



---

---

# Differential spectrum modeling and sensitivity for keV sterile neutrino search at KATRIN

---

---

Zur Erlangung des akademischen Grades eines  
DOKTORS DER NATURWISSENSCHAFTEN  
(DR. RER. NAT.)

von der KIT-Fakultät für Physik des  
Karlsruher Instituts für Technologie (KIT)  
angenommene

DISSERTATION

von

M. Sc. Martin Descher  
aus Künzelsau

Tag der mündlichen Prüfung: 07. Juli 2023

Referent: Prof. Dr. Guido Drexlin

Korreferentin: Prof. Dr. Susanne Mertens



This document is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0):  
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>

---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Neutrino physics</b>	<b>7</b>
2.1	History of Neutrino Physics . . . . .	8
2.2	Neutrino Oscillation . . . . .	15
2.2.1	Discovery of Neutrino Oscillation . . . . .	15
2.2.2	Oscillation Formalism . . . . .	16
2.2.3	Oscillation Parameters . . . . .	18
2.3	Neutrinos in the Standard Model . . . . .	19
2.3.1	Standard Model . . . . .	19
2.3.2	Electroweak sector . . . . .	19
2.3.3	Lepton masses . . . . .	21
2.4	Massive Neutrinos . . . . .	22
2.4.1	Dirac and Majorana Mass Terms . . . . .	22
2.4.2	Seesaw Mechanism . . . . .	23
2.4.3	Sterile Neutrino Mass Scales . . . . .	24
2.5	Absolute Neutrino Mass Measurements . . . . .	25
2.5.1	Cosmological Constraints . . . . .	25
2.5.2	Neutrinoless Double $\beta$ -Decay . . . . .	26
2.5.3	Direct Neutrino Mass Determination . . . . .	27
2.6	Sterile Neutrino Searches . . . . .	29
2.6.1	eV-scale Sterile Neutrinos . . . . .	29
2.6.2	keV-scale Sterile Neutrinos . . . . .	32
<b>3</b>	<b>The KATRIN Experiment</b>	<b>37</b>
3.1	Tritium $\beta$ -Decay . . . . .	38
3.2	Experimental Setup . . . . .	41
3.2.1	Source Section . . . . .	42
3.2.2	Rear Section . . . . .	43
3.2.3	Transport Section . . . . .	43
3.2.4	Spectrometer Section . . . . .	44
3.2.5	Detector Section . . . . .	46
3.3	Neutrino Mass Analysis . . . . .	48
3.4	keV Sterile Neutrino Search with TRISTAN . . . . .	50

3.4.1	Signature in Tritium $\beta$ -Decay . . . . .	50
3.4.2	Deep Spectrum Measurement . . . . .	51
3.4.3	Beamline Operation Modes . . . . .	52
3.4.4	TRISTAN Detector . . . . .	53
<b>4</b>	<b>Deep Tritium Spectrum Model</b>	<b>57</b>
4.1	Goals and Requirements . . . . .	58
4.2	Overview of Systematic Effects . . . . .	62
4.3	Modeling Approach . . . . .	71
4.3.1	Response Matrix Formalism . . . . .	71
4.3.2	Code Implementation . . . . .	73
4.4	Differential Spectrum Calculation . . . . .	79
4.4.1	$\beta$ -Decay Spectrum . . . . .	81
4.4.2	Propagation Model . . . . .	82
4.4.3	Rebinning . . . . .	88
4.4.4	Detector Response . . . . .	91
4.4.5	Normalization . . . . .	94
4.4.6	Data Acquisition Model . . . . .	96
4.5	Integral Spectrum Calculation . . . . .	99
<b>5</b>	<b>Experimental effects</b>	<b>101</b>
5.1	Source Effects . . . . .	102
5.1.1	Electron Scattering in the Source . . . . .	102
5.1.2	Magnetic traps . . . . .	113
5.2	Rear Wall Effects . . . . .	117
5.2.1	Backscattering . . . . .	117
5.2.2	Tritium activity . . . . .	124
5.3	Transport Effects . . . . .	126
5.3.1	Magnetic reflection . . . . .	127
5.3.2	Magnetic collimation . . . . .	128
5.3.3	Spectrometer transmission . . . . .	131
5.3.4	Post acceleration . . . . .	133
5.4	Detector Effects . . . . .	137
5.4.1	Backscattering and back-reflection . . . . .	137
5.4.2	Energy deposition losses . . . . .	141
5.4.3	Fano Noise . . . . .	143
5.4.4	Charge sharing . . . . .	146
5.5	Readout Effects . . . . .	151
5.5.1	Pileup . . . . .	152
5.5.2	Electronic Noise . . . . .	154
5.5.3	Detection Threshold . . . . .	155
5.5.4	Dead time . . . . .	155
5.5.5	Calibration . . . . .	156

<b>6</b>	<b>Sensitivity studies</b>	<b>157</b>
6.1	Sensitivity Calculation Method . . . . .	158
6.2	Model Parameters and Settings . . . . .	161
6.2.1	Parameters and Uncertainties . . . . .	161
6.2.2	Beamline Optimization . . . . .	165
6.3	Statistical Sensitivity . . . . .	167
6.3.1	Differential Mode . . . . .	167
6.3.2	Integral Mode . . . . .	168
6.4	Systematic Uncertainty Impact . . . . .	171
6.4.1	Individual Effect impacts . . . . .	174
6.4.2	Combined Impact . . . . .	175
6.4.3	Sensitivity Time Evolution . . . . .	176
<b>7</b>	<b>Conclusion and Outlook</b>	<b>179</b>
<b>A</b>	<b>Appendix</b>	<b>181</b>
A.1	Isotropic pitch angle distribution . . . . .	181
A.2	Rate optimization for sensitivity studies . . . . .	182
A.3	Sensitivity impact of individual parameters . . . . .	183
<b>B</b>	<b>Bibliography</b>	<b>187</b>



---

## Introduction

---

Neutrinos have been intriguing elementary particles ever since their postulation in 1930 [1] due to their peculiar properties. Three types (flavors) of neutrinos are known today, and they are the only known matter particles without electric charge and thus practically only interact via the weak force. This makes their interactions extremely rare, so they are notoriously challenging to detect through direct interactions. They are also the lightest of all matter particles and have been considered completely devoid of mass until the discovery of neutrino flavor oscillations around the year 2000 [2, 3]. This observation proved definitively that neutrino masses must be non-zero. While oscillation experiments have since determined the two mass splittings, the absolute mass scale is inaccessible by such experiments and remains an open question to this day. A successful determination of this scale would be a major achievement due to its consequences for particle physics and cosmology.

The Karlsruhe Tritium Neutrino experiment (KATRIN) [4, 5] aims to determine the absolute neutrino mass scale using precision tritium  $\beta$ -spectroscopy. This method was first suggested in Enrico Fermi's classic publication on the theory of weak decay from 1934 [6]. Even almost a century later, it can be regarded as the most robust method for investigating the neutrino mass, as it is a model-independent technique purely based on kinematics. KATRIN performs a precise measurement of the integral tritium spectrum in a narrow region around the kinematic endpoint energy  $E_0 \approx 18.6$  keV, where the imprint of the neutrino mass would be visible. With its ultra-luminous gaseous molecular tritium source and the largest, most precise MAC-E filter spectrometer ever built, the experiment can reach a sensitivity of  $m_{\bar{\nu}_e} < 0.2$  eV (90% CL) after five years of data-taking [4]. Already, with data from the first two physics runs, KATRIN sets the world-leading limit of  $m_{\bar{\nu}_e} < 0.8$  eV (90% CL) [7].

In the Standard Model of Particle Physics (SM), neutrinos are still assumed to be massless since the particular mechanism for generating the appropriate mass terms is unknown. One natural and theoretically well-motivated way of obtaining those mass terms is the extension of the SM with right-handed neutrinos. This type of neutrino would react even rarer in weak interactions and is thus called *sterile*, while the known flavors are referred to as *active*. The existence of sterile neutrinos could resolve several open questions concerning the SM neutrino sector and cosmology, such as the smallness

of the neutrino masses, the dominance of matter over anti-matter in the universe, or the question of whether neutrinos are their own antiparticles. Compared to active neutrinos, sterile neutrinos could be considerably more massive, and with a mass on the keV scale, they turn out to be excellent candidate particles for Dark Matter.

Such sterile neutrinos would also manifest in the  $\beta$ -decay spectrum with a tiny contribution whose size is determined by the mixing amplitude  $|U_{e4}|^2 = \sin^2 \theta$ . Therefore, after the neutrino mass measurement concludes, KATRIN will perform a high-statistics measurement of the differential tritium spectrum to energies deep below the endpoint. This enables the search for keV sterile neutrinos with masses  $m_s \lesssim 18$  keV at a sensitivity of  $\sin^2 \theta \sim 10^{-6}$ . This exceeds current laboratory limits by roughly two orders of magnitude. A novel detector system, the TRISTAN silicon drift detector [8], was developed specifically for this endeavor. It features improved high-rate capabilities and better energy resolution than the present KATRIN Detector. This allows for rapidly acquiring the necessary statistics and better identifying the sterile neutrino signature.

The new mode of operation comes with major challenges, one of which is the development of a model capable of precisely predicting the differential spectrum measured by the TRISTAN detector at the end of the KATRIN beamline. Such a model is first required to determine the impact of systematic uncertainties on the sensitivity. During the planning phase for the measurement, it becomes an indispensable tool for gauging the importance of particular systematic effects. It can thus be used for exploring optimized beamline settings and hardware modifications that significantly mitigate the influence of systematics. This thesis presents the first differential KATRIN model with the ability to combine all dominant systematics holistically. Furthermore, the model is employed to make a sensitivity estimate for Phase-1 of the measurement with TRISTAN, where all currently available systematics are included.

This work is structured into seven chapters, including this introduction. Chapter 2 summarizes relevant aspects concerning the history and current state of neutrino physics. At the end of the chapter, the focus is on sterile neutrinos and experimental searches. Chapter 3 introduces the KATRIN experiment in terms of its working principles, hardware components, analysis techniques, and the TRISTAN upgrade. The main part begins in chapter 4, where the differential model is explained in an overarching manner. Chapter 5 then continues, going into details concerning each systematic effect currently implemented. In chapter 6, the model is employed to estimate the sensitivity of TRISTAN Phase-1 with nine detector modules integrated into the KATRIN beamline. Finally, chapter 7 summarizes this work's key results and ends with an outlook concerning future developments.

## Neutrino physics

---

Of all known matter constituents, neutrinos are the universe's most abundant massive elementary particles. Yet, they barely contribute to the overall energy density of the universe as they are incredibly light. They are produced, for instance, in the cores of stars by the proton-proton fusion chain (solar neutrinos) or as relics of the Big Bang (cosmic neutrino background), affecting the structure formation in the early universe. Hence, they play a major role in cosmology and astroparticle physics. Here on Earth, they were studied first in weak radioactive decay, also known as  $\beta$ -decay. [9]

In the Standard Model (SM) of particle physics, neutrinos are the only fermions without electric charge and therefore do not take part in electromagnetic interactions that govern most of what humans perceive. Neither do they participate in interactions of the strong force, for they exhibit no color charge. Only via interactions of the weak force do they manifest. As a result of their subtle interaction strength, they mostly pass through matter unhindered and are notoriously difficult to detect directly. Still, the effort of studying neutrinos is worth the challenges since they offer many gateways to explore physics beyond the Standard Model via extensions of the neutrino sector, including Majorana mass terms and sterile neutrinos. [10]

This chapter summarizes several aspects of neutrino physics relevant to the contents of this work. First, an outline of the history with some important landmark discoveries about neutrinos is provided to set the stage for a brief description of the neutrino sector in the SM. Afterward, the groundbreaking discovery of neutrino flavor oscillation is discussed. This observation proves neutrinos have non-zero mass, which is incompatible with the SM, where neutrinos are assumed to be devoid of mass. The subsequent section explains the general method of extending the SM Lagrangian to include Dirac and Majorana mass terms for neutrinos. These imply fascinating new prospects, such as neutrinos acting as their own antiparticles or the existence of right-handed neutrinos. Lastly, it is discussed how the mass splittings between neutrino generations and the absolute scale can be determined experimentally.

## 2.1 History of Neutrino Physics

The history of neutrino physics begins with the first observation of the continuous energy spectrum of electrons emitted in radioactive  $\beta$ -decay by J. Chadwick in 1914 [11]. Chadwick's result was surprising at the time since energy spectra from  $\alpha$ - and  $\gamma$ -decay were both monoenergetic. And since seemingly just an electron was emitted from the decaying nucleus in  $\beta$ -decay, energy-momentum conservation would dictate that the electron always receives the same share of the fixed decay energy, leading to the expectation of a monoenergetic spectrum as well (see figure 2.1).

### Postulation by W. Pauli

Initially, it was disputed whether the continuous energy spectrum was, in fact, due to the decay itself or rather monoenergetic at first and broadened by secondary effects. Several years later, calorimetric measurements by J. Ellis et al. in 1927 [12] and L. Meitner et al. in 1930 [13] were able to rule out the second possibility, confronting the physics community with the possibility that energy conservation may be violated in individual decay processes [10]. However, in his famous letter from 1930 [1], W. Pauli proposed “a desperate way out” of this conundrum:

If an electrically neutral particle - the neutrino<sup>1</sup> - is created in addition to the electron,  $\beta$ -decay turns into a three-body process where the energy is randomly distributed among the constituents, resulting in a continuous energy spectrum for the electron. Due to the lack of charge, the neutrino would not manifest in the calorimetric measurements as it will likely escape, carrying away the missing energy. Furthermore, this idea solves another problem noticed at the time. In certain decays, the conservation of angular momentum seemed to be violated since the emitted electron carries away spin  $1/2$ , while the nuclear spin only changes in whole-numbered increments. This is resolved if the neutrino also carries spin  $1/2$ . [10]

### Fermi's theory of $\beta$ -decay

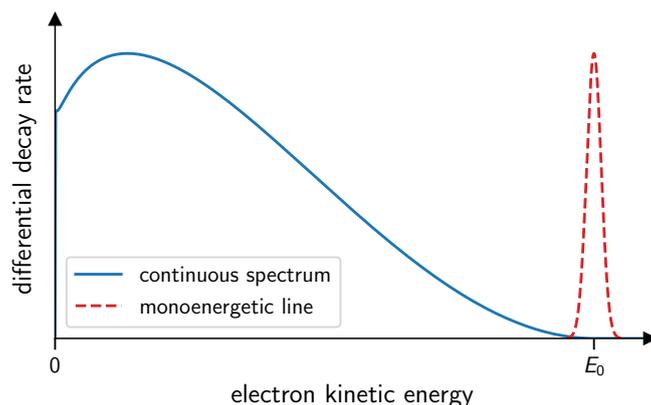
Based on Pauli's suggestion and the rapidly evolving understanding of atomic nuclei, E. Fermi formulated his classic theory of  $\beta$ -decay in 1934 [6]. Fermi's theory regarded neutrons and protons as two states of the same heavy nucleon. In the conversion from one to another, electrons and neutrinos (or their antiparticles) are created in a charge-conserving manner. Fermi's theory was very successful and correctly predicted the shape of the  $\beta$ -spectrum, including the impact of a potential neutrino mass. It is still widely used in nuclear physics, also serving as the basis for this work's  $\beta$ -spectrum calculations.

In today's terms, the process of  $\beta^-$ -decay can be summarized as follows: a nucleus  $N$ , unstable due to a surplus of neutrons, emits an electron and an electron antineutrino<sup>2</sup>,

---

<sup>1</sup>In his letter Pauli names this particle *neutron*. But by the time Pauli's idea gained popularity, the name was preempted by the discovery of the neutron - as it is known today - by J. Chadwick in 1932 [14]. E. Fermi, therefore, later renamed Pauli's particle to *neutrino* - the little neutral one. [15]

<sup>2</sup>The  $\beta$ -decay neutrino was later declared the *antineutrino* based on the concept of lepton number conservation and the prefix “electron” stems from the discovery of further lepton generations. [15]



**Figure 2.1:** The  $\beta^-$ -decay spectrum of the electron kinetic energy has a continuous shape ranging from zero to the kinematic endpoint energy  $E_0$ . If  $\beta$ -decay were a two-body process, a monoenergetic line spectrum at  $E_0$  would be expected.

while the atomic number  $Z$  increases by one:<sup>3</sup>



The decay energy  $Q$  is the mass difference between initial and final state particles and is distributed among the reaction products as kinetic energy. On the nucleon level, the decay is the conversion of a neutron to a proton

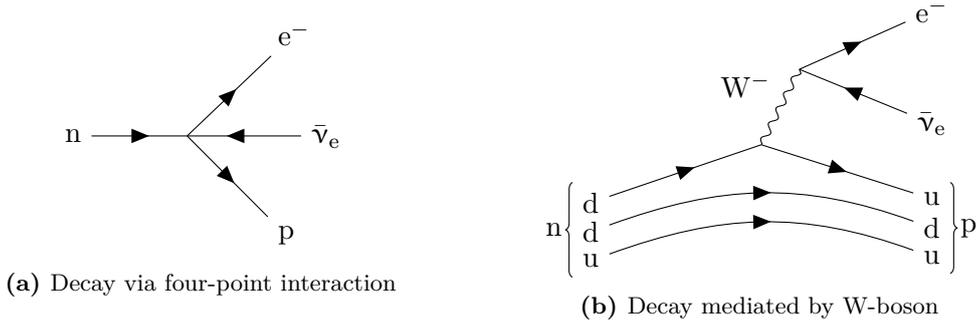


where the released energy stems from the difference in the binding energy of the mother and daughter nucleon configurations and the difference in neutron-proton mass. The decay of a free neutron is also possible since it is more massive than the sum of the proton, electron, and neutrino mass.<sup>4</sup> The reverse process,  $\beta^+$ -decay or *positron emission*, where a proton changes into a neutron, is only possible for bound protons where the mass deficit is accounted for by binding energy.

When, by chance, the neutrino is created nearly at rest, the electron receives the highest possible kinetic energy. This is the kinematic endpoint energy  $E_0 = Q - E_{\text{rec}}$ , where  $E_{\text{rec}}$  is the recoil energy the daughter nucleus receives. As already noted by Fermi in his original publication, the influence of a potential neutrino mass on the spectral shape is most prominent just below the endpoint. Given the experimental findings of his time, he concluded, “the rest mass of the neutrino is either zero, or, in any case, very small in comparison to the mass of the electron.”[16]. Frustratingly, no experiment observed the signature of neutrino mass in  $\beta$ -decay or similar processes to this day [17].

<sup>3</sup>In  $\beta^+$ -decay the electron and antineutrino are exchanged by their antiparticles, and the atomic number decreases by one.

<sup>4</sup>Interestingly, the free neutron lifetime of  $\approx 15$  minutes is abnormally large due to the comparatively small mass difference to the final state, leaving only little available phase space. [15]



**Figure 2.2:** First order Feynman graphs for neutron decay using Fermi’s four-point interaction (a), and the modern interpretation (b) where the decay is identified as a quark-transition under emission of a virtual W-boson that creates a lepton pair. The left picture can be recognized as an approximation of the right for energies much smaller than the W-boson mass [10].

On a more theoretical level, Fermi described the transition from initial to final state via a four-point interaction, i. e. a direct coupling of the spinor fields representing the free ingoing and outgoing particles. While this worked for the energy scale of  $\beta$ -decay (keV up to a few MeV), the theory was expected to fail at higher momentum transfers [15]. With today’s knowledge of the composition of nucleons from quarks, the decay is seen as a transition of a down-quark to an up-quark under the emission of a virtual W-boson that mediates the coupling between the fermions and creates the electron-neutrino pair (see figure 2.2). Fermi’s theory is regarded as the low-energy limit of the modern Glashow-Weinberg-Salam model, which describes the electroweak sector of the SM (more on this in section 2.3). [10]

### Discovery of the neutrino

In the years following its postulation, further indirect evidence for the existence of the neutrino was found in decay processes. For instance, Rodeback and Allen found the first evidence for neutrinos outside  $\beta$ -decay in 1952 [18] via recoil measurements in electron capture of  $^{37}\text{Ar}$ :



By measuring the recoil of the daughter nucleus with a time-of-flight method, they showed that the process is compatible with a two-body decay where a single neutrino is emitted.

Again, this result relied on the assumption of energy-momentum conservation. Going forward, only catching neutrinos in direct interactions would be considered the smoking gun of their existence. But unlike any other particle, neutrinos interact only via the weak interaction<sup>5</sup>, making this task extremely challenging.

In the mid-1950s, the direct discovery of antineutrinos was finally achieved by Reines

<sup>5</sup>Due to their nonzero mass, they could also interact through gravity. But since gravity is incredibly weak compared to the other forces, its effect is neglectable in ordinary particle physics [15, 19].

and Cowan et al. [20, 21] through observation of the inverse  $\beta$ -decay reaction:

$$\bar{\nu}_e + p \rightarrow n + e^+ . \quad (2.4)$$

In their decisive experiment from 1956, they used two key strategies: Firstly, they recognized that they needed a large flux of antineutrinos to offset the rarity of their interaction. This was achieved by positioning a large detector near the Savannah River nuclear reactor, which emits a constant stream of antineutrinos from  $\beta^-$ -decay of the neutron-rich fission products [10]. The antineutrino flux can be estimated from the generated reactor power, enabling a measurement of the inverse  $\beta$ -decay cross-section. Several water tanks served as targets<sup>6</sup>, interspersed with layers of liquid scintillator. Secondly, they chose to detect both the positron and the neutron signatures. Positrons almost immediately annihilate with electrons in their vicinity, emitting two back-to-back photons with 511 keV energy each. This caused a coincidence signal in two neighboring scintillators. For the neutron detection, they added Cadmium salt to the solution to capture the neutrons after their moderation to lower energy through multiple scattering on the hydrogen atoms of water. The capture process released  $\gamma$ -rays which were also detected in the scintillator as a delayed coincidence to the positron's signal. This approach allowed for powerful background discrimination by applying a sophisticated trigger that incorporated the time structure and allowed energy estimates from signal pulse heights.

### Neutrinos and their antiparticles

Another fundamental question raised in the 1950s concerned the neutrino-antineutrino relationship. Dirac's quantum theory of the electron from 1928 [22] showed that all spin- $1/2$  particles have a corresponding antiparticle of the same mass and opposite charge.<sup>7</sup> But both may be identical for uncharged particles, as is the case for the photon. If this were true for neutrinos as well, they would be classified as the only Majorana fermions<sup>8</sup> of the known elementary particles.

To investigate if  $\nu$  and  $\bar{\nu}$  are distinct particles, one can look at the crossed<sup>9</sup> reaction to (2.4), as well as one where the neutrino is exchanged by its antiparticle:

$$\nu_e + p \rightarrow n + e^- \quad (2.5)$$

$$\bar{\nu}_e + p \rightarrow n + e^- . \quad (2.6)$$

If  $\nu$  and  $\bar{\nu}$  were identical, both reactions should occur with the same cross-section [10]. The cross-section of the first reaction is implied through crossing symmetry from the measurement by Cowan and Reines. In 1953, R. Davis Jr. searched for the second reaction using reactor antineutrinos [28]. He found that the reaction does not occur and concluded that neutrino and antineutrino must be distinct particles.

<sup>6</sup>Water is used since it contains many isolated target protons.

<sup>7</sup>Dirac's theory led to his postulation of the positron in 1931 [23] which was discovered by C. D. Anderson in 1932 [24].

<sup>8</sup>Named after E. Majorana, who first hypothesized the existence of such particles in 1937 [25, 26].

<sup>9</sup>Crossing symmetry, first introduced by Gell-Mann and Goldberger in 1954 [27], is the concept of moving particles to the other side of the reaction and exchanging them by their antiparticle (see [15]).

Also around that time, the concept of lepton number conservation emerged<sup>10</sup>: Leptons ( $e^-$ ,  $\nu$ , and heavier generations) are assigned a value of  $L = +1$  and their antiparticles  $L = -1$ , and it is required that a reaction does not change the sum of the lepton numbers. Reaction (2.5) is therefore allowed, while reaction (2.6) is forbidden. So it would seem that the decisive difference between  $\nu$  and  $\bar{\nu}$  is their lepton number. However, once parity violation was discovered in weak interactions, it was realized that the difference also lies in their spin state. So, neutrinos may be Majorana particles after all, with the only distinction between neutrinos and antineutrinos being their helicity. [15]

### Parity violation and neutrino helicity

Until the 1950s, physicists had assumed that all physics processes would occur identically when mirrored. In other words, particle interactions should be invariant under a parity transformation of the spatial dimensions:

$$P : \vec{x} \rightarrow -\vec{x} . \quad (2.7)$$

This implies that the parity of an initial state of a reaction, which is the product of the incident particles' intrinsic parities<sup>11</sup>, equals that of the final state [9]. In 1956, T.D. Lee and C. N. Yang researched the validity of this assumption. They found evidence that this was true for the strong and electromagnetic interactions [15], while realizing that parity symmetry had never been tested in weak interactions. Inspired by the " $\theta-\tau$ " puzzle<sup>12</sup>, they raised the question of parity conservation in weak interactions and suggested possibilities for experimental tests [30].

Parity conservation can be tested by observing pseudoscalar quantities such as helicity, which is defined by the normalized projection of spin and momentum of a particle [10]:

$$\mathcal{H} = \frac{\vec{\sigma} \cdot \vec{p}}{|\vec{\sigma}||\vec{p}|} . \quad (2.8)$$

The eigenvalues of the helicity operator are  $+1$  for *right-handed* particles (spin aligned with momentum) and *left-handed* (spin opposed to momentum). The parity transformation turns right-handed into left-handed and vice versa.

Early in 1957, C. S. Wu et al. performed the first experimental test of parity conservation in  $\beta^-$ -decay of polarized  $^{60}\text{Co}$  and discovered that parity is maximally violated [31]. In their experiment, the direction of emitted electrons was measured and compared to the nuclear spin, which was aligned with a strong magnetic field at cryogenic temperature. The mirror configuration was achieved by flipping the direction of the magnetic field. Shortly after, a similar result was observed via measurement of the neutrino helicity in positive pion decays by Garwin et al. in 1957 [32], and in  $^{152m}\text{Eu}$  electron capture by Goldhaber et al. in 1958 [33].

---

<sup>10</sup>Konopinski and Mahmoud first introduced the idea for lepton number conservation in 1953 [29].

<sup>11</sup>Intrinsic parity is the eigenvalue ( $\pm 1$ ) of a particle's field under parity transformation (see [19]).

<sup>12</sup>The puzzle was about two mesons, known as  $\theta^+$  and  $\tau^+$ , that decayed into opposite states of parity but were otherwise identical. Lee and Yang proposed that they are the same particle (the  $K^+$  meson) and that one of the decays violates parity. [15]

Considering that neutrinos are solely produced in weak interaction, all these experiments concluded that neutrinos are always left-handed and anti-neutrinos are always right-handed.<sup>13</sup> And if (massless) right-handed neutrinos did exist, they could never interact with other SM particles. These so-called *sterile* neutrinos are thus not included in the current description of the electroweak sector of the SM.

### Theory of the weak interaction

The discovery of parity violation led to the development of the famous V-A theory (vector minus axial vector) of weak interaction in 1958 by several parties [34, 35]. Its predecessor, the Fermi interaction, is based on a pure vector current coupling of the fermion fields, conserving parity. The V-A theory suggests that the observed maximal parity violation can be achieved by mixing current terms of opposing parity, namely taking the difference of a vector current with positive parity and an axial vector current with negative parity (see [10] and section 2.3 for more explanation).

Another problem of Fermi’s theory was its direct coupling of the fermion fields through a point interaction instead of transmitting the force through a mediating boson, leading to the failure of the theory at high energy. Hence, an intermediate vector boson was required, but its properties were still unclear for several years. [15]

The V-A idea was further developed by S. L. Glashow [36], A. Salam [37], and S. Weinberg [38] into a theory, now known as the GWS model, that elegantly combines the weak and the electromagnetic forces into the unified electroweak interaction. The theory was a major success and culminated in the prediction of the W- and Z-bosons that transmit charged current and neutral current weak interactions, respectively. Neutral current interactions were first observed in the Gargamelle Neutrino Experiment at CERN in 1973 [39], and the vector bosons themselves were discovered at the UA1 and UA2 experiments at CERN’s Super Proton Synchrotron in 1983.

