sophisticated solution may produce non-uniform grids with varying template densities depending on the significance of different regions in parameter space.

Finally, machine learning models offer a promising path toward real time detection, given their widespread success in data analysis in modern physics [116, 117, 118]. Especially convolutional neural networks (CNNs) are interesting candidates for signal detection since their convolution operation is closely related to the concept of matched filters. A trained CNN can be interpreted as a hierarchical set of filters that are sensitive to certain features in the data [119]. A minimal set of optimal filters, encoded in the weights of the network, is found in the training phase. It has been demonstrated for LIGO that a CNN can achieve a detection performance comparable to a template bank while requiring only a fraction of its computational cost [120, 121]. In a first attempt using a similar CNN approach on simulated CRES data, a detection performance comparable to a template bank with 17% SNR loss was achieved [106]. Thus, these initial efforts should be intensified in the future.

6.5 Summary

In this chapter it was demonstrated that the optimal signal detection algorithm is accomplished with the generalized likelihood ratio test and matched filters (MFs). Moreover, it was found with a mathematical proof that the beamforming algorithm, which was the first detection algorithm under consideration for an antenna array detector, is a matched filter for a simplified signal model. Since this assumed

model only adequately describes a subset of the parameter space, this algorithm is inherently limited in its capabilities for detecting CRES signals. The theoretical limits of the detection performance with a matched filter were addressed in section 6.3.5. These limits only depend on SNR_{MF} , which can be calculated from simulated data. The relevant question for any realizable solution would be how close it approaches these optimal limits. Finally, it was shown in section 6.4.2 that an MF based detection approach is computationally infeasible at this moment, given the required large template bank and antenna array sizes. Suggestions for further research into computational performance optimizations were discussed in section 6.4.3.

Parameter Estimation of CRES Signals

“ Before I came here I was confused about this subject. Having listened to your lecture I am still confused. But on a higher level.

— **Enrico Fermi**

Pioneer in nuclear physics

In the previous chapter, the problem of detecting electrons was discussed. While the detection of an electron is a necessary first step, the goal of CRES is to measure the electron's energy. This problem falls into the category of parameter estimation, which is the task of determining the parameters θ that describe the probability distribution $p(\mathbf{x}; \theta)$ of observed data \mathbf{x} . For the parameter estimation of CRES electron tracks it is assumed that the observed data contain a signal and therefore the PDF in question is always under \mathcal{H}_1 as defined in section 6.1. The unknown parameters are $\theta = (E_{\text{kin}}, \alpha, \rho, \varphi, \phi_{\text{axial}}, \phi_{\text{cycl}}, t_0, \tau)$. Among these the only parameter of interest is the kinetic energy E_{kin} , whereas the remaining parameters are nuisance parameters.

The detection and parameter estimation problems share many mathematical similarities, as demonstrated in section 6.3.2, where it was shown that the detection of a signal with unknown parameters requires at least a rough parameter estimate. Nevertheless, it is worth discussing the parameter estimation problem separately with the focus on the statistical estimation limits.

7.1 Maximum Likelihood Estimation

A standard estimation approach is the maximum likelihood estimate (MLE), which is proven to have asymptotic efficiency [122], which means that for large data records it has minimum variance and is unbiased. With these properties the MLE is considered asymptotically optimal, since no other unbiased estimator has a lower

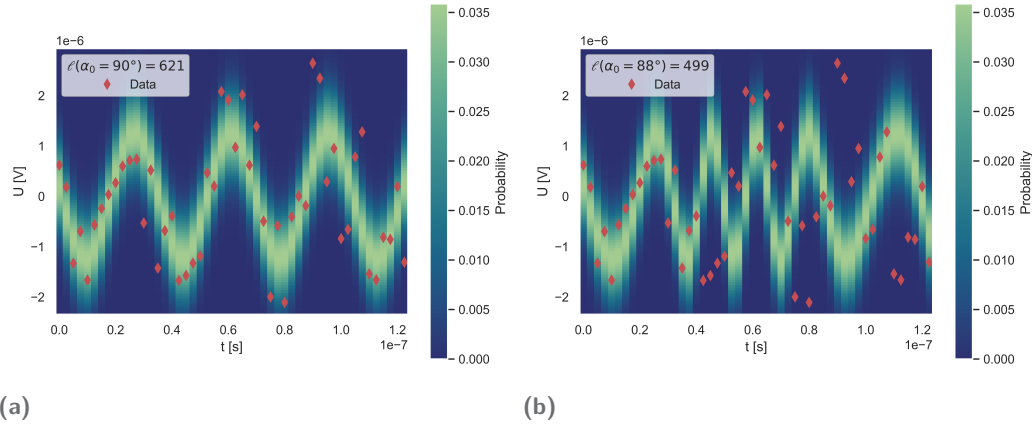


Fig. 7.1.: Likelihood calculation at a single point θ for a high SNR CRESana signal from a harmonic trap with added WGN. The plots show the first 50 data points of a single channel as orange points, with probability distributions based on the assumed θ represented by the underlying color scale. The distributions are evaluated at the data points. (a) illustrates the distributions for the true signal, while (b) depicts those for a modified pitch angle.

variance. However, an unbiased estimator with the same performance without the need of large data records may exist. However, even if such an estimator exists, there is no universal recipe that finds it in a given estimation problem. In contrast, the MLE offers a universal solution and is thus always the preferred approach in practical applications, where datasets are often sufficiently large.

The key concept is the likelihood function, which is the joint probability of the individual data observations viewed as a function of the parameters θ . In a CRES track, the observations are generated by the deterministic signal, which is parameterized by θ and subjected to additive random noise from a process that is similar across all observations. Therefore, all observations are random variables from the same family of distributions with PDF $p(x; \lambda)$ but with varying parametrization¹ λ that account for the different amplitudes of the underlying deterministic signal.

Thus, $p(x; \lambda_{ij}(\theta))$ denotes the PDFs of the individual observations x_{ij} of channel i and time sample j . This results in the likelihood function

$$\mathcal{L}(\theta; \mathbf{x}) = \prod_{i=0}^{n_a-1} \prod_{j=0}^{n_s-1} p(\mathbf{x}_{ij}; \lambda_{ij}(\theta)). \quad (7.1)$$

¹E.g. for a normal distribution the parametrization λ is given by its mean and variance.

In eq. (7.1) it is assumed that the observations are independent random variables, which is only true for uncorrelated noise. The calculation of $\mathcal{L}(\boldsymbol{\theta})$ with the varying PDFs is illustrated in fig. 7.1 for two different choices of $\boldsymbol{\theta}$ with the same data \mathbf{x} .

Numerically it is often more practical to use the log-likelihood function $\ell(\boldsymbol{\theta}; \mathbf{x}) = \log(\mathcal{L}(\boldsymbol{\theta}; \mathbf{x}))$ instead. The MLE of $\boldsymbol{\theta}$ is found by maximization of the likelihood, which is equivalent to maximizing the log-likelihood due to the monotonicity of the logarithm:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \mathbf{x}).$$

Once again assuming that the thermal noise in a CRES experiment is approximately complex WGN with variance σ^2 given by eq. (5.15), the observations $\mathbf{x}_{ij} = \mathbf{s}_{ij} + \mathbf{n}_{ij}$ are complex normal random variables with $\mathbf{x}_{ij} \sim \mathcal{CN}(\mathbf{s}_{ij}, \sigma^2)$. Hence, the PDFs for the likelihood function are² $p(\mathbf{x}_{ij}; \boldsymbol{\lambda}_{ij}(\boldsymbol{\theta})) = \frac{1}{\pi\sigma^2} e^{-\frac{1}{\sigma^2} |\mathbf{x}_{ij} - \mathbf{s}_{ij}(\boldsymbol{\theta})|^2}$ yielding the log-likelihood function

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = -\frac{1}{\sigma^2} \sum_{i=0}^{n_a-1} \sum_{j=0}^{n_s-1} |\mathbf{x}_{ij} - \mathbf{s}_{ij}(\boldsymbol{\theta})|^2 + \text{const}. \quad (7.2)$$

In theory the problem of determining the electron energy E_{kin} from the data is already solved by this approach in the best possible way. In practice this optimal solution suffers from issues that are similar to those encountered in the previous chapter. As before, an accurate signal model is required, which can only be provided by the simulation tool CRESana at this point. However, a simple grid search with pre-computed simulation data, similar to the template bank, is inadequate for the high resolution goals for the estimate of E_{kin} . Furthermore, the search for the global maximum is complicated since it is located in a narrow peak in an otherwise quasi-flat likelihood landscape with many side-maxima (Examples for this can be found in chapter 8, e.g. fig. 8.1). This is a challenging situation for typical gradient based optimization algorithms. Therefore, for a viable optimization strategy a good initial guess close to the global maximum is mandatory. This initial guess may be obtained from the signal detection stage with the template bank³.

²For a circularly symmetric complex random vector $\mathbf{x} \in \mathbb{C}^n$ the PDF is $p(\mathbf{x}) = \frac{1}{\pi^n \det(\mathbf{C})} \exp(-(\mathbf{x} - \boldsymbol{\mu})^H \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}))$ with location parameter $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} [99]. For WGN $\mathbf{C} = \sigma^2 \mathbf{1}$. Thus, the PDF takes the form: $p(\mathbf{x}) = \frac{1}{\pi^n \sigma^{2n}} \exp(-\frac{1}{\sigma^2} |\mathbf{x} - \boldsymbol{\mu}|^2)$

³These considerations highlight that the distinction between detection and parameter estimation in the implementation may be quite loose, despite being separate mathematical problems.

7.1.1 Uncertainties

A topic of high importance for a physics experiment is the treatment of statistical uncertainties of the measured quantities. For the MLE this means that for each estimated parameter θ_i a confidence interval $CI_\alpha = [c_a, c_b]$ for a specified confidence level α needs to be determined. The associated uncertainties are the distances between the estimate and the confidence interval bounds, i.e. $\Delta\theta_{i+} = c_b - \hat{\theta}_i$ and $\Delta\theta_{i-} = \hat{\theta}_i - c_a$ for the general asymmetric case. In most cases the uncertainties are stated for a confidence level of 68%, which is the $1-\sigma$ level of a 1-dimensional Gaussian distribution. There are several ways of constructing confidence intervals for the MLE. In this thesis, two methods are employed, yielding what are referred to as the *Minos* error and *Hesse* error, respectively. These names are consistent with the terminology used in the minimization software MINUIT [123, 124].

Minos Error

The Minos error is based on Wilks' theorem [125], which can be utilized in a way that assigns confidence levels to the values of the log-likelihood function. For this method it is important to divide the model parameters θ into the n parameters of interest ψ and nuisance parameters η . The MLE is then $\hat{\theta} = (\hat{\psi}, \hat{\eta})$ and the confidence region is given by [126, 127]

$$CI_\alpha(\hat{\psi}) = \left\{ \psi \mid \ell((\psi, \hat{\eta}(\psi)); \mathbf{x}) - \ell((\hat{\psi}, \hat{\eta}); \mathbf{x}) \geq -\frac{1}{2}Q_{\chi_n^2}(\alpha) \right\}, \quad (7.3)$$

where $Q_{\chi_n^2}(\alpha)$ is the quantile function of the χ^2 distribution with n degrees of freedom. $\ell((\psi, \hat{\eta}(\psi)))$ is the n -dimensional profile likelihood, which is found by maximizing the likelihood over the nuisance parameters η for a fixed ψ .

In words this means, with n parameters of interest, the boundary for an α confidence level of the estimated parameters is found at the iso-surface where the profile likelihood decreases by $\frac{1}{2}$ times the appropriate value of the quantile function from its maximum value. In general, although undesirable, this procedure is also capable of producing multiple disconnected confidence regions if the log-likelihood of side maxima exceeds the threshold of the confidence level [127]. This is illustrated in fig. 7.4.

For the Minos error $\Delta\theta_{i\pm}$ of a parameter estimate $\hat{\theta}_i$, it is assumed that θ_i is the only parameter of interest, and the corresponding 1-dimensional confidence interval containing the global maximum is calculated via eq. (7.3). The Minos error is

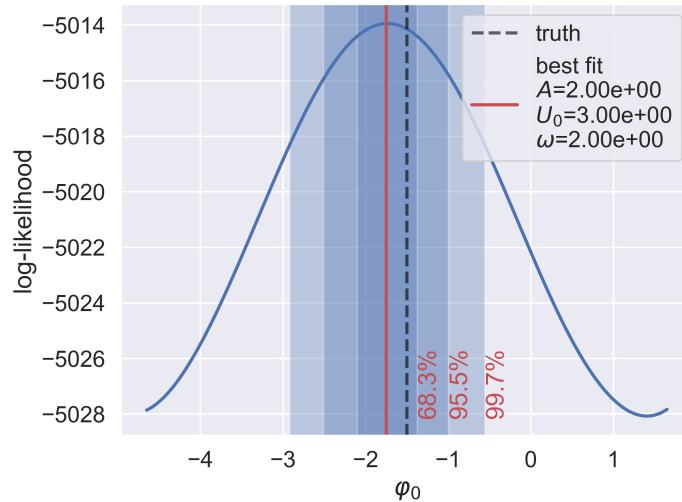


Fig. 7.2.: Log-likelihood function of a 1-D estimation problem with Gaussian noise. Solid red line marks the MLE and the dashed black line the true value, which is different due to the noise. The 68.3%, 95.5% and 99.7% confidence intervals are marked in different shades of blue.

then the (generally asymmetric) distance between the MLE and the bounds of this confidence interval.

An example of the calculation for a 1-dimensional estimation problem is depicted in fig. 7.2. Single channel data with added WGN from a model $s_j = A \sin(\omega t_j + \varphi_0) + U_0$ were generated. For the estimation it was assumed that all parameters are known with their true value, except for ω . $\ell(\omega)$ was calculated and the MLE $\hat{\omega}$ is found at its maximum (solid red line). The marked confidence intervals are found by calculating the difference $\ell(\omega) - \ell(\hat{\omega})$ (y-axis in fig. 7.2) while scanning ω left and right from the maximum until the difference exceeds the threshold set by the quantile function.

Figure 7.3 illustrates the calculation for a 2-dimensional estimation problem using the same model and data as before, but assuming that both ω and φ_0 are unknown. If both parameters are regarded as parameters of interest without additional unknown nuisance parameters, then eq. (7.3) produces isolines of the 2-dimensional log-likelihood function, which are determined with the same scan-and-thresholding approach from before. This is seen in fig. 7.3a where the interiors of the isolines mark the confidence regions of the joint estimation for the corresponding confidence levels.

However, if ω is the only parameter of interest in this scenario with φ_0 unknown, eq. (7.3) prescribes to perform the scan-and-threshold approach on the 1-dimensional profile likelihood. The profile likelihood is the result of maximization over φ_0 for

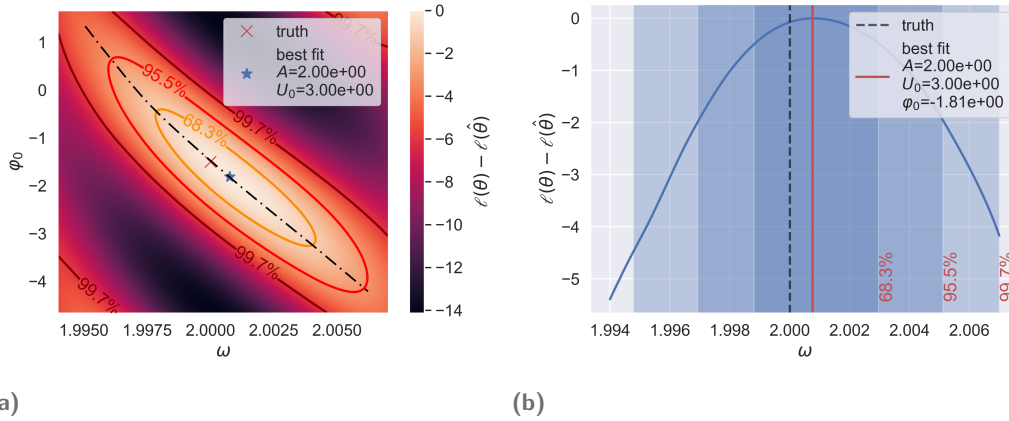


Fig. 7.3.: Calculation of confidence regions in a 2-dimensional estimation problem using the same data as fig. 7.2. (a) shows isolines indicating joint parameter estimation confidence regions. The star marks the MLE, the cross the true parameter values. (b) displays the profile likelihood obtained via maximization over φ_0 , which returns the likelihood values along the dash-dotted line in (a).

each ω , which yields the values of fig. 7.3a along the dash-dotted line. The result is depicted in fig. 7.3b. Compared to the case of the 1-dimensional estimation from fig. 7.2 the confidence intervals are much wider, which is the result of the correlation of ω and the unknown φ_0 .

It should be noted that Wilks' theorem only applies if a number of regularity conditions are met. While in practice the Wilks' confidence intervals often still provide good approximations, violations of these conditions can affect their accuracy. A detailed review of these conditions and possible solutions for failure cases can be found in [128].

Hesse Error

In the asymptotic limit the MLE $\hat{\theta}$ is an n -dimensional normal random variable [122] with $\mu = \theta_{\text{true}}$ and covariance matrix \mathbf{C} . An estimate of the covariance matrix can be found via the inverse Hessian matrix of the log-likelihood $\mathbf{C} \approx \mathbf{H}_{\ell(\theta; \mathbf{x})}^{-1}$ [129], which can be approximated numerically by finite differences of the log-likelihood at the estimated value⁴ $\hat{\theta}$. The covariance matrix contains the parameter variances in the diagonal entries as well as information about parameter correlations in the off-diagonal entries and it determines an n -dimensional error ellipsoid. Finally, the Hesse error for α confidence of a parameter estimate $\hat{\theta}_i$ is given by its standard

⁴Since the log-likelihood is a quadratic function under the normal distribution, the Hessian matrix is a constant. Thus, evaluation at $\hat{\theta}$ yields the same result as at θ_{true} .

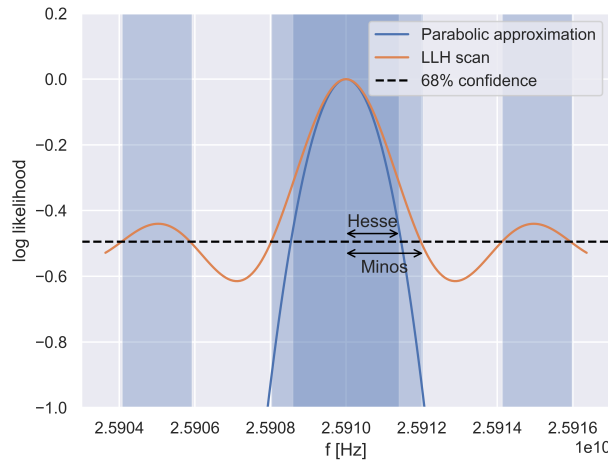


Fig. 7.4.: Comparison of Hesse and Minos errors for 68% confidence of sinusoidal frequency estimation. Dark (light) blue regions mark Hessian (Wilk's) confidence intervals. In this case the Wilks' confidence intervals are disconnected due to side maxima. Minos errors only describe the confidence interval around the global maximum.

deviation multiplied by the appropriate value of the standard normal quantile function, i.e. $\Delta\theta_i = Q_{\mathcal{N}(0,1)}\left(\frac{\alpha+1}{2}\right)\sqrt{C_{ii}}$.

Computationally this approach is comparatively fast, since it only requires the numeric calculation of the Hessian matrix, whereas the Minos error performs an expensive numeric optimization for each point ψ of the profile likelihood $\ell((\psi, \hat{\eta}(\psi)); \mathbf{x})$. Yet, the downside of this approach is that it is limited to cases where the amount of data is sufficient such that the estimation problem is in the asymptotic normal limit, where it is equivalent to the Minos result. Therefore, it is only suited for symmetric log-likelihood functions that are well approximated by a parabolic function near the maximum. The difference between the Hesse and Minos errors is illustrated in fig. 7.4.

7.2 Determination of Estimation Limits

In section 7.1, it was outlined that, similar to the detection problem, the optimal solution for estimation is provided by statistical theory and is easily adapted to CRES. However, the real challenge for a full-fledged implementation with real data lies in the details of the optimization strategy, which is outside the scope of this thesis. Instead, the remaining parts of this chapter prioritize the calculation of estimation limits from the information that is available in the signal with the assumed optimal

solution, while neglecting how hard it may be to recover that information from the signal without prior knowledge. This is still important given that an essential design criterion for a CRES experiment is its energy resolution $\Delta E_{\text{kin}}(\boldsymbol{\theta}_{\text{true}})$, which represents the spread of estimates \hat{E}_{kin} obtained from repeated measurements of the same electron with true electron parameters $\boldsymbol{\theta}_{\text{true}}$.

7.2.1 Analytic Approach: Cramér–Rao Lower Bound

From the theoretical standpoint the limits of parameter estimation are given by the Cramér–Rao lower bound (CRLB), which states that the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}$ must satisfy [122]

$$\text{Var}(\hat{\boldsymbol{\theta}}_i) \geq \mathbf{I}^{-1}(\boldsymbol{\theta})_{ii}. \quad (7.4)$$

In eq. (7.4) $\mathbf{I}^{-1}(\boldsymbol{\theta})$ denotes the inverse of the fisher information matrix, defined via second derivatives of the log-likelihood function as

$$\mathbf{I}(\boldsymbol{\theta})_{ij} = - \left\langle \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \theta_i \partial \theta_j} \right\rangle.$$

The CRLB provides a practical framework for the calculation of a theoretical energy resolution limit of a CRES experiment using the log-likelihood from eq. (7.2). However, once again this relies on a suitable model for the signal s .

A discussion of the CRLB assuming a linear chirp model for the electron signal is provided in [130] in detail⁵. Unlike the constant frequency sinusoidal model used for beamforming, the linear chirp takes into account the energy loss through a linearly increasing frequency, and the finite signal length with unknown start time. The model is given by

$$s_j = A g(t_j - t_0) e^{i(\omega_0(t_j - t_0) + \frac{\delta\omega}{2}(t_j - t_0)^2 + \varphi_0)}, \quad (7.5)$$

where $g(t)$ is the window function describing the transition from no signal to full amplitude and vice versa.

⁵Unfortunately, the final result given in the original calculations in [130] was found to be incorrect. The correct result is given here with eqs. (7.7) and (7.9).

In an idealized scenario the response of the antenna to a signal is instantaneous, resulting in the rectangular window function

$$g(t) = \begin{cases} 1 & t \in [0, \tau] \\ 0 & \text{else} \end{cases} .$$

Simulated signals from CRESana are always produced with a rectangular window due to the approximation of the transfer function with a constant gain. When properly implemented instead, the transfer function acts as a band-pass filter that creates a smooth window function, which corresponds to the non-instantaneous signal startup. Even so, the rectangular window should be a suitable approximation in most cases, given that the observation of the non-instantaneous signal startup phase of an antenna in real data is not expected, as antennas respond much faster than the sampling time step.

Assuming a fixed signal length τ , the unknown parameters are amplitude A , initial frequency ω_0 , frequency slope $\delta\omega$, initial phase φ_0 and start time t_0 . In [130] the following expression is given as CRLB for the estimate of the start frequency:

$$\text{Var}(\hat{\omega}_0) \geq \frac{\sigma^2}{2A^2} \left(\frac{\delta\omega^2}{\dot{M}} + \frac{M_0M_4 - M_2^2}{M_0M_2M_4 + 2M_1M_2M_3 - M_2^3 - M_1^2M_4 - M_3^2M_0} \right), \quad (7.6)$$

where $M_k = \sum_j t_j^k g^2(t_j)$ and $\dot{M} = \sum_j \dot{g}^2(t_j)$. Assuming the idealized rectangular window with $M_k = \sum_{j=0}^{n_\tau-1} \Delta t^k j^k$, and using Wolfram Mathematica [89] to evaluate eq. (7.6), produces

$$\text{Var}(\hat{\omega}_0) \geq \frac{\sigma^2}{2A^2} \left(\frac{\delta\omega^2}{\dot{M}} + \frac{12(8n_\tau - 11)(2n_\tau - 1)}{\Delta t^2 n_\tau (n_\tau^4 - 5n_\tau^2 + 4)} \right), \quad (7.7)$$

where $n_\tau = \lfloor \frac{\tau}{\Delta t} \rfloor$ is the number of signal samples. In leading order of n_τ , this simplifies to

$$\text{Var}(\hat{\omega}_0) \gtrsim \frac{\sigma^2}{2A^2} \frac{\delta\omega^2}{\dot{M}} + 96 \frac{\sigma^2}{\Delta t^2 A^2 n_\tau^3}. \quad (7.8)$$

Substituting the CRLB for the start time estimate, given in [130] as

$$\text{Var}(\hat{t}_0) \geq \frac{\sigma^2}{2A^2} \frac{1}{\dot{M}},$$

and $n_\tau = \lfloor \frac{\tau}{\Delta t} \rfloor$ into eq. (7.8), results in

$$\text{Var}(\hat{\omega}_0) \geq \delta\omega^2 \text{Var}(\hat{t}_0) + 96 \frac{\Delta t \sigma^2}{A^2 \tau^3}. \quad (7.9)$$

Therefore, $\text{Var}(\hat{\omega}_0)$ decreases for long tracks with a τ^{-3} scaling, but is ultimately limited by the start time estimate at $\delta\omega^2 \text{Var}(t_0)$ due to the linear time dependence of the frequency.

Finally, defining

$$\eta = \frac{SNR_{MF}}{\tau} = \frac{2P_T}{k_B T} \quad (7.10)$$

summarizes the constants in the result and relates it to the SNR defined in eq. (6.29). Substituting this normalized SNR into eq. (7.9) results in⁶

$$\text{Var}(\hat{\omega}_0) \geq \delta\omega^2 \text{Var}(\hat{t}_0) + 192 \frac{1}{\eta \tau^3}. \quad (7.11)$$

The CRLB on the start frequency estimate can be converted into an energy resolution limit using eq. (4.3) and Gaussian propagation of uncertainty⁷:

$$\Delta E_{\text{kin}} = \frac{\sqrt{\text{Var}(\hat{\omega}_0)}(m_0 c^2 + E_{\text{kin}})^2}{e c^2 B}. \quad (7.12)$$

In conclusion, eqs. (7.11) and (7.12) provide a straight forward way of calculating the energy resolution for an assumed noise variance. For an experimental setup, the signal amplitude A and the frequency slope $\delta\omega$, can be extracted from a simulation in CRESana. However, even using the optimal maximum likelihood solution, the CRLB is only achieved in the asymptotic limit and therefore a CRLB based energy resolution is possibly overly optimistic for finite data (i.e. for short τ). Furthermore, the assumed chirp model is only a viable approximation for electrons with $\alpha_0 = 90^\circ$. With the lack of a sufficient closed-form model for the modulated signals of smaller pitch angles, a similar treatment is unattainable for these cases.

7.2.2 Numeric Approach: Asimov Dataset

As an alternative to the analytic CRLB, the estimation limits can be obtained through brute-force using numeric MC experiments with simulated data $\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}_{\text{true}}) + \mathbf{n}$, where \mathbf{n} represents Gaussian noise. For each MC experiment the MLE can be obtained with the log-likelihood function of eq. (7.2), where the signal matrix $\mathbf{s}(\boldsymbol{\theta})$ is provided by the same model that generates the signal for the data. Repeating the MLE with many MC experiments yields the full distribution of estimates $\hat{\boldsymbol{\theta}}(\boldsymbol{\theta})$. Utilizing CRESana with a fixed experimental setup as the model for $\mathbf{s}(\boldsymbol{\theta})$, then allows

⁶The RMS power of the chirp is $P_T = \frac{A^2}{R}$, while $\sigma^2 = \frac{Rk_B T}{\Delta t}$.

⁷Assuming $\Delta B = 0$ in the given expression.

for an analysis of the achievable energy resolution $\Delta E_{\text{kin}}(\boldsymbol{\theta}_{\text{true}})$ of CRES for that setup.

However, this approach poses a significant computational challenge, since the goal is the determination of the energy resolution as a function of $\boldsymbol{\theta}_{\text{true}}$. Therefore, the number of CRESana simulations required for this, scales with $N_{\boldsymbol{\theta}} \cdot N_{\text{MC}} \cdot N_{\text{MLE}}$, where $N_{\boldsymbol{\theta}}$ represents the number of desired evaluation points in the 8-dimensional parameter space, N_{MC} the number of MC experiments and N_{MLE} the average number of evaluations of the likelihood function required to obtain the MLE.

In order to save the computation time of sampling data from many experiments it has been proven to be sufficient to analyze a single representative data set as discussed in [131]. This so-called *Asimov* data set is constructed as the expectation value of \mathbf{x} for the given $\boldsymbol{\theta}_{\text{true}}$. In the case of WGN this is a simple task, since the expectation value of the noise is zero and thus $\langle \mathbf{x} \rangle = \mathbf{s}(\boldsymbol{\theta}_{\text{true}})$. For this data set the MLE is always found at $\boldsymbol{\theta}_{\text{true}}$ with $\ell(\boldsymbol{\theta}_{\text{true}}; \langle \mathbf{x} \rangle) = 0$.

Furthermore, the α level confidence regions for $\boldsymbol{\theta}$ given by $\ell(\boldsymbol{\theta}; \langle \mathbf{x} \rangle)$ with Wilks' theorem correspond to the regions covering a probability of α in the distribution of MLEs $\hat{\boldsymbol{\theta}}(\boldsymbol{\theta})$ with actual observations \mathbf{x} . This is visualized in fig. 7.5, which compares the confidence regions obtained from an Asimov data set to the distribution of \hat{f} from MC experiments of estimating the frequency f of a sinusoid.

Therefore, with the Asimov data set, the issue of numerically finding the maximum can be neglected entirely. Instead, for the determination of the energy resolution it is sufficient to extract the parameter uncertainty ΔE_{kin} from the likelihood landscape of $\langle \mathbf{x} \rangle$ in the vicinity of $\boldsymbol{\theta}_{\text{true}}$. With this approach, the number of required simulations now scales with $N_{\boldsymbol{\theta}} \cdot N_{\Delta E_{\text{kin}}}$, where $N_{\Delta E_{\text{kin}}}$ represents the average number of evaluations of the likelihood function required to obtain ΔE_{kin} .

