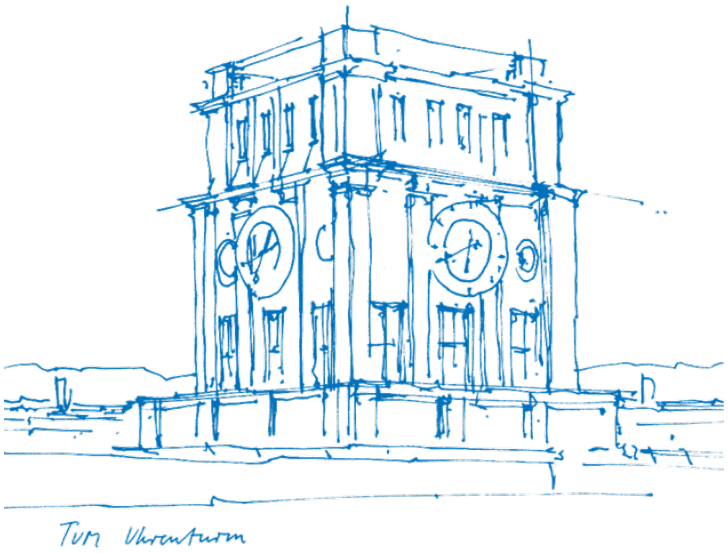
TUM School of Natural Sciences
Technical University of Munich

Fierz Interference Term in Neutron Beta Decay and New Approaches for Systematic Uncertainty Quantification

Max Lamparth



TUM School of Natural Sciences
Technische Universität München

Fierz Interference Term in Neutron Beta Decay and New Approaches for Systematic Uncertainty Quantification

Max Lamparth

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:

apl. Prof. Dr. Antonio Vairo

Prüfer*innen der Dissertation:

1. Prof. Dr. Bastian Märkisch
2. Prof. Dr. Lukas Heinrich

Die Dissertation wurde am 10.07.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 21.07.2023 angenommen.

Abstract

High-precision experiments of beta decays of unbound neutrons provide a strong lever to probe the structure of the weak interaction for potential tensor and scalar contributions. Such contributions would indicate physics beyond the Standard Model and can be measured as an energy-dependent shift in the electron energy distribution from neutron beta decay and are parametrized in the Fierz interference term b. Such studies require exact calibrations, understanding and control of measurement parameters, and quantifying systematic uncertainties.

This thesis focuses on determining systematic uncertainties and corrections in an electron energy measurement with the PERKEO III spectrometer conducted at the Institut Laue-Langevin in 2019/20. I present the data analysis of this measurement with the developed corrections to estimate systematic uncertainties, test hypotheses of their causes, and propose studies for further evaluation. To this end, I developed four analysis tools and studied new approaches for high-precision experiments, ranging from variational inference, deep learning, and causal inference to differentiable programming. Applications range from low-energy particle physics to state-of-the-art computational astrophysics simulations.

Contents

1 Introduction

High-precision experiments are one of the main drivers of searches for new fundamental forces and particles. Contrary to their high-energy counterparts that also use increasing energy scales and extensive collision data to look for rare, unusual events from new interactions or particles, high-precision experiments mainly focus on improving studies of known interactions and particles at lower energies to search for deviations and indicators of new physics models. These high-precision experiments require exact calibrations, understanding, and control of measurement parameters and related systematic uncertainties. While the Standard Model of particle physics successfully explains all known fundamental particles and the dominant forces at these scales, there are observations indicating its incompleteness that require physics models beyond the Standard Model. With high-precision experiments, we can gain insights into new processes to test existing models and advance our understanding.

As a pure weak interaction process at low-energy scales, studies of beta decays of free neutrons provide a strong lever to probe the structure of the weak interaction for potential tensor and scalar contributions, which would indicate physics beyond the Standard Model. Knowledge of the structure of the weak interaction is essential to guide new physics searches, similar to how the early studies of Fermi interaction in beta decay contributed to the discovery of the weak interaction and the formulation of the Standard Model. Constant improvements in high-precision measurements to continuously drive these searches require technological advances in experiment equipment and analysis tools. These advances make it more complex and critical to accurately and precisely estimate systematic uncertainties for new physical results.

To this end, this thesis focuses on determining and estimating systematic uncertainties and their causes for a precise electron energy measurement in neutron beta decay with the PERKEO III spectrometer to search for beyond Standard Model physics in left-handed tensor interactions. I summarize the related work to link the electron energy measurement to scalar and tensor interactions and recapture the measurement campaign and experiment setup conducted under my management at the Institut Laue-Langevin in 2019/20 [Lam19] to collect the required data from neutron beta decay. I discuss the data analysis of the PERKEO III measurement to estimate systematic uncertainties, test hypotheses of causes to increase understanding, and propose studies for further evaluation. The work in this thesis is also the result of collaborations and significant contributions of other theses [Ant19; Kro20; Fal22; Bes22].

In this thesis, I developed four analysis tools and frameworks to enable specific analysis and to impact work beyond this thesis. I study new analysis approaches for high-precision experiments, ranging from variational inference, deep learning, and causal inference to differentiable programming. My studies cover applications in low-energy particle physics to state-of-the-art computational astrophysics simulations and also lead to four scientific works [Wan+19; LBM22; Lam+22a; Dor+23]. Furthermore, I examine the systematic uncertainties of the electron energy spectrum measurement with PERKEO III and highlight areas with the highest potential for improvement to guide future high-precision experiments.

2 Physical Motivation

The *Standard Model* of particle physics successfully explains a wide range of physical phenomena in the universe. It describes all known fundamental particles and the dominant forces acting between them at these scales through quantum field theories (QFTs). It outlines the interactions of leptons (e.g., electrons), quarks (the constituents of, e.g., neutrons), and force-carrying bosons (e.g., photons). However, there are still observations that cannot be explained by it, suggesting that physics beyond the Standard Model is needed. It does not incorporate gravity, explain the imbalance of matter and antimatter in the universe, or predict the masses of neutrinos. It also cannot explain dark matter or dark energy, which comprise most of the universe.

This chapter focuses on searches for beyond Standard Model (BSM) physics through scalar and tensor contributions in the weak interaction in neutron beta decay. Beta decays have been essential throughout the development of the Standard Model [LY56; Cow+56; Wei58], e.g., to discover the parity-violating structure of the weak interaction [Wu+57]. They continue to be relevant to particle physics research through measurements of Standard Model parameters and their contribution to constraining BSM physics searches.

2.1 Beta Decay in the Standard Model

In the Standard Model, the weak interaction is the fundamental force mediating beta decay. It violates parity symmetry, allows quark flavor changes, and is transmitted by the heavy $W^{\pm}$ and Z bosons, the only force-carrying bosons with mass in the Standard Model. The weak interaction is an essential link between nuclear and particle physics.

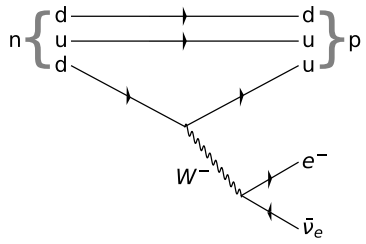
To study this interaction, we will particularly focus on the beta decay of a free neutron n into a proton p, an electron e, and an anti-electron neutrino $\bar{\nu}_e$ as

$$n \rightarrow p + e + \bar{\nu}_e.$$

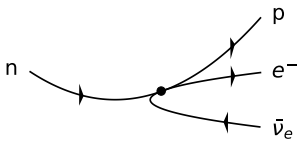
The decay is mediated through the weak interaction as illustrated in the Feynman diagram in Fig. 2.1a. The matrix element $\mathcal{M}_{\beta^-}$ for this beta decay with Dirac spinors ψ_i for $i \in \{u, d, e, \nu_e\}$, W boson mass m_W, and momentum transfer q^2 is given by

$$\mathcal{M}_{\beta^-} = -\left[\frac{g_W}{\sqrt{2}}\bar{\psi}_u \frac{1}{2}\gamma^\mu(1-\gamma^5)\psi_d\right]\left[\frac{g_{\mu\nu} - q_\mu q_\nu / m_W^2}{q^2 - m_W^2}\right]\left[\frac{g_W}{\sqrt{2}}\bar{\psi}_e \frac{1}{2}\gamma^\mu(1-\gamma^5)\psi_{\nu_e}\right]. \quad (2.1)$$

The left-handed chiral projection operator $\frac{1}{2}(1-\gamma^5)$ contains the parity-violating $V-A$ structure of the weak interaction. In beta decays, the moment transfer q^2 is limited by the available energy E_{β^-} in these processes and is a lot smaller than the mass of the W boson



(a) Feynman diagram of the neutron beta decay within the Standard Model.



(b) Four-point interaction of nucleon and leptonic currents within the EFT framework in Equ. (2.7).

$m_W = 80.377(12)$ GeV [Wor+22]. The available energy is given by the mass differences as [Wor+22]

$$E_{\beta^-} = m_n - m_p - m_e = 782.33341(46)\ \text{keV}.$$

At higher momentum transfers, the weak interaction must be unified with the electromagnetic force to an *electroweak* interaction. However, at the low-energy scales of beta decay, we can describe the interaction as a four-fermion interaction - where these four fermions directly couple to one another at a single vertex, as illustrated in Fig. 2.1b. The Fermi theory approximation of Equ. (2.1) is then given by

$$\mathcal{M}_{\beta^-} = \frac{G_F}{\sqrt{2}} g_{\mu\nu} \left[\bar{\psi}_u \gamma^\mu (1-\gamma^5)\psi_d\right] \left[\bar{\psi}_e \gamma^\mu (1-\gamma^5)\psi_{\nu_e}\right], \tag{2.2}$$

with Fermi constant G_F and $G_F/\sqrt{2} = g_W^2/8m_W^2$. In modern terms, Fermi theory is the low-energy effective field theory for the underlying Standard Model weak interaction process, which is explored in more detail in Sec. 2.2.

Using the Fermi theory framework and Equ. (2.2), we can derive a differential decay rate, integrating over angular dependencies and neutrino energy [Wil82] while neglecting electromagnetic interaction between the beta particles, as

$$\mathrm{d}\Gamma = \frac{G_F^2 |V_{ud}|^2}{32\pi^5} (1+3\lambda^2)\ p_e\ E_e \left(E_{\beta^-} - E_e\right)^2\ \mathrm{d}E_e, \tag{2.3}$$

with electron energy E_e and momentum p_e. Also, Equ. (2.3) illustrates that neutron beta decay can be parametrized by experimentally measuring two quantities: The neutron lifetime τ as integrated phase space and the ratio between axial-vector and vector coupling in the weak interaction λ. These experimental quantities are essential to check, e.g., the unitarity of the CKM matrix by setting limits on the first CKM quark mixing matrix element V_{ud}.

To conclude, we can use the Standard Model to describe beta decay. However, we need a different approach to incorporate BSM physics in experimental quantities.

2.2 Weak Effective Field Theory

To study BSM physics, scientists can use the Lagrangian $\mathcal{L}$ of the Standard Model and expand it to describe specific physical phenomena at particular energy scales using so-called

effective field theories (EFTs). They provide a framework to systematically investigate BSM physics by incorporating new degrees of freedom at the energy scale of interest while avoiding renormalization criteria in QFTs through suppression for energies larger than their energy scale. We quantify and parametrize the effects of the introduced, non-renormalizable operators of the EFT and its underlying high-energy theory in the so-called *Wilson coefficients* ϵ.

The Standard Model effective field theory (SMEFT) is the most common EFT and extends the Standard Model to higher energies by parameterizing new physics effects at lower energies, allowing us to study BSM physics in a model-independent way and providing guidance for new physics searches that can point us to the correct high-energy theory. However, when studying low-energy interactions, it is possible to formulate a weak EFT (WEFT) [Bha+12], which is constrained to energies at the electroweak scale (2 GeV), while SMEFT focuses on larger energies. The WEFT model assumes massless neutrinos and constrains all BSM physics in its Wilson coefficients ϵ at scales above 2 GeV.

Going to the even smaller energy scales of beta decays (3-momentum transfer $|q| \simeq 1 - 10$ MeV), we can use WEFT and formulate a *pionless EFT* [Kol99]. We describe the transitions in beta decays with non-relativistic quantum fields ψ_N for nucleons, i.e., the protons p and neutrons n, while using relativistic quantum fields for the electrons e and left-handed neutrinos ν_L. This distinction means going from quark-level vertices to nucleon-level four-point interactions [CB83].

2.2.1 EFT Lagrangian

The leading order subset of the pionless EFT Lagrangian $\mathcal{L}_{\not\pi\text{EFT}}$ for transitions in beta decays is given by [Kol99; Fal+21]

$$\mathcal{L}_{\not\pi\text{EFT}} \supset \mathcal{L}^{(0)} + O\left(\nabla/m_N\right) + \text{ h.c. }, \tag{2.4}$$

expanded in orders of spatial derivatives ∇ over nucleon mass $m_N = (m_p + m_n)/2$. The Lagrangian in Equ. (2.4) contains all Lorentz-invariant contributions from scalar S, pseudo-scalar P, axial-vector A, vector V, and tensor T in its nucleon-level Wilson coefficients C_i^+ with $i \in \{S, P, A, V, T\}$. Within the framework of the Standard Model, C_S^+, C_P^+, and C_T^+ are zero. The zeroth expansion term $\mathcal{L}^{(0)}$ is then given by [Fal+21]

$$\mathcal{L}^{(0)} = -\left(\psi_p^\dagger \psi_n\right)\left[C_V^+ \bar{e}_L \gamma^0 \nu_L + C_S^+ \bar{e}_R \nu_L\right] + \left(\psi_p^\dagger \sigma^k \psi_n\right)\left[C_A^+ \bar{e}_L \gamma^k \nu_L + C_T^+ \bar{e}_R \gamma^0 \gamma^k \nu_L\right], \tag{2.5}$$

with Pauli matrices σ^k, $\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}$, $\sigma^\mu = \left(\sigma^0, \sigma^k\right)$, and $\bar{\sigma}^\mu = \left(\sigma^0, -\sigma^k\right)$. At the zeroth level, the Lagrangian matches the historical Fermi and Gamow-Teller transitions and Lee-Yang notation $C_i^+ = C_i + C_i'$ [LY56]. In the limit $q \to 0$, we can match the nucleon-level Wilson coefficients C_i^+ with the quark-level WEFT Wilson coefficients ϵ and nucleon charges

$g_i = g_i(q^2 = 0)$ with $i \in \{S, A, V, T, IS\}$ (induced scalar IS), leading to [Wei58; Hol74; Bha+12; Fal+21]

$$\begin{aligned}
C_V^+ &= \frac{V_{ud}}{\sqrt{2}G_F}\left\{g_V\left(1+\epsilon_L+\epsilon_R\right)\sqrt{1+\Delta_R^V} - \frac{m_e}{m_N}g_T^{(1)}\epsilon_T\right\} \\
C_A^+ &= -\frac{V_{ud}}{\sqrt{2}G_F}\left\{g_A\left(1+\epsilon_L-\epsilon_R\right)\sqrt{1+\Delta_R^A} + 2\frac{m_e}{m_N}g_T^{(3)}\epsilon_T\right\} \\
C_S^+ &= \frac{V_{ud}}{\sqrt{2}G_F}\left\{g_S\epsilon_S + \frac{m_e}{2m_N}g_{IS}\left(1+\epsilon_L+\epsilon_R\right)\right\} \\
C_T^+ &= \frac{V_{ud}}{\sqrt{2}G_F}g_T\epsilon_T
\end{aligned} \tag{2.6}$$

with the first CKM matrix element V_{ud}, electron mass m_e, Fermi constant G_F, inner radiative corrections Δ_R^V and Δ_R^A [Wil82; Gor19; GS21; Cir+22]. Note that in the zeroth expansion term $\mathcal{L}^{(0)}$, there is no nucleon-level pseudoscalar contribution C_P^+. Higher order derivatives of $\mathcal{L}_{\pi\mathrm{EFT}}$ would also introduce effects such as weak magnetism, induced tensor terms, and quark-level tensor interactions. However, due to higher suppression with nucleon mass $1/m_N$, searches are more sensitive for BSM physics contributions through scalar and tensor contributions in the zeroth derivative leading order term.

2.2.2 Neutron Beta Decay Amplitude

With the Lagrangian for nuclear beta decay transitions in Equ. (2.5), we can calculate the beta decay amplitude $\mathcal{M}_{\beta^-}$. Expanded in orders of 3-momenta q over nuclear mass m_N, the leading order amplitude $\mathcal{M}_{\beta^-}^{(0)}$ is given with the Wilson coefficient from Equ. (2.6) and leptonic currents L [Fal+21]

$$\mathcal{M}_{\beta^-}^{(0)} = -\left\langle\psi_p^\dagger\psi_n\right\rangle\left[C_V^+L^0 + C_S^+L\right] + \left\langle\psi_p^\dagger\sigma^k\psi_n\right\rangle\left[C_A^+L^k + C_T^+L^{0k}\right]. \tag{2.7}$$

The leptonic currents L are given through the spinor wave functions of the electron e and electron neutrino ν with chirality L/R

$$\begin{aligned}
L &= \bar{u}_R\left(k_e\right)\nu_L\left(k_\nu\right) \\
L^\mu &= \bar{u}_L\left(k_e\right)\gamma^\mu\nu_L\left(k_\nu\right) \\
L^{0k} &= \bar{u}_R\left(k_e\right)\gamma^0\gamma^k\nu_L\left(k_\nu\right)
\end{aligned} \tag{2.8}$$

with $k \in \{1, 2, 3\}$. Using Fermi's golden rule, we can use Equ. (2.7) to calculate the differential decay width integrated over electron spin for the neutron decay as [JTW57]

$$\begin{aligned}
\frac{\mathrm{d}\Gamma}{\mathrm{d}E_e\,\mathrm{d}\Omega_e\,\mathrm{d}\Omega_\nu} = M_F^2\,F(Z,E_e)\,\left(1+\delta_R\right)\frac{p_eE_e\left(E_{\beta^-}-E_e\right)^2}{64\pi^5}\hat{\xi}\Bigg\{1 + b\frac{m_e}{E_e} + a\left(E_e\right)\frac{\boldsymbol{k}_e\cdot\boldsymbol{k}_\nu}{E_eE_\nu} \\
+\frac{\boldsymbol{J}\cdot\boldsymbol{k}_e}{JE_e}A\left(E_e\right) + \frac{\boldsymbol{J}\cdot\boldsymbol{k}_\nu}{JE_\nu}B\left(E_e\right) + \frac{\boldsymbol{J}\cdot\left(\boldsymbol{k}_e\times\boldsymbol{k}_\nu\right)}{JE_eE_\nu}D\left(E_e\right)\Bigg\},
\end{aligned} \tag{2.9}$$

with normalization $\hat{\xi}$, 3-momentum vector $\boldsymbol{k}_i$ for electron and neutrino, neutron spin $\boldsymbol{J}$, leading order electromagnetic corrections in Fermi function F and radiative corrections δ_R. Equ. (2.9) neglects any higher corrections on the parameters b, a, A, B, and D. Higher order corrections can be found in [Fal+21]. These parameters are the correlation coefficients directly caused by including all Lorentz-invariant contributions in the Lagrangian in

Equ. (2.4). Within the framework of the Standard Model, D and b are zero, and all other correlation coefficients are a function of $\lambda = C_A^+/C_V^+$.

We can further simplify Equ. (2.9) when integrating over all neutron spin directions, considering unpolarized neutrons, to

$$\mathrm{d}\Gamma \propto p_e E_e \left(E_{\beta^-} - E_e\right)^2 \hat{\xi} \left\{1 + b\frac{m_e}{E_e}\right\} \mathrm{d}E_e, \tag{2.10}$$

neglecting electromagnetic and radiative corrections. The Fierz interference term b is the focus of this thesis and will be further discussed in the following section.

2.3 Fierz Interference Term

The resulting energy spectrum in Equ. (2.10) contains only the phase space contribution of the so-called Fierz interference term b, which is a function of the Wilson coefficients C_i^+ from the pionless EFT Lagrangian in Equ. (2.5) [Fal+21]

$$\begin{aligned} \hat{\xi} &= \left|C_V^+\right|^2 + \left|C_S^+\right|^2 + 3\left[\left|C_A^+\right|^2 + \left|C_T^+\right|^2\right] \\ b\,\hat{\xi} &= \pm 2\,\mathrm{Re}\left[C_V^+ \bar{C}_S^+ + 3\,C_A^+ \bar{C}_T^+\right]. \end{aligned} \tag{2.11}$$

This expression simplifies to

$$b = 2\frac{C_S^+ + 3\lambda C_T^+}{1 + C_S^{+2} + 3\lambda^2 + 3C_T^{+2}} \simeq 2\frac{C_S^+ + 3\lambda C_T^+}{1 + 3\lambda^2} + O\left({C_{S/T}^+}^2\right). \tag{2.12}$$

By setting $C_T^+ = 0$, $C_S^+ = 0$, and $\lambda = C_A^+/C_V^+$, we can obtain the original Standard Model decay rate in Equ. (2.3) from Equ. (2.10) and (2.11).

As shown in Equ. (2.12), the Fierz interference term b is a function of BSM physics in the form of scalar and tensor contributions. Historically, the Fierz interference term refers to the interference when exchanging different spin bosons, i.e., virtual particles [Fie37; Fie39]. This interference modifies the overall coupling constant of the considered process and leads to small corrections in the predicted decay rates. For the pionless EFT framework we used, the interference is between different structures of nucleon-level currents instead of virtual particles. For visualization, Fig. 2.2 illustrates the effect of a non-zero Fierz interference term b on the electron beta spectrum.

We can therefore use the pionless EFT description of beta decay to quantify left-handed BSM physics by precisely measuring the differential decay rate.

2.3.1 Sensitivity of Different Decays

While Equ. (2.10) shows that we can use beta decay for studies of BSM physics, there are other beta decays besides that of the free neutron, that might be suitable to look for BSM physics.

The authors of [GN16] demonstrate that the beta decays of neutrons and ^{6}He nuclei are optimally sensitive probes to constrain the Fierz interference term b experimentally. Decays with endpoint energies in the 600 – 3800 keV range are within 20% [Δb] of the optimal statistical uncertainty, as illustrated in Fig. 2.3. The energy-dependent sensitivity arises from

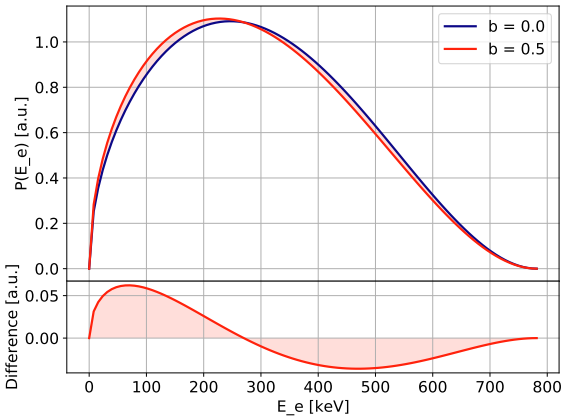


Figure 2.2 Effect of a non-zero Fierz interference term b on the electron beta spectrum. Both spectra are normalized. The value of $b = 0.5$ is exaggerated for visualization purposes.

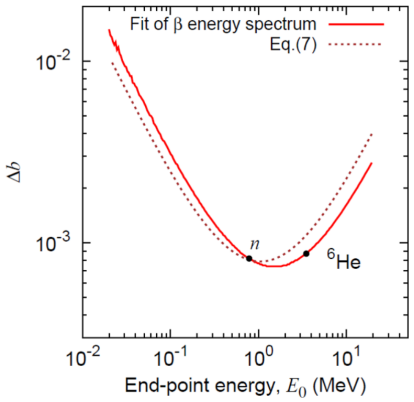


Figure 2.3 Graph and caption adapted from [GN16]: "The solid red line shows the 1σ statistical uncertainties obtained from fits of simulated beta energy spectra as a function of the endpoint energy E_0." The dotted line is neglected and the authors used 10^8 events for each fitted spectrum.

a balance between two competing factors. On the one hand, the magnitude of the m/E term in the decay probability in Equ. (2.10) grows with decreasing energy, enhancing sensitivity. On the other hand, the available phase space for BSM physics, i.e., the difference between Equ. (2.10) and (2.3), decreases with energy, reducing sensitivity. A reason to favor studying neutron decay over ^{6}He decay is that a measurement with free neutrons does not need to consider interference with nuclear wave functions of mother nuclei.

2.3.2 Related Work and Measurement Goal

Previous work already determined the Fierz interference term b in neutron beta decay. Currently, there is only one measurement directly from the electron energy spectrum in [Hic+17],

$$b = 0.067 \pm 0.005_{\text{ stat}} \, {}^{+0.090}_{-0.061}{}_{\text{ sys}} ,$$

and two from a combined analysis with the beta asymmetry A [Sun+20][1] and [Sau+20]

$$b = 0.066 \pm 0.041_{\text{stat}} \pm 0.024_{\text{sys}},$$
$$b = 0.017 \pm 0.020_{\text{stat}} \pm 0.003_{\text{sys}}.$$

The value from [Sau+20] is currently the most precise measurement with an overall result of $b = 0.017(21)$. Superallowed nuclear decays currently set the strongest limits for left-handed scalar interactions with [HT20] at 95% confidence level

$$-0.0020 < C_S^+/C_V^+ < 0.0020.$$

Using the scalar limit from [HT15] (predecessor result of [HT20]), the authors of [Mär+19] derived a limit on left-handed tensor interactions at 95% confidence level

$$-0.0044 < C_T^+/C_A^+ < 0.00023.$$

Compared to other decays, e.g., Li beta decay [Bur+22] at 95% confidence level

$$-0.087 < C_T^+/C_A^+ < 0.087,$$

neutron beta decay drives the limits on left-handed tensor interactions.

To further push the boundaries of BSM physics searches, we set out to measure the Fierz interference term with an absolute precision of $\sigma_b = 5 \cdot 10^{-3}$ from a direct electron energy spectrum measurement [Lam19], improving on the most precise result to date from [Sau+20] with $\sigma_b = 2.1 \cdot 10^{-2}$ by a factor of four. The resulting precision of [Sau+20] is limited by statistical uncertainties. Our proposed measurement [Lam19] is expected to be limited by systematical uncertainties, requiring improved methodologies in the analysis and a broader set of systematic effects to be studied. I was the main responsible for preparing and managing this measurement campaign. The analysis of that measurement campaign is the focus of my thesis, see Sec. 4.

The aspired precision of $\sigma_b = 5 \cdot 10^{-3}$ bridges the gap to or is competitive with proposed future measurements, e.g., [Sev14; Huy+16; Fry+19; Wan+19]. The PERKEO III spectrometer used, e.g., in [Sau+20], and for the measurement for this thesis, is followed-up by the PERC spectrometer [Dub+08], which is currently in development with my involvement [Wan+19].

[1] The authors of [Sun+20] also extract b directly from the electron energy spectrum, but focus on the beta asymmetry due to the resulting uncertainties.

3 Machine Learning for High-Precision Experiments

Particle physics experiments are often complex and intricate, with many tunable parameters. While deriving models of the underlying physical processes from first principles is not always possible, it is necessary to understand and model experiments to achieve precise results. Historically, this requirement led to the creation and utilization of Monte-Carlo-based simulations and advanced statistical tools. In this chapter, I intend to explore statistical machine learning and deep learning alternatives to enhance experimental work by integrating physical motivation and statistical goals where applicable. Specifically, I present applications in post-experimental data analysis, pre-experimental optimization of experiment design, and developing tools to gain insight into state-of-the-art large-scale simulations.

The presented work and some illustrations in this chapter is based on three scientific works with my contributions [LBM22; Lam+22a; Dor+23].

3.1 Methodology - Modeling

Without an analytical model based on first principles, we must rely on empirical data from experiments or simulations to construct a surrogate model to describe physical processes. Machine learning techniques can be employed to extract relationships from such data. These methods help analyze complex systems where theoretical models are intractable or nonexistent.

Two broad classes of machine learning methods are particularly relevant for surrogate modeling: regression and dimensionality reduction. Regression techniques like Gaussian process regression and neural networks can learn a functional mapping between inputs and outputs. They provide a surrogate model for emulating the system response. Dimensionality reduction methods, such as principal component analysis and autoencoders, can extract salient features from the data by reducing input redundancies.

We can use such data-driven approaches to improve data analysis, simulation efficiency, or experiment design. This section summarizes existing literature and recent scientific work and also highlights applicability to physics experiments.

3.1.1 Kernel Functions

Kernel functions k are defined as integral operators for $\boldsymbol{x} \in \chi$ on f with Lebesgue measure ν as

$$(T_k f)(\boldsymbol{x}) = \int_{\chi} k(\boldsymbol{x}, \boldsymbol{x}') \, \mathrm{d}\nu(\boldsymbol{x}').$$

Should k fulfill the Mercer condition [SS18], calculating $k(\mathbf{x}, \mathbf{x}')$ is equivalent to mapping $\mathbf{x}$ and $\mathbf{x}'$ to a feature space via a function ϕ and taking the inner product [SS18]

$$k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\nu}. \tag{3.1}$$

However, neither the projection nor the inner product needs to be calculated directly. This equivalence is called the *kernel trick*. Since the equivalent mapping of ϕ is primarily non-linear and usually into higher dimensions, we can use kernel functions to make data more separable for a more straightforward classification or analysis, e.g., with linear methods. Furthermore, we can construct the Gram matrix $\mathbf{K}$ for $K_{ij} = k(\mathbf{x_i}, \mathbf{x_j}')$ with respect to a data set $\mathcal{D}_n = \{\mathbf{x_1}, ..., \mathbf{x_n}\}$ of size n. Should $\mathbf{K}$ be positive semi-definite, then k is a valid covariance function and $\mathbf{K}$ the corresponding covariance matrix. We can interpret $\mathbf{K}$ as a similarity matrix because we score each point $\mathbf{x_i}$ to each point in $\mathcal{D}_n$ by a similarity criterion defined by the kernel function k.

Kernel functions can be, e.g., stationary ($k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$) and therefore invariant under spatial displacements, or isotropic ($k(\mathbf{x}, \mathbf{x}') = k(|\mathbf{x} - \mathbf{x}'|)$). In addition, we can create composite kernels, as a sum, product, or convolution of two kernel functions [SS18]. Given that k scores points by a similarity criterion, we can interpret composite kernels as combining similarity criteria with logical operators. As the kernel function parameters are related to data-specific characteristics such as length scales, kernel methods offer interpretable parameters. This benefit is unique compared to other machine learning methods and makes them attractive for physics applications [LBM22]. Thus, we can create physically motivated kernel functions with physical parameters and encode physical information like, e.g., symmetry ($k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') + k(-\mathbf{x}, \mathbf{x}')$) or local density changes of our problem. I want to highlight that the kernel function parameters, even if interpretable, still need to be set.

The computational complexity of $\mathbf{K}$ scales with $O(n^2)$. To use kernel functions with large data sets, we can use a low-rank matrix approximation called Nyström approximation [DM05; RW06]. For this, we chose a random subset $\mathcal{D}_m$ of m data points with $m < n$ to represent the data set and calculate the Gram matrix for $K_{ij} = k(\mathbf{x_i}, \mathbf{x_j})$ with $\mathbf{x_i} \in \mathcal{D}_n = \{\mathbf{x_1}, ..., \mathbf{x_n}\}$ and $\mathbf{x_j} \in \mathcal{D}_m \subset \mathcal{D}_n$.

To visualize kernel functions, we can consider a fictional problem: We have a measurement of K^+ decay and the experiment collected the decay particle charge Q and rest mass m in natural units with uncorrelated uncertainties. Our goal in this simple example is to separate the data set to only contain π^0. The dominant, non-leptonic decay channels are

$$\begin{aligned} K^+ &\to \pi^+\pi^0 \\ K^+ &\to \pi^+\pi^+\pi^- \\ K^+ &\to \pi^+\pi^0\pi^0. \end{aligned}$$

The mass resolution is sufficient to remove leptonic decay, but it is insufficient to separate the mass difference between $\pi^\pm$ and π^0 of $m_{\pi^\pm} - m_{\pi^0} \approx 4.6$ MeV. The measured data set is shown in Fig. 3.1. The branching ratios for different decays have been taken into account. Using the symmetric polynomial kernel, we can illustrate how the data set with $\mathbf{x} \in \mathcal{R}^2$ becomes linearly separable after applying the equivalent mapping function ϕ into a higher-dimensional space as

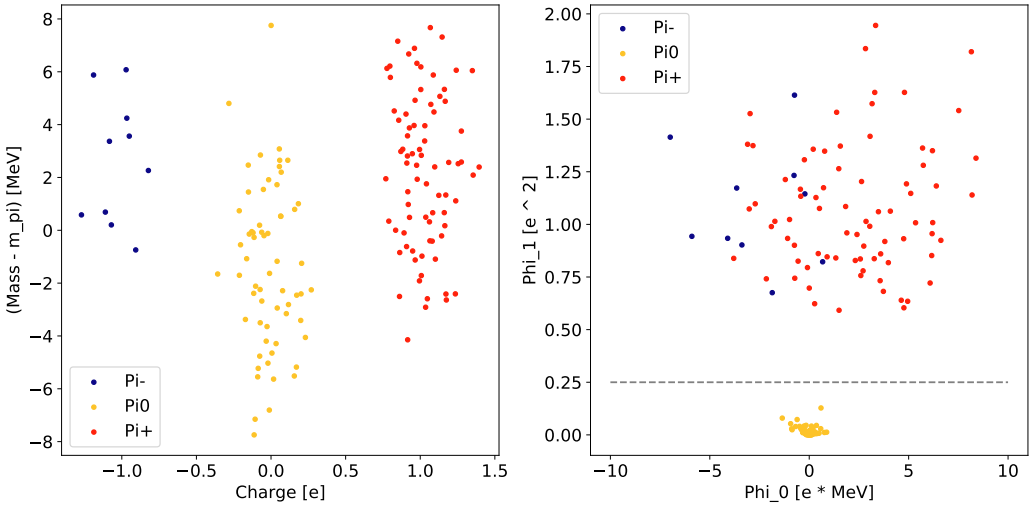


Figure 3.1 Illustrating kernel functions with the separation of $\pi^{\pm 0}$ produced in K^+ decay. (Left) Original data set of dominant non-leptonic K^+ decay. (Right) First two dimensions after mapping the data set with ϕ, see Equ. (3.2).

$$\text{for} \quad k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}')^d \qquad \xrightarrow{d=2} \qquad \phi(\mathbf{x}, \mathbf{x}') = \begin{pmatrix} \sqrt{2} x_0 x_1 \\ x_0^2 \\ x_1^2 \end{pmatrix} \equiv \begin{pmatrix} \sqrt{2} Q m \\ Q^2 \\ m^2 \end{pmatrix}. \tag{3.2}$$

The result is shown in Fig. 3.1. Even when only plotting the first two of the three dimensions in the feature space, we can separate the π^0 data points with a straight line in this constructed example.

Besides the polynomial kernel in Equ. (3.2), other common kernel choices are the *Matérn*$_\nu$ kernel class with $d = ||\mathbf{x} - \mathbf{x}'||^2$ [RW06; Sha+16]

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}')_{\nu=\frac{3}{2}} &= \sigma^2 \left(1 + \frac{\sqrt{3} d}{\rho}\right) \exp\left(-\frac{\sqrt{3} d}{\rho}\right) \\ k(\mathbf{x}, \mathbf{x}')_{\nu=\frac{5}{2}} &= \sigma^2 \left(1 + \frac{\sqrt{5} d}{\rho} + \frac{5 d^2}{2\rho^2}\right) \exp\left(-\frac{\sqrt{5} d}{\rho}\right), \end{aligned} \tag{3.3}$$

and the radial basis function kernel (RBF)

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{d}{2\rho^2}\right). \tag{3.4}$$

Both have a length scale parameter ρ, an output scale parameter σ and are isotropic kernel functions. The RBF kernel highlights the elegance of the kernel trick, as it's mapping function ϕ maps into an infinite-dimensional feature space as shown with the multinomial theorem.

Any algorithms relying on kernel functions to map data into high-dimensional feature spaces are called *kernel methods*. For example, we can combine kernel functions with data dimensionality reduction methods, e.g., principal component analysis (PCA). PCA simplifies complex data sets by transforming the original variables into new variables called

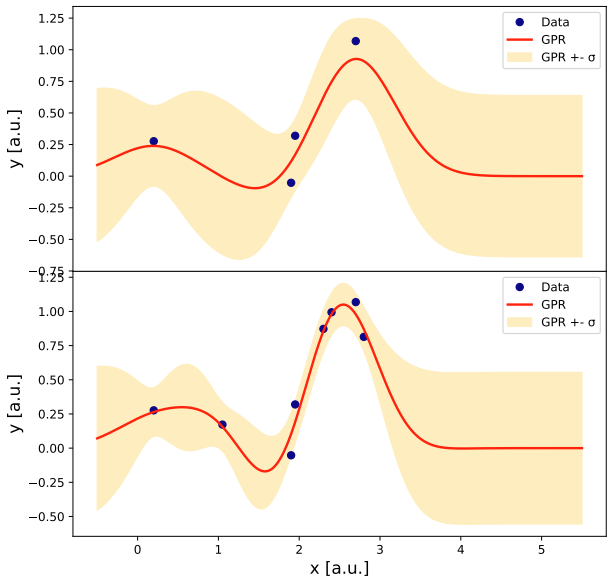


Figure 3.2 Caption and graph adapted from [LBM22]: "Example: Gaussian process posterior distribution for two different data sets. One with 4 (top) and updated for 8 (bottom) noisy points of the same unknown underlying distribution."

principal components. The principal components are orthogonal to each other and explain the maximum possible variance of the original data. By reducing the dimensionality of the data, principal component analysis can reveal hidden trends and patterns in the data. However, PCA is a linear technique, and if the data has significant non-linear effects, PCA will miss those relations. As presented in [SSM98], we can expand PCA to a non-linear component analysis by constructing $\mathbf{K}$ of our data set and performing PCA on the result, so-called *kernel PCA*. Another example for kernel methods are support vector machines and Gaussian processes, of which we will explore the latter in the next section.

3.1.2 Gaussian Processes

We can model the predictions and local uncertainties of a continuous input space by treating each point as a component of an infinite-dimensional Gaussian distribution. A sample from this Gaussian distribution is a function with a value for each point in the input space, defining a Gaussian process with function samples $\mathbf{f}$ as

$$\mathbf{f} \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) .$$

The mean and fluctuation of the sample function $\mathbf{f}$ state the predictive mean and uncertainty at a point $\mathbf{x}$, i.e., each point $\mathbf{x}$ in the input space has a mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$ to other points $\mathbf{x}'$.

We want to construct such a model using a finite data set of observations. Using Bayesian inference, we can define a prior over functions with a zero mean $m(\mathbf{x}) = 0$, covariance matrix

$\mathbf{K}$ with entries $K_{ij} = k(\mathbf{x_i}, \mathbf{x'_j})$, an observed data set of size n with input points $X \in \mathcal{R}^m$ and values $y \in \mathcal{R}$

$$\mathbf{f} \sim \mathcal{N}(\mathbf{0},\ \mathbf{K}(X, X)) .$$

Positive semi-definite kernel functions are used as covariance functions, see Sec. 3.1.1. We can expand this for observations with fixed noise $y = f(\mathbf{x}) + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. Expanding the following expressions for individual noise for each point is straightforward. In addition, we add testing points X_* in the input space where we want to predict function values $\mathbf{f}_*$ to get the joint prior over functions

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(X, X) + \sigma_n^2 \mathbf{I} & \mathbf{K}(X, X_*) \\ \mathbf{K}(X_*, X) & \mathbf{K}(X_*, X_*) \end{bmatrix}\right). \tag{3.5}$$

As we are using only Gaussian distributions, we can obtain the predictive distribution through conditioning

$$\mathbf{f}_* \mid X, \mathbf{y}, X_* \sim \mathcal{N}\left(\bar{\mathbf{f}}_*, \operatorname{cov}(\mathbf{f}_*)\right), \tag{3.6}$$

with predicted mean $\bar{\mathbf{f}}_*$ and covariance $\operatorname{cov}(\mathbf{f}_*)$

$$\begin{aligned} \bar{\mathbf{f}}_* &= \mathbf{K}(X_*, X)\left[\mathbf{K}(X, X) + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{y} \\ \operatorname{cov}(\mathbf{f}_*) &= \mathbf{K}(X_*, X_*) - \mathbf{K}(X_*, X)\left[\mathbf{K}(X, X) + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{K}(X, X_*) . \end{aligned} \tag{3.7}$$

Thus, Gaussian processes reconstruct the underlying signal without contaminating noise by computing a weighted average of noisy observations $\mathbf{y}$, making them equivalent to linear smoother by definition. Furthermore, they are a generalization of spline models, large neural networks, and support vector machines [Nea96; Wah78; RW06]. We can apply them to regression (GPR) or classification tasks.

The only parameters of a Gaussian process like described above are the noise parameter σ_n and any parameters of the covariance function k. Gaussian processes are non-parametric, as they are created from data sets and grow in complexity with them.[1] We can set these parameters by maximizing the marginal log-likelihood $\log p(\mathbf{y} \mid X)$ obtained through the Gaussian likelihood and prior as

$$\log p(\mathbf{y} \mid X) = -\frac{1}{2}\mathbf{y}^\top \left(\mathbf{K}(X, X) + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{y} - \frac{1}{2}\log\left|\mathbf{K}(X, X) + \sigma_n^2 \mathbf{I}\right| - \frac{n}{2}\log 2\pi. \tag{3.8}$$

The noise and kernel function parameters are optimized during the fit, implying statistically motivated results, unlike some other machine learning approaches. To illustrate the non-parametric nature, we give an example GPR for a one-dimensional input space for a data set with four and eight data points from the same data set in Fig. 3.2. Note that the predicted mean and covariance converge to the prior again away from the data. The predictive range depends on the length scale parameter of the kernel function, see Equ. (3.4).

The covariance function is the most critical model assumption for Gaussian processes. Its differentiability affects the smoothness of the samples drawn from the corresponding prior in Equ. (3.5). As natural phenomena usually have underlying non-continuous contributions, smooth samples struggle to model such phenomena, leading to a length scale collapse when

[1] Histograms are another example of a non-parametric model that grows in complexity with the data set to approximate probability density functions for a given set of hyperparameters, e.g., number of bins.

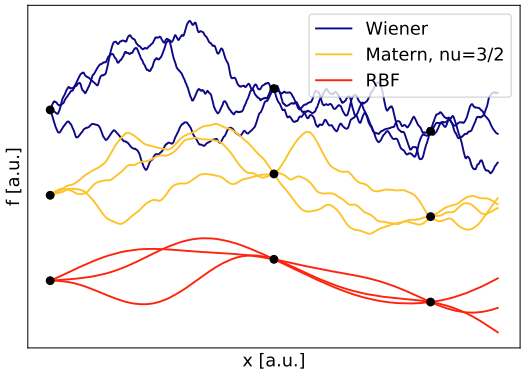


Figure 3.3 Illustrating the impact of mean-square differentiability of different kernel functions on sample smoothness by drawing samples from the joint prior in Equ. (3.5). The results are shifted on the y-axis for visualization purposes.

maximizing the marginal log-likelihood. This collapse, in turn, yields imprecise posterior distributions. To illustrate the impact of the mean-square differentiability of the covariance function, we can compare samples drawn from the prior for three kernel functions with the same length scale in Fig. 3.3. The RBF kernel function in Equ. (3.4) is infinitely mean-square differentiable, whereas the Matérn kernel functions are only $\lceil \nu \rceil - 1$ times [SS18] mean-square differentiable, and the Wiener kernel ($k(\mathbf{x}, \mathbf{x}') = \min(\mathbf{x}, \mathbf{x}')$) is not mean-square differentiable.

Besides prior knowledge about smoothness properties encoded in the covariance function, we also make other assumptions using Gaussian processes. Gaussian processes assume that similar inputs will have similar outputs, meaning the modeled function is smooth and continuous. For stationary covariance functions, the covariance function used in the Gaussian process depends only on the distance between inputs, not on the absolute values of the inputs, meaning the modeled function has the same smoothness properties everywhere. The Gaussianity of the underlying noise in the observations and latent function values is a requirement for the analytical tractability of the posterior predictive distribution. I discuss non-stationarity and adaptable metrics in covariance functions, data set size limitations, and non-Gaussian likelihoods in Sec. 3.1.3. In this section, we focus on scalar outputs, however, multi-output Gaussian processes are possible [AL08; Wil+16b], also see Sec. 3.1.3.

In conclusion, Gaussian processes allow us to incorporate prior knowledge about the function we want to model while offering a non-parametric but statistically motivated approach to model data while also offering uncertainty quantification.

3.1.3 Stochastic Variational Gaussian Processes and Deep Kernel Learning

The most significant limitation of Gaussian processes for (particle) physics applications is the insufficient scalability to large data sets. Due to the matrix inversion of $\mathbf{K}$ in the marginal log-likelihood in Equ. (3.8), Gaussian processes have a computational complexity of $O(n^3)$ for a dataset of size n. An approach to increasing Gaussian processes' scalability is a stochastic

variational Gaussian process (SVGP), as outlined by [HFL13; HMG14; Hen+15].[2] They combine sparse Gaussian processes [QR05] with posterior approximation via variational inference [BKM17]. Their approach enables stochastic gradient descent optimization on a lower bound of the marginal log-likelihood from Equ. (3.8) and non-Gaussian likelihoods.

The goal of this section is to outline how stochastic variational Gaussian processes (SVGP) are derived differently from *exact* Gaussian processes, as described in Equ. (3.7) and (3.8), and to scale Gaussian processes to large data sets. The authors of [QR05] use a sparse prior with m inducing points X_S as additional parameters to the model as

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_\mathrm{S} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(X, X) & \mathbf{K}(X, X_\mathrm{S}) \\ \mathbf{K}(X_\mathrm{S}, X) & \mathbf{K}(X_\mathrm{S}, X_\mathrm{S}) \end{bmatrix}\right). \tag{3.9}$$

The idea is to have the inducing points and their predictions $(X_\mathrm{S},\ \mathbf{f}_\mathrm{S}(X_\mathrm{S}))$ summarize the data set and add the inducing points X_S, but not their predictions $\mathbf{f}_\mathrm{S}$, as parameters to the model. In the exact Gaussian process joint prior over functions in Equ. (3.5), we want to use the information of $\mathbf{y}(X)$ to infer values $\mathbf{f}_*(X_*)$ during inference. With the sparse prior, we want to use the information of $\mathbf{f}_\mathrm{S}(X_\mathrm{S})$ to infer values $\mathbf{f}(X)$ during maximization of the likelihood. However, using the sparse prior from Equ. (3.9) results in a marginal log-likelihood $\log p(\mathbf{y} \mid X)$ still containing $\mathbf{K}(X, X)$ terms and also losing terms containing the added inducing points X_S. Therefore, we need another objective value besides the marginal log-likelihood to fit our model with all its parameters.

To this end, the authors of [HFL13; HMG14] use variational inference [BKM17] to solve this problem. The idea of variational inference is to approximate the posterior distribution with a variational distribution as

$$q\left(\mathbf{f}, \mathbf{f}_\mathrm{S}\right) \approx p\left(\mathbf{f}, \mathbf{f}_\mathrm{S} \mid \mathbf{y}\right).$$

We can choose the approximation as factorization

$$q\left(\mathbf{f}, \mathbf{f}_\mathrm{S}\right) = q\left(\mathbf{f} \mid \mathbf{f}_\mathrm{S}\right) q\left(\mathbf{f}_\mathrm{S}\right),$$

with marginal variational distribution $q\left(\mathbf{f}_\mathrm{S}\right)$ as Gaussian distribution with variational parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

$$q\left(\mathbf{f}_\mathrm{S}\right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The model has inducing points and variational parameters as additional model parameters. The number of inducing points is fixed and set manually. For variational inference, we can use a lower bound on the marginal log-likelihood as the optimization objective, the *evidence lower bound* or ELBO

$$\log p(\mathbf{y} \mid X) \geq \mathbb{E}_{\mathbf{f}, \mathbf{f}_s \sim q(\mathbf{f}, \mathbf{f}_s)}\left[\log\left(p(\mathbf{y} \mid \mathbf{f}) \frac{p\left(\mathbf{f}, \mathbf{f}_\mathrm{S}\right)}{q\left(\mathbf{f}, \mathbf{f}_\mathrm{S}\right)}\right)\right].$$

Which we can rewrite as

$$\mathrm{ELBO} = \sum_{i=1}^{n} \int \log\left(p\left(y_i \mid f_i\right)\right) q\left(f_i\right) \mathrm{d}f_i - \mathrm{KL}\left(q\left(\mathbf{f}_\mathrm{S}\right) \| p\left(\mathbf{f}_\mathrm{S}\right)\right), \tag{3.10}$$

[2] Generally, using variational inference with Gaussian processes is called *variational Gaussian processes* [TRB15], and this approach enables non-Gaussian likelihoods for many Gaussian process expansions, e.g., Deep Gaussian processes [SD17].

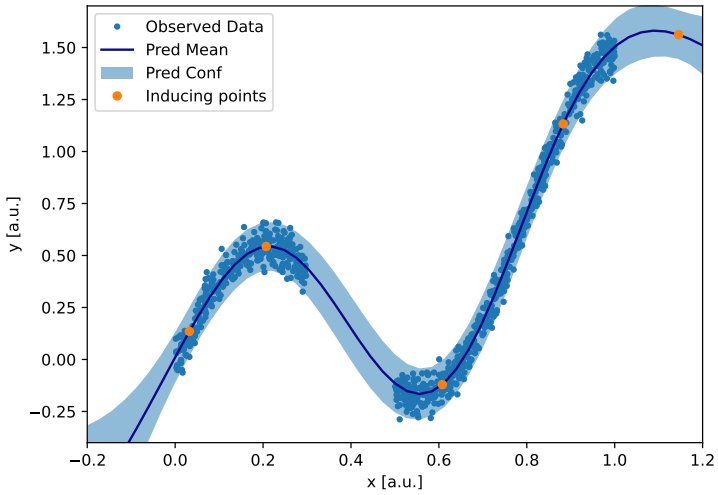


Figure 3.4 SVGP example: Predictive distribution with 95% confidence region from Equ. (3.11) after optimizing the *ELBO* from Equ. (3.10). The five inducing points X_S and their predictions $\mathbf{f}_S(X_S)$ summarize the data set. We plot only 10% of the observed data set for illustration.

with the Kullback–Leibler divergence or *relative entropy* KL [KL51]. When using the ELBO as the optimization objective with gradient methods, we can approximate the gradient of the sum in the first term in Equ. (3.10) with a random subset of the data set. This approximation is called stochastic gradient descent (SGD) and avoids using the entire data set, see Sec. 3.2.2 or [KW19]. The second term is independent of the data set. Maximizing the ELBO improves how well the variational distribution approximates the posterior and how well our model describes the original data set. In analogy to the exact Gaussian process in Equ. (3.7), the predictive distribution for the SVGP becomes

$$\mathbf{f}_* \mid X, \mathbf{y}, X_*, X_S \sim p\left(\mathbf{f}_* \mid \mathbf{f}_S\right) = \mathcal{N}\left(\bar{\mathbf{f}}_*, \operatorname{cov}\left(\mathbf{f}_*\right)\right),$$

with new predicted mean $\bar{\mathbf{f}}_*$ and covariance $\operatorname{cov}\left(\mathbf{f}_*\right)$ at testing points X_* and shorthand writing $\mathbf{K}\left(X_*, X_S\right) = \mathbf{K}_{*S}$ etc.

$$\begin{aligned} \bar{\mathbf{f}}_* &= \mathbf{K}_{*S}\mathbf{K}_{SS}^{-1}\boldsymbol{\mu} \\ \operatorname{cov}\left(\mathbf{f}_*\right) &= \left(\mathbf{K}_{*S}\mathbf{K}_{SS}^{-1}\right)\boldsymbol{\Sigma}\left(\mathbf{K}_{*S}\mathbf{K}_{SS}^{-1}\right)^{\intercal} + \mathbf{K}_{**} - \mathbf{K}_{*S}\mathbf{K}_{SS}^{-1}\mathbf{K}_{*S}^{\intercal}. \end{aligned} \tag{3.11}$$

The predictive mean and covariance do not require the original data set for calculation anymore. SVGPs achieve a computational complexity of $O(m^3)$. We illustrate SVGP on a simple, one-dimensional regression problem with a data set of $n = O(10^4)$ points using five inducing points X_S in Fig. 3.4. We generate the data set in Fig. 3.4 with an arbitrary, underlying function with observational noise. As can be seen, the inducing points successfully summarize the large data set. The SVGP model offers all advantages of Gaussian processes but scales to millions of data points, enabling regression and classification applications in particle physics. SVGPs are limited to problems where a smaller number of inducing points suffice to represent the data set's details and where regular kernels are expressive enough.

To overcome the limited flexibility of kernel functions, the authors of [Wil+16a] introduced scalable deep kernel learning (DKL) to utilize the adaptive basis functions of a neural

network (NN). To achieve computational complexity of $O(n)$ for n training points during inference, they use local kernel interpolation, inducing points, and structure exploiting algebra introduced in [WN15b]. The authors of [Wil+16a] use the output features of a deep kernel NN as input for a base kernel, e.g. RBF in Equ. (3.4), of a Gaussian process and jointly train the parameters of the deep kernel NN and the base kernel by maximizing the marginal log-likelihood. This approach combines, e.g., the infinite set of fixed basis functions of an RBF kernel with a set length scale with a fixed set of highly adaptive basis functions of NNs. Thus, the deep kernel NN learns statistical representations from the data set and enables non-euclidean similarity metrics throughout the input space.