### Neutrino flavors

Apart from nuclear  $\beta$ -decay, neutrinos are also produced in other weak processes such as pion decay:  $\pi^+ \rightarrow \mu^+ + \nu$ . In the 1960s, it was still unclear whether the neutrino produced in association with muons was the same as the one from  $\beta$ -decay. One hint towards them being different was the non-observation of the process  $\mu \rightarrow e + \gamma$ , while the decay  $\mu \rightarrow e + \bar{\nu} + \nu$  is possible. This problem could be solved by introducing a conserved quantum number for “ $\mu$ -ness” that distinguishes two flavors of neutrinos: Those produced in association with electrons ( $\nu_e$ ) and those with muons ( $\nu_\mu$ ). [15]

In 1959, B. Pontecorvo suggested discerning the difference with reactions induced by a pure beam of muon neutrinos and searching for the creation of electrons, which would falsify the idea of two flavors [40]. In 1960, M. Schwartz discussed how such beams could be created using accelerators [41]. The distinction between  $\nu_e$  and  $\nu_\mu$  was soon after

<sup>13</sup>Strictly speaking, this only applies to massless neutrinos. Using a Lorentz boost, there is always a reference frame with reversed helicity for massive particles.

demonstrated by L. M. Lederman, M. Schwartz, and J. Steinberger in an experiment at the Brookhaven AGS in 1962 [42].

The discovery of the  $\tau$ -lepton in 1975 [43] suggested the existence of a third neutrino flavor, the tau neutrino  $\nu_\tau$ . In 1989, the ALEPH experiment at CERN's Large Electron Positron Collider provided indirect evidence for its existence by measuring the  $Z$ -boson decay width [44], which showed that there must be three light neutrino flavors ( $m_\nu < m_Z/2$ ) that couple to the  $Z$ -boson. The direct discovery of the tau neutrino was achieved in 2000 by the DONUT collaboration [45].

### Solar neutrino problem

The sun emits a large flux of neutrinos from fusion processes in its core, mainly stemming from the proton-proton fusion chain (p-p chain) [46].<sup>14</sup> This chain consists of 11 sub-processes which result in the overall reaction equation



One neutrino is produced per  $\approx 13 \text{ MeV}$  energy output. Given the solar constant  $S = 8.5 \times 10^{11} \text{ MeV cm}^{-2} \text{ s}^{-1}$  this results in an estimated solar neutrino flux of  $\Phi_\nu = 6.5 \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$  on Earth's surface. The sub-reactions determine the energy spectrum, which is thus a collection of continuous spectra and monoenergetic lines up to 18.77 MeV. In measurements, only a fraction of the flux is accessible depending on the energy threshold. Most measurable neutrinos stem from  $\beta^+$ -decay of  ${}^8\text{B}$ . [10]

R. Davis Jr. achieved the first measurement of solar neutrinos in 1968 [49] with a radiochemical technique. A tank of 615 t of tetrachloroethylene ( $\text{C}_2\text{Cl}_4$ ) was positioned deep underground in the South Dakota Homestake gold mine to avoid cosmic rays. Through the reaction



radioactive Argon was created and retrieved from the tank after several months. By counting the  ${}^{37}\text{Ar}$  decays ( $\beta^-$ ,  $T_{1/2} \approx 35 \text{ d}$  [50]), the neutrino flux was estimated.

With only about one-third of the expected value, the measured flux revealed a significant deficit [51]. This caused much discussion about the validity of the measurement and the solar standard model [15]. But gradually, the *solar neutrino problem* was consolidated when the existence of a deficit was confirmed by real-time experiments using Cherenkov-light (e. g. Kamiokande-II [52]) and other radiochemical experiments (e. g. GALLEX [53] and SAGE [54]).

Eventually, the discovery of neutrino flavor oscillation resolved the solar neutrino problem (see next section 2.2). In 2002, R. Davis Jr. and M. Koshiba, the designer of the Kamiokande experiment, received the physics Nobel prize “for pioneering contributions to astrophysics, in particular for the detection of cosmic neutrinos” [55].

---

<sup>14</sup>The solar standard model [47] also includes the so-called CNO fusion cycle [48]. But since the sun mostly consists of hydrogen, the p-p chain accounts for 98.4% of the solar energy output. [10]

## 2.2 Neutrino Oscillation

### 2.2.1 Discovery of Neutrino Oscillation

In 1968, B. Pontecorvo suggested *neutrino oscillations* to solve the solar neutrino problem [56]. The general idea is that solar neutrinos are produced with electron flavor, but similar to the mixing in other systems of neutral particles<sup>15</sup>, neutrinos can change into a mixture of different types of flavors during propagation through space. This is possible if there is a mismatch between the eigenstates of weak interaction (flavor states) and the eigenstates of propagation (mass states).<sup>16</sup> Hence, flavor oscillations directly imply non-vanishing neutrino masses [10]. The term “oscillation” stems from the flavor transition probability being a sinusoidal function of time (or traveled distance), as demonstrated in section 2.2.

On Earth, the admixture of the electron flavor in solar neutrinos is reduced through this oscillation, decreasing the likelihood of electronic interaction. Since Davis’ detection method is only sensitive to the electron flavor, part of the solar neutrino flux is inaccessible. This causes the observed deficit, known as electron neutrino *disappearance*.

The first solid evidence for neutrino oscillation from a measurement of atmospheric muon neutrinos was reported by the Super-Kamiokande Collaboration (Super-K) in 1998 [2]. Super-K is a water-based Cherenkov radiation detector sensitive to charged particles created by incoming neutrinos of all known flavors through elastic neutrino-electron scattering [10]. For their discovery, they measured a deficit in events from up-going atmospheric muon neutrinos that passed through Earth, which is compatible with two-flavor  $\nu_\mu \leftrightarrow \nu_\tau$  oscillation.

Evidence for neutrino oscillation as the solution to the solar neutrino problem was first given by the Sudbury Neutrino Observatory Collaboration (SNO) in 2001 [3] using a similar Cherenkov radiation detector. A crucial difference to Super-K was the use of heavy water ( $D_2O$ ) to provide target neutrons, which enabled additional neutrino reactions: Inverse  $\beta$ -decay from charged current interactions exclusive to  $\nu_e$ , and neutron disintegration from neutral current interactions of all neutrino flavors [58]. This allowed for the separate determination of the total  $\nu$  flux and the  $\nu_e$  flux.

The measured total neutrino flux was consistent with the expectation from the solar standard model, and the flavors  $\nu_\mu$  and  $\nu_\tau$ , which cannot be produced in the sun, accounted for the majority of the flux. Thus, neutrino oscillation was shown as the origin of the solar neutrino deficit.<sup>17</sup> In 2015, T. Kajita and A. McDonald, the directors of Super-K and SNO, were jointly awarded the Physics Nobel prize “for the discovery of neutrino oscillations, which shows that neutrinos have mass” [61].

<sup>15</sup>Inspired by the particle-antiparticle oscillation in the neutral kaon system, Pontecorvo first explored the possibility of neutrino oscillation already in 1957 [57].

<sup>16</sup>The concept is the same as in quark mixing where the mass eigenstates also differ from the eigenstates of the weak interaction. Here, the sets of eigenstates are connected by the CKM matrix. [10]

<sup>17</sup>For the correct prediction of the solar neutrino flavor composition, the modification of neutrino oscillation in matter (MSW-Effect [59, 60]) must be considered. [10]

### 2.2.2 Oscillation Formalism

This section introduces the mathematical formalism of neutrino oscillation due to the mixing between flavor and mass eigenstates and the calculation of flavor transition probabilities. The topic is well covered in textbooks and other literature (see [10, 19, 62] or [63] for a more detailed explanation), so here only a brief summary is given. While the neutrino oscillation phenomenon is crucial for exploring neutrino masses, flavor mixing is also vital for explaining the signature of potential sterile neutrinos in  $\beta$ -decay, which will be important in section 3.4.1.

#### Mixing Matrix

For massive neutrinos, their eigenstates of propagation (mass states)  $|\nu_i\rangle$  ( $i = 1, 2, 3$ ) are not necessarily identical to their eigenstates of weak interaction (flavor states)  $|\nu_\alpha\rangle$  ( $\alpha = e, \nu, \tau$ ). Both sets of states form orthogonal bases, where each can be expressed in terms of the other via superposition: [10]

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i} |\nu_i\rangle \quad \text{and} \quad |\nu_i\rangle = \sum_\alpha U_{\alpha i}^* |\nu_\alpha\rangle . \quad (2.11)$$

The coefficients form the unitary Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix

$$U = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} . \quad (2.12)$$

The matrix is named after Z. Maki, M. Nakagawa, and S. Sakata, who developed the quantitative description of neutrino oscillation in 1962 [64], and B. Pontecorvo, who had suggested the idea and elaborated on experimental implications in 1967 [65].

The mass states are ordered by their contribution to the electron flavor  $|U_{e1}|^2 > |U_{e2}|^2 > |U_{e3}|^2$  [66], but it is still unclear whether this also corresponds to the ordering of masses (see normal vs. inverted hierarchy in section 2.2.3).

This  $3 \times 3$  unitary matrix has at most  $3^2 = 9$  free parameters, where 3 are interpreted as rotation angles or *mixing angles*  $\theta_{ij} \in [0, \pi/2]$  [62]. The remaining degrees of freedom are complex phases, 3 of which can be absorbed into the charged lepton mass eigenstates [62]. The last phases depend on the neutrino's Dirac or Majorana nature (see section 2.4). The pure Dirac case allows for one complex phase  $\delta_{\text{CP}}$  that causes CP violation in neutrino oscillation appearance [66]. In the Majorana case, two phases  $\eta_1$  and  $\eta_2$  are added, which may have an observable effect in neutrinoless double  $\beta$ -decay [66]. Altogether, the PMNS matrix can be written as [62]

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \cdot \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{\text{CP}}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{\text{CP}}} & 0 & c_{13} \end{pmatrix} \cdot \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} e^{i\eta_1} & 0 & 0 \\ 0 & e^{i\eta_2} & 0 \\ 0 & 0 & 1 \end{pmatrix} , \quad (2.13)$$

where  $s_{ij} = \sin \theta_{ij}$  and  $c_{ij} = \cos \theta_{ij}$ .

### Flavor Oscillation

Since neutrinos only interact in the weak interaction, they are always produced and detected in one of the weak eigenstates  $|\nu_\alpha\rangle$ . The initial state can be evolved over time by switching to the mass basis and using the Schrödinger equation to evolve the superposition of mass states. Employing a plane wave ansatz, this yields [10]

$$|\nu_\alpha(t)\rangle = \sum_i U_{\alpha i} e^{-iE_i t} |\nu_i\rangle, \quad (2.14)$$

where  $E_i = \sqrt{p_i^2 + m_i^2}$ . The probability for a neutrino to assume the flavor  $\beta$  during detection, i. e. the flavor transition probability, can be calculated via [10]

$$P_{\alpha\beta} = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2 = \sum_{ij} U_{\alpha i} U_{\alpha j}^* U_{\beta i}^* U_{\beta j} e^{-i(E_i - E_j)t}. \quad (2.15)$$

Using the relativistic approximation

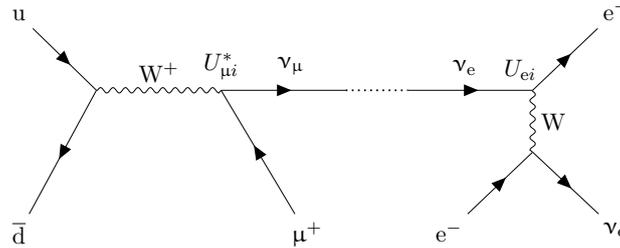
$$E_i = \sqrt{p_i^2 + m_i^2} \simeq p + \frac{m_i^2}{2E} \quad (2.16)$$

and the ‘‘same energy’’ and ‘‘same momentum’’ assumption  $p_i \simeq p_j \equiv p \simeq E$  (this also indicates  $L \simeq t$  with  $c = 1$ ) [63], the result can be written as [62]

$$P_{\alpha\beta}(L, E) = \delta_{\alpha\beta} - 4 \sum_{i < j} \text{Re}[U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}] \sin^2 \left( \frac{\Delta m_{ij}^2 L}{4E} \right) \quad (2.17)$$

$$+ 2 \sum_{i < j} \text{Im}[U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}] \sin \left( \frac{\Delta m_{ij}^2 L}{2E} \right) \quad (2.18)$$

with the mass splitting  $\Delta m_{ij} = m_i^2 - m_j^2$ . The first sum (2.17) is CP conserving, while the second sum (2.18) violates CP since it changes sign under  $\nu \rightarrow \bar{\nu}$ , which conjugates the matrix elements [62]. In either case, the transition probability is sinusoidal as a function of traveled distance  $L$ . The oscillation length depends on the mass splittings, while the PMNS matrix elements determine the amplitude.



**Figure 2.3:** Illustrative neutrino oscillation Feynman graph for  $\nu_e$  appearance: A muon neutrino is produced in  $\pi^+$ -decay, e. g. at an accelerator. After propagating some distance, it may interact with electronic flavor, for example, in charged current electron scattering.

### 2.2.3 Oscillation Parameters

For three neutrino mass states, oscillation phenomena are characterized by eight parameters: Two mass splittings ( $\Delta m_{21}^2$ ,  $\Delta m_{32}^2$ ), three mixing angles ( $\theta_{12}$ ,  $\theta_{23}$ ,  $\theta_{13}$ ), one CP-violating Dirac phase ( $\delta_{\text{CP}}$ ), and two Majorana phases ( $\eta_1$ ,  $\eta_2$ ). The latter are accessible through neutrinoless double  $\beta$ -decay ( $0\nu\beta\beta$ ) which is yet unobserved.

All other parameters can be determined via neutrino oscillation experiments that measure the *appearance* or *disappearance* of flavors in neutrino fluxes or beams with a known flavor composition. The primary sources are solar, atmospheric, reactor, and accelerator neutrinos [10]. Some important measurements are referenced here, and for a full listing see [62, pp. 1288]. Historically,  $\theta_{12}$  and  $\Delta m_{21}^2$  are connected to solar neutrinos. Precise measurements of these two parameters are obtained from a combination of solar neutrinos (Super-K, SNO) and reactor neutrinos (KamLAND) [67, 68].  $\Delta m_{32}^2$  and  $\theta_{23}$  are primarily measured using atmospheric (IceCube, Super-K) [69, 70] and accelerator neutrinos (T2K, NOvA, MINOS) [71–73], where the value of the mass splitting is supplemented by reactor neutrino measurements (Daya Bay, RENO) [74, 75].  $\theta_{13}$  is the smallest mixing angle of the bunch and has been determined with reactor neutrinos (Daya Bay, RENO, Double Chooz) [74–76]. Current values of these parameters, as calculated in a global analysis by the NuFIT group [77, 78], are presented in table 2.1.

Despite the multitude of measurements, two oscillation properties remain unclear: Firstly, the sign of  $\Delta m_{32}^2$  is still ambiguous, allowing for two potential ordering schemes for the mass states:

- Normal ordering:  $m_1 < m_2 \ll m_3$
- Inverted ordering:  $m_3 \ll m_1 < m_2$

According to the global analysis from [77, 78], normal ordering is slightly preferred at the  $1.5 - 2.5\sigma$  level depending on the data selection. Secondly, the CP violating phase  $\delta_{\text{CP}}$  is not precisely known and heavily depends on the mass order [79].

Since neutrino oscillation only depends on the squared difference of the mass values, the absolute mass scale is inaccessible by such measurements. Experimental methods of assessing the absolute mass scale are summarized in section 2.5.

**Table 2.1:** Oscillation parameters including  $\pm 1\sigma$  intervals from a global analysis by [77, 78].

Parameter	Normal ordering	Inverted ordering
$\sin^2 \theta_{12}$	$0.303_{-0.012}^{+0.012}$	$0.303_{-0.011}^{+0.012}$
$\sin^2 \theta_{23}$	$0.451_{-0.016}^{+0.019}$	$0.569_{-0.021}^{+0.016}$
$\sin^2 \theta_{13}$	$0.02225_{-0.00059}^{+0.00056}$	$0.02223_{-0.00058}^{+0.00058}$
$\Delta m_{21}^2/10^{-5} \text{ eV}^2$	$7.41_{-0.20}^{+0.21}$	$7.41_{-0.20}^{+0.21}$
$\Delta m_{32}^2/10^{-3} \text{ eV}^2$	$2.507_{-0.027}^{+0.026}$	$2.486_{-0.028}^{+0.025}$
$\delta_{\text{CP}}$	$232_{-26}^{+36}$	$276_{-29}^{+22}$

## 2.3 Neutrinos in the Standard Model

The Standard Model of particle physics (SM) describes the three generations of fundamental matter particles (quarks, charged leptons, and neutrinos) and their interactions, which are mediated by force carrier particles, the gauge bosons. Of the four fundamental interactions, the SM includes:

- the strong force mediated by gluons,
- the electromagnetic force mediated by the photon,
- and the weak force mediated by the  $Z^0$  and  $W^\pm$  bosons.

Gravity is not included. The last particle in the SM is the Higgs particle, whose field is responsible for giving mass to the fermions and the weak bosons. Figure 2.4 summarizes the SM particle content and properties. [19]

### 2.3.1 Standard Model

The SM is a quantum field theory that combines elements of classical field theory with quantum mechanics and special relativity [81]. Therein, particles are quantized excitations of a corresponding field. The matter particles (fermions), which are all spin  $1/2$  particles adhering to Pauli's exclusion principle [82], are represented by spinor fields, gauge bosons by vector fields, and the Higgs particle by a scalar field [10]. Also, as a natural consequence of quantum field theory, each particle has a corresponding antiparticle [83]. Furthermore, the SM can be classified as a non-abelian gauge theory (= Yang-Mills theory [84]) with the symmetry group  $U(1) \times SU(2) \times SU(3)$ . This means the SM Lagrangian is required to be invariant under local transformations from representations of these groups, which gives rise to the gauge vector fields and, ultimately, the interactions themselves. [19]

The  $SU(3)$  symmetry leads to quantum chromo dynamics (QCD) [62, ch. 9] that describes the strong force between quarks and gluons. The strong force does not mix with the others and can be treated independently [15]. The remaining symmetries constitute the electroweak theory, also called the Glashow-Weinberg-Salam (GWS) model [36–38]. The GWS model unifies the electromagnetic interaction described by quantum electrodynamics (QED) [35] and the weak interaction [6]. It incorporates the Brout-Englert-Higgs mechanism<sup>18</sup> [85–88] to generate the masses of the weak vector bosons by combining the concept of *spontaneous symmetry-breaking* with local gauge invariance [15].

### 2.3.2 Electroweak sector

The electroweak interaction is based on the combined symmetry group  $SU(2)_{I_3} \times U(1)_{Y_W}$ . Invariance of the SM Lagrangian to either of these symmetries results in the conserved Noether charges [89] of the weak isospin  $I_3$  and the weak hypercharge  $Y_W$ . However,

<sup>18</sup>Many scientists contributed to its formulation, but usually it simply referred to as “Higgs mechanism”.



concerning charged leptons and neutrinos. Defining the doublet as left-handed realizes the desired V-A coupling of the weak interaction. Since the charged leptons take part in electromagnetic interaction, which is not chiral, their right-handed partners must be included. Since these cannot participate in V-A interaction, they are SU(2) singlets. Right-handed (sterile) neutrino singlets are neglected due to missing evidence for their existence and properties. However, as discussed in the next paragraph and section 2.4, right-handed neutrinos can be included in models beyond the Standard Model to construct neutrino mass terms for the Lagrangian. [10, 19]

### 2.3.3 Lepton masses

For massive leptons, the Lagrangian of a quantum field theory must include an appropriate mass term of the general shape  $\mathcal{L}_{\text{mass}} = -m\bar{\ell}\ell$ , where  $m$  is the mass,  $\ell$  is the particle's spinor, and  $\bar{\ell}$  the adjoint spinor.<sup>19</sup> However, in the SM, the addition of such terms would break the  $SU(2)_{I_3} \times U(1)_{Y_W}$  gauge symmetry. Instead, the mass terms are generated during spontaneous symmetry-breaking by introducing a Yukawa coupling [92] with coupling strength  $G_\ell$  to the Higgs doublet  $\phi$ . The Yukawa coupling can be written by sandwiching  $\phi$  between the left-handed lepton doublet and the right-handed singlet [19]

$$\mathcal{L}_Y = -G_\ell (\bar{\nu}_\ell, \bar{\ell})_L \phi \ell_R + h.c. . \quad (2.20)$$

Electroweak symmetry-breaking (EWSB) shifts the Higgs field's vacuum expectation value away from zero. Mathematically, this is expressed by expanding the Higgs doublet around its new vacuum expectation value  $v \approx 246 \text{ GeV}$ <sup>20</sup> [10]

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \rho(x) \end{pmatrix} , \quad (2.21)$$

which splits this Lagrangian into a mass term and a Lepton-Higgs interaction term [19]

$$\begin{aligned} \mathcal{L}_Y &= -\frac{G_\ell}{\sqrt{2}} (v \bar{\ell}_L \ell_R + \bar{\ell}_L \ell_R \rho) + h.c. \\ &= \underbrace{-\frac{G_\ell}{\sqrt{2}} v \bar{\ell} \ell}_{\mathcal{L}_{\text{mass}}} - \underbrace{\frac{G_\ell}{\sqrt{2}} \rho \bar{\ell} \ell}_{\mathcal{L}_{\text{int}}} . \end{aligned} \quad (2.22)$$

So, the lepton mass is given by  $m_\ell = v G_\ell / \sqrt{2}$ , virtually replacing the mass with a coupling constant. While this doesn't give new insight into the particular values of lepton masses, the interaction term implies that, according to the SM, the Higgs couples to the mass with  $m_\ell/v$ . This is subject to experimental tests at the LHC, and so far, no deviation from this prediction has been found [93].

Notably, the absence of a right-handed neutrino in equation (2.20) makes it impossible to generate neutrino mass terms in the Standard Model. One method for including neutrino mass terms for models beyond the Standard Model is discussed in section 2.4.

<sup>19</sup>The inclusion of the mass term is required to end up with the Dirac equation, describing the motion of free fermions, when the Euler-Lagrange equation is applied to the Lagrangian. [15]

<sup>20</sup>The vacuum expectation value can be calculated from the Fermi constant via  $v = (\sqrt{2}G_F)^{-1/2}$ . [10]

## 2.4 Massive Neutrinos

As discussed in section 2.2, the phenomenon of neutrino oscillation proves that neutrinos have mass. In section 2.3 it is shown that the Standard Model is not able to account for this experimental fact since the omission of sterile neutrinos from the particle content does not allow for the same mass generation mechanism as for the other fermions. Section 2.4.1 addresses possible extensions to the standard model to account for massive neutrinos.

Another mystery concerning neutrinos is that in comparison to their respective charged leptons, their masses are extremely small (sub-eV scale, see section 2.5) without apparent reason. If neutrino fields are of pure Dirac-type and couple to the Higgs with Yukawa couplings, the coupling constants must be about six orders of magnitude smaller than those of the charged leptons [10]. A natural way of explaining the smallness of neutrino masses lies in the so-called *seesaw mechanism*, which is outlined in section 2.4.2.

Extensions of the Standard Model's neutrino sector often include sterile neutrinos of various mass scales. Section 2.4.3 discusses which mass values may be of particular interest to explain anomalies measured in reactor-based oscillation experiments<sup>21</sup>, or to provide suitable dark matter candidates for explaining cosmological observations<sup>22</sup>.

### 2.4.1 Dirac and Majorana Mass Terms

A simple method of generating neutrino mass terms is to add three right-handed neutrinos

$$(\nu_e)_R \quad (\nu_\mu)_R \quad (\nu_\tau)_R \quad (2.23)$$

to the SM particle content as singlets under the flavor SU(2) symmetry. Hence, they carry no weak charge and are *sterile* concerning all SM interactions. Nevertheless, they can mix with the other neutrino states via an extended PMNS matrix. [10]

With these additional fields, it is possible to construct so-called *Dirac mass terms* by coupling to the Higgs doublet like for the other fermions (see section 2.3.3). This results in mass terms of the form [10]

$$\mathcal{L}_D = -m_D(\bar{\nu}_L\nu_R + \bar{\nu}_R\nu_L) \quad \text{with} \quad m_D = G_\nu \frac{v}{\sqrt{2}}, \quad (2.24)$$

where  $v$  is the Higgs vacuum expectation value and  $G_\nu$  the Yukawa coupling constant.

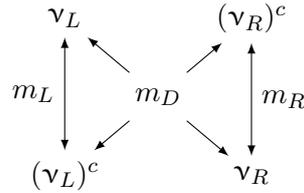
Unlike the other fermions, neutrinos carry no electric charge and could therefore be so-called Majorana fermions. First suggested by E. Majorana in 1937 [25, 26], these are hypothetical fermions whose spinor fields are equal to their charge conjugate<sup>23</sup>:  $\psi^c = \psi$ .<sup>24</sup>

<sup>21</sup>See [94] for detailed information on oscillation anomalies.

<sup>22</sup>A broad overview on neutrinos as dark matter can be found in [95].

<sup>23</sup>Charge conjugation transforms particle spinors into antiparticle spinors and vice-versa, but it also affects handedness of the chiral projections:  $\psi_L^c \equiv (\psi_L)^c = (\psi^c)_R$ . [19]

<sup>24</sup>As discussed in section 2.1, it has been established that there must be a difference between neutrino and antineutrino. If neutrinos were Majorana particles, solely their spin state/helicity would distinguish the particle from its antiparticle. [15]



**Figure 2.5:** Graphical representation of neutrino field couplings through the Dirac mass  $m_D$  and the Majorana masses  $m_{L/R}$ . Illustration based on a figure from [10].