The parameter uncertainties $\Delta \theta_i$ will be calculated using iminuit [132], a python interface to the C++ library MINUIT2 [133, 123, 124], the de facto standard for minimization and uncertainty determination in particle physics. It supports both the Hesse and Minos approach as discussed in section 7.1.1. The python interface has the advantage of a convenient implementation of the likelihood function in python, where $\mathbf{s}(\boldsymbol{\theta})$ is a function that directly calls CRESana with the experimental configuration under consideration. Therefore, CRESana is called once initially to generate $\langle \mathbf{x} \rangle = \mathbf{s}(\boldsymbol{\theta}_{\text{true}})$ and subsequently the minimization library decides on evaluation points $\boldsymbol{\theta}$ of the likelihood function which are forwarded to CRESana to produce the necessary signals. In this pure python implementation data is directly passed from the simulation tool to the minimizer and thus wait times for writing

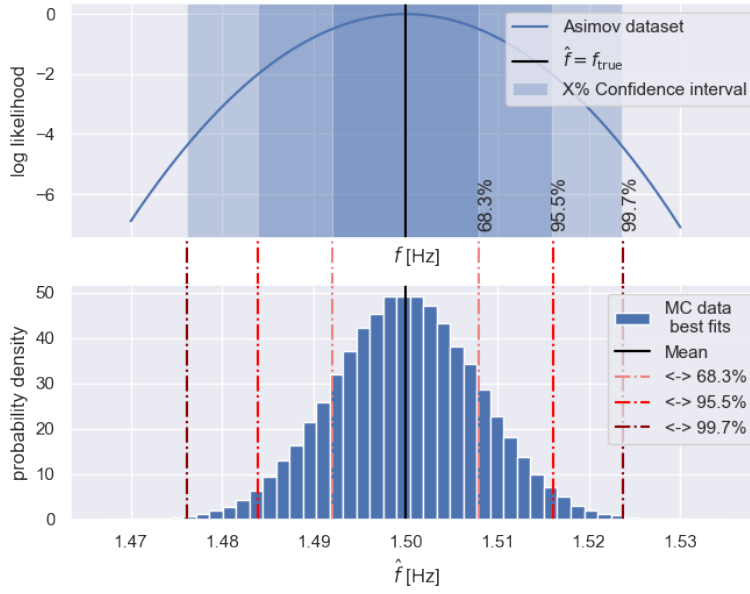


Fig. 7.5.: Determination of estimation limits for the frequency f of a sinusoid with an Asimov data set compared to MC experiments. Top is the log-likelihood of the Asimov data set with confidence intervals. Bottom is the distribution of \hat{f} from MC experiments. Lines mark interval boundaries containing 68.3%, 95.5% and 99.7% of observations. They align with the confidence intervals from the top.

and reading files are eliminated. On the contrary, using Locust as the simulation tool would only allow interfacing with the minimizer through files.

Despite that, the calculation of uncertainties with this setup can still end up in a major computational challenge. Assuming that an antenna array with thousands of channels and track lengths $\tau > 100 \mu\text{s}$ are required (see chapter 8), CRESana takes more than 100 s (see fig. 5.17) for the simulation in a single likelihood evaluation. For an N -dimensional parameter space the Minos error then requires an $(N - 1)$ -dimensional minimization for each calculated point in the profile likelihood. Under these considerations Minos errors seem infeasible at the moment. Therefore, it is important to validate that the Hesse errors yield a good approximation.

7.2.3 Validation of Gaussian Assumption for the Numeric Approach

The Hesse error assumes that near the maximum the likelihood function is well approximated by a Gaussian PDF for $\hat{\theta}$, resulting in the log-likelihood

$$\ell_{\text{approx}}(\boldsymbol{\theta}; \mathbf{x}) = -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{true}})^T \mathbf{C}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{true}}) + \text{const}, \quad (7.13)$$

where \mathbf{C} is the covariance matrix. The isosurfaces $\ell_{\text{approx}}(\boldsymbol{\theta}; \mathbf{x}) = c$ defined by an equation of this form are n -dimensional ellipsoids in the parameter space [134]. Wilks' theorem (eq. (7.3)) with the assumption that $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_{\text{true}}$ leads to

$$c = -\frac{1}{2}Q_{\chi_n^2}(\alpha), \quad (7.14)$$

which relates the ellipsoid of c to a confidence level α .

A convenient way for calculating points in the parameter space on the ellipsoid surface is provided through the mapping

$$f(\mathbf{r}) = \boldsymbol{\theta}_{\text{true}} + \mathbf{U}\sqrt{\boldsymbol{\Lambda}}\mathbf{r}, \quad (7.15)$$

where \mathbf{U} is the matrix of eigenvectors of \mathbf{C} and $\boldsymbol{\Lambda}$ are the corresponding eigenvalues arranged in a diagonal matrix [134]. Using this mapping, the ellipsoid surface for the m -sigma level of the distribution can be defined as

$$E_m(\mathbf{C}, \boldsymbol{\theta}_{\text{true}}) = \{f(\mathbf{r}) \mid \mathbf{r} \in S_n(m)\},$$

where \mathbf{r} is a vector on the sphere $S_n(m) = \{\mathbf{r} \in \mathbb{R}^n \mid \sum_{i=0}^{n-1} r_i^2 = m^2\}$ given in the eigenbasis.

The evaluation of eq. (7.13) for a $\boldsymbol{\theta}_E \in E_m(\mathbf{C}, \boldsymbol{\theta}_{\text{true}})$ produces the constant log-likelihood $c = \ell_{\text{approx}}(\boldsymbol{\theta}_E; \mathbf{x})$ at the m -sigma level. Replacing the covariance matrix by its eigenvalue decomposition $\mathbf{C} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$, the result is

$$c = -\frac{1}{2}(\mathbf{U}\sqrt{\boldsymbol{\Lambda}}\mathbf{r})^T \mathbf{C}^{-1}(\mathbf{U}\sqrt{\boldsymbol{\Lambda}}\mathbf{r}) = -\frac{1}{2}\mathbf{r}^T \mathbf{r} = -\frac{1}{2}m^2.$$

Hence, for a given log-likelihood value c , the corresponding sigma level of the assumed Gaussian likelihood function is⁸

$$m = \sqrt{-2c}. \quad (7.16)$$

The Hesse error of Minit produces an estimate of the covariance matrix \mathbf{C} of the assumed approximate likelihood function in eq. (7.13). Therefore, to validate that this assumption is reasonable after the estimation of \mathbf{C} , one can follow these steps: calculate parameter points on the confidence ellipsoid, evaluate the actual likelihood function for these points, and then determine the corresponding sigma level $m'(\boldsymbol{\theta}) =$

⁸This result connecting sigma level m and log-likelihood level c independent of the dimension n might appear unexpected if one is used to adjusting c for the dimension. This is only the case if a constant value of the associated confidence level α is required. The latter depends on n as seen in eq. (7.14).

$\sqrt{-2\ell(\boldsymbol{\theta}; \mathbf{x})}$. If the likelihood function is well approximated by a Gaussian at the assumed sigma level m , then $m'(\boldsymbol{\theta}_E) \approx m$ for all $\boldsymbol{\theta}_E \in E_m(\mathbf{C}, \boldsymbol{\theta}_{\text{true}})$.

While it is possible to arrive at the same conclusion by simply verifying that $\ell_{\text{approx}}(\boldsymbol{\theta}_E; \mathbf{x}) \approx \ell(\boldsymbol{\theta}_E; \mathbf{x})$, the comparison of m' and m has the advantage of providing a comprehensible metric for the approximation discrepancy if the values do not closely match. For example, if the confidence ellipsoid is calculated for a 1-sigma level error and the evaluation of the actual likelihood function yields $m' = 0.1$ at the evaluation point, it is obvious that the likelihood function deviates significantly from the Gaussian approximation, which underestimates the true error in this instance.

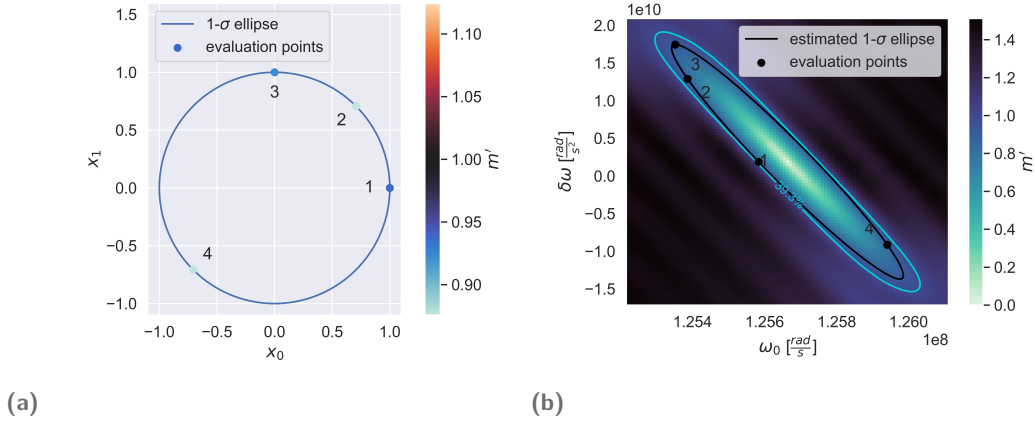


Fig. 7.6.: Demonstration of proposed Gaussian validation for a chirp with unknown parameters ω_0 and $\delta\omega$. (a) depicts the 1- σ ellipse in the x_0 - x_1 -plane of the eigenbasis and its minimum set of validation points. These are mapped into the parameter space using eq. (7.15) with \mathbf{C} estimated by minuit. Point colors indicate resulting values m' obtained with the original log-likelihood. (b) shows a 2D log-likelihood scan converted to m' , accompanied by the points and the estimated ellipse from (a) (both in black), which slightly deviate from the true 1- σ boundary (light-blue).

However, with a likelihood function that is expensive to evaluate, it is desirable to apply the steps above only to a small set of evaluation points. Furthermore, for a concise result it is beneficial to condense the outcome of the validation into a single number for each calculated uncertainty. Therefore, the following validation routine is proposed for this thesis:

1. A minimum set of evaluation points S_{min} in the eigenbasis of \mathbf{C} is used, consisting of the points $m\hat{\mathbf{r}}_i$ along the n coordinate axes, as well as two points in each coordinate plane, namely $\pm \frac{m}{\sqrt{2}}(\hat{\mathbf{r}}_i + \hat{\mathbf{r}}_j)$ for all $j > i$.
2. This generates the validation set in the parameter space via the mapping defined by the covariance matrix as $VS(\mathbf{C}, \boldsymbol{\theta}_{\text{true}}) = \{f(\mathbf{r}) \mid \mathbf{r} \in S_{\text{min}}\}$ with a total size of n^2 evaluation points.

3. The original likelihood function is evaluated for all parameter point combinations in $VS(\mathbf{C}, \boldsymbol{\theta}_{\text{true}})$, producing $M = \{m'(\boldsymbol{\theta}) | \boldsymbol{\theta} \in VS(\mathbf{C}, \boldsymbol{\theta}_{\text{true}})\}$.
4. The validation result is given by the value m' with the highest absolute deviation from the true value of m :

$$m'_{max} = \arg \max_{m' \in M} (|m' - m|) . \quad (7.17)$$

The choice of using eq. (7.17) as result assigns a high weight to a large deviation of just a single point to highlight potential misrepresentations in the covariance matrix, which could lead to a considerable underestimation of the uncertainty of interest.

Figure 7.6 visualizes the point set S_{\min} and its mapping into the parameter space for the calculation of M in a 2D estimation problem⁹.

7.2.4 Implementation of Numeric Approach for the Chirp Model

Before applying the numeric procedure described in the previous sections to signals from CRESana, it will be tested using the faster and relatively simple chirp model defined in eq. (7.5). The chirp model provides a valuable benchmark of the procedure considering that asymptotically its estimation limits are known through the CRLB. Furthermore, it is instructive to study possible complications of the procedure in this model. Since the model is a good approximation for the signals of a single channel with $\alpha_0 = 90^\circ$, complications encountered here must be anticipated for CRESana signals as well.

The chirp model has five unknown parameters $\boldsymbol{\theta} = (A, \omega_0, \delta\omega, t_0, \varphi_0)$. Among these, ω_0 is the only parameter of interest, and therefore, the goal is the calculation of $\Delta\omega_0$. However, as demonstrated in section 7.1.1, correlations of the parameter of interest with nuisance parameters broaden the profile likelihood and thus increase the uncertainty on the parameter of interest. Consequently, to compute $\Delta\omega_0$ all parameters must be accounted for unless their independence from ω_0 can be verified, which is done by inspecting 2D slices of the log-likelihood function as can be seen in fig. 7.7. The results show that only the amplitude is an independent parameter that can be neglected, whereas all other parameters exhibit significant correlations with each other.

⁹For higher-dimensional cases additional evaluation points are added from the other eigenbasis coordinate planes. Furthermore, the eigenbasis coordinate planes map to arbitrary planes that generally do not align with the parameter space coordinate planes, unlike fig. 7.6b.

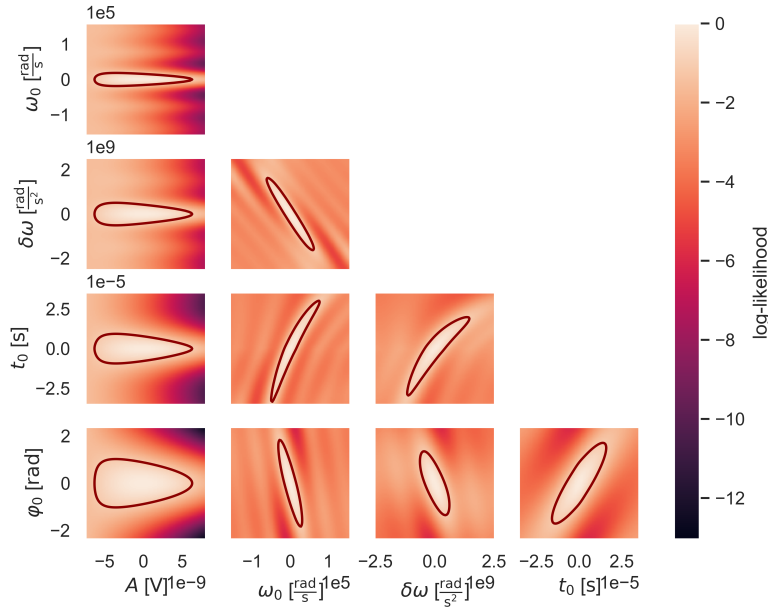


Fig. 7.7.: Asimov log-likelihood of chirp with $A = 7 \text{ nV}$, $\omega_0 = 2\pi 20 \text{ MHz}$, $\delta\omega = 2\pi 300 \text{ MHz s}^{-1}$, $t_0 = 100 \mu\text{s}$, $\varphi_0 = 0$ and with $\tau = 100 \mu\text{s}$, $T = 5 \text{ K}$, $f_s = 200 \text{ MHz}$. Each 2D plot fixes remaining parameters at their true values, with contour lines indicating the 68% confidence region from Wilks' theorem. The origin of the coordinate axes is set to the true parameter values.

Phase

The correlations with the phase φ_0 can be addressed in a simple way by a modification of the likelihood function. Using once again flat vectors $\tilde{\mathbf{x}}$, $\tilde{\mathbf{s}}$ for the data and signal matrices, one can identify that eq. (7.2) is equivalent to

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = -\frac{1}{\sigma^2} \left(|\tilde{\mathbf{x}}|^2 + |\tilde{\mathbf{s}}(\boldsymbol{\theta})|^2 - 2\text{Re} \left(\tilde{\mathbf{x}}^H \tilde{\mathbf{s}}(\boldsymbol{\theta}) \right) \right) + \text{const}. \quad (7.18)$$

The sensitivity of the likelihood function to the phase is introduced by taking the real part of the inner product. Replacing that by the absolute value leads to a modified likelihood function

$$\ell_{\text{mod}}(\boldsymbol{\theta}; \mathbf{x}) = -\frac{1}{\sigma^2} \left(|\tilde{\mathbf{x}}|^2 + |\tilde{\mathbf{s}}(\boldsymbol{\theta})|^2 - 2 \left| \tilde{\mathbf{x}}^H \tilde{\mathbf{s}}(\boldsymbol{\theta}) \right| \right) + \text{const}. \quad (7.19)$$

One can easily see that the modified likelihood function is invariant under φ_0 with $\ell_{\text{mod}}(\boldsymbol{\theta}; \mathbf{x}) = \max_{\varphi_0} (\ell(\boldsymbol{\theta}; \mathbf{x}))$, which makes this a proper profile likelihood of the remaining parameters. The suggested modification is consistent with the use of the

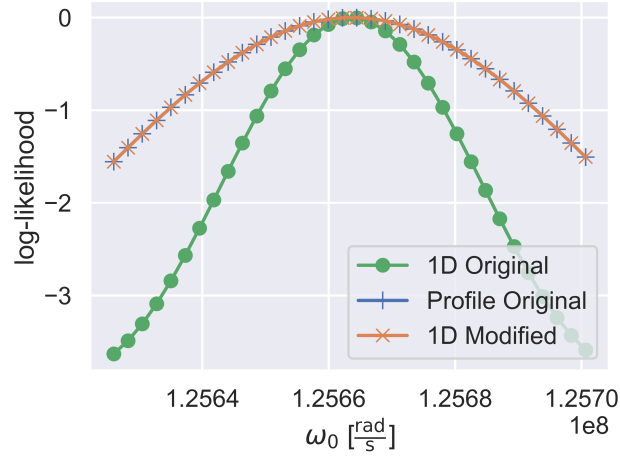


Fig. 7.8.: Comparison of different 1D likelihood functions of the chirp model. Blue represents the original likelihood with φ_0 fixed to the true value, orange the profile likelihood and green the modified likelihood. The last two overlap perfectly.

absolute value in the quadrature matched filter (eq. (6.20)) as detector for a signal with unknown phase, whereas the matched filter with known phase uses the real part (eq. (6.14)).

Figure 7.8 depicts the difference between the likelihood functions for the chirp model through scans along the ω_0 axis. The modified 1-dimensional function $\ell_{\text{mod}}(\omega_0; \mathbf{x})$ (eq. (7.19)) is identical to the profile likelihood for ω_0 obtained via maximization of the original function: $\max_{\varphi_0} (\ell(\omega_0, \varphi_0; \mathbf{x}))$. Compared to the original 1-dimensional function $\ell(\omega_0; \mathbf{x})$ (eq. (7.18)), which fixes the phase to its true value, the width of the maximum noticeably broadens.

Figure 7.9 shows the same correlation plots of fig. 7.7 with this modified likelihood function. As expected, this confirms that all parameters are now independent of φ_0 and in return the maxima generally increase in width. In addition, the last row shows that the modified likelihood function is incapable of estimating φ_0 . This is acceptable because φ_0 is only a nuisance parameter.

Considering that the electron parameter ϕ_{cycl} causes a similar constant phase shift of the recorded electron signal, i.e. $\mathbf{s}(\boldsymbol{\theta}) = e^{i\varphi_{\text{cycl}}} \mathbf{s}(\boldsymbol{\theta}_{\varphi_{\text{cycl}}=0})$, it can be expected that ϕ_{cycl} is correlated with the signal frequency and by extension E_{kin} , even outside of the idealized chirp model. Therefore, in order to disregard ϕ_{cycl} , eq. (7.19) should always be used for the likelihood function in CRES parameter estimation.

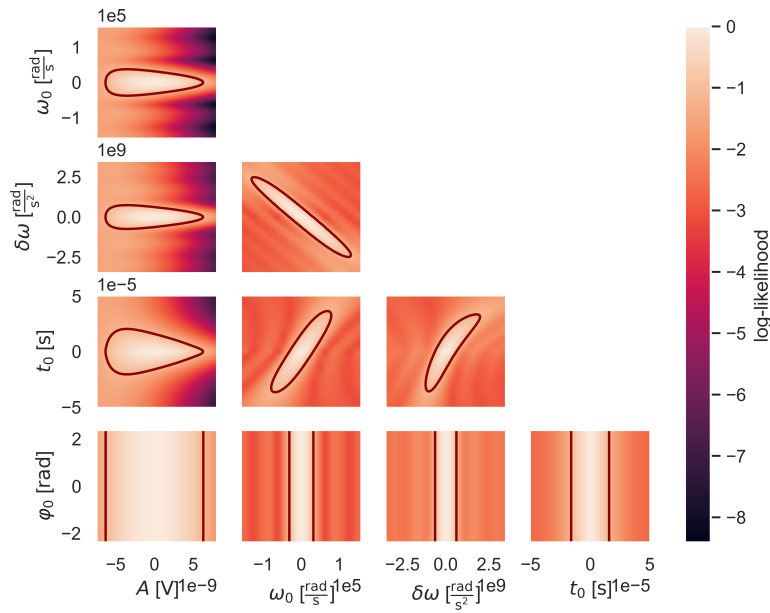


Fig. 7.9.: Modified asimov log-likelihood (eq. (7.19)) of chirp. Model parameters are identical to fig. 7.7.

Start Time

The single parameter that stands out the most is the start time t_0 . Figure 7.10a depicts the likelihood functions along the t_0 axis using a higher scanning resolution than before. The original phase sensitive likelihood function (eq. (7.18)) is highly oscillating along this axis, due to x and s going in and out of phase along t_0 . These oscillations are difficult to handle for a minimizer and severely complicate an accurate determination of the profile likelihood for Wilks' theorem and the Minos error $\Delta\omega_0$ (which requires a minimization over t_0). The Hesse error also performs inadequately, as the narrow maxima yield tiny uncertainties that do not account for the fact that the side maxima span a large range of t_0 before permanently dropping below the confidence threshold. In many cases (e.g. in fig. 7.7) it was also observed that the oscillations and full side-peak amplitudes do not get resolved due to a too high evaluation step size. This results in the likelihood function appearing smooth and oscillation-free along t_0 with a maximum of arbitrary width due to the missed peaks. All these observations reinforce the need for the modified likelihood function, which yields the envelope of the oscillations as depicted in fig. 7.10a.

However, a close-up of the global maximum of the modified likelihood function, as seen in fig. 7.10b, reveals that this function is non-differentiable at the maximum, which prevents a meaningful evaluation of the Hessian matrix for the uncertainty.

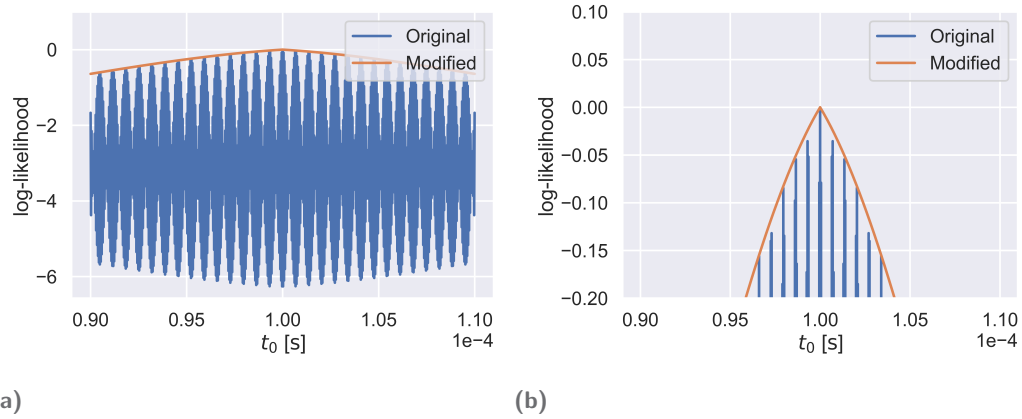


Fig. 7.10.: High resolution scans of likelihood functions of eqs. (7.18) and (7.19) for the chirp model along t_0 axis. All other parameters are fixed to their true values (same as in fig. 7.7). (b) shows a close up of the maximum in (a).

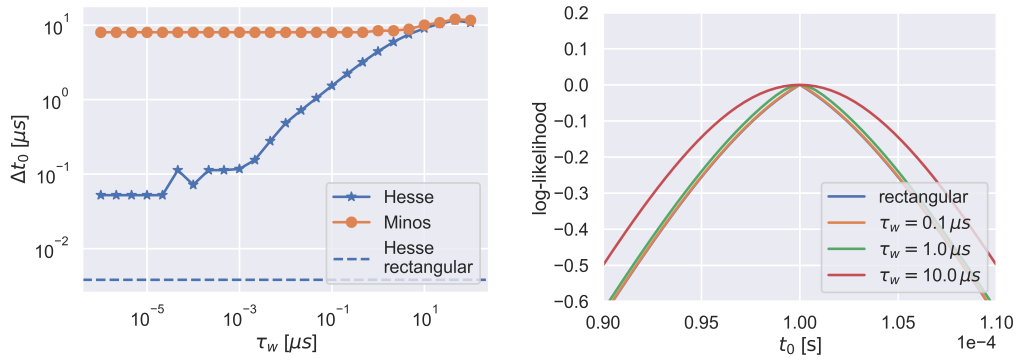
The reason for this is the rectangular window describing the instantaneous start and end of the signal. Without a simulation of the startup phase of a physical antenna, a straight forward fix to this issue is to use any smooth window function. Figure 7.11 demonstrates the effect of a window function based on the logistic function in the form

$$g(t) = \frac{1}{1 + e^{-4/\tau_w(t-t_0)}} \left(1 - \frac{1}{1 + e^{-4/\tau_w(t-(t_0+\tau))}} \right). \quad (7.20)$$

This is a function that creates smooth amplitude transitions near t_0 and $t_0 + \tau$ with the parameter τ_w controlling the steepness of the transition. Lower values of this time constant τ_w correspond to short response times, reminiscent of the rectangular window.

Figure 7.11a shows that with this window function the calculated Hesse error increases from the minuscule values obtained for the rectangular window to a reasonable scale. Nevertheless, below $\tau_w \approx 10 \mu\text{s}$ this error is still significantly smaller than the Minos error. The likelihood scans in fig. 7.11b verify that the window function smooths the peak, thereby making it differentiable, while the shape remains mostly unchanged for small τ_w . Therefore, the Hesse error still underestimates the true uncertainty given by the Minos error.

The reason for this is that even with a smooth maximum the log-likelihood is not approximately parabolic, and hence the Hessian matrix of the maximum contains no information about the function shape far from the maximum. This implies that the Hesse error only accounts for the additional uncertainty due to the window functions amplitude delay, which is, for $\tau_w \lesssim 10 \mu\text{s}$, overshadowed by the signal's inherent start time uncertainty under an instantaneous response. However, above



(a) Start time uncertainty with the likelihood function of eq. (7.19). Plot displays Hesse and Minos errors compared to Hesse error with the rectangular window. (b) Scan of the likelihood function of eq. (7.19) along the t_0 axis with different window functions.

Fig. 7.11.: Application of the window function from eq. (7.20). All parameters except t_0 fixed to their true values, which have the same true values as in fig. 7.7.

$10 \mu\text{s}$, the amplitude delay prevents the signal from reaching full amplitude within the $100 \mu\text{s}$ duration, which is unexpected for a real measurement. This suggests that the uncertainty contribution from the signal onset is very likely negligible¹⁰, and therefore, the addition of an artificial window function is superficial for the correct result.

Nevertheless, it was found that Minuit’s Minos procedure is incapable of producing any result at all without smoothing the maximum at least slightly. This is most likely due to Minuit’s minimization algorithm, which is gradient based and probably does not deal well with the non-differentiable maximum. Thus, for the calculation of Minos errors, the window function from eq. (7.20) was used with τ_w sufficiently small. For Hesse errors on the other hand, the inclusion of the windowing function is ineffective since the obtained results depend exclusively on the assumed window function, rendering them arbitrary while still underestimating the inherent uncertainty.

In a final attempt to enable the use of Hesse errors, an approximation scheme was tested. In this approach, the log-likelihood function was evaluated on an even grid around the maximum with four points per dimension. These evaluation points were used to construct a second order interpolating spline as a differentiable approximation of the log-likelihood. This was motivated by fig. 7.11b, where it appears that, apart from the maximum, the function is nearly parabolic along the t_0

¹⁰There may be scenarios with other window functions or true parameter settings of the model where this is different.

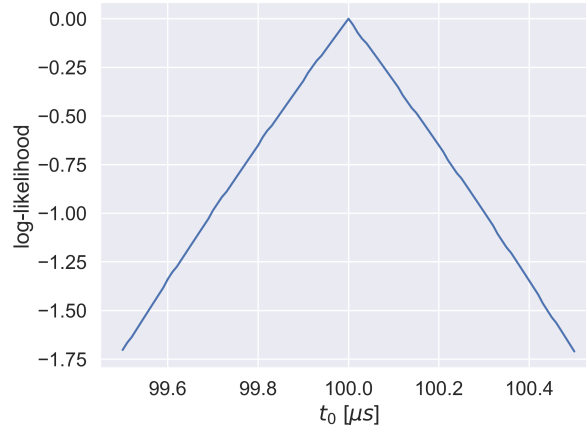


Fig. 7.12.: Scan of modified asimov log-likelihood for the chirp model along t_0 axis. The parameters are set corresponding to 10% of the electron's Larmor power for $B = 0.96$ T: $A = 73.6$ nV, $\omega_0 = 2\pi 20$ MHz, $\delta\omega = 2\pi 331$ MHz s $^{-1}$, $t_0 = 100$ μ s, $\varphi_0 = 0$. Relevant noise parameters are still $\tau = 100$ μ s, $T = 5$ K, $f_s = 200$ MHz.

axis. Similarly, in fig. 7.9 it seems that the corresponding 2D contours are close to being elliptic.

While this approximation scheme indeed produces accurate errors with the Hesse procedure in some cases, it proved to be inconsistent and non-universal. For example fig. 7.12 depicts a case with a higher SNR that is still well within expectations. In that case the log-likelihood is actually a triangular function and thus the approximation approach breaks down.

The discovery that the Hesse error is inadequate for the determination of Δt_0 is unfortunate due to the correlation of t_0 with all other parameters and thus using the Hesse error underestimates the uncertainty on the parameter of interest by an unknown amount. The only solution at this point would be to calculate the computationally more demanding Minos error, which currently seems to be impossible with CRESana as stated earlier.

7.2.5 Results of the Chirp Model Analysis

Finally, the Hesse and Minos errors $\Delta f = \frac{\Delta\omega_0}{2\pi}$ for the 1- σ level were evaluated in Minuit as a function of the track length τ using the full set of parameters of the chirp model. According to the CRLB (eq. (7.8)) results are expected to vary with the choice of the true amplitude A and chirp rate $\delta\omega$. In a CRES experiment, these abstract model parameters depend on the Larmor power of the electron, which is

mainly determined by the choice of the magnetic field¹¹ B and can be calculated with eq. (4.24).

Since the chirp rate is the time derivative of the cyclotron frequency (eq. (3.1)), expressed as

$$\delta\omega = \frac{d\omega_c}{dt} = -\frac{dE_{\text{kin}}}{dt} \frac{c^2 e B}{(m_0 c^2 + E_{\text{kin}})^2},$$

it is directly proportional to the Larmor power $P_{\text{Larmor}} = -\frac{dE_{\text{kin}}}{dt}$. Meanwhile, the amplitude is proportional to the detected power P_d , which can be modeled using an assumed detector power efficiency ϵ such that $P_d = \epsilon P_{\text{Larmor}}$.

Three scenarios were analyzed for noise with $T = 5$ K:

1. High field with $B = 0.96$ T and $\epsilon = 0.1$, corresponding to $f_c = 25.9$ GHz, $P_{\text{Larmor}} = 1.08 \times 10^{-15}$ W, $\delta\omega = 2\pi 331.3$ MHz s⁻¹ and $A = 23.3$ nV
2. High field with $B = 0.96$ T and $\epsilon = 1.0$, corresponding to $A = 232.8$ nV
3. Low field with $B = 0.04$ T and $\epsilon = 1.0$, corresponding to $f_c = 1.1$ GHz, $P_{\text{Larmor}} = 1.88 \times 10^{-18}$ W, $\delta\omega = 2\pi 0.024$ MHz s⁻¹ and $A = 9.7$ nV

The results of these scenarios are presented in figs. 7.13 to 7.15, including comparisons to the CRLB for fixed t_0 and the results of the proposed Gaussian validation from section 7.2.3 for the Hesse errors.