The DKL approach was expanded by [Wil+16b], introducing stochastic variational deep kernel learning (SV-DKL), effectively combining the benefits of DKL and SVGP. It allows for stochastic gradient training enabled by variational inference [BKM17] on the ELBO as for SVGP. For correlated multi-output predictions, the authors feed the outputs of the deep kernel NN into a set of additive base kernels, which they use for independent Gaussian processes. Furthermore, they mix these Gaussian processes linearly to induce correlations and feed the outputs into a softmax likelihood for classification. The SV-DKL model achieves a computational complexity of $O(m^{1+1/d})$ for fixed inducing points m and input dimensions of our data set d.

SV-DKLs are ideal for physics applications as they can handle large-scale physics datasets while enabling uncertainty quantification and non-Euclidean metrics often required in physics, e.g., to model local density changes. We can use them while retaining the advantages of Gaussian process models, like their statistical motivation. They offer a principled way of creating surrogate models from empirical data for further analysis by learning mappings between inputs and outputs. Also, they can be used to get an integrable model that accurately describes the underlying phenomena with the mentioned benefits.

3.2 Methodology - Optimization

To maintain continues improvements in experimental results, physicists keep innovating and optimizing their instruments, simulations, and data analysis. This development continuously increases the complexity and makes optimization more difficult. To tackle this issue, we will explore two optimization approaches based on machine learning methodology that are promising for physics applications.

Generally, optimization is finding the minimum or maximum of an objective function $\mathcal{L}$ that maps a set of tunable parameters $\mathbf{x}$ from the space of possible parameters χ to an objective value, e.g., maximizing the energy resolution of a detector, and the corresponding optimal parameters $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}} = \arg\min_{x \in \chi} \pm \mathcal{L}(\mathbf{x}), \tag{3.12}$$

where $\mathcal{L}$ incorporates the underlying physical properties of a system leading to the objective value, e.g., particle tracks and energy losses. $\mathcal{L}$ can be statistically motivated like a likelihood or another metric. This problem raises two questions: How to create $\mathcal{L}$ and how to minimize it to obtain $\hat{\mathbf{x}}$?

I want to focus on two specific cases and discuss solutions in this chapter.[3] In the first case, we cannot derive $\mathcal{L}$ from first principles, have little data, and creating more data is

[3] I recommend [KW19] for more optimization methodology.

expensive. This issue requires a sample-efficient, sequential approach that guides data-taking while optimizing. An example could be finding the ideal experiment parameters from simulations that take hours to test one set of parameters with enough precision. In the second case, we can create $\mathcal{L}$ via a surrogate model (or even from first principles), but the optimization is non-convex, and we want to optimize it with a gradient-based deep learning methodology. An example would be having a range of simulations of a detector for many different parameters, but finding the optimum is difficult due to the number of parameters and their correlation.

3.2.1 Bayesian Optimization

If we cannot derive $\mathcal{L}$ in Equ. (3.12) from first principles, we must construct a model based on the available data, e.g., previous detector designs or simulations, to approximate the function, e.g., using the methods in Sec. 3.1. However, if we have little or no data and data generation, i.e., testing a set of parameters, is very time-consuming or expensive, we must include the data acquisition for modeling $\mathcal{L}$ into the optimization, by updating $\mathcal{L}$ in a sample-efficient, sequential approach that guides data-taking while optimizing.

Specifically, Gaussian processes can be employed to create an approximate $\mathcal{L}$ from a limited number of noisy observations iteratively for each new observation, see Sec. 3.1.2 and 3.1.3. One key difference between Gaussian processes and other modeling methods, such as neural networks or splines, is the provided statistical uncertainty for predictions, see Equ. (3.7) and (3.11). We can leverage this aspect for statistically guided optimization based on criteria such as information gain or expected improvement. Selecting the next test point $n+1$ by leveraging the uncertainty of the current model for a data set $\mathcal{D}_n = \{(\mathbf{x}_l, \mathcal{L}_l) \mid l = 1, ..., n\}$ of n tested points sequentially and updating the model is called *Bayesian optimization* [Sha+16]. Such guided optimizations are essential when the goal is to use the fewest samples possible to find optimal parameters.

To determine the next set of untested parameters $\mathbf{x}$ to be sampled, we can formulate an *acquisition function* α that scores candidate data points' utility for the next evaluation, guiding the sequential search. There are two approaches to be balanced when guiding the sequential search: Exploiting the already available data and risk converging locally or exploring regions in the input space with little data, risking inefficiency. Acquisition functions must balance exploration of the input space versus exploitation of the current best result to achieve overall sample efficiency and reach the global optimum. Two commonly acquisition functions are expected improvement (EI) and lower confidence bounds (LCB)[Sha+16]. For Gaussian process posterior distribution with standard deviation σ_n and mean prediction μ_n for a given dataset $\mathcal{D}_n$, both acquisition functions are tractable. The LCB is

$$\alpha_{\mathrm{LCB}}(x; \mathcal{D}_n) = \mu_n(x) - \kappa \cdot \sigma_n(x) \tag{3.13}$$

with the exploration parameter κ. For Gaussian process posterior distribution f with normal distribution $\mathcal{N}$ and its cumulative distribution function Ψ, EI is

$$\begin{aligned} \alpha_{\mathrm{EI,\,GP}}(x; \mathcal{D}_n) &= \mathbb{E} \max\left(f(x) - f(x_n^+), 0\right) \\ &= \sigma_n(x) \cdot \gamma(x) \cdot \Psi\left(\gamma(x)\right) + \mathcal{N}\left(\gamma(x)\right) \end{aligned} \tag{3.14}$$

with $\gamma(x) = \left(f(x) - f(x_n^+) - \kappa\right) / \sigma_n(x)$.[4] Finally, we determine the best next candidate to be sampled x_{n+1} with a gradient-based approach on the acquisition function. As we are

[4] Other, less common acquisition functions are presented in [Sha+16; Ber+18], like entropy-based acquisition functions.

using gradient-based approaches on the Gaussian process surrogate of $\mathcal{L}$, these evaluations circumvent the original problem of expensive data generation.

To better visualize the sequential optimization and how the acquisition function guides it, we can look at a one-dimensional example where we optimize one parameter x in Fig. 3.5. For seven tested $(x,\ \mathcal{L})$ pairs, we can construct a Gaussian process model (Matérn kernel function ($\nu = 5/2$)) and use LCB as the acquisition function ($\kappa = 2$) in Fig. 3.5a. Using an L-BFGS algorithm [KW19], we can find the minimum of the acquisition function to determine the next parameter value to be tested. We test this parameter and repeat the pattern. After training the Gaussian process model on the additional data point in Fig. 3.5b, the uncertainty estimation grows in the previously unexplored region and the acquisition function guides the optimization in that region, ultimately finding the global optimum in following steps.

Bayesian optimization with Gaussian processes works without knowing the to-be-optimized black-box function analytically, removing the expensive amount of function evaluations that gradient-based or sampling-based approaches require. It is successful in correlated and higher-dimensional problems, as the underlying kernel function is well-suited for that. This method found applications in hyperparameter tuning [Ber+11], outperforming other methods, and has also found physics applications [Dur+20; Han+21; Jal+21; Lec18; Kis19; IWY17; Cis+20; Eks+19]. Furthermore, the authors of [RHE21] recently demonstrated how the approach could be expanded to multi-objective Bayesian optimization (MOBO) [PZŽ16] to optimally balance the trade-offs between multiple competing objectives for physics applications simultaneously.

3.2.2 Differentiable Programming

If we can construct a surrogate model of $\mathcal{L}$ in Equ. (3.12) or derive one from first principles, we can use two gradient-based deep learning optimization approaches to solve difficult optimizations: Stochastic gradient descent (SGD) and automatic differentiation.

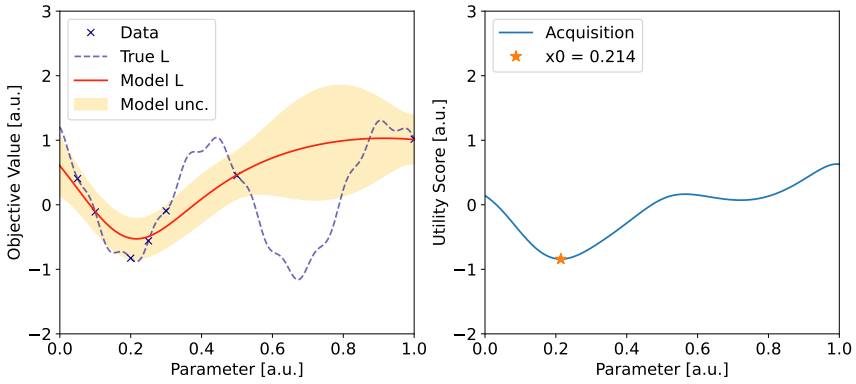
So far, we have only considered $\mathcal{L}(\mathbf{x})$ as a function of the experiment parameters $\mathbf{x}$, e.g., detector geometries, detector gain, or acting magnetic fields. More accurately, the objective value is also calculated for a training data set, e.g., a set of generated events of particles with distributed properties, and then averaged.[5] The objective value for one data point is $\mathcal{L}_i$. SGD is an optimization technique where we estimate the gradient of the objective function from subsets of the training data iteratively instead of calculating the exact gradient from the entire training set. When optimizing the parameters $\mathbf{x}$ of $\mathcal{L}(\mathbf{x})$, we can iteratively update the values of $\mathbf{x}$ at the k-th optimization step as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \frac{\alpha_k}{m} \sum_{i}^{m} \nabla_x \mathcal{L}_i(\mathbf{x}^{(k)}), \tag{3.15}$$

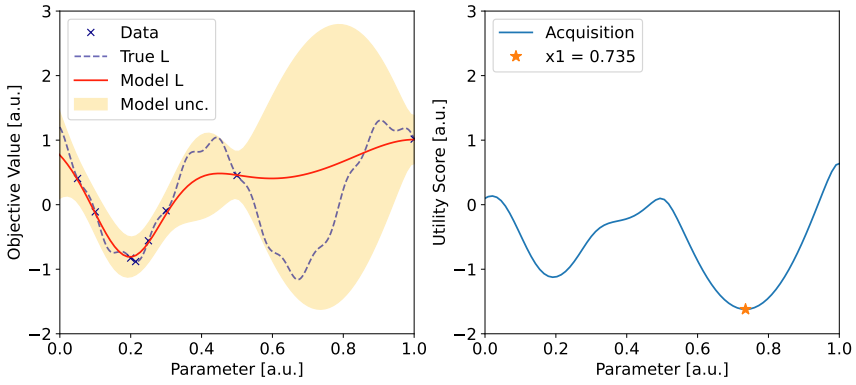
where m is the size of a random subset of the training data set of size n and *learning rate* α_k at step k. Generally, α_k is a hyperparameter that needs to be set manually or as a function of k. This basic formula can be adapted for better and faster convergence, e.g., by incorporating previous updates and their derivatives [KB14].

The stochastic gradients provide a noisy estimate of the true gradient, which can help the optimization escape local minima and plateaus. Also, this approximation enables us to use

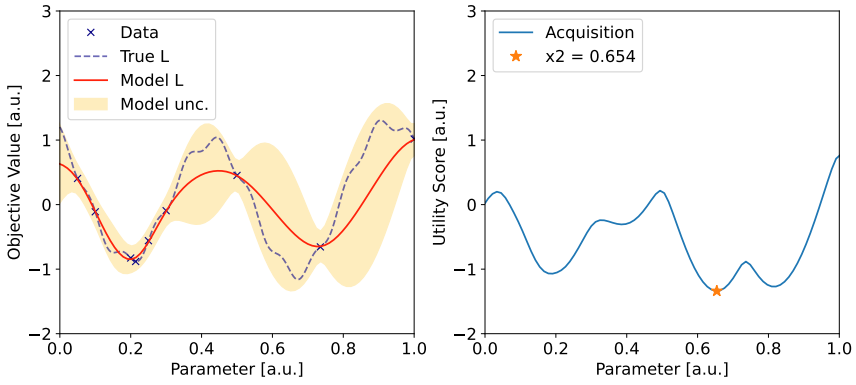
[5] The training data set can also be the data set used to create a surrogate model of $\mathcal{L}$ to incorporate the underlying physics.



(a) Step n



(b) Step $n + 1$



(c) Step $n + 2$

Figure 3.5 Bayesian optimization example for three consecutive optimization steps. The true objective function is unknown, but plotted for visualization. (Left) Gaussian process model (2σ uncertainty) trained on n data points at step n. (Right) Acquisition function uses the Gaussian process to chose the $n + 1$-th data point to be tested next.

large data sets, see Sec. 3.1.3 for an example, and adapt the optimization to changing or growing data sets. The latter is most relevant for particle physics applications, as we can only understand most state-of-the-art experiment setups with Monte-Carlo-based simulations. This requirement implies that we must rely on large Monte-Carlo-based data sets to construct $\mathcal{L}$ and optimize the experiment parameters.[6] Both benefits are also essential when optimizing high-dimensional and non-convex problems, as is the case for deep learning or when optimizing a complex detector design [Dor+23]. This similarity also implies that we can leverage the benefits of overparameterized models and their objective value landscape to find optima [MBB18; Bel+19; LZB22].

SGD requires the gradient of the objective function, and there are different ways to compute it. For large overparameterized models, automatic differentiation beats other approaches, such as numerical or symbolic differentiation, in efficiency and accuracy [Bay+17]. The idea is to automatically compute the partial derivatives of an implemented function sequentially by applying the chain rule of calculus to the program's elementary operations (addition, subtraction, multiplication, division), see [Bay+17; KW19] for details. It works for any function that can be expressed as a computer program. Combining automatic differentiation with gradient-based optimization makes up differentiable programming - a core ingredient of modern machine learning and deep learning approaches with valuable applications for physics [Dor+23].

We can leverage the framework of differential programming to optimize particle physics experiments. Differential programming works for implemented models of $\mathcal{L}$ from first principles or with surrogate models. Deep learning surrogate models are immediately usable with differentiable programming as they are trained with it. If there is a model from first principles, a surrogate model might be worth the trade-off in accuracy for faster computation of its values and derivatives [Kas+20]. Also, we can model individual parts of an experiment and combine the models to create an end-to-end pipeline when many simulation parameters are present, and a complete mapping of the parameter space would be too costly.

3.3 Application

We utilize the methodology from Sec. 3.1 and 3.2 for different physics applications, ranging from particle physics to computational astrophysics and cosmology. All presented example applications are based on the papers with my contributions [LBM22; Lam+22a; Dor+23] and partially reprinted where indicated.

3.3.1 Temperature Induced Gain Fluctuation Correction

As a specific example, we will optimize the analysis of an electron energy detector by improving its calibration by modeling an underlying systematic effect and their uncertainty with Gaussian processes. The detector and measurement setup is the same as in Sec. 4.2.1 and the result is used in Sec. 5.2.4. This section is a partial reprint of [LBM22].

Consider a simplified electron energy detector for an incoming electron with kinetic energy E_e. The detector is made up of a square scintillator, light guides at the sides, and a

[6] The main alternatives for sufficiently small data sets are quasi-Newton approaches, e.g., BFGS and L-BFGS [KW19].

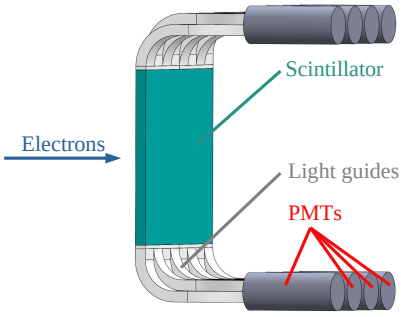


Figure 3.6 Caption and figure taken from [LBM22]: "Electron energy detector schematic. The spatial response measurement grid points are in front of the detector and parallel to the scintillator surface."

photomultiplier tube (PMT) at the end of each light guide, as shown in Fig. 3.6. When an electron hits the scintillator, it deposits its energy, and n_γ photons are produced proportional to the electron energy E_e. Photons travel through the detector to the PMTs and are converted to charge pulses with total charge A. The measured charge A is on an arbitrary scale, and we assume a linear relation to the original electron energy

$$\begin{aligned} A &\propto n_\gamma \propto E_e \\ E_e &= g \cdot A \end{aligned} \tag{3.16}$$

with gain g. Generally, we use a mono-energetic electron source with known energy to determine g. We must expand this model for a realistic data set measured over time to account for temperature-induced fluctuations $f_T(t)$ of the gain g over time.

$$g(t) = g \cdot f_T(t). \tag{3.17}$$

There is no analytical model for the temperature dependence of $g(T)$, as PMTs are inherently highly non-linear. However, we can characterize the systematic effect with a mono-energetic source. We measure the total charge A with the source in front of the detector at different times for temperature-induced gain fluctuations. The data set we use for this example problem stem from the recent PERKEO III measurement, see Sec. 4.

With Gaussian process regression (GPR), we can reconstruct underlying signals in a non-parametric and Bayesian way. We apply GPR with a Matérn kernel function ($\nu = 3/2$) and a mean prior $m(\mathbf{t}) = 1$ to determine the temperature-induced gain fluctuation correction $c_T(t)$ at different times t from Equ. (3.17) with $c_T(t) = f_T(t)^{-1}$. We use hourly measurements with the mono-energetic source over about two weeks. We assume a constant, underlying gain, which we extract as the weighted arithmetic mean of the hourly measurements. The relative deviation from the mean is the pointwise correction $c_T{}'(t_i)$ for time of measurement t_i. Therefore, the correction is given by

$$c_T(t_i) = c_T{}'(t_i) + \epsilon \qquad \text{with} \qquad \epsilon \sim \mathcal{N}(0, \sigma_n^2)$$

in analogy to Equ. (3.5).

The data set and the GPR results are shown in Fig. 3.7. Using GPR reconstructs the underlying correction with statistical uncertainty and avoids overfitting to noisy data. In particular,

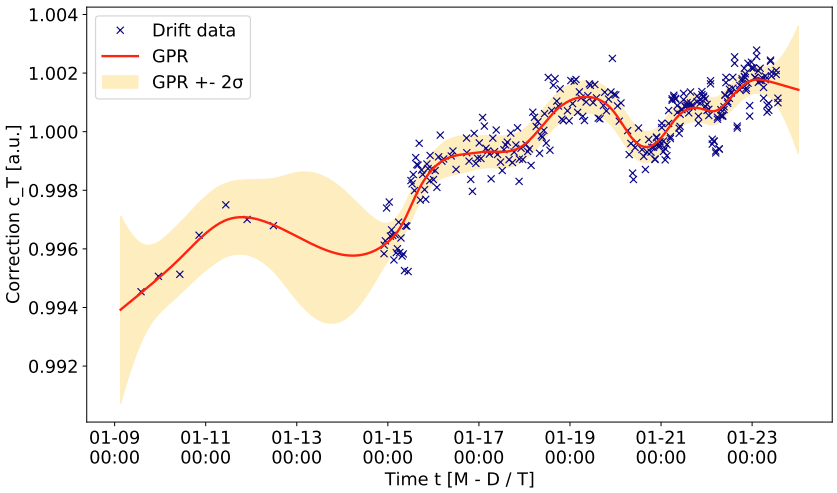


Figure 3.7 Temperature-induced gain fluctuation correction $c_T(t)$ with GPR, reconstructing the underlying signal.

the kernel function and its length scale parameter also guarantee this robustness in low-data regimes, see Fig. 3.7 before the 13th of January. Other standard tools in physics data analysis do not guarantee this benefit, e.g., splines or non-linear regression methods. Furthermore, even linear smoother will lack the uncertainty quantification and statistical motivation, as Gaussian processes generalize them. The uncertainty quantification specifically is essential for high-precision physics experiments. It enables accurate quantification of the systematic error on the final measurement, e.g., see Sec. 8. Furthermore, the GPR model parameters are interpretable, as the noise σ_n states the statistical fluctuation of the pointwise corrections $c_T{}'$ and the kernel function length scale ρ states the timescale for correction changes to occur. However, the interpretability depends on the choice of covariance function, as, e.g., stationary kernels fail to model data requiring varying local length scales.

The correction outperforms previous analyses in systematic uncertainty quantification [Sau18; Sau+20] and we directly improve the overall precision and quality of the PERKEO III analysis. Gaussian processes can model systematic corrections, e.g., by reconstructing unknown background time series, inferring energy-dependent corrections from simulations, or modeling spatial dependencies. We can also use Gaussian processes for event reconstruction through classification, see Sec. 3.3.3.

3.3.2 Detector Fine-Tuning

To build upon the electron detector example, we can correct another systematic effect that arises when taking data. This section is a partial reprint of [LBM22] and builds upon the collaboration with [Bes22], see Sec. 5.2.2.

Besides temperature-induced gain fluctuations, we must consider relative gain deviations between individual PMTs. PMTs are hard to tune for identical gain as they are inherently non-linear and because photon paths depend on the electron impact position on the scintillator. A different number of scintillation photons reaching each PMT leads to individual energy resolutions.[7] This spatial dependency, together with an energy-dependent spatial distribution

[7] This effect is worsened by the differences between PMT models, as it is not guaranteed that a detector uses PMTs of the same model and manufacturing batch. Also, PMTs degrade with age.

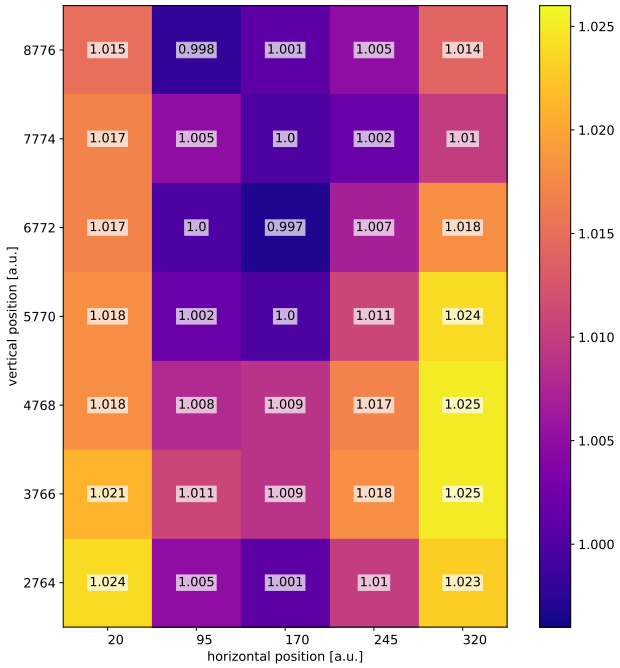


Figure 3.8 Unoptimized spatial response map with all peak positions μ_i shown relative to the central value μ_c. The objective values according to Equ. (3.19) are $\mathcal{L}_u = 22690$ for the uniformity and $\mathcal{L}_s = 9250$ for the symmetry.

of electrons on the detector, can lead to an energy-dependent gain $g(E)$. However, we can reduce this systematic effect by fine-tuning the individual PMT gain g_i to be equal or to create a spatially uniform and symmetric detector response. We can fine-tune each PMT i with a relative gain factor c_i to get the correct total charge A

$$A = \sum_i c_i A_i, \tag{3.18}$$

for the energy reconstruction in Equ. (3.16). We write all PMT fine-tuning factors as a vector $\mathbf{c}$.

We can also characterize the effect with a mono-energetic electron source with known energy. We measure the electron energy on a grid of different positions of the source in front of the detector, as such measurements contain more information about individual PMT contributions.

At each position k, we measure a spectrum of total charges $A' = A(\mathbf{c})$ for several electrons and extract the peak position μ_k with a fit of the histogram. For an ideal detector, the grid results are symmetric and uniform. An ideal fine-tuning should also lead to uniform and symmetric results. Without fine-tuning (i.e. $\mathbf{c} = \mathbf{1}$), we would accept non-uniform and asymmetric results. Such results would require hard-to-reach overall precision, as we must correct other effects depending on the detector positioning or the incoming electron distribution in the analysis. We illustrate the grid of measurements or *spatial response* of the detector as a map in Fig. 3.8 without fine-tuning.

We calculate the measured deviation from symmetry and uniformity of the resulting grid measurement as objective value $\mathcal{L}$ to turn it into an optimization problem as

$$\mathcal{L}(\mathbf{c}) = \mathcal{L}_u(\mathbf{c}) + \mathcal{L}_s(\mathbf{c}). \tag{3.19}$$

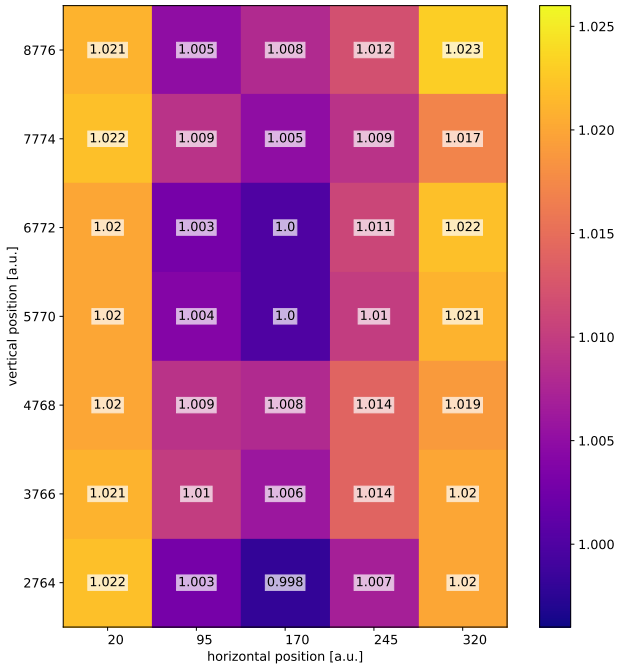


Figure 3.9 Spatial response map of Fig. 3.8 optimized with GPR resulting in the objective values from Equ. (3.19) with uniformity term $\mathcal{L}_\mathrm{u} = 6360$ and symmetry term $\mathcal{L}_\mathrm{s} = 710$. All peak positions μ_i are shown relative to the center value μ_c. Same color scaling as Fig. 3.8.

Minimizing $\mathcal{L}$ yields optimal gain factors $\mathbf{c}$. We define the uniformity term $\mathcal{L}_\mathrm{u}$ as the mean squared error of each grid point k for all n points to the central peak position μ_c for unoptimized gain factors, i.e. $\mathbf{c} = \mathbf{1}$.

$$\mathcal{L}_\mathrm{u}(\mathbf{c}) = \frac{1}{n}\sum_{k}^{n}\left(\mu_\mathrm{c} - \mu_k(\mathbf{c})\right)^2 \tag{3.20}$$

The reference point μ_c stabilizes the energy-channel relation before and after the optimization. To favor symmetric solutions, the square deviations from the m points in the left and rightmost columns are calculated in two groups: left to right columns and top to bottom halves of these columns are added as symmetry loss $\mathcal{L}_\mathrm{s}$

$$\begin{aligned}\mathcal{L}_\mathrm{s}(\mathbf{c}) = \alpha_1 \sum_{k}^{m}\left(\mu_{\mathrm{left},k}(\mathbf{c}) - \mu_{\mathrm{right},k}(\mathbf{c})\right)^2 \\ + \alpha_2 \sum_{k}^{m}\left(\mu_{\mathrm{top},k}(\mathbf{c}) - \mu_{\mathrm{bot},k}(\mathbf{c})\right)^2,\end{aligned} \tag{3.21}$$

with scaling factors $\alpha_1 = 0.5/m$ and $\alpha_2 = 0.5/(m-1)$. The symmetric deviation is only calculated for these columns, as they are most sensitive to any individual gain changes.

For a given set of gain factors $\mathbf{c} \in \mathbb{R}^8$ and the corresponding objective value $\mathcal{L}$, we need to recalculate all spectra for all grid points with processing and fit. This process requires time, and any solution method needs to guarantee symmetric and reproducible results. In addition, the PERKEO III data set has many such measurements to optimize, requiring high sample efficiency. However, the parameters $\mathbf{c}$ highly correlate, and the problem is hard to solve without brute force.

To predict minima of the objective function $\mathcal{L}(\mathbf{c})$ in Equ. (3.19), we require a model to infer values of $\mathcal{L}$ for untested gain factors $\mathbf{c}$. With an accurate model of $\mathcal{L}(\mathbf{c})$, we can evaluate it with little cost, needing fewer tested samples of $(\mathbf{c},\ \mathcal{L})$ pairs.

To this end, we use Bayesian optimization with Gaussian processes to fine-tune the individual PMT gain factors $\mathbf{c}$. We use GPR with a Matérn kernel function ($\nu = 5/2$). We employ a combination of improvement-based and optimistic policies as acquisition function by mainly using expected improvement (EI) and lower confidence bounds (LCB), see Sec. 3.2.1. We found $\kappa = 1$ for EI to balance a greedy search with the necessary exploration for the problem. When the EI utility is below a threshold and could lead to stagnation, we use LCB with $\kappa = 3$ for more exploration. We seed the approach with the best current optimum and random points at each iteration.

As comparison, we use empirically tuned random normal sampling ($\mathbf{c} \sim \mathcal{N}(\boldsymbol{\mu} = \mathbf{1}, \boldsymbol{\Sigma} = 0.05\ \mathbf{I})$) as baseline and a tree-structured Parzen estimator (TPE) [Ber+11]. TPE is also a case of Bayesian optimization, but instead of modelling the posterior distribution $p\,(y|x)$, it models the likelihood $p\,(x|y)$. The TPE model is a standard sequential model-based optimization approach for machine learning hyperparameter optimization, as it is less computationally heavy to compute than, e.g., GPR [Ber+11]. However, due to the covariance function, GPR is naturally more able to deal with highly correlated problems. We also tried the Monte-Carlo-Markov-Chain-based simulated annealing method, as it is commonly used in particle physics, but it is far too inefficient to be competitive. As an alternative approach to modelling the posterior distribution, we tried the novel deep adaptive design (DAD) method [Fos+21], a Bayesian optimal experimental design approach [Lin56] where an ideal strategy or optimization sequence is learned for a given number of steps, enabling rapid online inference and eliminating costly posterior update calculations. This would be useful to have an optimization policy for all 2D response maps, however, training turned out to be too costly for our example case.

All code is available on *GitHub*[8]. The spatial response maps are calculated with the PERKEO III analysis package *panter*, see Sec. 5.1. The Bayesian optimization algorithm is implemented with *Pyro* [Bin+18], which builds on *GPyTorch* [Gar+18] and we used the already implemented TPE from the *Optuna* [Aki+19] package.

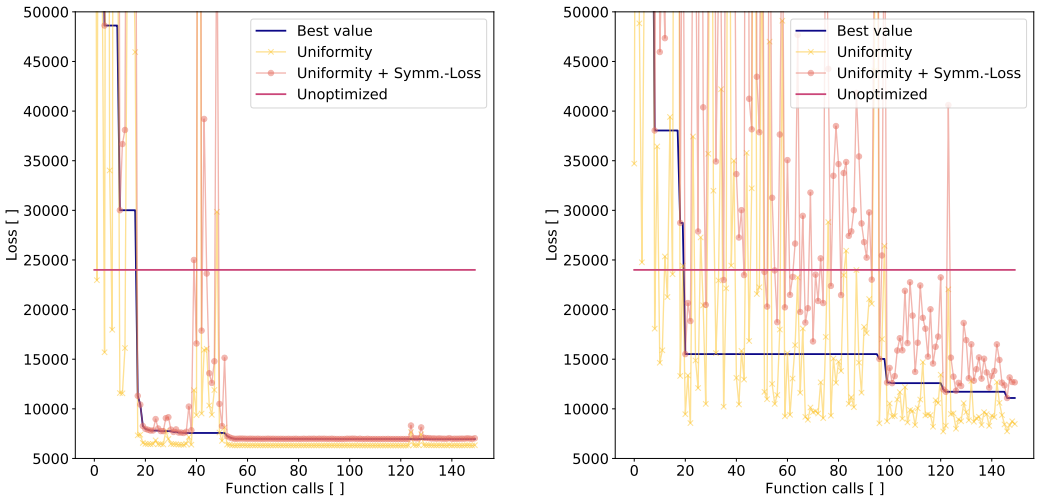
Symmetric uniformity is the overall goal, and for this problem, a minimized $\mathcal{L}_\mathrm{s}$ implies a small $\mathcal{L}_\mathrm{u}$. Therefore, we use the total loss $\mathcal{L}$ from Equ. (3.19) as objective value during the optimization and the symmetry loss $\mathcal{L}_\mathrm{s}$ for comparing results. Besides symmetry loss, we chose function evaluations as an additional metric. A function evaluation means calculating one pair $(\mathbf{c},\ \mathcal{L}(\mathbf{c}))$. We initialize the GPR with a few random data points from the random normal sampler, which we also count as function evaluations. To simplify the problem and compare methods, we combine PMTs and their gain factors c_i in symmetric groups of two (each corner). This grouping reduces the dimensionality and correlations of the problem by creating a fixed, embedded subspace of the original $\mathbb{R}^8$ input space at the cost of achievable minima quality. Tab. 3.1 presents the results for different PMT groupings and fixed function evaluations. One optimized spatial response map is shown in Fig. 3.9.

Bayesian optimization with GPR outperforms the other methods in overall best-achieved results and reproducibility. Contrary to GPR, the TPE approach suffers a high spread of results in Tab. 3.1. It is noteworthy how a well-tuned random normal sampler outperforms

[8] `https://github.com/maxlampe/detector-bayopt`

Table 3.1 Comparing optimization results for different methods and PMT groupings, i.e., different problem dimensionality. We fixed the number of function evaluations (calls) and run five uncorrelated optimizations for each setting for the map in Fig. 3.8, with the best achieved value $\mathcal{L}_s^*$ and the average achieved value $\bar{\mathcal{L}}_s$.

Dimensionality	Function calls	Method	$\mathcal{L}_s^*$	$\bar{\mathcal{L}}_s$	$\bar{\mathcal{L}}_s - \mathcal{L}_s^*$
		Unoptimized	9251		
$\mathbf{c} \in \mathbb{R}^4$	75	Random normal	1070	2900	1830
		TPE	1200	3800	2600
		GPR	**980**	**995**	**14**
$\mathbf{c} \in \mathbb{R}^8$	150	Random normal	4130	4640	510
		TPE	1540	2590	1050
		GPR	**643**	**675**	**32**



(a) Using Gaussian processes for Bayesian optimization

(b) Using the tree-structured Parzen estimator (TPE)

Figure 3.10 Optimization history with GPR for one run of Tab. 3.2. The current best optimization result for the given number of function calls is shown, as well as uniformity and symmetry loss terms. Between 40 and 55 function calls, we can see the acquisition function favoring exploration over exploitation, leading to an improvement.

the TPE approach, as the TPE approach struggles with the limited function calls. For all uncorrelated optimization runs, GPR would have required even fewer function calls as given, demonstrating its outstanding sample efficiency. This effect is shown in the optimization history for one run in Fig. 3.10, which is representative of all runs. In the Bayesian optimization with GRP run in Fig. 3.10a, we can see an exploration period leading to a better optimum at around 50 function evaluations. The most significant trade-off of using Bayesian optimization with GPR is the required computation time between testing samples to train the GPR and getting the next point in the sequential optimization. This computational requirement is why Bayesian optimization with GPR works best in cases where one evaluation is expensive, or the optimization is non-convex.

Bayesian optimization with GPR outperform previous analyses in systematic uncertainty quantification and achieve the best spatial response correction while being sample efficient and reproducible [Sau18; Sau+20]. In doing so, we improve the overall precision and quality of the PERKEO III analysis using GPR by modeling systematic corrections and optimizing analysis parameters. Gaussian processes provide well-calibrated predictive uncertainty for many tasks and are well-suited as an input model for Bayesian optimization. Bayesian optimization with GPR can optimize complex experiment setups and tune its parameters or determine energy cuts in correlated, costly analysis to maximize energy resolution.

3.3.3 Scalable Unsupervised Classification of Cosmological Shock Waves

The outlined methods in Sec. 3.1 enable an analysis of state-of-the-art-sized data sets, as they would also appear in particle physics. To illustrate this capability, I present the application of SV-DKL and kernel methods to a previously unsolved problem in computational astrophysics. The presented approach unites modeling flexibility with required statistical and physical motivation to better understand simulated data sets from modern supercomputers. This section is a partial reprint of [Lam+22a] and the extended abstract version of the same paper [Lam+22b]. As this example application is more complex than the others, I adhere to the paper structure and its sections for readability.

Introduction

Cosmological structures form by gravitationally accreting mass from their surroundings [Ber98; SD15; NO17]. As galaxies fall into clusters, they dissipate their energy in the form of shock waves in the diffuse gas between them, labeled as the intra-cluster medium (ICM) [Dol+05; Pfr+07; Pfr+17; Vaz+11; Vaz+16; SS15; Ste+22]. In these systems, the evolution of shock waves is the primary driver setting the global physical properties [Sed46; Tay50a; Tay50b; KO15; WN15a; Ste+20; Fie+17]. These shock waves are defined as discontinuities in density and temperature, propagating through the ICM. They are powerful accelerators of relativistic particles, which we can observe as synchrotron emission sources from merging galaxy clusters [Wee+19].

Modeling these cosmological systems with state-of-the-art simulations requires modern supercomputers, as there is large degeneracy in the possible geometry. The produced data sets contain up to $O(10^{10})$ particles, which we need to interpret to make conclusions about formation scenarios. However, shock wave structures in galaxy clusters form highly complex shapes and surfaces (see Fig. 3.11), and collisions between them lead to a superposition of different shock waves with overlapping geometries. From first principles, we can not make a simple, prior connection between in-falling substructures and shock wave surfaces. This

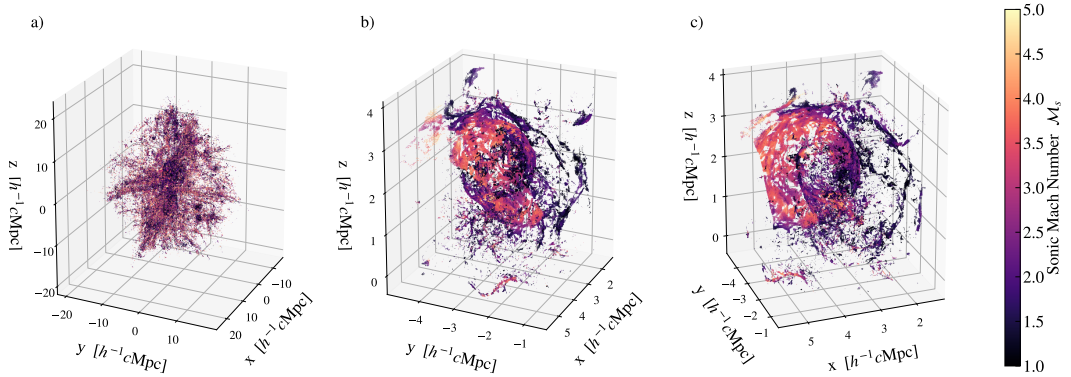


Figure 3.11 We show a reduced simulated data set CLUSTHD$_2$ containing all particles with a detected shock in the sonic Mach number range $\mathcal{M}_s \in [1, 5]$. *a)* Full simulation domain. The main cluster is in a cutout of the cosmic web, which shows internal and accretion shocks. *b)* Manual zoom into the cluster showing the complex shock structure of multiple ongoing merger shocks. We want to determine all shock surface particles as separate labeled groups and remove non-shock wave particles. *c)* Same as *b)*, but rotated by 45°.

setup poses a complex unsupervised classification problem for an unknown number of target classes in which we must find, separate, and label coherent shock wave structures in simulated data.

To this end, we propose a novel, physically motivated, and fully scalable machine learning pipeline to solve the outlined problem. We separate the unsupervised classification task by creating labels for a random subset of each data set and then training a classifier on that subset. For pre-processing, we exploit the non-stationarity of the problem with kernel principal component analysis (kernel PCA) [JC16; SSM98] and use Gaussian mixture models (GMM) [Bar12] to pre-clean the data from unwanted non-shock wave particles. For the subset classification, we further use physically motivated kernel functions with kernel PCA, Nyström approximation [DM05; RW06] and employ an agglomerative clustering (friends-of-friends, FoF [Dav+85]) algorithm with an automatically set linking length. Finally, we use the labeled subset to train a stochastic variational deep kernel learning (SV-DKL, DKL) [HFL13; HMG14; Wil+16a; Wil+16b] classifier to use our algorithm on the full data sets. For the first time, we can tackle this problem with our described pipeline and guarantee scalability for state-of-the-art and future data sets. To the best of our knowledge, this is the first application of DKL and SV-DKL to an astrophysical task. Furthermore, we are not aware of any preceding work having solved this unsupervised classification problem, in particular not with state-of-the-art resolution data sets.

Data Sets

We employ three data sets for the shock surface classification to gauge the flexibility of our pipeline for increasing physical complexity and different scales. All data sets were generated with OPENGADGET3[Spr05; Bec+16] (GNU). We name the data sets CLUSTHD (Fig. 3.11), CLUSTMHD and BOXMHD (Fig. 3.12). The first two are ultra-high-resolution (magneto) hydrodynamics simulations (MHD) of a single massive galaxy cluster ($\sim 10^9$ particles), and BOXMHD is a high-resolution simulation of a large cosmological volume with many clusters

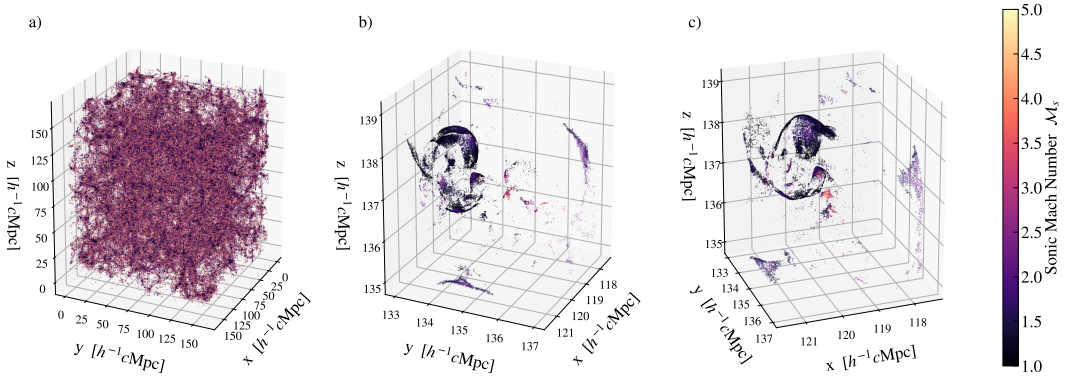


Figure 3.12 We show a reduced dataset for the BoxMHD simulation, as in Fig. 3.11. *a)* Full simulation domain. It contains the full cosmic web, as this is a cosmological box simulation. *b)* Manual zoom on one of the galaxy clusters forming at the intersection of cosmic filaments. *c)* Same as *b)*, but rotated by 45°.

($\sim 10^{10}$ particles), but at lower resolution [9]. The magnetic fields lead to additional motions perpendicular to the shock propagation and more patchy shock wave surfaces. BoxMHD is a simulation of a cosmological volume containing several cluster objects, whereas all other data sets consist of single cluster objects. For BoxMHD, our problem scales from an unknown number of target classes in one cluster to the same for an unknown number of clusters. We reduce the full data set of the simulation to only the parameters relevant for the actual shock surfaces. These are the spatial positions $\mathbf{x}$, the sonic Mach number $\mathcal{M}_\mathrm{s}$ and the shock normal vector $\hat{\mathbf{n}}_\mathrm{s}$.

VIRGO Model Pipeline

We propose a new analysis pipeline to solve the unsupervised classification of an unknown number of cosmological shock waves in a scalable, probabilistic and physically-motivated way. We separate our approach into four steps:

1) For data pre-processing, we remove data points above a conservative velocity threshold ($\mathcal{M}_s \leq 15$) and rescale the data set to a zero mean and unit variance. We can set this threshold, as large-velocity particles are rarely a part of shock waves surfaces. Our analysis only uses the particle position $\mathbf{x}$ and shock normal vector $\hat{\mathbf{n}}_s$. Each particle therefore is a 6-dimensional vector $\mathbf{q} = (x_x, x_y, x_z, \hat{n}_{sx}, \hat{n}_{sy}, \hat{n}_{sz})^\top$.

2) The raw simulation output is noisy with non-shock wave particles and not centered, as is illustrated in Fig. 3.11a and Fig. 3.12a. We use an RBF kernel with the Nyström approximation on the particle positions $\mathbf{x}$ for kernel PCA. We use GMM in the feature space with expectation maximization to separate the actual cluster of shock waves from non-shock wave particles by density estimation. This approach exploits the inherent local density changes of the problem by using a stationary kernel to separate non-shock wave particles from relevant shock wave particles.

[9] For more information, including a movie, please see `http://www.magneticum.org/complements.html#Compass`

3) We construct a physically motivated composite kernel k_{V} by adding two separate composite kernels made up of Matérn-5/2 kernels k_{M} and linear kernels k_{L}

$$k_1(\mathbf{q}, \mathbf{q}') = k_{\mathrm{M}}(\mathbf{x}, \mathbf{x}') \cdot k_{\mathrm{L}}(\mathbf{x}, \mathbf{x}') \tag{3.22}$$

$$k_2(\mathbf{q}, \mathbf{q}') = k_{\mathrm{M}}(\mathbf{x}, \mathbf{x}') \cdot k_{\mathrm{L}}(\hat{\mathbf{n}}_s, \hat{\mathbf{n}}'_s) \tag{3.23}$$

$$k_{\mathrm{V}}(\mathbf{q}, \mathbf{q}') = k_1(\mathbf{q}, \mathbf{q}') + k_2(\mathbf{q}, \mathbf{q}'). \tag{3.24}$$

k_1 creates a non-stationary kernel for spatial information, whereas k_2 combines local spatial information with shock normal directions of the particles. We combine k_{V} with the Nyström approximation and PCA, accepting a reduction of the data set to a random subset for computational limitations. The resulting feature space enables separation with a fixed linking length β FoF algorithm. We estimate β with the average n-next-neighbor distance in the resulting feature space. We automatically label every unclassified particle as non-shock wave particle.

4) We use this labeled subset to train an SV-DKL classifier. With its deep kernel network, we gain a locally adaptable similarity metric required for robust classification. The SV-DKL framework allows us to achieve fast inference and good scalability, as we are not limited by the size of the data set.

Our approach is distinctly scalable, as the proposed pipeline succeeds for data sets of extreme sizes: We can downsize the data set at each step only to recover full resolution with the SV-DKL at the end. We collect our analysis in a Python software package to be available for future analysis, called VIRGO (**V**ariational **I**nference package for unsupe**R**vised classification of (inter-)**G**alactic sh**O**ck waves)[10]. The package utilizes already implemented features of PyTorch [Pas+19](BSD), GPyTorch [Gar+18](MIT), scikit-learn [Ped+12](BSD) and pyfof [Gib19](MIT).

Experiments

We evaluate our pipeline on the generated data sets, as this is a previously unsolved problem and there are no benchmark data sets available. Different time steps of one simulation, indicated by an index, are quasi-independent data sets to be solved due to morphing structures and changing number of target classes. As there exist no labeled data sets, we must verify the results visually by the coherence of the shock wave surface classification and the removal of non-shock particles. In our studies, we observe that any other approach visibly over- or under-segments the shock waves.

First, we evaluate one single cluster data set where only one shock wave structure is present within many non-shock wave particles. We show the denoising and centering process representative for CLUSTHD$_2$ in Fig. 3.13. Our approach accurately separates the dense cluster region from the general simulation output. Should more structures be present, we increase the number of GMM components and obtain reliable results for all tested data sets. In doing so, we achieve good separation for all available and tested data sets. We classify the denoised result as described in step 3) while reducing the data set in size to a random subset. However, this step reduces the data set size from the original $O(10^6)$ to $O(10^4)$. We recover full resolution with the SV-DKL classifier trained on the labeled subset. This final classification does not depend on the choice of the random subset in the previous step. Fig. 3.16 shows the reconstructed and labeled data set of CLUSTHD$_2$. The complex

[10]The source code is available at https://github.com/maxlampe/virgo

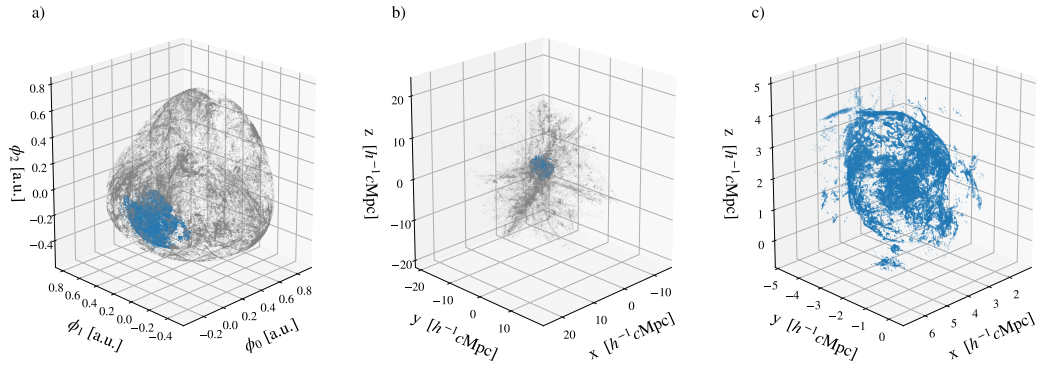


Figure 3.13 We show the denoising process of Sec. 3.3.3 step 2) and its impact on the CLUSTHD$_2$ data set from Fig. 3.11. *a)* GMM fitted in kernel-PCA space with low-density noise component (gray) and high-density cluster component (blue). We only plot the first three principal components of the kernel space dimensions. *b)* Labeled data from *a)*, but shown in physical space. *c)* Resulting denoised data set for further analysis.

morphology of the shock waves and its substructures are restored and correctly labelled. The figures show the clear separation and classification of the shock surfaces without over-segmentation. Any non-shock-wave particles are labeled as noise with a visibly low error rate.

To test the extent of VIRGO's capabilities, we run the pipeline on the more complex BOXMHD data set. We repeat step 2) for this data set twice, once with 35 GMM components, keeping the densest ten components, and then again on each remaining component with the parameters described above. This additional step is required to single out dense objects and do single cluster shock wave analysis like in for CLUSTHD. We illustrate the labeling of the subset from step 3) of the biggest substructure in BOXMHD in Fig. 3.15. The scalability with the SV-DKL works as well as for the other data sets. Some substructures are small enough for data size not to require the SV-DKL, as the Nyström approximation is not required, demonstrating that VIRGO is applicable to much larger and detailed data sets.

We compare the SV-DKL against a k-nearest-neighbor (k-NN, $k = 10$) classifier and a fully connected NN (like deep kernel NN) in Tab. 3.2 to test the propagated error for the resolution reconstruction. All models are trained from scratch and data sets are re-shuffled. The SV-DKL outperforms the other methods in accuracy. We expect the SV-DKL to perform better with non-shock wave particles near shock wave surfaces, due to the additive Gaussian processes. However, k-NN achieves decent accuracy, and we recommend it as a cost-effective replacement for online applications. VIRGO also successfully separates and labels shock waves on the more complex BOXMHD data set. We repeat step 2) for this data set twice to deal with the multiple cluster objects. This additional step is required to single out dense objects and do single cluster analysis. In addition, VIRGO shows signs of generalization, as we used the trained classifier from the labeled subset of CLUSTHD$_2$ on the full data set of CLUSTHD$_3$ and obtained good results as well. However, this requires the same amount of target shock wave classes. Overall, VIRGO solves the outlined classification problem of cosmological shock waves and delivers robust results on all tested data sets.

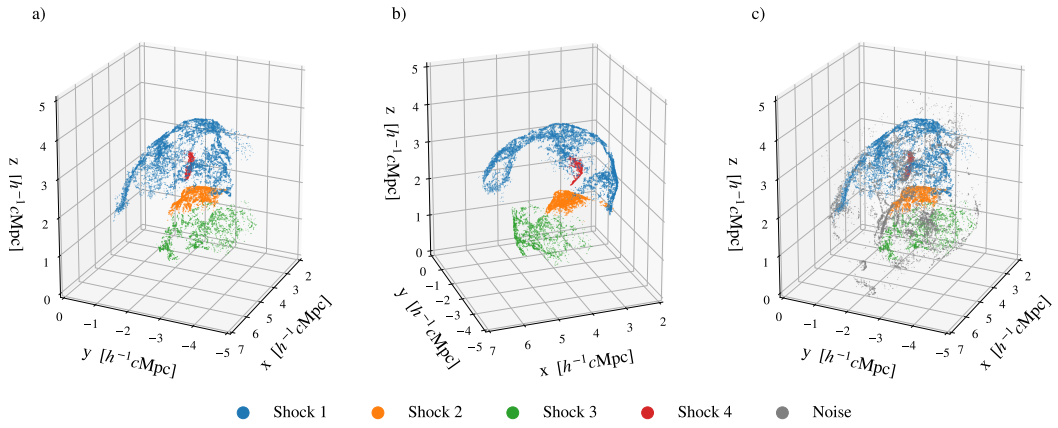


Figure 3.14 We show the labeled subset of CLUSTHD$_2$ after the denoising process. The clear separation of the shock wave surfaces from non-shock wave particles (gray) is visible. Also, the estimated error of labeling shock surface particles as non-shock wave particles is visibly negligible. *a)* Labeled subset with the FoF algorithm of step 3) in Sec. 3.3.3. *b)* same as a), but rotated by 45°. *c)* Same as a), but with non-shock wave particles plotted too.