In this case, the charge conjugated spinor  $\nu^c$  may be used to form the Lorentz scalars  $\bar{\nu}^c\nu$  and  $\bar{\nu}\nu^c$ , allowing for the construction of *Majorana mass terms*: [10, 19]

$$\mathcal{L}_M = -\frac{1}{2}m_L(\bar{\nu}_L^c\nu_L + \bar{\nu}_L\nu_L^c) - \frac{1}{2}m_R(\bar{\nu}_R^c\nu_R + \bar{\nu}_R\nu_R^c). \quad (2.25)$$

Here,  $m_L$  and  $m_R$  are Majorana masses. This equation implies that the interaction eigenstates  $\nu_{L/R}$  couple to their charge conjugate  $\nu_{L/R}^c$  through the respective Majorana mass, whereas the Dirac mass  $m_D$  from equation (2.24) leads to a coupling of  $\nu_L^{(c)}$  and  $\nu_R^{(c)}$  (see figure 2.5). As a result, the neutrino may act as its antiparticle leading to the possibility of lepton number violating interactions such as neutrinoless double  $\beta$ -decay (see section 2.5.2). [10]

Dirac and Majorana mass terms are not exclusive to each other and the most general mass term is a combination of both. It can be written as a matrix equation: [10]

$$\mathcal{L}_{\text{mass}} = \mathcal{L}_D + \mathcal{L}_M = -\frac{1}{2}(\bar{\nu}_L, \bar{\nu}_L^c) \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} \nu_R^c \\ \nu_R \end{pmatrix} + h.c. . \quad (2.26)$$

The actual mass eigenvalues and eigenstates are obtained via diagonalization. Since  $m_D$  and  $m_{L,R}$  are in principle unconstrained, a multitude of possible scenarios are allowed (e. g. pure Dirac  $m_{L,R} = 0$ , pure Majorana  $m_D = 0$ , pseudo-Dirac  $m_D \gg m_{L,R}$ ). Also, the mass term can be generalized to  $n$  flavors by replacing the spinors with  $n$ -component vectors. Accordingly,  $m_D$ ,  $m_L$ , and  $m_R$  are replaced by  $n \times n$  matrices. [10]

### 2.4.2 Seesaw Mechanism

A particularly enticing scenario is the choice  $m_L = 0$  and  $m_R \gg m_D$ , i. e. the inclusion of a heavy right-handed Majorana neutrino. It defines the so-called *type I seesaw mechanism* [96–100], which explains the smallness of the active neutrino masses. To illustrate the general idea, the mechanism for one neutrino generation is outlined here. Diagonalizing the mass matrix from equation (2.26) for this case yields the mass eigenvalues [10]

$$m_1 = \frac{m_D^2}{m_R} \quad \text{and} \quad m_2 = m_R \left( 1 + \frac{m_D^2}{m_R^2} \right) \approx m_R. \quad (2.27)$$

The effective mass  $m_1$ , belonging in mixture primarily to the active flavor, is effectively suppressed by the large Majorana mass associated with the sterile state. This scheme is motivated by the notion that  $m_D$  is preferably small to remain in accord with other SM Yukawa couplings. At the same time, the Majorana mass is, without further assumptions, not bound to any particular scale. The light neutrino masses on the sub-eV scale can then be explained via Majorana masses well above the electroweak scale. [10]

Apart from adding heavy right-handed Majorana fields, similar seesaw mechanisms can also be realized in models that feature Higgs triplets [100–102] (*type II seesaw*), or through the addition of lepton triplets [103, 104] (*type III seesaw*). And besides the type I-II-III seesaw mechanisms, which induce neutrino masses on the tree level [105], there are many other models capable of realizing naturally small neutrino masses. For example, neutrino mass generation via radiative quantum loops [106–108], or the *inverse seesaw mechanism* [109] where the type I seesaw is extended by additional singlets that are not sterile neutrinos. [110]

### 2.4.3 Sterile Neutrino Mass Scales

The vast landscape of neutrino mass theories offers many models introducing sterile neutrinos, and virtually any sterile mass scale can be realized [111]. The models can be constructed to explain experimental observations and indications, such as neutrino masses, mixing angles, oscillation anomalies, dark matter, and baryon asymmetry in the early universe [110]. From the experimental point of view, two mass scales for sterile neutrinos are particularly interesting due to their accessibility and physics motivation:

- $\mathcal{O}(0.1 - 10)$  **eV**: Sterile neutrinos with masses in the low eV range are motivated by anomalies in short-baseline oscillation experiments, where observed flux deficits may be explained by disappearance into a sterile state. These are the *LSND Anomaly* [112] and *MiniBooNE Low-Energy Excess* [113], the *Reactor Antineutrino Anomaly* [114], and the *Gallium Anomaly* [115]. A common model for this case is the minimalistic 3+1 scenario, where one light sterile neutrino is added to the active flavors. While this model can explain the anomalies, conventional explanations are not ruled out and there is tension between experiments (further information in section 2.6.1). [94, 116]
- $\mathcal{O}(1 - 100)$  **keV**: Sterile neutrinos in the keV mass range have been suggested as natural candidates for *Warm Dark Matter* (WDM) [117–121]. At the same time, the mass scale up to several MeV is accessible in weak decay laboratory experiments (further information in section 2.6.2). A popular model designed with cosmological observations of dark matter and baryon asymmetry in mind is the Neutrino Minimal Standard Model ( $\nu$ MSM) [122, 123]. It introduces three sterile neutrinos with masses below the electroweak scale (one on the keV scale, two on the GeV scale) in a mixed Dirac-Majorana case. [66, 95]

The experimental status regarding both mass scales is discussed in chapter 3, focusing especially on the keV mass scale due to the topic of this thesis.

## 2.5 Absolute Neutrino Mass Measurements

Since neutrino oscillation measurements are only sensitive to the squared difference of the neutrino masses, the absolute mass scale must be determined differently. Three main experimental approaches are pursued: Constraints from cosmological observations, neutrinoless double  $\beta$ -decay, and direct measurements based on kinematics [17]. Each method is sensitive to a different observable for the absolute mass scale, with varying degrees of model dependence. Oscillation parameter measurements (see also section 2.2.3) can be used to constrain the allowed regions for these observables as a function of the lightest neutrino mass (see [62, ch. 14], for example).

### 2.5.1 Cosmological Constraints

According to the current understanding of the universe’s evolution, the seeds for structure formation originate from primordial quantum fluctuations, which were enlarged by inflation. These early structures leave an imprint in the *Cosmic Microwave Background* (CMB), measured with great detail by the *Planck* satellite [124]. Anisotropies in the CMB correlate to today’s structure in the universe observed with astronomical surveys, such as the Sloan Digital Sky Survey (SDSS) [125]. Overall, the observations are well-described by the standard cosmological model, known as  $\Lambda$ CDM [62, ch. 25].

The freeze-out of the electroweak force in the early universe due to Hubble expansion<sup>25</sup> led to the decoupling of neutrinos from thermal equilibrium. These relic neutrinos constitute the *Cosmic Neutrino Background* (C $\nu$ B) [9]. Due to their low mass, relic neutrinos are relativistic and act as *Hot Dark Matter* (HDM), smearing out density fluctuations at small scales. The particular scale at which structure formation is affected is determined by the free streaming length of neutrinos, which depends on their mass [17]. Thus, the absolute neutrino mass scale can be constrained with measurements of the CMB multipole spectrum and the small-scale region of the matter power spectrum obtained from galaxy surveys. The observable is the sum of the neutrino masses  $m_i$  [62, ch. 26]

$$m_{\text{tot}} = \sum_i m_i , \quad (2.28)$$

since the degeneracy between individual neutrino masses cannot be resolved [126, 127]. Analyses of the Planck CMB spectra and lensing data [124, 128] in combination with various survey datasets by SDSS and DES [125, 129] yield limits in the range [130–134]

$$m_{\text{tot}} < 0.087 \text{ eV to } 0.54 \text{ eV} \quad (95 \% \text{ CL}) . \quad (2.29)$$

A summary of limits for different data selections is provided in [62, p. 480]. Besides the dependence on dataset selection, cosmological parameter estimation is also model and analysis dependent [17, 127]. Thus, direct laboratory-based measurements are required.

<sup>25</sup>This happened at a temperature of  $\mathcal{O}(1 \text{ MeV})$  corresponding to roughly 1 s after the big bang. [9]

### 2.5.2 Neutrinoless Double $\beta$ -Decay

Neutrinoless double  $\beta$ -decay ( $0\nu\beta\beta$ ) is the most sensitive probe for answering whether neutrinos are Majorana or Dirac fermions [62, ch. 14]. The corresponding reaction equation is

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- . \quad (2.30)$$

Figure 2.6 shows the lowest-order Feynman diagram for this process. The decay violates lepton number conservation by two units and may only occur if neutrinos are Majorana fermions [25, 135]. Non-zero Majorana mass is also needed since a spin flip is required [10]. Furthermore, the decay is only allowed in nuclei that undergo regular double  $\beta$ -decay ( $2\nu\beta\beta$ ), and for practical reasons, single  $\beta$ -decay must be energetically forbidden or suppressed otherwise. Hence, only 35 particular nuclei with  $(Z, A)$  in an even-even configuration are potential  $2\nu\beta\beta$ -emitters. [10]

Experiments search for  $0\nu\beta\beta$  by measuring the decay energy spectrum of a  $2\nu\beta\beta$ -emitter and looking for a peak at the  $Q$ -value of the transition. At the same time, regular double  $\beta$ -decay exhibits a continuous spectrum below  $Q$  since the neutrinos carry away some of the energy. At the time of writing, neutrinoless double  $\beta$ -decay has not been observed, and the best limits on the half-life are obtained from measurements with  $^{76}\text{Ge}$  by GERDA [136], and with  $^{136}\text{Xe}$  by KamLAND-Zen [137] (90 % CL):

$$T_{1/2}^{0\nu}(^{76}\text{Ge}) > 1.8 \times 10^{26} \text{ yr} \quad \text{and} \quad T_{1/2}^{0\nu}(^{136}\text{Xe}) > 1.07 \times 10^{26} \text{ yr}. \quad (2.31)$$

The half-life relates to the absolute neutrino mass scale via [138]

$$(T_{1/2}^{0\nu})^{-1} = G_{0\nu} |M_{0\nu}|^2 m_{\beta\beta} , \quad (2.32)$$

where  $G_{0\nu}$  is the phase space factor,  $M_{0\nu}$  is the nuclear matrix element, and  $m_{\beta\beta}$  is the *effective Majorana mass* of  $\nu_e$ . The latter relates to the neutrino mass states via

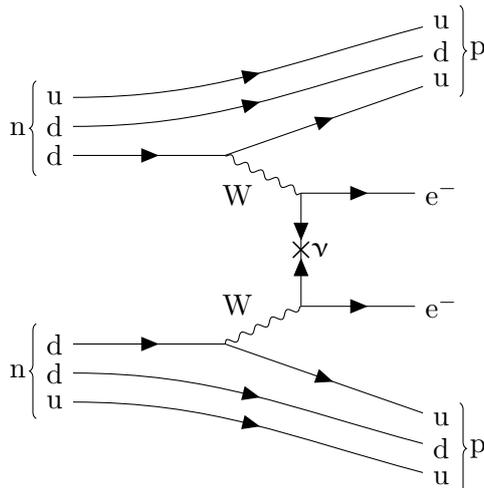
$$m_{\beta\beta} = \left| \sum_i U_{ei}^2 m_i \right| = \left| m_1 |U_{e1}|^2 + m_2 |U_{e2}|^2 e^{i(\eta_2 - \eta_1)} + m_3 |U_{e3}|^2 e^{i(-\eta_1 - 2\delta_{\text{CP}})} \right| . \quad (2.33)$$

Since this is a coherent sum, the complex Dirac phase  $\delta$  and the unknown Majorana phases  $\eta_{1,2}$  (see section 2.2.2) may lead to a (partial) cancellation of summation terms.

Neutrinoless double  $\beta$ -decay could also be induced by other new physics effects apart from Majorana masses. Hence, the observation of  $0\nu\beta\beta$  translates to the neutrino mass only under the assumption that the Majorana mass is the only source of lepton number violation [62, p. 307]. Furthermore, calculating the nuclear matrix element is tied to significant theoretical uncertainty with variations of up to factor three between predictions obtained via different methods [138]. Nevertheless, the half-life measurements from above lead to the respective limits (90 % CL) [136, 137]

$$m_{\beta\beta} < 0.079 - 0.180 \text{ eV} \quad \text{and} \quad m_{\beta\beta} < 0.061 - 0.165 \text{ eV} , \quad (2.34)$$

where the ranges account for the matrix element dependence.



**Figure 2.6:** Feynman graph for neutrinoless double  $\beta$ -decay: Two neutrons decay to protons under the emission of two  $\beta$ -electrons, while the neutrino emitted from one vertex acts as its antiparticle and cancels the neutrino of the other vertex, which is only possible if neutrinos are Majorana fermions [25].

### 2.5.3 Direct Neutrino Mass Determination

The ultimate method for probing the absolute neutrino mass scale is a direct measurement based on kinematics. One option is the time-of-flight technique using neutrinos from core-collapse supernovae. Neutrino observatories can measure the arrival time of the large neutrino flux emitted by such an event, and the distance is sufficient to be sensitive to the effective electron neutrino mass (see equation (2.35) below). Measurements of the supernova SN1987a neutrino flux set the limit  $m_{\bar{\nu}_e} < 5.7 \text{ eV}$  (95 % CL) [139].

Another approach is the laboratory-based precision measurement of weak decay spectra, which surpasses the sensitivity of time-of-flight measurements [17]. Unlike the supernova measurement, which is based on a neutrino emission model, precision spectroscopy has the additional benefit of being entirely model-independent. The neutrino mass's signature is a distortion at the endpoint of the electron energy spectrum from  $\beta$ -decay or the electron capture de-excitation spectrum (EC) [17].<sup>26</sup> The primary isotopes used for these measurements are tritium ( $\beta$ -decay) and holmium-163 (EC). Figure 2.7 demonstrates the signature by the example of tritium decay.

Since the electron neutrino is a mixture of the mass states, the  $\beta$ -decay spectrum is a superposition of spectra from all  $m_i$  [141]. However, the mass splittings are much smaller than the experimental energy resolution of  $\mathcal{O}(\text{eV})$ . The observable quantity is, therefore, a weighted average of the squared masses [17]:

$$m_{\nu_e} = \sqrt{\sum_i |U_{ei}^2| m_i^2}, \quad (2.35)$$

<sup>26</sup>This signature was suggested by E. Fermi in his original publication on  $\beta$ -decay [140] (see section 2.1).

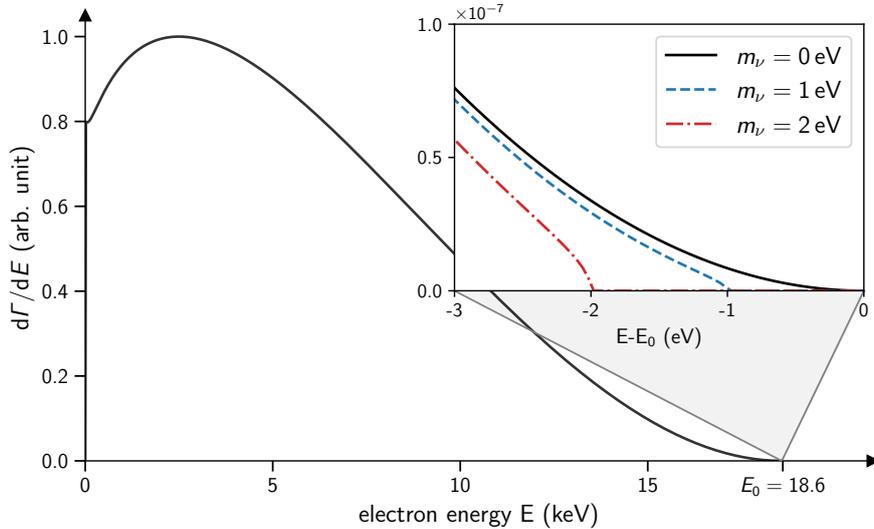
which is called the *effective electron neutrino mass*.<sup>27</sup> In contrast to  $m_{\beta\beta}$ , this is an incoherent sum, so the complex phases of the PMNS matrix elements do not matter. Oscillation experiments provide a lower limit on the effective mass, which depends on the mass ordering:  $m_{\nu_e} > 48 \text{ meV}$  ( $8.5 \text{ meV}$ ) for IO (NO) at 95 % CL [62].

The KATRIN experiment currently sets the lowest limit on the effective antineutrino mass by measuring the  $\beta$ -spectrum of molecular tritium with an integrating MAC-E filter [4, 5]. Using data from the first two measurement phases, the limit

$$m_{\bar{\nu}_e} < 0.8 \text{ eV} \quad (90 \% \text{ CL}) \quad (2.36)$$

is obtained [7]. This improves upon previous results from KATRIN's predecessors: the Mainz experiment with  $m_{\bar{\nu}_e} < 2.3 \text{ eV}$  (95 % CL) [142] and the Troitsk experiment with  $m_{\bar{\nu}_e} < 2.05 \text{ eV}$  (95 % CL) [143]. KATRIN data-taking will continue until the end of 2025, and the sensitivity goal for the final dataset is  $0.2 \text{ eV}$  [5]. With this precision, KATRIN could determine the absolute neutrino mass scale in the quasi-degenerate scenario where  $m_1 \gg \Delta m_{12}, \Delta m_{32}$  [17]. The Project 8 experiment [144] aims for even lower sensitivity by measuring the atomic tritium spectrum using cyclotron radiation emission spectroscopy. Its projected sensitivity of  $40 \text{ meV}$  could probe the entire inverted hierarchy scale.

Complementary searches for the effective electron neutrino mass are performed by measuring the de-excitation energy spectrum of  $^{163}\text{Ho}$  electron capture [17, 145]. Currently, three experiments are pursuing this approach [62]: ECHo [146], HOLMES [147], and NuMECS [148].



**Figure 2.7:** Imprint of the neutrino mass in the tritium  $\beta$ -decay energy spectrum. Some energy is required to account for the neutrino mass, which shifts the effective endpoint. Additionally, the spectral shape in the vicinity of the endpoint is distorted.

<sup>27</sup>EC and  $\beta$ -decay are sensitive to the effective masses of  $\nu_e$  and  $\bar{\nu}_e$ , respectively.

## 2.6 Sterile Neutrino Searches

The extension of the Standard Model of Particle Physics (SM) with right-handed neutrinos is a natural method of explaining neutrino masses (see section 2.4). Since right-handed neutrinos would not couple to any of the SM interactions, they are also called *sterile* neutrinos. In most extended models, introducing sterile neutrinos is tied to additional mass eigenstates that mix with the active neutrino flavors. Mixing makes it possible to search for sterile neutrinos even though they cannot interact directly. As section 2.4.3 mentions, the eV and keV mass scales are especially interesting for experimental searches. This section provides an overview concerning both. The keV scale is of particular interest to the topic of this thesis, and the search for keV sterile neutrinos at KATRIN is covered in more detail in section 3.4.

### 2.6.1 eV-scale Sterile Neutrinos

Searches for sterile neutrinos on the eV mass scale are motivated by four long-standing anomalies observed in short-baseline oscillation experiments. Two stem from an apparent excess in the appearance of  $\nu_e$  in  $\nu_\mu$  beams from pion decay, and two from a normalization deficit of registered events in  $\nu_e$  fluxes from nuclear reactors and radioactive sources. All anomalies can be explained by flavor transitions where a sterile neutrino and a fourth mass eigenstate are added. However, many alternative solutions to the anomalies exist, including other new physics effects and conventional explanations. [94]

A description of the anomalies and their current status is provided in the following. For a comprehensive summary of experimental findings and constraints concerning eV-scale sterile neutrinos, see [94].

- **LSND Anomaly:** In 2001, the LSND collaboration reported an excess in their  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  appearance measurement [112]. The neutrinos were created through pion decay at rest and detected in the Liquid Scintillation Neutrino Detector (LSND) at about 30 m distance from the source. To explain the observed excess with flavor oscillation, a relatively large neutrino mass splitting of  $\Delta m^2 \sim 1 \text{ eV}^2$  is required [94]. The similar KARMEN experiment [149] did not observe an oscillation excess, thus rejecting the LSND result and excluding most of the preferred parameter space.
- **MiniBooNE Low-Energy Excess:** The MiniBooNE experiment was constructed to provide an independent test for the LSND Anomaly. It probes approximately the same  $\Delta m^2$  region by using higher energy muon (anti)neutrinos from the Booster Neutrino Beam at Fermilab [150] and looking for electron (anti)neutrino appearance at a longer oscillation baseline with the MiniBooNE Cherenkov detector [151]. In their final result [113], the MiniBooNE collaboration reported a  $4.8\sigma$  excess of electron-like events over the background expectation. The excess is most prominent in the low-energy region as a function of the reconstructed electron energy. While the overall normalization is consistent with the preferred oscillation parameters by LSND, the spectral shape cannot be accounted for [94]. Thus, most of the preferred LSND parameter space is disfavored at 95 % CL, but the origin of the new anomaly remains unclear [94].

To investigate potential solutions for the low energy excess, the MicroBooNE liquid-argon time projection chamber [152] with better event reconstruction was constructed in the same neutrino beam at a shorter distance. So far, no explanation for the MiniBooNE result has been found [153]. Recently published results by the MicroBooNE collaboration [154] exclude large parts of the parameter regions preferred by LSND and the Gallium experiments.

- **Reactor Antineutrino Anomaly:** In short-baseline reactor experiments, the  $\bar{\nu}_e$  flux is measured and compared to an expectation of the reactor spectrum. In 2011, new estimations for reactor spectrum [155, 156] concluded that the total flux should be  $\approx 3\%$  larger than previously assumed. This led to a disagreement with reactor experiment data, referred to as the Reactor Antineutrino Anomaly (RAA) [114]. A potential explanation for the discrepancy could be the disappearance of neutrinos into a sterile state via oscillation  $\bar{\nu}_e \rightarrow \bar{\nu}_s$ . However, in light of recent flux model refinements, the tension was lifted considerably [157]. Furthermore, reactor fuel-dependent measurements at Daya Bay [158] and RENO [159] indicate mis-modeling of the flux [160]. In a statistical re-evaluation with the Feldman-Cousins method by [161], it is found that reactor data are consistent with the “no-sterile neutrino” hypothesis when combined with solar neutrino constraints.
- **Gallium Anomaly:** Complementary to the RAA, neutrino flux deficits were also observed in radiochemical experiments that use intense radioactive  $\nu_e$  sources ( $^{51}\text{Cr}$  and  $^{37}\text{Ar}$ ) and detect the rate with neutrino induced  $^{71}\text{Ga} \rightarrow ^{71}\text{Ge}$  transitions. Four combined measurements by the GALLEX [162] and SAGE [163, 164] experiments yield an average ratio of observed to predicted event rates of  $R = 0.86 \pm 0.05$  [94], which is a  $\sim 3\sigma$  deficit. In 2021, the anomaly was confirmed by the follow-up experiment BEST [165, 166], which reported an observed rate deficit of  $\approx 20\%$  with  $> 5\sigma$  statistical significance [167]. An explanation of the combined gallium data with the 3+1 sterile neutrino hypothesis requires a large mixing angle  $0.21 \leq \sin^2 2\theta \leq 0.47$  ( $2\sigma$  interval) and  $\Delta m^2 \gtrsim 1 \text{ eV}^2$  [161]. That being said, the region with  $\Delta m^2 \lesssim 10 \text{ eV}^2$ , which includes the best-fit point, is mostly excluded by reactor and accelerator data [161, 168] (see also [94]). Furthermore, the entire allowed parameter region is in severe tension with solar neutrino data [161, 169]. The remaining parameter space with  $\Delta m^2 \gtrsim 10 \text{ eV}^2$  is probed by KATRIN, which is the only ongoing laboratory experiment sensitive in this region [170] (see next paragraph).

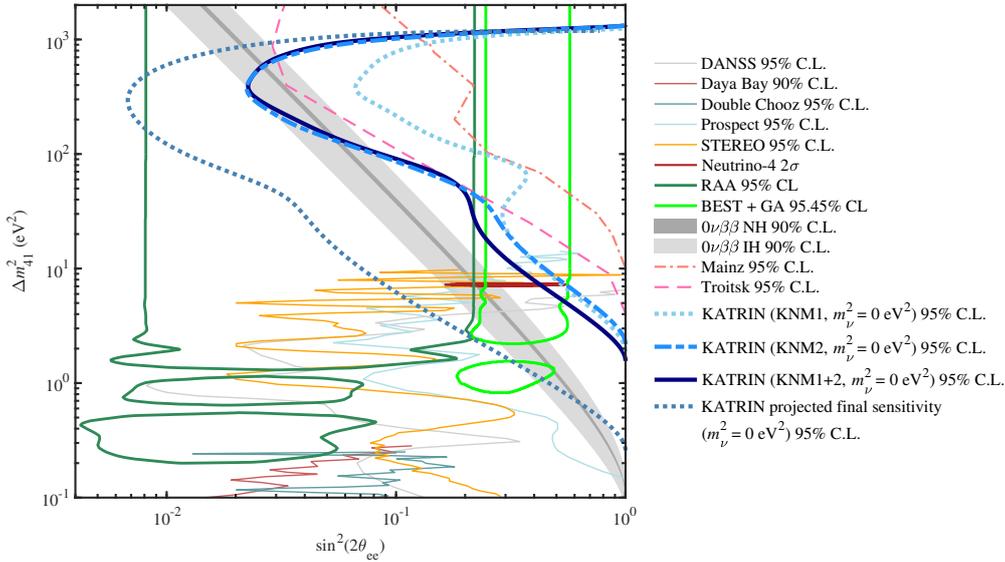
It should also be mentioned that the Neutrino-4 experiment reported positive evidence of a sterile neutrino based on detecting an oscillatory pattern in their reactor neutrino  $L/E$  spectrum [171, 172]. However, the region in the  $(\sin^2(2\theta), \Delta m^2)$  plane that can explain the pattern conflicts with solar neutrino measurements [169] and is also rejected by STEREO [173] and PROSPECT [174]. Furthermore, the data analysis was intensely debated [175, 176]. For instance, Wilk’s theorem was incorrectly applied [177] and a revised Monte Carlo-based evaluation yields a significance well below  $3\sigma$  [172].

### Direct Search at KATRIN

The KATRIN experiment [4, 5] (see chapter 3) measures the endpoint region of the tritium  $\beta$ -spectrum with unprecedented precision to determine the effective electron neutrino mass (see section 2.5.3). To achieve this, an intense gaseous molecular tritium source is combined with an integrating spectrometer (MAC-E filter) with  $\mathcal{O}(1 \text{ eV})$  energy resolution. The integral spectrum data is analyzed down to 40 eV below the endpoint  $E_0 = 18.57 \text{ keV}$  of the tritium spectrum.

Since the  $\beta$ -spectrum is a superposition of spectra corresponding to the mass states that mix with the electron antineutrino [141], further mass eigenstates introduced through sterile neutrinos would also contribute. This leads to a kink-like signature at  $E_0 - m_4$  in the spectrum, where  $m_4$  is the additional mass value (see section 3.4.1). Hence, KATRIN is sensitive to sterile neutrinos up to  $\Delta m_{41}^2 \lesssim (40 \text{ eV})^2 = 1600 \text{ eV}^2$  using the data from neutrino mass measurement campaigns. This probes especially the eV-scale's large  $\Delta m_{41}^2$  region, complementing oscillation experiments sensitive to small  $\Delta m_{41}^2$ . [170]

So far, the first [178] and second [170] measurement campaigns have been analyzed using a robust, blinded spectral shape analysis. The datasets have statistics of  $1.48 \times 10^6$  and  $3.76 \times 10^6$   $\beta$ -electrons, respectively. Figure 2.8 shows the resulting exclusion curves. The combined result of both campaigns already rejects a major portion of the parameter space with  $\Delta m^2 \gtrsim 10 \text{ eV}^2$  suggested by the Gallium Anomaly, and parts of the RAA region. The final sensitivity of KATRIN, combined with the projected sensitivity of future reactor experiments [179–181], fully covers the parameter space preferred by the Gallium Anomaly [160].



**Figure 2.8:** Exclusion contours from the first two KATRIN measurement campaigns and the projected final sensitivity for 1000 live measurement days. Also shown are the closed contours of the RAA and the Gallium Anomaly, the claimed signal by Neutrino-4, and exclusions from other experiments. Figure adapted from [170].

## 2.6.2 keV-scale Sterile Neutrinos

Sterile neutrinos on the keV mass scale are excellent candidates for *Warm Dark Matter* (WDM) as they are sufficiently heavy and cannot interact directly with Standard model particles. A mechanism for abundantly producing sterile neutrinos in the early universe was first proposed by Dodelson and Widrow in 1994 [117]. This section examines the implications of cosmology for the sterile neutrino parameter space, assuming they account for all of the observed dark matter ( $\Omega_{\text{DM}} = 0.12h^{-2} = 26.4\%$  [130]). Afterward, laboratory searches and current limits are discussed.