First, it should be noted that in most plots missing data points and sudden jumps that go against apparent trends can be observed. In the first case Minuit terminated its calculation in a failed state, whereas the jumps are most likely due to numerical instabilities encountered during the estimation of the Hessian matrix or the minimization for the profile likelihood. Manual inspection of the dubious points via scans of the likelihood function did not uncover any unexpected behavior that could explain the outliers.

Unfortunately, debugging these cases is hard since it is almost impossible to get insights into internal warnings and errors of Minuit with the python interface `iminuit`. Therefore, it must be accepted that the proposed calculation of estimation limits with Minuit does not consistently yield reliable results. Overall this is a minor issue considering that the majority of plot points is unaffected and that the order of magnitude is mostly correct as can be seen in the respective top plots with the log scale.

¹¹It only weakly varies over the range of relevant kinetic energies (few eV around the tritium endpoint at 18.6 keV) and pitch angles ($90^\circ \pm 10^\circ$)

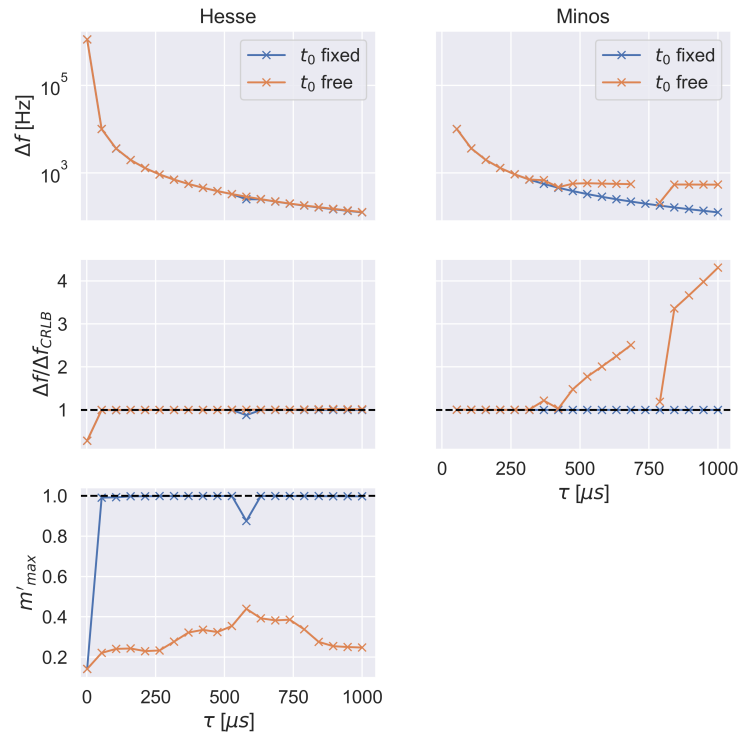


Fig. 7.13.: Minuit results for Δf of a chirp in scenario one. Hesse results are on the left, Minos on the right. The rows display Δf , a ratio plot as comparison to the CRLB with fixed start time, i.e. eq. (7.9) where $\text{Var}(\hat{t}_0) = 0$, and the validation result m'_{max} , i.e. eq. (7.17) from section 7.2.3 (only for Hesse error). For orange data points t_0 is a free parameter, while it is fixed to the truth for the blue ones.

The notable exception to the two cases discussed above is the jump in figs. 7.13 and 7.15 at $\tau = 1 \mu\text{s}$, which are characterized by a very low SNR. In these cases the profile likelihood does not drop below the $1\text{-}\sigma$ confidence level and thus the Minos error is technically infinite and no data point is displayed for it. At the same time, the Hesse error fails since this behavior is non-gaussian.

Figure 7.13 shows that the Hesse errors with fixed t_0 agree almost perfectly with the CRLB as evident from the ratio plot in the middle row except for the cases mentioned before. Furthermore, the result of $m'_{max} \approx 1$ in the last row proves that the Hesse error is accurate for the $1\text{-}\sigma$ level. With t_0 free on the other hand, the Hesse error of Δf does not change at all, which is obviously due to the minuscule error of t_0 as seen in fig. 7.11a. As expected for that case, the validation routine yields small values of m'_{max} indicating that the log-likelihood is indeed incompatible with the Gaussian assumption.

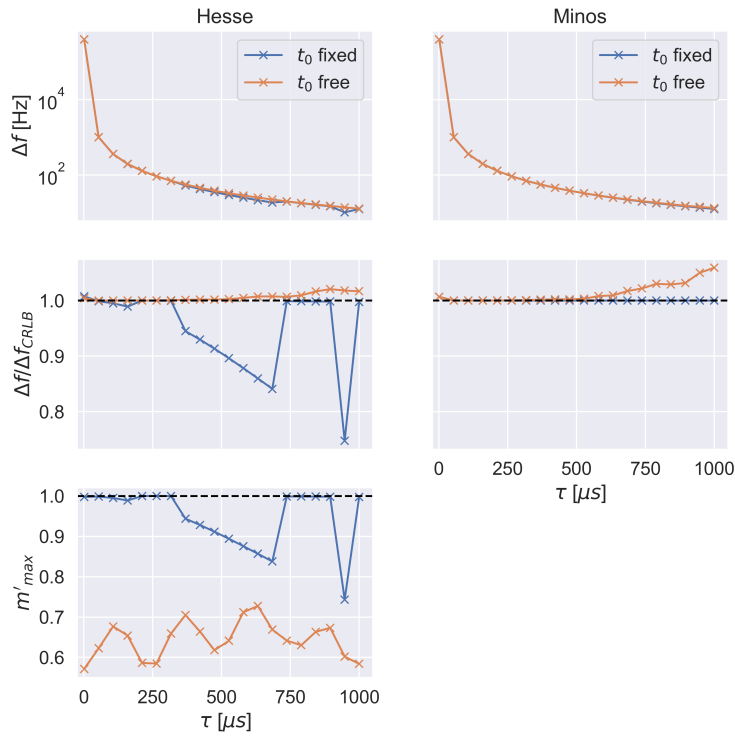


Fig. 7.14.: Minit results for Δf of a chirp in scenario two. For further details see caption of fig. 7.13.

The correct result for the case of t_0 unknown is only obtained through the Minos error, which (neglecting the calculation artifacts mentioned above) reaches a plateau for long tracks that is significantly higher than the displayed CRLB for fixed t_0 . This difference demonstrates the general importance of the inclusion of t_0 as a free parameter. The two different Minos results are consistent with the expectations set by eq. (7.8), which states that Δf^2 gets infinitely small with a τ^{-3} scaling, but is eventually limited by the contribution from Δt_0 .

However, the results presented in fig. 7.13 are for a case, where a signal with only 1% of the emitted Larmor power is recorded. The results for 100% Larmor power are depicted in fig. 7.14. The 10 times increase in amplitude reduces Δf by a factor of 10, which is consistent with the expectation from the CRLB. The observations for the Hesse error are mostly the same as for fig. 7.13, although the ratio plot displays many of the aforementioned artifacts in the case of fixed t_0 . While the result does appear to show a slight systematic disagreement with the CRLB in the range of 350–700 μs , this coincides with a similar trend in the validation plot in the last row, indicating that these results are incorrect. At the same time, the Hesse errors for

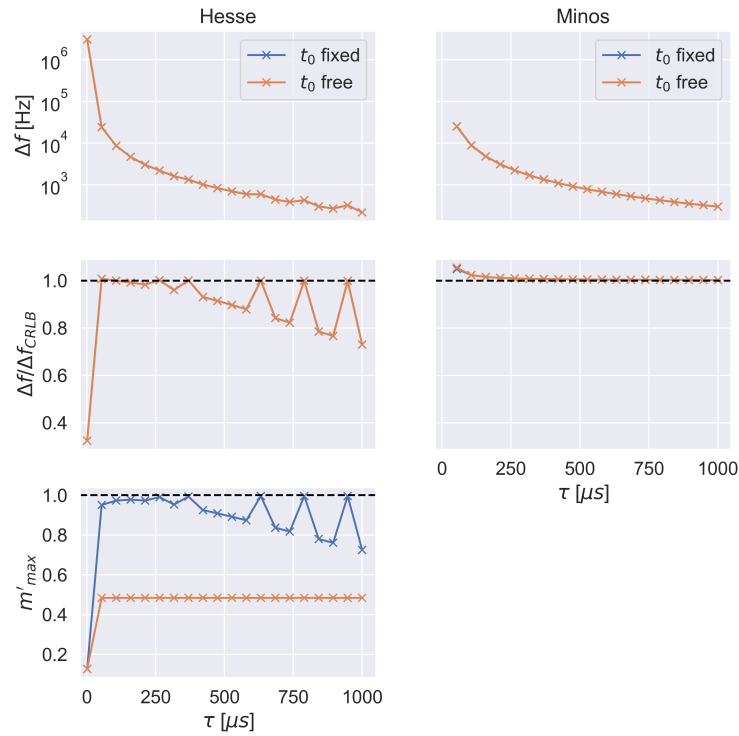


Fig. 7.15.: Minit results for Δf of a chirp in scenario three. For further details see caption of fig. 7.13.

t_0 free and Minos errors with t_0 fixed again closely match the CRLB. This further supports the assumption of an erroneous calculation of the Hessian matrix at the suspicious data points. Interestingly, the inclusion of t_0 as a free parameter appears to help Minit in the estimation of the Hessian matrix in this scenario, even though t_0 does not contribute to the uncertainty. Finally, the Minos error with t_0 free shows that the contribution from Δt_0 only marginally increases Δf for tracks longer than $500 \mu\text{s}$, not even reaching a 10% increase at $\tau = 1 \text{ ms}$.

The results for the last scenario with the lower magnetic field are depicted in fig. 7.15. In this case the Larmor power drops to 0.2% of the high field case, which increases Δf accordingly. While this represents a further five times reduction in detected power compared to scenario one, the main difference lies in the minuscule chirp rate: the cyclotron frequency only changes by 24 Hz over the full $1000 \mu\text{s}$ range of the plot. Given this negligible frequency increase, no significant contribution from the start time uncertainty is expected. Indeed, the Minos result indicates no difference when t_0 is included, and, neglecting the familiar artifacts of the Hesse error, all four

approaches to the uncertainty calculation yield the same result, which agrees very well with the CRLB.

The conclusions from figs. 7.14 and 7.15 imply that in these two cases the correlation of t_0 and the other parameters is weaker than previously anticipated. These claims are further supported by additional 2D correlation plots of the log-likelihood in appendix D.1 for these two cases. In the case with high magnetic fields, t_0 is indeed independent of the other parameters for short tracks, which turns into a non-negligible correlation for very long tracks. This contrasts with the low magnetic field case, where t_0 is uncorrelated even for long tracks. This proves that under the right circumstances, the impact of Δt_0 is mostly negligible and Hesse errors would be adequate.

As a final remark fig. 7.15 also shows an unfortunate weakness of the validation: the values of m'_{max} for the case of t_0 treated as a free parameter seem to indicate that the results might be incorrect due to the invalid Gaussian approximation. The obtained results are due to the inclusion of the t_0 axis and correctly indicate its non-Gaussian behavior. Despite that, the Minos errors prove that the Gaussian approximation still yields valid results. This apparent contradiction can be explained by the fact that t_0 is uncorrelated with the other parameters and thus its underestimation does not influence Δf while it is still included in the validation routine. Due to the decision to pick the maximum as the final result of the validation routine, a high weight is assigned to the deviation along the t_0 axis. Therefore, to prevent false conclusions uncorrelated parameters should be excluded from the validation routine.

7.3 Summary

This chapter outlined the utilization of the maximum likelihood method for the estimation of the parameters θ of CRES electrons. This solution is asymptotically optimal, but it again depends on a suitable signal model. For real data-taking conditions, where the global maximum is unknown a priori, further research is needed to develop efficient search strategies suited to the specific likelihood function with its narrow maxima and side maxima.

Additionally, in this chapter a method using simulated data was established for determining the estimation limits of the assumed optimal maximum likelihood approach, with the goal of determining the energy resolution $\Delta E_{\text{kin}}(\theta_{\text{true}})$. The procedure utilizes MINUIT [123, 124] for maximum likelihood estimation of an Asimov data set for CRES signals in white Gaussian noise (WGN) to evaluate

parameter uncertainties using the Hesse or Minos approach. It was tested for the frequency estimation in the chirp model, which serves as an approximation for CRES signals of $\alpha_0 = 90^\circ$ electrons.

Overall, the results are in very good agreement with the theoretical expectation set by the analytic Cramér–Rao lower bound (CRLB) (see section 7.2.1). The analysis of the chirp model demonstrated that a phase-invariant likelihood function, given in eq. (7.19), is recommended for CRES estimation. Furthermore, it was found that the inclusion of the start time t_0 is in general mandatory since it is correlated with the frequency. Unfortunately, the proper treatment of t_0 turned out to be challenging for a number of reasons and is only possible through the computationally demanding Minos errors. This is problematic considering that the final goal is the application to CRESana data of a full-scale experiment, for which the faster Hesse errors are preferred.

Nevertheless, the results obtained with the chirp model also demonstrated that t_0 can be disregarded under certain conditions, allowing the Hesse errors to yield good results. This is unconditionally true for experiments utilizing low magnetic fields, whereas high magnetic field settings require the detection of high fractions of the emitted electron power. The latter is problematic, as the power efficiency is realistically 10–30 % and varies with the detector design. Thus, a general treatment for any detector in a high field scenario necessitates the use of Minos errors. However, for a final Phase IV experiment that includes atomic tritium, a low magnetic field with $f_c < 1$ GHz is preferred due to loss mechanisms in the additional magnetic atom trap [30]. Therefore, the study of the full-scale experiment in chapter 8 will be conducted with a low magnetic field and Hesse errors.

Assessing the Limits of a Full Scale Detector: A Case Study

” *The whole is greater than the sum of its parts*

— Aristotle
Greek Philosopher

The preceding chapters laid all the groundwork for a numerical determination of the limits of detection and parameter estimation of CRES experiments. The next step is the combination with the newly developed simulation tool from chapter 5 in order to establish a simulation based framework that analyzes the theoretical limits of a given experimental setup. The goal of this chapter is to present a proof of concept of this framework utilizing a single demo setup of a full scale Project 8 Phase IV experiment. Since a complete design for a Phase IV experiment does not exist yet, an idealized hypothetical setup will be assumed. Ultimately, the final design will be determined by an optimization for the 0.04 eV neutrino mass sensitivity goal. The framework suggested in this thesis could prove useful for that task, as the results in the form of detection efficiency and energy resolution are crucial ingredients for the calculation of the neutrino mass sensitivity. Nevertheless, the evaluation of the neutrino mass sensitivity of the setup based on the results and the setup optimization are outside of the scope of this thesis.

8.1 Experimental Setup

8.1.1 Antenna Array

The antenna array part of the setup encompasses the design of the individual antennas and their placement in the detector. The antenna design depends on the target frequency of the experiment, which is 1 GHz or less for Phase IV due to unacceptable losses of the atom trap at higher magnetic fields [30]. While first prototype antennas operating at 26 GHz have already been designed, fabricated

and used to collect data in a test setup [135], the design work for a suitable 1 GHz antenna has not yet begun. Therefore, an idealized isotropic antenna with constant gain for all frequencies and directions will be assumed for the demo setup. This is certainly unrealistic, but more concrete assumptions about gain and directivity of such an antenna are considered too arbitrary at this point.

The antennas are placed on the lateral surface of a cylinder that surrounds the tritium gas cell (see fig. 3.2). No antennas are placed on the ends of the cylinder, which need to remain open at least from one side to permit access to the inside for calibration tasks and the injection of the atomic tritium beam [30]. The size of the cylindrical array is dictated by the required tritium source volume since the antennas surround the tritium cell. The required source volume, in turn, depends on the target for the statistical uncertainty, as a higher volume with a fixed gas density produces more events over the same time interval. Reaching Project 8's neutrino mass sensitivity target of 0.04 eV necessitates an *effective volume* of $V_{\text{eff}} = 10 \text{ m}^3$ [59]. The effective volume is the physical volume weighted by the detector efficiency ϵ for detecting any given decay electron. It can be interpreted as the equivalent physical gas volume that produces the same event rate, assuming each decay is registered. For any assumed detector setup with a fixed physical volume, this detector efficiency ϵ must be established through simulation.

Following these criteria an antenna array with diameter $D = 4 \text{ m}$ and length $L = 14 \text{ m}$ will be used. Compared to other large physics experiments, this seems reasonably sized, e.g. the main spectrometer of KATRIN has a diameter of 9.8 m and a length of 23.8 m [24]. The assumed antenna array size results in a physical volume of 176 m^3 for the enclosed tritium source. The larger physical volume compared to the target effective volume (10 m^3) is an initial estimate to account for expected losses in electron trapping and detection efficiency. Further analysis will determine if this setup meets the effective volume goal.

To maximize the amount of radiation received, the antenna array should consist of as many antennas as can possibly fit on the limited cylinder surface area. To lessen cost and complexity, the antennas should be distributed without overlapping their sensitive areas. The effective aperture, a measure for how much power a receiving antenna can capture from an incident plane wave, dictates the closest possible useful tiling of antennas on the surface. It is defined as [50]

$$A_e = G A_{iso} = \frac{G\lambda^2}{4\pi}, \quad (8.1)$$

where G is the antenna gain and $A_{iso} = \frac{\lambda^2}{4\pi}$ is the effective aperture of the isotropic receiver. Thus, covering the whole cylinder surface $A_{cylinder}$ with antennas as densely as possible requires a total antenna count of

$$n_a = \frac{A_{cylinder}}{A_e}.$$

The antennas are arranged on a number of n_r rings along the cylinder axis each encompassing n_c antennas around the circumference. Since $n_a = n_r n_c$, a reasonable choice would be

$$n_c = \frac{2\pi R}{\sqrt{A_e}},$$

and

$$n_r = \frac{L}{\sqrt{A_e}}.$$

With isotropic antennas with $G = 1$ the assumed detector size allows for a total of 23184 antennas arranged in 161 rings with 144 antennas each. It was found that the simulation of a single $100\mu s$ track with such an array takes 65–130 min in CRESana, depending on the electron parameters. Unfortunately, the analysis of estimation limits according to section 7.2 becomes computationally infeasible with individual simulation run times on this scale. In theory, the number of antennas can be reduced without decreasing the collected electron radiation by using higher gain antennas, which keeps the combined effective aperture of all antennas constant. However, the isotropic receiver has a gain of 1 by definition, as per eq. (8.1), and higher gains are associated with increased directivity [50]. Correctly assigning a directivity for an assumed gain necessitates a properly engineered antenna design, whereas with CRESana, gain and directivity can (unphysically) be configured independently of each other.

Therefore, it was decided to use a high-gain antenna with the directivity of the isotropic receiver for the demo setup in this thesis. This decision is motivated by the assumption that the signals of n nearby isotropic antennas are almost identical and combine to form a single signal with n times the power. Using a uniform distribution of antennas, the corresponding sparser array has a similar antenna arrangement, leading to the expectation of similar results¹. In fact it was found that the results for energy resolution and detected power do not change significantly between two different gain values (see appendix D.2).

¹The signal from a single isotropic antenna with gain G thus represents identical signals from G antennas with unity gain

The stated assumption is obviously an idealization that breaks down when G is pushed high enough to create vastly different and extreme antenna arrangements (e.g. using just a single antenna per ring). Ultimately, the demo setup will use antennas with $G = 56$, which results in 399 antennas arranged on 21 rings with 19 antennas each. This reduces the computation time for a single $100\mu\text{s}$ track to 48–93 s, depending on the electron parameters.

8.1.2 Magnetic Trap

Magnetic Field Design Requirements

It was shown in section 5.5 that the profile of the magnetic field significantly changes the signal spectrum since it influences the average cyclotron frequency and the axial frequency. Therefore, it is anticipated that the magnetic field choice impacts the reachable energy resolution. In 2D likelihood scans it was also found that the choice of the magnetic field profile affects the correlation between kinetic energy and other electron parameters, as can be seen for energy and pitch angle in fig. 8.1.

The plots show scans of the log-likelihood function from eq. (7.19) of an Asimov dataset for an electron with $E_{\text{kin}} = 18.6\text{ keV}$, $\alpha_0 = 90^\circ$, $r = 0\text{ m}$ and $\tau = 133.75\mu\text{s}$ in thermal noise with $T = 5\text{ K}$ simulated by CRESana with various setups. The signal model \tilde{s} in the likelihood function is provided through CRESana using the same simulation setup as used for the respective data \tilde{x} . All plots use an identical antenna array of 3 isotropic receivers with a gain of $G = 200$, uniformly distributed on a circle with a radius of $R = 0.5\text{ m}$ around the magnetic field's symmetry axis in the $z = 0$ plane (at the magnetic field minimum).

The simulation setups only differ in the magnetic field, which is defined via the profile along the symmetry axis $B_z(0, z)$ given by polynomial functions in the form

$$B_z(0, z) = az^D + B_0,$$

where D is the degree indicated in the plot. In all cases the constant has the value $a = 0.9\text{ T/m}^D$ and the background field is $B_0 = 0.96\text{ T}$. The profile $B_z(0, z)$ defines the full magnetic field vector $\mathbf{B}(\rho, z)$ through the analytic Legendre Polynomial solution as described in section 4.2.1. The exception is the plot with infinite degree, which uses the square well trap with $L = 0.5\text{ m}$ as described in appendix A.2, i.e. a constant field with $B_0 = 0.96\text{ T}$ and infinitely high walls at $z = \pm L/2$.

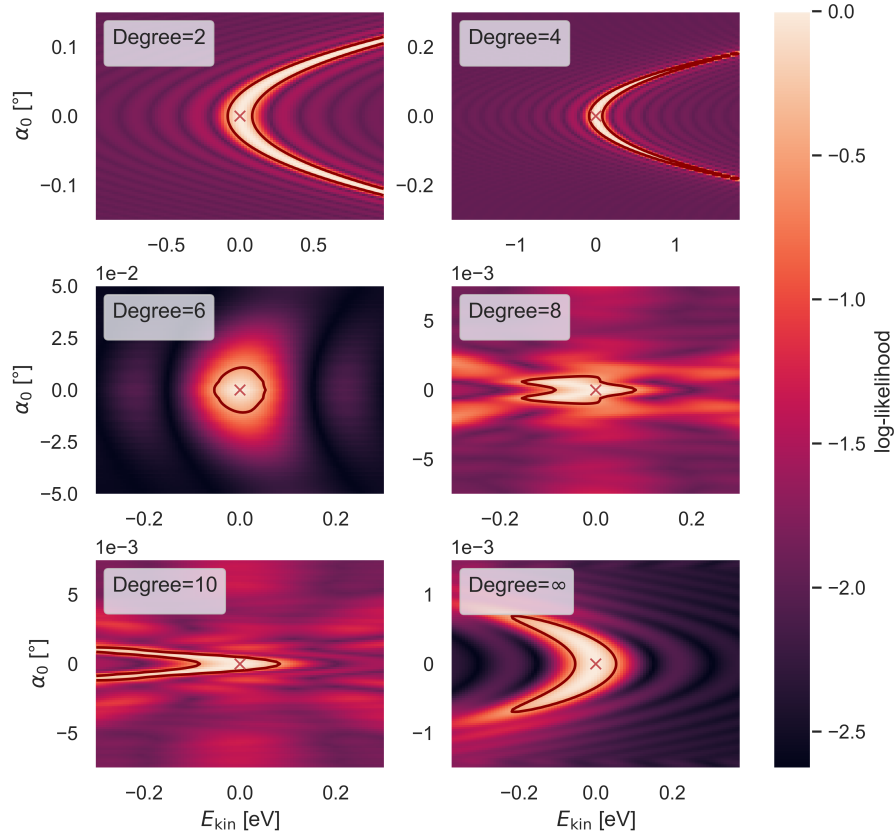


Fig. 8.1.: Scans of log-likelihood in the $E_{\text{kin}}-\alpha_0$ -plane produced with CRESana and various polynomial magnetic fields. See text for details on likelihood function and simulation settings. The red line indicates the 68% confidence level. The origin of the coordinate axes is set to the true parameter values (marked with "x"). Note the different axis scales for each plot.

The plots reveal non-linear correlations between kinetic energy and pitch angle. For a harmonic field (degree=2), the cyclotron frequency, which depends on both the kinetic energy and the magnetic field (see eq. (4.3)), creates an extreme case with a degenerate likelihood maximum along a curved line. In this harmonic case, the magnetic field varies with the electron's axial position, causing the average cyclotron frequency to depend on the pitch angle. Consequently, a small change in the pitch angle affects the signal similarly to a change in kinetic energy (see also fig. 5.18). The 68% confidence line in the plot shows that these ambiguous signals can extend to kinetic energies that are more than 1 eV above the true value.

Considering that the pitch angle is unknown, this ambiguity introduces an unacceptable uncertainty in the kinetic energy. Apart from the obvious drawbacks due to the higher energy uncertainty, such an extreme case of a non-linear correlation will also complicate the calculation of the energy resolution, as this is incompatible with the Gaussian assumption required for the Hesse errors.

The issue caused by the cyclotron frequency can be mitigated by reducing the dependence of B on the axial position. In the extreme case of a constant magnetic field (indicated with degree= ∞), the effect vanishes entirely because the cyclotron frequency remains constant for all pitch angles. However, in this scenario, the plot displays another type of correlation, likely due to the Larmor power, which directly influences the signal amplitude and the chirp rate, and depends on both the kinetic energy and the pitch angle (see eq. (4.24)). Yet, the added uncertainty on the kinetic energy from this correlation is notably smaller than in the harmonic case.

For the polynomial traps with degrees 2–10 both effects are present simultaneously. The first effect appears to dominate up to the $D = 6$ case, while the second effect becomes more prominent from $D = 8$ onward. The $D = 6$ scenario seems to be a sweet spot, where the first effect still dominates and visibly shapes the likelihood landscape. However, at the 68% confidence level, the two parameters appear uncorrelated.

During the development of experimental prototypes, correlations similar to the $D = 2$ case were found in simulations with realistic electromagnets that produce nearly harmonic field profiles [136]. This was an alarming discovery, as CRES is not a viable spectroscopy technique with these correlations. Consequently, there was a push for magnetic field profiles that are very flat in the central region.

Specific Field Configuration

The demo setup in this thesis uses a magnetic field developed within the Project 8 collaboration specifically to achieve a flat profile in the central region. The field is produced by eight electromagnetic coils, which are implemented as idealized current loops in CRESana. The electromagnets generate a negative trapping field that acts as a potential well when combined with the background field, which is set to $B_0 = 0.04$ T. This choice of magnetic field results in a cyclotron frequency of $f_c = 1.08$ GHz and a Larmor power of $P = 1.9 \times 10^{-18}$ W at the tritium endpoint. The smaller Larmor power produces negligible chirp rates, minimizing the correlation effect seen in the last plot. This will also benefit the analysis of the energy resolution by reducing the non-linear correlations.

The configuration parameters of the current loops are listed in table 8.1, and their three-dimensional arrangement is visualized in fig. 8.2a. Additionally, fig. 8.2b shows the produced magnetic field profile at three different radial positions. The maximum value of B , the background field, is only reached for $z \rightarrow \pm\infty$. Therefore, electrons with pitch angles near the lower trapping limit travel very long distances

Tab. 8.1.: Configuration parameters of the electromagnets that produce the trapping field

z [m]	R [m]	I [A]
-5.06	2.16	-824.2
-4.06	3.71	-1390.4
-2.63	4.83	-1805.6
-0.91	5.42	-2025.0
0.91	5.42	-2025.0
2.63	4.83	-1805.6
4.06	3.71	-1390.4
5.06	2.16	-824.2

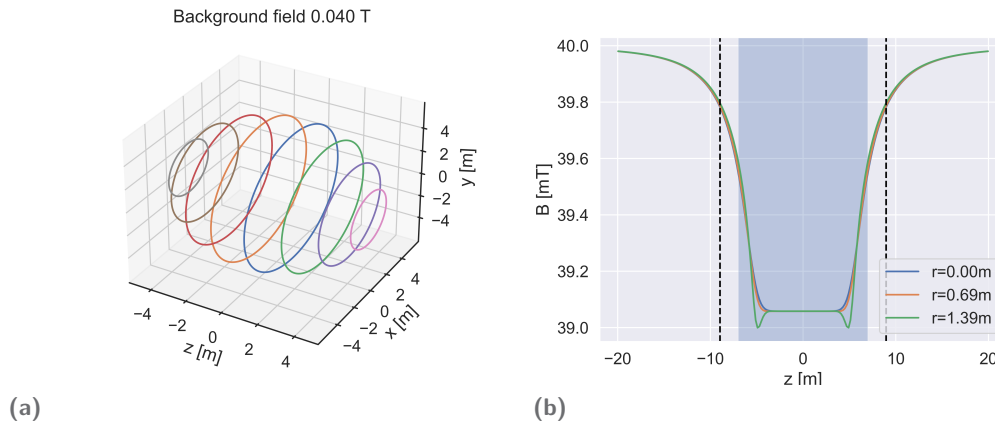


Fig. 8.2.: Magnetic trap for the demo setup. (a) shows the geometric arrangement of the 8 electromagnets producing the trapping field. (b) shows the absolute value $|B|$ of the field along the symmetry axis for three different radial positions r . The shaded blue region marks the extent of the antenna array, the dashed black lines the maximum z positions of electrons with $\alpha_0 = 82.17^\circ$.

before reaching their theoretical turning points. This issue can be addressed in future designs by creating potential barriers on top of the background field instead of a potential well.

Furthermore, fig. 8.2b reveals undesired side minima in the magnetic field at larger radii. These are problematic because they trap a subset of all electrons, which then produce signals different from those generated by electrons with identical parameters in the main minimum. This complicates the energy estimation, as it becomes necessary to distinguish between the two cases for each signal. In the worst case, this reduces the energy resolution on average if left unaccounted for.

While both of these issues are certainly non-optimal, the magnetic trap is still adequate for demonstrating the efficiency analysis of a setup. For example, for the

analysis in this chapter, it is sufficient to consider only electrons trapped in the main minimum, and a maximum travel distance can be set, which can be interpreted as a physical barrier in the experiment. Nevertheless, for more detailed neutrino mass sensitivity studies, these concerns will need to be addressed in future design iterations.

The magnetic field traps electrons with pitch angles greater than 81.17° . To set a reasonable limit to the axial travel distance, only electron's with pitch angles exceeding $\alpha_0 = 82.17^\circ$ are considered for the subsequent analysis, resulting in maximum turning points at $z = \pm 9.16$ m. The maximal radial position is limited to $r = 1.09$ m in the simulations to prevent electrons from approaching the coils too closely. This is necessary to maintain the adiabatic assumption that is fundamental for the simulation in CRESana. Given that the smallest coil has a radius of 2.16 m, this limit is quite conservative and was chosen to ensure robustness in the simulation. Despite this, the maximum physically allowed radial electron position would be $r = R$, the radius of the antenna array.