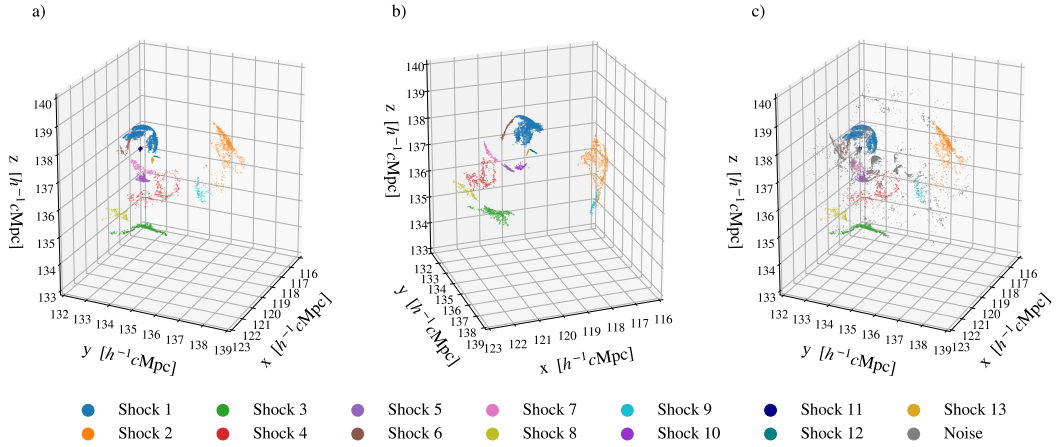


Figure 3.15 We show the labeled subset of BOXMHD after the denoising process. The clear separation of the shock wave surfaces from non-shock wave particles (gray) is visible. Also, the estimated error of labeling shock surface particles as non-shock wave particles is visibly negligible. *a)* Labeled subset with the FoF algorithm of step 3) in Sec. 3.3.3. *b)* same as a), but rotated by 45°. *c)* Same as a), but with non-shock wave particles plotted too.

Table 3.2 Comparing average test accuracies on the labeled subsets of the data after step 3) in Sec. 3.3.3 for different methods on different data sets for ten independent runs.

Method	CLUSTHD$_1$	CLUSTHD$_2$	CLUSTHD$_3$	CLUSTMHD$_1$	BOXMHD$_1$
k-NN	97.10 ± 0.34	96.61 ± 0.32	97.19 ± 0.30	96.57 ± 0.39	96.69 ± 0.48
FC-NN	95.33 ± 1.32	95.51 ± 0.84	96.63 ± 0.41	96.19 ± 0.50	95.05 ± 0.69
SV-DKL	**97.57 ± 0.54**	**97.00 ± 0.49**	**98.36 ± 0.18**	**98.08 ± 0.16**	**98.02 ± 0.37**

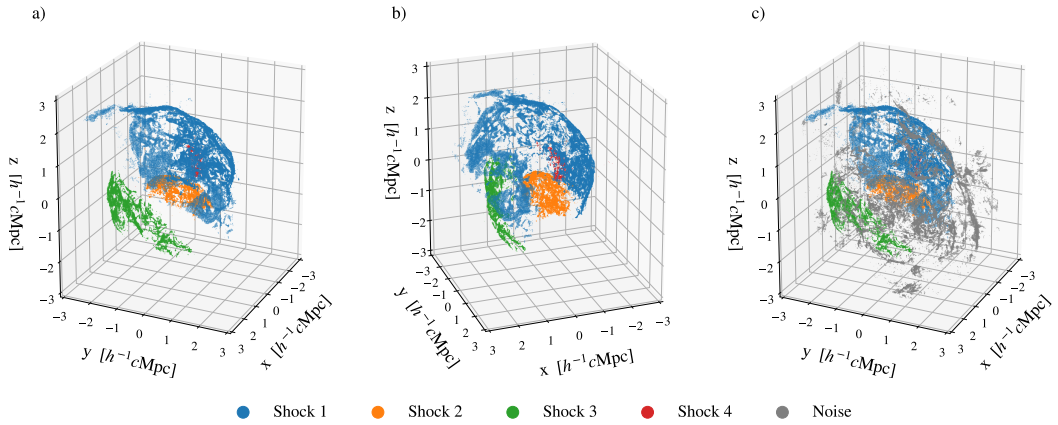


Figure 3.16 We show the full resolution reconstruction of CLUSTHD$_2$ with the SV-DKL classification. This result is the final VIRGO output for the raw input in Fig. 3.11 and allows scalability to large data sets. We state the accuracy of the recovery classification in Tab. 3.2. *a)* Labeled data set with SV-DKL classifier of step 4) in Sec. 3.3.3. *b)* same as a), but rotated by 45°. *c)* Same as a), but with non-shock wave particles.

Discussion

We demonstrated the capability of VIRGO to capture the irregular shapes of shock wave surfaces. For future work, we propose using VIRGO to improve large-scale galaxy-cluster simulations by increasing the efficiency of particle injections at shock structures [DB14; WVB17; Win+19; Bös+22] or to study supernovae remnants [Jan+12]. We determine the linking length estimator from step 3) to be most prone to error and limitation. However, our data sets are insufficient to construct an estimator for this hyperparameter without overfitting. Also, labeling errors in step 3) will be propagated by the SV-DKL. The Gaussian process might correct minor errors, but this will not fix larger misclassifications. For applications to more complex data sets, VIRGO should be combined with a better structure finder and a criterion for actual shock wave detection. Future work should verify the robustness of our chosen hyperparameters in a broader set of simulated data, as this might pose a challenge for users. We also propose training the same DKL over different data sets with SV-DKL to yield a more generalizable solution. The DKL could be combined with PCA and k-NN to achieve better computational scalability, classification for an unknown number of shock waves in a cluster, and robustness regarding the linking length hyperparameter, as the pre-trained DKL could even replace or at least improve the subset labeling of step 2).

Due to the probabilistic and non-parametric nature, Gaussian processes remain state-of-the-art for uncertainty quantification tasks [STM20; Abd+21] and are ideal for physics applications, as we can incorporate properties of physical systems. Furthermore, interpretable model parameters set Gaussian processes apart from other machine learning methods. Gaussian processes found applications in astrophysics [SKL12; HV14; Moo+16; Lec18; KCW22; Mil+22], condensed matter physics [HMK21], and dynamical systems [CK15; GFV21], but few in particle physics - presumably due to the size of data sets. However, expanded models like SVGP and SV-DKL should diminish previous limitations.

To conclude, we introduced a novel, physically motivated, and scalable pipeline. We are not aware of any preceding work having solved this unsupervised classification problem of cosmological shock waves, in particular not with state-of-the-art resolution data sets.

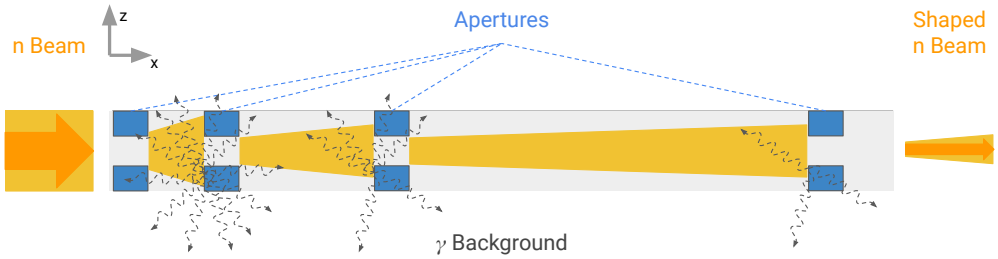


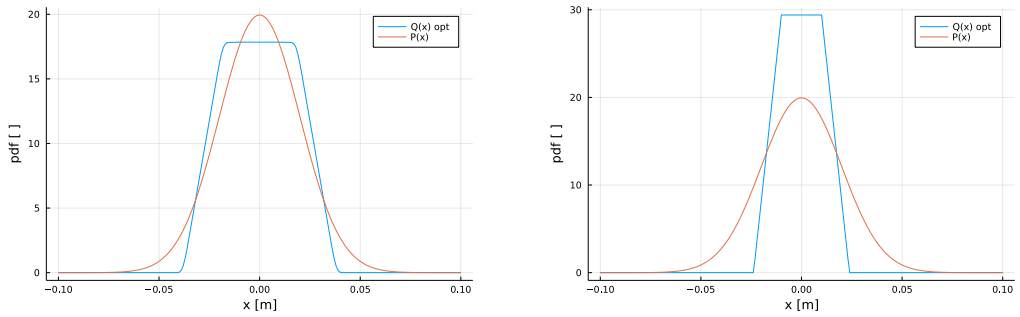
Figure 3.17 Graph and caption adapted from [Dor+23]: Schematic of the neutron beam-line optimization for $n = 4$ neutron absorbing quadratic apertures with position and width as parameters. Beam divergence and background are only drawn for illustration and not physically accurate.

3.3.4 Neutron Beam Line Design Optimization

As an example of how to use differentiable programming, see Sec. 3.2.2, I look at the example of optimizing the aperture positions in a neutron beam line to match a desired beam profile for high-precision particle physics experiments. Also, I outline the importance of differential programming for low-energy particle physics. This section is a reprint with minor adjustments of my contribution to [Dor+23]:

Low-energy particle physics provides many unique research opportunities to search for exotic particle candidates or beyond Standard Model physics. They range from decay [GNS19; DM21] to electric dipole moment (EDM) [Chu+19] and a variety of other measurements [JR10]. Specifically, tritium decay [Ake+19], neutron decay [Sau+20; Sun+20; Wan+19], neutron lifetime [Gon+21], and neutron EDM [Abe+20] experiments produce key results. These experiments are high-precision measurements designed for specific purposes, leading to complex designs. These experiments have many tunable parameters, and we must design and operate them optimally to maintain continuous improvements of experimental results. This quality requirement demonstrates the importance of advanced methods based on differentiable programming in this field. Parameters for the optimization of such experiments are case-specific. However, we aim to set global and local parameters to improve measurement precision by reducing uncertainties. We may choose whether to do end-to-end optimization for local or global experimental parameters, depending on the complexity and effective dimensionality of the experiment. In some cases, differentiable programming is not ideal, as other methods achieve better convergence, such as Bayesian optimization [Sha+16] as mentioned in Sec. 3.2.1 and [Dur+20; RHE21; LBM22].

To illustrate the capabilities of differentiable programming for low-energy particle physics, we choose the example of optimizing a neutron beam-line as used in [Sau+20; Wan+19] by tuning aperture placements and their width for desired beam characteristics. We can customize and adapt the optimization pipeline of [Dor+23] to the problem. Consider a source of cold neutrons reaching an experiment through neutron guides and a velocity selector. The neutrons have a known wavelength, transverse momentum, and position distribution depending on the neutron guides and the velocity selector. We position a set of n quadratic neutron apertures between the velocity selector and the experiment to shape the resulting beam distribution, as shown in Fig. 3.17. The beam distribution can be calculated analytically by trigonometry and a set of integrals, requiring no surrogate model [Abe+06; Mär06; Wan13]. We can calculate the beam distribution $Q(x)$ of the current setup for a fixed detector position on the beam axis, with x being the distance perpendicular to the beam center. Optimizing the beam distribution is essential to reduce systematic effects,



(a) **(b)**

Figure 3.18 Graph and caption taken from [Dor+23]: "Two different optimized beam-line distributions $Q(x)$ for $n = 3$ apertures. (a) uses only D_{KL} and $\mathcal{L}_2$, enforcing a good distribution approximation of $P(x)$ with equally sized apertures; (b) utilizes $\mathcal{L}$ as in Equation 3.25, sacrificing function approximation quality for lower background levels."

maintain experiment confinements, or other constraints like costs. The two most significant systematic effects are beam homogeneity and created background signals by the beam-line through neutron absorption. We may encode the desired beam homogeneity or shape in a target distribution $P(x)$. Therefore, we set the optimization objective as KL divergence or *relative entropy* D_{KL} [KL51] of $P(x)$ and the resulting beam distribution of the current setup $Q(x)$ as

$$D_{\mathrm{KL}}(Q \mid\mid P) = \sum_{x \in \mathcal{X}} Q(x) \log\left(\frac{Q(x)}{P(x)}\right).$$

Furthermore, we can expand the objective value with additional terms addressing different systematic effects. It is beneficial to place apertures further away from the experiment to minimize beam-line-induced background. We add the distance p_i of aperture i to the beam-line start as the first objective function adaption as

$$\mathcal{L}_1 = \frac{\alpha_1}{n} \sum_{i=1}^{n} p_n^2.$$

We also add aperture width w_i of aperture i as regularization

$$\mathcal{L}_2 = \frac{\alpha_2}{n} \sum_{i=1}^{n} w_n^2.$$

We use the weighting parameters α_1 and α_2 to tune the importance of each term. Therefore, the total optimization objective $\mathcal{L}$ for the differential programming pipeline is

$$\mathcal{L} = D_{\mathrm{KL}}(Q \mid\mid P) + \mathcal{L}_1 + \mathcal{L}_2. \tag{3.25}$$

We keep the model general by optimizing only aperture widths and positions, so that it can be adapted to specific experiments. We present example results in Fig. 3.18 with different regularization results for a Gaussian target distribution. The *Julia* code and pipeline is available on *GitHub*[11] and uses the *ForwardDiff* package [RLP16]. Beyond beam-line optimization, we can also optimize the geometry of scintillation energy detectors for optimal light transport, leading to better energy resolution and detector uniformity. Such detectors are often used for low-energy particle physics, and we propose using a differentiable programming optimization pipeline as described in [Dor+23] with a surrogate model from simulations to achieve optimal detector performance.

[11] https://github.com/maxlampe/NobleAD

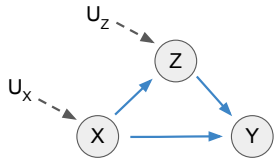


Figure 3.19 Structural causal model for an input variable X, a mediating variable Z, and an output variable Y represented as directed acyclical graph. The variables are shown as nodes, and causal relations are shown as directed edges. The functional relationships between variables are not shown. Also plotted are two exogenous variables U_Z and U_X that represent external, random influences not accounted for in the model.

3.4 Other Approaches

3.4.1 Causal Inference

Contrary to frequentist or Bayesian statistics that analyze the correlations between variables, causal inference studies causal relationships between variables using statistical methods. With causal inference, we can determine whether changing one variable causes another variable to change and establish causal relationships.

We use structural causal models as a framework for expressing causal relationships between variables [Pea09b]. Structural causal models can be abstracted as directed acyclical graphs that show the variables as nodes and relations as directed edges, as illustrated in Fig. 3.19 for an input variable X, a mediating variable Z, and an output variable Y. Also plotted are two exogenous variables U_Z and U_X that represent external, random influences that are not accounted for in the structural causal model. The required temporal order of variables for a structural causal model is encoded in the directed acyclical property of the graphs. However, the functional relationships f between variables, i.e., $Z = f_X(X, U_Z)$ or $Y = f_Y(X, Z)$, are not shown in Fig. 3.19. These functional relationships describe how changes in variables propagate through the structural causal model and how they influence each other.

To avoid biases and inaccurate results, the structural causal model must contain all relevant variables in temporal order and have independent exogenous variables. The functional relationships f are an input assumption to create the structural causal model, but not necessarily required when using causal discovery algorithms [GZS19].

Using the structural causal model in Fig. 3.19, we can calculate the total effect TE on Y for two different inputs x and x' as [Pea09a]

$$\mathrm{TE}_{x,x'}(Y) = \mathbb{E}\left[Y_x - Y_{x'}\right]$$

Also, we can conduct causal mediation analysis to calculate the direct DE and indirect IE effects of a mediating variable Z for two different inputs x and x' as [Pea09a]

$$\begin{aligned} \mathrm{IE}_{x,x'}(Y) &= \mathbb{E}\left[Y_{x,Z_{x'}} - \mathbb{E}\left[Y_x\right]\right] \\ \mathrm{DE}_{x,x'}(Y) &= \mathbb{E}\left[Y_{x',Z_x} - \mathbb{E}\left[Y_x\right]\right] \end{aligned} \tag{3.26}$$

by replacing the value of the mediating variable Z for one input, while evaluating the structural causal model for the respective other input. The principle of causal mediation

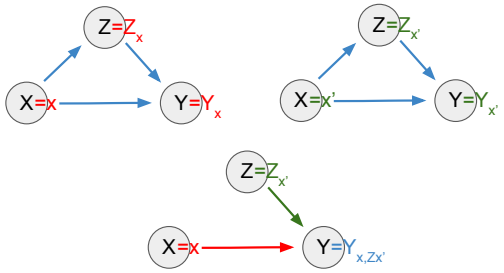


Figure 3.20 Demonstrating causal mediation analysis to calculate the direct or indirect effect for two different inputs x and x' for the mediating variable Z for the structural causal model in Fig. 3.19. Shown is the indirect effect IE calculation as stated in Equ. (3.26). We estimate the change in Y for setting $X = x$ but changing Z to the value it would have had for $X = x'$. The exogenous variables are omitted for visualization.

analysis is also illustrated in Fig. 3.20. Calculating the direct and indirect effect estimates the effect of a mediating variable Z on the total effect TE and can be interpreted as a measure of the importance of Z for these inputs. The calculations in Equ. (3.26) generalize to structural causal models of any sizes and number of mediating variables beyond the model in Fig. 3.20.

Causal inference is a separate research field from machine learning. However, we can use machine learning methods to construct structural causal models by training surrogate models from observations, as, e.g., neural networks are directed acyclical graphs by design. Alternatively, we can use causal mediation analysis to study the importance of hidden states as mediating variables for different inputs to make neural networks more interpretable. This approach is successfully used for interpretability research on internal states of advanced deep learning models, e.g., in large language models [Men+22; Wan+22; LR23].

We can also leverage these benefits for high-precision experiments. The detector response models and signal chains of experiments are inherently directed acyclical graphs that contain all or most of the relevant variables for the final signal. Therefore, we can determine individual detector components' direct and indirect effect on final quantities to determine functional behavior or guide optimization. I use this in Sec. 6.1.7 to determine the direct effect of individual signal charge integrator channels when studying the integrator non-linearity. Using causal mediation analysis led to the hypothesis and study of integration time-dependent systematic effects and the discovery of required future work.

3.4.2 Related Work

In the previous sections, I already discussed examples of existing physics applications of Gaussian processes, Bayesian optimization, and differentiable programming. Many more promising methods not covered in this work show significant potential for physical sciences. As explaining these methodologies goes beyond the scope of this work, I only give a curated overview of applications. These methodologies can be helpful for high-precision experiments, especially for modeling, inference, and simulation applications.

Due to their flexibility and expressiveness, neural networks have been widely used in physical sciences with applications ranging from quark tagging [Pea+17], to accelerator tuning [XRE22], and to planet formation studies [Pfe+22]. Also, they are used for symbolic regression [But+23] and for fluid simulations with Bayesian deep learning [Mue+22].

Graph-neural networks are promising for physics applications as they propagate information between neighboring nodes. This characteristic allows them to model local interactions that accurately capture global properties. Graph-neural networks are successfully used, e.g., for pattern recognition in the IceCube experiment [Abb+22] and to model the scattering behavior of cosmic muons in different materials [Lag+22]. Future studies might combine topological machine learning [HMR21] to enhance Graph-neural network approaches for physical sciences.

Generative models are advantageous for simulations of physical systems, as they learn a distribution from training data and can sample new data from the learned distribution for further analysis. One example is generative adversarial networks, which were successfully used for emulating cosmological multifields [AVH22] and for ultra-high-resolution detector simulations [Has+23]. Other works applied diffusion models to generate astronomical spectra [Doo+22] or normalizing flow models to gravitational wave studies [SSH22; Wil+23].

Learning policies for an agent with reinforcement learning by interacting with an environment by trial and error and reward maximization can lead to the discovery of novel strategies. Thus, reinforcement learning found applications for string theory vacua discovery [HNR19], for protein structure prediction [Jum+21], and for controlling the magnetic fields for tokamak plasma [Deg+22].

4 Experiment Setup

The physics goal is to look for BSM physics in neutron beta decay by precisely measuring the electron energy spectrum of these decays, see Equ. (2.9) and (2.12). To this end, we use the pre-existing spectrometer PERKEO III [Mär06; Mär+09], which was designed to study the neutron beta decay and the correlation coefficients in Equ. (2.9). Recent scientific highlights are high-precision measurements of the beta asymmetry A and the Fierz interference term b [Mär+19; Sau+20].

In this chapter, I present the experimental setup and how we adapted the spectrometer from previous measurements for a high-precision measurement of the electron energy spectrum at the Institut Laue-Langevin (ILL) in France. During the measurement campaign, I managed the preparation of the scientific equipment, the logistics, and the conduction of the measurement at the ILL. The ILL committee accepted our proposal for funding and instrument access at the ILL [Lam19] as a scientific highlight, enabling a measurement campaign from July 2019 to March 2020. Our team consisted of four constant scientists and up to twelve people during the setup period.

4.1 Measurement Concept

In analogy to the discovery of parity violation [Wu+57], the core idea of PERKEO III is to place the decaying neutrons in a magnetic field with the field lines going to two detectors, as illustrated in Fig. 4.1. The magnetic field has its maximum in the decay volume and decreases towards the detectors. Similar setups are also used in other neutron beta decay experiments, e.g., PERKEO II [Mun+13] or UCNA [Sun+20].

This setup has multiple benefits: Firstly, the charged decay products gyrate towards the detector in the direction of their initial emission due to the magnetic field. This guidance enables a full $2\times2\pi$ angular coverage, reducing statistical uncertainties of measurements and corrections from electrons missing the detectors based on their energy-dependent gyration radius. Secondly, this setup enables full energy reconstruction in cases of backscattering.

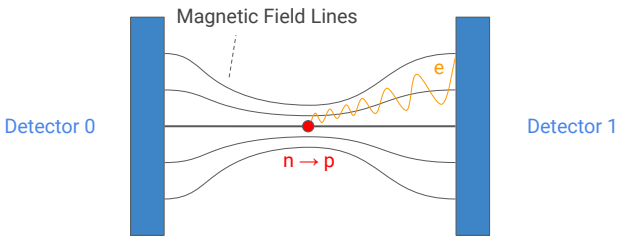


Figure 4.1 Schematic of the PERKEO III measurement concept. A neutron decays in a magnetic field and the electron gyrates toward one of the two detectors. Our setup is not sensitive to the other decay products, e.g., protons, and they are not illustrated. The gradient of the magnetic field leads to reflections of backscattered electrons due to the magnetic mirror effect.

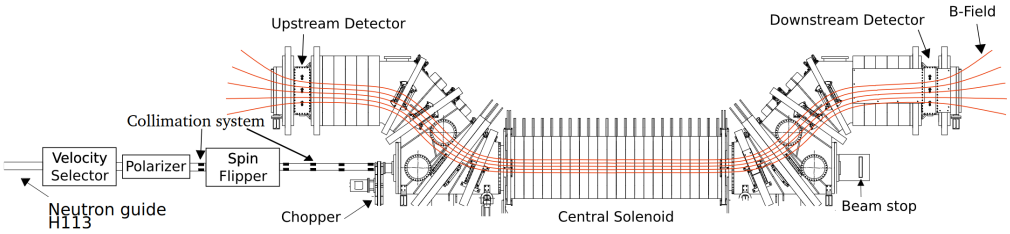


Figure 4.2 Graph taken from [Mes11]: PERKEO III schematic with beamline. The magnetic field lines are drawn in orange. The distance from detector to detector is about 8 m. We adapt this setup for our measurement by, e.g., removing the polarizer and spin flipper to increase the event rate.

If an electron backscatters off a detector, it only deposits a fraction of its kinetic energy in the primary detector, creating energy losses critical for high-precision energy measurements without a secondary detector. With our setup, the backscattered electron is either detected by the other secondary detector or is reflected onto the primary detector through the magnetic mirror effect induced by the magnetic field gradient, leading to the energy reconstruction of the event. Thirdly, the field's orientation from one detector to another can be used as a quantization axis for the neutron spin when measuring neutron-spin-dependent decay correlation coefficients. Besides these benefits, the setup reduces other systematic effects related to neutron spin-dependent measurements [Mes11; Sau18; Roi18].

4.2 PERKEO III Spectrometer

The PERKEO III spectrometer [Mär06; Mär+09] is a realization of the measurement concept in Fig. 4.1, and the experiment setup is shown in Fig. 4.2. We can divide the spectrometer into four components: The beamline, the central decay volume, and two detector vessels. The beamline connects the experiment to the ILL instrument site and produces a suitable neutron beam for the experiment, see Sec. 4.2.4. The 50 copper coils of PERKEO III create a magnetic field that separates the charged decay products from the neutron beam. The neutrons fly parallel to the magnetic field direction in the central decay volume. If a neutron decays in the magnetic field of the central decay volume, its charged decay products are guided to the detectors. A beam stop absorbs any remaining neutrons after exiting the central decay volume. More details on the PERKEO III setup are given in [Mär+09; Mes11; Wan13; Sau+20], and I highlight any changes specific to the 19/20 measurement campaign to determine the Fierz interference term in this section.

The support structure of the PERKEO III spectrometer is a vacuum vessel with mountings and flanges. The central decay volume is a cylindrical vessel with 2.7 m length and 50 cm inner diameter. The two detector vessels on each side have a length of 3 m and a rectangular cross-section of 50x50 cm^2. There is no direct line-of-sight from the beam axis to the detectors. We used turbomolecular pumps to reach a vacuum of 10^{-6} to 10^{-7} mbar throughout the experiment. This vacuum is sufficient to avoid perturbations of the electron energies.

The highest magnetic field B_0 = 152 mT is in the middle of the central decay volume, decreasing towards the detectors with B_d = 82 mT. The magnetic field gradients are locally homogeneous, allowing for adiabatic transitions with the Lorentz force, leading to circular movements of the charged particles perpendicular to the field lines, so-called gyration. The

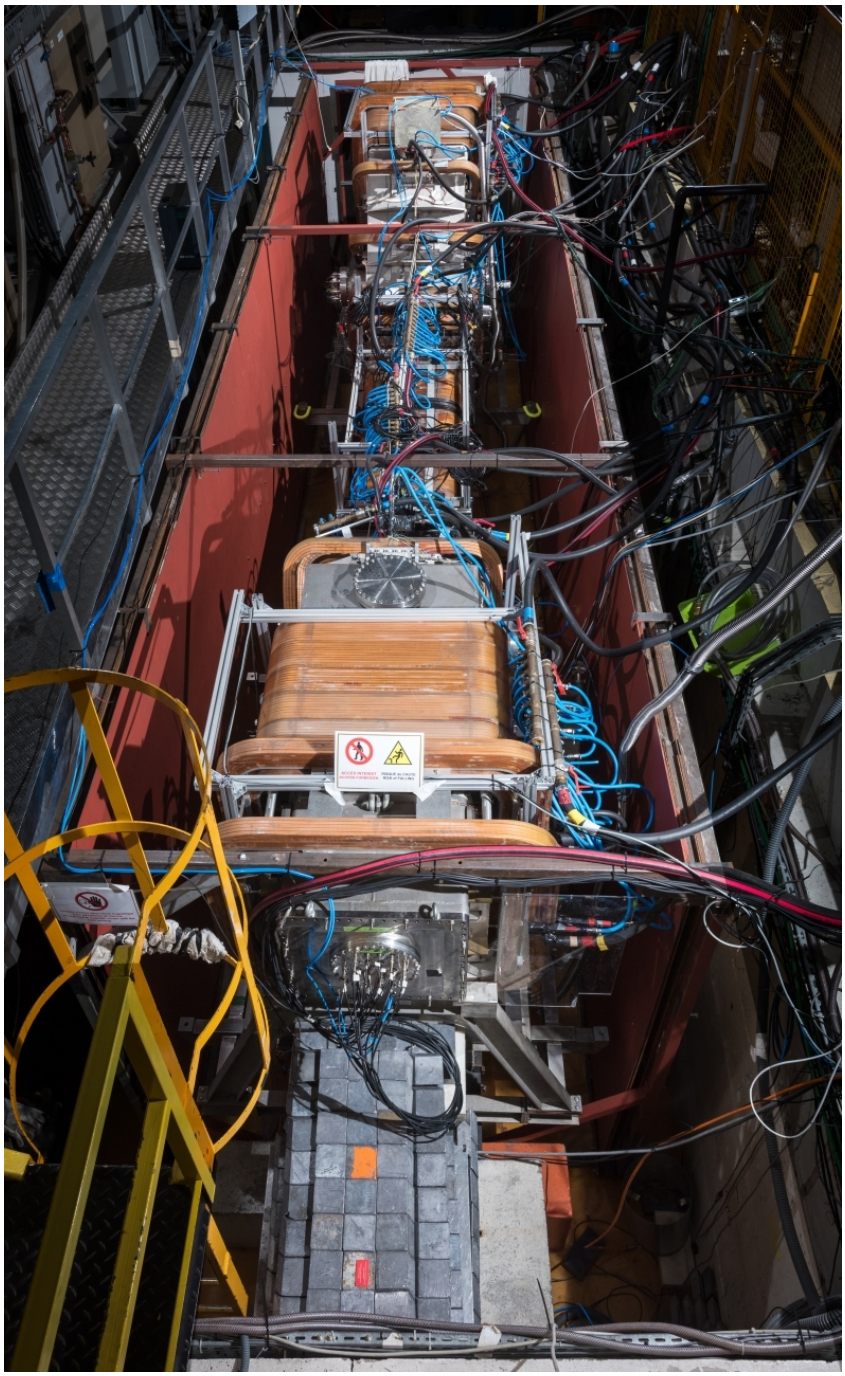


Figure 4.3 The PERKEO III spectrometer with magnetic shielding and connected supplies at the PF1B instrument site at the ILL, France. The PF1B instrument site is about 14 m long and 3.5 m deep and we setup the experiment in about 6 weeks, visualizing the magnitude of the endeavour. Picture taken standing on the beamline looking in the neutron beam direction. (Copyright: Laurent Thion, 2019)

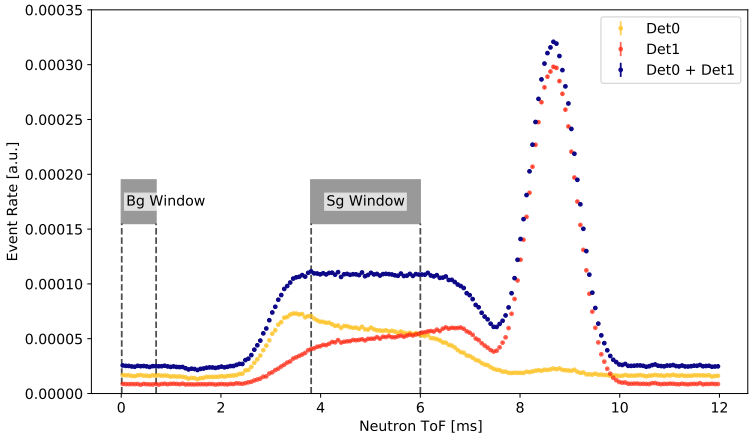


Figure 4.4 The number of detected events plotted against the neutron time-of-flight (ToF) to illustrate the principles of a pulsed beam measurement. The time-of-flight time is calculated as the difference between the detection time of an event to the last chopper rotation. The event rate corresponds to the neutron pulse propagating through the experiment, e.g., the rate increases between 2 and 4 ms when the neutron pulse enters the central decay volume and electrons from decays are guided to the detectors. We use time-of-flight information to extract signal and background spectra, enabling an accurate background subtraction every couple of ms. The signal and background measurement windows are indicated, and I discuss them in Sec. 5.3.2.

gyration radius in the central decay volume for an electron with kinetic energy E emitted at an angle θ to the beam axis, mass m_e, and charge e is given by [Jac75]

$$r_0 = \sqrt{E(E + 2m_e c^2)}\,\frac{\sin(\theta)}{ceB_0}. \tag{4.1}$$

The magnetic field gradient between the central decay volume and the detectors leads to a larger gyration radius at the detectors of $r_d = r_0\sqrt{B_0/B_d}$, derived through momentum conservation [Jac75; Wan13]. Also, the magnetic field gradient leads to a forward focusing of the charged particles. For example, an emitted electron with an approximately $\theta_0 = 90°$ angle to the beam and magnetic field axis has only a maximum angle of entry θ_c of 47° at the detectors

$$\theta_c = \arcsin\left(\sqrt{\frac{B_d}{B_0}}\sin(\theta_0)\right) = 47°. \tag{4.2}$$

The forward focusing is relevant to reduce the backscattering probability, see Sec. 6.1.4. For the PERKEO III experiment setup, [Wan13; Mes11; Sau18] discuss the magnetic field and related effects, especially for neutron spin-dependent asymmetry measurements where the deviation from symmetry of the magnetic field is a dominant systematic effect.

The experiment is situated at the PF1B instrument site at the ILL. The ILL reactor produces thermal neutrons (average velocity $v \approx 2200$ m/s), which are cooled to cold neutrons ($v \approx 800$ m/s) for the PF1B with a liquid deuterium moderator. The PF1B offers the strongest source of cold neutrons for particle physics in the world [Abe+06]. The high rate of neutrons enables using a pulsed beam. With a pulsed beam, we measure not only the background of the experiment itself but also from surrounding instruments between each neutron pulse. The reduction in measured events and increased statistical uncertainty is

a highly favorable trade-off, as it greatly reduces the systematic error from less accurate background measurements [Mär+09]. We can illustrate the principles of a pulsed beam measurement by plotting the number of detected events (event rate) against the time of a detected event relative to the last pulse (neutron time-of-flight)for a 5 min measurement, as seen in Fig. 4.4. For cold neutrons, the time-of-flight of electrons from the decay position to the detectors is negligible. The event rate corresponds to the neutron pulse propagating through the experiment. The event rate is the background rate when the neutron pulse enters the spectrometer. Electrons from decays will be detected when the neutron pulse reaches the central decay volume, increasing the event rate (at 3 ms). The event rate is approximately constant and plateaus (at $4-6$ ms) as the number of neutrons is not significantly reduced by decays. After the neutron pulse exits the central decay volume, the beam stop fully absorbs the neutron pulse, leading to a high event rate of produced gamma radiation (at $8-10$ ms). We can choose a neutron time-of-flight signal and background window to discriminate the measured data and create background-free spectra through subtraction. The validity of choice of the neutron time-of-flight windows is discussed in Sec. 5.3.2.

In Fig. 4.4, we can also differentiate between the event rates of the upstream detector (Det0) and the downstream detector (Det1). Due to the magnetic mirror effect, the event rate is higher for the upstream detector before the pulse reaches the maximum of the magnetic field in the central decay volume. Also, the background for the upstream detector is dominated by the constant background from the beamline, whereas the downstream detector has a smaller constant background but is dominated by the gamma radiation from the beam absorption.

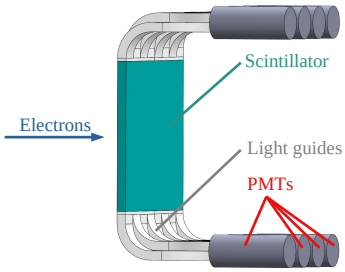
Shielding between the background sources and the detectors is necessary to minimize the impact of constant, instantaneous, and delayed background events. To this end, we placed shielding outside the vacuum vessel around the beamline and beam stop, which is already present in Fig. 4.4. We used lead bricks and specialized materials containing boron for shielding. The resulting signal-to-noise ratio is presented in Sec. 5.4.

The author of [Sau18] designed the beam stop that is made from $^{10}B^4C$ with a neutron backscattering suppressing structure in front. The chosen materials produce more instantaneous gamma emissions but less delayed neutron emissions. The latter would be harder to separate from the pulsed beam with the neutron time-of-flight windows, as shown in [Mes11].

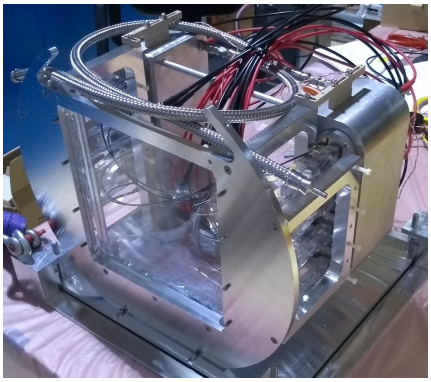
Difference to Previous PERKEO III Measurements

This measurement is the first application of PERKEO III with a setup optimized for purely measuring the electron energy spectrum. It is the first measurement of PERKEO III using an unpolarized beam, as we are not measuring a neutron spin-dependent quantity. This simplification increases the experiment's neutron flux and decreases the measurements' statistical uncertainties, see Sec. 4.2.4.

Furthermore, new experiment components were explicitly made for this measurement campaign: New detectors for a better energy resolution, a new robot to move the radioactive calibration sources, and a new cooling system for a more stable measurement and to enable higher magnetic fields. A picture of PERKEO III from the 2019/20 measurement campaign with magnetic shielding and connected supplies at the PF1B instrument site at the ILL is shown in Fig. 4.3, illustrating the size of the experiment.



(a) Re-used graph from Sec. 3.3.1: The PERKEO III scintillation detector schematic. Electrons deposit their energy in the scintillator and produced photons are detected by the PMTs.



(b) One of two assembled PERKEO III detectors in 2019 with the scintillator at the front, light guides at the sides, and PMTs in the cooling reservoir.

Figure 4.5 The PERKEO III 2019/20 detector design. Compared to previous designs, this design uses a thicker scintillator and light guides for improved uniformity and light yield. Also, the PMTs are mounted in a greater cooling reservoir, and the scintillator is read out at the sides.

4.2.1 Detector Design

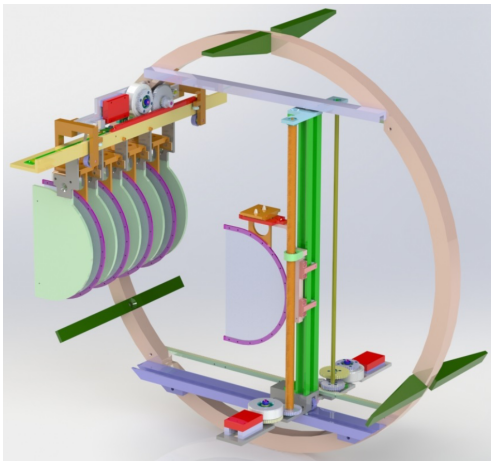
The PERKEO III detectors work on the principles described in Sec. 3.3.1: The detector comprises a plastic scintillator with acrylic light guides at the sides, connected to photomultiplier tubes (PMTs) at the end of each light guide, as shown in Fig. 4.5. Plastic scintillators are suited for this measurement, as they have a short response time, are less sensitive to gamma background, and have a lower backscattering probability than Si-based detectors. When an electron with kinetic energy E_e hits the detector, it deposits its energy, and n_γ photons are produced approximately proportional to the deposited electron energy E_e. A fraction of the photons travel through the detector to the PMTs, creating a charge pulse A through the photoelectric effect and multi-stage amplification. The fraction of photons reaching the detector for a fixed electron energy, the light yield, essentially defines the energy resolution of the detector. A more detailed and accurate description of the detector model is given in Sec. 6.1. The PMTs are fine-mesh type PMTs that can be operated in high magnetic fields. We chose their photoelectric absorption range to match the scintillator's emissions. The produced PMT charge pulses are approximately 20 ns wide.

The measured charge A is on an arbitrary scale, and contrary to the simplification in Sec. 3.3.1, we cannot assume the energy reconstruction to be entirely linear

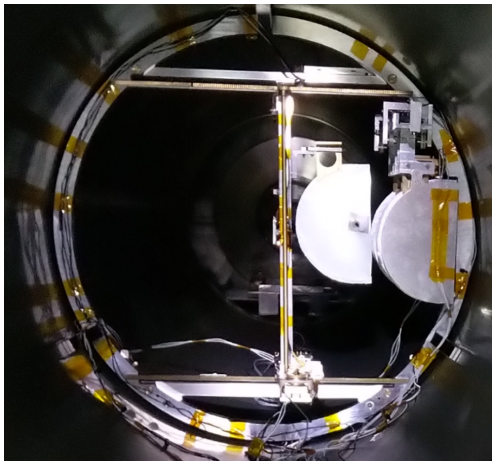
$$A = f_{\mathrm{Det}}(E_e) \tag{4.3}$$

with detector response function f_{Det}. To determine f_{Det}, we use calibrations sources, see Sec. 4.2.2. I discuss different non-linear effects and the energy reconstruction in Sec. 6.1.

The detector design considers the geometrical limitations of the experiment, e.g., a minimum scintillator size and thickness to detect all electron energies and positions, but also desired performance, e.g., uniformity of response and energy resolution. The author of [Ber18] did the core design of the detectors, designing new light-guide shapes for increased light yield, a photon readout from the sides, and a cooling reservoir for added temperature stability. The author of [Ant19] and I tested different scintillator designs for improved uniformity and light yield and tested light guide components for their transmission properties. These tests lead to the choice of using a thicker scintillator for the measurement campaign. Based on these



(a) Schematic from [Kro20]: Render of the CaliBot with calibration source holders. In this schematic, the CaliBot source holder picks-up a source from the storage arm and moves it to the center. The remaining sources are in a movable storage arm.



(b) CaliBot in the central decay volume of PERKEO III with a radioactive calibration source on its support foil and holder. The "hole" in the Mylar support foil is the ultra-thin C foil with the radioactive calibration source.

Figure 4.6 The PERKEO III 2019/20 *CaliBot* (Calibration Robot) to move and store the radiactive calibration sources. Compared to previous PERKEO III measurements, the chosen design only uses one holder arm to move the sources all in the same transversal plane to the beam axis and utilized more precise rotational encoders.

tests, we polished the light guides again for matching transmission properties and assembled the light guides from individual parts in a clean room. To enable the deployment of the detectors, I re-designed the support structure for the scintillator, light guides, and PMTs. The final detector comprises a $240\times240\times15$ mm^3 BC-408 plastic scintillator [Sai21] and acrylic light guides with the design from [Ber18]. Also, the detector is temperature stabilized using an external and separate water cooling circuit.

4.2.2 Calibration Sources and Robot

Generally, the detector response f_{Det} in Equ. (4.3) is non-linear due to multiple effects, which are discussed in Sec. 5 and 6, and must be measured to precisely measure the electron energy spectrum. To measure f_{Det}, we need calibration sources with known energies for references. These calibration measurements are essential for the overall energy-channel relation, its change over time due to temperature-induced gain fluctuations, see Sec. 5.2.4, and its dependence on the electron position on the detector, see Sec. 3.3.2.

We use five electron conversion sources that provide known peaks and auger electron emissions ranging from a few keV up to 1 MeV, exceeding the electron endpoint energy from the beta decay of $E_{\beta^-} \approx 782$ keV. Besides electrons, the sources also emit gamma radiation, whose related systematic effects were negligible in past PERKEO III experiments. Tab. 4.1 lists the sources with the approximate energies of the dominant energy peaks. The energies are only approximate, as each peak consists of multiple combinations of different transitions, see Sec. 7.1. The calibration sources are placed as a concentrated solution on ultra-thin carbon foils (12 and 30 μg/cm^2 area density, 10 mm diameter), supported by a thin Aluminum-coated Mylar foil (25 μm thickness, ca. 140 mm diameter) to minimize electron energy losses affecting the calibration measurement. The remaining energy losses from electrons interacting with the foils were negligible in past experiments [Sau18; Roi18].

Table 4.1 The calibration sources used for the PERKEO III 2019/20 measurement campaign with approximate energies of the dominant energy peaks, decay modes, half-life, and event rate. EC refers to electron capture transitions. Energy and half-life values taken from [Bé+16]. Event rate subtracted with background event rate of $\approx 530\ s^{-1}$.

Source Isotope	Decay Mode	Half-life	Energies [keV]	Event Rate [s^{-1}]
^{109}Cd	EC + Auger	461.9 d	75	4700
^{139}Ce	EC + Auger	137.6 d	127	2700
^{207}Bi	EC + Auger	32.9 y	503, 995	4000
^{113}Sn	EC + Auger	115.1 d	369	2000
^{137}Cs	EC + Auger + β^-	30.0 y	630	7200

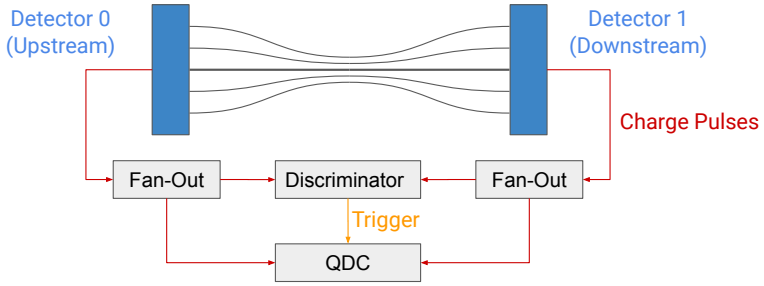


Figure 4.7 Schematic of the readout electronics and coincidence measurement setup that significantly reduces the detection of background events. The signals from each detector are fed into linear fan-out units to split the signal into two separate signal chains each. One half goes into a discriminator and logic unit that provides a trigger signal for the charge integrating analog-to-digital converter (QDCs). We feed the second half of the signals through delay lines (ca. 40 m), to allow for the processing time for the discriminator and the logic unit, and into the QDCs to be measured.

To switch between the measurement of the electron energy spectrum from neutron decay and the calibration sources, we need a robot to move the calibration sources from an idle position to positions covering the neutron beam distribution in the central decay volume - the *CaliBot*. This task is challenging, as the calibration sources are on fragile foils in a vacuum and a magnetic field. The vacuum is critical for overheating motor components and limits the choice of materials due to their outgassing. The magnetic field must stay uniform in the central decay volume, and any added parts must be sufficiently non-magnetic. The design and deployment of the CaliBot was the core topic of [Kro20], in which I was involved. A schematic of the CaliBot and a picture of it in PERKEO III are shown in Fig. 4.6. The CaliBot is located 50 cm out of center downstream in the neutron beam direction in the central decay volume.

With the CaliBot and while switching off the neutron beam, we can move the calibration sources to the center of the neutron beam distribution or move the sources on a 2D grid in the plane perpendicular to the beam axis to measure the spatial detector response. During the measurement campaign, we measured the detector calibration with all sources twice a day, every hour with ^{113}Sn for gain fluctuations and a daily 2D grid scan with ^{113}Sn for the spatial detector response.

4.2.3 Readout Electronics and Data Structure

We adapt the electronics from [Sau18] and [Roi18] to readout and digitize the electronic signals from the detectors. The measurement follows the coincidence measurement principle

and is illustrated in Fig. 4.7: The signals from all sixteen PMTs, eight per detector, are fed into linear fan-out units to split each signal into two separate signal chains. One half goes into a constant fraction discriminator unit with a timing resolution of 0.8 ns. It provides a start trigger signal for each PMT input and connects to a logic unit. If at least two PMTs from one detector trigger, the other half of the signals for all sixteen PMTs is registered in charge integrating analog-to-digital converter (QDCs). We feed the second half of the signals through delay lines (ca. 40 m) to allow for the processing time for the discriminator and the logic unit. This coincidence measurement setup significantly reduces the detection of background events and relies on sufficient photons being produced in the scintillators to trigger multiple PMTs. Measuring all PMTs on each trigger enables full energy reconstruction, where energies are split between PMTs or detectors to measure below the trigger threshold, e.g., backscattering or low-energy events close to the trigger threshold.

The QDCs sample with 100 MHz, i.e., every 10 ns, and integrate for 330 ns. The integration time needs to include the electronics induced own-signal and enough time to register backscattered electrons to the opposite detector (electron time-of-flight about $40 - 100$ ns). I discuss the effect of the electronics-induced own-signal, dependency on the last event, and QDC non-linearity in Sec. 5.2 and 6.1.7. After each trigger, the electronics system does not accept new signals for approximately 1.5 μs, a so-called non-paralyzable dead time. I discuss the dead time effects in Sec. 5.3.1. The electronics are stored in temperature-controlled cabinets.

The *Pudel* data acquisition software [Roi18] controls and monitors the readout electronics and incorporates all other automated experiment components, e.g., the supply voltage of the detectors. Each measurement is stored in ROOT[1] files and organized in cycles of 10 s duration. Each cycle stores global information, e.g., beamline instrument parameters or cycle validity. Each event within the cycles stores event-specific information, e.g., the trigger times for both detectors or the QDC values for all PMT channels.

Readout Modes

We can take data in different modes to trade information for storage efficiency. We can store each sampled value of the QDC, i.e., a value every 10 ns of the 330 ns integration time, only the first and the last QDC value, or the difference of the first and the last QDC value for each PMT channel. We label these modes "ALLMODE," "BOTHMODE," and "DELTAMODE," respectively. Only storing the difference gives the total measured integral of the QDC for each PMT for that event, as it subtracts the electronics-induced offset and leaves the integrated charge. However, storing all QDC samples enables analysis, e.g., of the non-linearity of the QDCs. For different event types and a few PMTs, I present "ALLMODE" data in Fig. 4.9. We took data in all three modes during the measurement campaign, but only a few hours with different sources in "ALLMODE" for the analysis of systematic effects.

Event Types

With the two-detector setup, a long integration time, and a trigger condition for each detector, we can differentiate events and analyze them accordingly. There are mainly four types of events: direct events, backscattering, correlated (quasi-instantaneous) events, and accidental

[1] https://root.cern/

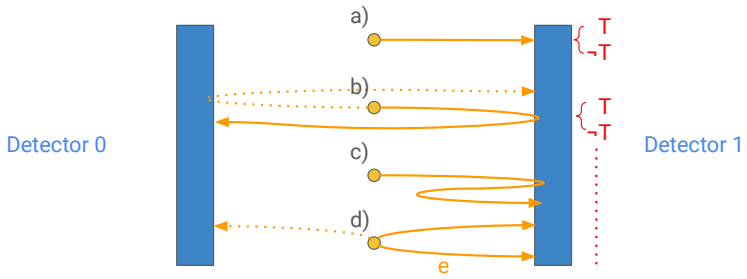


Figure 4.8 Different event types for detector 1 with a two-detector setup, a long integration time, and a trigger condition for each detector. Each detector interaction can lead to a trigger or not (T or ¬T) on that detector. Alternative scenarios are indicated with dotted lines. a) Direct event and electron deposits full energy in one detector. b) Backscattering without reflection on the magnetic field gradient. Detector 1 could be the primary or secondary detector. c) Backscattering with reflection on the magnetic field gradient. d) Multi-electron events from correlated electron emissions or accidental coincidentals.

coincidence events. I illustrate the different events and their possible trigger combinations in Fig. 4.8. The differentiation is essential to measure low-energy events, as the trigger is generally energy dependent. The energy dependence of the trigger, the trigger function, is discussed in Sec. 6.1.3.

Some of the trigger combinations in Fig. 4.8 pose a challenge for precise energy measurements. We can only differentiate backscattered events in the data by looking for events where both detectors have triggered. Correlated and accidental coincidence events can be removed by matching the time between the two detector triggers with the electron time-of-flight. However, for example, an electron could backscatter but only deposit a small fraction of its energy in the primary detector and not trigger the detector. The backscattered electron would be detected later on the same (mirror effect) or opposite detector, but the fraction of energy deposited on the first detector would be lost and undetected as there was no trigger. This effect is called undetected backscattering and is discussed in Sec. 6.1.4.

Multi-electron events, like correlated and accidental coincidence events, can create other issues. Correlated events are unavoidable in some cases, like with the calibration sources due to short-lived atomic excitation states. If two or more electrons reach the detectors within the time window of integration, their combined energy is detected. If their energies are sufficiently large, they are harder to distinguish from other events using electron time-of-flight information. This issue increases complexity when including trigger conditions and is discussed in Sec. 6.1. These issues also highlight the necessity of a two-detector setup for a high-precision measurement with PERKEO III and are a topic for future experiments with, e.g., the PERC spectrometer [Wan+19] that circumvent this necessity with a strong magnetic filter.

To illustrate different types of events, I present "ALLMODE" data examples in Fig. 4.9. For visualization, only two QDC channels are plotted for each detector, each corresponding to the readout of a single PMT. The variation in the first charge integration value is an electronics-induced and QDC channel-specific offset, the so-called pedestal, see Sec. 5.2.1. In the case of direct events, the electron fully deposits its energy in one detector, and the produced photons are distributed onto the PMTs of that detector, leading to a full charge integration of the pulses at the readout. If an electron backscatters, it can propagate to the opposite detector or be mirrored onto the same detector to deposit the remaining energy.