### Sterile neutrino decay

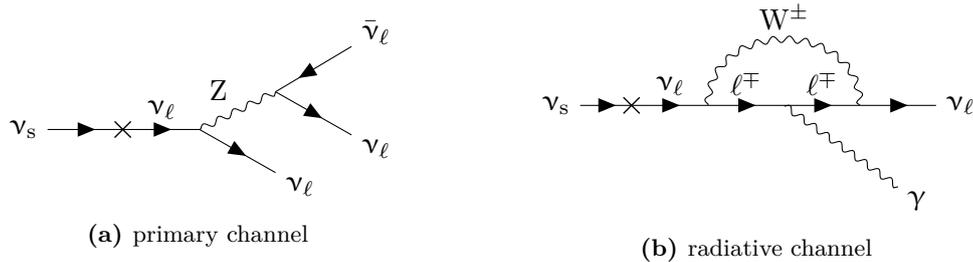
Heavy sterile neutrinos<sup>28</sup> are expected to decay into active neutrinos through mixing. For masses below twice the electron mass, the main decay channel is  $\nu_s \rightarrow \bar{\nu} \nu \nu$  (figure 2.9a), which determines the lifetime due to its dominance. For sterile neutrinos to be dark matter (DM) candidates, they must be stable on cosmological timescales ( $\tau > \tau_{\text{universe}}$ ), which sets an upper limit on the mixing amplitude. Still, tighter limits are obtained from requirements concerning the production mechanism. Subdominantly, sterile neutrinos decay radiatively  $\nu_s \rightarrow \nu \gamma$  (figure 2.9b), which is a much more testable prediction since a monoenergetic photon with energy equal to half the sterile neutrino mass is emitted. Searching for *unidentified X-ray lines* (UXL) using space-bound X-ray observatories thus represents a vital experimental test for sterile neutrino DM (see next paragraph). [66]

### Cosmological constraints

Cosmological observations lead to bounds on mass and mixing amplitude of sterile neutrino DM (shown in figure 2.10). The most important aspects are summarized here:

- **Production mechanism:** In the original Dodelson-Widrow mechanism [117], sterile neutrinos are thermally produced in the early universe through oscillation and scattering-induced decoherence (didactical overviews can be found in [66, 95, 182]). This can be used to construct an upper limit for the active-to-sterile mixing amplitude, above which too much DM would be produced. In the so-called Shi-Fuller mechanism [118], the production of sterile neutrinos is enhanced through lepton-number-driven resonant MSW conversion [59, 60] during big-bang nucleosynthesis (BBN). This approach allows a smaller mixing angle to create the appropriate DM amount. This mechanism is employed, for instance, in the *Neutrino Minimal Standard Model* ( $\nu\text{MSM}$ ), a popular reference model for cosmological observations. Based on BBN constraints of lepton asymmetry, the Shi-Fuller mechanism predicts a model-dependent lower bound for the mixing amplitude [182]. Further mechanisms include non-thermal production via new gauge interactions or decay of heavy particles [66, 95].

<sup>28</sup>For simplicity, the term “sterile neutrino” refers to either the flavor or the additional mass eigenstate, depending on context. This makes sense as long as the additional mass state is primarily sterile, i. e. the mixing with active flavors is small.



**Figure 2.9:** Leading Feynman diagrams for sterile neutrino decay (based on figures from [182] and [66]). In (b), the photon can also couple to the W boson instead of the charged lepton.

- Phase space considerations:** A robust lower bound on the mass of fermionic DM can be obtained from Pauli’s exclusion principle [82] since the phase space density is not allowed to exceed that of the degenerate Fermi gas. The limit can be further strengthened by including information on the observable coarse-grained phase space distribution (PSD) in today’s universe. Relating this coarse PSD to the primordial fine-grained PSD predicted by the DM model - the coarse PSD can never exceed the maximum of the fine PSD - gives the so-called Tremaine-Gunn bound [183]. [184]

The details in the construction of this bound vary, leading to several different lower limits between 0.2 keV and 6 keV [184, 185]. A conservative study [186] finds the limit  $m_s \gtrsim 0.1$  keV, which is weaker but more robust and independent of the DM formation model. [182]

- Structure formation:** The inclusion of WDM in the DM cocktail helps to explain problems concerning small-scale structure formation faced by standard cosmology, which mainly supposes cold dark matter (CDM). Differences between CDM and WDM emerge on the galactic sub-Mpc scale, while both DM types remain indistinguishable on large scales [66]. To summarize, the small-scale challenges of the CDM model concern the rotational curves of disk galaxies [187], the cusp-core problem [188], the missing satellites problem [189], and the too-big-to-fail problem [190]. A review addressing observations and possible solutions within the CDM or WDM models can be found in [191].

On the other hand, structure observations constrain the WDM free streaming length and, thus, the sterile neutrino mass. The main observational tool for this purpose is the Lyman- $\alpha$  forest method [192], which refers to measurements of absorption line spectra visible in spectra from distant light sources such as quasars. These lines stem from Lyman- $\alpha$  absorption in neutral hydrogen within the filaments of the *intergalactic medium* (IGM). The traversal of light through gas clouds at different redshifts results in a “forest” of lines. These measurements are used to construct a *flux power spectrum* probing the 0.1-1 Mpc scale, which can be compared to hydrodynamical structure formation simulations. The agreement between data and CDM simulations leads to sterile neutrino mass limits in the range  $m_s \lesssim 10$ -25 keV

[193, 194]. However, due to the potentially large effect of IGM gas dynamics on Lyman- $\alpha$  spectra [195], these bounds are often not displayed in literature [196]. [95] While the Lyman- $\alpha$  forest method yields the tightest constraints, another bound is obtained from Milky Way satellite galaxy counts. This method sets the limit  $m_s \lesssim 5\text{-}10\text{ keV}$  [197], depending on the degree of lepton asymmetry in resonant thermal DM production. Further methods are summarized in [95].

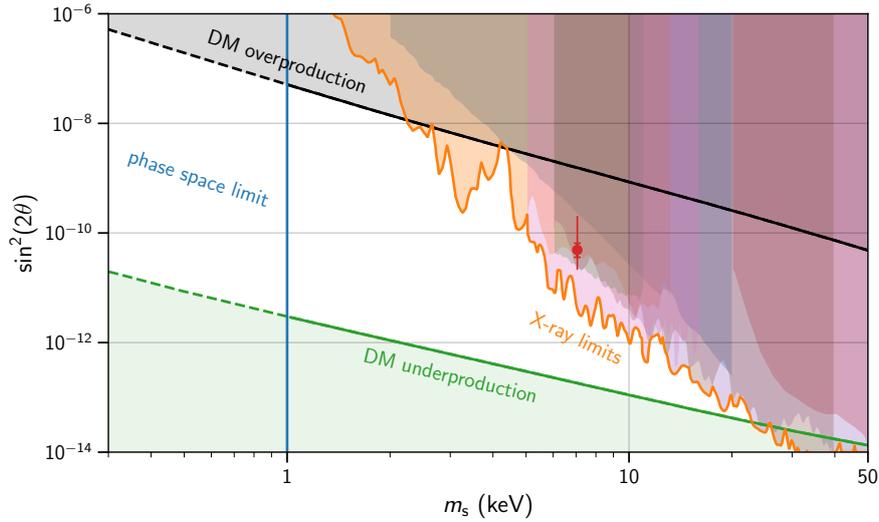
- **X-ray observations:** As mentioned above, keV sterile neutrinos can decay radiatively, emitting monoenergetic X-rays. The decay rate is proportional to  $\sin^2(2\theta)$  and  $m_s^5$  [182]. Consequently, the non-observation of unidentified cosmic X-rays constrains the mixing angle, especially for larger masses. Since cosmic X-rays cannot penetrate the atmosphere, they are measured using satellite telescopes. Some noteworthy X-ray observatories for this purpose are XMM-Newton, NuSTAR, Chandra, SRG, and Suzaku. For higher energies, the Fermi and INTEGRAL gamma-ray telescopes are also used. Current exclusions based on data from these missions are included in figure 2.10. These limits show that for masses  $\gtrsim 20\text{ keV}$ , X-ray observations cannot be reconciled with thermal DM production within the  $\nu\text{MSM}$  framework. Still, some parameter space is available depending on the lower mass limit from phase space arguments and structure formation.

Figure 2.10 also shows the signal candidate from the anomalous observation of an unidentified X-ray line with an energy of 3.5 keV, which was reported in 2014 [198, 199]. If this line is interpreted as a result of decaying sterile neutrinos, it corresponds to  $m_s \approx 7\text{ keV}$  and  $\sin^2(2\theta) \approx 5 \times 10^{-11}$  [198]. A summary of measurements concerning the line, including constraints and further positive results, can be found in [200]. More recent analyses based on data from Chandra [201], XMM-Newton [202, 203], and NuSTAR [204] exclude DM as the origin of the line.

### Laboratory Searches

Complementary to cosmological observations, keV sterile neutrinos can also be searched in terrestrial laboratory experiments. Usually, these investigations are based on kinematic principles in weak decays and provide model-independent exclusions, similar to direct neutrino mass search. Most existing limits are obtained from precision  $\beta$ -decay spectroscopy. As in the direct search for eV sterile neutrinos, sterile neutrinos would cause an additional spectral component below the energy  $E_0 - m_s$  (see also section 3.4.1), which leads to a characteristic kink-like signature. Current limits obtained from  $\beta$ -decay spectroscopy are shown in figure 2.11. This includes the KATRIN limit obtained from spectrum measurements during the first tritium campaign down to 1.6 keV below the endpoint [216]. In the near future, KATRIN's sensitivity will be significantly improved by measuring the differential tritium spectrum with unprecedented precision using the new TRISTAN detector (see section 3.4).

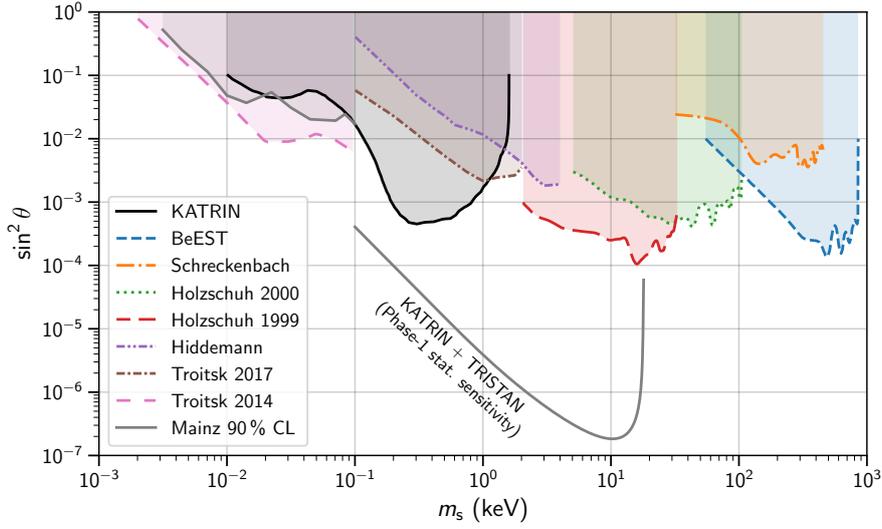
<sup>29</sup>Cosmological limits are usually reported in terms of  $\sin^2(2\theta)$ , while laboratory experiments set limits on  $\sin^2\theta$ . A quick conversion for small angles is to divide the cosmological limit by  $\approx 4$ .



**Figure 2.10:** Cosmological bounds on keV sterile neutrino dark matter.<sup>29</sup>Phase space limit and DM over-/underproduction limits are adopted from [184]. The upper line corresponds to non-resonant thermal DM production via the Dodelson-Widrow mechanism [117], while the lower limit is obtained with resonant Shi-Fuller production [118] using the maximal BBN lepton asymmetry allowed by the  $\nu$ MSM [122, 123]. X-ray limits (95% CL) were extracted from [201, 203–208]. The red dot indicates DM interpretation of the 3.5 keV X-ray line [198].

Another technique besides regular  $\beta$ -spectroscopy is to measure the atomic recoil in electron capture. This method is presently employed by the BeEST experiment [217], which uses superconducting tunnel junctions to measure the recoil energy from  ${}^7\text{Be}$  decay. The completed Phase II of BeEST sets an exclusion limit down to  $\sin^2\theta \approx 10^{-4}$  on the sub-MeV scale (also shown in figure 2.11). Phases III and IV of the BeEST experiment aspire to improve the statistics by more than two orders of magnitude [217].

Apart from KATRIN and BeEST, several other experiments aim to search for keV-scale sterile neutrinos in the future. These include the HUNTER experiment [218], which plans to perform total energy-momentum reconstruction in electron capture of  ${}^{137}\text{Cs}$  by suspending the decaying atoms in a vacuum with a magneto-optical trap. Another planned  $\beta$ -decay experiment is Magneto- $\nu$  [219], which measures the  ${}^{241}\text{Pu}$  spectrum using magnetic microcalorimeters. It operates in the same energy range as KATRIN and could potentially provide competitive limits [220]. Lastly, the calorimetric  ${}^{163}\text{Ho}$  electron capture experiments ECHo [146], HOLMES [147], and NuMECS [148], which are primarily designed for active neutrino mass measurements, can also provide sterile neutrino limits in the keV range.



**Figure 2.11:** Current laboratory limits on keV-scale sterile neutrino mixing [209–215], including KATRIN’s exclusion from the first tritium dataset [216] (95 % CL unless noted otherwise). The projected statistical sensitivity of KATRIN’s planned measurement using the new TRISTAN Phase-1 detector is also shown (from chapter 6 of this work).

Comparing figures 2.10 and 2.11, it is evident that cosmological constraints are much stronger than existing and even future laboratory limits. However, many unknowns exist concerning DM production, composition, and interactions. A study by Benso et al. [196] thus investigates two methods of relaxing cosmological bounds: The assumption of a DM cocktail, variation of the critical temperature for DM production, and the suppression of the sterile neutrino decay rate. It demonstrates cosmological constraints can be partially or even completely avoided, concluding that there are ample viable scenarios in which a laboratory discovery of sterile neutrinos can be reconciled with cosmological observations. For example, when sterile neutrinos account for 1-10 % of all DM, it is found that X-ray and DM overproduction constraints with  $m_s \lesssim 10$  keV are relaxed below the level accessible with KATRIN.

---

## The KATRIN Experiment

---

The Karlsruhe Tritium Neutrino (KATRIN) experiment [4, 5] aims to determine the absolute neutrino mass scale using tritium  $\beta$ -spectroscopy (see also section 2.5.3). The experiment is hosted by the Tritium Laboratory Karlsruhe (TLK) on Campus North of the Karlsruhe Institute of Technology. This facility is currently the only scientific laboratory operating a closed tritium cycle and is licensed to handle the required amount of tritium [5].

KATRIN measures the integral  $\beta$ -electron spectrum in a small region around the kinematic endpoint  $E_0 \approx 18.6$  keV, where the nonzero neutrino mass distorts the spectral shape. Unprecedented precision is achieved by combining an ultra-luminous gaseous tritium source with an activity of up to  $10^{11}$  Bq and a MAC-E filter spectrometer with an energy resolution in the eV range [5]. KATRIN started taking tritium data in 2019, and with data from the first and second measurement phases, the collaboration achieved the first direct neutrino mass measurement with sub-eV sensitivity in 2021 [7]. This sets the new best limit  $m_{\bar{\nu}_e} \lesssim 0.8$  eV (90% CL). At the time of writing, the ninth measurement phase has been completed, and data-taking will continue until 2025. The sensitivity goal for the final dataset of 1000 live days is 0.2 eV [5], which would improve upon the limit from predecessor experiments [142, 143] by one order of magnitude.

After the neutrino mass measurement concludes, KATRIN will perform a high statistics measurement of the differential tritium spectrum to energies deep below the endpoint. This enables the search for keV sterile neutrinos with  $m_s \lesssim 18$  keV at a sensitivity of  $|U_{e4}|^2 = \sin^2 \theta \sim 10^{-6}$  for the mixing angle [145], exceeding current laboratory limits by roughly two orders of magnitude (see section 2.6.2). A novel detector system, the TRISTAN silicon drift detector [8, 145], was developed specifically for this endeavor. It features improved high-rate capabilities and better energy resolution than the present Focal Plane Detector (FPD). This allows for rapidly acquiring the necessary statistics and better identifying the sterile neutrino signature.

This chapter begins with a theoretical description of the differential energy spectrum of  $\beta$ -electrons from molecular tritium decay in section 3.1. An account of KATRIN's experimental setup is provided in section 3.2. Aspects of the integral measurement mode and neutrino mass determination are covered in section 3.3, and details concerning the search for keV sterile neutrinos with TRISTAN are covered in section 3.4.

### 3.1 Tritium $\beta$ -Decay

Tritium  $\beta$ -decay is particularly well-suited for direct neutrino mass measurements for several reasons: [141]

- The tritium spectrum has a comparatively low kinematic endpoint. This emphasizes the signature of neutrino masses (sub-eV scale) and allows for energy determination with electrostatic retardation.
- Tritium decay is super-allowed. Thus, it exhibits a low half-life of 12.33 y, permitting the construction of a source with a large specific activity. Also, the decay matrix elements of super-allowed decays are energy independent, and the phase space factor exclusively determines the spectral shape.
- Tritium is bound in the simple molecule  $T_2$ , whose final state distribution for rovibrational excitations can be calculated.

Understanding the signature of neutrino masses in  $\beta$ -decay requires knowledge about the spectral shape, which is well described by Fermi's effective theory of weak decay [6]. However, it must also be considered that the decaying neutron is part of an atomic nucleus bound in  $T_2$  molecules. Hence, atomic and molecular corrections must be applied on top of the nuclear spectrum, which is relevant to the model described in chapter 4.

#### Nuclear $\beta$ -Spectrum

Tritium decays to  ${}^3\text{He}$  under the emission of an electron and electron antineutrino [221]

$$T \rightarrow {}^3\text{He}^+ + e^- + \bar{\nu}_e + Q(T) , \quad (3.1)$$

The mass difference between the mother and daughter nucleus is released as the surplus energy  $Q(T)$ . This energy is divided between the three decay products as the kinetic energy  $E$  of the electron, the total neutrino energy, including its mass, and the nuclear recoil energy  $E_{\text{rec}}$ . The kinematic endpoint of the electron energy spectrum is then

$$E_0 = Q - E_{\text{rec}} - m_i , \quad (3.2)$$

where  $m_i$  refers to any of the neutrino masses. Since  ${}^3\text{He}$  is much heavier than the electron and the neutrino, the recoil energy is comparatively small and reaches a maximum of 1.72 eV at the endpoint [221]. The shape of the nuclear  $\beta$ -electron energy spectrum can be calculated using Fermi's golden rule [141]:

$$\left(\frac{d\Gamma}{dE}\right)_{\text{nuc}} = \frac{G_F^2 |V_{\text{ud}}|^2}{2\pi^3} |M_{\text{nuc}}|^2 \cdot F(Z, E) \cdot E_e p_e E_\nu \sum_i |U_{ei}|^2 \underbrace{\sqrt{E_\nu^2 - m_i^2}}_{p_{\nu,i}} . \quad (3.3)$$

Here,  $G_F$  is the Fermi constant,  $V_{\text{ud}}$  the CKM-Matrix element, and  $M_{\text{nuc}}$  is the energy-independent nuclear matrix element. The Fermi function  $F(Z, E)$  accounts for the attractive electrostatic interaction between the electron and daughter nucleus. The remaining

elements constitute the phase space factor, where  $p_{e/\nu}$  and  $E_{e/\nu}$  are the momenta and total energies of neutrino and electron, respectively. Expressed explicitly with the kinetic electron energy  $E$ , these are  $E_e = E + m_e$ ,  $p_e^2 = (E + m_e)^2 - m_e^2$ , and  $E_\nu = E_0 - E$ .

Equation (3.3) reveals that the spectrum is a superposition of spectra regarding all neutrino masses, where the weights are the squared PMNS-Matrix elements (= mixing amplitudes). Since the mass splittings belonging to the active neutrinos are much smaller than the experimental energy resolution of  $\sim 1$  eV, the spectral signatures of individual  $m_i$  cannot be resolved. Therefore, the spectral shape is simplified to that of a single neutrino with an effective squared mass [141, 222]

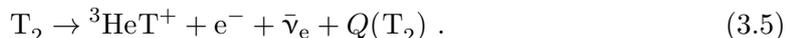
$$m_{\bar{\nu}_e}^2 = \sum_i |U_{ei}^2| m_i^2 . \quad (3.4)$$

The impact of the neutrino mass on the spectral shape is demonstrated in figure 2.7 in section 2.5.3. In contrast, heavier mass states from sterile neutrinos on the eV- to keV-scale can be resolved and lead to a distinct signature (covered in section 3.4.1).

### Atomic and molecular corrections

For a precise description of the tritium spectrum, several multiplicative higher-order correction factors must be applied to consider the atomic structure [221]. These account for radiative effects from real and virtual photons, screening of the nuclear coulomb field from shell electrons, energy-dependent recoil, the finite structure of the nucleus affecting the coulomb field, the motion of the coulomb field from recoil, and interactions with shell electrons. A calculation of the factors is provided by D.H. Wilkinson [223]. The combined effect of all corrections is on the order of 1 %, as demonstrated in figure 3.1.

Furthermore, tritium is typically bound in  $T_2$  molecules und thus the reaction equation (3.1) is altered to:

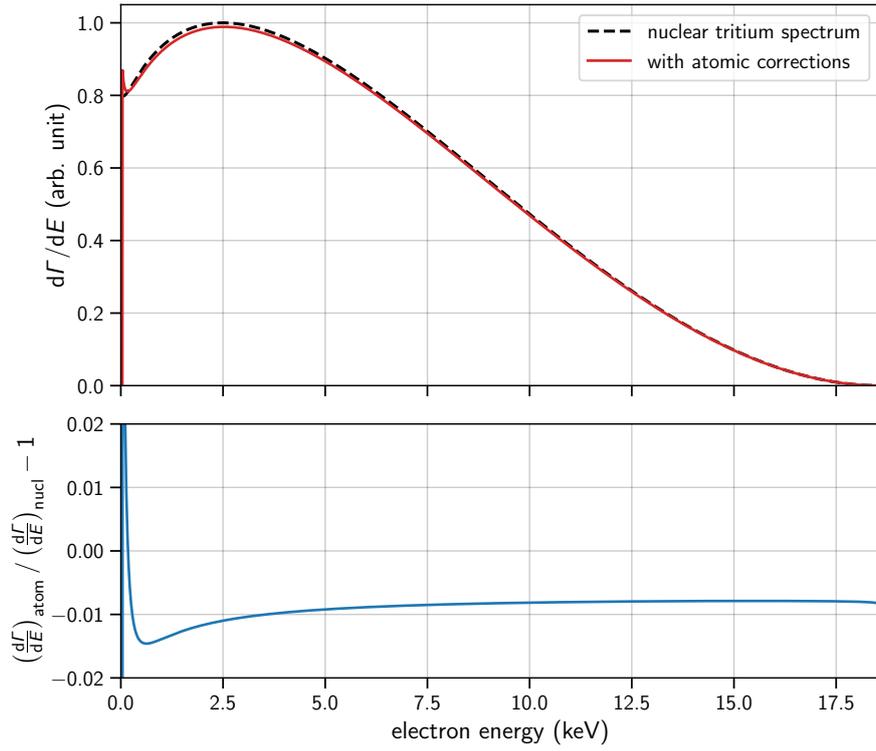


This affects the spectral shape in two ways: Firstly, the  $Q$ -Value is decreased by several eV because of the binding energy of mother and daughter molecules [141]. Secondly, the recoil can excite the daughter molecule to rotational and vibrational states or cause excitations to its electron shell, including ionization. The so-called *final state distribution* (FSD) encodes the excitation probability  $P_f$  as a function of the excitation energy  $V_f$  (see also [222]). The effect on the spectral shape can be computed by correcting the neutrino energy

$$E_\nu \rightarrow E_\nu = E_0 - E - V_f , \quad (3.6)$$

and summing the spectrum for all final states, weighing by  $P_f$ .<sup>1</sup> Lastly, it has to be considered that KATRIN's source gas also contains trace amounts of the other tritiated hydrogen isotopologues HT and DT. This affects the  $Q$ -Value and their FSD. [221]

<sup>1</sup>This is essentially a convolution of the atomic spectrum with the FSD.



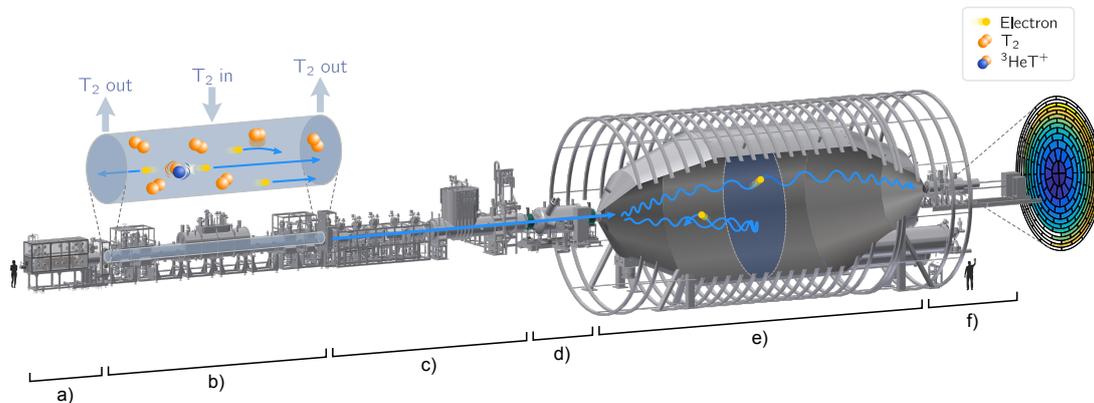
**Figure 3.1:** Tritium spectrum with and without atomic corrections. The correction factors are calculated according to [223] and implemented in the model described in chapter 4.

### Final state distribution for keV sterile search

The FSD is generally energy-dependent since the recoil changes with  $E$ . This can be neglected for the endpoint measurement since the recoil energy is approximately constant [221], which also enables the use of the sudden approximation in the FSD computation [224]. However, energy dependence must be considered for the keV sterile neutrino search, where the spectrum is measured to energies deep below the endpoint [225]. Recently, the influence of the fractional recoil momentum on the FSD has been investigated [226], and it was demonstrated that the width and the central energy of final states behave as smooth functions of energy. Further calculations concerning the electronic continuum of the energy-dependent FSD and its inclusion in the differential spectrum model (chapter 4) are ongoing.

### 3.2 Experimental Setup

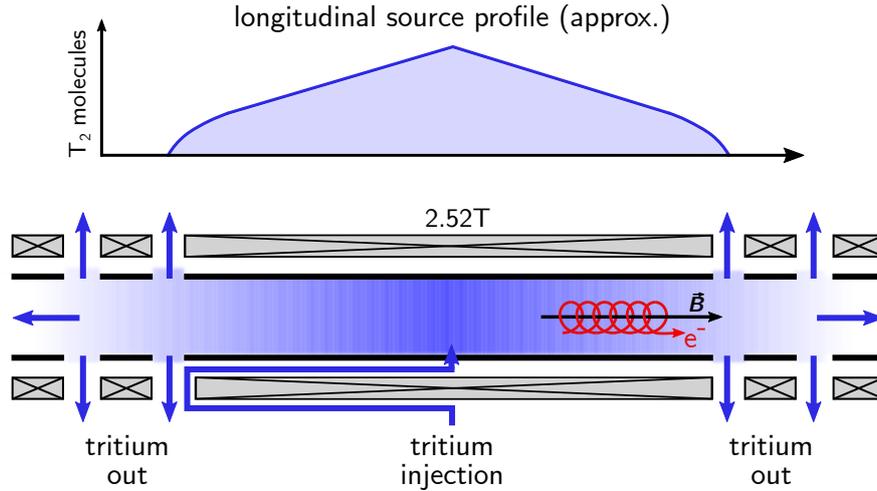
The KATRIN experiment [4, 5] has a 70 m long beamline, and it is subdivided into the Source and Transport Section (STS) and the Spectrometer and Detector Section (SDS). Each consists of several subcomponents, as shown in figure 3.2.



**Figure 3.2:** The KATRIN beamline and its components: (a) Rear Section, (b) Source Section, (c) Transport Section, (d) Pre-Spectrometer, (e) Main Spectrometer, (f) Detector Section. Components (a-c) comprise the STS and (d-f) the SDS. Figure kindly provided by L. Köllenberger.