8.1.3 Effective Trap Volume

Given the constraints of the setup of antenna array and trap, one can define the tritium gas volume that contributes to the statistical sensitivity by producing trapped electrons via the electrons' axial travel distance and the antenna array radius. For a fixed pitch angle α_0 , the corresponding cylindrical gas volume is obtained through the evaluation of the volume integral

$$V(\alpha_0) = \int_0^{2\pi} \int_0^R \int_{-z_{max}(r,\alpha_0)}^{z_{max}(r,\alpha_0)} r \, dz \, dr \, d\phi.$$

With electrons originating from radioactive decays, α_0 and, by extension, $V(\alpha_0)$ are random variables. The expectation value of $V(\alpha_0)$ with respect to α_0 defines the effective trap volume V_t . This volume is equivalent to the effective volume under the assumption that all trapped electrons are detected. Thus, V_t provides an upper limit for the reachable effective volume by summarizing the effect of the trap alone, without the need for simulating signals. The expectation value is derived in appendix C.3, resulting in

$$V_t = 2\pi \int_0^R \int_{\alpha_{min}}^{\pi/2} \int_{-z_{max}(r,\alpha_0)}^{z_{max}(r,\alpha_0)} \sqrt{\frac{B(r,z)(1 - \sin^2(\alpha_0))}{B(r,0)\left(\frac{B(r,0)}{B(r,z)} - \sin^2(\alpha_0)\right)}} \sin(\alpha_0) r \, dz \, d\alpha_0 \, dr. \quad (8.2)$$

The evaluation of eq. (8.2) with the proposed magnetic trap yields $V_t = 18.43 \text{ m}^3$. Therefore, reaching the effective volume of $V_{\text{eff}} = 10 \text{ m}^3$ that is required for the 0.04 eV sensitivity goal, demands an average detection efficiency of $\epsilon = 0.54$. This is an ambitious goal for detection with low backgrounds and suggests that the setup might need to be physically enlarged in a future iteration.

8.2 Energy Resolution

The energy resolution of the demo setup introduced in the previous section is evaluated using the maximum likelihood approach discussed in section 7.2, similarly to the chirp model demonstration in section 7.2.4. Parameter uncertainties $\Delta\theta_i(\theta_{\text{true}})$ are calculated with the Hesse error in iminuit [132].

8.2.1 Likelihood Function and Relevant Parameters

For the analysis the log-likelihood function from eq. (7.19) is used with an Asimov dataset. The signal model in the likelihood function $\tilde{s}(\theta)$ and the Asimov dataset $\tilde{x} = \tilde{s}(\theta_{\text{true}})$ are based on simulations of the respective electrons in CRESana with the demo setup. The variance σ^2 is provided by eq. (5.15), assuming thermal noise with $T = 5 \text{ K}$.

The likelihood function is an 8-dimensional function of the unknown electron parameters $\theta = (E_{\text{kin}}, \alpha_0, r, \phi, \varphi_{\text{axial}}, \varphi_{\text{cycl}}, t_0, \tau)$. Although E_{kin} is the only parameter of interest, in general the full parameter space with all nuisance parameters must be incorporated for the estimation of the covariance matrix or the profile likelihood calculation, unless they have no correlation (direct or indirect) with the parameter of interest (see section 7.1.1). As demonstrated in section 7.2.4, t_0 and φ_{cycl} can be excluded due to the choice of $B = 0.04 \text{ T}$ and the choice of eq. (7.19) as the likelihood function, respectively. From the remaining nuisance parameters notable correlations with E_{kin} are only expected for the pitch angle α_0 and the radial position r , since both affect the mean magnetic field and the signal modulation. The track length τ only sets the end time of the signal defined as $t_e = t_0 + \tau$. Thus, it is correlated with t_0 but cannot possibly have a correlation with the remaining parameters.

The azimuth angle ϕ represents the initial phase of the electron's drift motion and as such it is correlated with the drift motion's angular frequency ω_D since both

contribute to the absolute drift phase. This absolute phase takes the typical form of

$$\varphi(t) = \int_0^t \omega(t') dt' + \varphi_0 = \bar{\omega}t + \varphi_0, \quad (8.3)$$

where $\bar{\omega}$ denotes the average of the instantaneous frequency $\omega(t')$ and φ_0 an initial phase. Under a Gaussian likelihood function, the correlation coefficient of $\bar{\omega}$ and φ_0 is easily calculated, yielding $\rho(\bar{\omega}, \varphi_0) = -\frac{\sqrt{3}}{2} \approx -0.87$, which then also applies to the correlation of ϕ and the average drift frequency $\bar{\omega}_D$. Consequently, ϕ is potentially correlated with E_{kin} , α_0 , and r , as ω_D depends on these parameters (see eq. (4.16)). Nevertheless, in practice, it has been found that the required flat field profiles produce predominantly small radial gradients. This results in minimal drift motion with $\bar{\omega}_D \rightarrow 0$, regardless of E_{kin} , α_0 , and r , making this correlation negligible.

Besides the drift phase, ϕ affects the signal through the static phase shifts² caused by the time delay at the i -th antenna, given by $\phi_i = \frac{d_i}{c}\omega_c$. The distance d_i between the electron's position and the antenna is a function of r and ϕ , which together determine the position in the x - y -plane. Hence, it is possible to keep d_i , and consequently the signal, constant by varying both r and ϕ . However, in the rotationally symmetric antenna arrangement, this correlation disappears when the signals of many channels are included in the likelihood function, as it becomes impossible to vary both parameters such that all d_i remain similar. This is visualized for five antennas in fig. 8.3.

Lastly, the initial axial phase φ_{axial} , together with the average axial frequency $\bar{\omega}_{\text{axial}}$, defines the absolute phase of the electron's axial motion through eq. (8.3). This relationship again yields a correlation of $\rho(\bar{\omega}_{\text{axial}}, \varphi_{\text{axial}}) = -\frac{\sqrt{3}}{2}$. The average axial frequency is given by $\bar{\omega}_{\text{axial}} = \frac{2\pi}{t(z_{\text{max}})}$, where $t(z)$ is the solution of the axial trajectory, which depends on E_{kin} , α_0 , and r according to eq. (4.6). While the axial motion itself is not an observable, it shapes the observed signal modulation. Thus, correlations between φ_{axial} and E_{kin} , α_0 , or r are possible through the axial frequency.

In summary, ϕ , φ_{cycl} , t_0 , and τ can safely be excluded from the likelihood function, whereas E_{kin} , α_0 , r , and φ_{axial} are potentially correlated with each other. Whether these correlations significantly contribute to ΔE depends on the magnetic trap setup. For example, in an instance using the squarewell trap with constant field profile (see appendix A.2), the inclusion of φ_{axial} was found not to contribute to ΔE in any capacity. Conversely, in an instance using the harmonic trap (see appendix A.1), its

²The likelihood function from eq. (7.19) is only invariant under uniform phase shifts applied to all antenna signals. However, it remains sensitive to relative phase shifts between individual antenna channels.

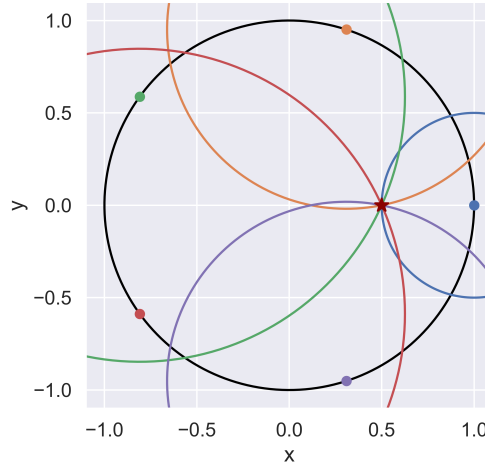


Fig. 8.3.: Illustration of distances between an electron and antennas in a rotationally symmetric array, in the x - y -plane. The red star marks the electron's position at $r = 0.5$ and $\phi = 0$. Distances d_i to the five antennas (colored dots) are represented by matching colored circles. Moving the electron along a circle keeps the corresponding d_i constant, but no r and ϕ combination can keep all d_i unchanged simultaneously.

exclusion underestimated ΔE by 15%. Unfortunately, the axial phase cannot be included in the likelihood function for the evaluation of the demo setup, since that parameter was only properly implemented for the analytic traps of CRESana³.

As a result of the considerations above, the energy resolution $\Delta E(\boldsymbol{\theta}_{\text{true}})$ is estimated with the reduced 3-dimensional likelihood function $\ell(E_{\text{kin}}, \alpha_0, r; \tilde{\mathbf{x}} = \tilde{\mathbf{s}}(\boldsymbol{\theta}_{\text{true}}))$.

8.2.2 Parameter Dependencies of the Energy Resolution

The last remaining open question concerns the relevant parameter dependencies of ΔE for the assumed true parameter values. This question differs from the discussion of the relevant parameters of the likelihood function in the previous section. The former focused on how ΔE changes with the same true signal depending on which parameters are assumed unknown, whereas the current discussion addresses how ΔE varies with different true signals.

³The axial phase was not prioritized during development, as it was never considered a relevant parameter for estimating the kinetic energy.

To address the parameter dependencies, it is illuminating to consult the theoretical expectation in the form of the CRLB. In the low field case, where $\text{Var}(\hat{t}_0) = 0$, the CRLB, as given by eq. (7.11), simplifies to

$$\text{Var}(\hat{\omega}_0) \geq 192 \frac{1}{\eta \tau^3}, \quad (8.4)$$

where η is the normalized SNR defined in eq. (7.10). Applying eq. (7.12) converts this lower limit for the initial frequency to the energy resolution. Therefore, ΔE only depends on τ , the dependencies of η , and E_{kin} due to the conversion with eq. (7.12). The dependencies of η are discussed in detail in section 8.3.1, where it is demonstrated that it mainly depends on α_0 and r with a small dependence on E_{kin} .

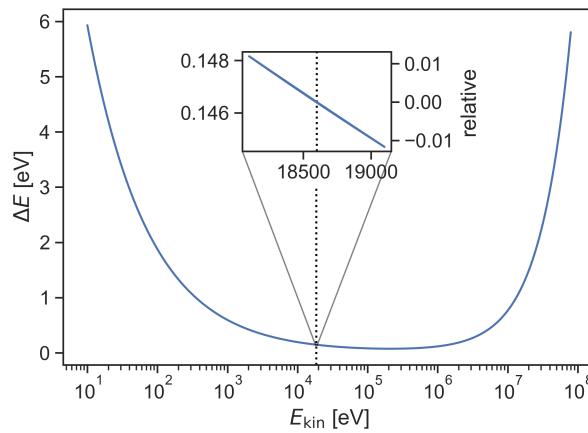


Fig. 8.4.: Energy resolution of a chirp for a $\alpha_0 = 90^\circ$ electron plotted as a function of the kinetic energy for $\tau = 1$ ms, $B = 0.04$ T, and $T = 5$ K. The inset plot shows a close-up around the endpoint of the tritium spectrum, which is marked by the dotted line. The second y-axis of the inset indicates the relative deviation from the value at the endpoint.

Figure 8.4 depicts ΔE as a function of E_{kin} , obtained using eqs. (7.12) and (8.4). For mildly relativistic energies, ΔE exhibits a broad, almost flat minimum, but increases rapidly in both the non-relativistic and highly relativistic limits. In the low energy regime, this increase is caused by a $\sqrt{\frac{1}{\beta^2} - 1}$ scaling in the Larmor power (see eq. (4.24)), and in the high energy regime it is the result of a E_{kin}^2 scaling in the frequency to energy conversion. The inset plot in fig. 8.4 shows that for ± 500 eV around the tritium endpoint in the valley region, ΔE varies by less than $\pm 1\%$, making ΔE approximately constant for the energies probed in a neutrino mass experiment. Thus, in summary, assuming the chirp model's CRLB, ΔE is considered to depend only on α_0 , r , and τ .

Even though the chirp model only approximates electrons with $\alpha_0 \rightarrow 90^\circ$, it is fairly simple to establish that the remaining parameters are also irrelevant for ΔE in the case of modulated signals produced with $\alpha_0 < 90^\circ$. The azimuth angle ϕ is irrelevant due to the rotational symmetry of the detector, and t_0 can be ignored because of the time invariance of the signals. Additionally, the two phases φ_{cycl} and φ_{axial} merely set arbitrary zero points of periodic signals that are observed over many cycles. While the exact signals do vary with these phases, the information contained in them remains independent of these zero points. In conclusion, ΔE is treated as a function of α_0 , r , and τ and is evaluated accordingly.

8.2.3 Result

The energy resolution of the demo setup was analyzed for the setup's four edge cases of (α_0, r) , each with track lengths ranging from 10–700 μs and using a fixed $E_{\text{kin}} = 18.6 \text{ keV}$. The results are presented in fig. 8.5, alongside a ratio plot that compares the results to the theoretic limit and the results of the Gaussian validation. The ratio plot is produced using the CRLB of the frequency of a pure sinusoid, which is given by [122]

$$\text{Var}(\hat{\omega}_0) \geq 6 \frac{\sigma^2}{\Delta t^2 A^2 (n_\tau^3 - n_\tau)} \approx 6 \frac{\sigma^2 \Delta t}{A^2 \tau^3} = \frac{12}{\eta \tau^3}. \quad (8.5)$$

This is a factor of 16 (a factor of 4 in ΔE) lower than the usual limit of the chirp. Remarkably, the results of the two $\alpha_0 = 90^\circ$ cases agree perfectly with this 4 times lower limit, which was unexpected. However, as expected, the $\alpha_0 = 82.17^\circ$ cases with modulated signals do not reach the lower limit predicted by their respective sinusoidal CRLB result⁴. The ratio of these cases is still nearly constant for τ , except for a couple of jumps that are again attributed to issues with Minuit (see section 7.2.4). This suggests that even in the modulated cases, the energy resolution has approximately the same scaling with τ as the original sinusoidal case.

Unfortunately, the validation results in the bottom panel appear discouraging. Only the $\alpha_0 = 90^\circ$ cases yield values of m'_{max} near 1, albeit with some larger deviations that might be incompatible with the Gaussian assumption underlying the Hesse error⁵. Moreover, the two cases with $\alpha_0 = 82.17^\circ$, especially with $r = 1.09 \text{ m}$, certainly seem to disagree with a Gaussian likelihood function.

⁴Each line in the ratio plot uses different values η for eq. (8.5). In other words, it would be incorrect to infer from fig. 8.5 that for the blue line ΔE is 1.3 times higher than the $\alpha_0 = 90^\circ$ cases.

⁵Reminder: For a perfectly Gaussian likelihood function m'_{max} would be 1.

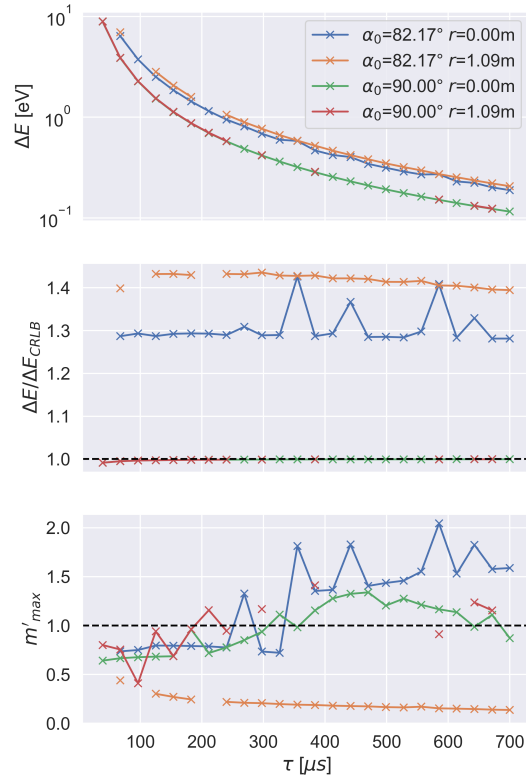


Fig. 8.5.: Results of energy resolution analysis of the demo setup as a function of track length τ . The first plot shows ΔE , the second the ratio to the CRLB calculated using eqs. (7.12) and (8.5) with the respective simulated $\eta(\alpha_0, r)$, and the last one the validation results (see section 7.2.3). Some data points are missing because Minuit occasionally failed to estimate the covariance matrix.

Likelihood Scans

To better understand the results, the likelihood function was inspected in the $E_{\text{kin}}-\alpha_0$, $E_{\text{kin}}-r$, and α_0-r -planes centered on θ_{true} (e.g., where $r = r_{\text{true}}$). The results for the four combinations of pitch angle and radius from fig. 8.5 are illustrated in figs. 8.6 and 8.7 with $\tau = 240 \mu\text{s}$. Each plot indicates the $1-\sigma$ error ellipse in the plane, derived from the estimated 3×3 covariance matrix, and the actual $1-\sigma$ confidence contour of the likelihood function, which ideally align closely. Furthermore, the plots display the contour of the validation result m'_{max} from the bottom panel of fig. 8.5, which has to intersect with the estimated 3D ellipsoid somewhere by the way it is calculated.

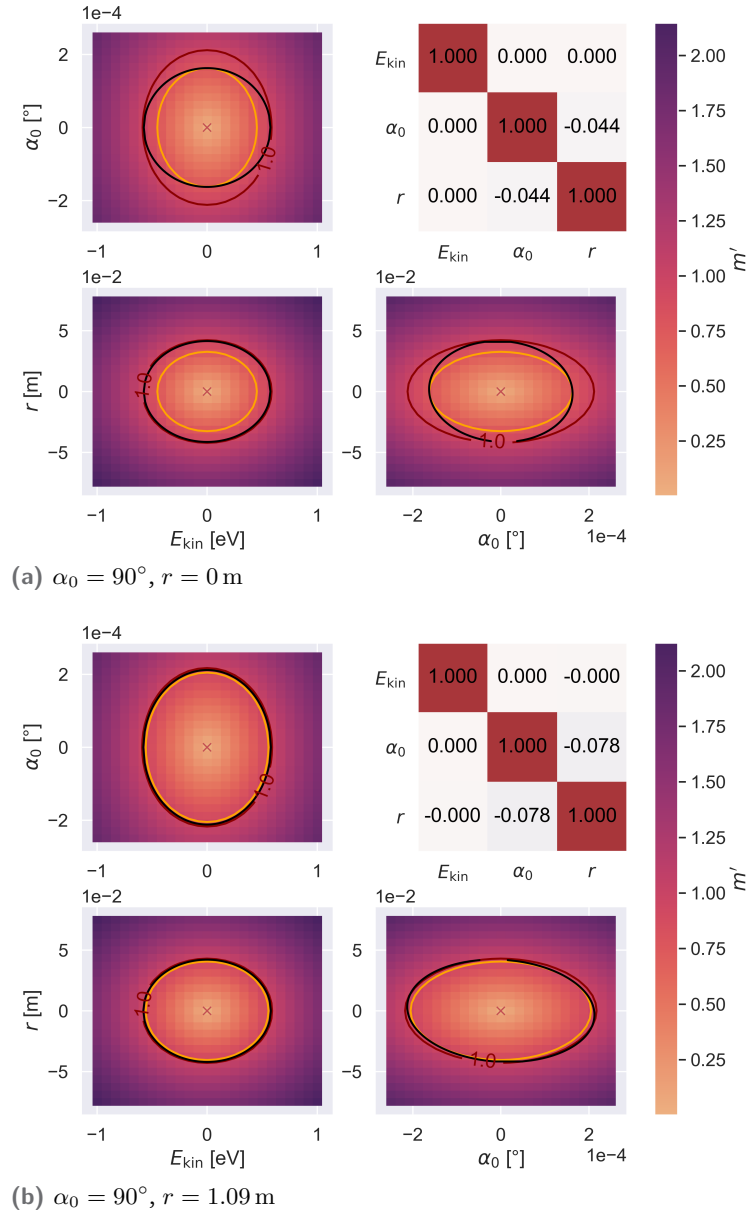


Fig. 8.6.: Scans of likelihood of the demo setup in the three 2D-planes for $\alpha_0 = 90^\circ$ and $\tau = 240 \mu\text{s}$. Likelihood values are converted to m' - σ levels (see eq. (7.16)). Red lines indicate the $1\text{-}\sigma$ confidence level, black lines the estimated $1\text{-}\sigma$ error ellipses, orange lines the contour of the validation result m'_{max} . The coordinate origin is shifted to the true parameter values (marked with "x"). Top right depicts estimated correlation matrix with colors ranging from blue (perfect anti-correlation) to red (perfect correlation).

However, the reported values of m'_{max} are not always obvious from the plots, as the intersection is not visible in all cases. This issue is particularly noticeable in the off-center and low pitch angle case (fig. 8.7b), where $m'_{max} \approx 0.2$. The corresponding contour, a barely visible circle in the center of the plot, is nowhere

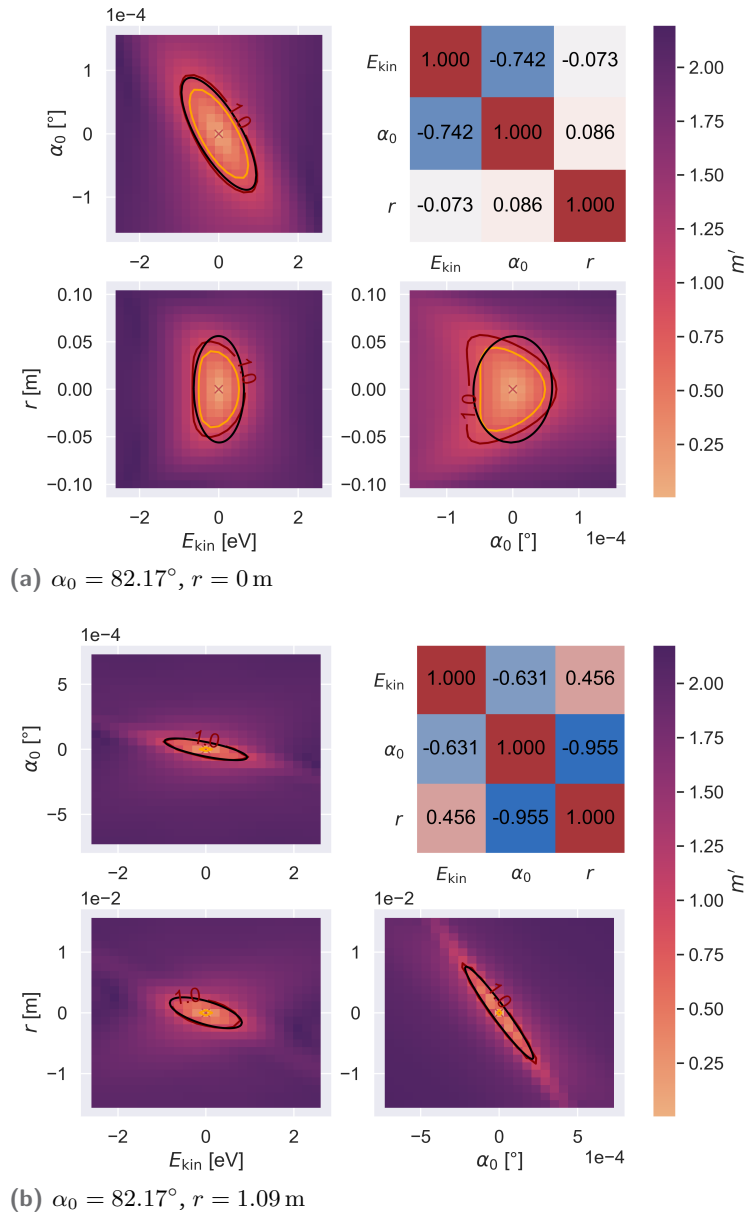


Fig. 8.7.: Scans of likelihood of the demo setup in the three 2D-planes for $\alpha_0 = 82.17^\circ$ and $\tau = 240 \mu\text{s}$. See fig. 8.6 for detailed description.

near the estimated ellipses. This discrepancy is most likely due to the fact that the plots only show slices of the three-dimensional ellipsoid. Additionally, the value m'_{max} is determined in the coordinate planes of the eigenbasis, which might not coincide with the planes of the physical coordinates. Further investigation of this case revealed that for most points examined by the validation routine, m' ranges from 0.8 to 0.9, with two outliers yielding values of 0.2 and 0.64. The validation routine places a high weight on such outliers by design. Ultimately, it is unknown why there is a single point in the eigenbasis where the likelihood function deviates so

far from the Gaussian assumption. To better understand these significant deviations in the validation results, it might be informative to conduct scans of the likelihood function in the coordinate planes of the eigenbasis instead. For now, however, it seems more appropriate to compare the estimated ellipses to the $1\text{-}\sigma$ likelihood contours.

A noteworthy deviation between the estimated ellipses and the contours is apparent only in the low pitch angle case with $r = 0$ (fig. 8.7a), where the contours in the planes involving r are not elliptical. This can be explained by the fact that $r = 0$ is technically on the edge of the parameter space. However, limiting the likelihood function at $r = 0$ prevents Minuit from successfully estimating a covariance matrix. Therefore, negative values were allowed, corresponding to electron positions in the opposite half of the cross-sectional plane with azimuth angle $\phi + \pi$. Nevertheless, the estimated ellipses approximate the actual contours sufficiently well, even though they are not elliptical. For the remaining cases, the ellipses match the contours very well, with almost perfect agreement in the off-center case with small pitch angle (fig. 8.7b) and only minor deviations in the two $\alpha_0 = 90^\circ$ cases. (fig. 8.6).

Additional results for $\tau = 671\ \mu\text{s}$ are presented in appendix D.3. The observations are mostly the same in this case. Except for a slight misestimation of the correlation between r and α_0 in the case $\alpha_0 = 90^\circ$, $r = 1.09\ \text{m}$, even the estimated correlation matrices are similar.

Alternative CRLB

Overall, the likelihood scans mostly justify the Gaussian assumption, despite the validation results. Therefore, the findings for ΔE in fig. 8.5 are rated as mostly accurate, including the surprisingly small values that are compatible with the CRLB of a pure sinusoid. A common hypothesis to explain this outcome suggests that this is a unique feature of low magnetic fields: with a field of $B = 0.04\ \text{T}$, the chirp rate is close to zero, and thus the signal is compatible with a pure sinusoid, making the smaller CRLB appropriate. However, using a trap with $B = 0.96\ \text{T}$, in which the signal has a considerable chirp rate, it was found that the results are compatible with the lower sinusoid CRLB as well in the $\alpha_0 = 90^\circ$ case.

These findings point at a crucial misconception in utilizing the chirp model for the derivation of the estimation limit: it is based on the assumption that the initial frequency and the chirp rate are two independent free parameters. In reality, both are functions of the electron parameters and the trap. In a setting where both

parameters are fully determined by the kinetic energy, the chirp rate is not a free parameter anymore but essentially a function of the frequency.

The CRLB for such a chirped sinusoid, where the chirp rate is a dependent parameter, is derived in appendix C.4 and is given by

$$\text{Var}(\hat{\omega}_0) \geq \text{Var}(\hat{t}_0) \delta\omega^2 + 6 \frac{\sigma^2}{\Delta t^2 A^2 (n_\tau^3 - n_\tau)}.$$

This is indeed the CRLB of the pure sinusoid with an additional contribution from the start time uncertainty. The latter can be neglected again due to the small chirp rate at $B = 0.04$ T.

Whether the chirp rate and the initial frequency are both fully determined by the kinetic energy depends on the magnetic trap and its parameter correlations. The fact that the demo setup exhibits no parameter correlations at all in the $\alpha_0 = 90^\circ$ case further supports that the limit of the pure sinusoid is appropriate. In contrast, it was confirmed that ΔE only achieves the higher CRLB of the chirp in the square well trap, which has a correlation between E_{kin} and α_0 at $r = 0$, $\alpha_0 = 90^\circ$, as seen in fig. 8.1.

So far the discussion of the results has focused on the cases of $\alpha_0 = 90^\circ$. For smaller pitch angles, the sinusoidal CRLB is not achieved, as expected due to the modulated signals. Nevertheless, the ratio plots of these cases displayed in fig. 8.5 are nearly constant in τ , indicating that the lower limits of the modulated signals only differ from the pure sinusoidal limit in their scaling constants. As a result, it can be assumed that the energy resolution for any electron track with parameters θ can be expressed as

$$\Delta E(\theta) = \frac{c(\alpha_0, r)}{\sqrt{\tau^3}}. \quad (8.6)$$

The scaling constant is defined by the parameter correlations of the setup and can be determined solely through numerical analysis. Consequently, the final result for a setup is obtained by evaluating its energy resolution $\Delta E(\alpha_0, r, \tau)$ on a sufficiently dense grid in α_0 and r , using a fixed track length τ . Subsequently, $c(\alpha_0, r)$ is calculated and turned into an interpolation function. While theoretically a single track length τ suffices for this procedure, using the mean of multiple track lengths can improve the accuracy.

8.3 Detection Efficiency

The detection efficiency of the demo setup is analyzed using the methodology outlined in section 6.3.5, assuming that a quadrature matched filter will be applied (eq. (6.20)). The relevant metric that characterizes the detection performance for an electron with parameters θ is its matched filter SNR defined via an inner product as $SNR_{MF}(\theta) = \frac{2}{\sigma^2} \tilde{s}(\theta)^H \tilde{s}(\theta)$ (see eq. (6.29)). This only requires the simulation of the electron's signal matrix \tilde{s} in CRESana.

8.3.1 Parameter Dependencies

In general the SNR can vary with the electron parameters, which are given by $\theta = (E_{\text{kin}}, \alpha_0, r, \phi, \varphi_{\text{axial}}, \varphi_{\text{cycl}}, t_0, \tau)$. However, the SNR is independent of the start time t_0 and the cyclotron phase φ_{cycl} , since the inner product is invariant under translations and phase shifts of the signal. The same is true for the polar phase ϕ of the electron's position in the cross-sectional plane due to the rotational symmetry.

Additional electron parameters can be eliminated using eq. (7.10), which states that $SNR_{MF} = \eta \cdot \tau$, where the normalized SNR η is proportional to the summed RMS power of all channels P_T . This power is determined by the electron trajectory, the antenna array efficiency and the electron's emitted Larmor power. The antenna efficiency can be considered independent of the energy due to the nearly flat frequency response (see fig. 5.8). Moreover, the Larmor power can be treated as constant over the small range of energies relevant to a neutrino mass measurement, as it varies by just $\pm 3\%$ across a window of ± 500 eV around the tritium endpoint. Variations of P_T over time, and any offset caused by the initial axial phase φ_{axial} are negligible when the RMS is taken over many cycles of the axial motion. Therefore, P_T varies only with α_0 and r due to the influence of the electron trajectory. In conclusion, the SNR scales linearly with τ and η , which primarily depends on α_0 and r with a minor effect of E_{kin} .