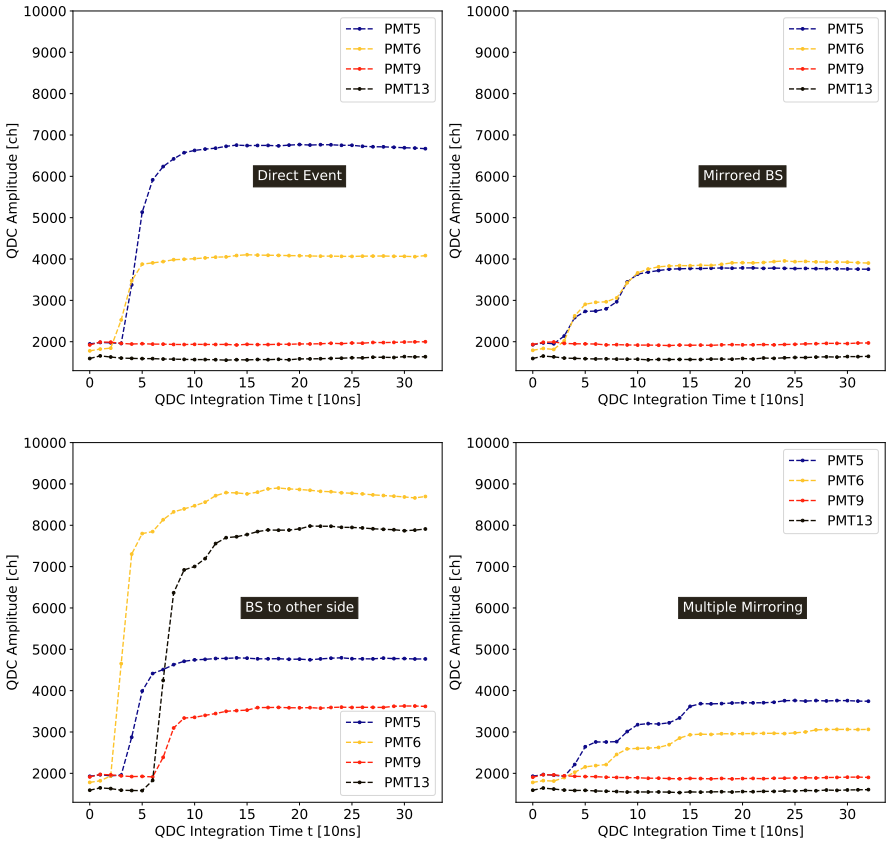


Figure 4.9 The charge integration with ALLMODE data for different event types, see Fig. 4.8. PMTs 5 and 6 are from detector 0 (upstream) and PMTs 9 and 13 are from detector 1 (downstream). (Top, left) Direct events: Energy is fully deposited in one detector. (Top, right) Backscattering with magnetic mirror effect: Energy is fully deposited in one detector but in two charge integration steps with the typical 40 – 50 ns electron time-of-flight width. (Bottom, left) Backscattering without magnetic mirror effect: Energy is split between the primary and secondary detector. The incoming charge pulses are the typical 40 – 50 ns electron time-of-flight apart between detectors. (Bottom, right) Higher-order backscattering: Energy is fully deposited in one detector but in three charge integration steps with the typical 40 – 50 ns electron time-of-flight width. The electron backscatters twice and is reflected each time by the magnetic mirror effect.

In both cases, the charge integration has two "steps", with a typical width of 40 – 50 ns electron time-of-flight, see also [Ber24]. The electron time-of-flight is similar for both cases, as most mirrored electrons are reflected close to the maximum of the magnetic field in the center of the central decay volume. Using the electron time-of-flight signature, we can also distinguish events with higher-order backscattering and mirroring, where the charge integration has multiple steps shifted by 40 – 50 ns electron time-of-flight. Multiple-electron events are not included in Fig. 4.9, as they are less straightforward to visually distinguish them from other events in a charge integration plot. They arrive within times of about our time resolution of 10 ns or shorter times at the detectors and we need to model their impact on the analysis with simulations, see Sec. 6.

4.2.4 Beamline Setup

We need a beamline to make the cold neutron beam match the experiment dimensions and allow for a pulsed beam measurement. PERKEO III is installed at the PF1B instrument site and connected to the ILL reactor through the neutron guide H113 [Abe+06]. The guide delivers a neutron capture flux ϕ_c at the exit of the guide of

$$\phi_c = 1.35 \cdot 10^{10}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}.$$

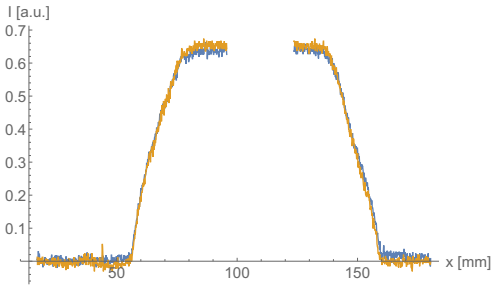
and a typical off-axis FWHM beam divergence of ≥ 7 mrad [Abe+06].

For the pulsed measurement, we need to reduce the velocity spread of the neutrons, as the neutron pulse would otherwise also spread over time and disturb the background measurements, see Fig. 4.4. To this end, we use a neutron velocity selector comprising a boron-coated turbine spinning at 425 Hz. It reduces the velocity distribution of the neutrons to a nearly symmetric triangular distribution with a mean of 5 Å and a 12% width.

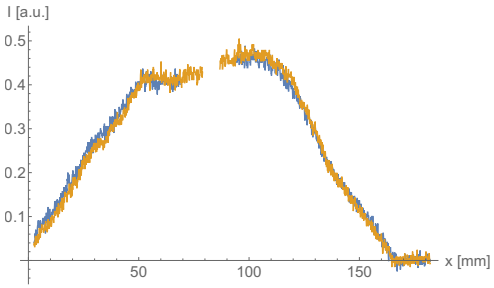
Deviating from past measurements that focus on neutron spin-dependent quantities, we did not use a polarizer for a polarized neutron beam. Measuring with an unpolarized beam increases the neutron flux in the experiment by a factor of four [Sau18]. We checked the polarization with ^{3}He cells and determined it to be sufficiently low with $P \leq 10^{-4}$, as the neutron guide might slightly polarize the neutron beam [Sol20]. Past measurements also combined the polarizer with a spin flipper, enabling a mirrored experiment setup from the neutron perspective. The spin flipper uses a rotating magnetic field in the neutron rest frame to flip the neutron spin every few seconds, greatly reducing systematic uncertainties of spin-dependent quantities [Mär+19; Sau+20]. However, we optimized our measurement for a pure electron energy spectrum measurement and did not use a polarizer or spin flipper.

We use five absorption apertures to shape and collimate the neutron beam to the experiment dimensions. Their positions determine the beam distribution and divergence in the experiment. Also, they affect the magnitude of the background close to the upstream detector. The resulting neutron beam distribution can be calculated, neglecting time dependencies, through geometric integrals [Wan13; Sau18]. I present an example of how to optimize the aperture positions in Sec. 3.3.4. For this measurement campaign, I tested different aperture setups with a custom simulation to also consider the electron distribution on the detectors. The apertures are made of ^{6}LiF, have a 6x6 cm^2 opening, and are positioned over the 4.3 m long beamline in distances of 70 – 100 cm.

Finally, we need to pulse the beam at the experiment entry. We used the pre-existing disk chopper from [Wer09] comprising ^{6}LiF plates with a 22.11 deg opening and a frequency of 83 Hz. The chopper frequency must match the neutron pulse velocity, i.e., the time



(a) Horizontal neutron beam distribution at the start of the central decay volume.



(b) Horizontal neutron beam distribution at the end of the central decay volume.

Figure 4.10 Horizontal neutron beam profile measurement with irradiated medical imaging foils in the plane perpendicular to the beam axis in PERKEO III for two different heights (blue, orange) at two different positions in the central decay volume. The horizontal non-linear divergence of the neutron beam is caused by different angular neutron beam components from the curved neutron guide.

that the neutron beam propagates through the experiment, and also guarantee the same background environment for the signal and background neutron time-of-flight windows in Fig. 4.4 since the chopper-produced background might vary with its rotation angle. I test these requirements in Sec. 5.3.2.

To verify the neutron beam distribution in the experiment, we placed two medical radiation testing foils[2] in the central decay volume. Due to the high neutron flux at the PF1B instrument site, we can darken these foils with irradiation damage. We use the magnitude and location of irradiation damage on these foils to calculate the neutron beam profile. Using high-resolution scanners and correcting the image for saturation effects, we can calculate the neutron flux in the plane perpendicular to the beam axis at two locations: 7 cm into the central decay volume and 6 cm before its end. The resulting horizontal distributions are shown in Fig. 4.10. Due to the curvature of the neutron guide H113 [Abe+06], the neutron beam comprises different angular components leading to the horizontal spread in Fig. 4.10. In comparison, the vertical spread diverges linearly. The full-width neutron beam cross-sections in the experiment are (horizontal × vertical, background to background)

$$\begin{aligned} &10.59\ (15) \times 10.41\ (10)\ \mathrm{cm}^2 \quad \text{(start)}, \\ &16.32\ (37) \times 16.11\ (22)\ \mathrm{cm}^2 \quad \text{(end)}. \end{aligned}$$

Using gold foil activation by the neutron beam in the center of the central decay volume, we determined the time-averaged neutron capture flux ϕ_c to be

$$\phi_c = 6.63(14) \cdot 10^6\ \mathrm{cm}^{-2}\mathrm{s}^{-1},$$

agreeing with the observed number of detected electrons from neutron beta decays.

[2] GAFCHROMIC EBT2, self-developing film for radiotherapy dosimetry

5 Creating Spectra From Measured Data

To precisely measure the electron energy spectrum from neutron beta decay, we must process the measured data, apply corrections where necessary, and finally create these spectra with quantified uncertainties to enable analyses of physical quantities. Non-energy spectra and specific analyses to characterize systematic effects also require a similar process. This chapter presents the data selection and application of corrections for spectra generation from the measured data sets and discusses the induced systematic uncertainties of these corrections.

The data reduction and processing is automated with a new data analysis framework, *Panter*, developed within this thesis. To create spectra, we must distinguish between single event corrections, e.g., rate dependency effects, and spectral corrections, e.g., background subtraction. Besides neutron beam data, i.e., measurements of electrons from neutron beta decay, we also collected calibration data sets with electron conversion sources, see Sec. 4, background measurements, and characterization measurements by varying hardware parameters. I explain the data sets used for each analysis. Much of this chapter's analysis and corrections build upon previous work in the PERKEO III collaboration [Sau18; Roi18]. These corrections have been improved upon by considering more systematic effects and developing better methodology.

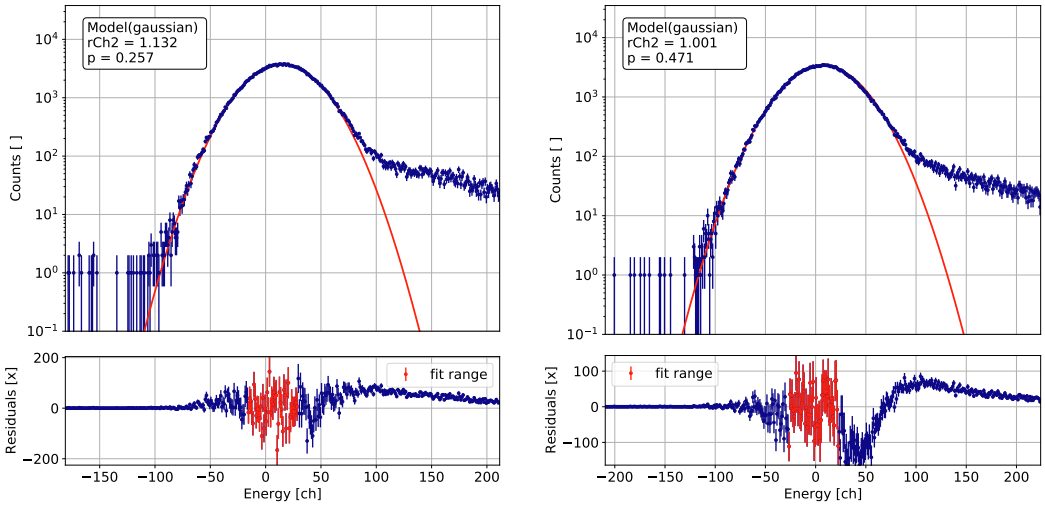
5.1 *Panter* Analysis Framework

Due to the custom data structures and required analysis for this high-precision measurement with PERKEO III, we need a data reduction and processing framework to enable the data analysis. To this end, I developed a python package called *Panter* (Perkeo ANalysis Tool for Evaluation and Reduction) for reducing, evaluating, and analyzing PERKEO III data with all necessary tools in an automated manner. It is designed as a user-friendly option for older PERKEO III data sets with similar structure, any analysis with the data set taken in my thesis, and any future PERKEO III data sets.

Panter has a general toolset for data analysis with custom classes, can process all PERKEO III data types and measurements with modular corrections, and automates PERKEO III specific analysis. Also, *Panter* has a broad set of unit tests to ensure accurate results using benchmarking and synthetic data sets, of which some were developed in collaboration with the author of [Bes22]. Unless otherwise indicated, all data analysis results are done with *Panter*.

5.2 Single Event Corrections

We must first reconstruct the electron energies from measured events to create electron energy spectra. To do so, we must account for systematic effects caused by our measurement setup on single-event detection. Specifically, I look at electronics-induced effects during readout, temperature-induced gain fluctuations, and detector fine-tuning. In previous PERKEO III



(a) Pedestal distribution for PMT2 on detector 0.

(b) Pedestal distribution for PMT6 on detector 0.

Figure 5.1 Determination of the pedestal position P_{ij} and width w_{ij} by looking at events where only the other detector triggered. The asymmetric fit range is necessary to reduce the bias for P_{ij} and w_{ij} towards larger values due to the long tails caused by undetected backscattering events visible in the log plots.

analyses, single-event corrections were unnecessary [Sau18]. For our analysis, most but not all systematic effects and corrections are still included in the theoretical description of the detector response model and not treated as single-event corrections, see Sec. 6.1. We reconstruct the electron energies of an event by summing up the integrated charges of all PMT readouts and I state the energy reconstruction formulae at the end of this section.

5.2.1 Electronics Induced Effects

Pedestal Correction

The electronics-induced and QDC channel-specific offset is called pedestal, see Sec. 4.2.3 for a description of the electronic readout system. Since the electronic readout system collects QDC samples before and after the integration of PMT pulses, we can subtract the first QDC sample of the integration time, the pedestal, from the last QDC sample to obtain the fully integrated charge of an event. However, the pedestal fluctuates, and subtracting the QDC sample values still leads to a residual and QDC channel-specific offset that we must correct for. For simplification, we label each QDC channel with the corresponding PMT channel.

We can determine this electronics-induced signal for a QDC channel from measured data by studying events without input on the respective QDC channel. Ideally, we want events with no signal on that QDC channel. Practically, we do this by looking at events where only the other detector is triggered and extract the pedestal value P_{ij} and width w_{ij} for each of the eight QDC channels j and each of the two detectors i. The resulting distribution for two QDC channels is shown in Fig. 5.1. As discussed in Sec. 4.2.3, backscattering events that fail to trigger a detector can lead to undetected energy losses [Roi+19]. Such events also contaminate the pedestal distribution by creating long tails in Fig. 5.1 and bias the extraction to larger values. This contamination was neglected in previous PERKEO III analysis [Sau18]. The pedestal needs to be determined at least on a measurement file-by-file

basis for each channel, and we lack further information to reduce the contamination with undetected backscattering events post-measurement. Studies to reduce the contaminating undetected backscattering with energy cuts on the other, triggered detector yields unreliable results.

Hence, we extract the values of P_{ij} and w_{ij} with iterated fits of Gaussian distributions. The first fit determines the fit ranges of the second one with an adaptive fit range as $[\mu - 1.1\sigma, \mu + 0.5\sigma]$ with μ and σ being the mean and with of the first Gaussian fit. I determined the asymmetric fit range by dividing the pedestal distribution summed over all PMTs for one detector with an upper fit range limit by the total pedestal spectrum. The upper fit range limit was set to minimize the amount of backscattering events while minimizing the cut into the actual pedestal distribution. The resulting second fit with a Gaussian distribution is shown in Fig. 5.1. With this approach, we can determine the values of P_{ij} and w_{ij} for each file and correct the integrated QDC values.

We can use the individual pedestal widths w_{ij} to calculate the total width of the integrated charge of an event as

$$w_{\text{tot}} = \sqrt{w_{\text{det0}}^2 + w_{\text{det1}}^2} = \sqrt{\sum_{j=1}^{8} w_{1j}^2 + \sum_{j=1}^{8} w_{2j}^2}.$$

We use w_{tot} to measure the electronics-induced fluctuation on the integrated charge of an event.

Using neutron beam data, we can estimate the individual QDC channel uncertainty $\sigma(P_{ij})$ for P_{ij} as $\sigma(P_{ij}) = 0.14(1)$ ch, leading to an uncertainty on the sum of $\sigma(P) = 0.56(4)$ ch. The individual QDC channel values P_{ij} are in a range of about ± 100 ch. Using a calibration measurement with all five calibration sources as verification, we can estimate the uncertainty for one detector as $\sigma(P_i) = 0.42(14)$ ch, leading to an uncertainty on the sum of $\sigma(P) = 0.59(20)$ ch. Thus, the calibration source results agree with the neutron beam result for the uncertainty of the pedestal correction.[1] This agreement also implies that we get the same systematic uncertainty from individual QDC channel corrections as from correcting the sum of all channels for one detector, as is done for the calculation with the calibration sources. The final systematic uncertainty on the measured amplitude offset is given by the neutron beam result $\sigma(P) = 0.56(4)$ ch. This correction is non-negligible for a precise measurement of the Fierz interference term b and I discuss its significance in Sec. 8.

To test a potential induced bias from the asymmetric fit range, I generated 500 synthetic data sets from an ideal Gaussian distribution with matching parameters to the pedestals, equal histogram parameters and equivalent event counts. Extracting the peak positions of these data sets with symmetric and asymmetric fits induced a mean shift of $\Delta(\mu) = 1.6 \cdot 10^{-3}$ ch and is negligible.

The pedestal width w_{ij} for individual PMT channels from neutron data is $w_{ij} = 29.25(22)$ ch and the total width is $w_{\text{tot}} = 117.20(88)$ ch. Using this result, I can estimate the significance of this uncertainty by varying w_{tot} when extracting the Fierz interference term b by fitting a neutron beta spectrum. The systematic uncertainty $\sigma_{\text{sys}}(b)$ from the uncertainty of the total width of the integrated charge of an event is $\sigma_{\text{sys}}(b) \leq 10^{-5}$ for all tested fit ranges and is thus negligible.

[1] The pedestal value is rate dependent and the pedestal values are different for each calibration source and the neutron beam data. The rate dependency correction is discussed in Sec. 5.2.1.

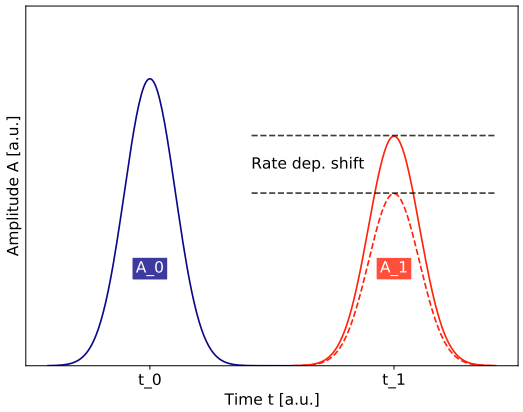


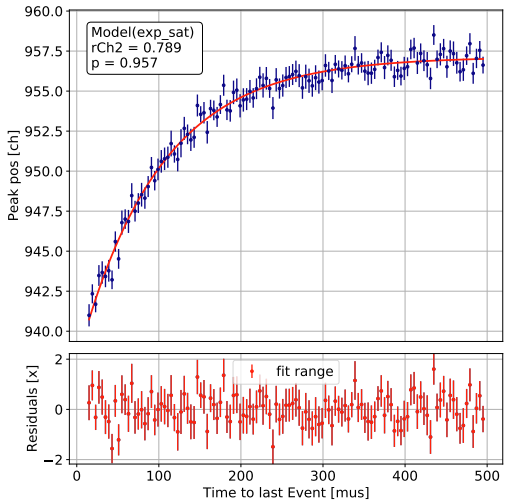
Figure 5.2 Schematic of variables to parametrize the rate dependency effect. Depending on the amplitude A_0 of and time $t_1 - t_0$ to the last event, the measured amplitude A_1 of an event is decreased.

For future measurements with a similar setup, the systematic uncertainty from the pedestal correction is limited by the asymmetric fit range, and we could reduce it by removing contaminating events, e.g., by shielding the detector from electrons for a pedestal measurement. Alternatively, using studies of undetected backscattering, future work could derive a fit model that includes the contamination to determine the pedestal position more precisely without limiting the fit range.

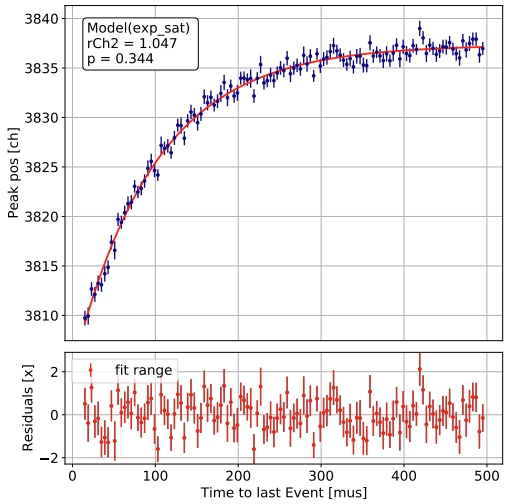
Rate Dependency Correction

The rate dependency effect is illustrated in the schematic in Fig. 5.2, and we must correct it to measure energy spectra precisely. An event's measured amplitude A_1 at time t_1 is shifted depending on the time $t = t_1 - t_0$ to the last event and its amplitude A_0. This effect is already analyzed in a previous PERKEO III result [Sau18] with a similar electronic readout system. However, previous work found the systematic effect of the rate dependency on the spin-dependent measurement of the beta asymmetry to be minor, and only a total correction was applied. In this thesis, I correct the rate dependency effect for each event and analyze the induced systematic uncertainty for greater precision.

To measure the systematic effect, we separated the detectors from the readout electronics and used two coupled pulse generators with slightly de-tuned frequencies and Gaussian pulses with different amplitudes as inputs. The de-tuned frequencies lead to a uniform distribution of times between events t. The results for two sets of amplitudes for a QDC channel are shown in Fig. 5.3. We filter the measured data for intervals of times to last events and extract a measured amplitude with a fit of a Gaussian distribution. For this analysis, we must also correct the pedestal first, see Sec. 5.2.1, as the pedestal is also affected by the rate dependency. For the measurements in Fig. 5.3, not all QDC channels were tested simultaneously. We used this to estimate the pedestal of QDC channels from measurements where others were tested. To reduce correlations between parameters, we determine the width of the Gaussian distributions for large times between events and fix it for the analysis.



(a) Testing the QDC channel of PMT0 with 50 and 300 mV pulses. Showing the rate dependency on the measured 50 mV pulse.



(b) Testing the QDC channel of PMT0 with 200 and 500 mV pulses. Showing the rate dependency on the measured 200 mV pulse.

Figure 5.3 Parametrizing the rate dependency effect by fitting the model in Equ. (5.1) to offline tests with pulse generators. The pulse generators have de-tuned frequencies to cover all possible times between events. The model describes the rate dependency effect well within statistical limits.

We can parametrize the effect in Fig. 5.3 as a function of the times t between successive events with an exponential saturation with a time scale k, unperturbed amplitude c_0 and rate dependency correction c as

$$f(t) = c_0 - c \cdot e^{-t/k}, \tag{5.1}$$

for different amplitudes. We hypothesize a linear relation between c and the current measured amplitude A_1, i.e.,

$$\eta = c/A_1 = \text{const.}, \tag{5.2}$$

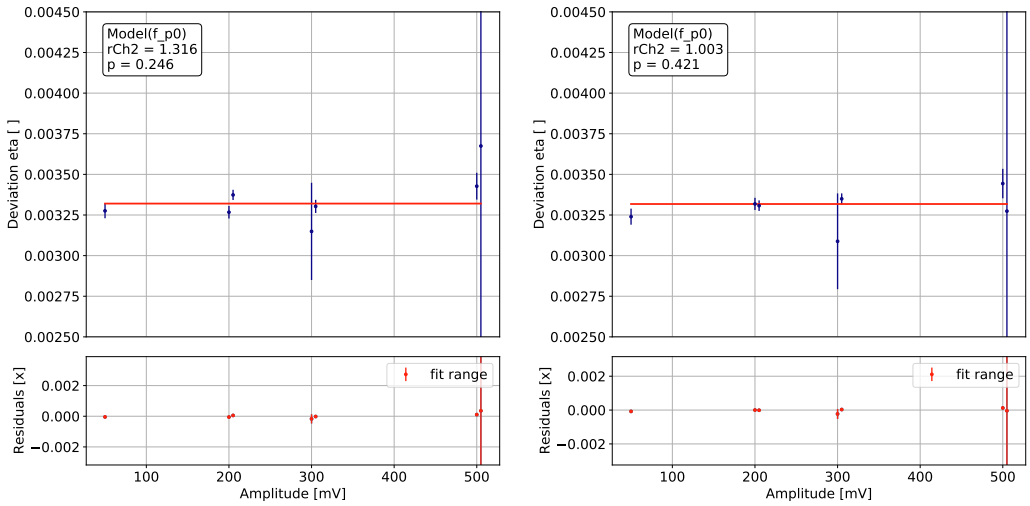
and that the effect has sufficient Markov properties, i.e., we only need to consider the last event instead of accumulating the contributions of all previous events. This hypothesis sufficed in previous analysis [Sau18].

We fit the rate dependency model in Equ. (5.1) to the data from Fig. 5.3. To further reduce the correlation between parameters, we determine an average time scale $k = 99$ μs for all QDC channels and determine c_0 and c from the fits. We use these results to check our hypothesis from Equ. (5.2) in Fig. 5.4. The hypothesis holds within statistical uncertainty for all QDC channels.

We can derive an additive rate dependency correction R for one event and for one QDC channel from our hypothesis as

$$R(t) = A_1 + A_0 \cdot \eta \cdot \exp^{-t/k}. \tag{5.3}$$

The resulting parameter values are shown in Tab. 5.1. Using the uncertainties in Tab. 5.1, I can estimate the significance of this correction by varying all η_{ij} in the spectrum generation and test how it affects the extraction of the Fierz interference term b when fitting a neutron beta spectrum. An estimated upper limit for the combined systematic uncertainty $\sigma_{\text{sys}}(b)$



(a) Results for QDC channel of PMT0.

(b) Results for QDC channel of PMT1.

Figure 5.4 Testing the rate dependency correction hypothesis from Equ. (5.2) that the rate dependency correction c is linear with the current measured amplitude A_1, making the deviation η constant. The hypothesis holds within statistical uncertainty for all QDC channels. The measured amplitude pairs are combinations of 50, 200, 300, and 500 mV.

Table 5.1 Resulting rate dependency correction parameters η_{ij} for detector i and QDC channel of PMT j. The combined systematic uncertainty on the Fierz interference term b of all η_{ij} is $\sigma_{\text{sys}}(b) \leq 10^{-4}$ and thus negligible.

	η_{0j} $[10^{-3}]$	η_{1j} $[10^{-3}]$
PMT0 | PMT8	3.333 ± 0.017	3.100 ± 0.018
PMT1 | PMT9	3.310 ± 0.017	3.581 ± 0.018
PMT2 | PMT10	3.543 ± 0.017	2.769 ± 0.017
PMT3 | PMT11	3.882 ± 0.016	2.884 ± 0.017
PMT4 | PMT12	3.204 ± 0.020	3.500 ± 0.017
PMT5 | PMT13	3.305 ± 0.020	2.352 ± 0.026
PMT6 | PMT14	2.877 ± 0.020	3.325 ± 0.019
PMT7 | PMT15	3.507 ± 0.018	3.469 ± 0.019

Table 5.2 Testing the rate dependency corrections on calibration sources, ordered by energy of the fitted peak. Comparing the different sources is not straightforward, as they have different rates and other energy peaks in their spectrum. Both effects lead to a different distribution of previous amplitudes A_0 and times to last events $t_1 - t_0$. Still, the rate dependency correction greatly reduces the systematic deviation and works as expected.

Calibration Source	Fitted Peak [keV]	Corrected [ch]	Uncorrected [ch]
^{109}Cd	75	-0.8 ± 1.7	-4.1 ± 1.7
^{139}Ce	127	-2.1 ± 1.9	-5.7 ± 1.9
^{113}Sn	369	-2.9 ± 2.1	-10.3 ± 2.1
^{137}Cs	630	-9.3 ± 7.8	-19.4 ± 7.8
^{207}Bi	995	-10.7 ± 5.1	-32.9 ± 5.1

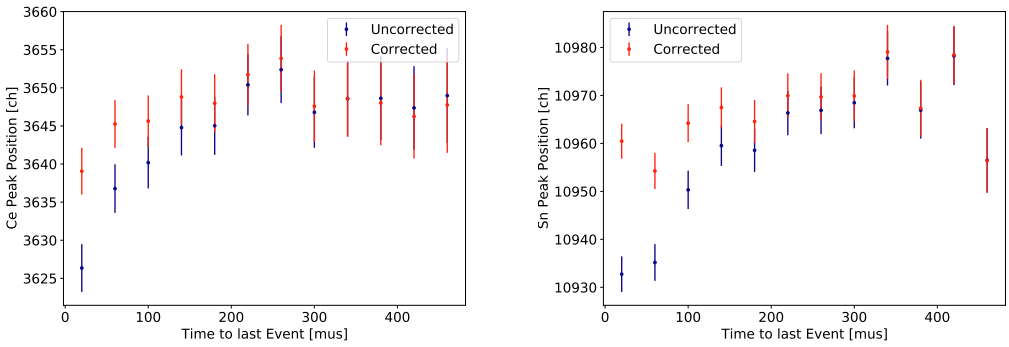
on the Fierz interference term b from the total rate dependency correction is $\sigma_{\text{sys}}(b) \leq 10^{-4}$ for all tested fit ranges and is thus negligible.

We use different calibration source measurements to verify the rate dependency correction. We extract the peak positions with fits with Gaussian distributions for intervals of times to last events, similar to Fig. 5.3, and show the results in Fig. 5.5.[2] The peak positions for an interval of times to last events for different measurements are combined by using a weighted arithmetic mean. The spectra are only corrected for pedestal shifts and rate dependency, i.e., without background subtraction. Assuming the peak position for large times to last events ($t \geq 300\,\mu$s) is the true peak position value, we can calculate the deviation δ with and without the correction using all times. The results are shown in Tab. 5.2, and the rate dependency correction works as expected, although it does not entirely remove the rate dependency bias. This effect might be caused by the smaller pulse generator with fixed frequencies leading to an event rate of 2 kHz. The calibration source event rates are $2 - 7$ kHz, and the times between events are distributed stochastically.

I conduct an additional check to rule out any non-linearity induced by the rate dependency correction with synthetic data. To generate the data, I use a Monte-Carlo-based simulation, inverse cumulative distribution function sampling, and representative PMT amplitudes. The correction adds a constant offset to the final spectra for frequencies up to 9 kHz and energies above 1 MeV, inducing no additional non-linearity.

There are options to improve the correction for future work with similar readout electronics. The individual QDC channel amplitudes are generally below channel 5000, and only the 50 and 200 mV signals from the pulse generator tests are in this range. More data in this region would improve the analysis by decreasing the systematic uncertainty. Also, more data would allow for better hypothesis testing and even determine if a linear dependence is accurate when aiming to further minimize this systematic effect. Alternatively, it is unclear why the pedestal is also rate dependent and why the subtraction of the first QDC sample does not automatically correct for the rate dependency in the first place. For a better understanding, analog electronic circuit simulation studies are being conducted within the PERKEO III collaboration.

[2] The 503 keV ^{207}Bi peak is missing in this analysis, as its position cannot be extracted accurately by fitting a Gaussian distribution.



(a) ^{139}Ce results from ten combined measurements. **(b)** ^{113}Sn results from ten combined measurements.

Figure 5.5 Validating the rate dependency correction of Equ. (5.3) for two calibration sources. We extract the peak positions for intervals of times to last events, similar to Fig. 5.3. As expected, shorter times are affected most, and the rate dependency correction significantly reduces the deviation at these times.

5.2.2 Spatial Response Gain Fine-Tuning

As discussed in Sec. 3.3.2, achieving perfectly matching gain for all PMTs is practically impossible. Besides imperfect detector components leading to a geometric asymmetry[3], relative gain deviations are a leading cause of a spatial dependency of the detector gain in the 2019/20 PERKEO III detectors for unprocessed data. This spatial dependency and the energy-dependent spatial distribution of electrons on the detector due to, e.g., gyration, lead to an energy-dependent gain $g(E)$.

We want to reduce the contributions of relative gain deviations to the spatial dependency of the detector gain by fine-tuning the individual PMT gain factors c_{ij}. Ideally, this reduction would only leave geometric asymmetry and other imperfections as the leading cause. However, we can also exploit the degrees of freedom of the individual PMT gain factors to form the spatial response to be symmetric. This approach allows for some compensation between contributions of both effects to the spatial dependency of the detector gain. Also, a more symmetric and uniform spatial response reduces the systematic uncertainties from effects related to detector misalignment and incoming electron distributions on the detectors.[4]

How to optimize the spatial dependency of the detector gain by fine-tuning individual PMT gain factors c_{ij} is discussed in Sec. 3.3.2, where one set of gain factors for one detector c_{0j} or c_{1j} is written as vector $\mathbf{c}$. We conducted the 2D grid measurement with the CaliBot using the ^{113}Sn calibration source, see Sec. 4.2.2. Fine-tuning c_{ij} is not a correction of a systematic effect but a change in the arbitrary scale between the incoming electron energy E_e and measured charge A of the detector response f_{Det} in Equ. (4.3). This approach slightly skews the detector response f_{Det} distribution but does not affect the shape of the final electron energy spectrum by design.

The resulting maps for both detectors for one spatial response measurement are shown in Sec. 5.4 with comparisons to previous PERKEO III detectors. I use the resulting maps to calculate the gain difference based on electron distributions as correction in Sec. 6.1.2.

[3] Examples of geometric asymmetries are scintillator and light guide surface roughness leading to local light yield losses or imperfect coupling of a light guide to a PMT.

[4] While other experiments want a sufficient spatial resolution, we see it as suffering from spatial dependence.

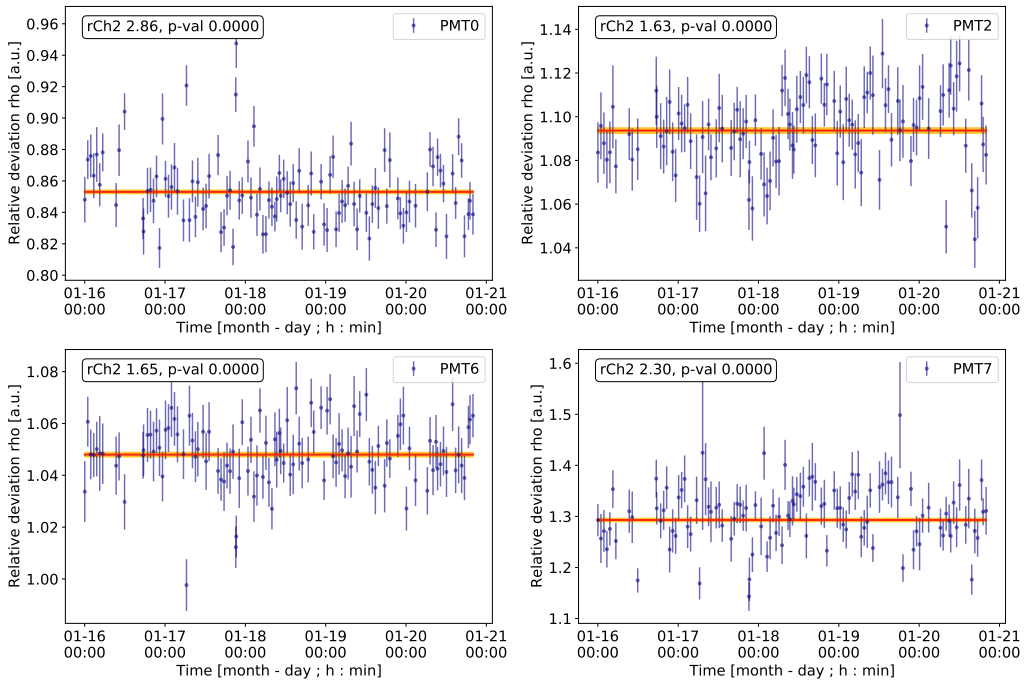


Figure 5.6 Testing the calibration source positioning uncertainty with individual PMT spectra from the hourly ^{113}Sn in both "center" positions, plotted for four PMTs. The fit result is shown as a red line with yellow uncertainty. The relative deviation ρ between both position measurements is not statistically constant, and thus, we cannot use individual PMT spectra for, e.g., calibrations.

The author of [Bes22] and I studied the optimization of the spatial response measurements and conducted simulation studies of the detector response. For the first time, spatial response measurements and simulations agree on the 1% level, Sec. 6.1. The comparison in [Bes22] of optimization approaches from previous work in PERKEO III [Mes11; Kre04], ultimately motivated our work [LBM22] presented in Sec. 3.3.2.

5.2.3 Source Positioning Uncertainty

Given that the new PERKEO III detectors have a light yield and energy resolution unprecedented in the collaboration, see Sec. 5.4, the question arises whether we can use individual PMT spectra for analysis and corrections. Besides energy resolution, the answer also depends on the precision of the calibration source positioning with the CaliBot. The CaliBot targets positions based on rotational encoders, and end switches limit its action space. The horizontal axis has a more coarse gear train, and is thus more susceptible to a systematic positioning uncertainty. The center of the neutron beam distribution is targeted in encoder steps, but an approximate center position is available via a switch on the horizontal axis as an alternative. We took hourly detector gain measurement with ^{113}Sn in both center positions as reference measurements to monitor the calibration source positioning for potential changes over time.

Theoretically, we know that both methods of moving to the center suffer systematic issues. Encoders can lose steps due to the custom control software and implemented multiplexer readout that was deployed. The motor moves faster with time of operation and gets slower after standby, leading to affected reaction times on switch contacts. However, we assume the switch targeted center position to be constant as its systematic effect should be smaller

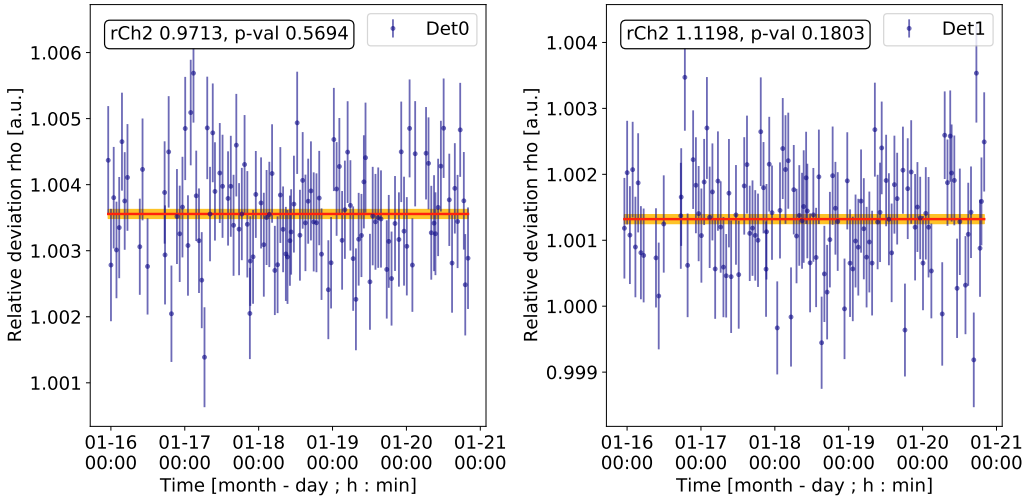


Figure 5.7 Testing the calibration source positioning uncertainty with detector spectra combining the signals of the eight detector PMTs from the hourly ^{113}Sn in both "center" positions. The fit result is shown as a red line with yellow uncertainty. The relative deviation ρ between both position measurements is statistically constant and thus, we can use combined spectra for, e.g., calibrations.

in comparison. If the encoder targeted center position is constant over time, the relative deviation ρ between both measurements must also be constant. I test this hypothesis for individual PMT data in Fig. 5.6. The ^{113}Sn peak positions are extracted by fitting the data with an exponentially modified Gaussian distribution.[5] The requirement for ρ = const. does not hold for individual PMT spectra, as the data significantly deviates from a constant.

Repeating the test, we can reconstruct the measured spectra on one detector by summing over all eight PMT signals of one detector for each event. The resulting spectrum is sufficiently Gaussian, and we can extract the measured energy with the fit of a Gaussian distribution, I repeat the test for the full detector data in Fig. 5.7, and ρ is constant within statistical limitations. We can verify the fitted value of ρ_i of detector i by comparing it to the measured spatial response after fine-tuning using the difference in rotational encoder steps with linear interpolation. The fitted values $\rho_0 = 1.003556(68)$ and $\rho_1 = 1.001321(68)$ agree with the linear interpolation of $\rho_0 = 1.003602$ and $\rho_1 = 1.001404$. This result implies that while the energy resolution of the detectors is sufficient to use individual PMT spectra for analysis and corrections, the calibration source positioning would need to be more precise. The detector is sensitive to horizontal positioning fluctuations, disabling, e.g., individual PMT temperature-induced gain corrections. Using detector spectra combining the signal of the detector PMTs for each event is robust to calibration source positioning uncertainties.

To further justify that using combined detector spectra is robust to positioning uncertainties, I conduct two additional tests. The detectors are placed opposite of each other in the experiment, and their measurements correlate regarding displacements perpendicular to the beam axis. Thus, the relative deviation between the detectors must also be constant. Also, the inter-detector spectra created by combining the signals of all PMTs from one side

[5] While more accurate descriptions of PMT spectra are available[Sau18; Diw20], determining their parameters is complex and can increase systematic uncertainty. I found exponentially modified Gaussian distributions to be good empirical models, motivated by [Diw20]. I hypothesize that the reason for this suitability is the voltage divider with capacitors in the high voltage base of the PMTs and the discharge behavior when producing a pulse.

(perpendicular to the beam axis) must have a constant relation to the equivalent spectra from the other side. Both tests are passed within statistical limitations.

5.2.4 Temperature Induced Gain Fluctuation Correction

We calibrate our detectors by precisely measuring the detector response f_{Det} with five electron conversion sources. However, the detector gain g fluctuates over time, and we must correct for these fluctuations between full calibration measurements. As described in Sec. 4.2.2, we monitor the detector gain hourly with a ^{113}Sn measurement between calibrations with all five sources. I extract the ^{113}Sn energy peak position by fitting a Gaussian distribution for each detector and plot the results over time in Fig. 5.8. The detector energy resolution is sufficient to correct individual PMTs for temperature-induced gain fluctuations, but the source positioning is not precise enough, see Sec. 5.2.3, and we correct the detectors as a whole.

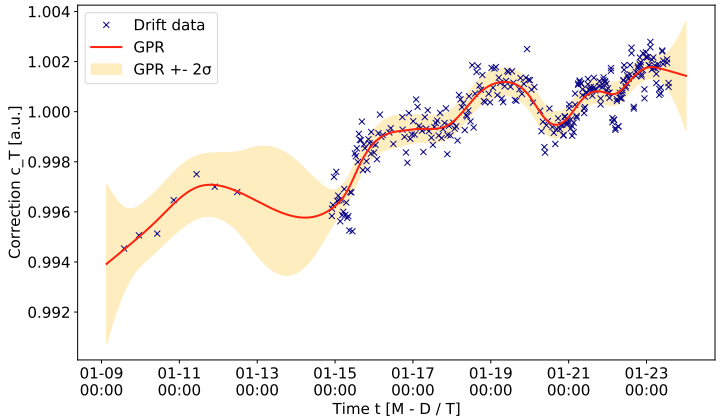
We hypothesize that the gain g mainly changes over time due to temperature-induced fluctuations. The PERKEO III spectrometer consumes a lot of power and has a closed cooling circuit. However, the heat exchanger is moderated with the thermally non-regulated ILL cooling water stored outside the facility. Therefore, we suspect a delayed correlation and long-time drift with the outside temperature. We expect that an increase in temperature leads to a higher gain and therefore a smaller correction factor c_T. We use Gaussian process regression (GPR) to reconstruct the underlying signals in a non-parametric and Bayesian way, as discussed in Sec. 3.3.1. The GPR constructs the temperature-induced gain fluctuation correction c_{Ti} for detector i that avoids correcting the noise, improving on previous PERKEO III analysis [Sau18], where a linear interpolation between the ^{113}Sn measurements was sufficient. The correction results are also plotted in Fig. 5.8. We calculate the correction as relative deviation from the weighted arithmetic mean of all ^{113}Sn peak positions.

The systematic uncertainty of the correction $\sigma(c_{Ti})$ between the 15th and 23rd of January 2020 is between $2 - 4 \cdot 10^{-4}$ and on average $3 \cdot 10^{-4}$ for each detector. The total systematic uncertainty $\sigma_{\mathrm{sys}}(g)$ on the detector gain g caused by temperature-induced gain fluctuations of both detectors is $\sigma_{\mathrm{sys}}(g) = 4.2 \cdot 10^{-4}$ ch / keV. As a precise energy measurement is highly sensitive to systematic corrections of the gain, this correction is non-negligible for a precise measurement of the Fierz interference term b and I discuss its significance in Sec. 8.

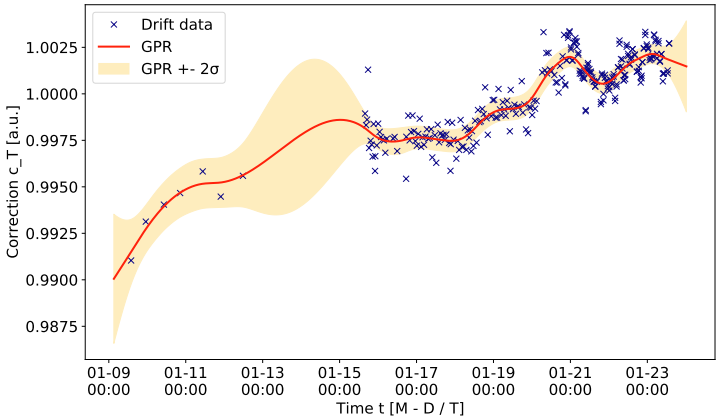
I check our hypothesis by verifying the correlation of the temperature-induced gain fluctuation correction with the GPR with the outside temperature in Grenoble in Fig. 5.8c and find a significant correlation matching our expectations. The correlation is not ideal, as we also expect other external events to influence the temperature of the cooling water and instrument. Examples of such events would be power-consuming instruments being turned on or off or user-caused changes, like an emergency shut-off of the experiment leading to a longer period of reaching a temperature equilibrium. Nevertheless, we can correct all these cases reliably with GPR, motivating our work [LBM22].

5.2.5 Energy Reconstruction Formula

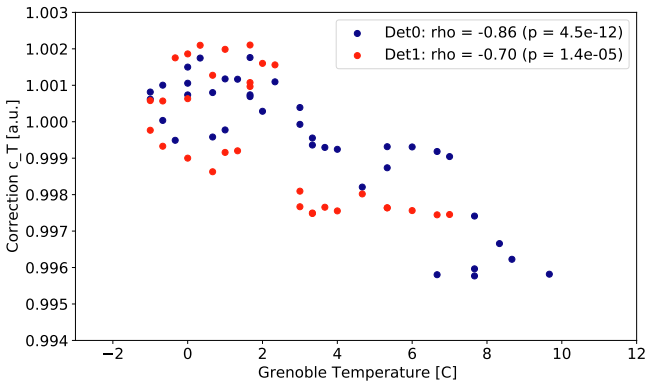
With all corrections on single events analyzed, I derive the formula for the energy reconstruction of a single event, combining previous sections. For a measured QDC sample $(A_1)_{ij}$ with detector i ($i \in \{0, 1\}$) from PMT j ($j \in \{0, ..., 7\}$), we subtract the pedestal P_{ij} and calculate the rate dependency correction amplitude R_{ij} according to Equ. (5.3). The rate dependency



(a) Re-used from Sec. 5.2.4: Temperature-induced gain fluctuation correction $c_T(t)$ with GPR, reconstructing the underlying signal for the upstream detector (Detector 0).



(b) Temperature-induced gain fluctuation correction $c_T(t)$ with GPR, reconstructing the underlying signal for the downstream detector (Detector 1).



(c) Correlation between drift correction and measured outside temperature in Grenoble (shifted by 7 h) between the 15th and 23rd of January 2020 for both detectors. Correlations are calculated with Pearson coefficient and are significant.

Figure 5.8 Overview of the temperature-induced gain fluctuation (drift) correction.

uses the previous measured sample $(A_0)_{ij}$, time to previous signal Δt, and parameters η_{ij} and k. Finally, we multiply the individual PMT amplitudes with fine-tuning factor c_{ij}, sum over the eight PMTs in one detector, and multiply the temperature-induced gain fluctuation correction c_{Ti} at measurement time t before summing both detector amplitudes to get the total event amplitude A as

$$\begin{aligned} A &= \sum_{i=1}^{2} A_i = \sum_{i=1}^{2} c_{Ti} \sum_{j=1}^{8} c_{ij} \cdot R_{ij} \\ R_{ij} &= (A_1 - P)_{ij} + (A_0 - P)_{ij} \cdot \eta_{ij} \cdot \exp^{-\Delta t/k} \end{aligned} \tag{5.4}$$

For the analysis for some of these corrections, I already used parts of Equ. 5.4, e.g., to create the ^{113}Sn spectra to determine the temperature-induced gain fluctuation correction c_{Ti}. Some analyses require studying detector signals independently and only summing over j to get the total detector amplitudes A_i.

Final Offset Correction

The initial pedestal correction in Sec. 5.2.1 removes the offsets for the individual PMT spectra on a sub-ch level. However, the consecutive corrections lead to an induced offset of 0-3 ch, and we correct them with a second and final offset correction with the same filter as for the pedestal calculation. This induced offset is caused by the single-event corrections implemented for this analysis and not necessary for previous work. The offset correction has the same uncertainty as the pedestal correction with $\sigma(O) = 0.56(4)$ ch. It is non-negligible for precisely measuring the Fierz interference term b, and I discuss its significance in Sec. 8.

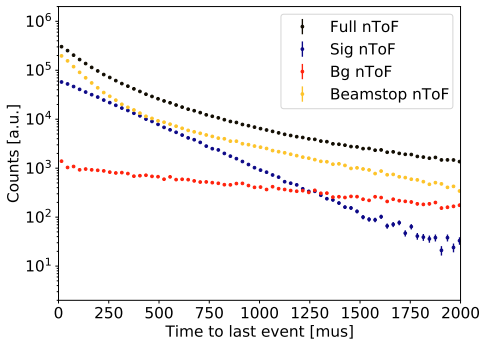
5.3 Spectral Corrections

After correcting single events and reconstructing their energy, we use histograms H to approximate the probability density function of the energy spectra. For neutron beam measurements and calibration measurements, we must subtract background events. We obtain the signal H_s and background H_b spectra in neutron beam measurements by discriminating with different neutron time-of-flight windows, see Sec. 4. For calibration measurements, background measurements are taken between measuring with the calibration sources. In both cases, we must correct the detected rate for each energy bin before subtracting H_s and H_b. This requirement also holds for other operations combining spectra.

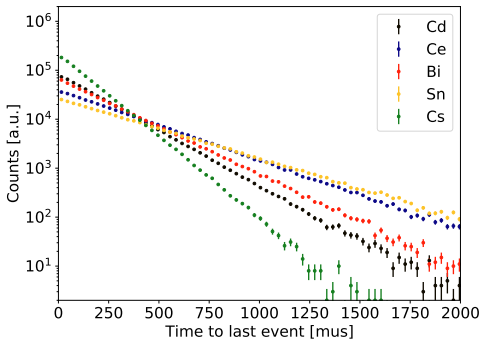
5.3.1 Dead Time Correction

Our measurement setup has a non-paralyzable dead time τ to allow for the electronics to reset between recorded events. Contrary to a paralyzable dead time, events occurring during the dead time are not registered and lost. Therefore, we must correct the number of measured events k in the time interval t, to correct the measured rate $N = k/t$ to obtain the true counts without dead time k_0 and true rate $N_0 = k_0/t$. The correction factor δ is given as [Leo94]

$$\delta = \frac{k_0}{k} = \frac{1}{1 - \tau N}.$$



(a) Neutron beam measurement times to last events distinguished for different neutron time-of-light windows.



(b) Times to last events for the five different calibration sources, see Sec. 4.2.2.

Figure 5.9 Validating the dead time correction of Equ. (5.5) requirement of constant true rates N_0 by plotting the observed counts for different times to last events. The observed rates are linear in the logarithmic plot and therefore in agreement with an exponential distribution, implying a constant rate. For the neutron measurement: The beamstop component (yellow) is shown to explain the distribution without neutron time-of-flight cuts (black). The neutron time-of-flight data before and after the signal window is not shown.

For our measurement, we apply neutron time-of-flight cuts Δ to our measurements, affecting the number of detected events. We must also adjust the formula with a correction δ_p for the chopper frequency ν as $\delta_p = (\Delta \cdot \nu)^{-1}$, leading to

$$\delta = \frac{1}{1 - \delta_p \tau \cdot N}. \tag{5.5}$$

The measured rate of histogram H is corrected with δ. A requirement for the dead time correction in Equ. (5.5) is a constant true rate N_0. To check this requirement, I present the distribution of times to last events for the neutron time-of-flight windows used to discriminate signal and background spectra and for the five calibration sources in Fig. 5.9. The observed rates are linear in the logarithmic plot and therefore in agreement with an exponential distribution, implying a constant rate.

We set the dead time digitally for the measurement as $\tau = 1.5\ \mu$s, however, we observe a dead time of $\tau = 1.660(3)\ \mu$s from neutron beam and calibration source measurements and use the measured value. We do not know the cause of this deviation. The correction factor δ is calculated for each measurement file. The average values of δ and their uncertainty are shown in Tab. 5.3 for different sources. I assume Δ to be exact and without error, while the fluctuations of τ, N, and ν are considered in the results.