The task of the STS is to provide a pure  $\beta$ -electron flux of up to  $10^{11}$  Bq with minimal distortion of the energy spectrum from systematic effects such as scattering. This is achieved with a Windowless Gaseous Tritium Source (WGTS), where molecular tritium is circulated by injecting the gas in the center and pumping it away at the front and rear. The  $\beta$ -electrons are guided downstream along the beamline by strong magnetic fields and leave the source without the energy loss a physical exit window would cause. The radioactive gas has to be retained in the closed loop of the Source Section and prevented from entering the spectrometer, where it would lead to background. This is achieved with two more pumping sections, which comprise the Transport Section: The Differential Pumping Section (DPS), which uses turbomolecular pumps, and the Cryogenic Pumping Section (CPS), which adsorbs the remaining tritium by cryo-sorption. On the other side of the source section is the Rear Section (RS), which houses monitoring and calibration tools and terminates the upstream end of the beamline.

In the SDS, the  $\beta$ -electron energy is analyzed with the Main Spectrometer (MS), which uses Magnetic Adiabatic Collimation and Electrostatic Retardation (MAC-E principle) [227, 228] to repel electrons whose energy is below the retarding potential  $qU_{\text{ret}}$ . Electrons with sufficient energy pass the potential barrier and are transported to the Detector Section, where they are counted by the Focal Plane Detector (FPD). The integral energy spectrum is measured by scanning  $qU_{\text{ret}}$  through a range of setpoints. The Pre-Spectrometer (PS) is a second spectrometer located upstream of the MS that can limit the flux of low-energetic  $\beta$ -electrons entering the SDS. In the following, several details concerning the subcomponents are provided.



**Figure 3.3:** Schematic drawing of the WGTS (based on a figure from [4]).

### 3.2.1 Source Section

The central piece of the source section is the ultra-luminous Windowless Gaseous Tritium Source (WGTS), which provides the flux of  $10^{11}$   $\beta$ -electrons per second. It is essentially a 10 m long steel tube with a diameter of 90 mm, filled with molecular tritium gas. Inside the tube, tritium decays, and the  $\beta$ -electrons are transported adiabatically along the beamline by a strong magnetic field. As the name implies, the WGTS has no physical windows at the ends, which would cause significant energy loss for the electrons.

The gas is continuously injected, at a constant inlet pressure of  $3 \times 10^{-3}$  mbar, through capillaries in the length-wise center and differentially pumped away at the ends, reducing the tritium density at these positions by a factor of about  $10^3$  [229]. This results in a stable longitudinal density profile, with a nominal column density of  $\rho d = 5.0 \times 10^{17}$  molecules/cm<sup>2</sup>. Since source stability is vital for the integral spectrum measurement, the WGTS is designed for fluctuations of  $< 0.1\%$ /h, which it exceeds by one order of magnitude [230]. The gas injection is fed by the inner tritium loop, which collects and purifies the tritium from the pump ports and replenishes the gas from a supply to keep a stable tritium purity of  $\varepsilon_T > 95\%$ . The purity is monitored with Laser Raman Spectroscopy (LARA), which can detect the relative amounts of each isotopologue.

The WGTS is enclosed by a cryostat, which keeps the beam tube at a constant temperature of 30 K in the standard configuration. There is a stringent requirement for temperature stability due to its impact on the column density. The temperature can be increased to 80 K, which improves source plasma conditions and enables the injection of gaseous  $^{83m}\text{Kr}$ , which is used for calibration purposes [231]. The cryostat also houses the liquid helium-cooled superconducting solenoid magnets [232] that create a homogeneous and stable guiding field for the electrons. The field is operated at a strength of  $B_{\text{src}} = 2.52$  T.

### 3.2.2 Rear Section

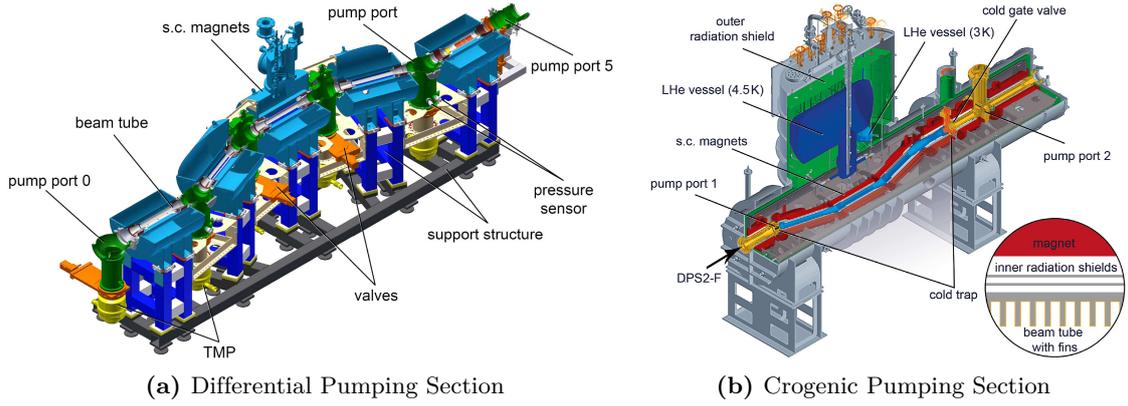
Due to the creation of charge carriers in the WGTS ( $\beta$ -electrons, secondary electrons from scattering, ions from decay and scattering), the source develops a cold plasma with an electric plasma potential on the mV scale [233]. This potential has to be regulated since an undefined or spatially inhomogeneous starting potential for the  $\beta$ -electrons would interfere with the energy measurement. Therefore, the magnetic field lines from the upstream end of the WGTS are mapped onto the Rear Wall (RW), a gold-plated steel disk with a diameter of 145 mm that is electrically connected to the WGTS beam tube with a controllable potential offset. Charged particles created in the source are transported to the RW, and homogeneous plasma conditions can be achieved by appropriately adjusting the RW potential under consideration of the WGTS and RW work functions. The RW is mounted inside a dedicated chamber which can be separated from the source with a gate valve. The chamber houses further monitoring devices, such as the BIXS system, which uses  $\beta$ -induced X-ray spectroscopy to watch the source activity. The RW can be illuminated with UV light through a UHV-proof quartz window. This can be used to clean the rear wall from residual tritium (via ozone production) [234] and to provide the source plasma with photoelectrons.

Behind the RW is a high-resolution angular selective electron gun (e-gun) whose beam is transmitted to the beamline through a 5 mm hole in the RW. The electrons are created photoelectrically with UV light from a stable Laser-Driven Light Source (LDLS) and accelerated with electric fields. The magnetic field guides the monoenergetic electrons through the KATRIN beamline, and the FPD detects them. As the beam connects the entire beamline, it is used for characterizing the electromagnetic field, especially concerning the transmission function of the Main Spectrometer. Furthermore, the e-gun beam is used for measuring the column density via the probability of scattering on source gas molecules, which causes an energy loss for the electrons.

### 3.2.3 Transport Section

The WGTS gas flow has to be reduced by 14 orders of magnitude to prevent tritium from entering the Spectrometer Section, where its decay would cause background events. This is achieved with differential and cryogenic pumping along the Transport Section, through which the  $\beta$ -electrons are adiabatically guided by magnetic fields of up to 4 T. The Differential Pumping Section (DPS) is located directly downstream of the WGTS and consists of five superconducting magnets. These are interspersed by pump ports equipped with turbomolecular pumps. The magnets and beam tube elements are tilted by  $20^\circ$  with respect to each other, which creates a chicane (see figure 3.4a). Electrons are transported through the corners by the guiding field, while the direct line-of-sight for neutral gas molecules is blocked, thus increasing their likelihood of hitting a wall and getting pumped away. The same principle is also used in the CPS.

Overall, the DPS reduces the gas flow by at least seven orders of magnitude, beyond which turbomolecular pumps are no longer efficient due to back diffusion. Therefore, the remaining tritium is pumped via cryo-sorption in the Cryogenic Pumping Section (CPS)



**Figure 3.4:** CAD drawings of the pumping sections showing, among other elements, the arrangement of beam tube elements, turbomolecular pumps (TMPs), and superconducting (s.c.) magnets. Figures adapted from [235].

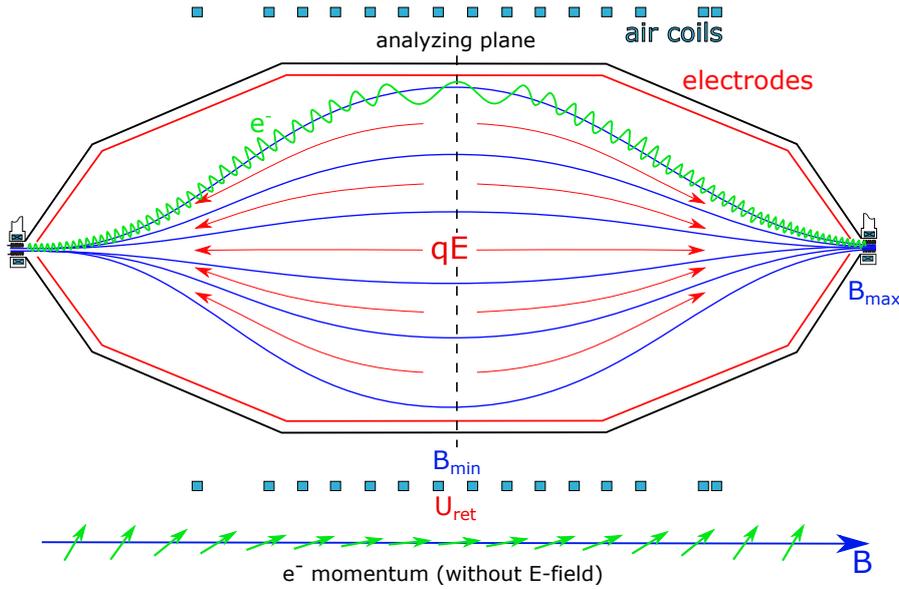
(see figure 3.4b). The beam tube elements of the CPS are cooled to 3 K with liquid helium. Gas molecules hitting the wall are therefore likely to adhere to the cold surface. Small gold-plated fins cover the beam tube to increase the surface area. Furthermore, a layer of argon snow is condensed on the gold plating, which increases the sticking coefficient. The argon layer and adsorbed tritium can be removed by heating to regenerate the system.

Apart from neutral gas, the source also emits a flux of positive tritium ions (mostly  $T_3^+$ ). Since the magnetic field guides these, they have to be filtered with electrodes positioned along the Transport Section. First, they are blocked with positive electrostatic potential barriers from ring electrodes with voltages of 5 V to 200 V. To prevent the accumulation of large amounts of stored ions, they are guided to the beam tube wall via  $E \times B$  drift in the electric field created by dipole electrodes. These are operated with offset negative voltages. Further dipole electrodes on positive voltage are used to remove low-energetic secondary electrons.

### 3.2.4 Spectrometer Section

The Spectrometer Section consists of the Pre-Spectrometer (PS) and the Main Spectrometer (MS), where the MS is used for  $\beta$ -scanning. Both spectrometers employ the principle of Magnetic Adiabatic Collimation with Electrostatic filtering (MAC-E) [227, 228] to let only electrons with energy above a specific threshold pass. The Spectrometer Section is operated at an ultra-high vacuum (UHV), which serves two purposes: Firstly,  $\beta$ -electrons have to be able to pass the spectrometers unhindered without scattering on gas molecules, which would constitute an energy loss. Secondly, a good vacuum limits the production of background events in the MS volume. Especially the second reason leads to the stringent pressure requirement of  $< 10^{-11}$  mbar. This pressure is achieved with a combination of cascaded turbomolecular pumps and non-evaporable getter pumps [5].

The general concept of the MAC-E filter is demonstrated in figure 3.5.  $\beta$ -electrons are



**Figure 3.5:** Working principle of the MAC-E filter. The  $\beta$ -electrons are guided by the magnetic field in cyclotron motion, and their momentum is gradually aligned with the field lines through adiabatic collimation. Their pitch angle is minimal in the analysis plane, where the electric retarding potential is the largest, and only electrons with sufficient parallel momentum can pass. Illustration based on figures from [4, 236].

emitted isotropically in the source in a magnetic field  $B_{\text{src}} = 2.52 \text{ T}$  and they are guided to the entrance of the spectrometer from the left. Their energy can be selected with an electric retardation field, which is aligned with the guiding magnetic field lines. However, simply applying a retarding potential would not suffice since the electric field will only affect the momentum's parallel component  $p_{\parallel}$  and leave the transverse component  $p_{\perp}$  unaffected. Therefore, the electron momenta must be aligned with the field without affecting the total energy. This is achieved with magnetic adiabatic collimation, which is the effect that the pitch angle  $\theta = \angle(\vec{p}, \vec{B})$  decreases when the electron transitions from a high magnetic field region to one with a low magnetic field due to conservation of the magnetic moment [237]

$$\frac{p_{\perp}^2}{B} = \text{const.} \quad (3.7)$$

The optimal energy resolution of the filter is achieved by aligning the maximum of the retarding potential  $qU_{\text{ret}}$  with the minimum magnetic field  $B_{\text{min}} = B_{\text{ana}} = 0.6 \text{ mT}$  in the so-called *analyzing plane*. Additionally, electrons with large initial pitch angles, for which the energy selection is the least precise, are rejected using the magnetic mirror effect (see section 5.3 for an explanation). The relation between the starting magnetic field at the source and the maximal magnetic field  $B_{\text{max}} = B_{\text{pch}} = 4.2 \text{ T}$  of the beamline at the

so-called pinch magnet (located downstream of the MS) defines the *acceptance angle*

$$\theta_{\max} = \arcsin \left( \sqrt{\frac{B_{\text{src}}}{B_{\max}}} \right). \quad (3.8)$$

Electrons with pitch angles below the acceptance angle are transmitted to the detector, while the others are reflected back. Altogether the filter width is then given by [141]

$$\frac{\Delta E}{E} = \frac{B_{\min}}{B_{\max}}. \quad (3.9)$$

A detailed explanation of magnetic collimation, magnetic reflection, and the transmission function of the MAC-E filter is provided in section 5.3.

Equation (3.9) shows that a small magnetic field in the analyzing plane is desirable. However, the conservation of magnetic flux

$$\Phi = \int \vec{B} \cdot d\vec{A}. \quad (3.10)$$

The so-called *flux tube* refers to the cross-section of the flux at any point of the beamline, whose field lines map to the detector. Equation (3.10) requires that the area of the flux tube increases when the magnetic field decreases. With equation (3.9), the maximal energy resolution is thus related to the diameter of the spectrometer. With its radius of  $\approx 5$  m, the MS can achieve an energy resolution of  $\Delta E = 0.93$  eV using the design field values [4]. To use this resolution, the retarding potential provided by KATRIN's high-voltage system must be stable and precise on the ppm level [5], which is achieved with resistive high-voltage dividers [238].

### Pre-Spectrometer and Monitor Spectrometer

The smaller PS served as a prototype for the MS and can be used in tandem to limit the flux of electrons before they enter the MS. However, it was realized that  $\beta$ -electrons can be trapped in the potential valley between the spectrometers  $\beta$ -electrons and accumulate over time, which causes additional spectrometer background [5]. Therefore, the PS is now operated at ground potential during  $\beta$ -scans [239].

A third spectrometer, the Monitor Spectrometer (MoS), is located in a separate building next to the KATRIN beamline. It is the spectrometer inherited from the predecessor experiment in Mainz [142], and it can be used to perform long-term stability tests of KATRIN's high-voltage system by periodically scanning the  $L_3$  line of a condensed  $^{83m}\text{Kr}$  source. Currently, the MoS is primarily utilized as a test bench for the integration of new hardware, such as the modules of the new TRISTAN detector (see section 3.4.4) or the e-gun.

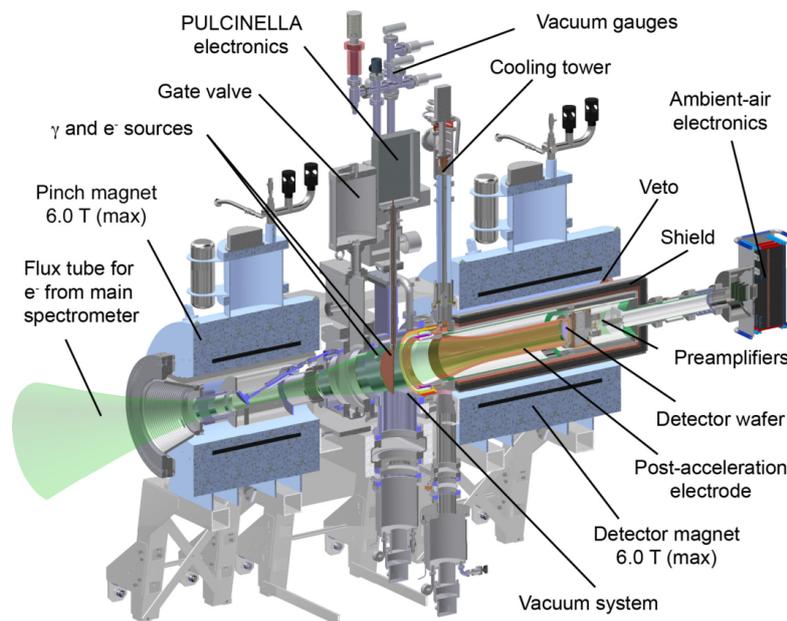
### 3.2.5 Detector Section

$\beta$ -electrons transmitted through the MS are guided to the Detector Section (see figure 3.6) and registered by the Focal Plane Detector (FPD). Since the Detector section is adjacent to the MS, a UHV with a pressure of  $< 10^{-9}$  mbar has to be maintained.

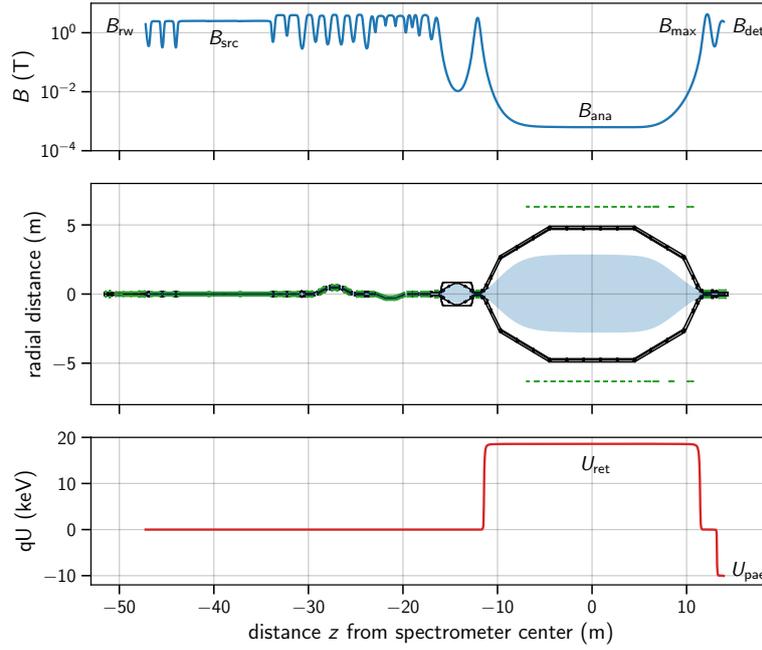
The FPD is a monolithic silicon  $p-i-n$ -diode array with 144 pixels on a diameter of 125 mm. Here, the incident electrons deposit their energy via multiple scattering, which creates a charge signal with an amplitude proportional to the energy. The signal is then amplified, fed out of the vacuum chamber, and read out with a custom data acquisition system (DAQ). The signal pulse height and event timing are extracted with three stages of consecutive digital trapezoidal filters operated with peaking times in the  $1.6\ \mu\text{s}$  to  $6.4\ \mu\text{s}$  range. The total rate limit for the FPD is approximately 1 Million counts per second (cps), and the energy resolution is  $\approx 2\ \text{keV}$  at full-width half maximum (FWHM).

Before electrons reach the detector, they get an energy boost of 10 keV by the Post Acceleration Electrode (PAE). This reduces detector systematics such as backscattering and energy loss in the dead layer (see section 5.4), and it shifts the signals into an energy region with fewer background events. The detector wafer is located inside the bore of a superconducting magnet and operated at a magnetic field of  $B_{\text{det}} = 2.52\ \text{T}$ . This matches the source magnetic field such that there is a complete mapping of the flux tube. The detector is surrounded by passive shielding and a muon veto to reduce background.

Next to FPD and PAE, the Detector Section also provides calibration sources. One is a  $^{241}\text{Am}$   $\gamma$  source that can be moved in front of the detector through a tube reaching into the UHV system. The other is a photoelectric electron source consisting of a titanium disk that can be positioned in front of the detector and illuminated with UV LEDs. Offsetting the electric potential of the disk allows for the generation of a large-area electron beam with energies in the range 0 keV to 25 keV. [5]



**Figure 3.6:** CAD drawing showing the primary components of the Detector Section. Figure adapted from [240].



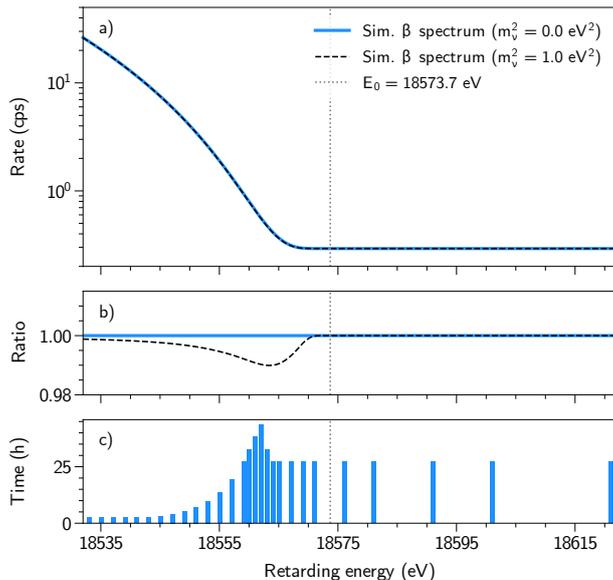
**Figure 3.7:** Summary of KATRIN’s electromagnetic field setup. The center plot shows a cross-section of the beamline geometry from above, where beam tube elements and electrodes are drawn in black and magnets in green. The flux tube mapping to the detector is shown in blue.

### 3.3 Neutrino Mass Analysis

The squared effective neutrino mass  $m_{\nu_e}^2$  and the endpoint energy  $E_0$  are extracted from the measured data via a spectral fit with a model prediction. A visualization of the integral spectrum and the signature of the neutrino mass are shown in figure 3.8. The integral spectrum model is calculated by convolving the differential tritium spectrum  $\frac{d\Gamma}{dE}$  (see section 3.1), with the instrumental response function  $R(E - qU_{\text{ret}})$ : [222]

$$\dot{N}(qU_{\text{ret}}; E_0, m_{\nu_e}^2) = A_s \cdot N_{T,\text{eff}} \int_{qU}^{E_0} \frac{d\Gamma}{dE} \cdot R(E - qU_{\text{ret}}) dE + R_{\text{bg}}. \quad (3.11)$$

The normalization factor  $N_{T,\text{eff}}$  is the effective number of tritium atoms in the source, corrected for acceptance and detection efficiency. The Signal amplitude  $A_s$  and the background rate  $R_{\text{bg}}$  are treated as free nuisance parameters. Besides these, the response function and differential spectrum depend on several systematic parameters [222]. The integration bounds can be set to the retarding energy  $qU_{\text{ret}}$  and  $E_0$  since either the response function or differential spectrum will be equal to zero otherwise. By doing so, the equation indicates that only electrons with sufficient surplus energy can pass the main spectrometer. The response function represents the probability for an electron of a given surplus energy  $E - qU_{\text{ret}}$  to be able to traverse through the entire beamline. While this is primarily determined by the transmission function (see section 5.3.3), the energy



**Figure 3.8:** Integral spectrum model for the KATRIN neutrino mass analysis. The ratio plot (b) shows the fraction of the spectra without and with neutrino mass. The bottom plot (c) shows the measurement time distribution that encodes the total time spent at each retarding potential setpoint. Figure adapted from [5].

loss from scattering on source gas molecules (see section 5.1.1) also plays a major role. See [221, 222] for details concerning the modeling of the response function.

The model can be fit to the data by optimizing a Poisson likelihood function. In the case of a fit result compatible with  $m_{\bar{\nu}_e}^2 = 0$ , the obtained  $m_{\bar{\nu}_e}^2$  value can be translated to an upper limit on  $m_{\bar{\nu}_e}$  via the Feldman-Cousins or the Lokhov-Tkachov construction [222, 241]. Using data from the first two measurement phases, KATRIN currently sets the world-leading limit on the effective mass [7]:

$$m_{\bar{\nu}_e} < 0.8 \text{ eV} \quad (90 \% \text{ CL}) \quad (3.12)$$

The sensitivity goal for the final dataset of 1000 live measurement days is 0.2 eV [5]. This would improve upon the limit  $m_{\bar{\nu}_e} \lesssim 2 \text{ eV}$  set by the Mainz [142] and Troitsk [143] experiments by one order of magnitude.

The neutrino mass measurement data can also be used for sterile neutrino searches on the eV mass scale. So far, data from the first two measurement campaigns has been analyzed, and limits on the mass and mixing amplitude are published in [170, 178]. Further information on this topic is provided in section 2.6.1.

A first search for sterile neutrinos on the keV scale was performed with data acquired during the first tritium (FT) campaign [216, 242]. In this phase, KATRIN was operated at 0.5 % of the nominal source luminosity, which allowed for a measurement of the integral tritium spectrum to 1.6 keV below the endpoint, and thus provided sensitivity towards sterile neutrino masses with  $m_s \lesssim 1.6 \text{ keV}$  (see section 2.6.2).

### 3.4 keV Sterile Neutrino Search with TRISTAN

The TRISTAN project (**TR**itium **I**vestigation on **ST**erile to **A**ctive **N**eutrino **M**ixing) of the KATRIN collaboration aims to measure the differential tritium  $\beta$ -spectrum to energies deep below the endpoint. The primary physics motivation is to search for the signature of sterile neutrinos on the keV mass scale and probe the sterile mixing amplitude with a sensitivity of  $\sin^2\theta < 10^{-6}$ . This would improve upon existing laboratory limits by two orders of magnitude (see section 2.6.2).  $\beta$ -spectroscopy allows for a model-independent search based on kinematics, a unique characteristic compared to cosmological limits that are based on further assumptions (see section 2.6.2).

KATRIN's ultra-luminous WGTS is perfectly suited for this challenge since it enables the acquisition of a high statistics dataset. At full source strength,  $10^{18}$   $\beta$ -electrons could be measured within three years of data-taking. This corresponds to a statistical sensitivity on the sterile mixing amplitude of  $\sin^2\theta \sim 10^{-8}$ . However, this would require a detection rate of  $10^{10}$  counts per second (cps), which is four orders of magnitude above the capabilities of the KATRIN FPD (see section 3.2.5). Using existing hardware, a statistical sensitivity of  $\sin^2\theta \sim 10^{-4}$  can be achieved. [238]

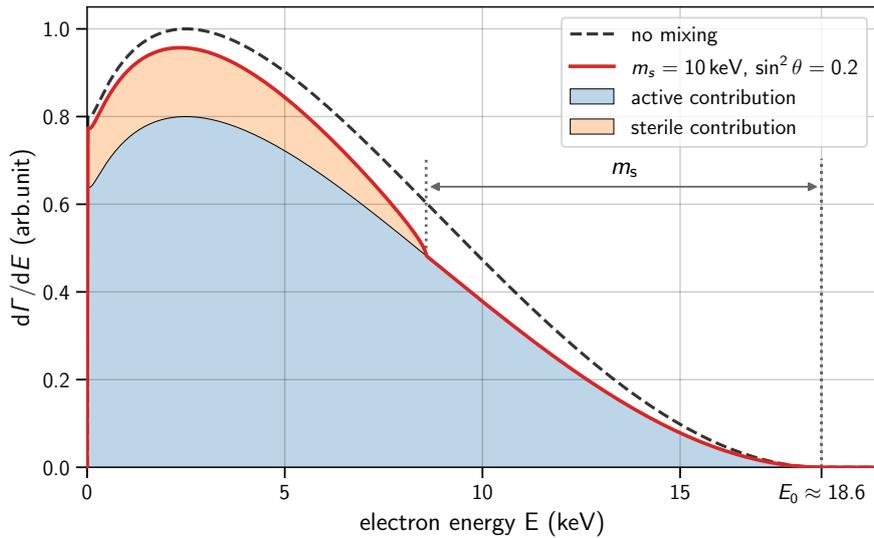
The novel TRISTAN silicon drift detector (SDD) [8, 145], developed by the KATRIN collaboration, has been designed with the ability to handle an event rate of  $10^8$  cps. Thus, up to  $10^{16}$   $\beta$ -electrons could be collected in one year, which would be sufficient to probe  $\sin^2\theta < 10^{-6}$  [145]. The SDD also has a much better energy resolution ( $< 300$  eV FWHM for electrons at 20 keV [238]) compared to KATRIN's existing FPD, which is a p-i-n diode type detector offering  $\approx 2$  keV energy resolution [5]. Consequently, the upgrade will permit a clearer identification of the sterile neutrino's characteristic spectral kink signature (see next section). Details concerning the detector's technology and the development process are given in section 3.4.4. The integration of the TRISTAN detector into the KATRIN beamline is planned to begin in 2026 after the neutrino mass measurement is completed.