8.3.2 Result

Electron signals of the demo setup were simulated in CRESana with pitch angles ranging from 82.17 – 90.0° and radii ranging from 0.0 – 1.09 m on a 51×51 grid. For each signal the total RMS power P_T was calculated and then converted to the normalized SNR η using eq. (7.10). The results of this procedure are displayed in

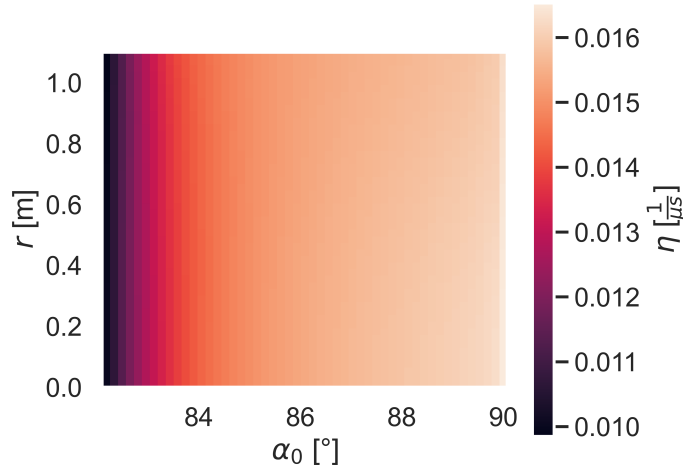


Fig. 8.8.: Normalized SNR η represented by color as a function of radius and pitch angle. Each pixel is calculated using eq. (7.10) with a $100\mu\text{s}$ long signal of the demo setup simulated in CRESana.

fig. 8.8 for an assumed noise temperature of $T = 5\text{ K}$. The plot demonstrates that reaching an SNR of 1 requires electron tracks that are at least $100\mu\text{s}$ long.

For all signals, observed RMS powers range from 3.4×10^{-19} – $5.7 \times 10^{-19}\text{ W}$, corresponding to 18–30 % of the emitted Larmor power in this scenario. This suggests that the assumed antenna array is quite efficient and operating near the theoretical limit. Surprisingly, the dependence on r is very weak. Conversely, simulations with the fabricated 26 GHz prototype antenna showed a notable reduction in detected power for high radii. This is most likely due to the use of the idealized isotropic antennas in the demo setup. For beam antennas with high directivity in the cross-sectional plane, on the other hand, electrons at higher radii are in the field of view of fewer antennas, resulting in a decrease in total detected power. Figure 8.9 illustrates this effect. This reinforces the fact that the obtained results are an optimistic idealization. Nevertheless, once a design for a 1 GHz antenna is available, more refined results can easily be produced with the same procedure.

By turning the simulation results into an interpolation function $\eta(\alpha_0, r)$, the matched filter SNR of any electron track with parameters θ can be evaluated quickly without running additional simulations employing

$$SNR_{MF}(\theta) = \eta(\alpha_0, r) \tau. \quad (8.7)$$

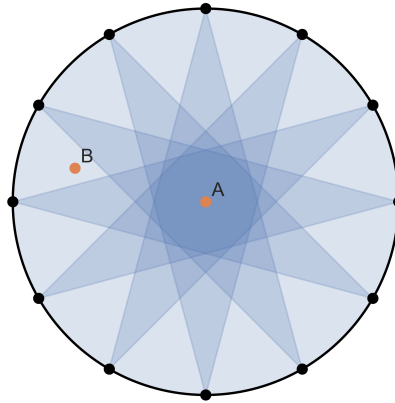


Fig. 8.9.: Schematic of a non-isotropic antenna array's sensitivity to radiation in the cross-sectional plane. Black dots represent antennas aimed at the center, each with a 30° -wide binary gain. A transparent blue light cone marks the gain=1 region of each antenna, with overlapping cones shown in deeper blue. At position A, all antennas are sensitive to the radiation, while at position B, most antennas are not. Note that this binary gain is a simplified, unphysical model for illustration.

Finally, the electron track's detection probability P_D is given by eq. (6.36) as a function of the false alarm probability P_{FA} . In practical applications, the matched filter concept is implemented using a template bank to account for the unknown parameters. For this reason, the detection probability given in eq. (6.36) also depends on the number of required templates n_t . The theoretical upper limit on P_D can always be determined using the optimistic assumption that $n_t = 1$, which essentially means assuming that the parameters are known in advance.

Figure 8.10 depicts the estimated detection probability as a function of P_{FA} and τ for the four corner cases of the simulated α_0 - r grid for the optimistic $n_t = 1$ assumption. The results show that electron detection is feasible with the demo setup with the required probability of 54%. This is attained for all examined electron parameters with track lengths of $\tau = 3500 \mu s$, assuming rigorous false alarm probabilities ($P_{FA} < 10^{-8}$). In cases where $\alpha_0 = 90^\circ$, detection is nearly certain ($P_D = 0.99$) with relaxed requirements for P_{FA} and long tracks.

Background Rate

Nevertheless, the significance of these results is limited without a requirement for the false alarm probability, which is given indirectly through the number of allowed

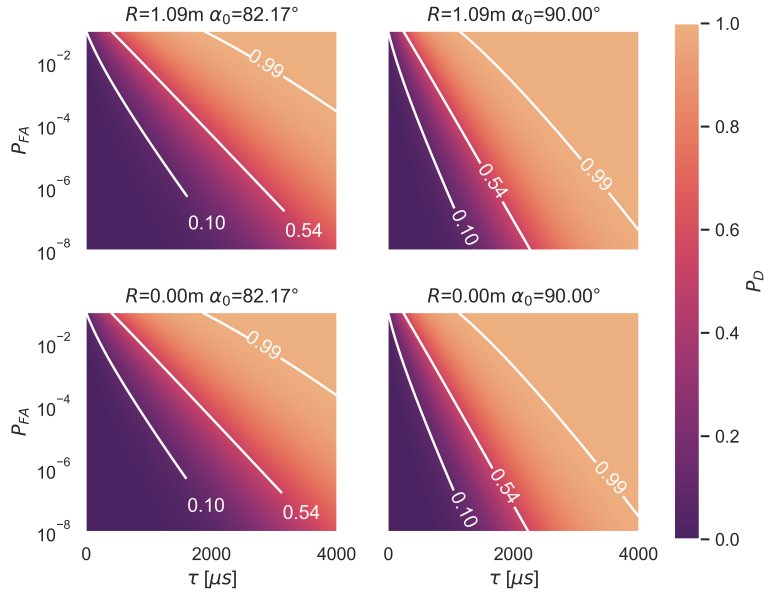


Fig. 8.10.: Detection probability P_D color-coded against track length τ and false alarm probability P_{FA} of simulated electrons in the demo setup for $n_t = 1$. Each plot represents a corner case for radius and pitch angle, as indicated by the subtitle. Contour lines indicate detection probabilities of 0.99, 0.54, and 0.1.

background events over the experimental runtime. This is incorporated into the neutrino mass sensitivity estimation as a background rate parameter b , which is the number of background events per unit of time and unit of energy. Hence, the background rate relates to the probability of false alarm of the entire template bank P_{FA} via

$$b = \frac{P_{FA}}{\tau E_{BW}}, \quad (8.8)$$

where E_{BW} is the energy bandwidth covered by the template bank. The probability P_{FA} depends on the probability of false alarm of an individual template P'_{FA} as defined in eq. (6.35). For $P'_{FA} n_t \ll 1$ the following approximation applies

$$P_{FA} = 1 - (1 - P'_{FA})^{n_t} = \left(1 - \left(1 - n_t P'_{FA} + \mathcal{O}(P'^2_{FA})\right)\right) \approx n_t P'_{FA}. \quad (8.9)$$

Unsurprisingly, the background rate depends on the number of statistically independent templates n_t . The construction of a template bank and estimation of the number of required templates are topics that require an in-depth treatment and are not explored in detail in this thesis. Instead, a simple heuristic is applied, assuming that the template bank evaluates n_E equidistant energies with n_θ templates to cover all remaining parameters for each energy such that $n_t = n_E \cdot n_\theta$. By utilizing dis-

tances of $2\Delta E$ between energies, any given electron signal falls within the margin of error of a probed energy. Since it was proven that ΔE is approximately constant with the kinetic energy (see section 8.2.2), it follows that the relationship $n_E = \frac{E_{BW}}{2\Delta E}$ holds⁶.

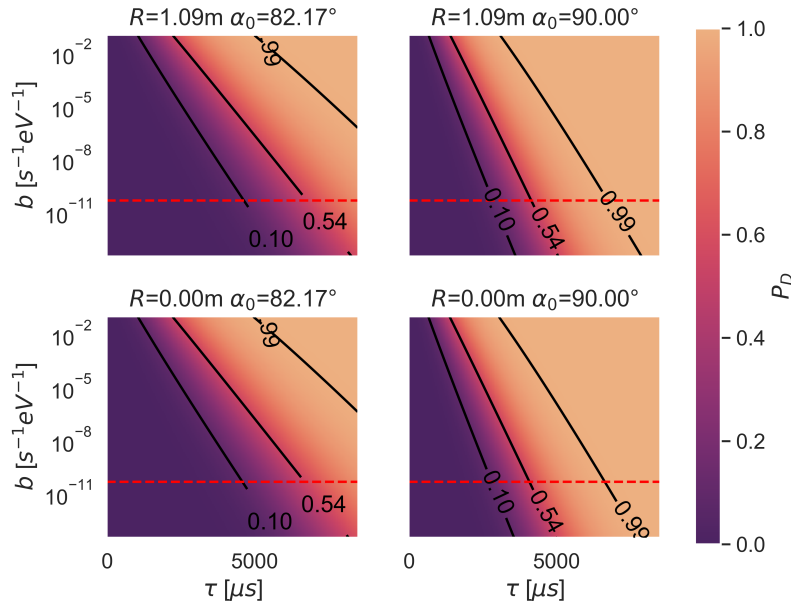


Fig. 8.11.: Detection probability P_D color-coded against track length τ and background rate b of simulated electrons in the demo setup for $n_\theta = 1$. Each plot represents a corner case for radius and pitch angle, as indicated by the subtitle. Contour lines indicate detection probabilities of 0.99, 0.2, 0.1 and 0.02. The dashed red lines mark $b = 4.7 \times 10^{-11} \text{ s}^{-1} \text{ eV}^{-1}$.

Applying this heuristic for n_t and substituting P_{FA} with the approximation from eq. (8.9) and ΔE with the expression from eq. (8.6) into eq. (8.8) results in

$$b = \frac{n_\theta P'_{FA}}{2\tau \Delta E} = \frac{n_\theta P'_{FA} \sqrt{\tau}}{2c(\alpha_0, r)}. \quad (8.10)$$

Rearranging eq. (8.10) for P'_{FA} yields the probability of false alarm of a single template as a function of b , which can be inserted into eq. (6.33), producing the detection probability of the template bank as a function of b :

$$P_D(b) = S_{\chi^2_2(\eta(\alpha_0, r)\tau)} \left(-2 \log \left(\frac{2b c(\alpha_0, r)}{n_\theta \sqrt{\tau}} \right) \right). \quad (8.11)$$

⁶The heuristic neglects the fact that n_θ could depend on the energy and that ΔE depends on r and α_0 .

Figure 8.11 shows the estimated detection probability as a function of b and τ . Without going into further details about the construction of the template bank, it is again optimistically assumed that $n_\theta = 1$.

8.4 Implications and Limitations of Results

8.4.1 Validation of Hesse Errors

A significant concern in the analysis of the energy resolution was the validity of applying Hesse errors. This concern arose due to the observation of distinctly non-linear correlations in likelihood scans for certain setups, such as those illustrated in fig. 8.1, which are incompatible with a Gaussian distribution of maximum likelihood estimates.

The proposed validation procedure outlined in section 7.2.3 aimed to assign a singular metric to each ΔE data point to assess its reliability. Unfortunately, the results of this validation did not strongly support the use of Hesse errors. However, this might suggest a design issue with the validation routine rather than a fundamental disagreement with the Gaussian likelihood assumption. A deeper understanding of these unfavorable results through an analysis of the likelihood function in the eigenbasis might help fix the validation routine and could be explored in future work.

Nevertheless, for now, the 2-dimensional likelihood scans, as depicted in figs. 8.6 and 8.7, appear to be a more effective diagnostic tool. For the demo setup, these scans are compatible with the assumed parabolic log-likelihood, thus supporting the application of Hesse errors. However, it is important to emphasize that in the analysis of any new experimental design, this conclusion must be reevaluated, as it may not generalize. The downside of applying this approach is the reduced convenience in interpreting results since one has to assess several plots per data point instead of a single number.

Moreover, the computational cost for generating these likelihood plots is quite high: depending on the parameters, generating a single plot using 18 CPU cores took from 2–6 h for the shorter track length (see figs. 8.6 and 8.7), and from 13–19 h for the longer one (see figs. D.5 to D.8). Given the close similarity of the likelihood landscape and the correlation matrix of the demo setup at the two different track lengths, however, it does not appear likely that these results would be highly dependent on

τ . Thus, it is recommended to generate these plots for the four corner cases of the α_0 - r -plane using just one or two track lengths.

8.4.2 Track Length Scaling and CRLB

A major achievement of the study in section 8.2 was the identification of eq. (8.6) as a simple scaling relation for the track length dependence of ΔE , which is also valid for $\alpha_0 < 90^\circ$. Furthermore, it was recognized that the lower limit of ΔE for $\alpha_0 = 90^\circ$ is in general determined by the CRLB of a pure sinusoid, which is four times better than the expected result of a linear chirp. The most plausible explanation is that the chirp rate is fully determined by the initial cyclotron frequency and not a free parameter. However, it must be emphasized that this result should not be assumed to hold for all experimental configurations.

Turning to the limitations of the results, it must be reiterated that the likelihood function did not include the initial axial phase φ_{axial} in the calculation of ΔE due to limitations of the current version of CRESana. It was proven with a harmonic trap that φ_{axial} can potentially contribute to ΔE , unlike the other neglected parameters. Therefore, its exclusion could lead to overly optimistic results.

Fortunately, this effect does not contribute to ΔE for electrons with a 90° pitch angle in the demo setup. This is because, for these electrons, indirect correlations of E_{kin} with φ_{axial} through α_0 and r are irrelevant, as E_{kin} is not correlated with these parameters (see fig. 8.6). Additionally, the correlation of φ_{axial} with E_{kin} itself is negligible since the axial frequency depends only weakly on the kinetic energy. For this reason, it can be concluded that the exclusion of φ_{axial} is not the reason for achieving a better energy resolution than the chirp CRLB would predict.

Nevertheless, the effect of φ_{axial} could still be relevant for smaller pitch angles. Generally, this effect is setup-dependent and unfortunately cannot be quantified without properly including the parameter in the analysis. Still, it seems unlikely that this effect would change ΔE by an order of magnitude, as it was found to contribute only 15 % of the total energy resolution in a harmonic trap.

The scaling relation in eq. (8.6) is not applicable if the two-dimensional plots of the likelihood function indicate non-Gaussian behavior. In that case, the only option is to calculate ΔE with Minos errors for each individual track length τ . Furthermore, the MLE is only Gaussian in the asymptotic limit, meaning that Hesse errors will not be accurate for tracks shorter than a certain length τ_g . Additionally, below a track length τ_∞ , it becomes impossible to determine an uncertainty at all because the low

SNR prevents the profile likelihood from ever falling below the required confidence threshold.

The minimum track lengths τ_s , and τ_∞ for these cases are unknown and can only be inferred by calculating the profile likelihood (Minos errors). Any analysis requiring Minos errors is highly undesirable, considering that computing each data point in fig. 8.5 already took up to 11 h to compute with the computationally relatively cheap Hesse errors. Fortunately, the issues at short track lengths are fairly irrelevant since the high requirements for ΔE are only reached with long track lengths. Nonetheless, for an ensemble analysis that takes the track length distribution into account, this means that the low end of the distribution is technically not correctly represented.

8.4.3 Neutrino Mass Sensitivity

The outcome of the studies performed in this chapter provides a preliminary assessment of the demo setup's performance. In prior studies it was demonstrated that a scenario compatible with achieving the $m_\beta > 0.04$ eV design sensitivity limit of Project 8 has an energy resolution of $\Delta E = 0.115$ eV [58], an effective volume of $V_{\text{eff}} = 10$ m³ [59] (which necessitates an average detection probability of 54% for the demo setup, as demonstrated in section 8.1), and a background rate of $b = 4.7 \times 10^{-11}$ s⁻¹ eV⁻¹ or less [66].

The data from the demo setup indicate that these goals can be achieved for all electrons without requiring unreasonably long tracks, needing only a few milliseconds instead of seconds. Specifically, the data in fig. 8.5 indicate that the energy resolution goal can be met if track lengths exceed $\tau \gtrsim 1000 \mu\text{s}$. Furthermore, the detection efficiency target is achieved without exceeding the background rate limit for electrons with track lengths of $7000 \mu\text{s}$ (see fig. 8.11). However, while these conclusions do look promising, the track lengths depend on other experimental parameters. Therefore, a meaningful assessment of the results requires a proper ensemble analysis using the track length distribution within the context of a neutrino mass sensitivity study.

8.5 Summary

This chapter presented a framework for estimating the limits of detection efficiency and energy resolution of individual electrons with parameters θ in a large antenna array, based on numerical simulations with CRESana. In both cases, the result is not

a single number but is instead parameterized by α_0 , r , and τ . The energy resolution is produced by analyzing the Hesse errors of simulated Asimov datasets on a two-dimensional grid in α_0 and r using a fixed track length τ_0 . The energy resolution scaling constant $c(\alpha_0, r)$ is evaluated and interpolated over the two-dimensional grid and can subsequently be applied in eq. (8.6) to yield the energy resolution $\Delta E(\boldsymbol{\theta})$ for any electron. For the detection efficiency, the normalized SNR $\eta(\alpha_0, r)$ is evaluated on a two-dimensional grid in α_0 and r using signal simulations with a fixed track length τ_0 and eq. (7.10). Interpolation of the result and the application of eqs. (8.7) and (8.11) yield a function $P_D(\boldsymbol{\theta}; b)$ to calculate the detection probability of any electron with parameters $\boldsymbol{\theta}$ for a required background rate b .

The main highlights of the results include the identification of a track length scaling relation for modulated CRES signals that is similar to the CRLB. Additionally, the analysis revealed that the CRLB of the standard chirp model is four times higher than the true lower bound. Lastly, an initial evaluation indicates that the results match the requirements for achieving the neutrino mass sensitivity goal, though further analysis is necessary.

Conclusion and Outlook

” *Some writers enjoy writing, I am told. Not me. I enjoy having written.*

— **George R. R. Martin**
Fantasy author

Antenna arrays are a promising avenue for large-volume detectors in high-statistics and high-resolution experiments using the CRES technology. However, given the expected high data rates and associated significant computational costs, they are not without challenges. In this thesis, techniques for algorithmic detection and parameter estimation of CRES signals within the noisy data generated by such detectors were explored. The primary focus was on establishing optimal solutions to these problems, specifically through matched filtering and maximum likelihood estimation. This includes the implementation of these methods to evaluate their efficiency limits using simulated data for a specific experimental design. Since simulating experimental designs with thousands of antenna channels was infeasible with the processing times of the existing simulation software Locust, the new simulation tool CRESana was developed to support the primary objectives. By employing well-motivated approximations, CRESana achieved an approximately 500-fold speedup compared to Locust, without compromising simulation quality.

Chapter 8 of this thesis integrated all these aspects to demonstrate the performance analysis of an idealized example setup applicable in a neutrino mass measurement using tritium β -decay spectroscopy in the Project 8 experiment. The results of this analysis are the detection efficiency as a function of the noise-induced background rate and the energy resolution, both of which are determined for individual electrons. These are crucial parameters as they directly affect the achievable neutrino mass sensitivity of Project 8. An initial evaluation suggested that the obtained results are compatible with the requirements of reaching Project 8's sensitivity goal of $m_\beta > 0.04$ eV.

The analysis procedure presented in chapter 8 is currently being applied in a detailed sensitivity study of a similar experimental setup for inclusion in an upcoming publication [135]. This simulation-based assessment of an experimental design

represents the first study of its kind and was made possible only by the significant improvements in simulation time achieved with CRESana. In all previous sensitivity studies, detection efficiency, background rate and energy resolution were incorporated through fixed assumptions, theoretical predictions from simplified models, or declared targets that were independent of each other.

Additionally, early design work on the next prototype experiments for Project 8 relied on the efficiency of suboptimal and underperforming detection algorithms as design objectives, which led to some problematic design decisions and false assumptions about the viability of a neutrino mass experiment based on antenna arrays. With the presented analysis, it is now possible to optimize upcoming experimental designs with respect to the true theoretical limits of detection efficiency and energy resolution, fully utilizing the available signal information.

Looking ahead, there are numerous opportunities to extend the presented work. While this dissertation primarily focused on establishing optimal solutions, many challenges remain regarding the practical aspects of computationally feasible solutions in real data-taking environments, which may need to operate in real time. For example, as discussed in section 6.4.2, a brute-force evaluation of the matched filter template bank for a Project 8 Phase IV experiment was estimated to be infeasible with the currently available computing resources. Therefore, areas of interest include not only smart approaches to constructing and evaluating the matched filter template bank but also effective minimization strategies for parameter estimation and the application of machine learning techniques. Many of these research areas could not be sufficiently explored for large-scale antenna arrays before, due to the lack of benchmarks for their effectiveness and a feasible way of generating data sets, which are now provided by the theoretical limits and CRESana. In this light, the continued development of CRESana should also be a focus moving forward.

Despite the many research opportunities utilizing CRESana, it should be noted that more recently, the Project 8 collaboration decided to retire the antenna array concept for the time being in favor of the resonant cavity concept (as discussed in section 3.1.1) [95], primarily due to the high expected monetary costs of constructing an experiment with hundreds to thousands of antennas. As a result, the immediate benefit of the fully developed framework from chapter 8 is diminished for the future experimental phases of Project 8. While the methodology for estimating the theoretical limits of detection efficiency and energy resolution is applicable regardless of the detector technology and physics use case, many parts of CRESana's existing simulation technique cannot be applied to a cavity detector, as described in

section 5.4.7. Therefore, in future phases, this analysis procedure will require an alternative source of simulated data tailored to the new detector concept.

The switch in the detector concept unfortunately transforms CRESana's greatest advantage — its highly specialized and efficient simulation system — into a significant limitation for Project 8. Nevertheless, while the research for this dissertation has been conducted within the context of the Project 8 experiment and neutrino mass determination, it is by no means limited to this application. CRES, in itself, represents a new technology for high-resolution spectroscopy of charged particles. Several other experimental efforts are already adapting CRES for various use cases, including the He6-CRES [73], QTNM [137], and PTOLEMY [138, 139] experiments, as well as potential applications in X-ray spectroscopy [140]. These initiatives are mostly in the early stages of experimental development, where detector technologies are still evolving. While only QTNM — incidentally also a tritium-based neutrino mass experiment — is currently pursuing an antenna array detector concept, the growing adoption of CRES highlights the potential for the simulation techniques developed in this thesis to be applied in experiments beyond Project 8.

Bibliography

- [1] B. Monreal and J. A. Formaggio. “Relativistic cyclotron radiation detection of tritium decay electrons as a new technique for measuring the neutrino mass”. In: *Phys. Rev. D* 80 (5 Sept. 2009), p. 051301 (cit. on pp. 1, 22).
- [2] Particle Data Group, P. A. Zyla, R. M. Barnett, et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020), p. 083C01. eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf> (cit. on pp. 1, 14).
- [3] E. Roulet and F. Vissani. “The Neutrino Story”. In: *Neutrinos in Physics and Astrophysics*. WORLD SCIENTIFIC, Oct. 2022. Chap. 1, pp. 1–20. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789811260940_0001 (cit. on p. 1).
- [4] I. Esteban, M. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou. “The fate of hints: updated global analysis of three-flavor neutrino oscillations”. In: *Journal of High Energy Physics* 2020.9 (Sept. 2020) (cit. on pp. 1, 7, 9, 16).
- [5] *NuFIT 5.2*. 2022. URL: <http://www.nu-fit.org/> (visited on Dec. 8, 2023) (cit. on pp. 1, 7, 9, 16).
- [6] M. Aker, D. Batzler, A. Beglarian, et al. *Direct neutrino-mass measurement based on 259 days of KATRIN data*. 2024. arXiv: 2406.13516 [nucl-ex] (cit. on pp. 1, 10, 16, 23).
- [7] V. Gligorov. “Real-time data analysis at the LHC: present and future”. In: *Proceedings of the NIPS 2014 Workshop on High-energy Physics and Machine Learning*. Ed. by G. Cowan, C. Germain, I. Guyon, B. Kégl, and D. Rousseau. Vol. 42. Proceedings of Machine Learning Research. Montreal, Canada: PMLR, 13 Dec 2015, pp. 1–18 (cit. on p. 2).
- [8] SKA consortium. *SKA Observatory – Big Data*. URL: <https://www.skao.int/en/explore/big-data> (visited on Aug. 23, 2024) (cit. on p. 2).
- [9] H. Schellman. “Computing for the DUNE Long-Baseline Neutrino Oscillation Experiment”. In: *EPJ Web Conf.* 245 (2020), p. 11002 (cit. on p. 2).
- [10] IceCube Collaboration. *IceCube Quick Facts*. URL: <https://icecube.wisc.edu/about-us/facts/#:~:text=IceCube%20is%20designed%20to%20detect,and%20about%20100%2C000%20per%20year>. (visited on Aug. 23, 2024) (cit. on p. 2).
- [11] W. Pauli. *Letter to Tübingen conference participants*. 1930. URL: <https://www.math.utah.edu/~beebe/talks/2015/qtm/pdf/pauli-1930-ltc.pdf> (visited on Aug. 1, 2024) (cit. on p. 5).

- [12]F. Reines and C. L. Cowan. “Detection of the Free Neutrino”. In: *Phys. Rev.* 92 (3 Nov. 1953), pp. 830–831 (cit. on p. 5).
- [13]C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. “Detection of the Free Neutrino: a Confirmation”. In: *Science* 124.3212 (July 1956), pp. 103–104 (cit. on p. 5).
- [14]Y. Fukuda, T. Hayakawa, E. Ichihara, et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567 (cit. on pp. 5, 6).
- [15]Q. R. Ahmad, R. C. Allen, T. C. Andersen, et al. “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”. In: *Phys. Rev. Lett.* 89 (1 June 2002), p. 011301 (cit. on pp. 5, 6).
- [16]S. Bilenky, C. Giunti, and W. Grimus. “Phenomenology of neutrino oscillations”. In: *Progress in Particle and Nuclear Physics* 43 (1999), pp. 1–86 (cit. on pp. 5, 6).
- [17]E. Roulet and F. Vissani. “Neutrino Oscillations”. In: *Neutrinos in Physics and Astrophysics*. WORLD SCIENTIFIC, Oct. 2022. Chap. 4, pp. 59–77. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789811260940_0004 (cit. on p. 7).
- [18]M. Kleesiek, J. Behrens, G. Drexlin, et al. “ β -Decay spectrum, response function and statistical model for neutrino mass measurements with the KATRIN experiment”. In: *The European Physical Journal C* 79.3 (Mar. 2019) (cit. on pp. 8–10, 18).
- [19]E. Fermi. “Versuch einer Theorie der β -Strahlen. I”. In: *Zeitschrift für Physik* 88.3–4 (Mar. 1934), pp. 161–177 (cit. on p. 8).
- [20]J. A. Formaggio, A. L. C. de Gouvêa, and R. H. Robertson. “Direct measurements of neutrino mass”. In: *Physics Reports* 914 (2021). Direct measurements of neutrino mass, pp. 1–54 (cit. on pp. 8, 10, 12–15, 17).
- [21]E. W. Otten and C. Weinheimer. “Neutrino mass limit from tritium β decay”. In: *Reports on Progress in Physics* 71.8 (July 2008), p. 086201 (cit. on p. 9).
- [22]L. I. Bodine, D. S. Parno, and R. G. H. Robertson. “Assessment of molecular effects on neutrino mass measurements from tritium β decay”. In: *Phys. Rev. C* 91 (3 Mar. 2015), p. 035505 (cit. on pp. 10, 12, 24).
- [23]KATRIN Collaboration. “KATRIN design report 2004”. In: *Wissenschaftliche Berichte. FZKA 7090* (2005). Ed. by KATRIN Collaboration. 51.54.01; LK 01 (cit. on pp. 10, 12, 17).
- [24]KATRIN Collaboration, M. Aker, K. Altenmüller, et al. “The design, construction, and commissioning of the KATRIN experiment”. In: *Journal of Instrumentation* 16.08 (Aug. 2021), T08015 (cit. on pp. 10–12, 160).
- [25]V. Lobashev and P. Spivak. “A method for measuring the electron antineutrino rest mass”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 240.2 (1985), pp. 305–310 (cit. on p. 10).

- [26] A. Picard, H. Backe, H. Barth, et al. “A solenoid retarding spectrometer with high resolution and transmission for keV electrons”. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 63.3 (1992), pp. 345–358 (cit. on p. 10).
- [27] M. Arenz, M. Babutzka, M. Bahr, et al. “Commissioning of the vacuum system of the KATRIN Main Spectrometer”. In: *Journal of Instrumentation* 11.04 (Apr. 2016), P04011 (cit. on p. 11).
- [28] C. Kraus, B. Bornschein, L. Bornschein, et al. “Final results from phase II of the Mainz neutrino mass search in tritium beta decay”. In: *Eur. Phys. J. C* 40 (2005), pp. 447–468. arXiv: hep-ex/0412056 (cit. on pp. 12, 22).
- [29] V. N. Aseev, A. I. Belesev, A. I. Berlev, et al. “Upper limit on the electron antineutrino mass from the Troitsk experiment”. In: *Phys. Rev. D* 84 (11 Dec. 2011), p. 112003 (cit. on pp. 12, 22).
- [30] Project 8 Collaboration, A. A. Esfahani, S. Böser, et al. *The Project 8 Neutrino Mass Experiment*. 2022. arXiv: 2203.07349 [nucl-ex] (cit. on pp. 12, 16, 17, 21–23, 89, 157, 159, 160).
- [31] E. Roulet and F. Vissani. “Neutrinos in Cosmology”. In: *Neutrinos in Physics and Astrophysics*. WORLD SCIENTIFIC, Oct. 2022. Chap. 10, pp. 191–212. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789811260940_0010 (cit. on p. 13).
- [32] The Planck Collaboration. *The Scientific Programme of Planck*. 2006. arXiv: astro-ph/0604069 [astro-ph] (cit. on p. 13).
- [33] N. Aghanim, Y. Akrami, M. Ashdown, et al. “Planck2018 results: VI. Cosmological parameters”. In: *Astronomy & Astrophysics* 641 (Sept. 2020), A6 (cit. on pp. 13, 16).
- [34] E. Roulet and F. Vissani. “Neutrino Masses and Mixings”. In: *Neutrinos in Physics and Astrophysics*. WORLD SCIENTIFIC, Oct. 2022. Chap. 3, pp. 41–57. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789811260940_0003 (cit. on p. 14).
- [35] S. Abe, T. Araki, K. Chiba, et al. *Search for Majorana Neutrinos with the Complete KamLAND-Zen Dataset*. 2024. arXiv: 2406.11438 [hep-ex] (cit. on pp. 14, 16).
- [36] T. J. Loredo and D. Q. Lamb. “Bayesian analysis of neutrinos observed from supernova SN 1987A”. In: *Phys. Rev. D* 65 (6 Feb. 2002), p. 063002 (cit. on pp. 15, 16).
- [37] F. An, G. An, Q. An, et al. “Neutrino physics with JUNO”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.3 (Feb. 2016), p. 030401 (cit. on p. 15).
- [38] F. Rossi-Torres, M. M. Guzzo, and E. Kemp. *Boundaries on Neutrino Mass from Supernovae Neutronization Burst by Liquid Argon Experiments*. 2015. arXiv: 1501.00456 [hep-ph] (cit. on p. 15).
- [39] H.-K. Proto-Collaboration, : K. Abe, et al. *Hyper-Kamiokande Design Report*. 2018. arXiv: 1805.04163 [physics.ins-det] (cit. on p. 15).