Despite the impression in Fig. 5.9, the detected rate is not entirely constant when looking at very short times to last events (below 25 μs) and the cause of this deviation is unknown. The last PERKEO III analysis [Sau18] estimated a shift from this effect on δ of $\delta_c = \delta + 1.8 \cdot 10^{-5}$. Using this result to increase the uncertainty of δ and combining the systematic uncertainties for the signal and background neutron time-of-flight cuts in Tab. 5.3, I can estimate the significance of this correction by varying the background subtraction in the spectrum generation and test how it affects the extraction of the Fierz interference term b when fitting a neutron beta spectrum. I obtain an upper limit on the combined systematic uncertainty $\sigma_{\text{sys}}(b)$ on the Fierz interference term b from the dead time correction as $\sigma_{\text{sys}}(b) \leq 5 \cdot 10^{-5}$ for all tested fit ranges, making it negligible.

Table 5.3 Resulting dead time correction factors δ for all calibration sources and neutron beam data detector. The different values are caused by different rates N.

Source	δ [a.u]
Beam (Full nToF)	1.006143(12)
Beam (Signal nToF)	1.007787(13)
Beam (Bg nToF)	1.002422(11)
^{109}Cd	1.00887(16)
^{139}Ce	1.00563(18)
^{207}Bi	1.00770(18)
^{113}Sn	1.00487(24)
^{137}Cs	1.01294(17)

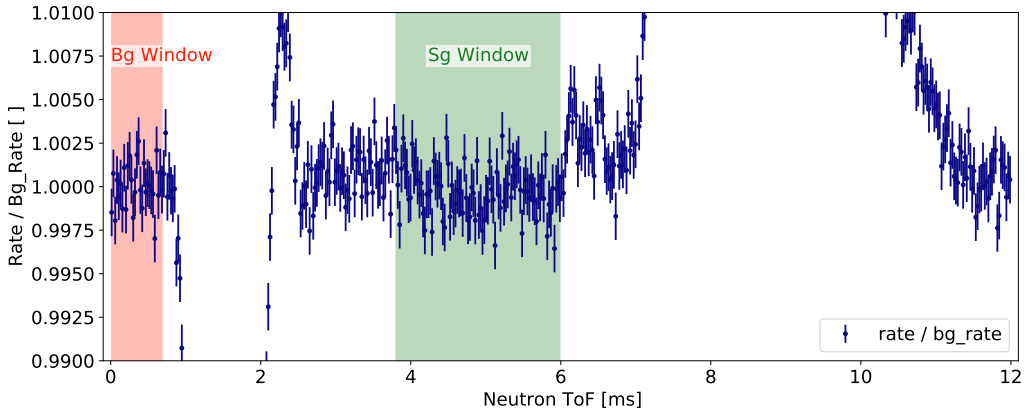


Figure 5.10 Combined background measurement without magnetic fields as detected rate vs neutron time-of-flight. The detected rate is normed with the fitted constant rate in the background window. The final signal and background neutron time-of-flight windows are also shown. The signal time-of-flight window cuts at 6 μs to avoid the time-dependent background rate increase.

5.3.2 Signal and Background Time-of-Flight Dependence

We must verify the background subtraction for the pulsed neutron beam measurement, presented in Sec. 4. While the pulsed neutron beam measurement allows for a background subtraction every chopper rotation, the measured background spectra must represent the background when measuring the signal spectra. We expect the background measurement to contain all pulsed beam-independent background sources, e.g., from other experiments or gamma radiation from neutron absorption in the beamline. However, we must verify if the pulsed beam-dependent background is representative, e.g., if delayed signals from the beam stop or background variations from individual chopper disc tiles are included and if the background created by the chopper is uniform.

To study the pulsed beam-dependent background, we used the PERKEO III setup as described and measured for three days, but without magnetic fields leading to no electrons reaching the detectors from neutron decays. With this data set, I can create a combined neutron time-of-flight spectrum from background measurements as shown in Fig. 5.10.[6] The sharp decrease in detected events between 1 and 2 ms stems from the chopper opening and not absorbing any neutrons. Delayed emissions from the beam stop, e.g., signals from the

[6]Before combining the data collected over three days with weighted arithmetic means, I verified that the distribution of detected rates in each neutron time-of-flight bin is Gaussian within statistical limits.

absoption of scattered neutrons, are causing the long tail up to 12 ms. These emissions motivate a background neutron time-of-flight window between 12/0 ms and the chopper opening again. The minor rate increase at 6.5 ms presumably stems from a chopper disc tile and must be avoided in the signal neutron time-of-flight window. Thus, I chose the neutron time-of-flight signal as 3.8 to 6.0 ms and the background window as 0 to 0.7 ms to discriminate the measured data and create background-free spectra through subtraction. The neutron time-of-flight signal window must also guarantee that the neutron pulse is within the homogeneous region of the magnetic field to avoid systematic uncertainties from the mapping of the beta decay electrons onto the detectors. The latter can be checked with the detected rate in Fig. 4.4 and is fulfilled for the chosen window.

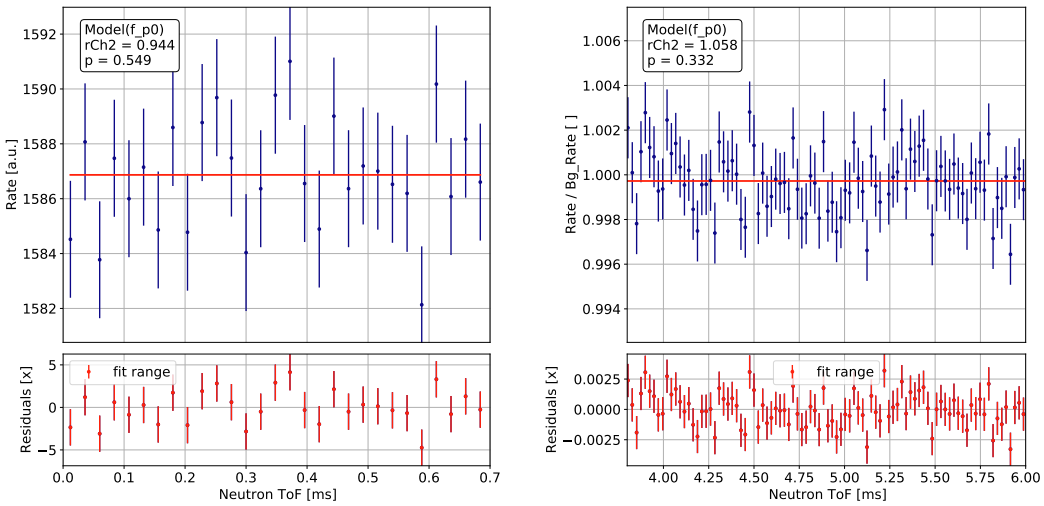
Three requirements must be fulfilled to verify if the measured background spectra represent the background when measuring the signal spectra. The detected rate must be constant and equal in both windows, and the energy spectrum must match. To this end, I fit the rate in the background window and use it to normalize the detected rates in Fig. 5.10. The fit of the rate is shown in Fig. 5.11 as $c = 1586.87(40)$ and is statistically constant with a reduced $\chi^2 = 0.94$ and p-value of $p = 0.55$. The fit of the rate in the signal window is also statistically constant with a reduced $\chi^2 = 1.06$ and p-value of $p = 0.33$, but the fitted value of $c = 0.99972(14)$ deviates from 1. Similar to the dead time correction, I use this deviation as systematic uncertainty to estimate an upper limit on the combined systematic uncertainty $\sigma_{\text{sys}}(b)$ on the Fierz interference term b from the background subtraction as $\sigma_{\text{sys}}(b) \leq 3.6 \cdot 10^{-4}$ for all tested fit ranges, making it negligible in comparison, see Sec. 8.

To test the energy dependence of the correction and if the energy spectra match in both windows, I subtract both spectra, like for neutron beam data with magnetic fields, and verify if the resulting spectrum is constant and in agreement with 0 for all energies. The results are shown in Fig. 5.12 and statistically in agreement with 0 with $c_0 = 2.5(3.1)$ and $c_1 = 3.2(2.4)$. Moving the lower fit range up to channel 7000, ≈ 227 keV, the values further improve $c_0 = 1.9(3.8)$ and $c_1 = 1.5(2.2)$. Thus, there is no added energy dependence from the background subtraction, especially when fitting above channel 7000.

Compared to previous PERKEO III measurements [Sau18; Mes11], we have no measurable slope of the rates in the background window. However, our selected background window is 30% smaller. The signal window is similar in previous measurements, as they also observed the background rate increase at 6.5 ms [Sau18]. In contrast to the measurement with neutron spin-dependent quantities highly sensitive to the magnetic mirror effect [Sau18], we do not need to center our signal window around the neutron time-of-flight rates of both detectors. Also, our measurement is less sensitive to systematic background corrections. For future work with a similar setup, the time dependence of the background analysis can be improved by exploiting the shorter signal window by moving the beam stop further inwards, allowing for a longer background measurement and better statistics.

5.4 Data Quality

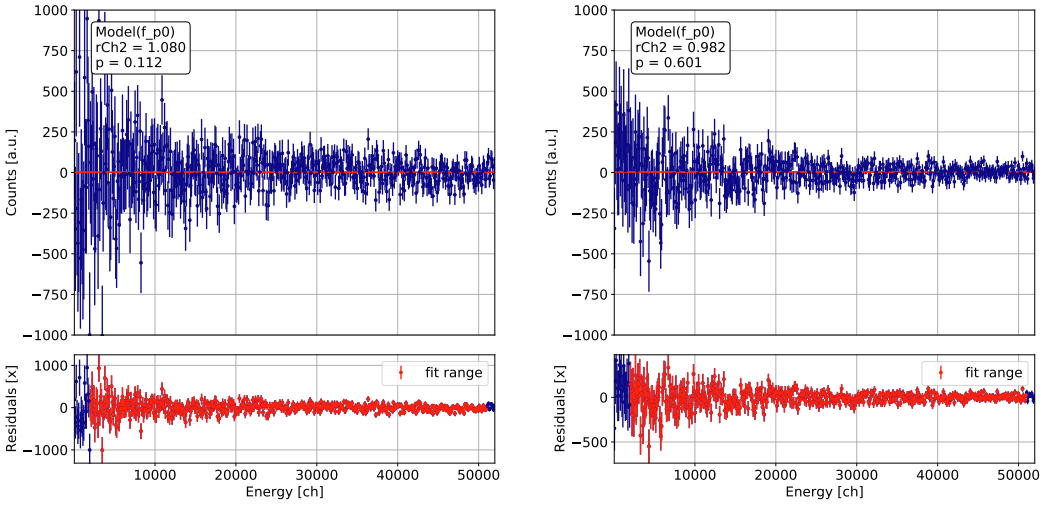
During the 2019/20 PERKEO III measurement campaign, about $5 \cdot 10^8$ neutron beta decay events were detected, neglecting characterization measurements, making it the most extensive unpolarized data sets of its kind. Besides using the data set to measure the Fierz interference term b, other applications are possible for future analysis that will build upon the data set and its analysis tools of this thesis, e.g., looking for deviations in spacetime



(a) Fitting the detected rate in the background window to norm the detected rates.

(b) Fitting the detected rate in the signal window after norming the rate with the fitted background rate. The rate in the signal window hsould be 1.

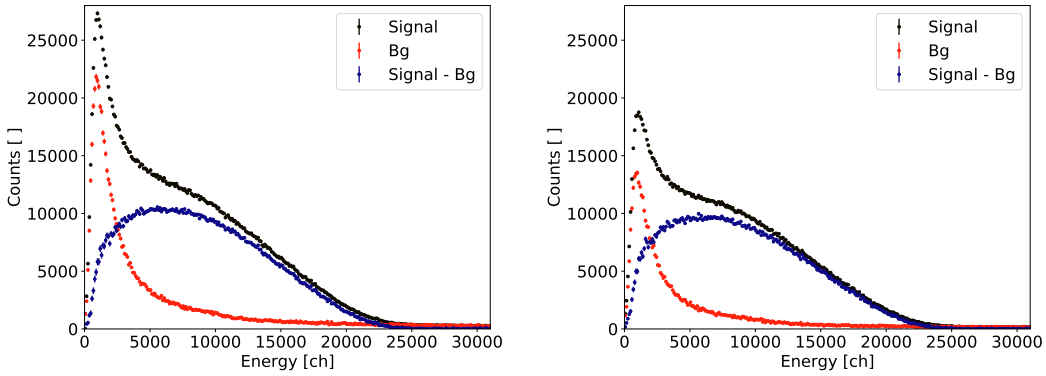
Figure 5.11 Testing the detected rates in the background and signal neutron time-of-flight windows from the background measurements without magnetic fields from Fig. 5.10 to verify that the detected rate in the background window is representative for the background rate in the signal window.



(a) Background subtraction for detector 0 (upstream).

(b) Background subtraction for detector 1 (downstream).

Figure 5.12 Testing the energy dependence of the background subtraction with the background measurements without magnetic fields and verifying that there is no added energy dependence from the background subtraction.



(a) Electron beta spectrum from neutron measurements on detector 0 (Upstream)

(b) Electron beta spectrum from neutron measurements on detector 1 (Downstream)

Figure 5.13 Signal and background measurement of six combined 5 min neutron beam measurements with all corrections. The signal-to-noise ratio for the whole energy range is 3.007(4) for detector 0 and 3.968(7) for detector 1, improving on previous results by 50 – 100%. The background measurement is corrected for the shorter measurement time from the neutron time-of-flight cut.

Table 5.4 Signal-to-noise ratios for different lower energy cuts for both detectors. A signal of 10000 ch is about 323 keV.

Lower cut-off [ch]	S/N Detector 0 [a.u]	S/N Detector 1 [a.u]
0	3.007 ± 0.004	3.968 ± 0.007
7000	5.393 ± 0.013	7.936 ± 0.025
10000	4.893 ± 0.014	7.278 ± 0.027

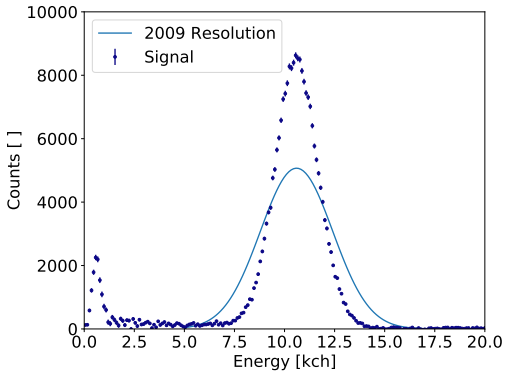
symmetry to probe Lorentz invariance [Dia14]. Thus, I discuss the quality of the data set, compare it to previous PERKEO III results, and present consistency checks.

5.4.1 Detector Performance

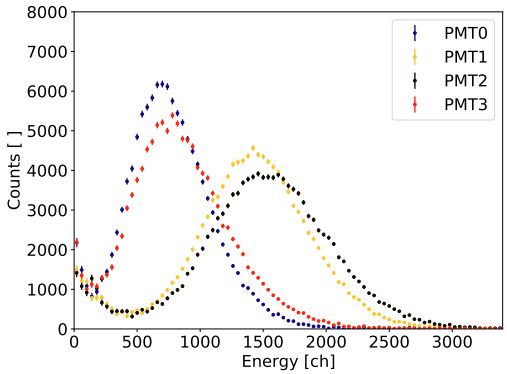
A higher signal-to-noise ratio is a key indicator for the statistical precision and accuracy of a measurement, as more information can be extracted from it. Due to optimizing our measurement setup to measure the electron beta spectrum precisely, we did not use a neutron spin polarizer. This setup leads to a higher neutron flux and overall detected rate, removes the strong gamma background created from the neutron spin polarizer, and allows for the first neutron absorbing aperture in the beamline to be further away from the detector. The greater distance reduces the noise on the upstream detector, and all effects combined directly improve the signal-to-noise ratio.

Compared to the 2009 PERKEO III measurement with the current most precise result of the Fierz interference term b [Sau+20] with a signal-to-noise ratio of ≈ 2 [Sau18], we improve the signal-to-noise ratio by 50-100 %. Our signal-to-noise ratio results are presented in Tab. 5.4 and illustrated in Fig. 5.13. The rate of background events is lower for detector 1, as the leading background source during the signal neutron time-of-flight cut stems from the beamline further upstream.

The light yield f_{pe} is defined as the number of photons detected for one keV and is directly related to the energy resolution of the measurement. A more detailed and accurate description of the detector model is given in Sec. 6.1. Compare to the PERKEO III 2009 measurement with



(a) A ^{113}Sn measurement with detector 0 (Upstream), reconstructing the signal with all PMTs.



(b) Individual PMT spectra of the ^{113}Sn measurement with detector 0 (Upstream).

Figure 5.14 Illustrating the light yield improvement compared to previous detectors in 2009 with ^{113}Sn and the caused increase in energy resolution after subtracting background.

$f_{\text{pe}} = 0.230\ \text{keV}^{-1}$ [Sau18], our improved detectors[7] reach a light yield of $f_{\text{pe}} = 0.660\ \text{keV}^{-1}$. The light yield improvement is shown in Fig. 5.14 for a ^{113}Sn measurement. Also, the light yield is sufficiently high to potentially analyze individual PMT spectra, which was previously not a viable option.

In the last PERKEO III measurement, the temperature-induced gain fluctuation correction is one of the largest detector-related systematic uncertainties [Sau+20]. The last measurement had a day-night variation of about 2% [Sau18], while our measurement has no visible day-night variation and day-to-day changes of about 0.03%, Sec. 5.2.4. The significant reduction is the cause of better detector design [Ber18] and cooling system, a better-suited environment with colder cooling water, and due to more sophisticated analysis tools that avoid correcting statistical fluctuations. Also, fine-tuning the individual PMT gain factors before correcting temperature-induced gain fluctuations should reduce individual PMT gain fluctuations to a certain extent.[8] This improvement is essential for reaching the targeted precision for measuring the Fierz interference term b.

The spatial dependence of the detector gain, quantified as detector uniformity, leads to an energy-dependent gain $g(E)$. Comparing the uniformity of the spatial response measurements between the PERKEO III 2009 (Fig. 5.15) and our measurement (5.16), the maximum deviation is decreased and the symmetry greatly increased. This improvement enables a more precise systematic correction for the spatial dependence of the detector gain and light yield in Sec. 6.1.2.

5.4.2 Data Consistency and Quality Checks

Event Permutation

I verify that measurements are robust to event permutation to rule out specific correlations between events from the readout electronics and data storing. Specifically, I extract the measured signal by fitting individual PMT ^{113}Sn spectra with exponentially modified Gaussian distributions while using only every n-th event. The total number of events used after the

[7] Detector calibration parameters are discussed in Sec. 7.2.
[8] The 2009 PERKEO III analysis also fine-tunes the individual PMT gain factors making it a fair comparison.

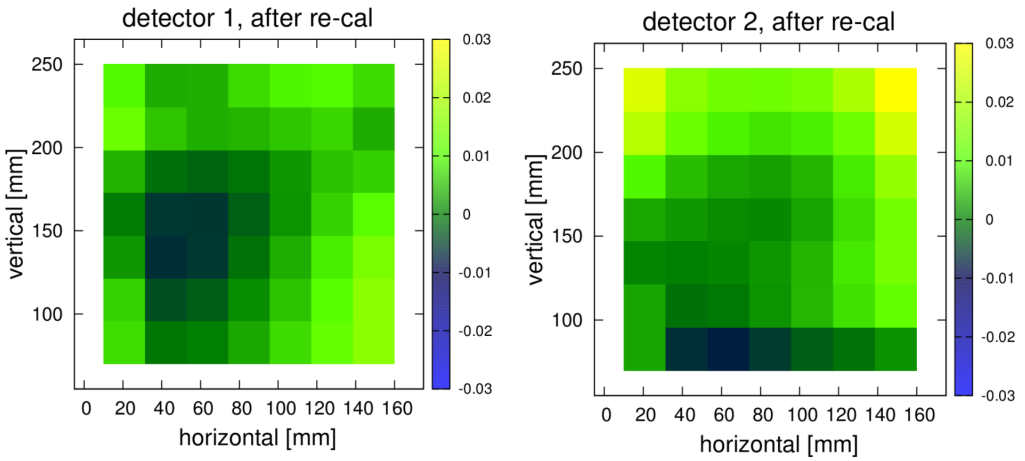
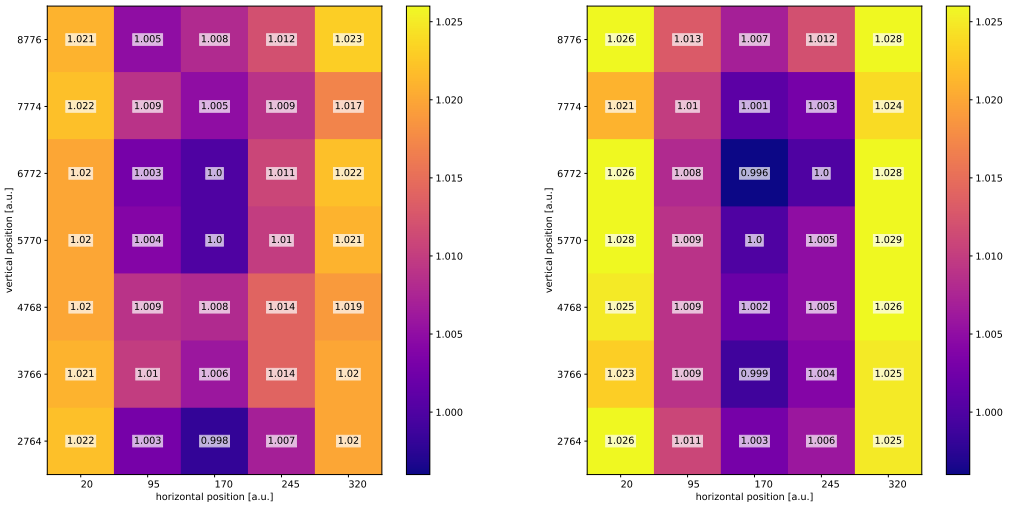


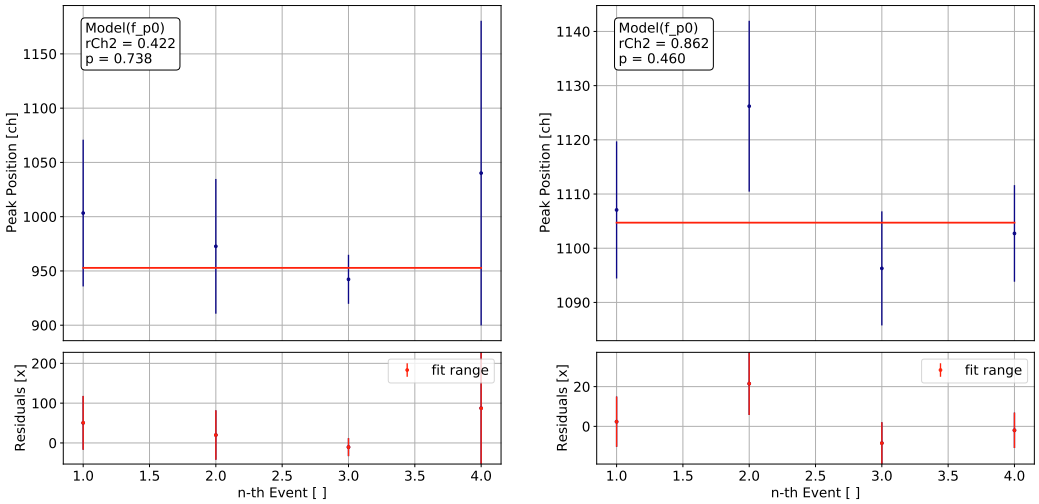
Figure 5.15 Taken from [Mes11; Sau18]: Spatial response measurement of the 2009 PERKEO III detectors after fine-tuning the detector response.



(a) Reused from Sec. 3.3.2: Spatial response measurement of detector 0 (upstream).

(b) Spatial response measurement of detector 1 (downstream).

Figure 5.16 Spatial response for one ^{113}Sn grid measurement with current data for both detectors after fine-tuning the individual PMT gain factors. The resulting response for both detectors is significantly more symmetric and uniform than previous PERKEO III measurements, see Fig. 5.15. Detector 1 is less uniform than detector 0, requiring a larger spatial response correction in Sec. 6.1.2. The same color scale was used for both plots.



(a) Results for data of PMT7

(b) Results for data of PMT10

Figure 5.17 Event permutation checks: Extracting measured signals of ^{113}Sn using only every n-th event in individual PMT spectra. The total number of events used after the permutation is kept constant and only shown for two example PMTs that are representative of all PMTs. No permutation-dependent effect is observed within statistical limitations.

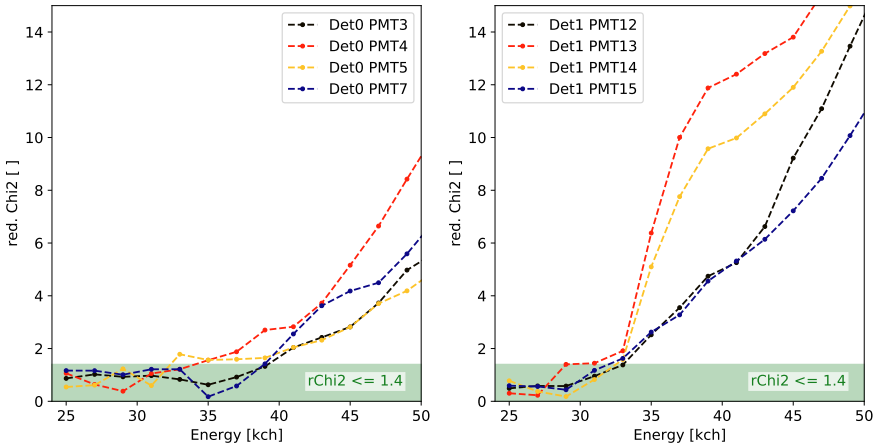
permutation is kept constant to eliminate statistical uncertainty differences. The results are shown in Fig. 5.17 and I detect no deviation for different event permutations. Also, I verify that the fluctuations of every second ($n = 2$) and third ($n = 3$) event relative to the fitted mean pass a random number generator test to rule out correlations.

Individual PMT Channel Saturation

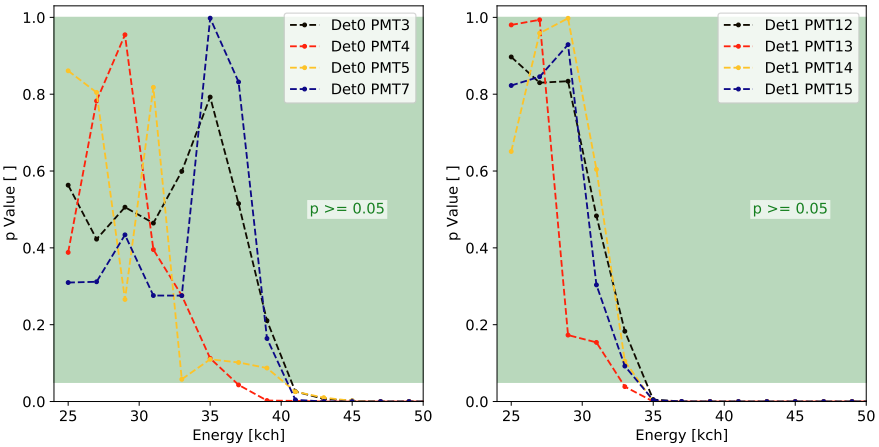
We must verify that individual PMT channels do not saturate and induce additional non-linearity. To this end, I study the behavior of the last 20 bins of individual PMT spectra when measuring ^{207}Bi. The maximum bin is given by the dynamic range of the QDC charge integrator. The hypothesis is that the last bins follow an exponential decay as the tail of the actual signal peak. The results of this test are shown in Fig. 5.18 for the PMT channels most affected in the test for each detector. Using a cut-off at 34 kch (about 1.1 MeV) for both detectors eliminates the effect and guarantees that it is not present in the measured data.

Anomalous PMT Behavior

When analyzing the number of detected events of individual PMT channels as integral normed by the total number of events of that detector, PMT13 of detector 1 shows anomalous behavior between the 13th and 15h of January 2020. All other PMTs are stable and fluctuate in unison, while PMT13 strongly deviates and loses events or measured amplitude. The effect is localized to that period and does not affect other measurements. To guarantee the validity of this analysis, I do not use the data between the 13th and 15h of January 2020.



(a) Fit result reduced χ^2 value for different PMTs with chosen cut-off at $p \leq 1.4$.



(b) Fit result p value for different PMTs with chosen cut-off at $p \geq 0.05$.

Figure 5.18 Test until which detected energy the last 20 bins in individual PMT spectra follow exponential decay when measuring ^{207}Bi. The reduced χ^2 and p values are used to set a cut-off energy to rule out individual PMT saturation effects. Only showing PMT channels with significant saturation effects. Only including detected energies up to 34 kch ensures avoiding any saturation effects.

Correction Universality

A simple test of splitting a measured data set into disjunct sets, applying corrections, and combining the spectra again must produce identical spectra without splitting the dataset and applying the same corrections. This test must be fulfilled for all corrections and possibilities to split the data set into disjunct tests. When using the *Panter* analysis framework, this test is fulfilled for all tested combinations within numerical precision.

Different Measurement Modes

We must verify which QDC samples from data measured in the ALLMODE, e.g., used in Fig. 4.9, produce the same results as data measured in the other modes, specifically DELTA-MODE. The first and last QDC samples produce the same results when using six ^{113}Sn measurements for each mode and studying the quadratic deviations for all measurements. Overall, no systematic deviation between the measurement modes is observed.

6 Theoretical Model of Measured Spectra

To extract physical quantities like the Fierz interference term b from measured data, we must formulate a theoretical model to fit it to the data and determine its uncertainties. We construct our hypothesis model by forward modeling from binned electron line spectra to scintillator interactions and the charge integration in the QDCs. This approach is more precise than applying all corrections to measured data, as unfolding the data distribution leads to information losses, limits statistical accuracy, and insufficient precision for some correction parameters. Instead, we assume the underlying models, e.g., for the non-linearity, from physical motivation. In this chapter, I giev an overview how to derive the theoretical spectra based on previous work, compare results to simulations to verify our understanding of the data, and derive corrections that expand the theoretical spectra. Where applicable, I point out future work for corrections.

6.1 Model of the Detector Response

The energy spectra for the electrons from neutron beta decay or the calibration sources must be modeled under the scope of the detector response and experiment setup. This section summarizes the work of [Sau18; Roi18].

The theoretical shape of former is described in Equ. (2.10) and the author of [Roi18] implemented the software to go from conversion source transition energies to binned histograms for the latter, see Sec. 7. The code considers emission combinations that could happen simultaneously in the QDC integration time window and Auger electron emissions that correlate with conversion electron emissions. The code also includes beta decay spectrum generation, as for the neutron beta decay or the decay of ^{137}Cs.

These single and multielectron events hit the scintillator of a detector. Electrons deposit their energy E, and due to fluor ex- and de-excitations, a number of photons $n_\gamma(E)$ is produced. The number of produced photons $n_\gamma(E)$ follows a Poisson distribution. These photons propagate in the scintillator, and $n = t \cdot n_\gamma(E)$ reach a PMT through the light guides with a transmission probability t in a binomial process. Generally, the transmission probability $t = t(\mathbf{x})$ is not constant and depends on the impact position $\mathbf{x}$ of the electrons, i.e., the photon source. We assume t to be constant and correct resulting effects afterwards, see Sec. 6.1.2. At the PMT, the photosensitive layer converts the photons into $n_{\text{pe}} = \sigma \cdot t \cdot n_\gamma(E)$ photo-electrons in another binomial process with the photon-conversion probability, also referred to as quantum efficiency, $\sigma \approx 0.22$ [Pho96]. A PMT produces charge pulses from single photo-electrons through nineteen avalanche amplification stages with high-voltage dividers. The amplification between stages is described with a Poisson distribution, and the overall amplification can be modeled with a Galton-Watson process for population growth [Sau18].

We can combine all these processes into a final Poisson distribution using the moment-generating functions of the individual processes with the characteristic number of photo-

electrons per detected energy $f_{\mathrm{pe}} = a \cdot t \cdot \sigma$, where a is the linear relation between the number of photons per energy $n_\gamma = a \cdot E$, as [Sau18]

$$p_{\mathrm{pe}}(n_{\mathrm{pe}}; E, f_{\mathrm{pe}}) = \frac{(f_{\mathrm{pe}} \cdot E)^n}{n_{\mathrm{pe}}!} \exp\left(-f_{\mathrm{pe}} \cdot E\right). \tag{6.1}$$

The relation between incoming charge pulse A for an initial electron energy E is Gaussian distributed for a channel c. Due to the pedestal subtraction in Sec. 5.2.1, we can assume the gain g as linear relation without offset as $\mu = A \cdot g$ and σ_{Q}, leading to [Sau18]

$$p_{\mathrm{Q}}(c; A, g, \sigma_{\mathrm{Q}}) = \mathcal{N}(c; A \cdot g, \sigma_{\mathrm{Q}}).$$

This model gives us a simple framework to describe our spectra and implies that the widths of peaks in the final spectra solely depend on the light output, and the relation of measured peak positions and electron energies determines the gain. For example, this dependence means we must correct the temperature-induced gain fluctuations in Sec. 5.2.4 as a gain correction with an uncertainty, as neglecting it would lead to a broader electron beta spectrum with a different energy resolution to the calibration measurements. We can expand the model in Equ. (6.1) for a non-linear energy-channel relation by replacing E with an effective $E_{\mathrm{f}}(E)$. The model in Equ (6.1) holds for a readout with multiple PMTs where the light is distributed between them. The final distribution combining all PMTs will have the same form but uses an average value of f_{pe}.

Scintillator Non-Linearity

Scintillators are known to have a non-linear energy-light output relation [Bir51; Pös+21] due to processes like ionization quenching and other local saturation effects from high stopping powers for low-energy electrons. Commonly, the Birks quenching model [Bir51] is used to describe scintillator non-linearity. Ideally, we would assume a linear relation between electron energy E and light output $L = a \cdot E$. However, we must assume a non-linear $L = a \cdot E_{\mathrm{f}}(E)$ relation with the effective energy E_{f} deposited in the scintillator that is proportional to the light yield. Writing the differential light yield along an infinitesimal track length $\mathrm{d}x$ and the deposited energy $\mathrm{d}E$ as the effective energy E_{f} deposited in the scintillator, we obtain [Sau18]

$$\frac{\mathrm{d}E_{\mathrm{f}}}{\mathrm{d}x} = \frac{\frac{\mathrm{d}E}{\mathrm{d}x}}{1 + k_B \cdot \frac{\mathrm{d}E}{\mathrm{d}x}},$$

with the particle and scintillator dependent Birks coefficient k_B. The differential effective energy E_{f} for the initial electron energy is then

$$\frac{\mathrm{d}E_{\mathrm{f}}}{\mathrm{d}E} = \frac{1}{1 + k_B \cdot \frac{\mathrm{d}E}{\mathrm{d}x}}.$$

We obtain the effective energy by integrating over the full deposited electron energy E as

$$E_{\mathrm{f}}(E) = \int_0^E \frac{1}{1 + k_B \cdot \frac{\mathrm{d}E'}{\mathrm{d}x}(E')} \, \mathrm{d}E'. \tag{6.2}$$

In my work, I use the implementation of Equ. (6.2) from [Sau18; Roi18] that uses a log-log-linear interpolation for $\mathrm{d}E'/\mathrm{d}x$ from the ESTAR data set [Ber+17] that is also used in [Mär+19; Sau+20].

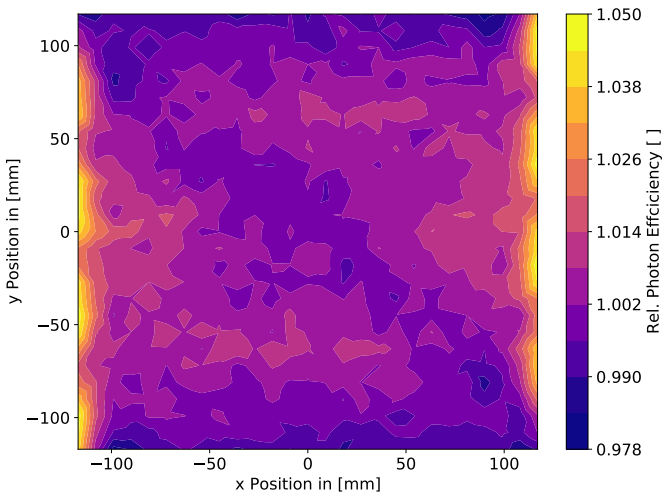


Figure 6.1 Light transmission efficiency, also referred to as photon efficiency, map of one detector combining the signal of all eight PMTs. The resulting map shows an estimated geometrical limit to the achievable uniformity, as the distribution has a local asymmetry and is skewed over the entire scintillator surface due to the light guide design [Ber18], readout at the sides, and surface roughness. The simulated light sources are 3 mm below the scintillator surface, the surface roughness parameter is $\rho = 0.8$, and the results are normalized to the value in the center.

The author of [Sau18] conducted experimental tests and quantified the Birks coefficient k_B for the scintillator used for the 2019/20 PERKEO III measurement as $k_B = 123 \pm 14$ nm/keV. Possible scintillator surface effects, e.g., a non-scintillating dead-layer which would increase the non-linearity for low-energy electrons, could not be discovered within statistical limits and are therefore neglected within my thesis.

6.1.1 Spatial Dependence of the Light Yield

In Equ. (6.1), we assumed the light transmission probability t to be uniform and constant over the scintillator surface. However, in reality, $t = t(\mathbf{x})$ is not constant and depends on the impact position $\mathbf{x}$ of the electrons, i.e., the photon source. To study systematic effects caused by this simplification, we use simulations of the light transmission efficiency in the scintillator and light guides to create a model of $t(\mathbf{x})$ for corrections.

There are multiple causes of the spatial dependence of t. For example, one cause is the localized readout with light guides at the sides, and the critical reflection angle at the scintillator surface lead to different angular acceptances depending on the distance of a light source to a light guide readout. Also, scintillators have an attenuation length of about 250 mm [Sai21] and non-ideal surfaces. Both effects lead to light losses depending on the photon path length and the number of reflections on the path to a PMT. Furthermore, the light guide components leading to one PMT for the detectors used on the 2019/20 PERKEO III measurement campaign are not left-right symmetric and have different light transmission probabilities.

The data sets were created in collaboration with [Bes22] using Geant4 [Ago+03], and more details on the simulation setup and software can be found there. The non-ideal scintillator surface is described with a surface roughness parameter ($[0, 1]$, 1 is an ideal surface) of the

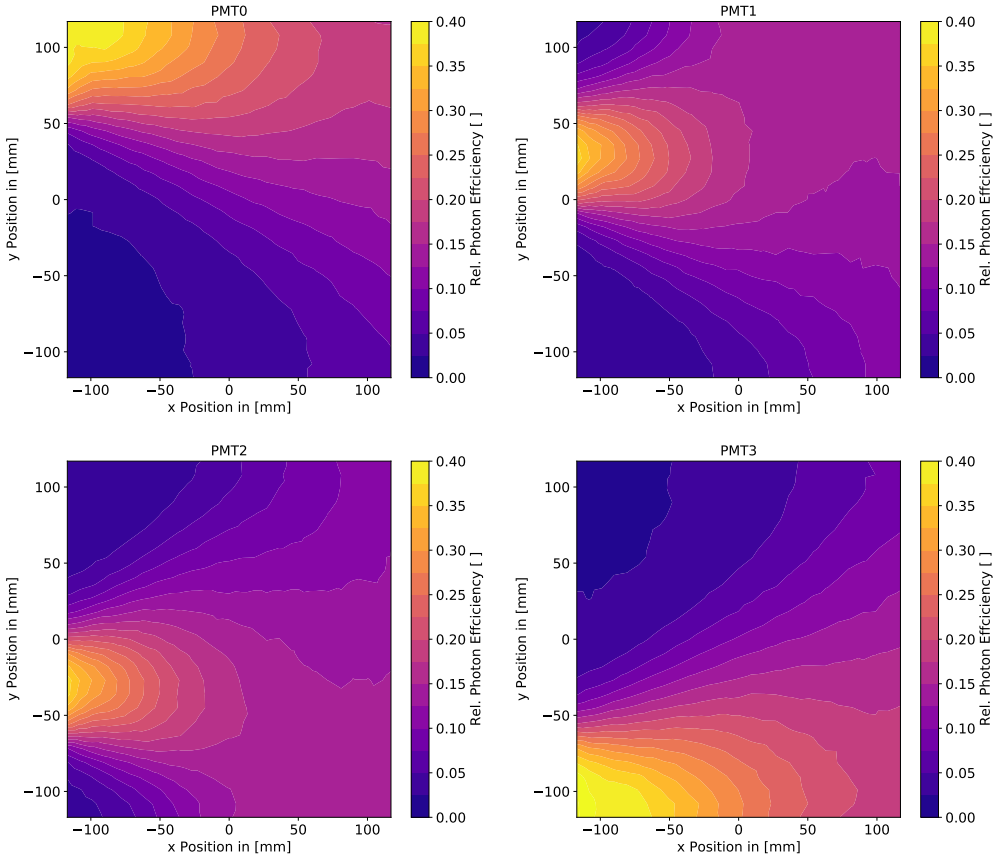


Figure 6.2 Light transmission efficiency, also referred to as photon efficiency, map for individual PMTs. The resulting map shows the causes of the skewed distribution over the entire scintillator surface for the combined signal of all eight PMTs in Fig. 6.1 due to the asymmetric light guide designs, readout at the sides, and surface roughness. The simulated light sources are 3 mm below the scintillator surface, the surface roughness parameter is $\rho = 0.8$, and the results are normalized to the value in the center. These individual PMT maps can be used for advanced trigger models, see Sec. 6.1.3 and [Roi18].

Glisur model that parametrizes random micro-facets tilted against an ideal surface, leading to light extraction. We modeled the light transmission efficiency for a grid of points in the plane perpendicular to the incoming electron direction at a fixed depth. The resulting light transmission efficiency maps are shown in Fig. 6.1 when combining all eight PMTs of one detector and for individual PMTs in Fig. 6.2 for a light source depth of 3 mm.

The light transmission efficiency map for a whole detector in Fig. 6.1 shows a slight bathtub shape horisontally with higher light transmission probabilities closer to sides with the light guides. The light extraction rapidly increases close (about 10 mm) to the light guides and more so for one of the two components making up a light guide for one PMT. These effects lead to a local asymmetry and a skewed distribution over the entire scintillator surface. The structure of the light transmission efficiency map is similar for other simulated light source depths and mainly changes in the transmission probabilities.

For individual PMTs, the light transmission efficiency maps in Fig. 6.2 are normalized to the center value of that of the whole detector in Fig. 6.1. The observed spatial dependence verifies the observations in Fig. 6.1 and shows similar behaviors. These individual PMT maps are essential for advanced trigger models, see Sec. 6.1.3 and [Roi18].

The observed uniformity with the ^{113}Sn grid measurements in Sec. 5.4 and Fig. 5.16 is locally weighted with the electron distribution resulting from the magnetic field setup in PERKEO III, see Sec. 6.1.2. Compared to the ^{113}Sn grid measurements, the simulated light transmission probabilities are closer to the observed values than before in the PERKEO III collaboration, highlighting the quality of the detectors. However, deviations remain.

The symmetry and uniformity-based objective value for fine-tuning the individual PMT gain factors in Sec. 3.3.2 removes any nuances during the optimization by design. We tried using the simulated light transmission probability maps as ground truth to be used for the optimization. However, the simulated maps depend on the surface roughness parameter that we could not verify with simulations alone [Bes22] and need to be targeted with future experimental work to refine future analyses. Also, a critical result of [Bes22] is that the different surfaces (scintillator top and sides and light guides) should have individual surface roughness parameters. This result makes sense, as the surfaces are manufactured differently, e.g., milled and polished, but it was neglected in previous work. Due to the remaining ambiguity, we used a surface roughness parameter of 0.8 for all parts for the simulated light transmission probability maps, in agreement with estimations in previous work [Ber18].

Additionally, we tried to determine whether $t = t(\mathbf{x})$ is also energy-dependent $t(\mathbf{x}, E)$, as we must be aware of any non-linearity to determine the Fierz interference term b accurately. Thus, the author of [Bes22] used simulations to estimate the energy deposition at different depths for the electron incident angles and combined the light transmission efficiency maps for different depths as weighted sums for the energies of the calibration sources. Using the distribution of electrons on the detector, see Sec. 6.1.2, we estimated the light transmission probability for a given energy averaged over the corresponding spatial extent. However, we found no significant non-linearity above 300 keV, and for lower energies, we observed a dependence on the surface roughness parameter that could not be verified with simulations alone [Bes22]. Therefore, further experimental work is needed to determine the surface roughness parameters accurately.

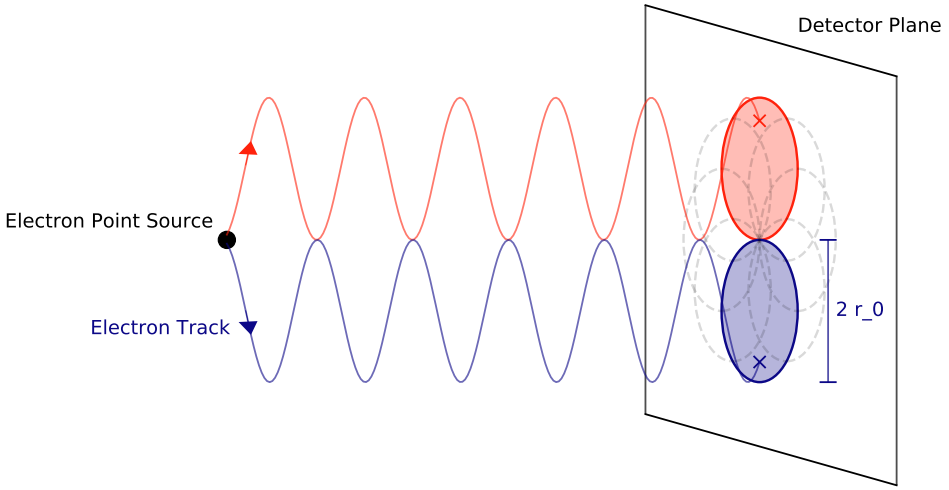


Figure 6.3 Schematic to illustrate the spatial distribution of an electron point source in a homogeneous magnetic field on a detector, the so-called point-spread function. The energy and emission angle-dependent gyration of the electrons lead to a symmetric distribution on the detector that diverges in the center and has a maximum extent of twice the gyration radius r_0. Studying this effect analytically or with simulations is crucial for the analysis of systematic effects in PERKEO III. The magnetic field gradient is neglected in this schematic.

6.1.2 Spatial Dependence of the Detector Gain

Point-Spread Function

Due to the magnetic field, charged decay particles propagate with gyrating tracks to the detectors, and due to the gradient of the magnetic field, their angular distribution is forward-focussed, see Equ. (4.2) in Sec. 4. Generally, we require simulations with a model of the magnetic field setup to describe the electron propagation and distribution at the detectors accurately. However, an analytical approach enables studying individual effects directly, isolated from other processes.

As studying the electron distribution at the detectors is essential for analyzing PERKEO III systematic effects, this process has been studied extensively within the PERKEO III collaboration, e.g., [Dub+14; DS16; Dub15; Roi18; Bes22]. For a point-like source with position x_0, the probability density function ϕ for a position x on the detector surface is called point-spread-function (PSF) as illustrated in Fig. 6.3. It diverges around x_0 on the detector, has a maximum extent of twice the gyration radius r_0 in Equ. (4.1), and is energy-dependent. It is sufficient to neglect effects for the calibration sources with a fixed distance z to the detectors as studied in [Roi18], due to the long distances relative to the gyration radii. In collaboration with [Bes22], I use a Monte-Carlo-based simulation to model the PSF ϕ for different energies and the magnetic field setup as in the PERKEO III 2019/20 measurement.

Calculating Detector Gain For Neutron Beam Data

Contrary to the point-like calibration sources, the neutron beam in PERKEO III has a spatial extent transversal to the beam direction, see Sec. 4. This spatial extent leads to a broader electron distribution on the detectors for electrons from neutron beta decay than for the electrons from the calibration sources. As the spatial response of the detectors is not perfectly uniform, i.e., the light yield is not constant over the detector surface, a broader

distribution receives a larger effective gain g_s. Hence, to use a gain g determined with the calibration sources for the neutron measurement, we must correct g for a neutron beam measurement with a correction factor as $g_s = c_s \cdot g$. Correcting this effect as a pure spatial gain correction c_s neglects that the increased gain for a broader distribution stems from a higher light transmission efficiency and therefore a larger value of the light yield f_{pe} in Equ. (6.1). This gain correlation thus also requires a correction of the light yield as $f_{pe,s} = c_{f,s} \cdot f_{pe}$. In previous PERKEO III work, the spatial light yield correction was neglected, as the neutron spin-dependent beta asymmetry is not sensitive to it.

First, we estimate the electron distribution from neutron beta decays on the detector. Using the neutron beam profile measurements in Sec. 4, we construct a trapezoidal approximation of the distribution at the start and end of the central decay volume and linearly interpolate between these measurements along the neutron beam axis as described in [Bes22]. Using the Monte-Carlo-based simulation of the PSF ϕ, covering all electron emission angles and a uniform electron energy distribution, we obtain a data set with impact positions (x, y) and energies E. We create a subset of the data set weighted by the probability density function of the electron spectrum from neutron beta decay in Equ. (2.1).[1]

Secondly, I construct a model for the spatial response from the measurements with ^{113}Sn. I estimate the uniformity function L to reconstruct the true underlying uniformity function $t(x)$ at scan point x_0 as

$$L_{Sn}(x_0) = \int_{-\infty}^{\infty} t(x) \cdot \phi_{Sn}(x - x_0)\, \mathrm{d}x,$$

with the PSF ϕ_{Sn} for electrons with the ^{113}Sn peak energy as weighting function for the integral. We can rewrite this integral using the equivalent uniform distribution of ϕ_{Sn} with the standard deviation σ_{Sn} as

$$L_{Sn}(x_0) = \frac{1}{2\sigma_{Sn}} \int_{-\sigma_{Sn}-x_0}^{\sigma_{Sn}-x_0} t(x)\, \mathrm{d}x.$$

For a different source with a different energy and therefore different PSF ϕ_1 and standard deviation σ_1, the integral becomes

$$L_1(x_0) = \frac{1}{2\sigma_1} \cdot \left(\sigma_{Sn} L_{Sn}(x_0) + \int_{\sigma_{Sn}-x_0}^{\sigma_1-x_0} t(x)\, \mathrm{d}x + \int_{-\sigma_1-x_0}^{-\sigma_{Sn}-x_0} t(x)\, \mathrm{d}x \right).$$

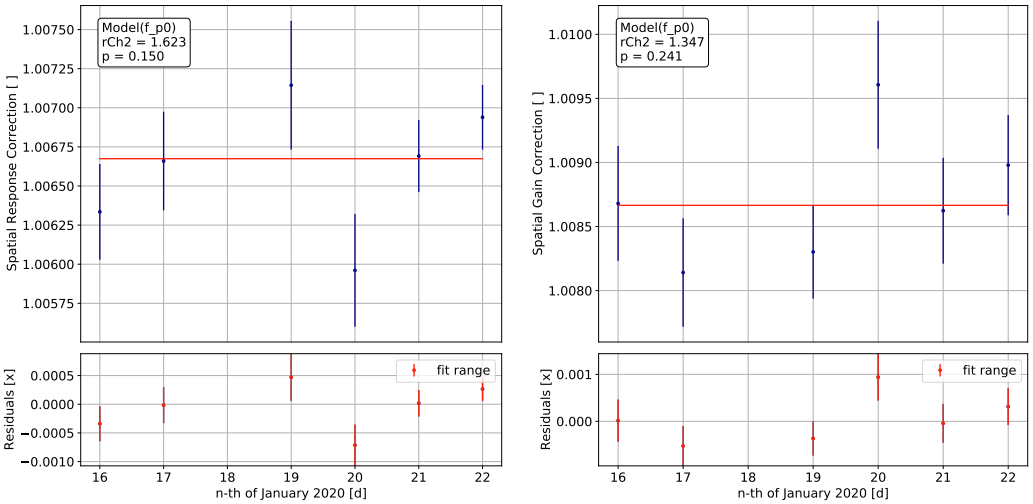
To finally calculate the relative uniformity $U(x)$ at position x using the ^{113}Sn spatial response measurements as

$$U(x) = \frac{L_1(x)}{L_{Sn}(x_0)}.$$

To accurately determine σ_i, I use a data set of Geant4 simulations of PERKEO III of the spatial distributions on the detectors for the calibration sources from [Ber24].[2] I use $t_h = a \cdot \cosh(x)$ and $t_v = a \cdot \cosh(y) - b \cdot y^4$ to approximate the horizontal and vertical shape of the uniformity distribution t from the ^{113}Sn spatial response measurements with fits. Separating

[1] This correction is not sensitive to recoil corrections or BSM contributions of the electron beta spectrum.

[2] Ideally, we would also like to use such a simulation for the electron distribution from the neutron beam. However, an accurate beam distribution requires extensive modeling of the pulsed beam with time resolution, the beam line components, and the output distribution of the neutron guide itself. Such simulations are planned for future work in [Ber24].



(a) Detector 0 (Upstream)

(b) Detector 1 (Downstream)

Figure 6.4 Testing the stability of the spatial response correction over time to extract a correction for each detector. The correction is stable for each detector within statistical fluctuations. This test also verifies the form-stability of the individual PMT fine-tuning from Sec. 3.3.2. The correction is calculated for each detector for all available ^{113}Sn spatial response measurements between the 16th and 22nd of January 2020.

the correction in this approach is only possible due to the symmetry of the spatial detector response after fine-tuning the individual PMTs, see Sec. 3.3.2.