#### 3.4.1 Signature in Tritium $\beta$ -Decay

As discussed in section 3.1, the  $\beta$ -spectrum can be interpreted as a superposition of spectra from all mass states that mix with the electron neutrino flavor via the PMNS matrix (equation (3.3)). While the mass states corresponding to the active flavors are small and cannot be distinguished in the spectrum, the kinematic endpoint of the contribution from a keV-scale sterile neutrino would be well separated. The differential decay rate can thus be written as [145]

$$\frac{d\Gamma}{dE} = \cos^2\theta \frac{d\Gamma}{dE}(m_{\bar{\nu}_e}) + \sin^2\theta \frac{d\Gamma}{dE}(m_s), \quad (3.13)$$

where  $m_{\bar{\nu}_e}$  is the effective mass of the light eigenstates,  $m_s = m_4$  is the mass of the keV-scale sterile neutrino, and  $\sin^2\theta = |U_{e4}|^2$  is the sterile mixing amplitude concerning the electron flavor. The spectrum is visualized in figure 3.9. It shows that the sterile neutrino adds a spectral contribution of a distinct shape, which also causes a characteristic kink to appear at the energy  $E_0 - m_s$ .



**Figure 3.9:** Signature of a keV sterile neutrino in tritium  $\beta$ -decay, with an exaggerated mixing amplitude. The endpoint of the spectrum contribution belonging to the heavy mass state is well separated from that of the small masses, which generates a characteristic kink signature.

### 3.4.2 Deep Spectrum Measurement

The energy range of the measurement determines the accessible mass range for the sterile neutrino measurement. The accessible mass range is ultimately limited to  $m_s < E_0 = 18.6$  keV in tritium  $\beta$ -decay. This means that the retarding potential in KATRIN's main spectrometer must be decreased as much as possible for the goal of broadening the range. While this is possible in general, systematic effects in the low energy range must be considered when choosing the value for  $qU_{\text{ret}}$ . Currently, there is no definitive specification concerning the retarding potential setpoint for the deep spectrum measurement, but a value in the range 100 eV to 500 eV could be acceptable from the systematics point-of-view. This is further discussed in section 4.3.

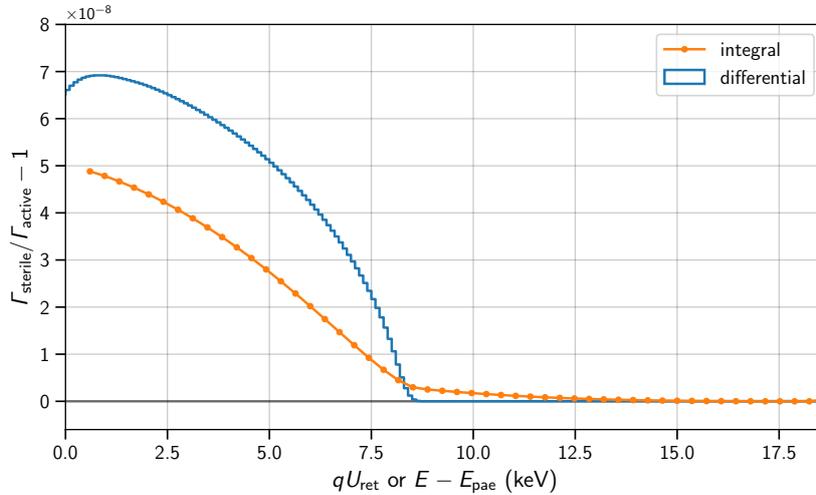
Another constraint stems from the requirement of fully adiabatic transport from the source to the detector. In a previous simulation study [243] it was shown that for high surplus energies, i. e. at low retarding potentials, there is a significant chance for electrons to be guided non-adiabatically in the low magnetic field of the main spectrometer. This can cause unwanted, chaotic reflections and trapping of incoming electrons. Thus, the effect has an unpredictable influence on the energy spectrum. One way to decrease the risk of non-adiabaticity is to choose a larger retarding potential, which would not be desirable as it would limit the accessible mass range. However, the study also shows that fully adiabatic transport can be achieved by increasing the field in the main spectrometer or by decreasing the detector magnetic field. Investigations concerning the magnetic field requirements and measurements demonstrating the effect of non-adiabaticity are currently ongoing [244].

### 3.4.3 Beamline Operation Modes

In a previous study [225] it was shown that a measurement of the differential tritium spectrum would yield much better sensitivity than KATRIN's standard integral measurement mode (about one order of magnitude). This work confirms this, based on a more advanced model of the experimental response (see chapter 4 and section 6.3.2). A differential measurement can be achieved at KATRIN by using the detector instead of the main spectrometer to estimate the electron energy.

There are two main reasons for the improved sensitivity in differential mode over the integral mode: Firstly, in an integral spectrum scan, a significant amount of time is spent in the high-energy region where the rate is low. On the other hand, the rate is always constant in the differential mode and can be adjusted for optimal detector performance by tuning the source column density. Secondly, the sterile neutrino signature is much more distinct in the differential spectrum, as demonstrated in figure 3.10.

Another potential mode of operation is to perform a semi-differential spectrum measurement using Time-Of-Flight (TOF) spectroscopy [245]. This method exploits that electrons with energies just above the retarding energy are significantly slower than those with more surplus energy. The method was simulated in a study by Steinbrink et al. [246], which showed that statistical sensitivity surpassing that of the integral mode can be achieved. However, the technical realization is still a work in progress. Ultimately, all operation modes can be used for cross-checking each other, since each is sensitive to different systematic effects (see also 4.2 in chapter 4). The TOF method could be particularly useful for probing specific sub-regions of the accessible mass range [225].



**Figure 3.10:** Comparison of the sterile neutrino signature in integral and differential modes of operation. Shown is the spectrum with a sterile neutrino ( $m_s = 10$  keV,  $\sin^2 \theta = 10^{-7}$ ) divided by the expectation for the spectral contribution of the active branch. The model introduced in chapter 4 (TRModel) is used for this calculation, and spectral distortions from systematic effects are enabled.

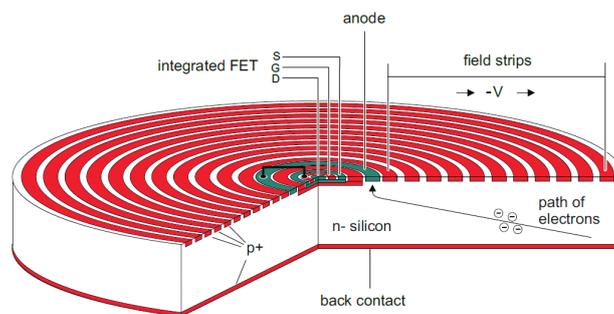
### 3.4.4 TRISTAN Detector

The TRISTAN detector has to fulfill several requirements: It has to have good energy resolution to retain details of the spectral shape, it must be capable of handling rates up to  $10^8$  cps, the entrance window should be about 50 nm thin to minimize energy loss, and it has to be able to cover a large area with a diameter of  $\approx 20$  cm to fully map the flux tube in KATRIN's focal plane. These requirements led to the Silicon Drift Detector (SDD) [247] as the technology of choice for TRISTAN. [8]

#### Silicon Drift Detector Array

Silicon detectors use reverse biasing of a pn-junction to create a large depletion zone. Energy depositions from incident ionizing particles<sup>2</sup> create electron-hole pairs that are drifted to the contacts by the depletion field. This produces a measurable charge signal with an amplitude proportional to the primary particle's energy. The key idea behind the SDD is to drift the charges with an additional drift field to a point-like anode, which is much smaller than the detector pixel itself (see figure 3.11). This significantly reduces the contact capacitance, resulting in a larger signal amplitude and also a shorter rise time. Consequently, it is possible to manufacture relatively large pixels while maintaining optimal energy resolution close to the Fano limit (see section 5.4.3). Furthermore, the fast rise time allows for better event separation at high rates.

The pixel diameter of the TRISTAN SDD is chosen as 3 mm based on an optimization concerning charge-sharing, and detector backscattering and back reflection [248] (see also section 5.4). Each pixel features an integrated n-type field effect transistor (nJFET), which is part of the first amplification stage [145]. This is crucial since it allows the pre-amplifier to be at a distance of several cm, which enables the construction of a pixel array covering a large area. Design and production of the TRISTAN SDD chip are handled by the semiconductor laboratory of the Max-Planck-Society (MPG-HLL).



**Figure 3.11:** Schematic view of a silicon drift detector with integrated field effect transistor. The electrons are funneled to the central anode by the drift field, which is created by the field strips. Figure adapted from [8].

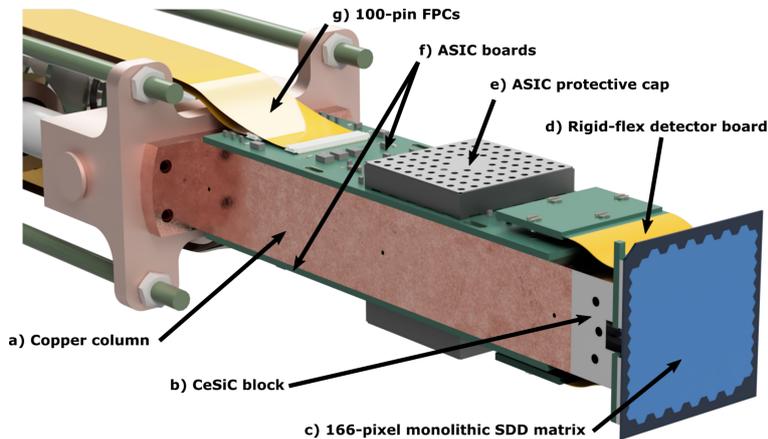
<sup>2</sup>Usually SDDs are employed for detecting X-rays, i. e. photons in the keV range.

### TRISTAN 166-pixel module

The maximal rate per pixel is  $\approx 10^5$  cps due to limitations of the data acquisition system from dead time and pileup (see section 5.5). Therefore, the incident electron flux has to be distributed over approximately 1000 pixels to meet the rate requirement. Meanwhile, the number of pixels per detector chip is limited by the complexity and cost [8]. This led to a modular detector design where each unit is equipped with a 166-pixel SDD array. The pixels have a hexagonal shape and cover a  $4\text{ cm} \times 4\text{ cm}$  area. Several modules are then combined to fill the area of the focal plane at KATRIN (see next paragraph).

Figure 3.12 shows a drawing of the TRISTAN module. A central design aspect is the non-planar geometry with a front-facing SDD chip, allowing for the side-by-side arrangement of modules since there are no obstructing circuit boards. The SDD is glued to the so-called CeSiC block, which is made from a silicon carbide ceramic. This block is mounted on a copper column that connects to a cooling structure. Both materials offer great thermal conductivity, which is required to cool the detector and front-end electronics to the operating temperature of about  $-40^\circ\text{C}$  for thermal noise reduction [238]. The CeSiC material is chosen as a thermal link to the copper since its thermal expansion coefficient is similar to that of silicon. Gluing the chip directly to the copper could damage the detector due to mechanical stresses when the temperature changes. [8]

The bond pads on the back of the chip are connected to the two front-end boards (ASIC boards) via a connector made from rigid PCBs with a flexible piece in between. The ASIC boards are each equipped with seven 12-channel ETTORE pre-amplifier ASICs [249]. This amplification circuit has two output stages: The first stage is a pulsed reset pre-amplification circuit that completes the feedback loop with the SDD JFET. The second stage further amplifies the signal to drive an analog-to-digital conversion (ADC) circuit that can be placed at a significant distance. The second stage adds a  $15\ \mu\text{s}$  exponential decay to the pulse shape (see also section 5.5).

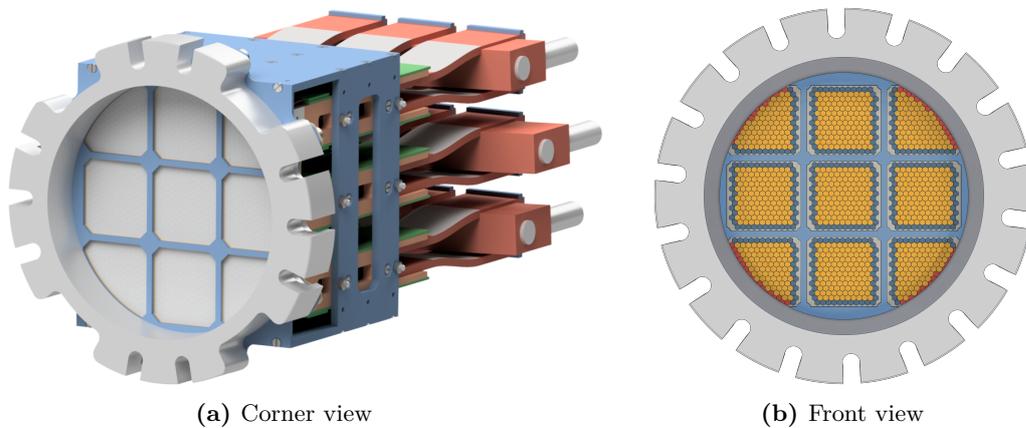


**Figure 3.12:** CAD drawing of the TRISTAN 166px-module with labels for the primary components. Figure adapted from [8].

### Beamline integration phases

The integration into the KATRIN beamline is structured into three phases: [8]

- **Phase-0:** One detector module is integrated at the KATRIN Monitor Spectrometer (MoS) for testing the performance in an environment with strong magnetic fields and UHV conditions. Phase-0 was started in November 2020 and is ongoing. In 2022, the first fully operational 166-pixel module was successfully integrated and tested with X-rays from a  $^{55}\text{Fe}$  source and electrons from the  $^{83m}\text{Kr}$  source located at the opposite side of the MoS [250].
- **Phase-1:** Nine modules are integrated into the KATRIN beamline to replace the existing focal plane detector. A schematic view of the mechanical arrangement of the modules is shown in figure 3.13. This design offers 1264 obstruction-free pixels, 936 of which are inner pixels (“golden pixels”) that should experience even illumination with electrons. The ambient air electronics are also replaced with a new DAQ system developed by the IPE institute at KIT [8]. Phase-1 already offers significant physics reach (see section 2.6.2), and it will start after the neutrino mass measurements are completed by the end of 2025. The sensitivity studies presented in this work are based on the Phase-1 specifications (see chapter 6).
- **Phase-2:** In the last stage, the number of modules is increased to 21 for a total of 3486 pixels, which allows for the acquisition of an even larger dataset. In this phase, the detector is positioned closely behind the detector magnet in a lower field of  $B = 0.8\text{ T}$ , which improves the backscattering and back-reflection systematics.



**Figure 3.13:** Mechanical design prototype of the TRISTAN Phase-1 detector. The modules are held in place by a steel structure (light blue) which includes a vacuum aperture covering the gaps between modules. The assembly is attached to a vacuum flange (gray) in the front. In the back, the modules are thermally coupled to a cooling structure. Between modules and the cooling structure must be sufficient space for cable routing from the modules to the electrical feed-throughs of the chamber. In the front view (b), the pixels are colored by groups of inner pixels (yellow), edge pixels (blue), and partially obstructed pixels (red).



---

## Deep Tritium Spectrum Model

---

Probing small mixing angles  $\sin^2 \theta < 10^{-6}$  with an optimized keV sterile neutrino search at KATRIN requires significant changes to the beamline hardware and its operation. However, the data analysis and sensitivity estimation procedure, which are based on a model prediction of the measured  $\beta$ -spectrum, also differ from the neutrino mass analysis and must be improved to reach ppm-level precision. Otherwise, a possible sterile signal is hidden beneath systematic uncertainties of the model description. Also, if the model is inaccurate, a spectral distortion arising from experimental effects may be mistaken as a sterile neutrino signature.

Several studies concerning model building and sensitivity estimation for TRISTAN were conducted in the past [225, 243, 248, 251–253]. These were usually individual studies where the impact of a subset of systematics is modeled, and their uncertainty propagated to calculate the corresponding loss in sensitivity. In these models, it is difficult to consider the interplay of effects or to find the optimal operating parameters, such as magnetic field settings for the beamline. A model describing all effects simultaneously with a complete description of the KATRIN beamline is mandatory. What was previously missing, in particular, were models for transitions between magnetic fields and electric potentials at different points along the beamline. Together these effects govern the electron propagation starting in the source, through the beamline until they are removed at the detector or the rear wall.

This chapter describes a novel model for the differential tritium spectrum at KATRIN. The code implementation of the new model is named TRModel [254] and provides a structure capable of holistically combining all relevant effects. Through this approach, the interplay between effects is appropriately considered. The primary goal of the model's application is to enable extensive sensitivity studies. A variety of such studies is presented in chapter 6. The focus of this chapter is to provide an overview of all parts of the model and the general calculation procedure. Chapter 5 then continues with a detailed description of each effect's implementation.

## 4.1 Goals and Requirements

Two main goals were defined for the development of the TRModel code. Namely, providing a realistic assessment of the sensitivity for TRISTAN Phase-1 and investigating hardware requirements. It is urgent to address these questions since hardware-related developments may need considerable lead time before the measurement begins. These two overarching goals can be divided into several subgoals that imply requirements for the model:

1. Sensitivity estimation for TRISTAN Phase-1
  - a) Identification and modeling of significant experimental effects
  - b) Consideration of interplay between effects
  - c) Ability to vary systematic parameters for propagating uncertainties
  - d) Minimal calculation time for model evaluation to enable parameter fitting
  - e) Quantification of the sensitivity impact of systematic parameter uncertainties
2. Hardware requirements and settings
  - a) Optimization of the electromagnetic field configuration
  - b) Determination of the optimal source activity
  - c) Testing the impact of hardware modifications that mitigate sensitivity loss

While these goals are essential, it is also important to mention what the model is not required to do. Namely, it does not need to be perfectly accurate, i. e. no exact agreement to calibration data is needed. Although accuracy is eventually necessary to fit data, sensitivity studies can be performed by comparing the model to itself as long as the general spectral shape and the uncertainty stemming from systematic parameters are sufficiently representative. Making the model accurate is another big challenge, and it would require many calibration measurements that cannot be performed without the TRISTAN detector.

### Requirements

Another question is: Can the existing model used for neutrino mass analysis also be employed for the keV sterile measurement? The short answer is: Unfortunately, no. There are multiple reasons why developing a new spectrum model is beneficial for analyzing TRISTAN detector data as opposed to using the existing integral KATRIN spectrum model. Here are three particular reasons, which are expanded upon in the subsequent paragraphs:

1. **Differential detector spectrum:** Analyzing the differential detector spectrum instead of the integral spectrum yields better sensitivity for keV sterile neutrinos. Calculating the differential spectrum requires detailed knowledge of the detector response.

2. **Increased energy range:** In several instances, the integral model assumes a fixed electron energy. This is a valid approximation for the energy range of the neutrino mass analysis, but, evidently, it does not hold for the entire spectrum. Furthermore, the existing model can completely neglect several experimental effects required to describe the entire spectrum. For instance: Electron backscattering at the rear wall or electrons traversing the beamline multiple times before they are registered at the detector.
3. **Dynamic pitch angle distribution:** For many of the new systematic effects, the dependence on the pitch angle of the  $\beta$ -electrons is much more critical. For example, any scattering effect, such as electron scattering on source gas, at the rear wall, or the detector, depends on and affects the angular distribution. Likewise, the angular distribution is not static along the beamline and is mostly non-isotropic and energy-dependent for electron energies well below the endpoint.

### Differential detector spectrum

As mentioned in section 3.4 and demonstrated in chapter 6, measurement and analysis of the differential spectrum results in considerably better sensitivity for keV sterile neutrinos than applying the integral method. This is mainly due to significant statistics gain and less dependence on source stability.

The integral spectrum model used for neutrino mass analysis [221, 222] predicts the detected rate at the focal plane detector as a function of the retarding potential  $U$  inside the main spectrometer. The calculation is performed via a one-dimensional convolution. Here, the product of differential  $\beta$ -decay rate  $\frac{d\Gamma}{dE}$  and an experimental response function  $R(E, U)$  are integrated over the kinetic energy  $E$  of the  $\beta$ -electron:

$$\dot{N}(U) = \frac{1}{2} N_T \int_{qU}^{E_0} \frac{d\Gamma}{dE} \cdot R(E, U) dE . \quad (4.1)$$

The normalization factor  $N_T$  equals the number of tritium atoms in the source and a factor  $\frac{1}{2}$  accounts for the source emitting only half of all decay electrons in the direction of the detector. The lower integration bound is equal to the retarding energy  $qU$  where  $q$  denotes the negative elementary charge of the electron. This lower bound represents that only electrons with sufficient surplus energy can make it past the main spectrometer, which acts as a high-pass filter.

Experimentally, the integral count rate is obtained by summing the registered events in a particular region of interest (ROI) in the differential detector spectrum. The latter term refers to the distribution of deposited energies in the focal plane detector as the DAQ reads them out. Not accounting for distortions from systematic effects, the differential detector spectrum represents the actual energy spectrum of the incoming electrons. However, there exist tiny distortions, and they must be considered in the model. In particular, they arise from detector backscattering, energy loss in the detector's entrance window, statistical and electronic noise, signal pileup, and dead time. In the case of the integral model, any changes to the count rate caused by detector- or DAQ-related systematics

are considered through efficiency factors. While these efficiency factors may depend on the retarding potential, they are by construction independent of the deposited energy in the detector.

To conclude, the current integral spectrum model does not calculate the differential detector spectrum and instead directly predicts the expected count rate. Calculating the differential detector spectrum instead of the count rate requires much more detailed knowledge of the detector’s response. The new detector response model within TRModel addresses this critical fact.

### Increased energy range

Due to the low mass scale of active neutrinos, a narrow energy window around the spectral endpoint is chosen for KATRIN neutrino mass measurement. With the endpoint energy  $E_0 \approx 18\,575\text{ eV}$  the analysis window for the most recent analysis is [7]

$$E_0 - 40\text{ eV to } E_0 + 135\text{ eV} . \quad (4.2)$$

This window contains all the information about the effective neutrino mass. For a keV sterile neutrino, however, the characteristic spectral kink can be located anywhere in the  $\beta$ -spectrum, given that the mass is smaller than the endpoint energy. For a measurement and a corresponding model description, the energy range should ideally be extended as far below the endpoint as technically possible and reasonable from an analysis perspective.

The main technical limitations for the minimum energy are the maximum count rate capability of the focal plane detector, adiabatic transport in the spectrometer section, and ion retention in the transport section. While the concerns of high rates at the detector are addressed with the new TRISTAN detector, the adiabaticity is still under investigation [243]. Negative blocking potentials of up to 100 V are located along the transport section for source ion retention [255]. This threshold is a limit that cannot be circumvented, but compared to the demands from the data analysis, it is entirely within reason.

For the data analysis, a lower limit of around 100 eV to 500 eV appears to be reasonable according to simulations performed for TRModel. Below this soft threshold, scattering effects in the source and on the rear wall produce many secondary electrons leading to large systematic uncertainties (see sections 5.1.1 and 5.2.1). Therefore, a model description below this limit is nonsensical, and a lower bound of 100 eV is chosen for all TRModel simulations. In the following, the term “deep spectrum” will refer to the tritium spectrum in the energy range

$$100\text{ eV to } 18\,600\text{ eV} . \quad (4.3)$$

With the extension of the energy range, several experimental effects, which were neglectable or treated with simplified models before, become relevant because their associated energy loss no longer moves the events outside the analysis window. The most significant effects to be considered are rear wall backscattering, the shape of the energy spectrum from tritium deposited on the rear wall surface, and the energy dependence of source scattering cross-sections. For a more comprehensive list, see section 4.2.

### Dynamic pitch angle distribution

Apart from the kinetic energy of  $\beta$ -decay electrons, the angle at which electrons are emitted in the gaseous tritium source plays an important role for KATRIN. In particular, the critical quantity is the pitch angle  $\theta$  which is defined as the angle between the electron momentum and the direction of the magnetic guiding field

$$\theta = \sphericalangle(\vec{p}_e, \vec{B}) \Leftrightarrow \cos \theta = \frac{\vec{p}_e \cdot \vec{B}}{|\vec{p}_e| \cdot |\vec{B}|} . \quad (4.4)$$

If the magnetic field is parallel to the central beamline axis of KATRIN, as is the case in the WGTS, the pitch angle is equal to the polar angle. Often it is helpful to use the cosine of the pitch angle because an isotropic pitch angle distribution is flat in  $\cos \theta$ . Further information can be found in section 4.4.2. Due to axial symmetry, the azimuthal angle  $\phi$  can be neglected.

The pitch angle is considered several times in the integral KATRIN spectrum model. First, the magnetic field at the pinch magnet determines the acceptance angle

$$\theta_{\max} = \arcsin \sqrt{\frac{B_{\text{src}}}{B_{\text{max}}}} \Leftrightarrow (\cos \theta)_{\min} = \sqrt{1 - \frac{B_{\text{src}}}{B_{\text{max}}}} \quad (4.5)$$

Second, the pitch angle distribution determines the shape of the main spectrometer transmission function (see section 5.3.3). Third, the initial pitch angle is considered in the description of source gas scattering since it determines the path length the electrons travel within the source (see section 5.1.1). [221, 256]

The source electrons' pitch angle distribution is generally considered to be static, meaning that even if an electron scatters, its pitch angle stays unchanged. This is a secure assumption as long as the electron energy is sufficiently close to the endpoint. In a single scattering, the electron changes its direction of travel depending on the energy loss and type of scattering process. Since the elastic scattering cross-section is small and the energy loss by inelastic scattering is  $> 10$  eV [257], few scatterings are possible until the electron energy falls outside the analysis window from equation (4.2) [256]. For this reason, a static angular distribution is sufficient, and for the construction of the overall experimental response function  $R(E, U)$  from equation (4.1) all angles within the acceptance angle (4.5) can be integrated over.

However, for the deep spectrum (4.3), many more scatterings are possible on average, and higher energy losses and angular changes due to ionization scattering have to be considered. Hence, the assumption of a static angular distribution is no longer valid. A dynamic and energy-dependent angular distribution is required. Furthermore, several new systematic effects, as explained in the following section, are also angle-dependent. Most importantly, backscattering off the detector and the rear wall requires information about the incident angle. And because there are now multiple effects that change both the energy and the pitch angle of electrons, the interplay of effects becomes a major concern and needs to be addressed by the model.

## 4.2 Overview of Systematic Effects

When analyzing the KATRIN tritium spectrum, it must be considered that experimental effects distort the decay spectrum (see section 3.1). Any description of these effects usually depends on parameters that carry systematic uncertainties. Overall, the uncertainty of the measured data points of the energy spectrum is composed of statistical and systematic error, which can be expressed in a covariance matrix if the uncertainty is sufficiently small.