- [40]A. De Rújula and M. Lusignoli. “Calorimetric measurements of ^{163}Ho decay as tools to determine the electron neutrino mass”. In: *Physics Letters B* 118.4 (1982), pp. 429–434 (cit. on p. 16).
- [41]L. Gastaldo, K. Blaum, K. Chrysalidis, et al. “The electron capture in ^{163}Ho experiment – ECHO”. In: *The European Physical Journal Special Topics* 226.8 (June 2017), pp. 1623–1694 (cit. on p. 16).
- [42]M. De Gerone, B. Alpert, M. Balata, et al. “Status of the HOLMES Experiment”. In: *Journal of Low Temperature Physics* 209.5–6 (Nov. 2022), pp. 980–987 (cit. on p. 16).
- [43]C. Velte, F. Ahrens, A. Barth, et al. “High-resolution and low-background ^{163}Ho spectrum: interpretation of the resonance tails”. In: *The European Physical Journal C* 79.12 (Dec. 2019) (cit. on p. 16).
- [44]I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler, and T. Schwetz. “Updated fit to three neutrino mixing: exploring the accelerator-reactor complementarity”. In: *Journal of High Energy Physics* 2017.1 (Jan. 2017) (cit. on p. 17).
- [45]NuFIT 3.1. 2022. URL: <http://www.nu-fit.org/> (visited on Dec. 8, 2023) (cit. on p. 17).
- [46]M. Aker, A. Beglarian, J. Behrens, et al. “Direct neutrino-mass measurement with sub-electronvolt sensitivity”. In: *Nature Physics* 18.2 (Feb. 2022), pp. 160–166 (cit. on p. 17).
- [47]J. D. Jackson. *Classical Electrodynamics*. 3rd ed. Nashville, TN: John Wiley & Sons, July 1998 (cit. on pp. 18, 20, 27–32, 35, 37, 53, 78).
- [48]O. Aberle, I. Béjar Alonso, O. Brüning, et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2020 (cit. on p. 18).
- [49]A. A. Esfahani, S. Böser, N. Buzinsky, et al. “SYNCA: A Synthetic Cyclotron Antenna for the Project 8 Collaboration”. In: *Journal of Instrumentation* 18.01 (Jan. 2023), P01034 (cit. on p. 19).
- [50]C. A. Balanis. *Antenna Theory: Analysis and design*. en. John Wiley & Sons, 2005 (cit. on pp. 20, 41, 67, 160, 161).
- [51]J. S. Herd and M. D. Conway. “The Evolution to Modern Phased Array Architectures”. In: *Proceedings of the IEEE* 104.3 (2016), pp. 519–529 (cit. on p. 20).
- [52]H. Steyskal. “Digital beamforming antennas - An introduction”. In: *Microwave Journal* 30 (Dec. 1986), p. 107 (cit. on p. 20).
- [53]B. Van Veen and K. Buckley. “Beamforming: a versatile approach to spatial filtering”. In: *IEEE ASSP Magazine* 5.2 (1988), pp. 4–24 (cit. on pp. 20, 93).
- [54]X. Zhang, A. Molisch, and S.-Y. Kung. “Variable-phase-shift-based RF-baseband code-sign for MIMO antenna selection”. In: *IEEE Transactions on Signal Processing* 53.11 (2005), pp. 4091–4103 (cit. on p. 20).

- [55]F. Sohrabi and W. Yu. “Hybrid Analog and Digital Beamforming for mmWave OFDM Large-Scale Antenna Arrays”. In: *IEEE Journal on Selected Areas in Communications* 35.7 (2017), pp. 1432–1443 (cit. on p. 20).
- [56]R. N. Clarke and C. B. Rosenberg. “Fabry-Perot and open resonators at microwave and millimetre wave frequencies, 2-300 GHz”. In: *Journal of Physics E: Scientific Instruments* 15.1 (Jan. 1982), p. 9 (cit. on p. 20).
- [57]S. Ramo, J. R. Whinnery, and T. Van Duzer. *Fields and waves in communication electronics*. en. 3rd ed. Nashville, TN: John Wiley & Sons, Jan. 1994 (cit. on p. 20).
- [58]A. Ashtari Esfahani, M. Betancourt, Z. Bogorad, et al. “Bayesian analysis of a future β decay experiment’s sensitivity to neutrino mass scale and ordering”. In: *Phys. Rev. C* 103 (6 June 2021), p. 065501 (cit. on pp. 22, 184).
- [59]A. A. Esfahani, D. M. Asner, S. Böser, et al. “Determining the neutrino mass with cyclotron radiation emission spectroscopy—Project 8”. In: *Journal of Physics G: Nuclear and Particle Physics* 44.5 (Mar. 2017), p. 054004 (cit. on pp. 22, 160, 184).
- [60]D. M. Asner, R. F. Bradley, L. de Viveiros, et al. “Single-Electron Detection and Spectroscopy via Relativistic Cyclotron Radiation”. In: *Phys. Rev. Lett.* 114 (16 Apr. 2015), p. 162501 (cit. on pp. 22, 89).
- [61]A. Picard, H. Backe, J. Bonn, et al. “Precision measurement of the conversion electron spectrum of ^{83m}Kr with a solenoid retarding spectrometer”. In: *Zeitschrift für Physik A Hadrons and Nuclei* 342.1 (Mar. 1992), pp. 71–78 (cit. on p. 23).
- [62]R. G. H. Robertson, T. J. Bowles, G. J. Stephenson, et al. “Limit on $\bar{\nu}_e$ mass from observation of the β decay of molecular tritium”. In: *Phys. Rev. Lett.* 67 (8 Aug. 1991), pp. 957–960 (cit. on p. 23).
- [63]W. Stoeffl and D. J. Decman. “Anomalous Structure in the Beta Decay of Gaseous Molecular Tritium”. In: *Phys. Rev. Lett.* 75 (18 Oct. 1995), pp. 3237–3240 (cit. on p. 23).
- [64]K. Altenmüller, M. Arenz, W.-J. Baek, et al. “High-resolution spectroscopy of gaseous ^{83m}Kr conversion electrons with the KATRIN experiment”. In: *Journal of Physics G: Nuclear and Particle Physics* 47.6 (May 2020), p. 065002 (cit. on p. 23).
- [65]A. Ashtari Esfahani, S. Böser, N. Buzinsky, et al. “Tritium Beta Spectrum Measurement and Neutrino Mass Limit from Cyclotron Radiation Emission Spectroscopy”. In: *Phys. Rev. Lett.* 131 (10 Sept. 2023), p. 102502 (cit. on pp. 23, 89, 90).
- [66]A. Ashtari Esfahani, S. Böser, N. Buzinsky, et al. “Cyclotron radiation emission spectroscopy of electrons from tritium β decay and ^{83m}Kr internal conversion”. In: *Phys. Rev. C* 109 (3 Mar. 2024), p. 035503 (cit. on pp. 23, 89, 90, 92, 98, 184).
- [67]L. Thorne. “Building an atomic source for the Project 8 experiment”. In: *Proceedings of XVIII International Conference on Topics in Astroparticle and Underground Physics — PoS(TAUP2023)*. TAUP2023. Sissa Medialab, Feb. 2024, p. 231 (cit. on p. 23).
- [68]T. G. Northrop. “Adiabatic charged-particle motion”. In: *Reviews of Geophysics* 1.3 (1963), pp. 283–304. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/RG001i003p00283> (cit. on p. 27).

- [69]A. A. Esfahani, V. Bansal, S. Böser, et al. “Electron radiated power in cyclotron radiation emission spectroscopy experiments”. In: *Physical Review C* 99.5 (May 2019) (cit. on pp. 30, 201, 205).
- [70]E. W. Weisstein. *Convective Operator*. URL: <https://mathworld.wolfram.com/ConvectiveOperator.html> (visited on Jan. 7, 2024) (cit. on p. 34).
- [71]J. Johner. “Angular distribution of the total cyclotron radiation of a relativistic particle with parallel velocity”. In: *Phys. Rev. A* 36 (3 Aug. 1987), pp. 1498–1501 (cit. on p. 39).
- [72]B. A. Trubnikov. “Radiation of plasma in magnetic field”. In: *Dokl. Akad. Nauk SSSR* 118 (5 1958), pp. 913–916 (cit. on p. 40).
- [73]W. Byron, H. Harrington, R. J. Taylor, et al. “First Observation of Cyclotron Radiation from MeV-Scale e^\pm following Nuclear β Decay”. In: *Phys. Rev. Lett.* 131 (8 Aug. 2023), p. 082502 (cit. on pp. 41, 189).
- [74]D. Furse, S. Groh, N. Trost, et al. “Kassiopeia: a modern, extensible C++ particle tracking package”. In: *New Journal of Physics* 19.5 (May 2017), p. 053012 (cit. on pp. 50, 55).
- [75]V. Aseev, A. Belev, A. Berlev, et al. “Energy loss of 18 keV electrons in gaseous T and quench condensed D films”. In: *The European Physical Journal D* 10.1 (Mar. 2000), pp. 39–52 (cit. on p. 50).
- [76]KATRIN Collaboration. *Kassiopeia*. 2017. URL: <https://github.com/KATRIN-Experiment/Kassiopeia> (visited on Apr. 9, 2023) (cit. on p. 50).
- [77]KATRIN Collaboration. *Kassiopeia’s documentation*. 2017. URL: <https://katrin-experiment.github.io/Kassiopeia/index.html> (visited on May 8, 2024) (cit. on p. 50).
- [78]A. A. Esfahani, S. Böser, N. Buzinsky, et al. “Locust: C++ software for simulation of RF detection”. In: *New Journal of Physics* 21.11 (Nov. 2019), p. 113051 (cit. on p. 51).
- [79]Project 8 Collaboration. *locust_mc*. 2019. URL: https://github.com/project8/locust_mc (visited on Apr. 9, 2023) (cit. on p. 51).
- [80]F. Thomas. *CRESana*. 2023. URL: <https://github.com/MCFlowMace/CRESana> (visited on Apr. 9, 2023) (cit. on p. 52).
- [81]C. R. Harris, K. J. Millman, S. J. van der Walt, et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362 (cit. on p. 52).
- [82]P. Virtanen, R. Gommers, T. E. Oliphant, et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272 (cit. on pp. 52, 56, 60, 62).
- [83]J. Simpson, J. Lane, C. Immer, and R. Youngquist. *Simple Analytic Expressions for the Magnetic Field of a Circular Current Loop*. Tech. rep. NASA Technical Reports Server (NTRS), 2001 (cit. on p. 54).

- [84]S. Tou. *Visualization of fields and applications in engineering*. en. Nashville, TN: John Wiley & Sons, Apr. 2011 (cit. on p. 55).
- [85]C. De Boor. *A practical guide to splines*. en. 1st ed. Applied Mathematical Sciences. New York, NY: Springer, Nov. 2001 (cit. on pp. 56, 62).
- [86]R. Piessens, E. de Doncker-Kapenga, C. W. Überhuber, and D. K. Kahaner. *Quadpack*. Springer Berlin Heidelberg, 1983 (cit. on p. 56).
- [87]H. Takahasi and M. Mori. “Double exponential formulas for numerical integration”. In: *Publications of the Research Institute for Mathematical Sciences* 9.3 (1973), pp. 721–741 (cit. on p. 58).
- [88]The mpmath development team. *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.3.0)*. <http://mpmath.org/>. 2023 (cit. on p. 58).
- [89]Wolfram Research, Inc. *Mathematica, Version 13.3*. Champaign, IL, 2023 (cit. on pp. 59, 139, 219).
- [90]R. C. Johnson and H. Jasik. *Antenna Engineering Handbook*. en. 2nd ed. New York, NY: McGraw-Hill, Mar. 1984 (cit. on p. 62).
- [91]Ansys, Inc.f. *Ansys HFSS*. URL: <https://www.ansys.com/products/electronics/ansys-hfss> (visited on May 8, 2024) (cit. on p. 65).
- [92]C. Shannon. “Communication in the Presence of Noise”. In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21 (cit. on pp. 70, 123, 209).
- [93]H. Nyquist. “Thermal Agitation of Electric Charge in Conductors”. In: *Phys. Rev.* 32 (1 July 1928), pp. 110–113 (cit. on pp. 71, 89).
- [94]J. R. Barry, E. A. Lee, and D. G. Messerschmitt. *Digital Communication*. en. 3rd ed. Dordrecht, Netherlands: Springer, Sept. 2003 (cit. on pp. 71, 89).
- [95]J. Stachurska. *Going Big for Phase III of the Project 8 Neutrino Mass Experiment*. 2023. arXiv: 2311.16415 [physics.ins-det] (cit. on pp. 78, 188).
- [96]S. Faruque. *Radio Frequency Modulation Made Easy*. Springer International Publishing, 2017 (cit. on pp. 80, 81).
- [97]X. Ba, Y. Yang, J. Li, and J. Chen. “Differentially coherent acquisition algorithm for indoor GPS”. In: *2006 IET International Conference on Wireless, Mobile and Multimedia Networks*. 2006, pp. 1–4 (cit. on p. 89).
- [98]D. Manandhar, Y. Suh, and R. Shibasaki. “GPS Signal Acquisition and Tracking-An Approach towards Development of Software-based GPS Receiver”. In: *Proceedings of The Institute of Electronics, Information and Communication Engineers, Technical Report of IEICE, German* (Jan. 2006) (cit. on p. 89).
- [99]S. M. Kay. *Fundamentals of statistical signal processing, volume II*. Philadelphia, PA: Prentice Hall, Jan. 1998 (cit. on pp. 91, 105–109, 114, 118, 119, 133).

- [100]A. M. Elbir, K. V. Mishra, S. A. Vorobyov, and R. W. Heath. “Twenty-Five Years of Advances in Beamforming: From convex and nonconvex optimization to learning techniques”. In: *IEEE Signal Processing Magazine* 40.4 (2023), pp. 118–131 (cit. on p. 93).
- [101]S. Krusevac, P. Rapajic, and R. Kennedy. “Mutual Coupling Effect on Thermal Noise in Multi-Element Antenna Systems”. In: *Progress in Electromagnetics Research* 59 (2006), pp. 325–333 (cit. on p. 107).
- [102]A. A. Esfahani, Z. Bogorad, S. Böser, et al. “Viterbi decoding of CRES signals in Project 8”. In: *New Journal of Physics* 24.5 (May 2022), p. 053013 (cit. on p. 111).
- [103]A. Viterbi. “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”. In: *IEEE Transactions on Information Theory* 13.2 (1967), pp. 260–269 (cit. on p. 111).
- [104]S. Babak, R. Balasubramanian, D. Churches, T. Cokelaer, and B. S. Sathyaprakash. “A template bank to search for gravitational waves from inspiralling compact binaries: I. Physical models”. In: *Classical and Quantum Gravity* 23.18 (Aug. 2006), p. 5477 (cit. on pp. 112, 115).
- [105]C. Van Den Broeck, D. A. Brown, T. Cokelaer, et al. “Template banks to search for compact binaries with spinning components in gravitational wave data”. In: *Phys. Rev. D* 80 (2 July 2009), p. 024009 (cit. on p. 115).
- [106]A. A. Esfahani, S. Böser, N. Buzinsky, et al. *Real-time Signal Detection for Cyclotron Radiation Emission Spectroscopy Measurements using Antenna Arrays*. 2023. arXiv: 2310.02112 [physics.ins-det] (cit. on pp. 116, 129).
- [107]. *BLAS (Basic Linear Algebra Subprograms)*. URL: <http://www.netlib.org/blas/> (visited on Feb. 13, 2024) (cit. on p. 124).
- [108]Intel Corporation. *Intel MKL (Intel Math Kernel Library)*. URL: <https://software.intel.com/content/www/us/en/develop/tools/math-kernel-library.html> (visited on Feb. 13, 2024) (cit. on p. 124).
- [109]NVIDIA Corporation. *cuBLAS - Basic Linear Algebra on NVIDIA GPUs*. URL: <https://developer.nvidia.com/cublas> (visited on Feb. 13, 2024) (cit. on p. 124).
- [110]J. Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural Networks* 61 (2015), pp. 85–117 (cit. on p. 125).
- [111]I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016 (cit. on p. 125).
- [112]NVIDIA Corporation. *NVIDIA H100 Tensor Core GPU datasheet*. URL: <https://resources.nvidia.com/en-us-tensor-core/nvidia-tensor-core-gpu-datasheet> (visited on Feb. 13, 2024) (cit. on p. 125).
- [113]NVIDIA Corporation. *Matrix Multiplication Background User’s Guide*. URL: <https://docs.nvidia.com/deeplearning/performance/dl-performance-matrix-multiplication/index.html> (visited on Feb. 13, 2024) (cit. on p. 125).

- [114]TOP500. *TOP500 November 2023*. URL: <https://www.top500.org/lists/top500/2023/11/> (visited on Feb. 17, 2024) (cit. on pp. 127, 128).
- [115]J. O. Smith. *Mathematics of the Discrete Fourier Transform (DFT)*. online book, 2007 edition. <http://ccrma.stanford.edu/~jos/mdft/>, accessed 2024-02-16 (cit. on p. 128).
- [116]K. Albertsson, P. Altoe, D. Anderson, et al. “Machine Learning in High Energy Physics Community White Paper”. In: *Journal of Physics: Conference Series* 1085.2 (Sept. 2018), p. 022008 (cit. on p. 129).
- [117]F. Psihas, M. Groh, C. Tunnell, and K. Warburton. “A review on machine learning for neutrino experiments”. In: *International Journal of Modern Physics A* 35.33 (2020), p. 2043005. eprint: <https://doi.org/10.1142/S0217751X20430058> (cit. on p. 129).
- [118]D. Bourilkov. “Machine and deep learning applications in particle physics”. In: *International Journal of Modern Physics A* 34.35 (2019), p. 1930019. eprint: <https://doi.org/10.1142/S0217751X19300199> (cit. on p. 129).
- [119]L. Stanković and D. Mandić. “Convolutional Neural Networks Demystified: A Matched Filtering Perspective-Based Tutorial”. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 53.6 (2023), pp. 3614–3628 (cit. on p. 129).
- [120]D. George and E. A. Huerta. “Deep neural networks to enable real-time multimessenger astrophysics”. In: *Phys. Rev. D* 97 (4 Feb. 2018), p. 044039 (cit. on p. 129).
- [121]D. George and E. Huerta. “Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data”. In: *Physics Letters B* 778 (2018), pp. 64–70 (cit. on p. 129).
- [122]S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice-Hall, Apr. 5, 1993 (cit. on pp. 131, 136, 138, 171, 218).
- [123]F. James and M. Roos. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”. In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367 (cit. on pp. 134, 141, 156).
- [124]F. James. “MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1”. In: (1994) (cit. on pp. 134, 141, 156).
- [125]S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62 (cit. on p. 134).
- [126]J. Soch, T. B. of Statistical Proofs, Maja, et al. *StatProofBook/StatProofBook.github.io: StatProofBook 2023*. Version 2023. Jan. 2024 (cit. on p. 134).
- [127]F. James. *Statistical methods in experimental physics (2nd edition)*. 2nd ed. Singapore, Singapore: World Scientific Publishing, Dec. 2006 (cit. on p. 134).
- [128]S. Algeri, J. Aalbers, K. D. Morå, and J. Conrad. “Searching for new phenomena with profile likelihood ratio tests”. In: *Nature Reviews Physics* 2.5 (Apr. 2020), pp. 245–252 (cit. on p. 136).

- [129]V. Dovì, O. Paladino, and A. Reverberi. “Some remarks on the use of the inverse hessian matrix of the likelihood function in the estimation of statistical properties of parameters”. In: *Applied Mathematics Letters* 4.1 (1991), pp. 87–90 (cit. on p. 136).
- [130]N. Buzinsky. “Statistical Signal Processing and Detector Optimization in Project 8”. PhD thesis. Massachusetts Institute of Technology, 2021 (cit. on pp. 138, 139, 218).
- [131]G. Cowan, K. Cranmer, E. Gross, and O. Vitells. “Asymptotic Formulae for Likelihood-Based Tests of New Physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011), pp. 1–19. arXiv: 1007.1727 (cit. on p. 141).
- [132]H. Dembinski and P. O. et al. “scikit-hep/iminuit”. In: (Dec. 2020) (cit. on pp. 141, 167).
- [133]M. Hatlo, F. James, P. Mato, et al. “Developments of mathematical software libraries for the LHC experiments”. In: *IEEE Trans. Nucl. Sci.* 52 (2005), pp. 2818–2822 (cit. on p. 141).
- [134]M. Grotschel, L. Lovasz, and A. Schrijver. *Geometric algorithms and combinatorial optimization*. en. 2nd ed. Algorithms and combinatorics. Berlin, Germany: Springer, Dec. 2011 (cit. on p. 143).
- [135]Project 8 collaboration. *Antenna Arrays for CRES-based Neutrino Mass Measurement [Manuscript in preparation]*. 2024 (cit. on pp. 160, 187).
- [136]A. Ziegler. “Development of Scalable Approaches to Neutrino Mass Measurement with the Project 8 Experiment”. PhD thesis. Pennsylvania State University, 2023 (cit. on p. 164).
- [137]R. Saakyan. *Determination of Absolute Neutrino Mass Using Quantum Technologies*. Talk at the UK HEP Forum 2020: Quantum leaps to the dark side. Nov. 10, 2020 (cit. on p. 189).
- [138]M. Betti, M. Biasotti, A. Boscá, et al. “Neutrino physics with the PTOLEMY project: active neutrino properties and the light sterile case”. In: *Journal of Cosmology and Astroparticle Physics* 2019.07 (July 2019), p. 047 (cit. on p. 189).
- [139]Y. Iwasaki, A. Tan, and C. G. Tully. *Towards CRES-Based Non-destructive Electron Momentum Estimation for the PTOLEMY Relic Neutrino Detector*. 2024. arXiv: 2404.00817 [physics.ins-det] (cit. on p. 189).
- [140]K. Kazkaz and N. Woollett. “Using cyclotron radiation emission for ultra-high resolution x-ray spectroscopy”. In: *New Journal of Physics* 23.3 (Mar. 2021), p. 033043 (cit. on p. 189).

A.1 The Harmonic Trap

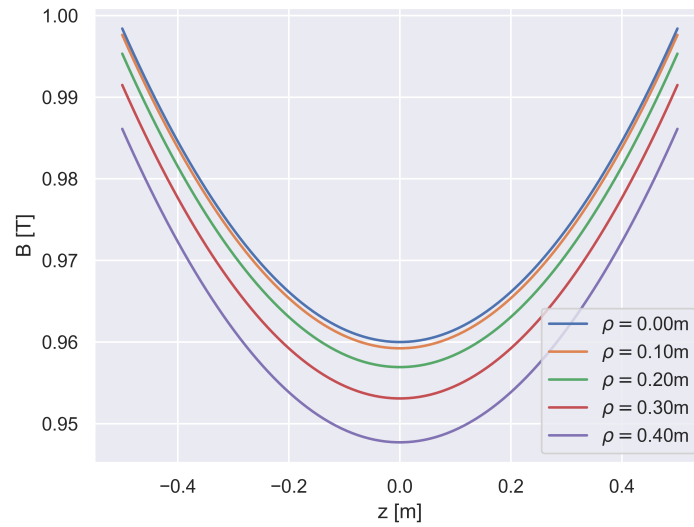


Fig. A.1.: Field profiles of a harmonic trap for several different radii with $B_0 = 0.96$ T and $L_0 = 2.5$ m.

The harmonic trap is a magnetic trap with a field that has a quadratic profile along z . With the solutions eqs. (4.12) and (4.13) a field with that characteristic is produced with $a_l = 0 \forall l \in \mathbb{N} \setminus \{1, 3\}$ which yields $\mathbf{B}_H = B_\rho(\rho, z)\hat{e}_\rho + B_z(\rho, z)\hat{e}_z$ with

$$\begin{aligned}
 B_\rho(\rho, z) &= -3\rho z a_3, \\
 B_z(\rho, z) &= a_1 - \frac{3}{2}(\rho^2 - 2z^2)a_3 \\
 B(\rho, z) &= \sqrt{9\rho^2 z^2 a_3^2 + \frac{1}{4}(2a_1 - 3(\rho^2 - 2z^2)a_3)^2}. \tag{A.1}
 \end{aligned}$$

Figure A.1 shows the field profile $B(\rho, z) = |\mathbf{B}_H(\rho, z)|$ for several radii ρ . This trap type is a generalization of the harmonic trap in [69] extending it to radii $\rho > 0$. This

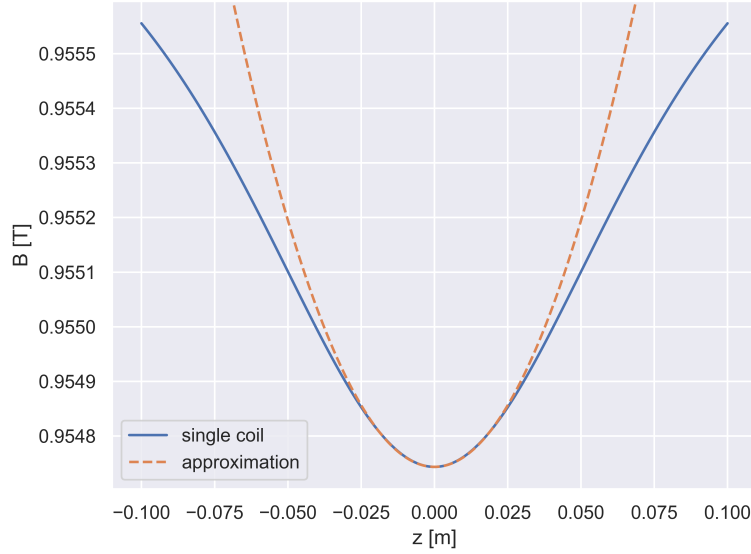


Fig. A.2.: Field profile of a single coil and a harmonic approximation at $\rho = 0$ with $L_0 = 2.3$ m and $B_0 = 0.95$ T.

is clear by evaluating the field at $\rho = 0$ and setting the coefficients appropriately to $a_1 = B_0$, $a_3 = \frac{B_0}{3L_0^2}$ yielding

$$B_z(z) = B_0 \left(1 - \frac{z^2}{L_0^2} \right),$$

where L_0 is called the characteristic length.

The main motivation for this field is the fact that it is a simple analytic approximation to the field on the symmetry axis of a single electromagnetic coil, which is demonstrated in fig. A.2. This is useful because it is possible to find a closed form solution for the axial motion in this trap. As mentioned earlier the integral that solves the axial motion in eq. (4.6) only has a closed form solution if $B(z)$ is a polynomial with order ≤ 2 . This is actually not exactly true for $B(\rho, z)$ in eq. (A.1). However, since it is required that the background field B_0 is much stronger than the trapping component of \mathbf{B}_H it is implied that $\frac{B_0}{2L_0^2} (\rho^2 - 2z^2) \ll B_0$ for ρ and z small. Therefore, it is reasonable to assume that $B_\rho \ll B_z$ for the interesting region of ρ, z and thus $B(z) \approx B_z(\rho, z)$.

Using this approximation, the axial motion is determined by

$$t(z) = \frac{\sqrt{B(z_{max})}}{v_0} \int_0^z \frac{dz'}{\sqrt{B(z_{max}) - B_0 \left(1 - \frac{1}{2L_0^2} (\rho^2 - 2z'^2) \right)}}.$$

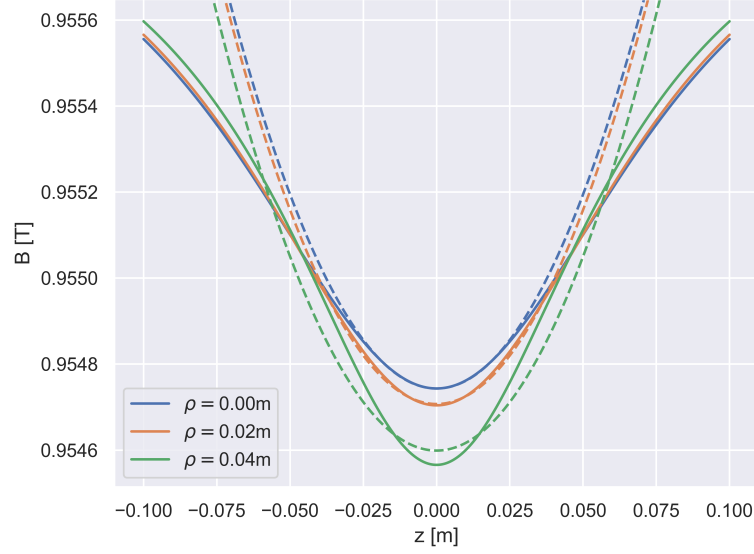


Fig. A.3.: Field profile of a single coil and a harmonic approximation for several different radii with $L_0 = 2.3$ m and $B_0 = 0.95$ T. The approximation quality severely degrades for $\rho > 0$.

This integral can be rearranged as

$$t(z) = a \int_0^z \frac{dz'}{\sqrt{b - cz'^2}},$$

with

$$a = \frac{\sqrt{B(z_{max})}}{v_0}, \quad b = B(z_{max}) - B(0), \quad c = \frac{B_0}{L_0^2}.$$

Under the condition that $0 < z < \sqrt{\frac{b}{c}}$ the integral has the solution

$$t(z) = \frac{a}{\sqrt{c}} \arctan \left(z \sqrt{\frac{c}{b - cz^2}} \right). \quad (\text{A.2})$$

Solving eq. (A.2) for z yields

$$z(t) = z_{max} \sin(\omega_H t), \quad (\text{A.3})$$

with

$$z_{max} = \sqrt{\frac{b}{c}}, \quad \omega_H = \frac{\sqrt{c}}{a}.$$

Due to the above restriction of z , this solution is strictly speaking only valid for $0 < t < \frac{\pi}{2\omega_H}$ but it is straightforward to extend that using the symmetry and

periodicity of the motion (see section 5.4.1). Finally, calculating a, b, c requires $B(0)$ and $B(z_{max})$. First

$$B(0) = B_0 \left(1 - \frac{1}{2L_0^2} \rho^2 \right).$$

Second $B(z_{max})$ is calculated using eq. (4.5) which yields

$$B(z_{max}) = \frac{B(0)}{\sin^2(\alpha_0)}.$$

Therefore, the final solution for the motion is eq. (A.3) with

$$z_{max} = \frac{\sqrt{L_0^2 - \frac{1}{2}\rho^2}}{\tan(\alpha_0)}, \quad \omega_H = \frac{v_0 \sin(\alpha_0)}{L_0} \frac{1}{\sqrt{1 - \frac{\rho^2}{2L_0^2}}}. \quad (\text{A.4})$$

and it requires $\rho < \sqrt{2}L_0$. In eq. (A.4) α_0 is the electron's pitch angle at $z = 0$ and ρ its initial radial position.