Finally, I use the generated data set for the electron distribution on the detector from neutron beta decay and the relative uniformity $U(x)$ to calculate the spatial gain correction $c_{\mathrm{s},i}$ for each detector from one ^{113}Sn spatial response measurement as $c_{\mathrm{s},0} = 1.00694(21)$ and $c_{\mathrm{s},1} = 1.00898(39)$ assuming a conservative total positioning uncertainty of ±5 mm in the central decay volume, and combining the fit parameter uncertainties, horizontal and vertical estimates and both detectors.

The individual PMT fine-tuning factors from Sec. 3.3.2 optimize for a symmetric and uniform spatial detector response. However, we must verify that the resulting spatial response has a stable form and does not add a systematic effect. Therefore, I calculate the spatial gain correction $c_{\mathrm{s},i}$ for each daily ^{113}Sn spatial response measurement to verify that it is constant over time within statistical fluctuations, as the correction should be caused by geometric effects alone. The spatial gain corrections for each detector are plotted for each day with an available ^{113}Sn spatial response measurement in Fig. 6.4. Both corrections are constant within statistical fluctuations over the seven days, implying that the individual PMT fine-tuning maintains form stability. The mean spatial gain correction $c_{\mathrm{s},i}$ for each detector is then

$$\overline{c}_{\mathrm{s},0} = 1.00667(11)$$
$$\overline{c}_{\mathrm{s},1} = 1.00867(17).$$

The correction value of detector 1 is larger than that of detector 0, as detector 1 is less uniform than detector 0, and the broader electron distribution from the neutron beam has a higher impact. The combined total systematic uncertainty $\sigma_{\mathrm{sys}}(g)$ on the detector gain g from the spatial response correction is $\sigma_{\mathrm{sys}}(g) = 2.0 \cdot 10^{-4}$ ch/keV.

Repeating the analysis for the spatial light yield correction $c_{\mathrm{f,s}}$ to construct a light yield uniformity and the generated data set for the electron distribution on the detector from neutron beta decay, I obtain for each detector i the correction as

$$c_{\mathrm{f,s},0} = 1.00210(20)$$
$$c_{\mathrm{f,s},1} = 1.00264(25).$$

The spatial response correction must be corrected for the gain and light yield together. Therefore, this correction is non-negligible for a precise measurement of the Fierz interference term b and I discuss its significance in Sec. 8.

Induced Edge Effect

Due to the broad electron distribution on the detectors for electrons from neutron beta decay, some electrons with large gyration radii $r_0(E)$ can miss the scintillator and are not detected. This effect leads to an energy-dependent spectral correction, affecting the end-point region of the electron beta spectrum, and is called *edge effect*. Contrary to previous PERKEO III measurements, the scintillator cross section is reduced in size to optimize for uniformity and background reduction. They no longer ensure a full coverage of the electron distribution, see [Raf16] for a discussion. In addition, contrary to the spatial gain correction estimation, this analysis is strongly affected by recoil corrections or BSM contributions of the electron beta spectrum. This analysis requires a more accurate model of the spatial distribution on the detector, as the correction calculation only consists of events in the tails of the spatial electron distribution. Such an accurate model will be studied with PERKEO III Geant4 simulations in [Ber24].

6.1.3 Trigger Function

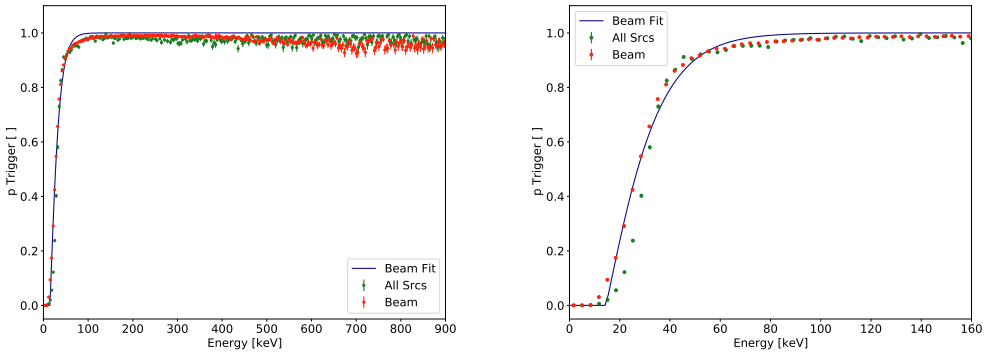
The readout system registers an event when two out of the eight PMTs of each detector measure a signal above the trigger threshold, as described in Sec. 4. For larger amplitudes A, i.e., electron energies, the probability for the readout system to trigger should be 100%. For amplitudes closer to the trigger threshold, the probability decreases until it reaches zero. The probability p of a single photon for an amplitude A in channel is summarized in the empirical trigger function [Mun06] from the binomial distribution as

$$T(A) = 1 - (1-p)^{a\cdot A}\left(1 + \frac{a\cdot A\cdot p}{1-p}\right), \tag{6.3}$$

with the number of photons per measured amplitude a. The analysis in this thesis builds upon the *p3fit* tool that uses the trigger function as described in Equ. (6.3). This model does not accurately describe counting statistics and non-linearity [Roi18]. Besides reducing the detected rates for lower energies, the trigger function affects types of events, see Fig. 4.8, differently, as they differ in how their energy is distributed between detectors. In particular, undetected backscattering leads to systematic energy losses that we must correct for measurements with PERKEO III [Roi+19].

We can construct $T_i(A)$ for detector i from binned amplitudes A from measured data by dividing the following spectra:

$$T_i(A) = \frac{P_{\mathrm{B}}(A)}{P_{\mathrm{O}}(A) + P_{\mathrm{B}}(A)},$$



(a) Full energy range from 0 – 900 keV.

(b) Low-energy range from 0 – 160 keV.

Figure 6.5 The fitted trigger function in Equ. (6.3) to neutron beam data for the upstream detector and measured data from neutron beam data and a combined calibration with all five sources. The energy scale is given in keV and for orientation, but it neglects the non-linearity of the energy-channel relation. The fitted model and the observed data deviate significantly. Potential causes are the short trigger time window of 180 ns and the higher light yield of the detectors, requiring advanced trigger models.

with $P_{\mathrm{B}}(A)$ being the measured spectrum of detector i where both detectors have triggered and $P_{\mathrm{O}}(A)$ being the spectrum of detector i where only the opposite detector $(1-i)$ triggered. The resulting trigger functions $T(A)$ from beam data and from a combined calibration with all five sources are shown in Fig. 6.5, as well as the fit of the empirical model in Equ. (6.3) to beam data. The energy scale is given in keV for orientation, but it neglects non-linearity.

As can be seen in Fig. 6.5, both data sets do not reach 100% trigger probability for large energies and therefore deviate from the expectation. This deviation implies either contamination in $P_{\mathrm{O}}(A)$ or missing events in $P_{\mathrm{O}}(B)$, or both. The trigger time window is the time interval for the individual detector triggers to be registered as one double trigger event. Ideally, the trigger time window should be as long as the QDC integration time of 330 ns. We observe that the trigger time window in the measurement campaign is about 180 ns, which is smaller than the trigger time window from previous PERKEO III measurements, e.g., [Sau18].[3] The shorter trigger time window affects $P_{\mathrm{B}}(A)$ and $P_{\mathrm{O}}(A)$, effectively "moving" events from $P_{\mathrm{B}}(A)$ to $P_{\mathrm{O}}(A)$. This effect could lead to deviations and the implications of a shorter trigger time compared to a longer integration time must be studied with a complete simulation of PERKEO III with higher-order backscattering and electron time-of-flight, as is currently done for [Ber24].

The goodness-of-fit for the trigger function in Equ. (6.3) deviates significantly and statistically rules out the model. The fitted trigger function deviates for both data sets especially for energies below 140 keV. The trigger functions from both data sets decrease faster than the fitted model with a larger non-linearity between 50 and 140 keV but maintain a higher trigger probability between 30 and 50 keV. This deviation between the trigger function from data and the fitted model is probably caused by the significant increase in light yield n_γ of the 2019/20 PERKEO III detectors, see Sec. 4. A greater light yield invalidates the simplification in Equ. (6.3), as more photons are distributed to multiple PMTs and not all PMTs have the same probability for a photon to reach them, see Fig. 6.2. A better trigger function model

[3] The trigger time window is independent of the QDC integration time window. The shorter trigger time window was an oversight during the measurement campaign.

using the spatial light transmission probabilities for individual PMTs and accurate photon statistics is described in [Roi18] and should improve the description of the trigger function below 140 keV.

The deviation between the trigger function constructed from beam data and the five calibration sources combined has two causes. For one, the calibration sources emit correlated multielectron events that increase $P_{\mathrm{O}}(A)$ more than $P_{\mathrm{B}}(A)$ for lower energies where $p < 0.8$. Also, the calibration source holders lead to absorption and energy losses that perturb the distributions, see Sec. 6.2.

While the fitted trigger function does not accurately explain the observed data, we can still extract trigger function parameters and verify stability over time to some extent. Fitting the trigger function to beam data over the core measurement period in January 2020, the fitted trigger function parameters are constant within statistical limits. We can extract the trigger function values for each detector for the probabilities p_i as

$$p_0 = 0.8106(17)$$
$$p_1 = 0.7497(23)$$

and for the conversion parameter a_i as

$$a_0 = 2.014(6) \cdot 10^{-3}$$
$$a_1 = 2.158(9) \cdot 10^{-3},$$

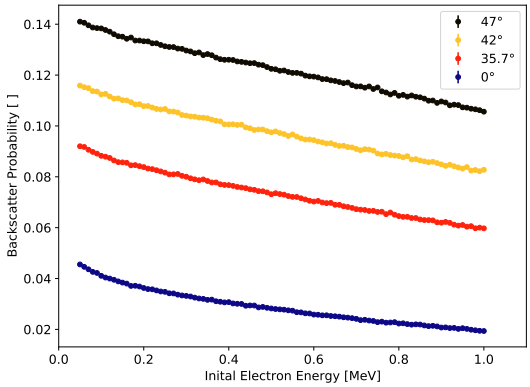
neglecting the bad goodness-of-fit of the trigger model. While it does not explain the observed data, the fit results are stable, indicating promising results with the right trigger function model. An estimation of the induced systematic uncertainty from these parameters on the Fierz interference term b does not make sense.

In conclusion, we currently do not have an accurate model to describe the trigger function behavior of the measurement. Such a model is essential to analyze systematic effects related to undetected backscattering [Roi+19; Roi18]. The authors of [Roi+19] calculate a systematic correction shift of about $-2.6 \cdot 10^{-3}$ for measuring the Fierz interference term b with PERKEO III from an electron beta spectrum measurement with a fit range from 180 keV upward. The authors also noted that the correction stems from incomplete energy reconstruction, and a decreased trigger probability would have negligible influence. Currently, we cannot verify this estimate and related systematic uncertainties. Furthermore, the inaccurate fitted trigger function model should lead to visible deviations of the theoretical models and measured data in Sec. 7, especially for energies below 200 keV, considering backscattering.

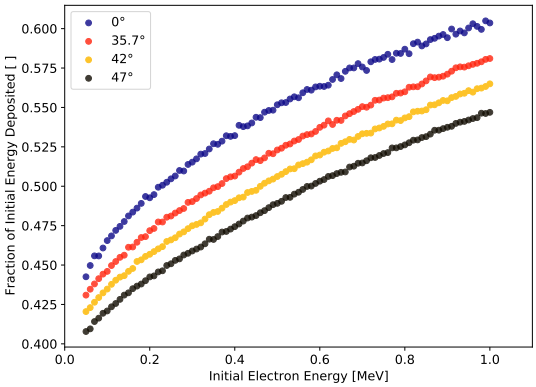
Future work can significantly improve these issues and enable analysis for energies below 200 keV by studying the application of the advanced trigger function in [Roi18] to this measurement. Also, the effect of the shorter trigger time window must be studied, requiring a detailed simulation of PERKEO III with magnetic field setup, which will be studied in [Ber24].

6.1.4 Electron Backscattering

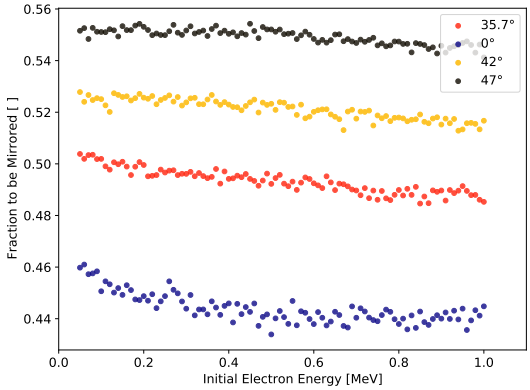
Using the simulation framework Geant4, we studied electron backscattering off a scintillation detector and its effects for the Fierz interference term b with the setup described in [Fal22]. The goal was to understand electron backscattering as an isolated process and its



(a) Probability of an electron to backscatter depending on its initial energy and for different incident angles θ perpendicular to the detector plane. The probability decreases with steeper incident angles and higher energies, as the electrons propagate deeper into the scintillator and the scattering cross-section increases at lower energies.



(b) Average fraction of the initial electron energy being deposited on the primary detector when backscattering. The average fraction increases for higher energies and steeper incident angles, as electrons that have penetrated deeper into the scintillator also deposit more energy when scattering back out.



(c) The fraction of backscattered electrons being reflected due to the magnetic mirror effect back onto the primary detector. This effect mainly depends on the initial electron angle, as it affects the angular distribution when backscattering.

Figure 6.6 Results of the simulated backscattering studies using data and code from [Fal22]. The initial electron angular distribution at the detector has its maximum at $\theta_c = 47°$ and the expected value $\hat{\theta} = 35.7°$.

parametrization for this analysis. For fixed incident angles θ of the electrons to the surface normal of the scintillator, different quantities can be studied, e.g., the probability of an electron to backscatter, the emission angle of an electron after backscattering, and the fraction of the original energy being deposited in the scintillator. The magnetic field is assumed to have two values: the magnitude at the detectors or in the central decay volume. The results can then be expanded with the angular distribution of electrons due to the magnetic field gradient and the two-detector setup in PERKEO III to estimate the probability of electrons being reflected by the magnetic mirror onto the same detector.

Using the data sets and analysis code from [Fal22], I reconstruct different backscattering quantities in Fig. 6.6. The backscatter probability of an electron depends on its incident angle θ and its energy. For steeper incident angles and higher energies, the probability decreases, as the electrons propagate deeper into the scintillator, making backscattering less likely. Also, the stopping power increases in the scintillator for lower energies. As a result, the maximum angle of entry $\theta_c = 47°$ due to Equ. (4.2) has the maximum backscatter probability of approx. 14%. The expected value of the incident angle $\hat{\theta} = 35.7°$ has a backscatter probability of approx. 9% and decreases to 7%.

Studying the mean fraction of the initial electron energy being deposited when backscattering in Fig. 6.6b, a similar effect is observed, as electrons that have penetrated deeper into the scintillator also deposit more energy when scattering back out. Although not shown, the fractions of the initial electron energy being deposited range from 0 to 1 for all energies. The results are also in agreement with previous PERKEO III results of electron backscattering [Roi18; Roi+19]. Based on the results in Fig. 6.6, I create a surrogate model of the backscattering process to be used for the simulation framework *Smelt* in Sec. 6.1.5.

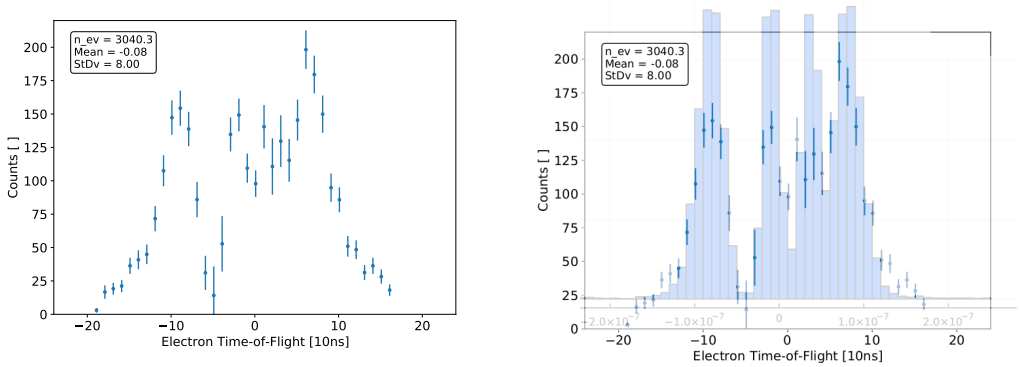
As mentioned in Sec. 4 and 6.1.3, undetected backscattering leads to undetected electron losses and skews the measured energy spectra to lower energies. Based on the calculated, systematic effect of undetected backscattering on a Fierz interference term b measurement with PERKEO III in [Roi+19], we must accurately determine the impact on this measurement. Analyzing this effect to reach lower fit ranges and extracting accurate and precise corrections from the measured electron energy spectra is essential. Such studies are conducted for [Ber24] and must expand on the results in [Roi18; Fal22] with more elaborate simulations of the PERKEO III experiment, including multiple backscattering, angular distributions for higher-order backscattering, and a more accurate model of the magnetic field without the simplifications for the calculations here.

6.1.5 *Smelt* Simulation Tool

Motivated by the observed phenomena related to detector timing with the trigger function and electron backscattering, I developed a *Julia*-based simulation tool called *Smelt*[4] (SiMulation of ELectron detection Timing) to study coincidence behavior and other effects for calibration source measurements with PERKEO III. The goal was to understand the electron time-of-flight spectra of the calibration sources, estimate rates for different event types, and study systematic effects that require detection time resolution.

Compared to a full PERKEO III simulation with Geant4 as currently studied for [Ber24], *Smelt* cannot calculate systematic corrections to be used on the Fierz interference term b directly. However, *Smelt* is designed for rapid hypothesis testing and systematic effect estimation to reduce the ambiguity of results of more extensive simulations that are hard to verify. This

[4] https://github.com/maxlampe/smelt



(a) Measured ^{109}Cd electron time-of-flight spectrum. **(b)** Measured and simulated ^{109}Cd spectrum overlaid.

Figure 6.7 Number of detected events for the differences in trigger times of the two detectors, the so-called electron time-of-flight spectrum, of ^{109}Cd. The *Smelt* generated simulation data with correlated electron emissions explains the multipeak structure and time-of-flight distances inbetween them. The asymmetric peak positions are explained by the CaliBot position about 50 cm out-of-center towards the downstream detector. The simulation is shifted by one 10 ns time-of-flight bin.

is a crucial approach for high-precision measurements, as large, complex simulations can induce their own biases.

The tool models the detection timing response with a 5 ns resolution, the angular distribution of electrons at the detectors, and backscattering with the parametrization and interpolation from Sec. 6.1.4. It also calculates the relativistic electron time-of-flight in PERKEO III for backscattering with and without the magnetic mirror effect, applies the Birks non-linearity from Sec. 6.1, and considers the fitted trigger function from Sec. 6.1.3 for the detection. Calibration spectra are approximated with Gaussian distributions and the possibility of correlated electron with instantaneous or stochastic emission times.[5] This approach simplifies the complexity of the electron emission spectra of the calibration sources but is sufficient for first-order studies of the correlated emissions to reveal the intricate time-of-flight dependence.[6] The distribution of energies to individual PMTs is done based on the light transmission simulations from Sec. 6.1.1. Furthermore, *Smelt* has a broad set of physically motivated unit tests to verify the results for individual processes and calculations.

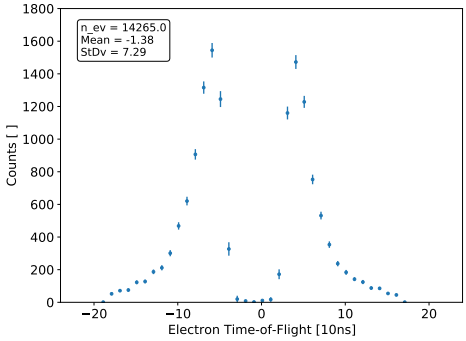
With *Smelt*, we have a simulation tool to understand the detector response based on the timing information and energy distribution between events and to study the different event types in Fig. 4.8. For example, we can model the energy-dependent electron time-of-flight and the energy distribution between detectors to set a limit on accidental coincidences and verify correlated electron emissions. The listed processes can be switched on or off to precisely link observed phenomena and effects to individual processes. The *Smelt* tool is used for the analysis in Sec. 6.1.6, 6.1.7, and 6.1.8.

6.1.6 Electron Time-of-Flight

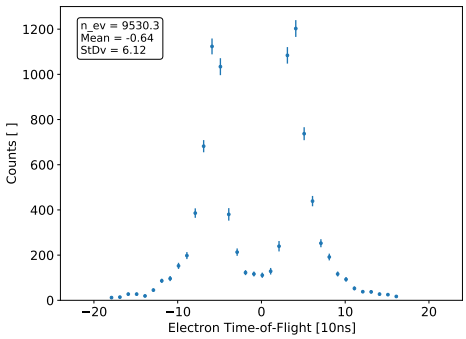
In the previous analyses in the PERKEO III collaboration, the electron time-of-flight spectra from the calibration sources were not studied in detail. For this analysis, the understanding

[5] Using a Gaussian distribution for the total detector signal as the sum of eight Poisson distributions also works well on measured data.

[6] See Sec. 7.1 for an overview and especially [Roi18] for a rich discussion of accurate and detailed spectra generation of calibration sources.



(a) Measured electron time-of-flight spectrum from neutron beam data with no accidental coincidences or correlated electron emissions at zero times-of-flight.



(b) Measured ^{113}Sn electron time-of-flight spectrum with correlated electron emissions around zero times-of-flight.

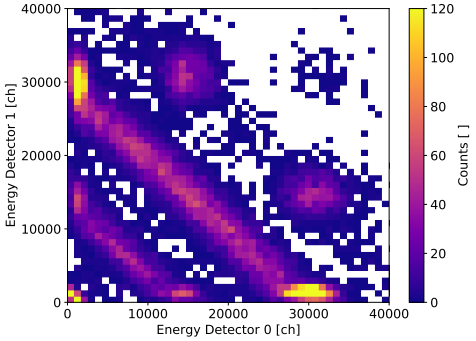
Figure 6.8 Measured electron time-of-flight spectrum from neutron beam and a ^{113}Sn measurement. A naive expectation is that there should be no events for zero times-of-flight due to the finite maximum velocity of the electrons. Compared to the conversion electrons from the main peak, the correlated Auger electrons in ^{113}Sn do not have enough energy to create times-of-flight shorter than 30 ns. This observation led to discovering previously neglected systematic effects of the calibration source holder in Sec. 6.2.

of the calibration sources is more relevant, as the detectors have a higher energy resolution, are sensitive to potential new systematic effects, and have a shorter trigger time window than before.

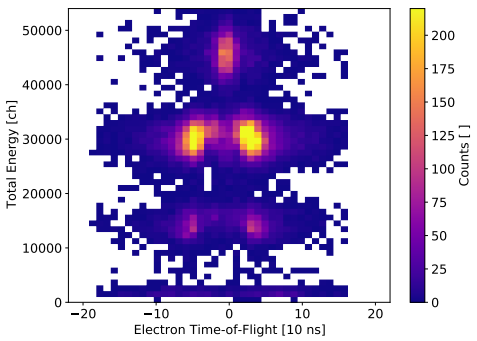
To visualize the analysis of electron time-of-flight and different event types, I studied measured and simulated 1D and 2D spectra where both detectors have triggered. This double-trigger requirement limits the analysis to a subset of events but also yields a more precise lever on phenomena like accidental coincidences, and it completely removes the requirement of background subtraction. Regarding nomenclature, I refer to the difference in trigger times between detectors as times-of-flight for historical reasons, which is accurate for backscattering events between detectors but less straightforward when referring to trigger time differences from two separate electrons as, e.g., in accidental coincidences. The *Smelt* simulations in this section include all processes except the underlying non-linearity.

When looking at the measured electron time-of-flight spectra of calibration sources, in particular, ^{109}Cd in Fig. 6.7 or ^{207}Bi in Fig. 6.9b, we observe multiple time-of-flight peaks. From a naive perspective, one would expect two peaks symmetric around zero times-of-flight with a time distance given by the spatial distances in PERKEO III from backscattering and the relativistic velocity limit. The electron energy distribution then sets the shape and width of these two peaks. There should be no events around zero times-of-flight due to the finite maximum velocity of the electrons. The time-of-flight is positive or negative depending on which detector triggered first in case of backscattering. We observe this structure for the electron time-of-flight from neutron beta decay in Fig. 6.8a.

With *Smelt*, we can explain the observed structure in a principled way. It is straightforward to prove that single electron emissions cannot explain multiple peaks, as they follow the naive expectation. The calibration sources also emit, e.g., Auger electrons, whose emission is correlated with the conversion electron emission. Some excitation states have short lifetimes, also leading to correlated electron emission given the length of our QDC integration time. For ^{109}Cd, we can model the greatly simplified energy spectrum with a Gaussian peak and



(a) Measured 2D histogram of the energies detected in each detector for ^{207}Bi. The diagonal lines between the peaks show the fractions of energy deposited in the case of backscattering.



(b) Measured electron time-of-flight and detected energy in a 2D histogram for ^{207}Bi. Marginalizing the histogram onto the x-axis creates the equivalent time-of-flight plot as for ^{109}Cd in Fig. 6.7.

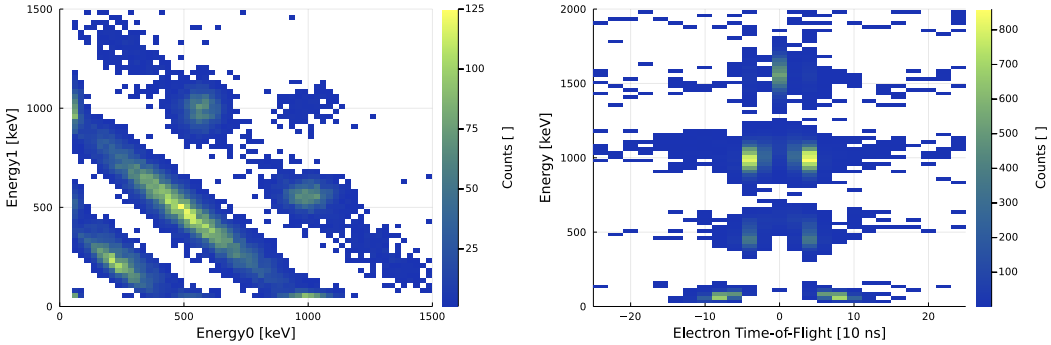
Figure 6.9 Measured ^{207}Bi spectra with a double-trigger requirement to study different event types in, e.g., electron time-of-flight studies. The correlated electron emissions from the 995 keV and 503 keV due to a 12 ps lifetime of the excited nuclear state forms islands (left) and a peak around zero times-of-flight (right). Accidental coincidences form also isolated islands (left), as for two accidental 995 keV electrons, and broad bands to longer times-of-flight for each peak (right). The short trigger time window of 180 ns is visible as electron time-of-flight cut-off on the right.

add instantaneous, correlated electron emission with a low percentage and the corresponding Auger electron energies. With *Smelt*, we simulate the detection response with timing and energy resolution and reconstruct the measured ^{109}Cd electron time-of-flight spectrum in a first-order approximation. As seen in Fig. 6.7, the simulated time-of-flight spectrum creates matching asymmetric peak positions and time-of-flight distances, verifying that the correlated electron emissions explain the multipeak structure of the time-of-flight spectrum. Additionally, this result explains the asymmetric in the peaks solely with the CaliBot position about 50 cm out-of-center downstream towards detector 1 in the central decay volume.

However, the widths of the time-of-flight peaks in Fig. 6.7 differ. Introducing decay times to the correlated electron emissions of about 40 ns would produce matching peak widths, but they are not physically meaningful as these nuclear processes work on shorter time scales. Initial Geant4 simulations of a full PERKEO III setup show that a more detailed description of the magnetic field setup explains the broader time-of-flight peaks and longer tails. This result highlights the importance of future work in [Ber24] for precise studies of correlated emissions with low energies in PERKEO III.

To further study correlated electron emissions, I look at the ^{207}Bi electron time-of-flight and energy detection, as shown in Fig. 6.9. Unlike ^{109}Cd, ^{207}Bi has three main energy peaks with energies up to 1 MeV. Thus, we must look at 2D plots that also consider the detected energy and its distribution between the detectors to distinguish different electron emissions due to the finite velocity of the electrons that causes overlay of the electron time-of-flight peaks for larger energies.

In Fig. 6.9b, I show the measured ^{207}Bi electron time-of-flight plotted against the detected energy summed over both detectors. The trigger time window of 180 ns is visible as a electron time-of-flight cut-off. Projecting the histogram onto the x-axis creates the equivalent time-of-flight plot as for ^{109}Cd in Fig. 6.7. In doing so, the finite electron velocity is visible for times-of-flight from backscattering for larger detected energies in Fig. 6.9b. Also, we observe



(a) Simulated 2D histogram of detected electron energy between detectors for ^{207}Bi with *Smelt*.

(b) Simulated 2D histogram of electron time-of-flight and detected energy for ^{207}Bi with *Smelt*.

Figure 6.10 Verifying the hypothesis with the simplified models in *Smelt* of Fig. 6.9 by toggling modeled processes, e.g., removing the correlation between the high-energy peaks removes the respective islands, and reducing the simulated QDC integration time removes the accidental coincidences. The simulated measurement time is shorter than the equivalent real measurement time in Fig. 6.9.

a peak around zero times-of-flight due to the higher energies of the ^{207}Bi electrons. This structure is similar to ^{109}Cd with two peaks, but they overlap for ^{207}Bi due to greater electron energies. This peak is from correlated electron emissions of a conversion electron from the 1 MeV state that transitions with a 12 ps lifetime and about 500 keV in the child nucleus [Bé+16]. These two electrons can go to opposing detectors, creating the time-of-flight peak with the double-trigger requirement.

This effect is also clearly visible when plotting the energy distribution onto the detectors, as in Fig. 6.9a. Close to each axis, we see the expected main peaks detected on each detector with an Auger electron from the opposite detector fulfilling the double-trigger requirement without contributing much energy. The diagonal lines between the peaks stem from the fraction of energy deposited in the primary detector in case of backscattering, as described in Sec. 6.1.4. The most likely fractions are between 40 and 60%. Correlated events form islands in Fig. 6.9a. Also, accidental coincidence events of the 1 MeV electrons are seen as another island or as broad time-of-flight tails in Fig. 6.9b for each energy peak. The broad time-of-flight tail is most prominent for the 1 MeV electron energy peak.

All explanations can be verified with 2D histograms simulated with *Smelt*, like the ones shown in Fig. 6.10. The ^{207}Bi energy spectrum is simplified and approximated with a Gaussian mixture model, and correlations are included as for ^{109}Cd with an additional correlation of the 1 MeV and the 500 keV peak for the 12 ps excitation state. The simulated 2D histograms for the energy distribution between detectors and the electron time-of-flight against detected energy agree with the histograms from measured data.

We can verify our hypothesis by toggling modeled processes in *Smelt*, e.g., removing the correlation between the high-energy peaks removes the respective islands in Fig. 6.9a, and reducing the simulated QDC integration time removes the accidental coincidences. This approach also works for backscattering, as we can remove the diagonals by turning off the backscattering model or reduce the diagonal to points and lines when only allowing fixed angles or using only the mean fraction of energy deposited in the primary detector. Without correlated electron emissions and backscattering, we see single counts of accidental coincidences in the backscattering diagonal of energy deposition between detectors. However,

the number of events is negligible given the calibration source rate, and we can exclude accidental coincidences as relevant contributions to all calibration spectra.

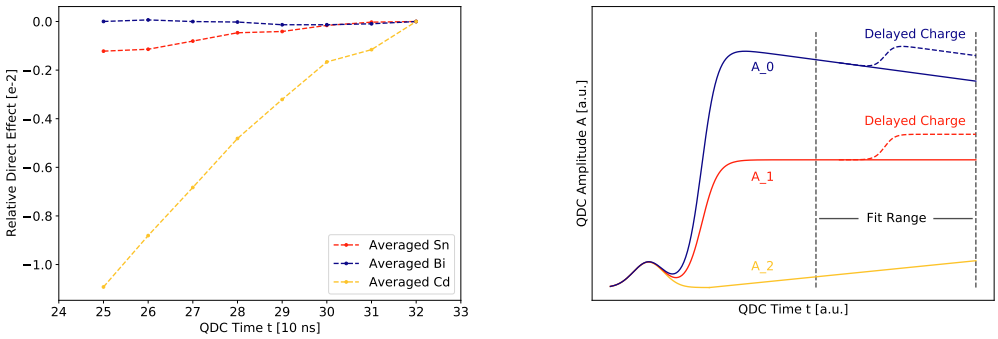
Another observed deviation from the naive expectation in all calibration sources is that there should be no events around zero times-of-flight due to the relativistic velocity limit. Correlated electron emissions can explain this deviation for some calibration sources, e.g., ^{109}Cd and ^{207}Bi. However, these explanations do not work for ^{113}Sn where the bins in the range close to zero times-of-flight contain events, as shown in Fig. 6.8. Although ^{113}Sn also has correlated low-energy electron emissions, the energy difference between the electrons from the main peak with about 369 keV and the Auger electron energies is too large to create such small electron time-of-flight differences between triggers. With *Smelt*, we determine that detecting events in that time-of-flight range is impossible without another electron-creating process with electron energies above 100 keV. This result motivated searches for previously neglected systematic effects from the calibration source holders: Mainly secondary electrons produced in the thin support foil and frame from gamma emissions of the calibration sources and electron losses in the thin support foil and frame. The corrections from these effects explain the observed deviation and are discussed in Sec. 6.2.

The findings of the electron time-of-flight analysis for calibration sources with *Smelt* show the importance of hypothesis testing tools that allow for control beyond what general simulation frameworks offer for high-precision experiments. These general frameworks are essential for constructing precise and accurate results but can also offer hazy explanations and not enough customizability, e.g., partially turning interactions and processes to test a specific hypothesis. In addition, studying the measured data of PERKEO III in 2D histograms was not done before in the collaboration and might offer important levers to study, e.g., the non-linearity as curvature in the diagonal of backscattering events or room for advanced 2D model fitting, see Sec. 6.1.7. A preliminary study to determine the non-linearity from 2D calibration source histograms was inconclusive due to statistical limitations.

6.1.7 Charge Integrator Non-Linearity

Previous work established using the Birks non-linearity from Sec. 6.1 to model the non-linearity of the PERKEO III detector response motivated by the assumption that the scintillator is the source of the non-linearity. The author of [Sau18] measured the Birks coefficient for a scintillator (BC408 [Sai21]) as $k_B = 123(14)$ nm/keV in offline measurements. However, in the most recent PERKEO III results [Mär+19; Sau+20], an effective non-linearity of $k_B = 430.5(10.3)$ nm/keV was observed from fits with calibration sources that use the Birks model to describe the non-linearity. To probe potential causes of this deviation, the author of [Sau18] used offline tests with pulse generator and PMT data to rule out linear fan-outs in the readout signal chain as the cause. The electron readout was assumed to be the cause of the non-linearity. Also, other empirical non-linearity models were tested, mainly a quadratic and an exponential model, in combination with a Birks model with a fixed k_B value from the scintillator measurement, but no model could be selected within statistical certainty. Ultimately, the Birks model described the observed non-linearity with the calibration sources sufficiently well within statistical limits.

Contrary to previous measurements of neutron spin-dependent quantities, we want to precisely measure the pure electron beta spectrum and are more sensitive to non-linearities. To better understand and analyze the QDC integrator behavior and potentially induced non-linearities, we measured a subset of the data during the 2019/20 PERKEO III measurement campaign in the ALLMODE measurement mode, as described in Sec. 4. In this mode, we



(a) Averaged, relative direct effect of the QDC sample time of individual PMT channels on the detected energy for different calibration sources using causal mediation analysis. The direct effect increases linearly with t, and the gradient of the relation is energy-dependent.

(b) Schematic of the hypothesis that the gradient of the QDC integrator sample value A over QDC sample time t is amplitude-dependent and causes a previously unidentified QDC non-linearity.

Figure 6.11 Initial observation of a QDC non-linearity that depends on the integration time and a schematic of the formulated hypothesis (effect exaggerated) for an amplitude-dependent gradient of the QDC sample time t dependence. The gradient is extracted in a channel-specific analysis, and delayed charge events are treated separately.

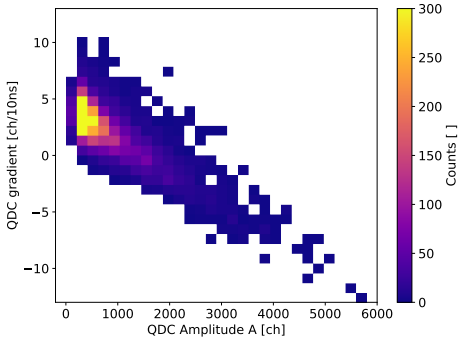
stored each sample of the 100 MHz integrator and measured a complete calibration with all five sources and a few hours of neutron beam data with ^{113}Sn drift measurements.

The signal chain of the measured QDC samples for individual PMT channels to the final peak position fit in the energy spectrum forms a directed acyclical graph, including the data processing and detector response model. Thus, we can view the signal chain as a structural causal model [Pea09b], see Sec. 3.4.1. This view allows us to conduct causal mediation analysis [Pea09a] and calculate the direct effect of a QDC sample at a chosen time t for an individual PMT on the total detected energy. The results of this test are shown in Fig. 6.11a for three calibration sources of different energies. I plot the average relative direct effect for all PMTs to visualize the different sources. The estimated direct effect increases linearly with t, and the gradient of the linear relation is energy-dependent. This energy-dependence of the QDC integrator must be studied further.

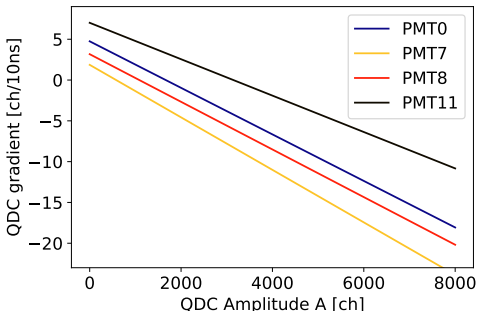
Based on the results of the direct effect calculations, I formulate the hypothesis that the gradient of the QDC integrator sample value over QDC sample time t is amplitude-dependent, i.e., energy-dependent, as illustrated in the schematic Fig. 6.11b, and causes the observed non-linearity.[7] To test this hypothesis, I analyze the ALLMODE data combined for ^{207}Bi, ^{113}Sn, and ^{139}Ce, for the QDC channels of individual PMTs to extract the relation between the QDC gradient m and the amplitude A with pedestal subtraction at the last QDC time. This amplitude A is the value that we also extract in the other measurement modes.

To extract the gradient, I fit a linear model to the last 180 ns of QDC samples for events without delayed charge in the integrator due to, e.g., backscattering or correlated electron emissions. I do a separate linear fit to the last 110 ns of QDC samples for events with delayed charge in the integrator. These cut-offs were estimated to separate events with and without delayed charge and also guarantee enough integration time to avoid including the integration of the signal rise time in this first-order analysis. This amplitude A is corrected for pedestal

[7] A potential cause could be an amplitude-dependent leakage current over the silicon used in the QDC integrator.



(a) Fit result without delayed-charge events for the QDC channel of PMT0.



(b) Regression results for four PMTs. The results of all other PMTs lie between PMT7 and PMT11.

Figure 6.12 Using ALLMODE data from ^{113}Sn, ^{207}Bi, ^{139}Ce, I extract the amplitude-dependence of the QDC gradient with linear regression from this 2D histogram to parametrize the QDC non-linearity in a first-order approximation. The measured amplitudes A are corrected for pedestal and rate dependency. As both corrections are additive, the extracted gradients need no correction. The energy-channel relation is approximately 30 ch / keV for a full detector summing over all PMTs.

and rate dependency, as in Sec. 5. As both corrections are additive in our approach, the extracted gradients need no correction. The relation of the resulting (A, m) pairs is extracted with linear regression and mean-square deviation minimization. The histogram for the (A, m) pairs for the QDC channel of PMT0 and the regression results for the QDC channels of four PMTs are shown in Fig. 6.12. Repeating this test separately for ^{207}Bi and ^{113}Sn data yields similar results, ruling out a higher-order energy or rate dependency.

All QDC channels have a positive offset for small amplitudes, and the QDC gradient m decreases linearly with larger amplitudes. The linear regression offset and gradient determine at which detected amplitude m changes sign. This relation also explains the causal mediation analysis results in Fig. 6.11a, as for smaller detected amplitudes A, the QDC gradient m is positive for all or most QDC channels, leading to an amplitude increase when varying the QDC sample time t. For large amplitudes, this effect reverses.

With a parametrization of the amplitude dependence of the QDC gradient $m_i = m_i(A)$ over integration time t, we can create a first-order correction. With the correction, we reconstruct the detected amplitudes A to the estimated QDC sample value A_c at the start of the fit ranges Δt_i to study the induced non-linearity of the amplitude dependence as

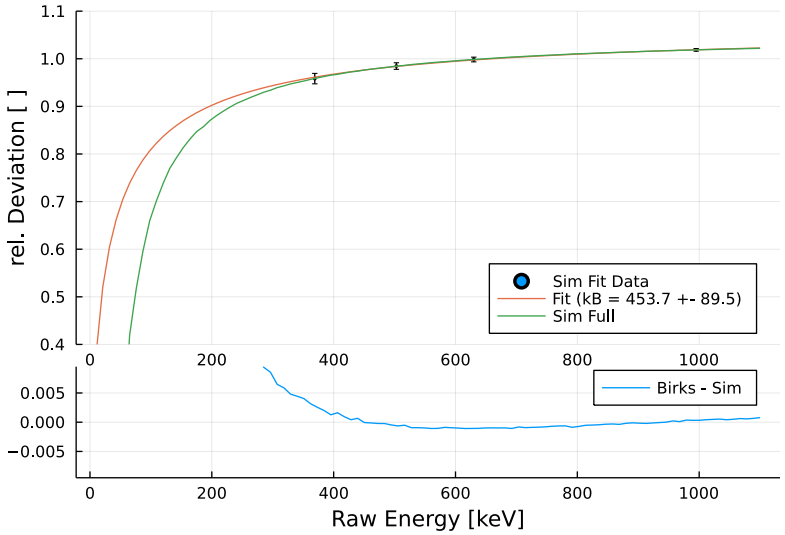
$$A_c = A - m_0 \cdot \Delta t_0 \tag{6.4}$$

$$A'_c = A - m_0 \cdot \Delta t_0 - (m_1 - m_0) \cdot \Delta t_1, \tag{6.5}$$

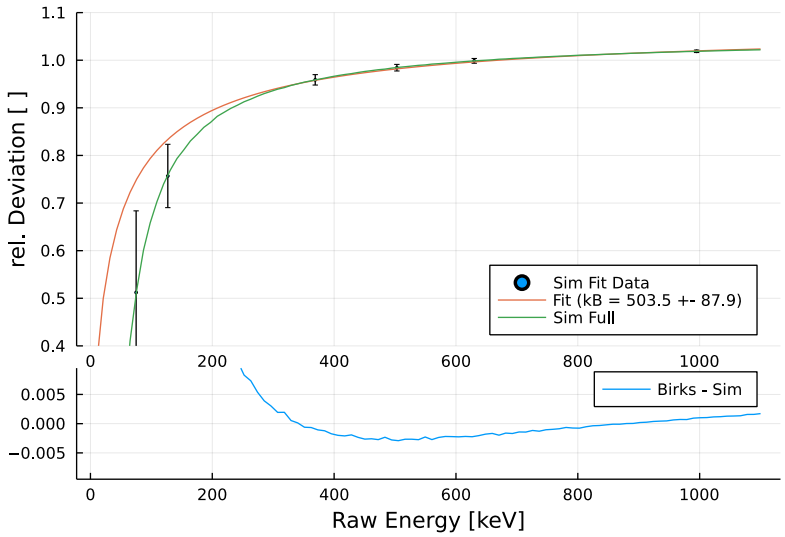
with A'_c being the corrected amplitude for delayed charge events, which make up about 10% of all events. To accurately include the delayed charge events, I use *Smelt* and its modeling of backscattering, energy distribution based on light transmission probabilities, and empirical model of the energy-dependent energy resolution. Also, the Birks non-linearity model with $k_B = 123$ nm/keV is applied as an underlying scintillator non-linearity.[8]

In Fig. 6.13, I use the simulation setup to model the effective QDC non-linearity for all energies as a continuous line without uncertainty. Also, I create data points for the energy

[8]The fitted effective k_B values in Fig. 6.13 barely change with other values, e.g., $k_B = 150$ nm/keV [Abe+11].



(a) Correcting synthetic data in *Smelt* with the four energy peaks of ^{207}Bi, ^{113}Sn, and ^{137}Cs. Showing the relative deviation between the corrected and uncorrected amplitudes from Equ. (6.5).



(b) Correcting synthetic data in *Smelt* with the energy peaks of the calibration sources. Showing the relative deviation between the corrected and uncorrected amplitudes from Equ. (6.5).

Figure 6.13 Determining the effective QDC non-linearity with simulated data in *Smelt* in a first-order parametrization. The simulated data is modeled using a continuous line (green, "Sim Full") with energy resolution. The calibration source energy peaks are plotted as data points and used to fit the Birks non-linearity model (orange, "Fit"), see Sec. 6.1 The effective non-linearity deviates systematically from the Birks model.

peaks of the calibration sources using the energy resolution and spread over different events and both detectors as uncertainty estimation. Fitting a Birks non-linearity model to the points from the calibration sources yields an estimate for the QDC non-linearity as it would be measured with our calibration sources and parametrized by the implemented Birks non-linearity model. Given the ambiguity with the trigger function that skews the measured data below 200 keV, see Sec. 6.1.3, affecting ^{109}Cd and ^{139}Ce, the result fitting all six electron energy peaks of the five sources is also ambiguous. The estimated QDC non-linearity assuming a Birks model from a simulated first-order correction for high-energy calibration sources is $k_B = 454(87)$ nm/keV.

Fitting the detector response from Sec. 6.1 to our measured high-energy calibration spectra, we obtain $k_B = 471(23)$ nm/keV from a single calibration, see Sec. 7.2. Our fitted result from measured data and the 2009 result agree within 1 to 3σ with the simulated estimate. This result strongly indicates that the observed non-linearity is linked to the QDC integrator in a first-order approximation. Furthermore, the simulated QDC non-linearity deviates systematically from the Birks model for smaller amplitudes, indicating that a different functional form is required. This additional non-linearity must be applied after the Poisson modeling in Equ. (6.1), while the scintillator non-linearity is applied to the energy going into the Poisson distribution. The deviation effectively "tilts" the fitted non-linearity and causes a larger k_B value, which we currently cannot compare to measured data meaningfully due to the trigger ambiguity. These results on the QDC non-linearity indicate that the model based on a sole Birks non-linearity in Sec. 6.1 will not be able to accurately describe the measured data for energies below 350 keV. It is unclear whether the previous measurement of neutron spin-dependent quantities in [Mär+19; Sau+20] were sensitive to and affected by this effect, but our current understanding and the results from [Sau18] indicate that they were not.

While these new insightful results create an understanding of the observed QDC non-linearity, they cannot be used directly for correction. The simulated QDC non-linearity might be biased for smaller amplitudes, where the linearity approximation, chosen fit ranges, and the fitted linear model of the QDC sample gradient are difficult to verify. Thus, using the simulated result to construct a correction would lead to significant systematic uncertainties beyond the set precision goal. For the same reasons, deducing a functional form to fit the calibration sources might also induce biases.

Future work can take three steps to accurately and precisely determine the QDC non-linearity. First, using detailed offline tests with pulse generators should yield a more precise estimation of the non-linearity and can motivate an empirical QDC non-linearity model. Second, studying the electronic response in isolated simulations of the integrator hardware boards should give an accurate model for the non-linearity. Third, extensively studying the 2D energy deposition between detectors or groups of PMTs for backscattering events in an analysis of all calibration measurements should lead to additional information on QDC channel-specific behavior. Especially, data from ^{139}Ce and ^{113}Sn have a high potential for this analysis as the energy distribution between the detectors covers energies ranging between 40 and 100% of the main peak energy in the case of backscattering. The energy range between 50 – 369 keV is highly impacted by the QDC non-linearity and potentially allows for a parametrization of the non-linearity to enable fitting the electron energy spectrum from neutron beta decay for lower energies. However, these studies must be linked with full calibrations to extract exact values for the gain g and the effective number of photoelectrons f_{pe} to reduce correlations between parameters.

6.1.8 Synchrotron Radiation Losses

Another potential source for an induced non-linearity on measured electron energy spectra is losses from synchrotron radiation. The calculations in this section are based on the equations in [Wal94]. An electron with charge e propagating in a circular motion in a magnetic field with local curvature radius ρ emits synchrotron radiation with an instantaneous power P of

$$P = \frac{2}{3}\frac{e^2 c\beta^4\gamma^4}{4\pi\epsilon_0\rho^2},$$

with speed of light c, relativistic factors β and γ, and vacuum permeability ϵ_0. Thus, the energy per turn u_0 of circular motion is given by the path integral

$$u_0 = \oint \frac{P}{\beta c}\,\mathrm{d}s = \frac{2}{3}\frac{e^2 c\beta^3\gamma^4}{4\pi\epsilon_0}\oint \frac{1}{\rho^2}\,\mathrm{d}s.$$

Assuming a uniform bending radius in the bending magnets, i.e., an isomagnetic lattice, u_0 becomes

$$u_0 = \frac{e^2}{3\epsilon_0}\frac{\beta^3\gamma^4}{\rho}.$$

For u_0 in eV, we get the relation

$$u_0 = 2.65 \cdot 10^4 \cdot E^3 \cdot B$$

for E given in GeV and B given in T.

Using the time-of-flight calculations, distances, and backscattering model in *Smelt*, I estimate the synchrotron losses for initial electron energies by averaging over the angular distribution. Even though the synchrotron losses scale with E^3, for the magnetic field strength in PERKEO III with $B_{\max} = 152$ mT, the number of gyration rotations u_0 calculated with *Smelt*, and maximum energies of $E_{\max} = 1$ Mev, the effect only reaches a maximum loss of $4.16(41) \cdot 10^{-4}$ eV. These losses are negligible, given that one QDC integrator channel, i.e., the resoltion, from the electronic readout is about 32 eV, and we do not need to model synchrotron emission losses in our theoretical models as non-linearity. Even for future experiments like PERC [Wan+19] with stronger magnetic fields and longer decay volumes, this effect is negligible.

6.2 Calibration Source Holder Systematic Effects

The results of the *Smelt* electron time-of-flight studies in Sec. 6.1.6 motivated searches for secondary electrons from the gammas emitted by the calibration sources and energy losses of primary electrons through the calibration source holders. To this end, the author of [Ber24] used a Geant4 simulation of the full PERKEO III setup with an accurate description of the magnetic fields and backscattering to create data sets with these processes for a fixed initial electron or gamma energy in two separate simulation studies. Using these data sets, I create spectra matching the electronic readout and processing signature to create spectral corrections to the theoretical spectra of the calibration sources.

The calibration sources are suspended on ultra-thin carbon foils (12 and 30 μg/cm^2 area density, 10 mm diameter) that are supported by a thin Aluminum-coated Mylar foil (25 μm thickness, ca. 140 mm diameter). The Mylar foils are suspended on an arc-shaped 5 mm

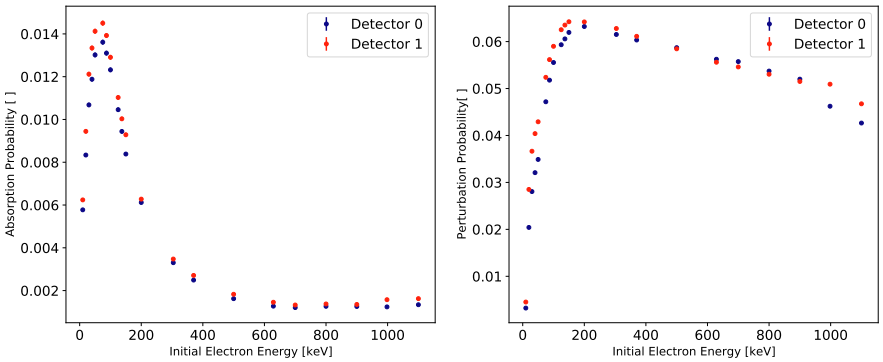


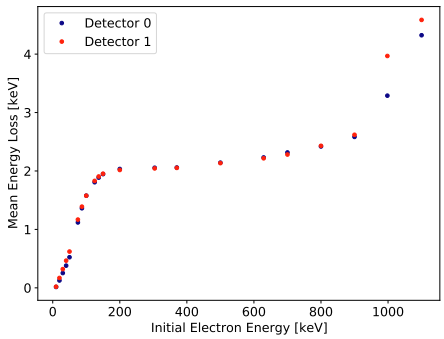
Figure 6.14 Simulated probabilities of electrons from calibration sources to be absorbed in the calibration source holder or to interact with it, i.e., be perturbed and lose energy, for each detector and different energies. The increasing gyration radius causes the energy-dependent trends of the probabilities as the interaction with the source holder requires a minimum radius to bridge the distance from the calibration source position to the Mylar support foil. In some cases, the perturbation can occur before interacting with a detector but mainly happens after backscattering.

wide and 2 mm thick Aluminum frame, see Sec. 4.2.2. The contribution of the ultra-thin carbon foils as a source of secondary electrons or energy losses is insignificant and therefore neglected [Roi18; Ber24]. In the simulations, the calibration source holder is placed at the calibration position in the plane transversal to the beam axis and the CaliBot position about 50 cm out of the center of the central decay volume. Also, the source holder is tilted by 10° towards the beam axis to account for the local waviness of the support foils and general angle uncertainty of the holder.