In the deep tritium spectrum measurement, the set of known systematics deviates from those relevant to the endpoint measurement, and effects that are relevant to both may differ significantly in importance. From the model-building perspective, all effects that cannot be neglected must be represented by model components. This work will focus only on the components most relevant to the deep spectrum measurement. For a list of effects considered in the neutrino mass analysis, see [7, 222, 258]. Table 4.1 summarizes the systematic effects of the deep spectrum measurement. While this list may not be conclusive, it contains all dominant systematics to the current state of knowledge. The effects are listed roughly in sequential order along the path from source to detector readout, representing the approximate order in which they are addressed in TRModel. The following subsections go through the categories of the table and give an overview of each effect and how they interconnect. See figure 4.1 for a graphical representation. Not all systematics listed in table 4.1 are implemented yet in TRModel, as some require more input or further external studies.

Furthermore, some effects are only relevant to the integral or differential mode. Table 4.2 summarizes these two aspects. In the following, all effects are explained broadly to give an overview. For more information on particular effects, see the corresponding section in chapter 5.

### Tritium decay

$\beta$ -electrons are emitted in the source from tritium decay, their decay energy following the differential energy spectrum according to Fermi's theory of weak decay. As described in section 3.1, several atomic corrections, such as charge shielding effects or nuclear recoil, must be considered. Furthermore, as for the endpoint measurement, the tritium decay daughter molecules may be excited to rovibrational states. The probability density for this effect is called the final state distribution, and for the extensive energy range of the deep spectrum measurement, its energy dependence cannot be neglected.

### Source

Within the WGTS,  $\beta$ -electrons can scatter off gas molecules in the source. Multiple processes are possible: elastic scattering, molecular excitation, and electron impact ionization. In each scattering process, electrons will lose energy, and their pitch angle is altered slightly. Compared to the endpoint measurement, a larger number of scattering processes have to be considered since electron energies are less likely to fall out of the

**Table 4.1:** Systematic effects for the deep tritium spectrum measurement.

Category	Effect	Systematic
Tritium decay	Atomic corrections	Theoretical uncertainty
	Final state distribution	Theoretical uncertainty
Source	T <sub>2</sub> scattering	Column density Cross-sections
	Magnetic traps	Column density Magnetic field Non-adiabatic escape
	Plasma potential	Electric potential
	Activity fluctuations	Column density
Rear wall	Backscattering	Amplitude & shape
	Tritium decays	Amplitude & shape
Transport & Spectrometer	Magnetic reflection	Magnetic fields
	Magnetic collimation	Magnetic fields
	Synchrotron energy loss	Path length
	Transmission function	Retarding potential
	Adiabaticity	Spectrometer B-field
Detector	Post acceleration	Electric potential
	Backscattering and reflection	Pulse pair resolution Reflection time-of-flight Pixel-change probability
	Energy deposition	Dead layer Backscattering escape
	Charge sharing	Charge cloud size
Background	Fano noise	Fano factor
	Energy spectrum	Amplitude & shape
Readout	Pileup	Pulse pair resolution
	Electronic noise	Broadening width
	Detection threshold	Cutoff energy and width
	Dead time	Efficiency
	ADC-nonlinearity	Distortion amplitude & shape
	Calibration	Gain and offset

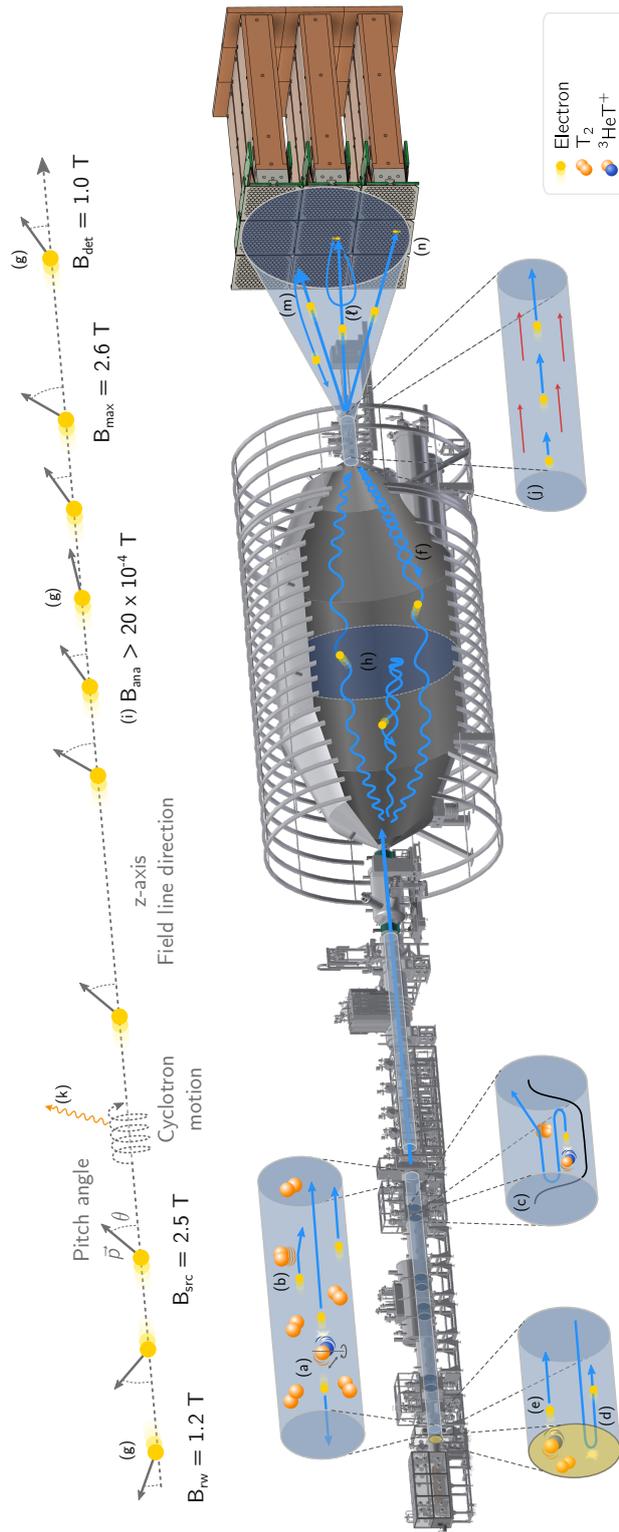
energy range of interest due to energy loss. Also, significant energy loss in ionization scattering cannot be neglected. The scattering probability is further increased for electrons trapped in dips of the magnetic field, which are present within the source volume. Electrons emitted with large pitch angles inside these dips will be trapped through magnetic reflection until their pitch angle is sufficiently decreased to leave the trap.

Tritium decay and scattering effects lead to a large quantity of charged particles in the WGTS volume, which form a plasma. This creates an electric potential on the millivolt scale. It defines the starting potential of decay electrons and is generally time-dependent. Furthermore, the overall activity of the source, which is proportional to the tritium column density, will fluctuate and slowly drift over time. These time-dependent effects are essential in the case of an integral measurement since the conditions must be sufficiently constant for all retarding potentials within at least one scan. These effects may be negligible or simpler to model for the differential measurement mode. Assuming no large instabilities exist, an inhomogeneous plasma potential would slightly broaden the energy distribution, similar to the Doppler effect. A sufficiently small broadening can be neglected compared to the detector resolution. Small fluctuations in the source activity can also be neglected in differential mode since they would only affect the normalization of the spectrum but not its shape.

### **Rear wall**

$\beta$ -electrons are emitted isotropically within the source. Hence, only half of all electrons are guided toward the detector, while the other half moves toward the rear wall. There they should ideally be absorbed, but a small chance to scatter back from the rear wall has to be considered, resulting in the deposition of some energy within the rear wall material. For the gold-plated rear wall, the overall probability for backscattering is rather large due to the high nuclear charge of gold, which leads to a backscattering coefficient on the order of 40 % to 50 % (see section 5.2.1). Depending on their pitch angle, backscattered electrons may either be trapped at the rear wall by magnetic fields or propagate back to the source and toward the detector. Because of the scattering energy loss, a sterile neutrino signature imprint can be considered lost. Due to the magnitude of the effect and the high uncertainty of backscattering simulations, rear wall backscattering is one of the dominant systematics.

Although most tritium gas is pumped away at the ends of the WGTS, the remaining tritium density at the rear wall is not negligible. It leads to tritium adsorption on all beam tube surfaces, including the rear wall, which causes an activity accumulation over time. Tritium on the rear wall surface is likely to be bound in hydrocarbons [259], which is expected to have a different final state distribution. The hydrocarbon layer can be removed by producing ozone within the rear wall chamber[234], but cleaning is time-consuming. It will likely only be performed in larger time intervals, i. e. days or months. Coupled with the fact that electrons from the rear wall are also likely to scatter before they reach the detector, the rear wall activity may be considered among the dominant systematics.



**Figure 4.1:** Overview of beamline-related experimental effects for the deep spectrum measurement: (a) Final state distribution, (b)  $T_2$  scattering, (c) Magnetic traps, (d) Rear wall backscattering, (e) Rear wall tritium decays, (f) Magnetic reflection, (g) Magnetic collimation, (h) Magnetic collimation, (i) Spectrometer adiabaticity, (j) Post acceleration, (k) Synchrotron energy loss, (l) Backscattering and back-reflection, (m) Charge collection and backscattering escape, (n) Charge sharing

### Transport and spectrometer

Electrons are adiabatically guided along the beamline by magnetic fields of the order of 1 T. During propagation, electrons move through regions of lower and higher magnetic fields compared to the field at their origin, i. e. the source or rear wall magnetic fields. When moving to a lower magnetic field region, the electron angular distribution undergoes magnetic collimation, the principle the MAC-E filter is designed upon. Reversely, when moving to a higher field region, the angular distribution is de-collimated, and the part of the angular distribution below the critical angle  $\theta_{\text{crit}}$  is reflected. The source and pinch magnetic field ratio defines the angular acceptance in the integral mode. But since electrons enter the beamline at multiple locations, there is not only one acceptance angle. Rather, electrons from the rear wall and source have individual acceptance angles (see section 5.2.1).

When electrons reach the spectrometer, they can be transmitted or reflected by the retarding potential. Whether an electron of a particular energy and angle is transmitted can be calculated analytically and is given by the transmission condition described in section 5.3.3. For the differential measurement, the transmission function is well-approximated by a step-like cut of the spectrum at the retarding energy because the 2.7 eV width of the transmission edge in nominal operation [5] is much smaller than the TRISTAN detector energy resolution of about 300 eV [8]. And since the transmission edge in the measured differential spectrum is at the edge of the region of interest, small instabilities of the retarding potential would be inconsequential. In an integral measurement, stability plays a role, however.

Another effect connected to the main spectrometer is the occurrence of non-adiabatic motion. For large surplus energies, as would be the case in a deep spectrum measurement, the adiabatic motion condition (see section 5.3) is violated in the spectrometer's low magnetic field. This can lead to an unwanted loss of transmission through the spectrometer. Unfortunately, then the energy- and angle-dependence is chaotic and cannot be accurately modeled [243]. Therefore, this effect has to be avoided, for example, by increasing the magnetic field of the air coils [260].

Downstream of the spectrometer, electrons are boosted in energy with the post-acceleration electrode before they hit the detector. Apart from shifting the energy spectrum by a fixed amount, this also strongly collimates their angular distribution.

Altogether, the electromagnetic fields govern the propagation of electrons along the beamline. Due to scattering and reflections, a large portion of  $\beta$ -electrons makes multiple passes through the experiment before they hit the detector or are absorbed in the rear wall. TRModel considers this in the so-called propagation model, approximating this process in an iterative convolution (see section 4.4.2). During their propagation in the magnetic field of the beamline, electrons continuously lose energy through the emission of synchrotron radiation. In work by Huber [243], it was shown that the energy loss is neglectable compared to the overall energy resolution. However, this should be reevaluated for electrons making multiple passes through the beamline until they are detected.

## Detector

Detector effects arguably introduce the most significant distortions to the measured tritium spectrum. In the integral spectrum measurement, these effects can be treated by applying efficiency factors to the measured count rate. But a detailed description of the spectral shape, as measured by the detector, is needed for the differential measurement mode.

Upon impact onto a detector pixel, electrons propagate into the bulk of the active detector material and deposit their energy through multiple scattering. Free charges in the form of electron-hole pairs are created, and the electrons drift to the readout anode. The creation of electron-hole pairs is a statistical process and thus randomly fluctuates between events, which leads to an energy broadening referred to as Fano noise [261]. This broadening represents the ultimate limit in energy resolution for silicon drift detectors.

Apart from this effect, the surface layer of the detector called the entrance window, is not fully active. This means the energy deposited here does not create electron-hole pairs, or the electrons can not be drifted to the anode. The surface layer may therefore be referred to as the dead layer. However, this name is sometimes misleading since the layer is not entirely insensitive, as the name might suggest. Instead, its effect is better described by a gradual increase in charge collection efficiency over depth [262, 263].

As the primary electron moves through the bulk via multiple scattering, it can be reflected and leave the detector bulk in the direction it came from, depositing only part of its energy. This is referred to as *backscattering*. The backscattered electron escapes in many detector test environments, and the deposited energy spectrum, referred to as the backscattering tail, has a characteristic shape. In the KATRIN environment, however, the backscattered electron re-enters the beamline. Depending on its pitch angle and energy, the electron is reflected yet again onto the detector by the electromagnetic fields, producing a second hit. If the return time is longer than the readout time resolution of about 100 ns (see section 5.5), or the position change due to drift and Larmor motion is large enough for the electron to migrate to another pixel, it is counted as a new event. However, it will likely return to the same detector pixel within the time resolution. For this fast back-reflection, the signal from the second hit merges with the one from the initial hit.

Lastly, the segmentation of the detector into pixels has to be taken into account. If an incident electron hits the detector close to a pixel boundary, the produced charge cloud distributes over both pixels causing a so-called charge-sharing event in each channel. Even charge-sharing between triple-pixel junctions is possible for the TRISTAN detector with its hexagonal pixel pattern [248]. Since the readout records event energy and time, a coincidence or multiplicity analysis can be performed to mitigate both charge sharing and back-reflection to adjacent pixels.

## **Background**

Background at KATRIN is the subject of extensive investigations due to its importance for measuring the neutrino mass scale. Without going into detail, the background can be grouped into two categories: spectrometer-related background and detector-related backgrounds such as intrinsic radioactivity and cosmic muon events [5, 240]. Work by Harms [264] provides an in-depth review of the differential background spectrum at KATRIN and a differential background rate of  $6 \times 10^{-4}$  cps/keV is found. This translates to a  $10^{-2}$  cps background rate over the deep spectrum range. Yet, the background for the deep spectrum measurement may slightly differ from the current expectation due to the adjusted field setting and deviating vacuum conditions stemming from the integration of the TRISTAN detector. Measurements simulating the new conditions are therefore vital.

Overall, the significance of the background is expected to be low because of the significant signal rate at the detector. Compared to the endpoint measurement, where the rate at 40 eV below the endpoint is on the order of 10 cps [5], the tolerated rate for the TRISTAN detector is more than a factor  $10^6$  higher [8]. Therefore, the background rate is also minor when compared to the tritium event rate. Nevertheless, the non-tritium background is included in TRModel with a simplified description. Any unknown shape of the background can be simulated by artificially increasing the uncertainty.

## **Readout**

The readout or data acquisition stage concludes the measurement. The signals from all channels are processed in pulse height analysis to extract the energy and timestamps of events registered as a charge in the detector. During the transfer from the detector to the digitization circuit outside the vacuum chamber, the signal picks up electronic noise. This degrades the energy resolution and manifests as a broadening of the energy spectrum in addition to the Fano broadening. The analog-to-digital converter is placed as close to the detector as possible to limit the noise to a minimum. At the same time, the rest of the data acquisition is positioned further away from the detector magnet and on electric ground potential. [8]

Unfortunately, digitization itself introduces a small distortion to the measured spectrum. In contrast to the integral method, this effect is specific to the differential measurement, which relies on pulse height extraction. Due to the unavoidable nonlinearity of the ADC transfer function, some pulse heights are more likely to occur than others. This results in a practically random fluctuation of the measured spectrum on the eV-scale. Accurate modeling of this effect is unrealistic due to the number of channels and modeling uncertainties. Instead, mitigation techniques, such as Gatti sliders, pixel averaging, and correction methods, are investigated to reduce the spectral distortion to an insignificant level. [252]

At the targeted source activity and a resulting event rate of about  $10^5$  cps per pixel, rate-dependent readout effects play an essential role. Each detector pixel is hit by a considerable fraction of electrons so close in time that their signal pulses overlap. As a result, their pulses cannot be distinguished from an event with higher energy. This pileup

phenomenon leads to an energy spectrum contribution overlapping with and extending past the measured tritium spectrum. The rate per pixel and the pulse pair time resolution determine the amplitude of the pileup spectrum. The fraction of pileup events compared to the number of input events is expected to be around 1% for the TRISTAN detector operated at its optimal input rate for the highest statistics [253].

Each registered event also causes dead time in its channel, which blocks the recording of another event for a particular time window. This mitigates the amount of pileup but also diminishes the overall efficiency. The trigger dead time window depends on the settings for the pulse shaping parameters of the digital filters used, and it is around 1 to 2  $\mu\text{s}$  for the TRISTAN readout (see section 5.5.4). Another source of dead time is the pulsed reset of the preamplification circuit, which adds a rate-independent component to account for the time required to discharge the preamplifier capacitance in regular intervals. Luckily, the effect of dead time on the spectrum is independent of energy and can be treated by applying an efficiency factor to the amplitude. In an analysis where the amplitude is fitted, it consequently does not matter as a systematic effect. In TRModel, the dead time efficiency is included regardless. This allows calculating the detected rate to estimate the collected statistics for a given beamline setting and measurement time.

**Table 4.2:** Relevance of systematic effects for differential and integral measurement modes. Relevance should not be mistaken for the significance of the effects, which is covered in chapter 6. The rightmost column indicates the implementation status in TRModel Release 1.0 [254]. ✓ indicates an implementation exists, and if enclosed by brackets, further extensions of the model are necessary. ✗ labels effects that are currently not available. ∼ denotes that the effect is covered by external studies that show it can be neglected or avoided experimentally.

Category	Effect	Differential	Integral	TRModel
Tritium decay	Atomic corrections	relevant	relevant	✓
	Final state distribution	relevant	relevant	✗
Source	T <sub>2</sub> scattering	relevant	relevant	(✓)
	Magnetic traps	relevant	relevant	(✓)
	Plasma potential	broadening	relevant	∼
	Activity fluctuations	irrelevant	relevant	∼
Rear wall	Backscattering	relevant	relevant	✓
	Tritium decays	relevant	relevant	✓
Transport & Spectrometer	Magnetic reflection	relevant	relevant	✓
	Magnetic collimation	relevant	relevant	✓
	Synchrotron energy loss	relevant	relevant	∼
	Transmission function	relevant	relevant	✓
	Adiabaticity	relevant	relevant	∼
Detector	Post acceleration	relevant	relevant	✓
	Backreflection	relevant	efficiency	(✓)
	Energy deposition	relevant	efficiency	✓
	Charge sharing	relevant	efficiency	✓
Background	Fano noise	relevant	efficiency	✓
	Energy spectrum	relevant	relevant	
Readout	Pileup	relevant	efficiency	✓
	Electronic noise	relevant	efficiency	✓
	Detection threshold	relevant	relevant	✓
	Dead time	efficiency	efficiency	✓
	ADC-nonlinearity	relevant	irrelevant	∼
	Energy calibration	relevant	relevant	✓

## 4.3 Modeling Approach

This section outlines the approach for developing TRModel, which aims to fulfill the goals and requirements listed in section 4.1.

### 4.3.1 Response Matrix Formalism

As a general approach, a forward convolution method is chosen in which the differential tritium  $\beta$ -spectrum is folded with an experimental response function to calculate a prediction of the measured spectrum. A comprehensive overview of the method can be found in the book by Blobel and Lohrmann [265].

The ultimate goal of model building is to compare the model to measured data, which requires high accuracy. However, for sensitivity studies, the spectrum prediction of the “no-mixing” case ( $\sin^2 \theta = 0$ ) can be compared - using a statistical test - to other predictions where a sterile neutrino component ( $m_s \in [0, 18.6 \text{ keV}]$  and  $\sin^2 \theta > 0$ ) is added to the input spectrum. This alleviates the constraint on accuracy since the no-mixing model prediction is used as a proxy for the data. Of course, the model prediction should still be as accurate as possible for realistic sensitivity estimations.

### Convolution

In mathematical terms, the kinetic energy  $E$  of the decay electron is a random variable whose distribution is equal to the differential tritium spectrum from Fermi’s  $\beta$ -decay theory. Considering the decay spectrum may also contain a sterile component, the input distribution can be written as

$$S_{\text{in}}(E; \sin^2 \theta, m_s) = \frac{d\Gamma}{dE}(E; \sin^2 \theta, m_s) . \quad (4.6)$$

For simplicity, the dependence on  $\sin^2 \theta$  and  $m_s$  will be suppressed in the following passages. The energy measured by the detector, denoted here as  $\hat{E}$ , is also a random variable and follows the distribution  $S_{\text{out}}(\hat{E})$ . Both distributions relate to one another via the convolution integral:

$$S_{\text{out}}(\hat{E}) = \int_{E_{\text{min}}}^{E_{\text{max}}} R(E_{\text{det}}, E; \vec{p}) \cdot S_{\text{in}}(E) dE . \quad (4.7)$$

The integration kernel  $R(\hat{E}, E; \vec{p})$  is the complete response function of the experiment, which can be obtained through a combination of Monte Carlo methods and analytical models for the sub-components. For any value of the measured energy  $\hat{E}$ , the response yields the probability of measuring that particular value if the actual value is  $E$ . Ideally, the response would be a delta function, and the measured energy would always equal the actual energy. In practice, the experimental effects outlined in section 4.2 will result in uncertainties, biases, and efficiency losses in determining the energy spectrum. Furthermore, the response function depends on nuisance parameters, denoted by the vector  $\vec{p}$ , which may carry systematic uncertainties.

### Discretization

In practice, the measured spectrum will not be a smooth distribution but rather a histogram. And similarly, for the numerical solution of the convolution integral, the input and output distributions are represented as binned histograms. Equation (4.7) can then be re-written in discrete form in terms of a matrix multiplication

$$\vec{S}_{\text{out}} = R(\vec{p}) \vec{S}_{\text{in}} , \quad (4.8)$$

where  $\vec{S}_{\text{in}}$  and  $\vec{S}_{\text{out}}$  are now vectors of length  $m$  and  $n$  respectively, which contain the number of events in every bin. The response corresponds to an  $m \times n$  matrix that maps each input bin to all output bins. Any off-diagonal elements of the matrix will cause a re-distribution of events from one energy bin to another. It should be noted that neither bin width nor the actual number of bins, and, therefore, the energy range needs to be the same for input and output. In fact, due to technical limitations in some parts of TRModel, the output spectrum has considerably finer binning compared to the input. Also, the output extends to higher energies to accommodate for increased event energy from post-acceleration and pileup.

### Factorization

The response matrix can be subdivided into matrices describing the individual effects listed in section 4.2. The complete response is then the product of all component responses

$$R(\vec{p}) = R_N(\vec{p}_N) \times R_{N-1}(\vec{p}_{N-1}) \times \cdots \times R_2(\vec{p}_2) \times R_1(\vec{p}_1) , \quad (4.9)$$

where the rightmost matrix acts on the spectrum first. The component responses may depend on their own set of parameters  $\vec{p}_i$  that all concatenate to  $\vec{p}$ . Again, the shape of the matrices is arbitrary - even the number of dimensions - as long as they interface correctly with each other and  $R_1$  and  $R_N$  match the number of bins of  $\vec{S}_{\text{in}}$  and  $\vec{S}_{\text{out}}$  respectively. This way, it is possible to extend the dimensionality starting at the tritium source to include a dimension for the pitch angle by giving the corresponding response three dimensions - one for the input energy and two for the corresponding output energy and angle. The angular dimension is later collapsed at the response describing the energy deposition in the detector, which has two input dimensions and one output dimension.

### Calculation

The subdivision of the response into factors further enables the use of different approaches for determining the response of each component. For example, the rear wall and detector backscattering responses can be computed with Monte Carlo simulations, whereas the magnetic collimation and mirroring responses can be calculated analytically. Among other information, table 4.3 summarizes which responses are modeled analytically or with Monte Carlo simulations. The advantages and drawbacks of both approaches are covered in the following passage.

The Monte Carlo method (MC) is a widely employed class of computational algorithms to treat statistical problems. It allows for the modeling of complex systems with comparative ease by random sampling to numerically approximate the prediction of an observable

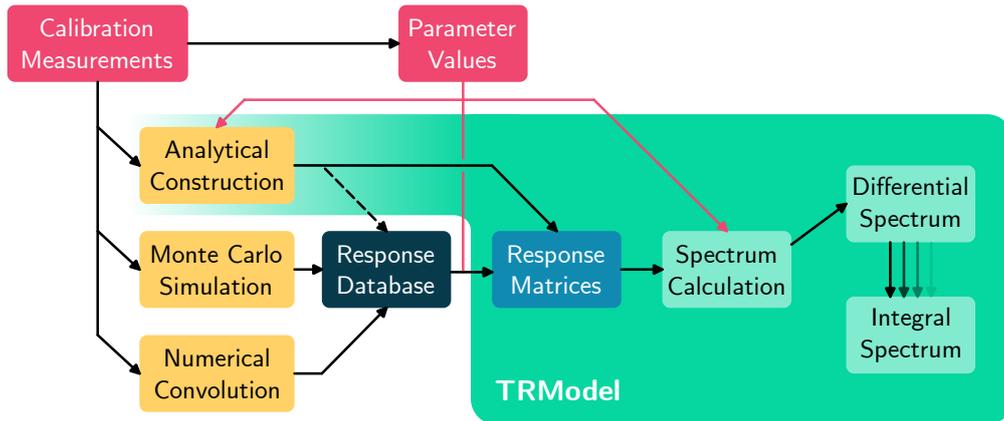
[265]. Many MC-based tools exist, and they are widely used at KATRIN. For example, Geant4 [266] for the simulation of particle passage through matter, KESS [267] for the simulation of electron interactions in silicon detectors, Kassiopeia [268] for tracking of electrons through the electromagnetic field of the KATRIN beamline, and Drips [269] or other trace-level signal processing software [253] for simulating the readout chain.

Analytical models may be preferred over Monte Carlo simulations for several reasons. Firstly, analytical do not contain statistical errors, and secondly, they are, in some cases, faster to compute depending on the required statistics for a corresponding MC simulation. But even if an appropriate analytic description already exists, a Monte Carlo simulation can still be helpful to validate the analytical model. But as soon as experimental effects increase in complexity, as for example in detector backscattering, the MC method is often favored since the effect may come to be too complicated to handle analytically. Depending on the simulation result, finding an analytical parametrization that matches the simulation output can be feasible.

The parameterization of MC data may be necessary for the deep tritium measurement in the future since it might be unfeasible to run Monte Carlo simulations with sufficient statistics to reach the target precision [270]. To illustrate the problem: For the deep tritium spectrum measurement, the collection of up to  $10^{16}$  decay electrons is envisioned [225]. Assuming an unrealistically low simulation time per event of 1 s and parallel computation on a million CPU threads, it would still take over 300 years to complete the simulation of the MC dataset with equal statistics as the measurement. Usually, the statistical error of the simulation should at least be smaller than the statistical error of the measured observable. A general rule of thumb is to have even 10 times the amount of MC statistics compared to the measurement [270].

### 4.3.2 Code Implementation

Adhering to the response matrix formalism, the main task for the new model code is matrix multiplication, a math operation for which efficient implementations are available in many programming languages. TRModel is implemented in Python 3 [271], which is well suited to the task when optimized libraries such as NumPy [272] are utilized. TRModel can be installed following the instructions on the KATRIN Gitlab [254] and imported as a package in Python. An overview of the code structure of TRModel is shown in figure 4.3. The code consists of four main parts: The global model, several sub-packages, response databases, and a shared library.