With these results the motion of an electron in a general harmonic trap can be described also for the case that $\rho > 0$. The results are given as closed form expressions, meaning that they can be evaluated without any expensive numeric integration and they are valid solutions of Maxwell's equations. Even so, the practical applicability of this trap is very limited. Figure A.3 shows that the approximation quality for a coil decreases for $\rho > 0$ if a single harmonic field is used.

A.2 The Square Well Trap

The square well trap is a magnetic trap that has a constant value B_0 over a length L and is infinitely high outside of that well:

$$\mathbf{B}_{SW}(\rho, z) = \begin{cases} B_0 \hat{\mathbf{e}}_z & |z| < L/2 \\ \infty & \text{otherwise} \end{cases}$$

Figure A.4 shows the field profile $B(z) = |\mathbf{B}_{SW}(\rho, z)|$, which is the relevant function for the solution of the axial motion (eq. (4.6)). The solution $z(t)$ is a triangle wave with frequency

$$\omega_{SW} = \frac{v_0 \cos(\alpha_0) \pi}{L},$$

where v_0 is the absolute velocity of the electron and α_0 is the pitch angle.

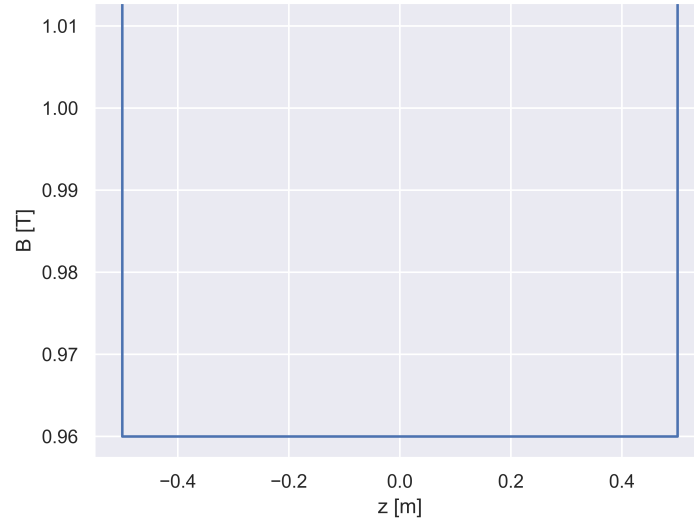


Fig. A.4.: Field profile of a square well trap with $L = 1$ m and $B_0 = 0.96$ T.

The constant part of the square well trap is an allowed solution for the magnetic field description in section 4.2. It is accomplished with the solutions in eqs. (4.12) and (4.13) by setting $a_l = 0 \forall l \in \mathbb{N} \setminus \{1\}$ and $a_1 = B_0$. However, this field is constant everywhere and is lacking the sharp walls. The exact solution with infinitely sharp walls that is presented here cannot be accomplished with Maxwell's equations.

A.3 The Bathtub Trap

The bathtub trap is a combination of the harmonic trap and the square well trap. Its purpose is to provide a trap with the constant field over a long distance L but without the infinities that prohibit the square well trap by smoothly transitioning into harmonic walls. The idea is that the flat portion is achieved as a background field provided by a large solenoid, which would be approximately constant in the central field region. The trap is generated by adding two coils at position $|z| > L/2$ which are approximated by harmonic fields as usual. This kind of approximation is used for example in [69]. The resulting field is defined as

$$\mathbf{B}_{BT}(\rho, z) = \begin{cases} \mathbf{B}_H(\rho, z + L/2) & z < -L/2 \\ \mathbf{B}_H(\rho, z - L/2) & z > L/2 \\ \mathbf{B}_{SW}(\rho, z) & \text{otherwise} \end{cases} .$$

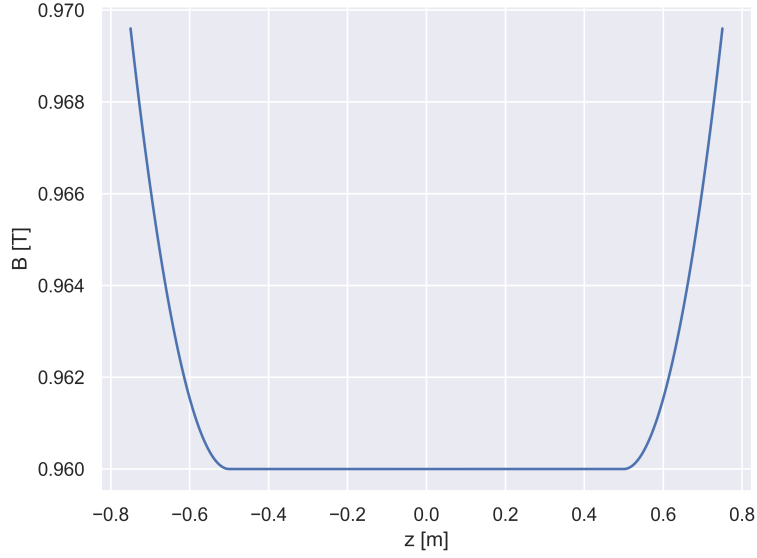


Fig. A.5.: Field profile of a bathtub trap at $\rho = 0$ with $L = 1$ m, $B_0 = 0.96$ T and $L_0 = 2.5$ m.

Figure A.5 depicts the field profile $B(z) = |\mathbf{B}_{BT}(0, z)|$ on the symmetry axis. However, this field has the issue that it is discontinuous at $z = \pm L/2$ for $\rho \neq 0$. This is because the square well field has no ρ dependence since $\mathbf{B}_{SW}(\rho, z) = a_1 \hat{\mathbf{e}}_z$, but $\mathbf{B}_H(\rho, 0) = (a_1 - \frac{3}{2}a_3\rho^2)\hat{\mathbf{e}}_z$ (see also fig. A.6).

This can be resolved by defining the bathtub trap using the Heaviside function $\Theta(x)$ in the following way:

$$\mathbf{B}_{BTc}(\rho, z) = \mathbf{B}_H(\rho, (1 - \Theta(z + L/2))(z + L/2) + \Theta(z - L/2)(z - L/2)) .$$

With this definition the bathtub trap is the harmonic trap with a coordinate transformation for z . The discontinuity is eliminated since $\mathbf{B}_{BTc}(\rho, z) = (a_1 - \frac{3}{2}a_3\rho^2)\hat{\mathbf{e}}_z$ even for $-L/2 \leq z \leq L/2$.

Still, this solution is now incompatible with eqs. (4.12) and (4.13) and thus it is not a valid magnetic field for the assumptions in section 4.2. This is because the term proportional to ρ^2 without powers of z that now exists in the flat part is only possible for $a_3 \neq 0$. But $a_3 \neq 0$ always produces a term proportional to z^2 , which is

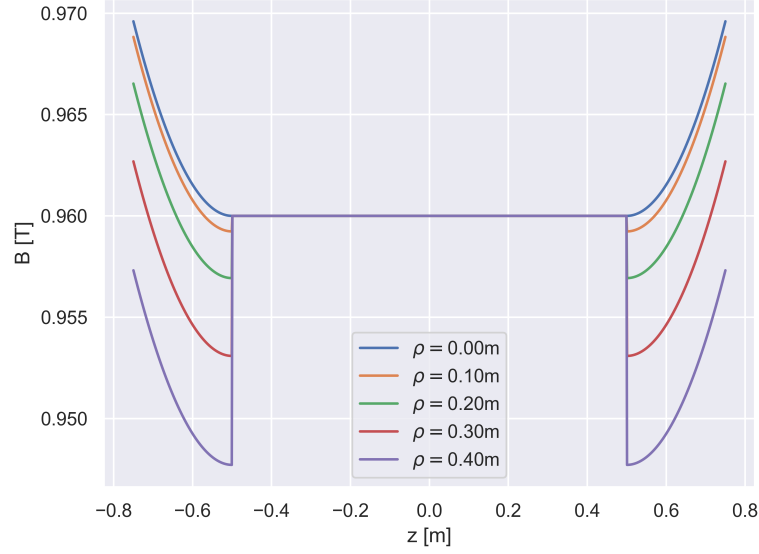


Fig. A.6.: Field profiles of a bathtub trap for several different radii with $L = 1$ m, $B_0 = 0.96$ T and $L_0 = 2.5$ m. This shows that the profile is discontinuous for $\rho > 0$.

missing. In fact, it is possible to show that this solution violates the assumption of $j = 0$ from section 4.2:

$$\begin{aligned}
 \nabla \times \mathbf{B}_{BTc}(\rho, z) &= \left(\frac{1}{\rho} \frac{\partial B_z}{\partial \phi} - \frac{\partial B_\phi}{\partial z} \right) \hat{\mathbf{e}}_r + \left(\frac{\partial B_\rho}{\partial z} - \frac{\partial B_z}{\partial \rho} \right) \hat{\mathbf{e}}_\phi \\
 &\quad + \frac{1}{\rho} \left(\frac{\partial}{\partial \rho} (\rho \cdot B_\phi) - \frac{\partial B_\rho}{\partial \phi} \right) \hat{\mathbf{e}}_z \\
 &= -\frac{\partial B_z}{\partial \rho} \hat{\mathbf{e}}_\phi \\
 &= 3\rho a_3 \hat{\mathbf{e}}_\phi .
 \end{aligned}$$

With Ampère's law (eq. (4.10)) this means the field is only possible with a current density $\mathbf{j} = 3\rho \frac{a_3}{\mu_0} \hat{\mathbf{e}}_\phi$ in the region $-L/2 \leq z \leq L/2$. However, for a CRES experiment long mean free paths are desired for the electrons. This is incompatible with a dense structure of wires that can produce the required current distribution.

Finally, the initial assumption that a bathtub trap is a good approximation for a trap consisting of a background field with two trapping coils is also inaccurate. This is demonstrated in fig. A.7. In order to create a potential wall in this setup the trapping coils have to add positive field contributions. The field of this kind of coil can only be approximated with a negative harmonic function at the maximum. Conversely, the transition region from a flat field into a field that is dominated by a coil is actually not the region of the field where a coil is well approximated by a harmonic function.

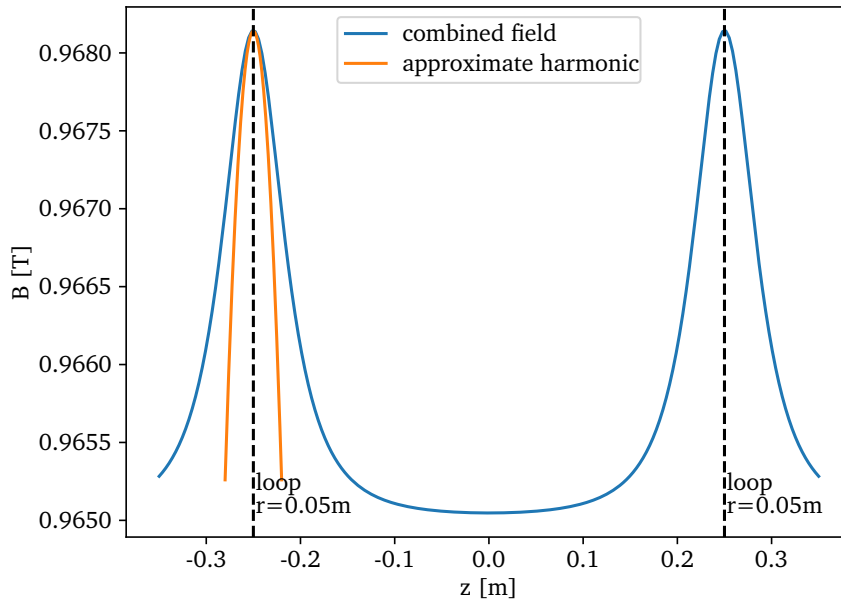


Fig. A.7.: Field profile of a realistic trap with a background field and two electromagnetic coils at $\rho = 0$ together with a harmonic approximation to one of the coil fields. The approximation that can be used for the coil has no value for the calculation of the motion in this trap because it is only valid at the unstable field maximum.

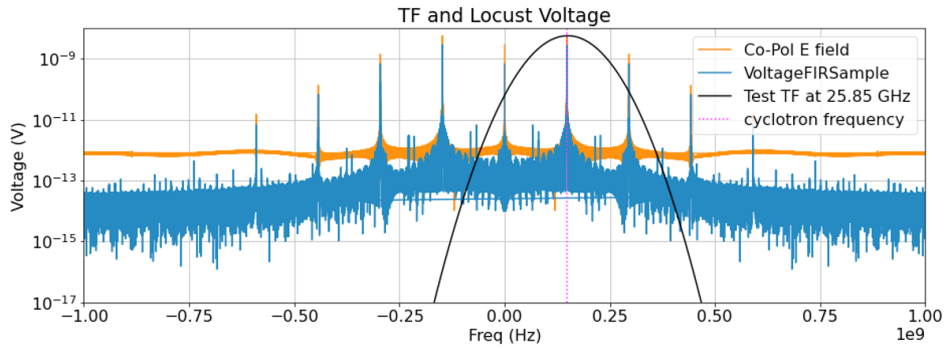
In conclusion the two findings here should recommend against the use of the bathtub trap in simulations of CRES as results have very little explanatory power even on a conceptual level. Even for cases where this trap might be able to provide a good approximation for the field of a real trap at $\rho = 0$, the real field will certainly behave very differently for $\rho \neq 0$.

Bugs in Locust

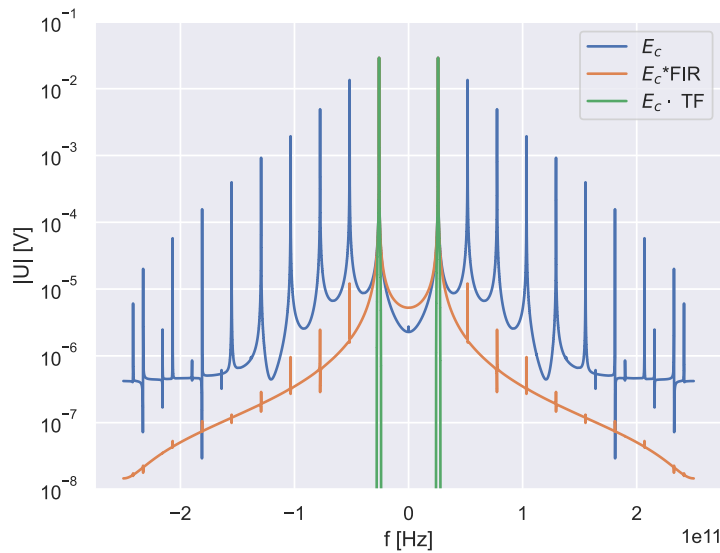
B.1 Applying Antenna Response to Undersampled Signals

One complication of numerically calculating the electromagnetic fields comes from the cyclotron frequency of 26 GHz for an experiment with a 1 T magnetic field. According to the Nyquist–Shannon sampling theorem [92] these fields can only be properly sampled with sampling rates $f_s > 52$ GHz which is required for the application of all the high frequency digital filters in the simulation (that includes the antenna response). However, the actual frequencies of interest for a neutrino mass measurement are a band only 100s of MHz wide around the target frequency of 26 GHz. In a laboratory measurement the analog voltage signal can be reduced to such a bandwidth with frequency mixers and low-pass filters prior to digitization (see section 5.4.4). Therefore, digitally sampling the intermediate signals and filters in the simulation requires significantly higher sampling rates and thus more computational resources than sampling of the final voltages even though the fields always remain analog signals in a real life measurement. The situation is even worse if one considers the harmonics of the cyclotron frequency as seen in fig. 4.6a which seemingly require sampling rates of well over 100 GHz. In the experiment the harmonics would be eliminated from the analog signal since they are outside of the antenna’s passband. But in the digitally sampled numeric fields the harmonics can show up as aliased frequencies inside of the passband.

In order to reduce the computational cost of sampling intermediate signals the implementation in Locust uses an approach where the electromagnetic fields are undersampled with a sampling rate of 2 GHz. A high resolution signal as input for the digital filters is constructed by interpolating with a mono-frequency sinusoidal signal between each field sample. A Hilbert transform on the original field samples is used to estimate instantaneous amplitude and phase for that sinusoidal signal.



(a) Incorrectly applied digital filter (black) on an undersampled signal (orange) in Locust. The undersampled signal is interpolated as described in the text. The peaks outside of the passband are not suppressed in the output (blue). For visualization purposes the filter was rescaled and moved to its undersampled frequency alias.



(b) Correct way of applying digital filter (green) to an appropriately sampled signal (blue) yields an output (orange) where the peaks outside of the passband are suppressed.

Fig. B.1.: Filtering of the sampled field signal.

This approach is not working as intended. This can be seen in fig. B.1a. The orange spectrum is the undersampled field signal which aliases to lower frequencies including the harmonics. The signal is then passed through a narrow filter at 25.85 GHz by the interpolation approach described above which produces the blue output. The filter is unable to remove the peaks outside of its passband and it seems to act as a constant factor on the spectrum instead. In contrast fig. B.1b depicts what the result should look like without undersampling and with correct filtering. This proves that Locust is unable to correctly apply any high-frequency filters to the undersampled input signal, which is especially concerning for narrow filters that should act on smaller scale modulations of the base frequency. As a result of this, the aliases of the harmonics shown here can sometimes still appear in the final spectrum after the final low pass filter (see fig. B.2).

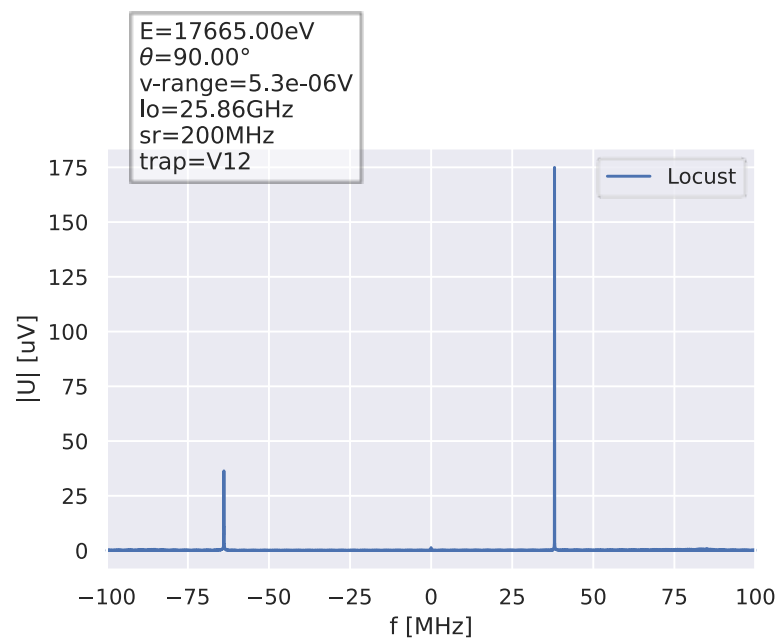
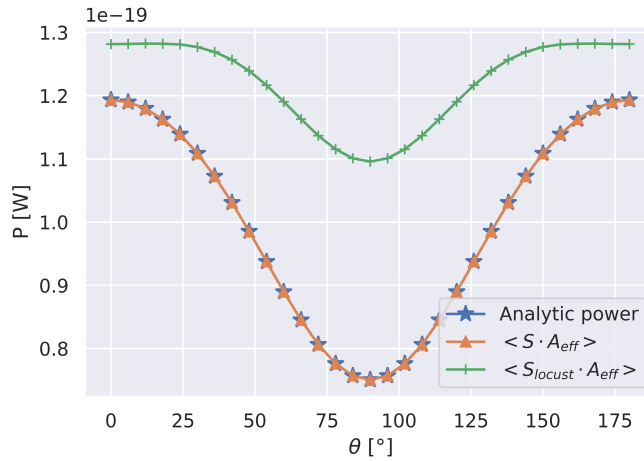


Fig. B.2.: Aliased harmonic inside a Locust spectrum. In this simulation only a single peak is expected at the cyclotron frequency which is the higher one. The smaller peak in the left can be traced back to an aliased frequency of one of the harmonics of the cyclotron frequency.

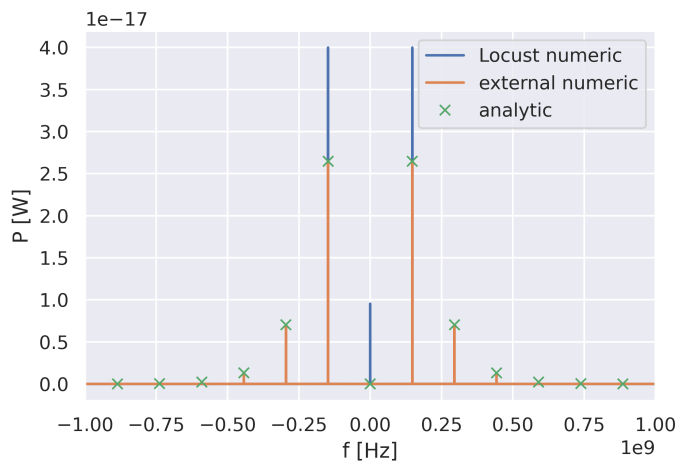
B.2 Incorrect Field Results

The second issue that was found are the numeric solutions of the electromagnetic fields with eqs. (4.21) and (4.22). From section 4.3 the angular distribution of the radiated power and the power spectrum including harmonics are known (see

figs. 4.5 and 4.6). Both can be calculated from the numeric result as well via the Poynting vector \mathbf{S} and the effective area A_{eff} . Figure B.3a shows a comparison of the angular power distribution between the Locust field solutions, an external implementation of the field solutions and the theory from fig. 4.5. The power in Locust is overall too high and the angular distribution does not follow the same shape. Figure B.3b shows the power spectrum of the field for an observer at $\theta = 90^\circ$. The Locust solution has an unexpected peak at $f = 0$ Hz (DC) and otherwise does not distribute the correct fractions of power into the different harmonics.



(a) Comparison of angular power distribution between Locust field solutions and theory.



(b) Comparison of the power spectrum between Locust field solutions and theory for a polar angle of $\theta = 90^\circ$.

Fig. B.3.: Comparison of angular power distribution and power spectrum of field solutions between Locust and theory.

Additional Proofs and Calculations

C.1 Calculation of the Orthogonal Gradient for the Drift Motion

In this section the orthogonal gradient that appears in the drift motion is calculated for a rotationally symmetric magnetic field.

In eq. (4.7) $\nabla_{\perp} B$ denotes the gradient of the magnitude B in the plane orthogonal to \mathbf{B} . This plane is defined by the spherical basis vector of \mathbf{B} , which is

$$\frac{\mathbf{B}(\rho, z)}{B(\rho, z)} = \hat{\mathbf{e}}_r = \begin{pmatrix} \sin(\theta) \cos(\phi) \\ \sin(\theta) \sin(\phi) \\ \cos(\theta) \end{pmatrix},$$

with $\theta = \arccos\left(\frac{B_z(\rho, z)}{B(\rho, z)}\right)$ and $\phi = \arctan 2(B_y(\rho, z), B_x(\rho, z))$. The other two spherical basis vectors are

$$\hat{\mathbf{e}}_{\theta} = \begin{pmatrix} \cos(\theta) \cos(\phi) \\ \cos(\theta) \sin(\phi) \\ -\sin(\theta) \end{pmatrix} = \frac{1}{B(\rho, z)} \begin{pmatrix} \cos(\phi) B_z(\rho, z) \\ \sin(\phi) B_z(\rho, z) \\ -B_{\rho}(\rho, z) \end{pmatrix}, \quad \hat{\mathbf{e}}_{\phi} = \begin{pmatrix} -\sin(\phi) \\ \cos(\phi) \\ 0 \end{pmatrix} \quad (\text{C.1})$$

and they define the gradient as

$$\nabla_{\perp} B = D_{\hat{\mathbf{e}}_{\theta}} B(\rho, z) \hat{\mathbf{e}}_{\theta} + D_{\hat{\mathbf{e}}_{\phi}} B(\rho, z) \hat{\mathbf{e}}_{\phi}. \quad (\text{C.2})$$

In eq. (C.2) $D_{\mathbf{v}} f = \nabla f \cdot \mathbf{v}$ denotes the directional derivative of f with respect to \mathbf{v} . Hence, the calculation of these directional derivatives requires the gradient of the field magnitude $B(\rho, z)$ which in cylindrical coordinates is

$$\nabla B(\rho, z) = \frac{\partial B(\rho, z)}{\partial \rho} \hat{\mathbf{e}}_{\rho} + \frac{1}{\rho} \frac{\partial B(\rho, z)}{\partial \phi} \hat{\mathbf{e}}_{\phi} + \frac{\partial B(\rho, z)}{\partial z} \hat{\mathbf{e}}_z. \quad (\text{C.3})$$

With $B(\rho, z) = |\mathbf{B}(\rho, z)| = \sqrt{B_\rho(\rho, z)^2 + B_z(\rho, z)^2}$ the derivatives for the gradient are

$$\frac{\partial B(\rho, z)}{\partial \rho} = \frac{1}{B(\rho, z)} \left(B_\rho(\rho, z) \frac{\partial B_\rho(\rho, z)}{\partial \rho} + B_z(\rho, z) \frac{\partial B_z(\rho, z)}{\partial \rho} \right), \quad (\text{C.4})$$

$$\frac{\partial B(\rho, z)}{\partial \phi} = 0, \quad (\text{C.5})$$

$$\frac{\partial B(\rho, z)}{\partial z} = \frac{1}{B(\rho, z)} \left(B_\rho(\rho, z) \frac{\partial B_\rho(\rho, z)}{\partial z} + B_z(\rho, z) \frac{\partial B_z(\rho, z)}{\partial z} \right). \quad (\text{C.6})$$

Using eqs. (C.3) to (C.6) one can calculate the directional derivatives of eq. (C.2). For the second term one finds that

$$D_{\hat{e}_\phi} B(\rho, z) = \left(\frac{\partial B(\rho, z)}{\partial \rho} \hat{e}_\rho + \frac{\partial B(\rho, z)}{\partial z} \hat{e}_z \right) \cdot \hat{e}_\phi = 0$$

because $\hat{e}_\rho \cdot \hat{e}_\phi = \hat{e}_z \cdot \hat{e}_\phi = 0$ as the three vectors are orthogonal to each other. The other derivative requires the inner product of \hat{e}_θ from eq. (C.1) with the two standard cylindrical basis vectors

$$\hat{e}_\rho = \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \\ 0 \end{pmatrix}, \quad \hat{e}_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The results are

$$\hat{e}_\rho \cdot \hat{e}_\theta = \frac{B_z(\rho, z)}{B(\rho, z)}, \quad \hat{e}_z \cdot \hat{e}_\theta = -\frac{B_\rho(\rho, z)}{B(\rho, z)},$$

which yields

$$\begin{aligned} D_{\hat{e}_\theta} B(\rho, z) &= \left(\frac{\partial B(\rho, z)}{\partial \rho} \hat{e}_\rho + \frac{\partial B(\rho, z)}{\partial z} \hat{e}_z \right) \cdot \hat{e}_\theta \\ &= \frac{1}{B(\rho, z)} \left(\frac{\partial B(\rho, z)}{\partial \rho} B_z(\rho, z) - \frac{\partial B(\rho, z)}{\partial z} B_\rho(\rho, z) \right). \end{aligned}$$

Finally, using the derivatives from eqs. (C.4) to (C.6) the final result is

$$\begin{aligned} D_{\hat{e}_\theta} B(\rho, z) &= \frac{1}{B(\rho, z)^2} \left(B_z(\rho, z)^2 \frac{\partial B_z(\rho, z)}{\partial \rho} - B_\rho(\rho, z)^2 \frac{\partial B_\rho(\rho, z)}{\partial z} \right. \\ &\quad \left. + B_z(\rho, z) B_\rho(\rho, z) \left(\frac{\partial B_\rho(\rho, z)}{\partial \rho} - \frac{\partial B_z(\rho, z)}{\partial z} \right) \right). \end{aligned} \quad (\text{C.7})$$

Thus, the orthogonal gradient for the grad-B drift (eq. (4.7)) points in direction \hat{e}_θ and has a magnitude $|\nabla_\perp B| = D_{\hat{e}_\theta} B(\rho, z)$ which can be calculated with eq. (C.7) for any rotationally symmetric field.

C.2 Proofs for Matched Filter SNR

This section addresses the derivation of the expectation value and variance of the matched filter test statistic for section 6.3.5. The test statistic given in eq. (6.14) is

$$T(\mathbf{x}) = \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{x} \right).$$

Under \mathcal{H}_1 $\mathbf{x} = \mathbf{s} + \mathbf{n}$, which results in the following for the expectation value

$$\langle T(\mathbf{x}; \mathcal{H}_1) \rangle = \langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} (\mathbf{s} + \mathbf{n}) \right) \rangle = \langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} \right) \rangle + \langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right) \rangle.$$

In the above expression $\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} = (\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s})^H$, which means that it is real and thus $\langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} \right) \rangle = \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}$. The component $\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n}$ is a complex random vector \mathbf{n}' whose entries $n'_i = \sum_j \mathbf{C}_{ij}^{-1} n_j$ are normally distributed with means $\mu_i = \sum_j \mathbf{C}_{ij}^{-1} \mu_j$. The μ_j are the means of the original n_j , which are zero. Consequently, $\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} = \sum_j \mathbf{s}_j^* n'_j$ is a normally distributed complex random variable with mean $\mu = \sum_i \mathbf{s}_i^* \mu'_i = 0$. Therefore, $\langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right) \rangle = 0$ and the result for the expectation value is

$$\langle T(\mathbf{x}; \mathcal{H}_1) \rangle = \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}.$$

The variance is given by

$$\begin{aligned} \text{Var} (T(\mathbf{x}; \mathcal{H}_1)) &= \langle T(\mathbf{x}; \mathcal{H}_1)^2 \rangle - \langle T(\mathbf{x}; \mathcal{H}_1) \rangle^2 \\ &= \langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} (\mathbf{s} + \mathbf{n}) \right)^2 \rangle - \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} \right)^2 \\ &= \langle \left(\text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} \right) + \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right) \right)^2 \rangle - \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} \right)^2 \\ &= \langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right)^2 \rangle + 2 \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s} \langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right) \rangle \\ &= \langle \text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right)^2 \rangle. \end{aligned}$$

Using polar notation with $\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} = e^{i\varphi} \left| \mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right|$ yields

$$\text{Re} \left(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right) = \cos(\varphi) \left| \mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \right|. \quad (\text{C.8})$$

As demonstrated above, $\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n}$ is a complex normal random variable with zero mean. Moreover, its real and imaginary parts are uncorrelated. This is due to the fact that \mathbf{n} is a complex random vector characterized by a real covariance matrix and thus the real and imaginary parts of its components are initially uncorrelated

which persists after the linear transformation. Therefore, the angle φ in eq. (C.8) follows a uniform distribution on the interval $[0, 2\pi]$ and it is independent of the absolute value $|\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n}|$ which has a Rayleigh distribution.