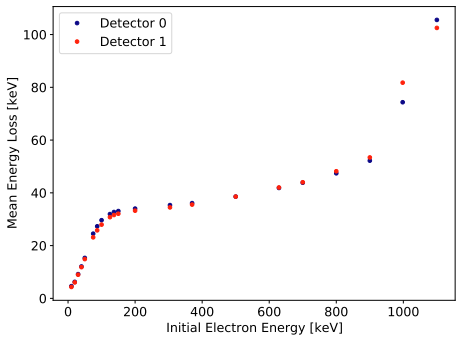
6.2.1 Support Foil Induced Energy Losses

A number of n electrons are simulated for a set of initial energies in separate simulations, with a starting point at the center of the calibration source holder and isotropic emission. Besides the detected electron energies, the simulation tracks the number of electrons interacting with the source holder, giving the probability p_t of an electron being perturbed for each simulated electron energy. Also, we track the number of electrons being absorbed, giving the absorption probability p_a for each energy. The results for the probabilities are shown in Fig. 6.14 and the average energy losses scaled for all or only the perturbed electrons in Fig. 6.15.

Both probabilities p_t and p_a show an increasing trend for low energies. The increasing gyration radius causes this trend as the interaction with the source holder requires a minimum gyration to bridge the distance from the calibration source position to the Mylar support foil. The absorption probability p_a decreases again for higher energies, as the probability of transmitting through the source holder without being absorbed increases with higher energies. We see a minor increase for the highest energies up to 1 MeV, caused by large enough gyration radii to interact with the comparatively thick (2 mm thick, 5 mm wide) aluminum frame. The same reasons explain the trends of the average energy loss in Fig. 6.15. The average energy loss calculated for all electrons, perturbed and unperturbed, shows this correction is non-negligible, as most electrons lose 2 – 3 keV.



(a) Mean energy loss calculated over all electrons, perturbed and unperturbed, for different initial energies.



(b) Mean energy loss for perturbed electrons for different initial energies.

Figure 6.15 Simulated mean energy losses from interactions with the calibration source holder. The observed energy dependence is explained by the minimum gyration radus required to interact with the Mylar support foil. At large energies, some electrons start interacting with the Aluminum frame, leading to greater energy losses.

Generally, both detectors show very similar behavior. Slight differences are caused by the asymmetric CaliBot position about 50 cm out-of-center in the central decay volume towards the downstream detector. This asymmetry causes effects, e.g., due to electrons being emitted towards the center of the central decay volume (upstream) and being reflected by the magnetic mirror effect at the maximum value of the magnetic field in the center.

To create a spectral correction of the source holder-induced energy losses, we must create a surrogate model from the simulated data of fixed initial energies to get the perturbed electron spectrum for the continuous space of possible initial energies. As the probabilities for perturbation and absorption are already calculated, I normalize the spectra of perturbed electrons to preserve details in the shapes for lower energies. A selection of the perturbed electron spectra is shown in Fig. 6.16. The plot depth indicates the initial electron energy, visualizing the changing shape of the perturbed electron spectrum towards higher energies. The trend of a sharper energy peak at the initial electron energy continues for the tested energies up to 1.1 MeV. Due to the shape of the perturbed electron distributions, the resulting corrections mostly affect the low-energy tails of energy peaks in the calibration spectra.

The shown perturbed electron spectra are slices in the 2D space that we must interpolate with a surrogate model. Due to their flexibility, I use neural networks as a surrogate model with pre-training on the low-energy data to preserve the nuances of the changing shapes and a mean square error deviation as an objective value. With a surrogate model, we can correct each line E in the theoretical line spectra as described in Sec. 6.1 by adding the perturbed electron spectra for that energy E with norm p_t and rescale the original probability of E by the factor $\delta = 1 - p_\mathrm{t} - p_\mathrm{a}$. Thus, this correction has no direct free parameters in a calibration source fit. However, this correction asymmetrically affects calibration energy peaks, leading to non-negligible systematic uncertainties as it changes energy peak position, width, and through its energy-dependence, also the non-linearity.

To verify the surrogate model interpolation of the perturbed electron spectra, I remove two data sets for initial energies when training the model to see how accurate the interpolation would model these energies. The results are shown in Fig. 6.17 for initial electron energies of 250 and 700 keV. The surrogate model successfully models the missing energies by

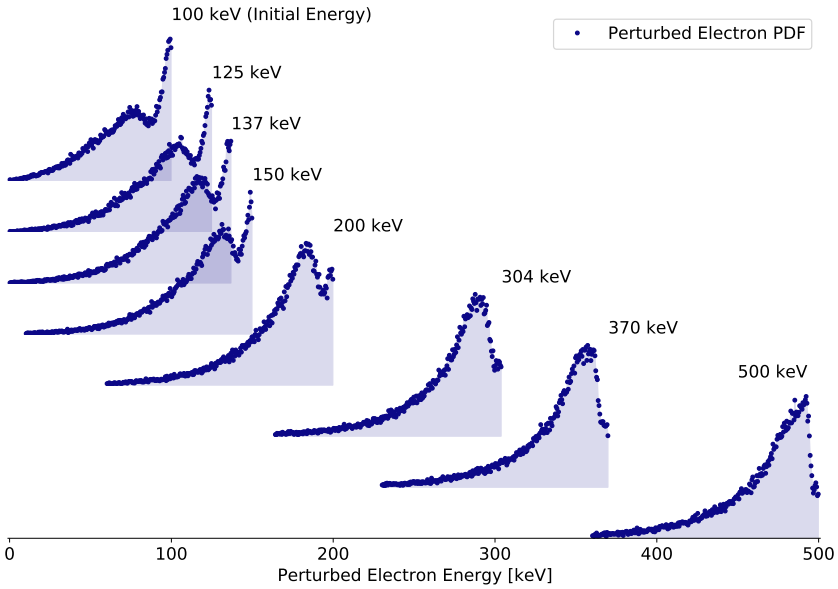


Figure 6.16 Visualization of the changing shapes of the normalized perturbed electron spectra towards higher energies, as used for the interpolation for the correction. The plot depth indicates different initial electron energies. The spectra are normalized to preserve details in low-energy spectra which have smaller total probabilities for the interpolation. The trend of a sharper energy peak at the initial electron energy continues for the tested energies of up to 1.1 MeV.

interpolating from the closes initial electron energies in the training data set, ±50 keV for the 250 keV and ±100 keV for the 700 keV electron. Small deviations from the surrogate model are caused by the statistical fluctuation of the training process and higher precisions of the correction require more data.

I use this fluctuation and related effects to make a preliminary estimate of the systematic uncertainty of the correction on calibration parameters of the detector response. The total relative systematic uncertainty $\sigma_{\text{sys}}(g)$ on the detector gain g caused by induced energy losses for both detectors is $\sigma_{\text{sys}}(g) = 2.59 \cdot 10^{-4}$ ch/keV. The total absolute systematic uncertainty $\sigma_{\text{sys}}(f_{\text{pe}})$ on the effective number of photoelectrons per keV f_{pe} for one detector is $\sigma_{\text{sys}}(f_{\text{pe}}) = 3.21 \cdot 10^{-3}$ keV^{-1}. The total absolute systematic uncertainty $\sigma_{\text{sys}}(k_B)$ on the Birks coefficient k_B for one detector is $\sigma_{\text{sys}}(k_B) = 3.19$ nm/keV. As a precise energy measurement is highly sensitive to systematic corrections of the detector response parameters, this correction is non-negligible for a precise measurement of the Fierz interference term b and I discuss its significance in Sec. 8.

6.2.2 Induced Secondary Electrons

A number n of gammas are simulated with an initial energy linked to one of the gamma emissions from the calibration sources. The gamma source is at the center of the calibration source holder. During the simulation, we track the number and energies of produced secondary electrons, i.e., photoelectric and Compton electrons, that reach the detectors. The probability p_{s} of a secondary electron reaching a detector given a number of emitted gammas is shown in Fig. 6.18 for each gamma energy. The probability increases with increasing

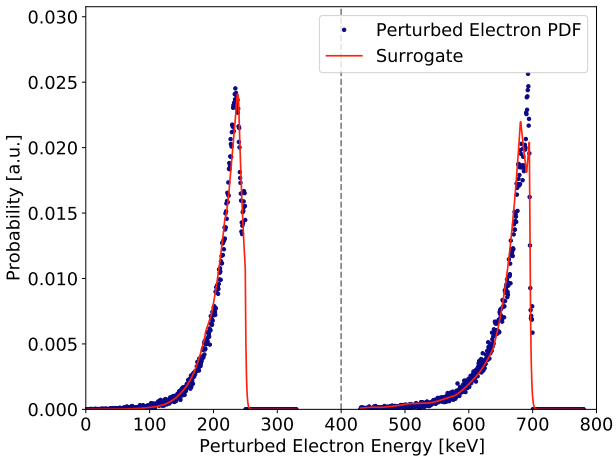


Figure 6.17 Verification of the surrogate model interpolation of the perturbed electron energy PDF for initial electron energies not present in the training data. The closest initial electron energies in the training data are 200 and 300 and 700 and 800 keV, respectively. The surrogate model successfully interpolates between energies in the training data, guaranteeing a meaningful model over the full 2D plane of initial electron energies and perturbed electron energy spectra. The minor deviations between the surrogate model and the data are related to statistical fluctuations of the training process, and I use them to estimate the induced systematic uncertainty.

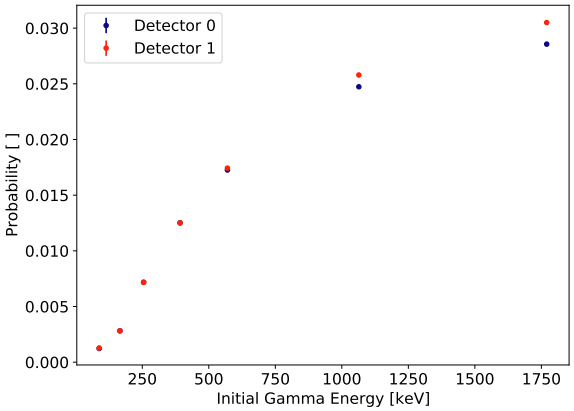


Figure 6.18 Simulated probability of a secondary electron reaching a detector for a produced gamma for different initial gamma energies. The probability increases with increasing gamma energies, as the produced secondary electrons also have a higher available energy. The secondary electrons are produced within the Mylar support foil and aluminum frame and have a higher chance of escaping the material with higher energies.

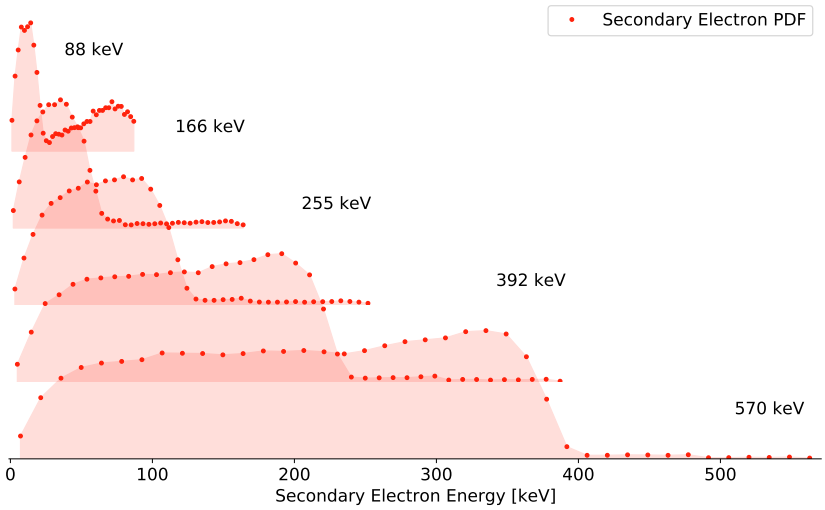


Figure 6.19 Visualization of the changing shape of the normalized energy spectrum for the gamma-induced secondary electrons, i.e., photoelectric and Compton electrons, in the source holder for different initial gamma energies, indicated by the plot depth. The spectra are normalized to preserve details in low-energy spectra which have smaller total probabilities for the interpolation and correction. The distinct Compton edge is visible for all energies.

gamma energies, as the produced secondary electrons also have a higher available energy. The secondary electrons are produced within the Mylar foil and aluminum frame and have a higher chance of escaping the material with higher energies. The difference between the behavior of both detectors is caused by the asymmetric setup and the magnetic mirror effect at the center of the decay volume.

We add the spectrum of the gamma-induced secondary electrons in the source holder to the theoretical calibration spectra to create a correction for each calibration source. All normalized secondary electron spectra from the simulated gamma energies are shown in Fig. 6.19. The plot depth indicates the initial gamma energy, visualizing the changing shape of the secondary electron spectrum towards higher gamma energies. The distinct Compton edge is visible for all energies, and the resulting correction affects almost all energies below the gamma energy due to the long tails over the entire energy range.

I estimate the uncertainty of this correction by splitting the simulated data set for the interpolation into subsets and creating a set of corrections with greater statistical variety. Using these varied corrections, I can estimate the significance of the overall correction by varying it during the spectrum generation and test how it affects the extraction of the Fierz interference term b when fitting the resulting neutron beta spectrum. I obtain an upper limit on the combined systematic uncertainty $\sigma_{\text{sys}}(b)$ on the Fierz interference term b for both detectors from the secondary electron emission as $\sigma_{\text{sys}}(b) \leq 5 \cdot 10^{-5}$ for all tested fit ranges, making it negligible. The small systematic uncertainty is intuitive, given that the correction mainly consists of long, almost flat energy tails below the energy of the causing gammas. This shape and the distance of the Compton edge from the main peak cause the systematic uncertainty to mainly affect the lower 500 keV ^{207}Bi peak below the 995 keV peak in calibration fits. However, the secondary electron spectra are necessary for our theoretical model to successfully describe the observed calibration data.

7 Fitting Measured Spectra

After creating a theoretical description of our detector response model as our hypothesis, we can check if it agrees with the observed and processed data by fitting. We use the measurements with the five calibration sources to extract the remaining detector parameters to fit the electron beta spectrum from neutron beta decay to ultimately extract the Fierz interference term b. Due to the observed issues with the trigger function model and QDC non-linearity discussed in Sec. 6.1.3 and 6.1.7, we expect statistically significant deviations of our hypothesis. In this chapter, I conduct these fits, discuss the systematic deviations, and link them to the studies in previous chapters.

7.1 *p3fit* Fitting Tool

The analysis and fitting tool *p3fit* was developed within the PERKEO III collaboration over several data analyses and years, with the current version being mainly developed in [Roi18; Sau18], and used for recent PERKEO III results, e.g., [Mär+19; Sau+20]. It comprises the outlined detector response model from Sec. 6.1 and other detector models for comparison, like non-linearity descriptions, from past analyses. In *p3fit*, energy spectra are modeled with a broadening of theoretical line spectra through convolutions and represented as histograms, as there are no simple closed-form solutions. The tool uses a detailed description to accurately model the theoretical spectra of the electron beta decay and the calibration sources with radiative, recoil, and Coulomb interaction corrections. The internal transitions and correlated electron emissions of the calibration sources are modeled with decay channels, consider the integration time and apply non-linearity correctly to multi-electron events. I created all theoretical spectra and fits in this chapter with *p3fit*.

In the current implementation, I neglect multi-electron non-linearity handling to allow for the corrections on the theoretical energy line spectra from the calibration source holder systematic effects from Sec. 6.2. This description leads to minor deviations in the non-linearity description that must be accounted for in future work. However, given the systematic uncertainty of the calibration source holder corrections, see Sec. 8, and the required QDC non-linearity measurements, see Sec. 6.1.7, these deviations are currently negligible. Also, multi-electron handling and the calibration source holder's systematic effects only apply to the calibration sources and not to the electron beta spectrum measurement from neutron decay. In addition, *p3fit* is blinded for an analysis of the Fierz interference term b with a relative and absolute distortion [Mär22].

Due to the significant increase in energy resolution, see Sec. 5.4, the previously implemented internal conversion coefficients in *p3fit* are not accurate enough. Thus, I updated the internal conversion coefficients for all calibration sources based on values in the DDEP [Bé+16] database and using the BrIccFO [Kib+08] software. In particular, higher-order contributions from other multipolarities for the internal conversion coefficients are essential for a theoretical description of our measured calibration spectra. These contributions were

negligible in past measurements and are only necessary now due to the increased energy resolution.

7.2 Calibration Fits

High Energy Fits

Based on the known issues with the trigger function model and the QDC non-linearity description, which significantly increase at lower energies (below 350 keV), I study the high-energy calibration sources separately. For the high energy fits, I look at the 995 keV and 503 keV peaks of ^{207}Bi, the 630 keV peak of ^{137}Cs, and the 369 keV peak of ^{113}Sn. The combined and simultaneous fit of all four energy peaks for the upstream detector is shown in Fig. 7.1.

The theoretical model describes the measured calibration sources accurately within statistical limitations (reduced $\chi^2 = 1.1$, $p = 0.10$). We also observe that the 503 keV ^{207}Bi peak is only successfully described by the theoretical model with the new internal conversion coefficients and both calibration source holder corrections, highlighting the necessity of these new corrections. The extracted detector parameters are gain $g = 30.934 \pm 0.021$ ch/keV, non-linearity Birks coefficient $k_B = 485 \pm 18$ nm/keV, and light yield $f_{\mathrm{PE}} = 0.664 \pm 0.015$ keV^{-1}.

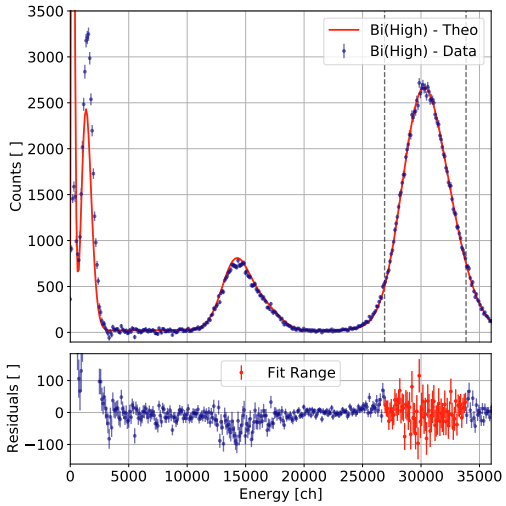
Systematic deviations between our hypothesis and the processed data are visible. While both ^{207}Bi peaks fit well individually, the respective other peak deviates as seen in the residuals due to non-matching branching ratios. This deviation was already observed in the past in other PERKEO III data sets and we fit these two peaks with a separate normalization. Also, all theoretical descriptions of the detector response deviate at lower energies. This deviation is mainly caused by the insufficient trigger function model. Additionally, the positive residuals for the Auger electron peaks below channel 2000 may be explained by systematic deviation of the measured and fitted trigger function in Fig. 6.5.

In addition, the residuals of ^{113}Sn and ^{137}Cs show some systematic deviations, presumably due to the non-linearity description based on the Birks model. As seen in Fig. 6.13, the Birks model does not accurately describe the QDC non-linearity in a first-order approximation, and the deviation between the Birks and observed non-linearity diverges below 400 keV. Reducing the statistical weight of the ^{113}Sn peak for the fit by reducing its fit range lessens the deviation for ^{137}Cs as indicated in the residuals, and the Birks coefficient decreases to $k_B = 471 \pm 23$ nm/keV.

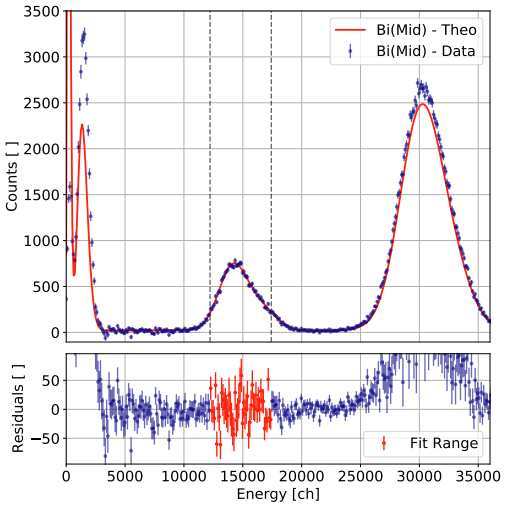
Due to the significant deviations at lower energies, the spectral analysis below 400 keV is not meaningful with the current description of the detector response. This issue should be mitigated with a more accurate model of the QDC non-linearity, see Sec. 6.1.7. Given the sensitivity to higher-order effects in the theoretical description of the conversion sources due to the excellent energy resolution, future high-precision measurements with similar calibration schemes will benefit from new advances in nuclear studies of these calibration sources.

Full Energy Range Calibration

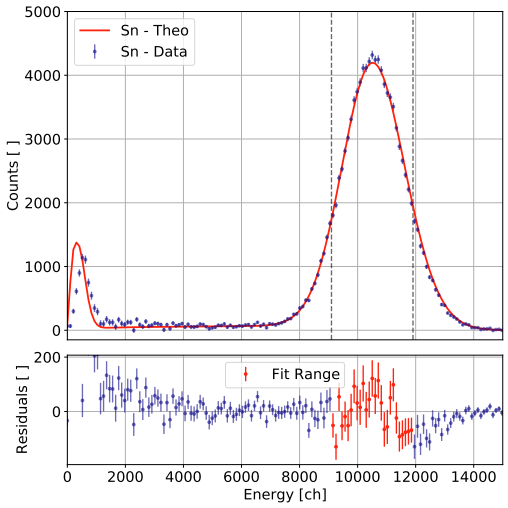
Expanding the calibration fit from three to all five sources by also fitting ^{109}Cd (75 keV peak) and ^{139}Ce (127 keV peak), we observe the same issues already hinted at by the high-energy



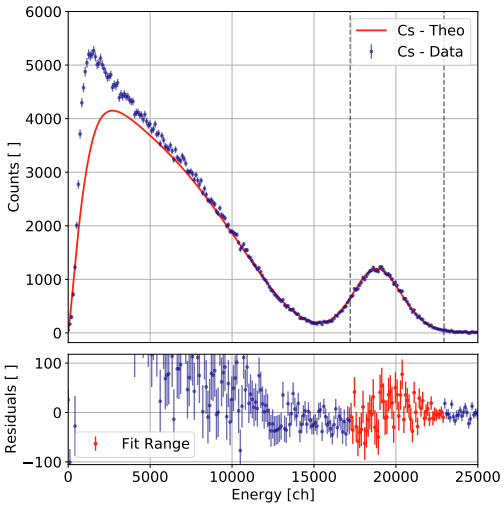
(a) Fitting the high-energy peak of ^{207}Bi with 995 keV.



(b) Fitting the mid-energy peak of ^{207}Bi with 503 keV.

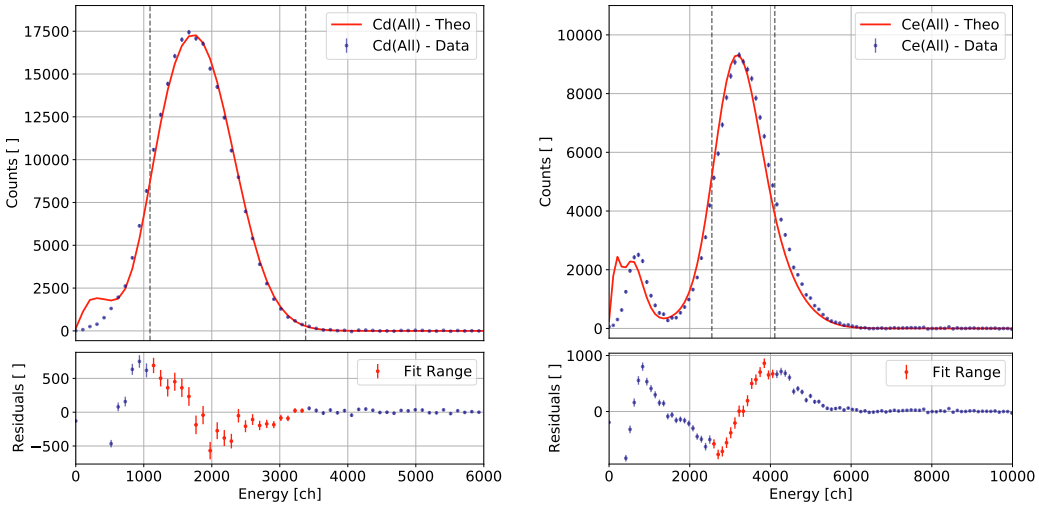


(c) Fitting the 369 keV peak of ^{113}Sn.



(d) Fitting the 630 keV peak of ^{137}Cs.

Figure 7.1 Combined fit of the high-energy calibration sources of one calibration measurement on the upstream detector (detector 0). The theoretical model describes the observed data within statistical limitations (reduced $\chi^2 = 1.1$, $p = 0.10$). The theoretical model only successfully describes the 503 keV ^{207}Bi peak when using the new internal conversion coefficients and considering both calibration source holder corrections, highlighting the necessity of these new corrections. The ^{207}Bi peaks fit well individually but require separate normalization due to non-matching branching ratios. The insufficient trigger function model causes significant deviations and positive residuals at lower energies. The residuals of ^{113}Sn and ^{137}Cs show some systematic deviations, presumably due to the non-linearity description based on the Birks model. As the deviation between the Birks and observed QDC non-linearity diverges below 400 keV, we see systematic tensions in the residuals of ^{113}Sn and ^{137}Cs.



(a) Fitting the 75 keV peak of ^{109}Cd.

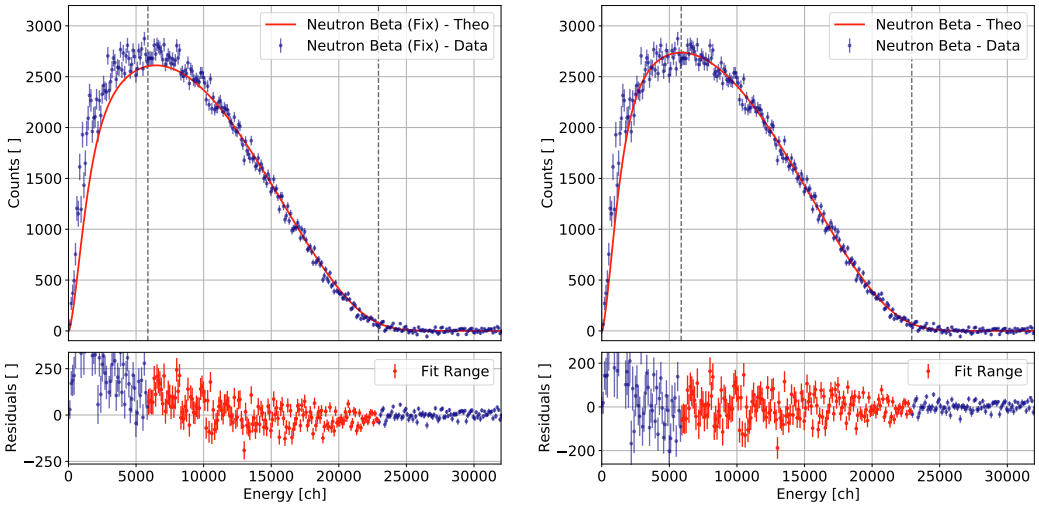
(b) Fitting the 127 keV peak of ^{139}Ce.

Figure 7.2 Combined fit of all calibration sources of one calibration measurement on the upstream detector (detector 0), only showing the two low-energy sources. The theoretical model does not describe the observed data within statistical limitations (reduced $\chi^2 = 5.0$, $p = 0.0$). The systematic deviations from the QDC non-linearity and the trigger function cause the peak positions of both low-energy sources to pull the residuals and effective non-linearity in different directions. These pulls are also seen as a tendency in the high-energy sources, similar to the "tilt" in Fig. 6.13 with all five sources for the estimated QDC non-linearity with *Smelt*.

calibration sources regarding QDC non-linearity and the trigger function. Fig. 7.2 shows the fits for Cd and Ce for the upstream detector resulting from a combined fit with all sources The systematic deviations cause the peak positions of both low-energy sources to pull the fitted non-linearity in different directions. Separating the effects of the QDC non-linearity and the trigger function for these energies is not straightforward, as these effects are highly correlated. The QDC non-linearity affects the peak position and the trigger function the peak height differently at different energies. The theoretical model does not accurately describe the measured calibration sources within statistical limitations (reduced $\chi^2 = 5.0$, $p = 0.0$).

The extracted detector parameters are gain $g = 30.916 \pm 0.022$ ch/keV, non-linearity Birks coefficient $k_B = 439 \pm 6$ nm/keV, and light yield $f_{\mathrm{PE}} = 0.663 \pm 0.011$ keV^{-1}. The resulting gain g and light yield f_{PE} are unchanged within statistical limits. However, the non-linearity Birks coefficient k_B decreases, and its uncertainty is significantly reduced. The latter is intuitive, as the non-linearity parametrization is dominantly determined at lower energies. The systematic deviations cause the smaller value of k_B and can be verified by only adding one of the two low-energy sources to the combined fit with the high-energy sources: Only adding ^{109}Cd results in a $k_B = 472 \pm 7$ nm/keV, agreeing with the high-energy fit, and adding only ^{139}Ce leads to $k_B = 385 \pm 9$ nm/keV. This discrepancy highlights that a different functional shape is required for the QDC non-linearity.

A sufficient theoretical description of all five measured calibration sources is required to accurately describe non-linearity and use the extracted detector parameters to measure the Fierz term b. Currently, this analysis is not possible and requires the future work outlined in Sec. 6.



(a) Fitting with the extracted detector parameters from the high-energy calibration fit of Fig. 7.1.

(b) Same as a), but with k_B as free fit parameter to illustrate the remaining systematic non-linearity.

Figure 7.3 Fitting the electron beta spectrum from neutron decay (approximately $3.7 \cdot 10^5$ events) with the extracted parameters from the calibration. The fit range is approximately 250 - 750 keV, but depends on the value of k_b. The beta spectrum requires a significant non-linearity that is not accounted for in the current theoretical description of the detector model, and we see deviations below channel 5000 based on the trigger model and QDC non-linearity as with the calibration measurements. Further analysis requires future work as outlined in Sec. 6 and a new detector response model.

7.3 Electron Beta Spectrum Fits

Even though spectral analysis of the electron beta spectrum from neutron decays with the current calibration is not meaningful, we can still test our theoretical description and verify our previous observations. The fit using the detector parameters from the calibration with the high-energy sources and correcting the gain g and the light yield f_{PE} with the spatial corrections is shown in Fig. 7.3. I use the extracted non-linearity value of k_B from the calibration in Fig. 7.3a and release k_B as a free fit parameter in Fig. 7.3b (k_B = 854 ± 65 nm/keV). In both fits, the theoretical model does not accurately describe the measured calibration sources within statistical limitations (reduced $\chi^2 = 2.6$ and $\chi^2 = 2.0$, respectively). As expected, the beta spectrum requires a significant non-linearity to fit, and we see similar deviations below channel 5000 based on the trigger model and QDC non-linearity.

Adjusting the fit range to only higher energies and leaving k_b unconstrained, we get $k_B \approx 440$ nm/keV which is close to the value for the high-energy calibration fit. The upper fit range limit does also affect the non-linearity parameter k_B, indicating the necessity of an edge effect correction, as discussed in Sec. 6.1.2 and planned in [Ber24]. Further analysis requires future work as outlined in Sec. 6 and a new detector response model, especially a more accurate QDC non-linearity description.

However, we must accurately determine the systematic uncertainties on the Fierz interference term b of the dominant corrections without biasing the results with the remaining systematic effects. To this end, I developed a new simulation-based tool, which is described in the next chapter.

8 Systematic Uncertainty Estimation

Using a pure fitting tool like *p3fit* and a data set to calculate the systematic uncertainties gives results specific to the used detector model and data set. While giving accurate results, this approach does not allow for comparisons between different detector models, i.e., experiments, and data sets, i.e., studied beta decays, beyond the final systematic uncertainty.

To tackle this issue, I present a new simulation-based analysis tool that can be used to generate and fit beta decay spectra with a detector model for systematic uncertainty quantification. With this tool, I can determine the systematic uncertainty budget for the corrections studied in this thesis while avoiding a potential bias of the remaining systematic effects of the trigger function and QDC non-linearity. In addition, I study the scaling laws for the sensitivity of individual detector parameter variations, look at their correlation, and compare the results to previous measurements to discuss future optimization approaches.

8.1 *Freya* Analysis Tool

To study the systematic uncertainties of detector parameters precisely and without bias from specific measurements, I developed a *Julia* package called *Freya*[1] (FieRz Error and sYstematics Analysis). The goal is to complement previous work on statistical kinematic sensitivity of different decays [GN16] and statistical sensitivity studies [GJL95] with a more experimental perspective with the estimation of systematic uncertainties. Such a perspective would also allow for comparisons of experiments.

The *Freya* package implements the detector model in Equ. (6.1) with Poisson broadening of a theoretical energy spectrum with an effective energy $E_\mathrm{f}(E; k_B)$ from the Birks non-linearity[2] as

$$p_\mathrm{pe}(n_\mathrm{pe}; E, f_\mathrm{pe}) = \frac{[f_\mathrm{pe} \cdot E_\mathrm{f}(E; k_B)]^n}{n_\mathrm{pe}!} e^{-f_\mathrm{pe} \cdot E_\mathrm{f}(E; k_B)} \tag{8.1}$$

and Gaussian broadening for the electronics readout and pedestal width σ_Q as

$$p_\mathrm{Q}(c; A, g, \sigma_\mathrm{Q}) = \mathcal{N}(c; A \cdot g, \sigma_\mathrm{Q}). \tag{8.2}$$

More details are given in Sec. 6.1.

To extract systematic uncertainties $\sigma_\mathrm{sys}(b)$ of the Fierz interference term b, we must include all energy-dependent contributions in the electron beta spectrum. Thus, I use the theoretical neutron decay beta spectrum from Sec. 2 in Equ. (2.9) with corrections as

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}E_e} \propto F(Z, E_e) \cdot \left[1 + \delta_R(E_e, E_{\beta^-})\right] \cdot p_e E_e \left[E_{\beta^-} - E_e\right]^2 \left[1 + b\frac{m_e}{E_e}\right], \tag{8.3}$$

[1] https://github.com/maxlampe/freya

[2] We use the Birks non-linearity model to describe the effective total non-linearity and not only for the scintillator non-linearity, as it is used in the literature [Bir51; Pös+21].

with electron-proton Coulomb interaction correction with the Fermi function $F(Z, E_e)$ and radiative and recoil correction $\delta_R(E_e, E_{\beta^-})$ as function of the electron energy E_e and end point energy E_{β^-}.

I use the implementation of outer radiative corrections (based on [Sir67]) and recoil corrections (based on [Rei99; IPT13]) from *p3fit* [Roi18; Sau18] and port them to *Julia*. The Fermi function is approximated as [MM33]

$$F(Z, E_e) \approx \frac{2\pi\eta}{1 - e^{-2\pi\eta}} \quad \text{with} \quad \eta = \frac{\alpha Z}{\beta},$$

with fine-structure constant α, relativistic factor β, and nuclear charge Z. This approximation works well for neutron beta decays and leads to a relative deviation of $\leq 10^{-3}$ for the neutron decay spectrum [Roi18]. This deviation must be considered when fitting measured data but is negligible for neutron decay studies with *Freya*, as the generating function is also the fit function.

Potential Future Work

This framework can be easily expanded with other detector response models and beta decay rates to generalize the results and compare experiments, reducing ambiguities. Different systematic sensitivities of beta decays, e.g., for ^{14}O or ^{6}He, can be compared for the same detector response model with the same parameters. However, such studies for larger values of Z required a more accurate implementation of the Fermi function as given in [Fer34; BB82].

Also, probing the systematic uncertainty of new QDC non-linearity models is also possible by expanding p_{Q} with effective amplitude A_{f} and non-linearity parameter k_{Q} as

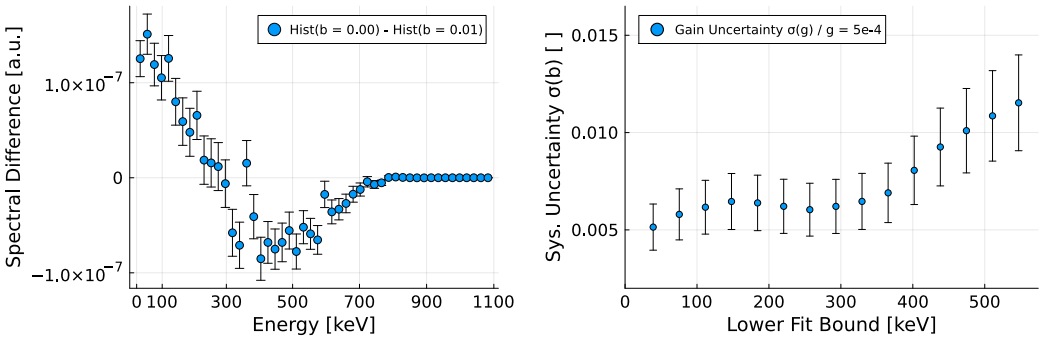
$$p_{\mathrm{Q,\,f}}(c; A, g, \sigma_{\mathrm{Q}}, k_{\mathrm{Q}}) = \mathcal{N}(c; A_{\mathrm{f}}(A, k_{\mathrm{Q}}) \cdot g, \sigma_{\mathrm{Q}}).$$

This approach can also be used to compare different electronic readout systems before a measurement. In general, *Freya* can be used to compare experiment design decisions or for the data analysis of future high-precision experiments of beta decay searches of the Fierz interference term b and potentially other energy-dependent quantities.

Setup and Characterization

For the studies in this chapter, I determine the systematic uncertainty $\sigma_{\mathrm{sys}}(b)$ by sampling one or multiple parameters i, e.g., the detector gain g, with a fixed variation σ_i and study the induced deviations of b. I assume a prior distribution of $b = 0.017 \pm 0.021$ based on the current most precise result of [Sau+20]. This choice might induce a bias of the resulting systematic uncertainties towards non-Standard Model values of b. An initial consistency test limits this bias at $\leq 12\%$ of the resulting uncertainties.

Each simulated electron beta spectrum is generated from the theoretical spectrum in Equ. (8.3) with the broadening of the detector model as in Equ. (8.1) and (8.2) with $2.2 \cdot 10^8$ events. The extraction of the induced deviations of b with *Freya* achieves a bias and baseline precision limit of $5 \pm 12 \cdot 10^{-5}$. This precision is tested for deviations up to $\Delta b = 5 \cdot 10^{-2}$. Therefore, we can conduct reliable studies of $\sigma_{\mathrm{sys}}(b)$ with the current implementation of *Freya* in the range of $1.2 \cdot 10^{-4}$ to $5 \cdot 10^{-2}$. The baseline uncertainty of $\sigma_{Freya}(b) = 1.2 \cdot 10^{-4}$ is included in all results.



(a) Difference of two normed histograms with $4 \cdot 10^8$ events and different values of the Fierz term b. The b value difference is twice the target precision ($5 \cdot 10^{-3}$).

(b) Fit range dependence of the systematic uncertainty $\sigma_{\text{sys}}(b)$ for a set gain uncertainty. The errorbars do not show statistical fluctuations between data points.

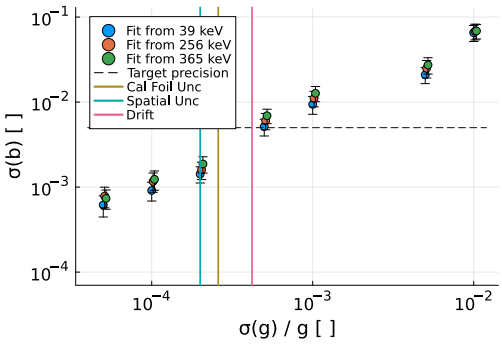
Figure 8.1 Simulated spectral dependence for the systematic uncertainty $\sigma_{\text{sys}}(b)$ determination with *Freya*. Due to the spectral shift for non-zero Fierz terms, the estimated systematic uncertainty depends on the fit range. This fit range dependence is stronger for the dominant detector parameter variations, e.g., as for the gain g in Sec. 8.2.

I use the detector parameters from the high-energy calibration fit of Sec. 7 with minimum tensions between the residuals of ^{113}Sn and ^{137}Cs for the studies in this chapter: Gain $g = 30.931$ ch/keV, effective Birks coefficient $k_B = 470$ nm/keV, and light yield $f_{\text{PE}} = 0.658$ keV^{-1}. The offset, i.e., pedestal, is given as $P = 0$ and the pedestal width as $\sigma_{\text{Q}} = 117.2$, from the corrections in Sec. 5. The width σ_{Q} is not varied in these studies as it is already shown to be insignificant given its uncertainties.

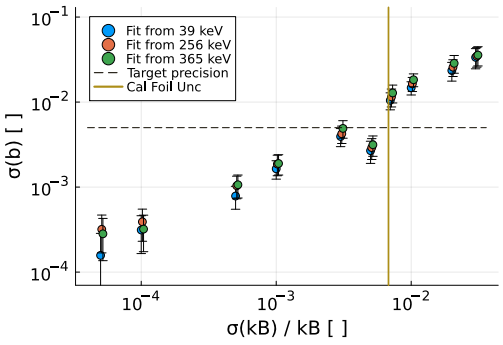
I chose three fit ranges to compare the results for the systematic uncertainty $\sigma_{\text{sys}}(b)$ estimation. Firstly, I use 39 – 743 keV (5 – 95% of endpoint energy E_{β^-}) as the fit range would allow for comparisons of the statistical uncertainty studies in [GN16]. The second fit range is from 256 – 743 keV (33 – 95%), which shows good sensitivity, see Fig. 8.1, and is our target fit range based on previous work [Sau18]. In the past, fits below $\approx$ 250 keV lead to increasing systematic uncertainties related to the trigger function and non-linearity. Finally, I also fit from 365 – 743 keV (47 – 95%) to probe the sensitivity for a fit range that avoids the critical trigger and QDC non-linearity issues. I focus on varying the lower fit bound, as it significantly affects the resulting systematic uncertainties. When varying k_B, the systematic uncertainty on b could be reduced by decreasing the upper fit bound, but the lower fit bound is still more impactful. Also, for previous PERKEO III analyses, the edge effect correction sets the upper fit bound, which is currently studied for [Ber24], and not the corrections included here.

8.2 Individual Systematic Effects

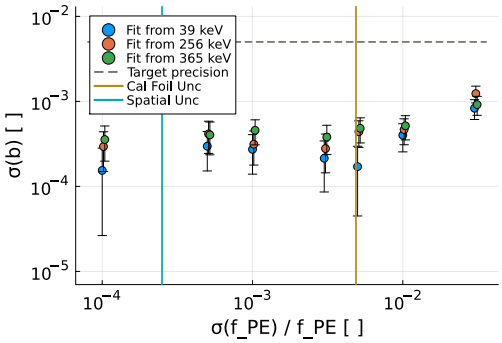
First, I focus on the sensitivity to variations of individual detector parameters for different orders of magnitude to obtain scaling laws for the systematic uncertainty $\sigma_{\text{sys}}(b)$ and understand the importance of these parameters. Such scaling laws help predict systematic uncertainties of future experiments with similar detector response models and give precise estimations of how much to improve problematic corrections to reach a desired precision. We can also directly obtain the systematic uncertainties for uncorrelated corrections, such as the pedestal and temperature-induced gain fluctuation corrections that affect only the channel offset and gain, respectively. The results for the four detector parameters are shown



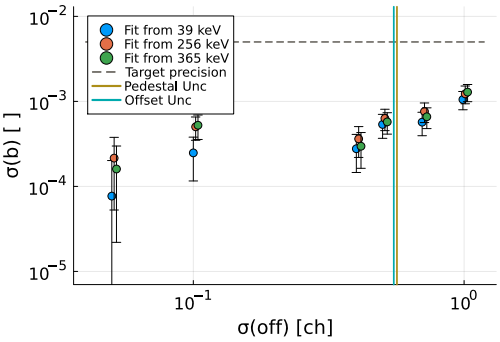
(a) Variation of gain g (relative)



(b) Variation of effective Birks coefficient k_B (relative)



(c) Variation of light yield f_{PE} (relative)



(d) Variation of channel offset P (absolute)

Figure 8.2 Simulated scaling behavior of the systematic uncertainty $\sigma_{\mathrm{sys}}(b)$ for varying individual detector parameters with *Freya*. The results for different fit ranges are slightly shifted for better visualization. I indicate the targeted precision of $5 \cdot 10^{-3}$ as a dashed line, and all relevant corrections from Sec. 5 and 6 are shown as vertical lines. The most significant systematic correction is the temperature-induced gain fluctuation correction ("Drift"). The calibration support foil uncertainty ("Cal Foil Unc") is only a preliminary estimate, and the correction is limited by simulation data availability.
The systematic uncertainty $\sigma_{\mathrm{sys}}(b)$ is most sensitive to variations of the gain g and effective non-linearity Birks coefficient k_B. However, both parameters are correlated, see Sec. 8.3, and typically obtained from calibrations together. This effect makes interpreting the k_B result ambiguous. The systematic uncertainties grow when increasing the lower fit bound for systematic uncertainties $\sigma_{\mathrm{sys}}(b) \leq 1 \cdot 10^{-3}$.

in Fig. 8.2. Besides the uncorrelated corrections of individual detector parameters, the correlated corrections from previous sections are also plotted for comparison and analyzed in Sec. 8.3.

The scaling of $\sigma_{\text{sys}}(b)$ for individual detector parameter variations follows a linear relation in the log-log plots of Fig. 8.2 above the saturation at the precision limit of *Freya* at $1.2 \cdot 10^{-4}$. This relation indicates a slight exponential relation, which is intuitive. The systematic uncertainty $\sigma_{\text{sys}}(b)$ is most sensitive to variations of the gain g and effective non-linearity Birks coefficient k_B, while the variations of the light yield f_{PE} and the channel offset P have little effect. For the parameters with the biggest impact on $\sigma_{\text{sys}}(b)$, the systematic uncertainties slightly grow with when increasing the lower fit bound for systematic uncertainties $\sigma_{\text{sys}}(b) \geq 1 \cdot 10^{-3}$. With these results, we can accurately estimate the systematic uncertainties of uncorrelated corrections.

It is important to highlight that interpreting the results for individually varying the effective Birks coefficient k_B is ambiguous. In this study, the other parameters are fixed, and for realistic calibration corrections, the gain g and the effective Birks coefficient k_B are extracted together and also correlated, see Sec. 8.3. Varying k_B can also induce an offset P, but due to the final offset correction, see Sec. 5, P is fixed at $P = 0$ for our calibration and neutron beam fits. Furthermore, the Birks non-linearity model does not accurately describe the QDC non-linearity, as seen in Fig. 6.13. Another non-linearity model should have a different behavior in terms of the systematic uncertainty of its parametrization on $\sigma_{\text{sys}}(b)$.

Temperature-Induced Gain Fluctuation Correction

Based on the results in Sec. 5.2.4 of a relative gain uncertainty of $\sigma(g)/g = 4.2 \cdot 10^{-4}$ for the temperature-induced gain fluctuations, I obtain a systematic uncertainty $\sigma_{\text{sys}}(b)$ for different fit ranges as

$$\begin{aligned}
\sigma_{\text{sys}}(b) &= 3.58 \pm 0.66 \cdot 10^{-3} \quad (39 - 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 4.24 \pm 0.78 \cdot 10^{-3} \quad (256 - 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 4.59 \pm 0.83 \cdot 10^{-3} \quad (365 - 743 \text{ keV}),
\end{aligned}$$

making it a significant correction close to, but still below, the targeted precision of $5 \cdot 10^{-3}$. The systematic uncertainty increases with a higher lower fit bound.

Pedestal and Offset Correction

Based on the results in Sec. 5.2.1 of a channel offset uncertainty of $\sigma(P) = 0.56$ ch for the pedestal and final offset correction, I obtain a combined systematic uncertainty $\sigma_{\text{sys}}(b)$ for both effects and different fit ranges as

$$\begin{aligned}
\sigma_{\text{sys}}(b) &= 0.78 \pm 0.22 \cdot 10^{-3} \quad (39 - 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 1.00 \pm 0.25 \cdot 10^{-3} \quad (256 - 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 0.93 \pm 0.24 \cdot 10^{-3} \quad (365 - 743 \text{ keV}),
\end{aligned}$$

making it a relevant, but not critical, correction for all tested fit ranges.

Radiative Corrections

Besides uncertainties from experimental and data processing corrections, we must also consider theoretical uncertainties of radiative corrections δ_R and the Coulomb interaction in F. Using the combined relative uncertainty of radiative, Coulomb and recoil corrections of $\sigma(\Delta_{\text{tot}})/\Delta_{\text{tot}} = 3.48 \cdot 10^{-3}$ based on the values from [Cir+23] for the integrated beta spectrum, I can estimate the systematic uncertainty $\sigma_{\text{sys}}(b)$ for different fit ranges as

$$\begin{aligned}
\sigma_{\text{sys}}(b) &= 1.8 \pm 1.2 \cdot 10^{-4} \quad (39 - 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 2.1 \pm 1.3 \cdot 10^{-4} \quad (256 - 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 2.8 \pm 1.3 \cdot 10^{-4} \quad (365 - 743 \text{ keV}),
\end{aligned}$$

making it a minor correction.

8.3 Correlation of Systematic Effects

Expanding the studies from variations of individual detector parameters to correlated variations, we can construct a correlation matrix for the relevant detector model parameters and estimate the systematic uncertainty for correlated corrections, i.e., the spatial response correction and the energy losses at the calibration holder support foils. The correlation matrix is

$$\rho(g, k_B, f_{\text{PE}}, P) = \begin{pmatrix} 1.000 & 0.035 & -0.221 & -0.149 \\ 0.035 & 1.000 & -0.090 & -0.093 \\ -0.221 & -0.090 & 1.000 & 0.225 \\ -0.149 & -0.093 & 0.225 & 1.000 \end{pmatrix} \tag{8.4}$$

and robust under fit range variations. Most notably, the parameters with the biggest impact on $\sigma_{\text{sys}}(b)$, the gain g and effective Birks coefficient k_B, have a positive correlation. The correlations of k_B to other variables are comparatively smaller, but k_B and its uncertainty are, e.g., approximately 16 times larger than the gain g and its uncertainty, respectively. To estimate the systematic uncertainty of correlated corrections, I vary all affected parameters combined with their respective uncertainty for each correction.

Spatial Response Correction

Based on the results in Sec. 6.1.2 of the spatial response correction uncertainty on the gain $\sigma(g)/g = 2.0 \cdot 10^{-4}$) and on the light yield $\sigma(f_{\text{PE}})/f_{\text{PE}} = 2.5 \cdot 10^{-4}$, I obtain a combined systematic uncertainty $\sigma_{\text{sys}}(b)$ for different fit ranges as

$$\begin{aligned}
\sigma_{\text{sys}}(b) &= 1.63 \pm 0.30 \cdot 10^{-3} \quad (\text{fit } 39 \text{ - } 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 1.93 \pm 0.35 \cdot 10^{-3} \quad (\text{fit } 256 \text{ - } 743 \text{ keV}) \\
\sigma_{\text{sys}}(b) &= 2.15 \pm 0.40 \cdot 10^{-3} \quad (\text{fit } 365 \text{ - } 743 \text{ keV}),
\end{aligned}$$

making it a relevant, but not critical, correction similar to the offset correction.

Calibration Support Foil Energy Loss Correction

Based on the initial estimated result in Sec. 6.2 of the correction induced by the energy losses in the calibration source holder with an uncertainty on the gain $\sigma(g)/g = 2.6 \cdot 10^{-4}$, the light

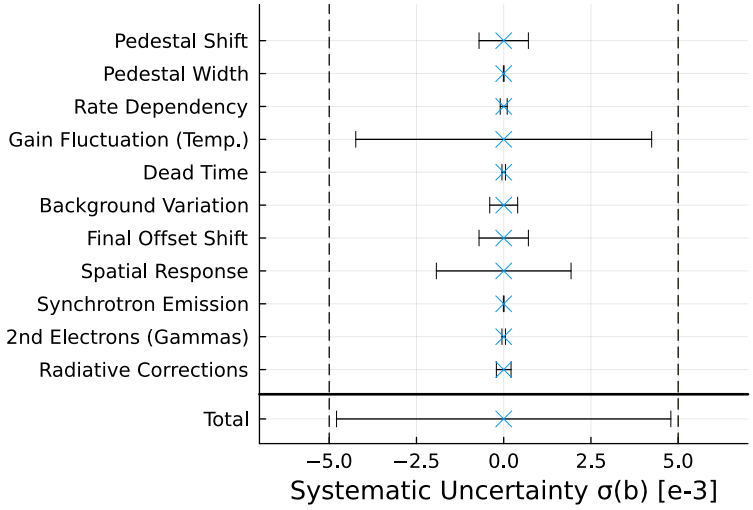


Figure 8.3 Preliminary systematic uncertainty budget for the Fierz interference term measurement with PERKEO III for a fit range of 256 - 743 keV. The overall systematic uncertainty for this fit range is $\sigma_{\text{sys}}(b) = 4.79 \pm 0.91 \cdot 10^{-3}$ and within the targeted precision of $\sigma_{\text{sys}}(b) \leq 5 \cdot 10^{-3}$ in [Lam19]. The result is mainly set by the temperature-induced gain fluctuation and spatial response correction.

yield $\sigma(f_{\text{PE}})/f_{\text{PE}} = 3.2 \cdot 10^{-3}$, and the effective Birks coefficient $\sigma(k_B)/k_B = 6.8 \cdot 10^{-3}$, I obtain a combined systematic uncertainty $\sigma_{\text{sys}}(b)$ for different fit ranges as

$$\begin{aligned}
\sigma_{\text{sys}}(b) &= 1.11 \pm 0.13 \cdot 10^{-2} \quad \text{(fit 39 - 743 keV)} \\
\sigma_{\text{sys}}(b) &= 1.23 \pm 0.17 \cdot 10^{-2} \quad \text{(fit 256 - 743 keV)} \\
\sigma_{\text{sys}}(b) &= 1.35 \pm 0.19 \cdot 10^{-2} \quad \text{(fit 365 - 743 keV),}
\end{aligned}$$

making it the potentially largest systematic uncertainty. However, this correction is presently set by fluctuations in the interpolation function that are caused by a lack of available data from the simulation studies. This correction has a great potential for improvement and is studied for [Ber24]. Also, solving the QDC non-linearity issue and enabling calibration fits with the lower energy sources ^{109}Cd and ^{139}Ce will decrease the systematic uncertainty and set limits on the Birks coefficient, whose correlated uncertainty is one dominant cause of the effect.