The considerations above imply that

$$\langle \text{Re}(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n})^2 \rangle = \langle \cos^2(\varphi) |\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n}|^2 \rangle = \langle \cos^2(\varphi) \rangle \langle |\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n}|^2 \rangle$$

The first of the two expectation values can be calculated with a simple integration as

$$\langle \cos^2(\varphi) \rangle = \int_0^{2\pi} \frac{1}{2\pi} \cos^2(\varphi) d\varphi = \frac{1}{2}.$$

Using this result, the final expression for the variance can be derived

$$\begin{aligned} \text{Var}(T(\mathbf{x}; \mathcal{H}_1)) &= \frac{1}{2} \langle |\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n}|^2 \rangle \\ &= \frac{1}{2} \langle (\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n})(\mathbf{s}^H \mathbf{C}^{-1} \mathbf{n})^H \rangle \\ &= \frac{1}{2} \langle \mathbf{s}^H \mathbf{C}^{-1} \mathbf{n} \mathbf{n}^H (\mathbf{C}^{-1})^H \mathbf{s}^H \rangle \\ &= \frac{1}{2} \mathbf{s}^H \mathbf{C}^{-1} \langle \mathbf{n} \mathbf{n}^H \rangle (\mathbf{C}^{-1})^H \mathbf{s}^H \\ &= \frac{1}{2} \mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}^H. \end{aligned}$$

In the last step above it was used that $\langle \mathbf{n} \mathbf{n}^H \rangle = \mathbf{C}$ with $\mathbf{C}^H = \mathbf{C}$.

C.3 Calculation of Effective Trap Volume

The effective trap volume introduced in section 8.1.2 is defined as the expectation value

$$V_t = \langle V(\alpha_0) \rangle_{\alpha_0} = \int_{\alpha_{min}}^{\pi - \alpha_{min}} V(\alpha_0) p_{\alpha_0}(\alpha_0) d\alpha_0 = 2 \int_{\alpha_{min}}^{\pi/2} V(\alpha_0) p_{\alpha_0}(\alpha_0) d\alpha_0, \quad (\text{C.9})$$

where $V(\alpha_0)$ is the volume integral

$$V(\alpha_0) = \int_0^{2\pi} \int_0^R \int_{-z_{max}(r, \alpha_0)}^{z_{max}(r, \alpha_0)} r dz dr d\phi.$$

The calculation of this expectation value requires $p_{\alpha_0}(\alpha_0)$, the probability density of producing a decay electron with pitch angle α_0 . Given the joint probability density

$p_{r\phi z\alpha_0}(r, \phi, z, \alpha_0)$ of electron position in cylindrical coordinates and pitch angle, $p_{\alpha_0}(\alpha_0)$ is the marginal distribution of α_0 :

$$p_{\alpha_0}(\alpha_0) = \int_0^R \int_0^{2\pi} \int_{-z_{max}(r, \alpha_0)}^{z_{max}(r, \alpha_0)} p_{r\phi z\alpha_0}(r, \phi, z, \alpha_0) r \, dz \, d\phi \, dr. \quad (\text{C.10})$$

The initial electron positions are drawn from a uniform distribution that spans the volume $V(\alpha_0)$, whereas the pitch angles are determined by the direction of the initial momentum, which is isotropically distributed. The probability density for a decay electron's initial direction can be defined as

$$p_{\phi_*\alpha_*}(\phi_*, \alpha_*) = \frac{1}{4\pi} \sin(\alpha_*), \quad (\text{C.11})$$

where α_* is the electron's instantaneous pitch angle at the time of its creation¹ and ϕ_* is the corresponding azimuth angle.

The construction of the joint distribution $p_{r\phi z\alpha_0}(r, \phi, z, \alpha_0)$ from eq. (C.11) requires a change of variables from α_* to α_0 . According to eq. (4.5) the conversion from α_* to α_0 is defined as

$$\sin(\alpha_*) = \sin(\alpha_0) \sqrt{\frac{B(r, z)}{B(r, 0)}}.$$

Therefore, the joint distribution is

$$\begin{aligned} p_{r,\phi,z,\alpha_0}(r, \phi, z, \alpha_0) &= \frac{1}{V_a(\alpha_0)} \int_0^{2\pi} p_{\phi_*,\alpha_*}(\phi_*, \alpha_*(\alpha_0)) \, d\phi_* \quad (\text{C.12}) \\ &= \underbrace{\frac{1}{V_a(\alpha_0)}}_{\text{Position normalization}} \underbrace{\frac{1}{2} \sin(\alpha_0) \sqrt{\frac{B(r, z)}{B(r, 0)}}}_{\sin(\alpha_*)} \underbrace{\sqrt{\frac{1 - \sin^2(\alpha_0)}{\frac{B(r, 0)}{B(r, z)} - \sin^2(\alpha_0)}}}_{\frac{d\alpha_*}{d\alpha_0}}. \end{aligned}$$

Finally, combining eqs. (C.9), (C.10) and (C.12) yields the result

$$V_t = 2\pi \int_0^R \int_{\alpha_{min}}^{\pi/2} \int_{-z_{max}(r, \alpha_0)}^{z_{max}(r, \alpha_0)} \sqrt{\frac{B(r, z) (1 - \sin^2(\alpha_0))}{B(r, 0) \left(\frac{B(r, 0)}{B(r, z)} - \sin^2(\alpha_0) \right)}} \sin(\alpha_0) r \, dz \, d\alpha_0 \, dr.$$

¹Reminder: α_0 is the pitch angle at the center of the trap, which is different from α_* .

C.4 CRLB of a Chirp with Dependent Chirp Rate

In this section the CRLB of a chirp with a dependent chirp rate will be derived. The derivation takes inspiration from the approach taken in [130] for the CRLB of the regular chirp.

The signal model is defined as

$$\mathbf{s}(\boldsymbol{\theta})_j = A g(\mathbf{t}_j - t_0) e^{i\left(\omega_0(\mathbf{t}_j - t_0) + \frac{\delta\omega(\omega_0)}{2}(\mathbf{t}_j - t_0)^2 + \varphi_0\right)},$$

where the free parameters $\boldsymbol{\theta} = (A, \omega_0, \varphi_0, t_0)$ are amplitude A , start frequency ω_0 , initial phase φ_0 , and t_0 and g is the window function. Notably, the chirp rate $\delta\omega(\omega_0)$ is not a free parameter.

According to eq. (7.4), the variance of ω_0 has the lower limit

$$\text{Var}(\omega_0) \geq \mathbf{I}^{-1}(\boldsymbol{\theta})_{11},$$

where \mathbf{I} is the Fisher information matrix. For complex WGN the Fisher information matrix can be calculated with [122]

$$\mathbf{I}(\boldsymbol{\theta})_{ij} = \frac{2}{\sigma^2} \text{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \boldsymbol{\theta}_i} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \right). \quad (\text{C.13})$$

The required derivatives are:

$$\begin{aligned} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial A} &= \frac{\mathbf{s}(\boldsymbol{\theta})}{A}, \\ \frac{\partial \mathbf{s}(\boldsymbol{\theta})_j}{\partial \omega_0} &= i \mathbf{s}(\boldsymbol{\theta})_j (\mathbf{t}_j - t_0) \left(1 + \frac{1}{2} \frac{\partial \delta\omega(\omega_0)}{\partial \omega_0} (\mathbf{t}_j - t_0) \right), \\ \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \varphi_0} &= i \mathbf{s}(\boldsymbol{\theta}), \\ \frac{\partial \mathbf{s}(\boldsymbol{\theta})_j}{\partial t_0} &= -\mathbf{s}(\boldsymbol{\theta})_j \left(i (\omega_0 + \delta\omega(\omega_0)(\mathbf{t}_j - t_0)) + \frac{g'(\mathbf{t}_j - t_0)}{g(\mathbf{t}_j - t_0)} \right). \end{aligned}$$

Applying these derivatives to evaluate eq. (C.13), certain sums emerge that can again be identified as the moments $M_k = \sum_{j=-m}^{n_s-1-m} \mathbf{t}_j^k g^2(\mathbf{t}_j)$ and $\dot{M}_k = \sum_{j=-m}^{n_s-1-m} \mathbf{t}_j^k \dot{g}^2(\mathbf{t}_j)$, where n_s is the absolute number of samples, and m is the index of the start time such that $t_0 = \Delta t m$ with sampling time step Δt . Assuming a rectangular window, the moments simplify to $M_k = \sum_{j=0}^{n_\tau-1} \Delta t^k j^k$, where n_τ is the number of signal

samples. Defining the shorthand $m_k = \sum_{j=0}^{n_\tau-1} j^k$, the non-zero elements of the Fisher information matrix can be expressed as:

$$\begin{aligned}
\mathbf{I}(\boldsymbol{\theta})_{00} &= \frac{2}{\sigma^2} \operatorname{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial A} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial A} \right) = \frac{2}{\sigma^2} m_0 \\
\mathbf{I}(\boldsymbol{\theta})_{11} &= \frac{2}{\sigma^2} \operatorname{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \omega_0} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \omega_0} \right) \\
&= \frac{2A^2 \Delta t^2}{\sigma^2} \left(m_2 + \left(\frac{\partial \delta \omega}{\partial \omega_0} \Delta t \right) m_3 + \frac{1}{4} \left(\frac{\partial \delta \omega}{\partial \omega_0} \Delta t \right)^2 m_4 \right) \\
\mathbf{I}(\boldsymbol{\theta})_{22} &= \frac{2}{\sigma^2} \operatorname{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \varphi_0} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \varphi_0} \right) = \frac{2A^2}{\sigma^2} m_0 \\
\mathbf{I}(\boldsymbol{\theta})_{33} &= \frac{2}{\sigma^2} \operatorname{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial t_0} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial t_0} \right) \\
&= \frac{2A^2}{\sigma^2} \left(\dot{M} + \omega_0^2 m_0 + \delta \omega^2 \Delta t^2 m_2 + 2\omega_0 \delta \omega \Delta t m_1 \right) \\
\mathbf{I}(\boldsymbol{\theta})_{12} &= \frac{2}{\sigma^2} \operatorname{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \omega_0} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \varphi_0} \right) \\
&= \frac{2A^2 \Delta t}{\sigma^2} \left(m_1 + \frac{1}{2} \left(\frac{\partial \delta \omega}{\partial \omega_0} \Delta t \right) m_2 \right) \\
\mathbf{I}(\boldsymbol{\theta})_{13} &= \frac{2}{\sigma^2} \operatorname{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \omega_0} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial t_0} \right) \\
&= -\frac{2A^2 \Delta t}{\sigma^2} \left(\omega_0 m_1 + \delta \omega \Delta t m_2 + \frac{1}{2} \left(\frac{\partial \delta \omega}{\partial \omega_0} \Delta t \right) (\omega_0 m_2 + \Delta t m_3) \right) \\
\mathbf{I}(\boldsymbol{\theta})_{23} &= \frac{2}{\sigma^2} \operatorname{Re} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \varphi_0} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial t_0} \right) \\
&= -\frac{2A^2}{\sigma^2} (\omega_0 m_0 + \delta \omega \Delta t m_1)
\end{aligned}$$

For the usual range of magnetic fields the derivative $\frac{\partial \delta \omega}{\partial \omega_0}$ is small compared to the sampling rate, e.g. $\frac{\partial \delta \omega(\omega_0)}{\partial \omega_0} = -0.0006 \text{ s}^{-1}$ for $B = 0.04 \text{ T}$ and $\frac{\partial \delta \omega(\omega_0)}{\partial \omega_0} = -0.35 \text{ s}^{-1}$ for $B = 0.96 \text{ T}$. This implies that $\Delta t \frac{\partial \delta \omega(\omega_0)}{\partial \omega_0} \ll 1$. Therefore, all terms in the Fisher information matrix that scale with powers of $\Delta t \frac{\partial \delta \omega(\omega_0)}{\partial \omega_0}$ are negligible. Setting these terms to zero, inversion of the matrix and evaluation of the sums with Mathematica [89] yields the CRLB of ω_0 :

$$\operatorname{Var}(\hat{\omega}_0) \geq \frac{\sigma^2}{A^2} \left(\frac{\delta \omega^2}{2\dot{M}} + \frac{6}{\Delta t^2 (n_\tau^3 - n_\tau)} \right). \quad (\text{C.14})$$

For the start time t_0 the CRLB result is

$$\text{Var}(\hat{t}_0) \geq \frac{\sigma^2}{2A^2M}. \quad (\text{C.15})$$

Thus, inserting eq. (C.15) into eq. (C.14) results in:

$$\text{Var}(\hat{\omega}_0) \geq \text{Var}(\hat{t}_0) \delta\omega^2 + 6 \frac{\sigma^2}{\Delta t^2 A^2 (n_\tau^3 - n_\tau)}.$$

The result has a familiar structure: It is the CRLB for a pure sinusoid with constant frequency (see eq. (8.5)) with an additional contribution from the uncertainty on the start time, which is structurally similar to the CRLB of the chirp (see eq. (7.9)).

Additional Plots

D.1 Chirp Model Likelihood Scans

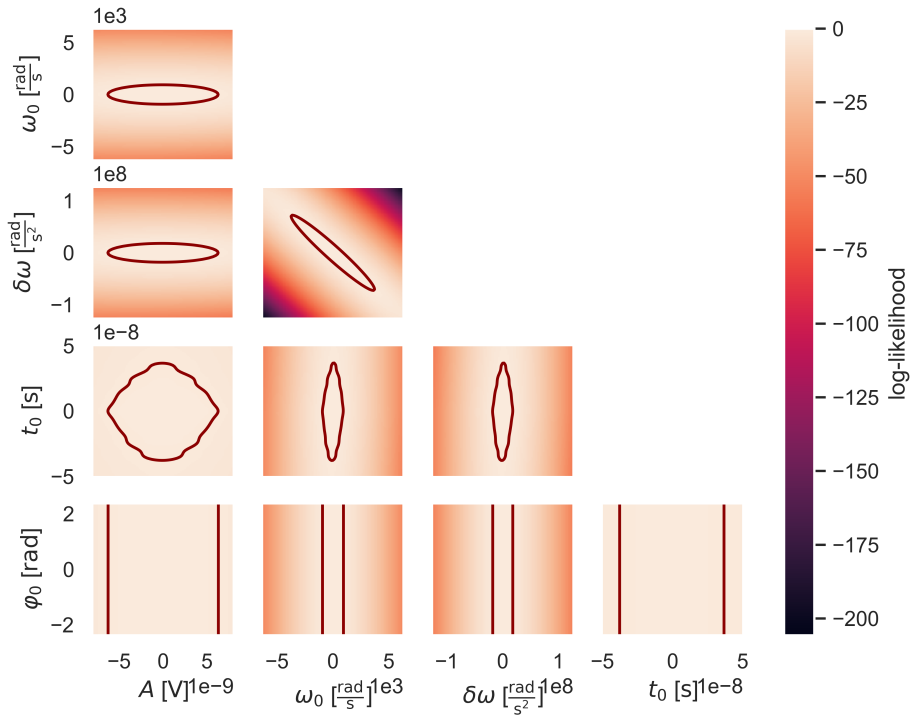


Fig. D.1.: Modified asimov log-likelihood (eq. (7.19)) of a chirp. Parameters A , $\delta\omega$ are set for electron with $E_{\text{kin}} = 18.6$ keV, $\alpha_0 = 90^\circ$ in $B = 0.96$ T. Other model parameters are $\omega_0 = 2\pi \cdot 20$ MHz, $t_0 = 100 \mu\text{s}$, $\varphi_0 = 0$ and with $\tau = 100 \mu\text{s}$, $T = 5$ K, $f_s = 200$ MHz. Each 2D plot fixes remaining parameters at their true values, with contour lines indicating the 68% confidence region from Wilks' theorem. The origin of the coordinate axes is set to the true parameter values.

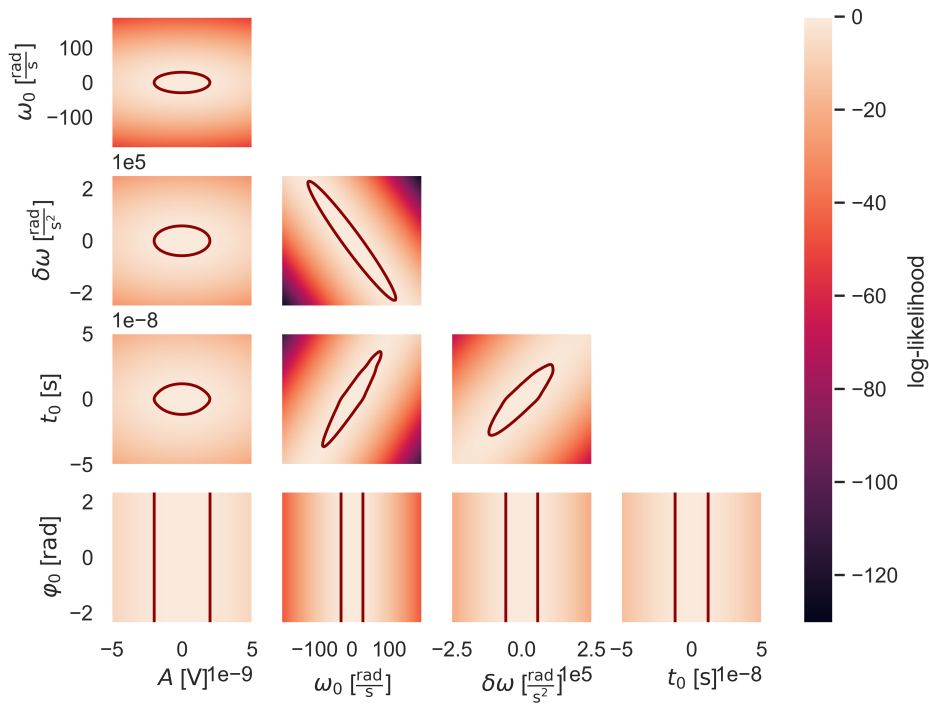


Fig. D.2.: Modified asimov log-likelihood (eq. (7.19)) of a chirp. Parameters A , $\delta\omega$ are set for electron with $E_{\text{kin}} = 18.6 \text{ keV}$, $\alpha_0 = 90^\circ$ in $B = 0.96 \text{ T}$. Other model parameters are $\omega_0 = 2\pi 20 \text{ MHz}$, $t_0 = 100 \mu\text{s}$, $\varphi_0 = 0$ and with $\tau = 1000 \mu\text{s}$, $T = 5 \text{ K}$, $f_s = 200 \text{ MHz}$. Each 2D plot fixes remaining parameters at their true values, with contour lines indicating the 68% confidence region from Wilks' theorem. The origin of the coordinate axes is set to the true parameter values.

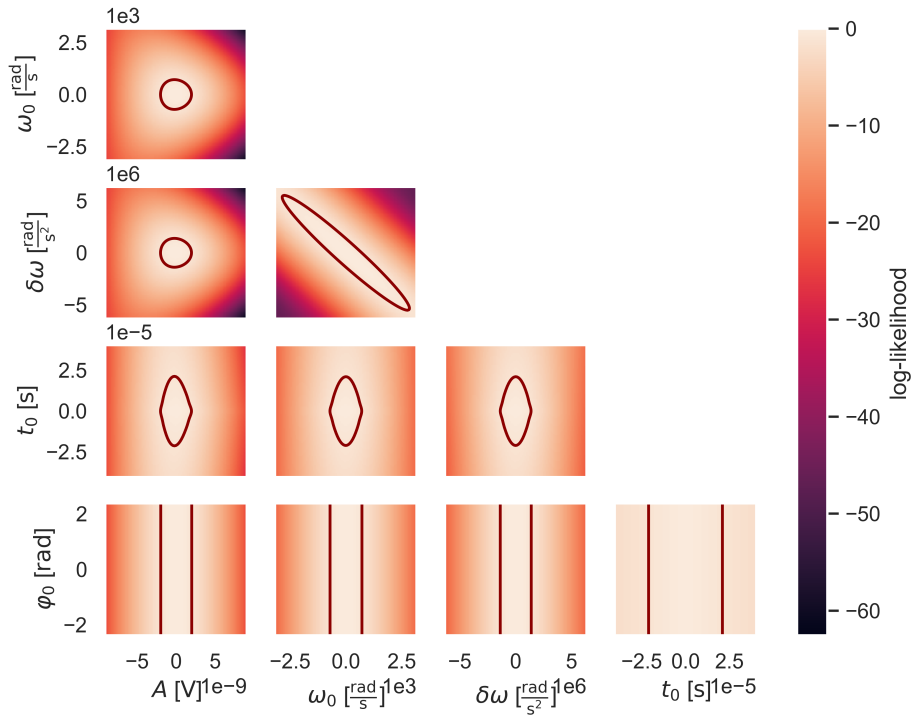


Fig. D.3.: Modified asimov log-likelihood (eq. (7.19)) of a chirp. Parameters A , $\delta\omega$ are set for electron with $E_{\text{kin}} = 18.6$ keV, $\alpha_0 = 90^\circ$ in $B = 0.04$ T. Other model parameters are $\omega_0 = 2\pi 20$ MHz, $t_0 = 100 \mu\text{s}$, $\varphi_0 = 0$ and with $\tau = 1000 \mu\text{s}$, $T = 5$ K, $f_s = 200$ MHz. Each 2D plot fixes remaining parameters at their true values, with contour lines indicating the 68% confidence region from Wilks' theorem. The origin of the coordinate axes is set to the true parameter values.

D.2 Comparison of High and Low Gain Antennas

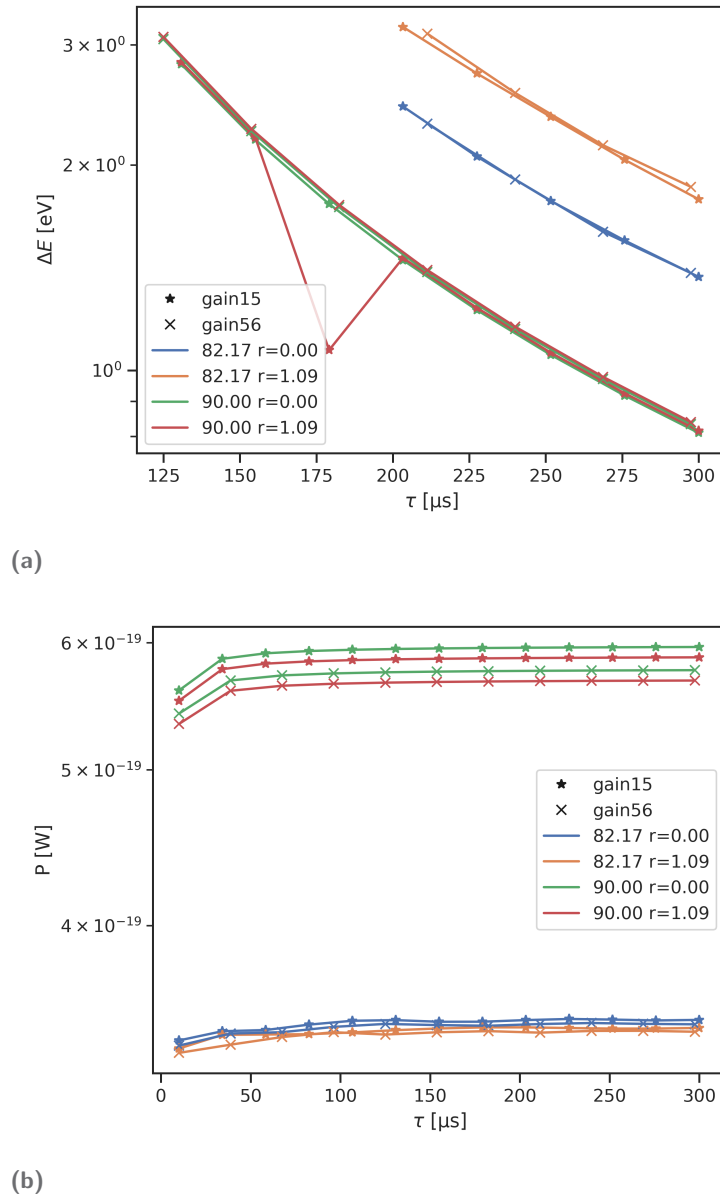


Fig. D.4.: Energy resolution ΔE (a) and detected power P (b) as function of track length τ for 4 different electrons. Setup and calculation details in chapter 8. Results for identical setups with the same total effective aperture with gains of 15 and 56 are compared.

D.3 Full Scale Detector Likelihood Scans

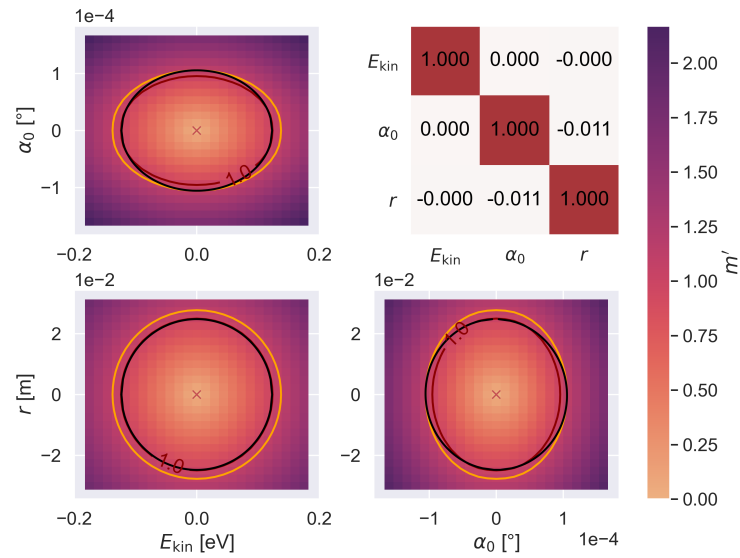


Fig. D.5.: Scans of likelihood of the demo setup in the three 2D-planes for $\alpha_0 = 90^\circ$, $r = 0$ m, $\tau = 671 \mu\text{s}$. Likelihood values are converted to m' - σ levels (see eq. (7.16)). Red lines indicate the $1\text{-}\sigma$ confidence level, black lines the estimated $1\text{-}\sigma$ error ellipses. The coordinate origin is shifted to the true parameter values (marked with "x"). Top right depicts estimated correlation matrix with colors ranging from blue (perfect anti-correlation) to red (perfect correlation).

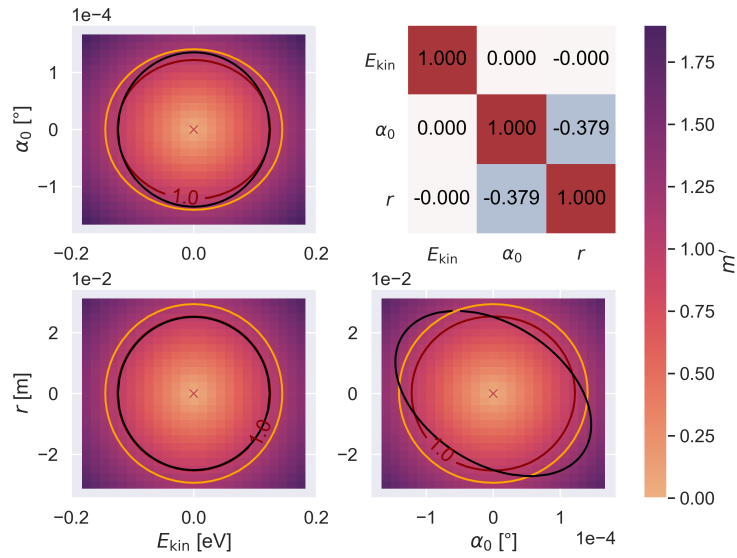


Fig. D.6.: 2D scans of the likelihood function of the demo setup for $\alpha_0 = 90^\circ$, $r = 1.09$ m, $\tau = 671 \mu\text{s}$. For further details see caption of fig. D.5.

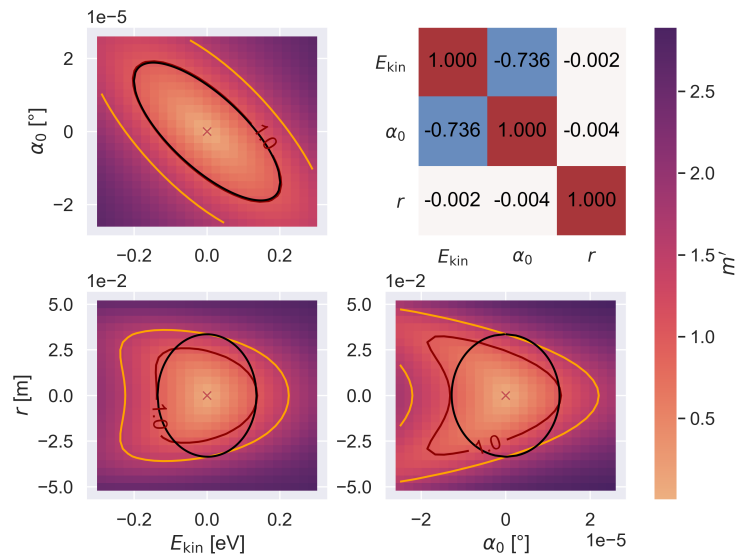


Fig. D.7.: 2D scans of the likelihood function of the demo setup for $\alpha_0 = 82.17^\circ$, $r = 1.09$ m, $\tau = 671 \mu\text{s}$. For further details see caption of fig. D.5.

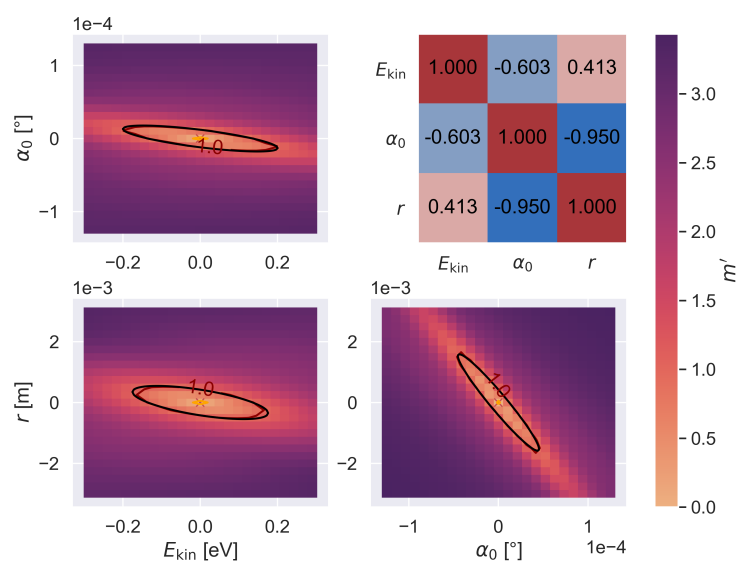


Fig. D.8.: 2D scans of the likelihood function of the demo setup for $\alpha_0 = 82.17^\circ$, $r = 1.09$ m, $\tau = 671 \mu\text{s}$. For further details see caption of fig. D.5.

Lebenslauf

[In der elektronischen Version aus Datenschutzgründen entfernt]

Colophon

This thesis was typeset with $\text{\LaTeX}2_{\epsilon}$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at <http://cleanthesis.der-ric.de/>.