8.4 Systematic Uncertainty Budget

With the systematic uncertainties for all dominant corrections estimated with *Freya* and the upper limits of minor corrections in Sec. 5 and 6, I can calculate the preliminary systematic uncertainty for the corrections studied in this thesis as

$$\begin{aligned}
\sigma_{\text{sys}}(b) &= 4.04 \pm 0.78 \cdot 10^{-3} \quad \text{(fit 39 - 743 keV)} \\
\sigma_{\text{sys}}(b) &= 4.79 \pm 0.91 \cdot 10^{-3} \quad \text{(fit 256 - 743 keV)} \\
\sigma_{\text{sys}}(b) &= 5.18 \pm 0.97 \cdot 10^{-3} \quad \text{(fit 365 - 743 keV).}
\end{aligned}$$

The systematic error budget is illustrated in Fig. 8.3. For fits from 39 and 256 keV, the overall systematic uncertainty is within the targeted precision of $\sigma_{\text{sys}}(b) \leq 5 \cdot 10^{-3}$ as stated in the proposal [Lam19]. Studies to extract the Fierz interference term b with lower fits

bounds of 365 keV would exceed the targeted precision. For all fit ranges, the result and uncertainty are mainly set by the temperature-induced gain fluctuation and spatial response correction. In this work, the statistical uncertainty $\sigma_{\text{stat}}(b)$ limitations are not discussed. Based on the studies in [GN16], a simulation with 10^8 events leads to a statistical uncertainty of $\sigma_{\text{stat}}(b) = 8 \cdot 10^{-4}$. Neglecting the assumptions and simplifications in [GN16], this result implies a required $2.5 \cdot 10^6$ events of detected neutron beta decays to reach $\sigma_{\text{stat}}(b) = 5 \cdot 10^{-3}$. Our collected data set surpasses this requirement by two orders of magnitude, see Sec. 5.4.

Not included in the systematic uncertainty budget is the correction from the calibration support foil induced energy losses, as it is currently limited by the available data and including it is not meaningful. Also, not studied in the analysis are the discussed corrections in Sec. 6 that require the further simulation studies for the edge effect and the trigger function, and new experimental work for the QDC non-linearity. The simulation studies for the calibration support foil induced energy losses, edge effect, and trigger function are currently conducted for [Ber24].

8.5 Optimizing Detector Parameters

How can future measurements optimize their setup in order to reach a precision $\sigma_{\text{sys}}(b) \leq 5 \cdot 10^{-3}$? With *Freya*, we can test the potential of optimizing individual detector parameters, such as light yield f_{PE}, to probe how this would change the sensitivity of the systematic uncertainty $\sigma_{\text{sys}}(b)$.

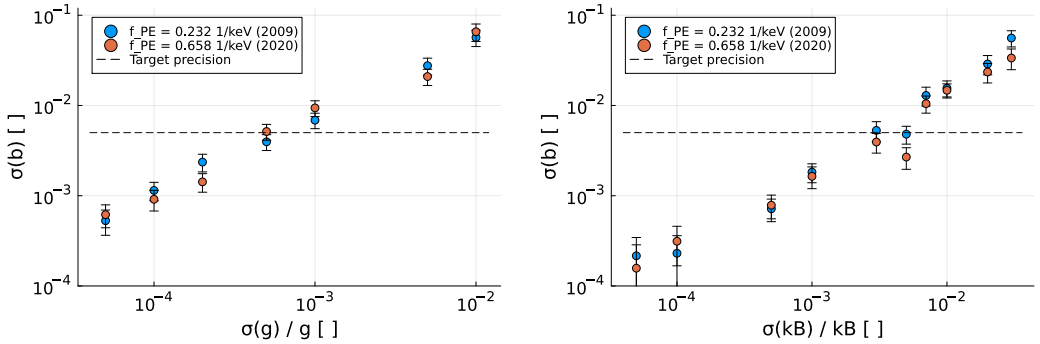
We have a concrete example for a light yield f_{PE} improvement when comparing the 2009 and 2020 PERKEO III detectors: The light yield improved from $f_{\text{PE}} \approx 0.232\,\text{keV}^{-1}$ to $f_{\text{PE}} \approx 0.658$ keV^{-1}. I simulate the systematic uncertainty $\sigma_{\text{sys}}(b)$ for two detectors with these two values to see if this affects the scaling laws from Fig. 8.2 for the two parameters with the biggest impact on $\sigma_{\text{sys}}(b)$, the gain g and the effective Birks non-linearity coefficient k_B. To only capture the direct effect of the light yield change, all other parameters are the same as in the previous section.

The results are shown in Fig. 8.4. The relative sensitivity to gain variations stays constant with a weighted average change of +5(11)%, which is expected, given the functional relation in Equ. (6.1). For the Birks coefficient, I obtain an indicated improvement of the relative sensitivity of −29(16)% for $\sigma(k_B)/k_B \geq 10^{-3}$. I neglect the smaller k_B variations, as they are within the baseline precision limit σ_{Freya}. This result indicates that the extraction of the Fierz interference term b from the electron beta spectrum is slightly improved by the increased energy resolution, and complements the relative sensitivity result in Fig. 8.2.

I repeat the test for a hypothetical non-linearity reduction, e.g., with new electronic readout systems, from $k_B = 470$ nm/keV to half the value $k_B = 235$ nm/keV. The relative sensitivity to gain variations shows indications of a slight improvement with a weighted average change of −15(11)%. For the Birks coefficient, I obtain no measurable improvement when using the same absolute uncertainty. Thus, an effective k_B improvement could slightly reduce the sensitivity of $\sigma_{\text{sys}}(b)$ to gain variations for future measurements.

Future Experiments and Goodhardt's Law

For previous PERKEO III measurements, the main focus of improvement for detectors was light yield, i.e., energy resolution, improvements, followed by temperature stability. Im-



(a) Variation of gain g (relative) **(b)** Variation of effective Birks coefficient k_B (relative)

Figure 8.4 Comparing the sensitivity of the systematic uncertainty $\sigma_{\text{sys}}(b)$ when changing the detector light yield with *Freya*. The sensitivity to gain variations does not change within statistical limits, and there is a slight trend of improvement for Birks coefficient variations of $-29(16)\%$ of the resulting systematic uncertainties.

proving the energy resolution improves the precision of calibration measurements with the gain and non-linearity parametrization and also indirectly improves critical corrections as the temperature-induced gain fluctuations and, arguably, the spatial response corrections.[3] Furthermore, while the extraction of the Fierz interference term b is barely sensitive to energy resolution improvements, such improvements are still essential to test our detector response model hypothesis with calibration measurements and discover yet unknown systematic effects that we must correct for.

Still, whether this policy suits future measurements that seek greater precision is worth questioning. Better energy resolution requires more corrections to be considered and also a higher precision of these corrections. As seen in this analysis of the 2019/20 PERKEO III measurement, one of the two new calibration source holder corrections, albeit limited by simulation data availability, affects the systematic uncertainty of b and was negligible in previous measurements. Also, the required higher-order contributions to the internal conversion coefficients highlight the general sensitivity to systematic effects. Further energy resolution increases will lead eventually to an even greater reliance on simulation results and will require estimations for systematic uncertainties of the theoretical description of the calibration sources.

This trend is a case of Goodhardt's law: *"When a measure becomes a target, it ceases to be a good measure"* [Hos96]. Not optimizing for greater energy resolutions for well-understood follow-up experiments might be counter-intuitive but lead to higher overall precision. The studies in this section support this approach. From a pure spectral perspective, improving individual detector parameters does not significantly improve the sensitivity of $\sigma_{\text{sys}}(b)$ to detector parameter uncertainties. However, improving the precision of individual corrections and the relative uncertainties of these detector parameters has the greatest potential for improvement for future measurements of b instead of, e.g., reducing the overall non-linearity value of new readout electronics or further increasing the light yield. This approach would simplify the optimization problem, as we can factorize it into individual optimizations and analyses of systematic effects to a certain extent. This strategy only works for searches of

[3] A greater light yield can correlate with better uniformity and, therefore, smaller spatial response correction, but not necessarily the precision of the correction.

gradient deviations, e.g., as the Fierz interference term b, and not for searches of localized spectral deviations, e.g., as for light-particle searches.

To conclude, the optimization of high-precision experiments is becoming more complex, requiring advanced modeling and optimization approaches, as discussed in Sec. 3 and presented in our works [LBM22; Lam+22a; Dor+23]. We can continue driving the precision boundaries by improving the systematic uncertainties of corrections and detector parameters. Better energy resolutions are certainly necessary to verify any results, but innovative ideas for the experiment design could also achieve crucial improvements.

9 Summary

High-precision experiments are vital tools to test the existing Standard Model of particle physics, gain insights into unknown fundamental processes, and advance our understanding. These experiments are often complex and intricate, with many tunable parameters that require exact calibrations, understanding, and precise quantification of systematic uncertainties. Studies of neutron beta decay are prime examples of such high-precision experiments. They provide an essential perspective to study the structure of the weak interaction and search for left-handed scalar and tensor contributions. These contributions are quantified in the Fierz interference term b, which is zero in the Standard Model. If non-zero, it would add an energy-dependent phase-space correction to the differential decay rate. With a precise measurement of the electron beta spectrum, we can probe the structure of the weak interaction for these contributions.

Within this thesis, we successfully conducted such a measurement with the PERKEO III spectrometer at the Institut Laue-Langevin in France in 2019/20. I present the experimental setup and how we designed new scientific equipment for this high-precision electron energy measurement. Also, I explain the newly developed data analysis tool and pipeline with corrections that expands on previous work for all studied corrections, introduce single-event corrections, and discuss three new ones. In doing so, the individual corrections are refined, leading to improvements of up to two orders of magnitude, and significant overall quality advances of the data set are highlighted and compared to the previous PERKEO III data set from 2009. Furthermore, my studies achieved important progress in understanding systematic effects, such as correlated electron emissions and their times-of-flight, the electronics integrator non-linearity, and spatial response corrections. The causes of remaining systematic issues are discussed and tested with spectral analysis of calibration source and neutron beta decay measurements. Relevant future studies with experiments and simulations are discussed and pointed out.

In this work, I studied new methodologies for high-precision experiments, specifically statistical machine learning and deep learning approaches to enhance experimental work by integrating physical motivation and statistical goals where applicable. I present techniques for post-experimental data analysis, pre-experimental optimization of experiment design, and enhancing the analysis of large-scale simulation data. My studies cover applications in low-energy particle physics to state-of-the-art computational astrophysics simulations [LBM22; Lam+22a; Dor+23]. In doing so, I successfully developed four analysis tools for data analysis, simulations, and hypothesis testing essential for analyzing the PERKEO III 2019/20 measurement. I derive the preliminary systematic uncertainty budget for the Fierz interference term b of all studied corrections as

$$\sigma_{\text{sys}}(b) = 4.79 \pm 0.91 \cdot 10^{-3},$$

which is in the targeted precision limit [Lam19] for the 2019/20 PERKEO III measurement campaign. I determine limitations and single out areas of improvement with the greatest potential to pave the way for improved searches of left-handed tensor contributions in the

weak interaction with this data set. Also, I discuss experiment optimization strategies to guide future high-precision experiments using new methodologies.

Bibliography

[Abb+22] R. Abbasi et al. “Graph Neural Networks for low-energy event classification and reconstruction in IceCube”. In: *Journal of Instrumentation* 17.11 (Nov. 2022), P11003. DOI: 10.1088/1748-0221/17/11/p11003.

[Abd+21] M. Abdar et al. “A review of uncertainty quantification in deep learning: Techniques, applications and challenges”. In: *Information Fusion* 76 (2021), pp. 243–297. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2021.05.008.

[Abe+20] C. Abel et al. “Measurement of the Permanent Electric Dipole Moment of the Neutron”. In: *Phys. Rev. Lett.* 124 (8 Feb. 2020), pp. 081803-1–081803-6. DOI: 10.1103/PhysRevLett.124.081803.

[Abe+06] H. Abele et al. “Characterization of a ballistic supermirror neutron guide”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 562.1 (2006), pp. 407–417. DOI: 10.1016/j.nima.2006.03.020.

[Abe+11] C. Aberle et al. “Light output of Double Chooz scintillators for low energy electrons”. In: *Journal of Instrumentation* 6.11 (Nov. 2011), P11006. DOI: 10.1088/1748-0221/6/11/P11006.

[Ago+03] S. Agostinelli et al. “GEANT4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8.

[Ake+19] M. Aker et al. “Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN”. In: *Physical Review Letters* 123 (22 Nov. 2019), pp. 221802-1–221802-10. DOI: 10.1103/PhysRevLett.123.221802.

[Aki+19] T. Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019.

[AL08] M. Alvarez and N. Lawrence. “Sparse convolved Gaussian processes for multi-output regression”. In: *Advances in Neural Information Processing Systems* 21 (2008).

[AVH22] S. Andrianomena, F. Villaescusa-Navarro, and S. Hassan. “Emulating cosmological multifields with generative adversarial networks”. In: *preprint arXiv:2211.05000* (2022).

[Ant19] M. Antony. “Optimization and commissioning of detectors for beta spectroscopy of the decay of the free neutron”. BA Thesis. Technical University Munich, 2019.

[Bar12] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012. DOI: 10.1017/CBO9780511804779.

[Bay+17] A. G. Baydin et al. “Automatic Differentiation in Machine Learning: A Survey”. In: *Journal of Machine Learning Research* 18.1 (2017), pp. 5595–5637. ISSN: 1532–4435. DOI: 10.5555/3122009.3242010.

[Bé+16] M.-M. Bé et al. *Table of Radionuclides*. Vol. 8. Monographie BIPM-5. Pavillon de Breteuil, F-92310 Sèvres, France: Bureau International des Poids et Mesures, 2016. ISBN: 978-92-822-2264-5.

[Bec+16] A. M. Beck et al. "An improved SPH scheme for cosmological simulations". In: *Monthly Notices of the Royal Astronomical Society* 455.2 (Jan. 2016), pp. 2110–2130. DOI: 10.1093/mnras/stv2443.

[BB82] H. Behrens and W. Bühring. *Electron Radial Wave Functions and Nuclear Beta-Decay*. Vol. The international series of monographs on physics. Oxford University Press, 1982.

[Bel+19] M. Belkin et al. "Reconciling modern machine-learning practice and the classical bias-variance trade-off". In: *Proceedings of the National Academy of Sciences* 116.32 (2019), pp. 15849–15854.

[Ber+17] M.J. Berger et al. "ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions (version 2.0.1)". In: *National Institute of Standards and Technology* (2017).

[Ber+11] J. Bergstra et al. "Algorithms for hyper-parameter optimization". In: *Advances in neural information processing systems* 24 (2011).

[Ber+18] J. Berk et al. "Exploration enhanced expected improvement for bayesian optimization". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2018, pp. 621–637.

[Ber18] K. Bernert. "Development of a new scintillation detector for beta spectroscopy". MA Thesis. Technical University Munich, 2018.

[Ber24] K. Bernert. *Personal communication: Upcoming dissertation. TUM*. 2024.

[Ber98] E. Bertschinger. "Simulations of Structure Formation in the Universe". In: *Annual Review of Astronomy and Astrophysics* 36 (Jan. 1998), pp. 599–654. DOI: 10.1146/annurev.astro.36.1.599.

[Bes22] M. Bestehorn. "Analysis of detector-based corrections for measuring the Fierz interference term from neutron beta-decay". MA Thesis. Technical University Munich, 2022.

[Bha+12] T. Bhattacharya et al. "Probing Novel Scalar and Tensor Interactions from (Ultra)Cold Neutrons to the LHC". In: *Physical Review D* 85 (5 2012), pp. 054512-01–054512-26.

[Bin+18] E. Bingham et al. "Pyro: Deep Universal Probabilistic Programming". In: *Journal of Machine Learning Research* (2018).

[Bir51] J. B. Birks. "Scintillations from Organic Crystals: Specific Fluorescence and Relative Response to Different Radiations". In: *Proceedings of the Physical Society. Section A* 64.10 (1951), pp. 874–877. DOI: 10.1088/0370-1298/64/10/303.

[BKM17] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. "Variational inference: A review for statisticians". In: *Journal of the American statistical Association* 112.518 (2017), pp. 859–877.

[Bös+22] L. M. Böss et al. "CRESCENDO: an on-the-fly Fokker–Planck solver for spectral cosmic rays in cosmological simulations". In: *Monthly Notices of the Royal Astronomical Society* 519.1 (Dec. 2022), pp. 548–572. ISSN: 0035-8711. DOI: 10.1093/mnras/stac3584.

[Bur+22] M. T. Burkey et al. "Improved Limit on Tensor Currents in the Weak Interaction from ^{8}Li β Decay". In: *Physical Review Letters* 128 (20 2022), pp. 202502-1–202502-6. DOI: 10.1103/PhysRevLett.128.202502.

[But+23] A. Butter et al. “Back to the Formula – LHC Edition”. In: *preprint arXiv:2109.10414* (2023).

[Chu+19] T. E. Chupp et al. “Electric dipole moments of atoms, molecules, nuclei, and particles”. In: *Reviews of Modern Physics* 91 (1 Jan. 2019), pp. 015001-1–015001-55. DOI: 10.1103/RevModPhys.91.015001.

[Cir+22] V. Cirigliano et al. “Pion-induced Radiative Corrections to Neutron β Decay”. In: *Physical Review Letters* 129.12 (Sept. 2022), pp. 121801-1–121801-7. DOI: 10.1103/PhysRevLett.129.121801.

[Cir+23] V. Cirigliano et al. “Effective field theory for radiative corrections to charged-current processes I: Vector coupling”. In: *preprint arXiv:2306.03138* (2023).

[Cis+20] E. Cisbani et al. “AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case”. In: *Journal of Instrumentation* 15.05 (2020), P05009.

[CB83] E. D. Commins and P. H. Bucksbaum. *Weak interactions of leptons and quarks*. Cambridge University Press, 1983.

[Cow+56] C. L. Cowan Jr. et al. “Detection of the free neutrino: A Confirmation”. In: *Science* 124.3212 (1956), pp. 103–104. DOI: 10.1126/science.124.3212.103.

[CK15] J. Cui and R. V. Krems. “Gaussian Process model for collision dynamics of complex molecules”. In: *Physical review letters* 115.7 (2015), p. 073202.

[Dav+85] M. Davis et al. “The evolution of large-scale structure in a universe dominated by cold dark matter”. In: *ApJ* 292 (May 1985), pp. 371–394. DOI: 10.1086/163168.

[Deg+22] J. Degrave et al. “Magnetic control of tokamak plasmas through deep reinforcement learning”. In: *Nature* 602 (Feb. 2022), pp. 414–419. DOI: 10.1038/s41586-021-04301-9.

[Dia14] J. Diaz. “Tests of Lorentz Symmetry in Single Beta Decay”. In: *Advances in High Energy Physics* 2014 (2014), pp. 1–7. DOI: 10.1155/2014/305298.

[Diw20] M. Diwan. “Statistics of the Charge Spectrum of Photo-Multipliers and Methods for Absolute Calibration”. In: *Journal of Instrumentation* 15.P02001 (2020), pp. 1–12. DOI: 10.1088/1748-0221/15/02/P02001.

[Dol+05] K. Dolag et al. “Turbulent gas motions in galaxy cluster simulations: the role of smoothed particle hydrodynamics viscosity”. In: *Monthly Notices of the Royal Astronomical Society* 364 (3 Dec. 2005), pp. 753–772. DOI: 10.1111/j.1365-2966.2005.09630.x.

[DB14] J. Donnert and G. Brunetti. “An efficient Fokker-Planck solver and its application to stochastic particle acceleration in galaxy clusters”. In: *Monthly Notices of the Royal Astronomical Society* 443.4 (Oct. 2014), pp. 3564–3577. DOI: 10.1093/mnras/stu1417.

[Doo+22] L. Doorenbos et al. “Generating astronomical spectra from photometry with conditional diffusion models”. In: *preprint arXiv:2211.05556* (2022).

[Dor+23] T. Dorigo et al. “Toward the end-to-end optimization of particle physics instruments with differentiable programming”. In: *Reviews in Physics* 10 (2023), p. 100085. ISSN: 2405-4283. DOI: 10.1016/j.revip.2023.100085.

[DM05] P. Drineas and M. W. Mahoney. “On the Nystrom Method for Approximating a Gram Matrix for Improved Kernel-Based Learning”. In: *Journal of Machine Learning Research* 6.72 (2005), pp. 2153–2175.

[Dub15] D. Dubbers. “Magnetic guidance of charged particles”. In: *Physics Letters B* 748 (2015), pp. 306–310. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2015.07.004.

[DM21] D. Dubbers and B. Märkisch. “Precise Measurements of the Decay of Free Neutrons”. In: *Annual Review of Nuclear and Particle Science* 71 (2021), pp. 139–163. DOI: 10.1146/annurev-nucl-102419-043156.

[DS16] D. Dubbers and U. Schmidt. “Generation of narrow peaks in spectroscopy of charged particles”. In: 837 (2016), pp. 50–57. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.09.004.

[Dub+08] D. Dubbers et al. “A clean, bright, and versatile source of neutron decay products”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 596.2 (2008), pp. 238–247. ISSN: 0168-9002. DOI: 10.1016/j.nima.2008.07.157.

[Dub+14] D. Dubbers et al. “The point spread function of electrons in a magnetic field, and the decay of the free neutron”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 763 (2014), pp. 112–119. ISSN: 0168-9002. DOI: 10.1016/j.nima.2014.06.020.

[Dur+20] J. Duris et al. “Bayesian optimization of a free-electron laser”. In: *Physical review letters* 124.12 (2020), p. 124801.

[Eks+19] A. Ekström et al. “Bayesian optimization in ab initio nuclear physics”. In: *Journal of Physics G: Nuclear and Particle Physics* 46.9 (2019), p. 095101.

[Fal+21] A. Falkowski et al. “Constraints on subleading interactions in beta decay Lagrangian”. In: *preprint arXiv:2112.07688* (2021).

[Fal22] L. Fallboehmer. “Analysis of Backscattering Processes for Electron Energy Detection in Neutron Beta Decay”. BA Thesis. Technical University Munich, 2022.

[Fer34] E. Fermi. “Versuch einer Theorie der Beta-Strahlen. I”. In: *Zeitschrift für Physik* 88 (3 1934), pp. 0044–3328. DOI: 10.1007/BF01351864.

[Fie+17] D. Fielding et al. “How supernovae launch galactic winds?” In: *Monthly Notices of the Royal Astronomical Society: Letters* 470.1 (Sept. 2017), pp. L39–L43. DOI: 10.1093/mnrasl/slx072.

[Fie37] M. Fierz. “Zur Fermischen Theorie des Beta-Zerfalls”. In: *Zeitschrift für Physik* 104 (1937), pp. 553–565. DOI: 10.1007/BF01330070.

[Fie39] M. Fierz. “Über die relativistische Theorie kräftefreier Bose-Felder mit beliebigem Spin”. In: *Helvetica Physica Acta* 12 (1939), pp. 3–37.

[Fos+21] A. Foster et al. “Deep adaptive design: Amortizing sequential bayesian experimental design”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 3384–3395.

[Fry+19] J. Fry et al. “The Nab Experiment: A Precision Measurement of Unpolarized Neutron Beta Decay”. In: *EPJ Web of Conferences*. Vol. 219. EDP Sciences. 2019, pp. 04007-1–04007-6. DOI: 10.1051/epjconf/201921904002.

[Gar+18] J. Gardner et al. “Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration”. In: *Advances in neural information processing systems* 31 (2018).

[GFV21] A. Ghouse, L. Faes, and G. Valenza. “Inferring directionality of coupled dynamical systems using Gaussian process priors: Application on neurovascular systems”. In: *Physical Review E* 104.6 (2021), p. 064208.

[Gib19] S. Gibbons. *pyfof - Friends of Friends cluster finding in python*. 2019. URL: https://github.com/simongibbons/pyfof (visited on 05/08/2022).

[GJL95] F. Glück, I. Joó, and J. Last. "Measurable parameters of neutron decay". In: *Nuclear Physics A* 593.2 (1995), pp. 125–150. ISSN: 0375-9474. DOI: 10.1016/0375-9474(95)00354-4.

[GZS19] C. Glymour, K. Zhang, and P. Spirtes. "Review of Causal Discovery Methods Based on Graphical Models". In: *Frontiers in Genetics* 10 (2019). ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00524.

[Gon+21] F. M. Gonzalez et al. "Improved Neutron Lifetime Measurement with UCNτ". In: *Physical Review Letters* 127 (16 Oct. 2021), pp. 162501-1–162501-6. DOI: 10.1103/PhysRevLett.127.162501.

[GN16] M. González-Alonso and O. Naviliat-Cuncic. "Kinematic sensitivity to the Fierz term of β-decay differential spectra". In: *Physical Review C* 94 (3 Sept. 2016), pp. 035503-1–035503-7. DOI: 10.1103/PhysRevC.94.035503.

[GNS19] M. González-Alonso, O. Naviliat-Cuncic, and N. Severijns. "New physics searches in nuclear and neutron β decay". In: *Progress in Particle and Nuclear Physics* 104 (2019), pp. 165–223. DOI: 10.1016/j.ppnp.2018.08.002.

[Gor19] M. Gorchtein. "γW Box Inside Out: Nuclear Polarizabilities Distort the Beta Decay Spectrum". In: *Physical Review Letters* 123 (4 June 2019), pp. 042503-1–042503-5. DOI: 10.1103/PhysRevLett.123.042503.

[GS21] M. Gorchtein and CY. Seng. "Dispersion relation analysis of the radiative corrections to g_A in the neutron β-decay". In: *Journal of High Energy Physics* 2021.10 (2021), pp. 1–27. DOI: 10.1007/JHEP10(2021)053.

[HV14] R. van Haasteren and M. Vallisneri. "New advances in the Gaussian-process approach to pulsar-timing data analysis". In: *Phys. Rev. D* 90.10 (2014), p. 104012. DOI: 10.48550/arXiv.1407.1838.

[HMK21] A. Hajibabaei, C. W. Myung, and K. S. Kim. "Sparse Gaussian process potentials: Application to lithium diffusivity in superionic conducting solid electrolytes". In: *Physical Review B* 103.21 (2021), p. 214102.

[HNR19] J. Halverson, B. Nelson, and F. Ruehle. "Branes with brains: exploring string vacua with deep reinforcement learning". In: *Journal of High Energy Physics* 2019.6 (June 2019). DOI: 10.1007/jhep06(2019)003.

[Han+21] A. Hanuka et al. "Physics model-informed Gaussian process for online optimization of particle accelerators". In: *Physical Review Accelerators and Beams* 24.7 (2021), p. 072802.

[HT15] J. C. Hardy and I. S. Towner. "Superallowed $0^+ \to 0^+$ nuclear β decays: 2014 critical survey, with precise results for V_{ud} and CKM unitarity". In: *Phys. Rev. C* 91 (2 Feb. 2015), p. 025501. DOI: 10.1103/PhysRevC.91.025501.

[HT20] J. C. Hardy and I. S. Towner. "Superallowed $0^+ \to 0^+$ nuclear β decays: 2020 critical survey, with implications for V_{ud} and CKM unitarity". In: *Physical Review C* 102 (4 Oct. 2020), pp. 045501-1–045501-29. DOI: 10.1103/PhysRevC.102.045501.

[Has+23] H. Hashemi et al. "Ultra-High-Resolution Detector Simulation with Intra-Event Aware GAN and Self-Supervised Relational Reasoning". In: *preprint arXiv:2303.08046* (2023).

[HMR21] F. Hensel, M. Moor, and B. Rieck. "A Survey of Topological Machine Learning Methods". In: *Frontiers in Artificial Intelligence* 4 (May 2021), p. 681108. DOI: 10.3389/frai.2021.681108.

[HFL13] J. Hensman, N. Fusi, and Neil D. Lawrence. “Gaussian Processes for Big Data”. In: *preprint arXiv:1309.6835* (Sept. 2013).

[HMG14] J. Hensman, A. Matthews, and Z. Ghahramani. “Scalable Variational Gaussian Process Classification”. In: *preprint arXiv:1411.2005* (Nov. 2014).

[Hen+15] J. Hensman et al. “MCMC for Variationally Sparse Gaussian Processes”. In: *preprint arXiv:1506.04000* (June 2015).

[Hic+17] K. P. Hickerson et al. “First direct constraints on Fierz interference in free-neutron β decay”. In: *Physical Review C* 96 (4 2017), pp. 042501-1–042501-6. DOI: 10.1103/PhysRevC.96.042501.

[Hol74] B. R. Holstein. “Recoil effects in allowed beta decay: The elementary particle approach”. In: *Reviews of Modern Physics* 46 (4 Oct. 1974), pp. 789–814. DOI: 10.1103/RevModPhys.46.789.

[Hos96] K. Hoskin. *The ‘Awful Idea of Accountability’: Inscribing People into the Measurement of Objects.* Vol. Accountability: Power, Ethos and the Technologies of Managing. International Thomson Business Press, 1996.

[Huy+16] X. Huyan et al. “Toward a measurement of weak magnetism in 6He decay”. In: *Hyperfine Interactions* 237.93 (2016), pp. 1–7. DOI: 10.1007/s10751-016-1303-8.

[IWY17] P. Ilten, M. Williams, and Y. Yang. “Event generator tuning using Bayesian optimization”. In: *Journal of Instrumentation* 12.04 (2017), P04028.

[IPT13] A. N. Ivanov, M. Pitschmann, and N. I. Troitskaya. “Neutron β^- decay as a laboratory for testing the standard model”. In: *Phys. Rev. D* 88 (7 Oct. 2013), p. 073002. DOI: 10.1103/PhysRevD.88.073002.

[Jac75] J. D. Jackson. *Classical electrodynamics*. John Wiley & Sons, 1975. ISBN: 0-471-30932-X.

[JTW57] J. D. Jackson, S. B. Treiman, and H. W. Wyld. “Possible Tests of Time Reversal Invariance in Beta Decay”. In: *Physical Review* 106 (3 May 1957), pp. 517–521. DOI: 10.1103/PhysRev.106.517.

[JR10] J. Jaeckel and A. Ringwald. “The Low-Energy Frontier of Particle Physics”. In: *Annual Review of Nuclear and Particle Science* 60 (2010), pp. 405–437. DOI: 10.1146/annurev.nucl.012809.104433.

[Jal+21] S. Jalas et al. “Bayesian Optimization of a Laser-Plasma Accelerator”. In: *Physical review letters* 126.10 (2021), p. 104801.

[Jan+12] H. T. Janka et al. “Core-collapse supernovae: Reflections and directions”. In: *Progress of Theoretical and Experimental Physics* 2012.1, 01A309 (Dec. 2012), pp. 1–33. DOI: 10.1093/ptep/pts067.

[JC16] I. T. Jolliffe and J. Cadima. “Principal component analysis: a review and recent developments”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), p. 20150202. DOI: 10.1098/rsta.2015.0202.

[Jum+21] J. Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (Aug. 2021). Number: 7873 Publisher: Nature Publishing Group, pp. 583–589. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2.

[KCW22] K. Karchev, A. Coogan, and C. Weniger. “Strong-lensing source reconstruction with variationally optimized Gaussian processes”. In: *MNRAS* 512.1 (May 2022), pp. 661–685. DOI: 10.1093/mnras/stac311. arXiv: 2105.09465 [astro-ph.IM].

[Kas+20] M. Kasim et al. “Up to two billion times acceleration of scientific simulations with deep neural architecture search”. In: *APS Division of Plasma Physics Meeting Abstracts* 2020 (2020), BO05–001.

[Kib+08] T. Kibedi et al. “Evaluation of theoretical conversion coefficients using BrIcc”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (May 2008), pp. 202–229. DOI: 10.1016/j.nima.2008.02.051.

[KO15] C. G. Kim and E. C. Ostriker. “Momentum Injection by Supernovae in the Interstellar Medium”. In: *The Astrophysical Journal* 802.2, 99 (Apr. 2015), pp. 1–25. DOI: 10.1088/0004-637X/802/2/99.

[KB14] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *preprint arXiv:1412.6980* (Dec. 2014).

[Kis19] O. Kiss. “Bayesian Optimization for machine learning algorithms in the context of Higgs searches at the CMS experiment”. arXiv:1911.02501. 2019.

[KW19] M. J. Kochenderfer and T. A. Wheeler. *Algorithms for Optimization*. MIT Press, 2019.

[Kol99] U. van Kolck. “Effective field theory of nuclear forces”. In: *Progress in Particle and Nuclear Physics* 43 (1999), pp. 337–418.

[Kre04] M. Kreuz. “Messung von Winkelkorrelationen im Zerfall polarisierter Neutronen mit dem Spektrometer Perkeo II”. Diploma Thesis. Heidelberg University, 2004.

[Kro20] A. Kropf. “Energy calibration of beta detectors and proton asymmetry in neutron beta decay”. MA Thesis. Technical University Munich, 2020.

[KL51] S. Kullback and R. A. Leibler. “On information and sufficiency”. In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.

[Lag+22] M. Lagrange et al. “TomOpt: Differentiable Muon-Tomography Optimization”. In: *EGU General Assembly Conference Abstracts*. 2022, EGU22–9470.

[Lam19] M. Lamparth. “Determination of the Fierz interference term from neutron beta decay”. Experiment Proposal. Institute Laue-Langevin, 2019.

[LBM22] M. Lamparth, M. Bestehorn, and B. Märkisch. “Gaussian Processes and Bayesian Optimization for High Precision Experiments”. In: *preprint arXiv:2205.07625* (May 2022). DOI: 10.48550/arXiv.2205.07625.

[LR23] M. Lamparth and A. Reuel. “Analyzing and editing inner mechanisms of backdoored language models”. In: *preprint arXiv:2302.12461* (2023).

[Lam+22a] M. Lamparth et al. “Virgo: Scalable Unsupervised Classification of Cosmological Shock Waves”. In: *preprint arXiv:2208.06859* (2022).

[Lam+22b] M. Lamparth et al. “Virgo: Scalable Unsupervised Classification of Cosmological Shock Waves”. Machine Learning and the Physical Science Workshop at the 36th conference on Neural Information Processing Systems. 2022.

[Lec18] F. Leclercq. “Bayesian optimization for likelihood-free cosmological inference”. In: *Phys. Rev. D* 98.6, 063511 (Sept. 2018), p. 063511. DOI: 10.1103/PhysRevD.98.063511. arXiv: 1805.07152 [astro-ph.CO].

[LY56] T. D. Lee and C. N. Yang. “Question of Parity Conservation in Weak Interactions”. In: *Physical Review* 104 (1 Oct. 1956), pp. 254–258. DOI: 10.1103/PhysRev.104.254.

[Leo94] W. R. Leo. *Techniques for Nuclear and Particle Physics Experiments*. Vol. 2. Springer, 1994, pp. 122–127. ISBN: 3-540-57280-5. DOI: 10.1007/978-3-642-57920-2.

[Lin56] D. V. Lindley. “On a measure of the information provided by an experiment”. In: *The Annals of Mathematical Statistics* 27.4 (1956), pp. 986–1005.

[LZB22] C. Liu, L. Zhu, and M. Belkin. “Loss landscapes and optimization in over-parameterized non-linear systems and neural networks”. In: *Applied and Computational Harmonic Analysis* 59 (2022), pp. 85–116. DOI: 10.1016/j.acha.2021.12.009.

[MBB18] S. Ma, R. Bassily, and M. Belkin. “The power of interpolation: Understanding the effectiveness of SGD in modern over-parametrized learning”. In: (2018), pp. 3325–3334.

[Mär06] B. Märkisch. “Das Spektrometer PERKEO III und der Zerfall des freien Neutrons”. Ph.D. Dissertation. Heidelberg University, 2006.

[Mär22] B. Märkisch. *Personal communication*. 2022.

[Mär+09] B. Märkisch et al. “The new neutron decay spectrometer PERKEO III”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 611.2 (2009), pp. 216–218. ISSN: 0168-9002. DOI: 10.1016/j.nima.2009.07.066.

[Mär+19] B. Märkisch et al. “Measurement of the Weak Axial-Vector Coupling Constant in the Decay of Free Neutrons Using a Pulsed Cold Neutron Beam”. In: *Physical Review Letters* 122 (24 June 2019), pp. 242501-1–242501-7. DOI: 10.1103/PhysRevLett.122.242501.

[Men+22] K. Meng et al. “Locating and Editing Factual Associations in GPT”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2022.

[Mes11] H. Mest. “Measurement of the Beta-Asymmetry in the Decay of Free Polarized Neutrons with the Spectrometer PERKEO III”. Ph.D. Dissertation. Heidelberg University, 2011.

[Mil+22] A. C. Miller et al. “Mapping Interstellar Dust with Gaussian Processes”. In: *preprint arXiv:2202.06797* (Feb. 2022).

[Moo+16] C. J. Moore et al. “Improving gravitational-wave parameter estimation using Gaussian process regression”. In: *Phys. Rev. D* 93.6, 064001 (Mar. 2016), p. 064001. DOI: 10.1103/PhysRevD.93.064001. arXiv: 1509.04066 [gr-qc].

[MM33] N. F. Mott and H. S. W. Massey. *The Theory of Atomic Collisions*. Vol. The international series of monographs on physics. Oxford: Clarendon Press, 1933, p. 283.

[Mue+22] M. Mueller et al. “Leveraging Stochastic Predictions of Bayesian Neural Networks for Fluid Simulations”. In: *preprint arXiv:2205.01222* (2022).

[Mun06] D. Mund. “Messung der Betaasymmetrie A im Neutronenzerfall”. Ph.D. Dissertation. Heidelberg University, 2006.

[Mun+13] D. Mund et al. “Determination of the Weak Axial Vector Coupling $\lambda=g_A/g_V$ from a Measurement of the β-Asymmetry Parameter A in Neutron Beta Decay”. In: *Phys. Rev. Lett.* 110 (17 Apr. 2013), p. 172502. DOI: 10.1103/PhysRevLett.110.172502.

[NO17] T. Naab and J. P. Ostriker. “Theoretical Challenges in Galaxy Formation”. In: *Annual Review of Astronomy and Astrophysics* 55.1 (Aug. 2017), pp. 59–109. DOI: 10.1146/annurev-astro-081913-040019.

[Nea96] R. M. Neal. *Bayesian learning for neural networks*. Springer Science & Business Media, 1996.

[PZŽ16] P. M. Pardalos, A. Zhigljavsky, and J. Žilinskas. *Advances in stochastic and deterministic global optimization*. Springer, 2016.

[Pas+19] A. Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *preprint arXiv:1912.01703* (Dec. 2019).

[Pea+17] J. Pearkes et al. “Jet Constituents for Deep Neural Network Based Top Quark Tagging”. In: *preprint arXiv:1704.02124* (2017).

[Pea09a] J. Pearl. “Causal inference in statistics: An overview”. In: *Statistics Surveys* 3 (2009), pp. 96–146. DOI: 10.1214/09-SS057.

[Pea09b] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009. ISBN: 052189560X.

[Ped+12] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *preprint arXiv:1201.0490* (Jan. 2012).

[Pfe+22] T. Pfeil et al. “A Neural Network Subgrid Model of the Early Stages of Planet Formation”. In: *preprint arXiv:2211.04160* (2022).

[Pfr+07] C. Pfrommer et al. “Simulating cosmic rays in clusters of galaxies - I. Effects on the Sunyaev-Zel’dovich effect and the X-ray emission”. In: *Monthly Notices of the Royal Astronomical Society* 378.2 (June 2007), pp. 385–408. DOI: 10.1111/j.1365-2966.2007.11732.x.

[Pfr+17] C. Pfrommer et al. “Simulating cosmic ray physics on a moving mesh”. In: *Monthly Notices of the Royal Astronomical Society* 465.4 (Mar. 2017), pp. 4500–4529. DOI: 10.1093/mnras/stw2941.

[Pho96] Hamamtsu Photonics. *Photomultiplier Tube R5924 (Datasheet)*. 1996.

[Pös+21] T. Pöschl et al. “Measurement of ionization quenching in plastic scintillators”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 988 (2021), pp. 164865-1–164865-8. ISSN: 0168-9002. DOI: 10.1016/j.nima.2020.164865.

[QR05] J. Quiñonero-Candela and C. E. Rasmussen. “A Unifying View of Sparse Approximate Gaussian Process Regression”. In: *Journal of Machine Learning Research* 6.65 (2005), pp. 1939–1959.

[Raf16] L. Raffelt. “Measurement of the Proton Asymmetry (C) in free neutron Beta-decay with Perkeo III”. Ph.D. Dissertation. Heidelberg University, 2016.

[RW06] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Second. MIT Press, 2006. DOI: 10.7551/mitpress/3206.001.0001.

[Rei99] J. Reich. “Winkelkorrelationen Im Zerfall Polarisierter Neutronen”. Ph.D. Dissertation. Universität Heidelberg, 1999.

[RLP16] J. Revels, M. Lubin, and T. Papamarkou. “Forward-Mode Automatic Differentiation in Julia”. In: *preprint arXiv:1607.07892* (2016).

[Roi18] C. Roick. “Particle Detection and Proton Asymmetry in Neutron Beta Decay”. Ph.D. Dissertation. Technical University Munich, 2018.

[Roi+19] C. Roick et al. “Undetected Electron Backscattering in Perkeo III”. In: *EPJ Web of Conferences* 219.04005 (2019), pp. 1–6.

[RHE21] R. Roussel, A. Hanuka, and A. Edelen. “Multiobjective Bayesian optimization for online accelerator tuning”. In: *Phys. Rev. Accel. Beams* 24 (6 June 2021), p. 062801. DOI: 10.1103/PhysRevAccelBeams.24.062801.

[Sai21] Saint-Gobain. “BC-400 / BC-404 / BC-408 / BC-412 / BC-416 Premium Plastic Scintillators”. In: (2021).

[SD17] H. Salimbeni and M. Deisenroth. “Doubly stochastic variational inference for deep Gaussian processes”. In: *Advances in neural information processing systems* 30 (2017).

[Sau18] H. Saul. “Energy Dependence of the Beta Asymmetry in Neutron Beta Decay”. Ph.D. Dissertation. Technische Universität Wien and Technische Universität München, 2018.

[Sau+20] H. Saul et al. “Limit on the Fierz Interference Term *b* from a Measurement of the Beta Asymmetry in Neutron Decay”. In: *Phys. Rev. Lett.* 125 (11 Sept. 2020), pp. 112501-1–112501-7. DOI: 10.1103/PhysRevLett.125.112501.

[SS15] K. Schaal and V. Springel. “Shock finding on a moving mesh - I. Shock statistics in non-radiative cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 446.4 (Feb. 2015), pp. 3992–4007. DOI: 10.1093/mnras/stu2386.

[SSM98] B. Schölkopf, A. Smola, and K.-R. Müller. “Nonlinear Component Analysis as a Kernel Eigenvalue Problem”. In: *Neural Computation* 10.5 (1998), pp. 1299–1319. DOI: 10.1162/089976698300017467.

[SS18] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2018. DOI: 10.7551/mitpress/4175.001.0001.

[Sed46] L. I. Sedov. “Propagation of strong shock waves”. In: *Journal of Applied Mathematics and Mechanics* 10 (Jan. 1946), pp. 241–250.

[Sev14] N. Severijns. “Correlation and spectrum shape measurements in Beta-decay probing the standard model”. In: *Journal of Physics G: Nuclear and Particle Physics* 41.11 (Oct. 2014), p. 114006. DOI: 10.1088/0954-3899/41/11/114006.

[SKL12] A. Shafieloo, A. G. Kim, and E. V. Linder. “Gaussian process cosmography”. In: *Phys. Rev. D* 85.12 (2012), p. 123530. DOI: 10.1103/physrevd.85.123530.

[Sha+16] B. Shahriari et al. “Taking the Human Out of the Loop: A Review of Bayesian Optimization”. In: *Proceedings of the IEEE* 104.1 (2016), pp. 148–175. DOI: 10.1109/JPROC.2015.2494218.

[STM20] J. Shi, M. Titsias, and A. Mnih. “Sparse orthogonal variational inference for Gaussian processes”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 1932–1942.

[Sir67] A. Sirlin. “General Properties of the Electromagnetic Corrections to the Beta Decay of a Physical Nucleon”. In: *Phys. Rev.* 164 (5 Dec. 1967), pp. 1767–1775. DOI: 10.1103/PhysRev.164.1767.

[Sol20] T. Soldner. *Personal communication*. 2020.

[SD15] R. S. Somerville and R. Davé. “Physical Models of Galaxy Formation in a Cosmological Framework”. In: *Annual Review of Astronomy and Astrophysics* 53 (Aug. 2015), pp. 51–113. DOI: 10.1146/annurev-astro-082812-140951.

[Spr05] V. Springel. “The cosmological simulation code GADGET-2”. In: *Monthly Notices of the Royal Astronomical Society* 364.4 (Dec. 2005), pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x.

[Ste+20] U. P. Steinwandel et al. “Hot phase generation by supernovae in ISM simulations: resolution, chemistry, and thermal conduction”. In: *Monthly Notices of the Royal Astronomical Society* 495.1 (June 2020), pp. 1035–1060. DOI: 10.1093/mnras/staa821.

[Ste+22] U. P. Steinwandel et al. “On the Small-scale Turbulent Dynamo in the Intracluster Medium: A Comparison to Dynamo Theory”. In: *The Astrophysical Journal* 933 (2 July 2022), pp. 1–25. DOI: 10.3847/1538-4357/ac715c.

[SSH22] V. Stimper, B. Schölkopf, and J.M. Hernández-Lobato. “Resampling Base Distributions of Normalizing Flows”. In: *preprint arXiv:2110.15828* (2022).

[Sun+20] X. Sun et al. “Improved limits on Fierz interference using asymmetry measurements from the Ultracold Neutron Asymmetry (UCNA) experiment”. In: *Physical Review C* 101 (3 Mar. 2020), pp. 035503-1–035503-7. DOI: 10.1103/PhysRevC.101.035503.

[Tay50a] G. Taylor. “The Formation of a Blast Wave by a Very Intense Explosion. I. Theoretical Discussion”. In: *Proceedings of the Royal Society of London Series A* 201 (1065 Mar. 1950), pp. 159–174. DOI: 10.1098/rspa.1950.0049.

[Tay50b] G. Taylor. “The Formation of a Blast Wave by a Very Intense Explosion. II. The Atomic Explosion of 1945”. In: *Proceedings of the Royal Society of London Series A* 201 (1065 Mar. 1950), pp. 175–186. DOI: 10.1098/rspa.1950.0050.

[TRB15] D. Tran, R. Ranganath, and D. M. Blei. “The variational Gaussian process”. In: *preprint arXiv:1511.06499* (2015).

[Vaz+11] F. Vazza et al. “A comparison of cosmological codes: properties of thermal gas and shock waves in large-scale structures”. In: *Monthly Notices of the Royal Astronomical Society* 418.2 (Dec. 2011), pp. 960–985. DOI: 10.1111/j.1365-2966.2011.19546.x.

[Vaz+16] F. Vazza et al. “Constraining the efficiency of cosmic ray acceleration by cluster shocks”. In: *Monthly Notices of the Royal Astronomical Society* 459.1 (June 2016), pp. 70–83. DOI: 10.1093/mnras/stw584.

[Wah78] G. Wahba. “Improper priors, spline smoothing and the problem of guarding against model errors in regression”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 40.3 (1978), pp. 364–372. DOI: 10.1111/j.2517-6161.1978.tb01050.x.

[WN15a] S. Walch and T. Naab. “The energy and momentum input of supernova explosions in structured and ionized molecular clouds”. In: *Monthly Notices of the Royal Astronomical Society* 451.3 (Aug. 2015), pp. 2757–2771. DOI: 10.1093/mnras/stv1155.

[Wal94] R.P. Walker. “Synchroton Radiation”. In: *CAS - CERN Accelerator School : 5th General Accelerator Physics Course* (1994), pp. 437–459. DOI: 10.5170/CERN-1994-001.437.

[Wan+19] C. Wang et al. “Design of the magnet system of the neutron decay facility PERC”. In: *EPJ Web of Conferences*. Vol. 219. EDP Sciences. 2019, pp. 04007-1–04007-6. DOI: 10.1051/epjconf/201921904007.

[Wan+22] K. Wang et al. “Interpretability in the Wild: a Circuit for Indirect Object Identification in GPT-2 small”. In: *preprint arXiv:2211.00593* (2022).

[Wan13] X. Wang. “The Free Neutron Beta-Decay: A Powerful Tool for the Investigation in Particle Physics”. Ph.D. Dissertation. Technical University Munich, 2013.

[Wee+19] R. J. van Weeren et al. “Diffuse Radio Emission from Galaxy Clusters”. In: *Space Science Reviews* 215.1 (Feb. 2019), pp. 1–75. DOI: 10.1007/s11214-019-0584-z.

[Wei58] S. Weinberg. “Charge Symmetry of Weak Interactions”. In: *Phys. Rev.* 112 (4 Nov. 1958), pp. 1375–1379. DOI: 10.1103/PhysRev.112.1375.

[Wer09] D. Werder. “Development and Characterization of a Pulsed Beam for Neutron Decay Experiments”. Diploma Thesis. Heidelberg University, 2009.

[Wil+23] J. Wildberger et al. “Adapting to noise distribution shifts in flow-based gravitational-wave inference”. In: *Physical Review D* 107.8 (Apr. 2023). DOI: 10.1103/physrevd.107.084046.

[Wil82] D. H. Wilkinson. “Analysis of neutron β-decay”. In: *Nuclear Physics A* 377 (2–3 1982), pp. 474–504. DOI: 10.1016/0375-9474(82)90051-3.

[WN15b] A. G. Wilson and H. Nickisch. “Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP)”. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. ICML’15. Lille, France: JMLR.org, 2015, pp. 1775–1784.

[Wil+16a] A. G. Wilson et al. “Deep Kernel Learning”. In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. Ed. by Arthur Gretton and Christian C. Robert. Vol. 51. Proceedings of Machine Learning Research. Cadiz, Spain: PMLR, May 2016, pp. 370–378. DOI: 10.48550/arXiv.1511.02222.

[Wil+16b] A. G. Wilson et al. “Stochastic Variational Deep Kernel Learning”. In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee et al. Vol. 29. Curran Associates, Inc., 2016. DOI: 10.48550/arXiv.1611.00336.

[Win+19] G. Winner et al. “Evolution of cosmic ray electron spectra in magnetohydrodynamical simulations”. In: *Monthly Notices of the Royal Astronomical Society* 488.2 (Sept. 2019), pp. 2235–2252. DOI: 10.1093/mnras/stz1792.

[WVB17] D. Wittor, F. Vazza, and M. Brüggen. “Testing cosmic ray acceleration with radio relics: a high-resolution study using MHD and tracers”. In: *Monthly Notices of the Royal Astronomical Society* 464.4 (Feb. 2017), pp. 4448–4462. DOI: 10.1093/mnras/stw2631.

[Wor+22] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[Wu+57] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Physical Review* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[XRE22] C. Xu, R. Roussel, and A. Edelen. “Neural Network Prior Mean for Particle Accelerator Injector Tuning”. In: *preprint arXiv:2211.09028* (2022).

Acknowledgements

First and foremost, I want to thank Anka Reuel and my parents, Bettina and Waldemar Lamparth, for their love, understanding, and endless support over these years.

I want to thank my supervisor Bastian Märkisch and mentor Torsten Soldner for their guidance, wisdom, and the opportunity to contribute to the field and grow as a researcher.

The realization of the work in this thesis was only possible thanks to the efforts and support ranging from minor to essential contributions by (in alphabetical order):

> Hartmut Abele, Matthias Antony, Karina Bernert, Mattis Bestehorn, Ludwig Böss, Andreas Doblhammer, Klaus Dolag, Tommaso Dorigo, Martin Eigner, Luca Fallböhmer, Karin Frank, Daniel Greenwald, Boris Grube, Lukas Heinrich, Rudi Hotz, Erwin Jericha, Jens Klenke, Michael Klopf, Ulrich Köster, Annabel Kropf, Manuel Lebert, Kathrin Lehmann, Martin Losekamm, Robert Neagu, Laszlo Papp, Christoph Roick, Heiko Saul, Ulrich Schmidt, Jacob Steinhardt, Ulrich Steinwandel, Giles Strong, Pietro Vischia, Lukas Werner, and Ralf Ziegler.
>
> I thank all of you!

I would also like to thank these organizations for a productive membership, valuable scientific exchange, or other support: PERKEO Collaboration, PERC Collaboration, MODE Collaboration, Origins Cluster of Excellence, and the people of the TUM Physics Department Workshop.