**Figure 4.2:** Strategy for the development of TRModel: Response matrices are constructed via Monte Carlo simulations (e. g. detector and rear wall scattering), numerical convolution (e. g. source scattering), or analytical construction (e. g. post-acceleration). These methods, as well as any parameter values associated with them, are informed by calibration measurements. The response calculation is performed in advance and stored in databases if necessary. TRModel retrieves all required responses and applies them in the spectrum calculation. In the current release of TRModel, all analytically constructed responses are calculated at runtime, but in principle, they could also be stored in a database. The integral spectrum is computed through repeated calculation of the differential spectrum with different retarding potential values.

## Structure

The global model contains the main executable functions for calculating the spectrum prediction. For a set of parameters the user gives in a Python dictionary, the model evaluation function `model_differential` calculates the differential spectrum as the detector readout registers it. It does so by gathering all necessary responses from the sub-packages and applying them in the appropriate order to the tritium spectrum. A detailed description of the entire procedure is given in section 4.4. Another function called `model_integral` can be used to calculate the integral spectrum. This is accomplished by evaluating the differential model for a set of retarding potentials and summing the counts in a region of interest.

The sub-packages contain the models of the categories listed in table 4.3. Each package provides functions for creating or loading their respective response matrices and methods for applying them to the output spectrum of the preceding convolution step. At the time of writing, there are seven packages. Most packages are self-sufficient and can be imported, used, and tested as standalone. One exception is the propagation package which can be seen as an extension of the global model. In it, the responses from other packages are combined to form an abstracted version of the beamline between the source and the detector surface, i. e. it encompasses all processes where the  $\beta$ -electrons propagate in space and their pitch angle matters. Lastly, the shared library contains any code not specific to one of the sub-packages or the global model. For example, the library includes a list of relevant physical constants, standard methods for caching, rebinning, or other technical aspects, and extensions for performing parameter fits and sensitivity studies.

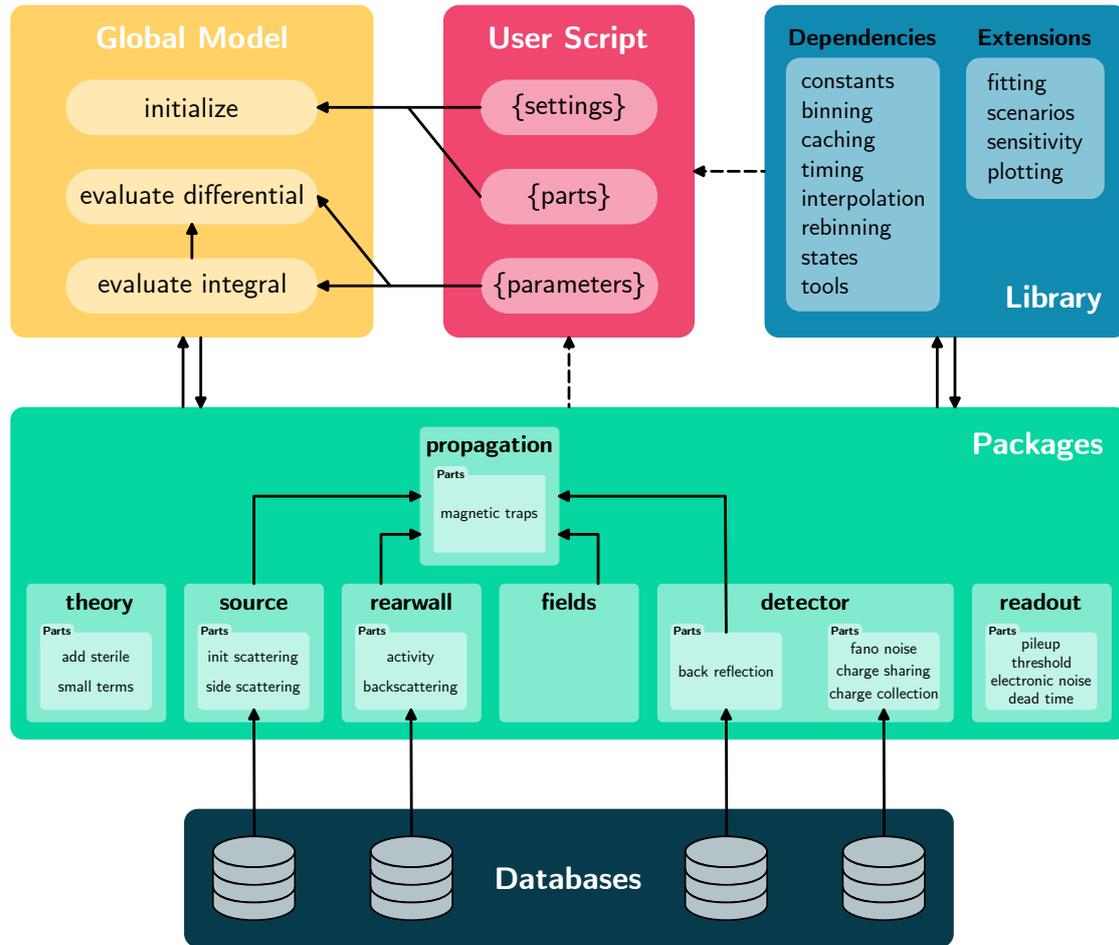


Figure 4.3: TRModel code structure

### Performance

An essential requirement to execute parameter fits and sensitivity studies is the speed of calculation. For each parameter fit, the model may be evaluated on the order of 10-1000 times depending on the particular choice and number of fit parameters. In the sensitivity studies presented in chapter 6, fits for several nuisance parameters are performed for each sterile neutrino hypothesis. Since there are two parameters of interest, namely the mixing angle  $\sin^2 \theta$  and the sterile neutrino mass  $m_s$  which are scanned in a grid of dimension  $50 \times 20$ , in total 1000 hypotheses are checked per study. Altogether, the number of model evaluations equals the product of hypotheses times the average number of evaluations per fit, resulting in  $10^4 - 10^6$  evaluations. So the calculation speed of a model evaluation should be between milliseconds and seconds to perform studies within hours or days of computation time.

A significant factor determining the calculation speed lies in the approach. Using the response matrix method, responses for most effects can be calculated once and used for any input spectrum, i. e. spectra with different sterile neutrino hypotheses. But apart from the choice of the base method, TRModel utilizes several other techniques to reduce calculation time:

- **Response database:** For most components listed in table 4.3, the responses can be calculated at runtime, i. e. when the model evaluation function is called. This is impossible for large matrices and ones based on external simulations since running these simulations for each function call would require too much time. Instead, the simulations are run in advance for a range of input parameters. From the simulations, a database of responses is constructed that has one additional dimension per parameter. At the time of writing, there are databases for five types of responses: Initial source scattering, source scattering for side-entry electrons, rear wall backscattering, detector back reflection, and detector charge collection.
- **Model initialization:** The calculation of the spectrum prediction with TRModel is split into two steps. The first step is the initialization, during which any computing-intensive task must be performed only once. Most notably in this step, TRModel loads the databases - or parts of them - into the computer memory. In the second step, the model is evaluated as many times as required without rerunning the initialization tasks every time.
- **Vectorization:** Using NumPy [272] arrays and corresponding methods for array operations, calculation speed can be significantly improved since the core NumPy functionalities are pre-compiled and programmed in highly optimized C code. This method can be regarded as standard practice in Python data analysis and development. The process of adjusting the code to accommodate the usage of the NumPy library is called vectorization, and it is employed throughout TRModel.
- **Parallelization:** As an additional benefit, NumPy automatically parallelizes its operations to run on multiple CPU threads. On the Kalinka cluster at KIT, a model evaluation can be distributed over up to 80 threads simultaneously, depending on the particular computation node. Additionally, separate instances of TRModel can be run on multiple cluster nodes.
- **Caching:** Whenever a response matrix is calculated or read from a database, it is stored in a last-recently-used cache (LRU cache). This functionality is implemented using the Memoization package [273], which provides decorator functions [271] that can be attached to any python function. For all functions decorated in this way, the function's input argument values and the corresponding return object are stored in a memory cache. If the function is later called with the same set of arguments as in a prior evaluation, the function directly returns the corresponding cached object, resulting in a significant speedup. The LRU cache has a specified length, and it operates in a first-in-first-out manner, i. e. when the length is exceeded, the oldest entry is deleted so that the overall memory consumption is kept in check.

In TRModel, all functions returning response matrices are cached in this way. Furthermore, the caching is applied to intermediate energy spectra in the chain of response applications, leading to a speed-up as long the initial spectrum stays the same. This is especially useful for fitting parameters applied in later stages of the spectrum calculation, such as the overall spectrum normalization or the data acquisition gain, since most calculation steps are skipped.

- **Just-in-time compilation:** As an interpreter-based language, Python is slower than compiled languages like C/C++. This may become a problem when it is difficult or impossible to vectorize the code for use with NumPy. A standard method for achieving performance comparable to faster languages is the just-in-time (JIT) compilation using the Numba package [274]. With the decorators offered by the package, particular functions within TRModel are labeled for a compilation step, translating the function to fast machine code. The term “just-in-time compilation” stems from the fact that the compilation happens at runtime when the function is called for the first time. The compiled and much faster version is used for subsequent function calls. Functions that conduct numerical tasks and use many loops are particularly suitable, and performance can be increased by up to one or two orders of magnitude [274]. Not all functions are compatible with Numba’s machine code compilation since only a subset of Python, and NumPy features are supported. In TRModel, the `fields` package makes full use of Numba for the analytical construction of its response matrices.

### Initialization

As mentioned above, the spectrum calculation is split into the initialization and model evaluation steps. With the initialization function, the user specifies which experimental effects - or parts - are to be included by passing a Python dictionary of bool values - the “parts” dictionary in figure 4.3 - corresponding to the model components. The global model relays the information to the packages. Each effect can be toggled, and the user can tailor the model to what is relevant to their analysis. Disabling components generally lessens the memory requirements and decreases computation time, especially if any of the 3- or 4-dimensional responses are removed.

Furthermore, it is required to pass another dictionary - the “settings” dictionary - to the initialization function to choose between a variety of options that are not allowed to be changed during model evaluation. These settings consist of fixed parameters, debug options, the names of the databases to load, binning specifications, and string-type specifiers that are used to choose between different implementations of detector effects.

**Table 4.3:** Responses available with TRModel Release 1.0 [254]: Listed are the matrix dimensions  $N_d$  including input and output dimensions, their construction method, whether they are calculated at runtime or read from a database on disk, and their input parameters. In the cases of magnetic traps and rear wall tritium decays, the effects are described using combinations of the other responses in TRModel. Pileup is described via a self-convolution of the detected spectrum and not with a response matrix calculated in advance. The deadtime response is scalar and therefore listed with zero number of dimensions.

Category	Component	$N_d$	Method	Runtime	Parameters
Tritium decay	Atomic corrections	1	Analytic	yes	
Source	Scattering initial	3	Analytic	no	$\rho d_{\text{init}}$
	Scattering side	4	Analytic	no	$\rho d_{\text{side}}$
	Magnetic traps	-	Analytic	-	$\rho d_{\text{trap}}, B_{\text{trap}}, B_{\text{src}}$
Rear wall	Backscattering	4	Geant4	no	$A_{\text{rwbs}}$
	Tritium decays	-	Geant4	-	$A_{\text{act}}$
Transport & Spectrometer	Magnetic reflection	1	Analytic	yes	$B_1, B_2$
	Magnetic collimation	2	Analytic	yes	$B_1, B_2$
	Transmission	1	Analytic	yes	$qU$
	PAE reflection	1	Analytic	yes	$E_{\text{pae}}$
	PAE energy boost	2	Analytic	yes	$E_{\text{pae}}$
	PAE collimation	2	Analytic	yes	$E_{\text{pae}}$
Detector	Backreflection	4	Geant4	no	$A_{\text{refl}}$
	Energy deposition	3	Geant4	no	$\lambda$
	Charge sharing	2	Analytic	yes	$w_{\text{cc}}$
	Fano noise	2	Analytic	yes	$F_{\text{fano}}$
Readout	Pileup	-	Analytic	-	$\tau_{\text{min}}$
	Electronic noise	2	Analytic	yes	$\sigma_{\text{en}}$
	Detection threshold	1	Analytic	yes	$E_{\text{thr}}, \sigma_{\text{thr}}$
	Dead time	0	Analytic	yes	$\tau_{\text{dead}}$
	Calibration	2	Analytic	yes	$G, E_{\text{offset}}$

## 4.4 Differential Spectrum Calculation

Calculating the differential spectrum is the core goal of TRModel. Other functionalities, such as integral spectrum calculation and sensitivity estimation, are based on it. This section explains all technical aspects of the spectrum calculation, while details concerning individual effect descriptions and response matrix construction are given in chapter 5. An example of the resulting spectrum illustrating the most prominent spectral features is shown in figure 4.4. The evaluation of the differential spectrum is structured into four high-level steps. In order of execution, they are:

1. Calculation of the binned tritium  $\beta$ -decay energy spectrum.
2. Propagation of the energy spectrum from the source to the detector.
3. Application of the detector response.
4. Evaluation of the data acquisition model.

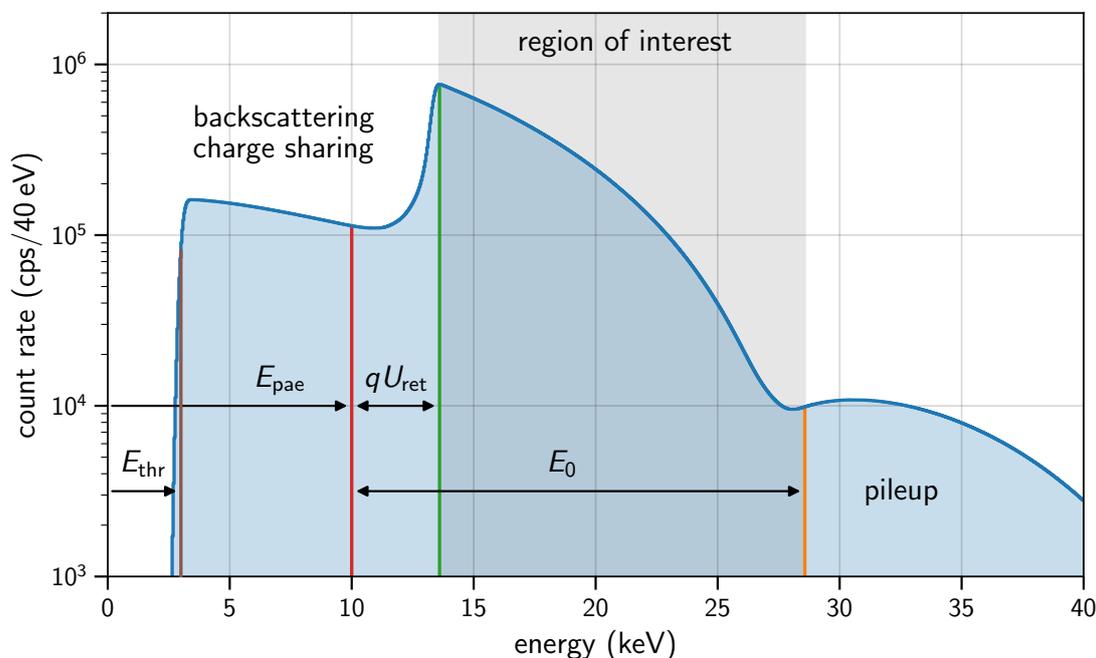
These steps are invoked in the `model_differential` function of the global model, and the corresponding packages carry out calculations involving several substeps. Each step is explained in detail in subsections 4.4.1 through 4.4.6.

Additionally, the global model performs two other vital tasks. The first task concerns the interfacing between response matrices of different binning - particularly the interface between the propagation model output and the detector response. A common binning with a width of 185 eV is used for the initial energy spectrum and the propagation model.<sup>1</sup> This is relatively large compared to the detector resolution of about 300 eV FWHM [8]. The detector response matrices are therefore constructed with a finer bin width of 10 to 185 eV depending on the setting. A rebinning method that assumes a smooth underlying energy distribution is utilized to scale up the spectrum from the propagation step to match the detector response. How this is done while conserving the bin contents is covered in subsection 4.4.3.

The second task is the normalization of the spectrum such that the integral is equal to the rate of registered hits per detector pixel. This is vital for the rate-dependent readout effects pileup and dead time and the correct overall predicted rate after the readout. Correctly estimating the total detected rate is crucial for gauging the statistics used in sensitivity studies. The specifics of the normalization procedure are covered in subsection 4.4.5.

---

<sup>1</sup>This binning is chosen since it is sufficient for sensitivity studies and computationally feasible with the current TRModel implementation. The same bin width is used in a previous sensitivity study [225].



**Figure 4.4:** Example of the TRModel differential spectrum: For this demonstration, a retarding potential of  $qU = 3.6$  keV, a post-acceleration energy of  $E_{\text{pae}} = 10$  keV, and a detection threshold of  $E_{\text{thr}} = 3$  keV are chosen. The retarding potential cuts off the lower energetic part of the spectrum during propagation along the beamline. The  $\beta$ -spectrum - including the cutoff - is shifted along the energy axis by the post-acceleration energy. Due to detector effects - primarily backscattering and charge sharing - the predicted spectrum extends below the cutoff with a characteristic shape, often referred to as a backscattering tail. However, other detector systematics, such as back-reflection and energy loss in the dead layer, also contribute to the tail. The spectrum above the cutoff and below the shifted endpoint energy  $E_0 + E_{\text{pae}}$  is labeled as the region of interest since the shape in this region closest resembles the original tritium spectrum shape. The high signal rate per detector pixel leads to a significant probability of signal pileup, which causes a spectral contribution that extends above the shifted endpoint. Lastly, the detection threshold energy leads to a broadened cut-off at the lower end of the spectrum.

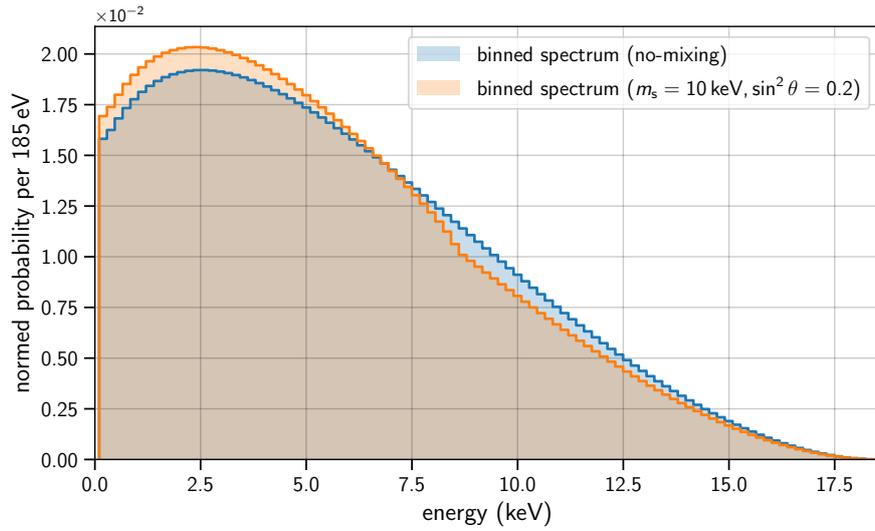


Figure 4.5: TRModel binned tritium  $\beta$ -decay spectrum.

#### 4.4.1 $\beta$ -Decay Spectrum

The differential tritium  $\beta$ -spectrum is implemented according to the description in section 3.1. To summarize, the decay is described with Fermi's point-like weak interaction and the golden rule [141]. Any energy-independent constants can be neglected since the spectrum is later normalized according to the experimental tritium half-life from [275]. Multiplicative atomic corrections are also considered and implemented according to [223]. Their impact on the keV sterile neutrino sensitivity is quantified in a previous publication by Mertens et al. [225]. As stated in section 3.1, energy loss from the excitation of the daughter molecule is neglected due to missing information on the energy-dependent final state distribution (FSD). Calculations concerning the FSD are ongoing [226]. The effect on the tritium spectrum will be implemented in TRModel as soon as the FSD is available. The sterile neutrino's contribution to the  $\beta$ -spectrum is implemented following equation (3.13) in section 3.4.1.

In TRModel, the differential decay rate is numerically integrated over each energy bin interval. For this work, a binning of 100 equally wide intervals over the energy range [100, 18600] eV is chosen, resulting in a bin width of 185 eV. This choice is in close agreement with the previous study by Mertens et al. [225]. But more importantly, the binning is chosen according to the binning of the response matrices of the propagation model. After the bin integration, the spectrum is represented as a vector of length 100. In the last step of the initial spectrum calculation, the vector is normalized such that its sum, corresponding to the total decay probability, is equal to 1. The normalization adjustment to match the expected tritium decay rate is handled later, as described in section 4.4.5. Figure 4.5 shows the final binned tritium  $\beta$ -spectrum.

### 4.4.2 Propagation Model

The propagation model is structurally the most complex part of TRModel, combining many response matrices into an abstract version of the beamline. Its purpose is to describe the energy and pitch angle-dependent transport of  $\beta$ -electrons, starting with the binned tritium spectrum from the previous section, to calculate the spectrum of  $\beta$ -electrons hitting the detector at the downstream end of the beamline. The propagation of the electrons is emulated with an iterative convolution method where the initial distribution is injected into a network of response matrices and intermediate stages - here referred to as *states*. These states temporarily store the distribution and correspond to locations along the beamline. In each iteration step, the distribution is transferred from one state to the adjacent states by convolving with the appropriate response matrices that describe beamline components or systematic effects occurring between the state locations. The transfer along the state chain continues until an output state at the detector is reached, where the distribution content is accumulated. A similar approach is employed in work by [276], and in the TRISTAN Source Simulation [277].

#### Pitch angle binning

During propagation, the  $\beta$ -spectrum is represented as a 2-dimensional binned distribution over energy and  $\cos\theta$ , where  $\theta$  is the pitch angle. This 2D approach is necessary since almost all propagation effects depend strongly on the pitch angle. The 2-dimensional electron distribution is represented by a  $100 \times 50$  matrix associated with 100 energy bins in the range 100 eV to 18 600 eV and 50  $\cos\theta$  bins in the range  $[-1, 1]$ . Here,  $\cos\theta = -1$  corresponds to electrons moving upstream toward the rear wall,  $\cos\theta = 1$  corresponds to electrons moving downstream toward the detector, and  $\cos\theta = 0$  corresponds to electrons with pitch angle perpendicular to the magnetic field lines. The binning is chosen such that one bin edge exactly coincides with 0. This way, the matrix can be split into a downstream or *forward* propagating component with  $\cos\theta \in [0, 1]$ , and an upstream or *backward* moving component with  $\cos\theta \in [-1, 0]$ . The cosine of the pitch angle is chosen as a basis for equidistant binning instead of binning in  $\theta$  for several reasons. Most notably, the isotropic angular distribution is uniform in  $\cos\theta$ . For further explanation and a list of advantages and drawbacks of this choice, see appendix A.1.

#### Building blocks

The propagation model is built from 4 different object types: Response matrices, input states, intermediate states, and the output state. In figure 4.6, these objects are represented by differently colored blocks, and together, they form an abstract version of the KATRIN beamline. Before the iterative convolution begins, input distributions are fed into the input states. The primary input is the binned tritium spectrum, represented by a 1D vector, written into the input state for source gas tritium decay. The vector is directly convolved with the 3-dimensional initial source scattering response which describes the effect of electrons scattering on their way out of the source region when they originate from within. The response has one input dimension for energy and two output dimensions

for energy and angle. This way, the 1D input vector is transformed into a 2D distribution written into the states adjacent to the source. Other optional input states are defined for  $\beta$ -electrons originating from magnetic traps and the rear wall surface.

During the iterative convolution, the 2D electron distribution is stored in intermediate states and moved between them. All states have a forward and backward component. The states are associated with z-positions and electromagnetic field conditions along the beamline. Ordered from upstream to downstream they are located as follows:

- State 1: Directly in front of the rear wall surface
- State 2: Upstream of the WGTS
- State 3: Downstream of the WGTS
- State 4: In between Spectrometer and post-acceleration electrode
- State 5: Directly in front of the detector surface

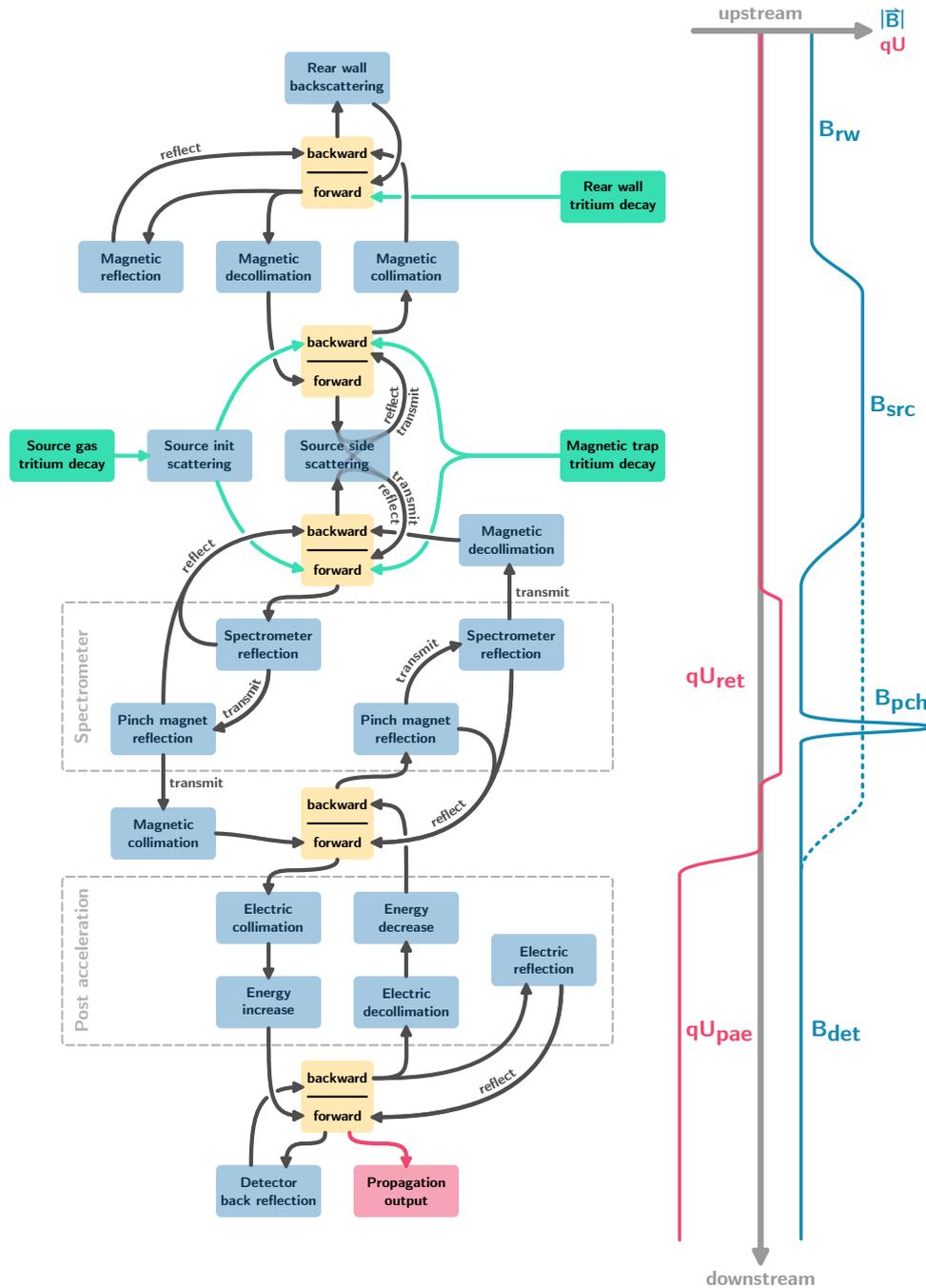
The electromagnetic field conditions at these state locations are indicated in figure 4.6. To account for a post-acceleration of up to 20 keV, the energy binning of state 5 and the output state is increased to 216 bins which span the range from 100 eV to 40 060 eV.

The output state represents the distribution of electrons that create distinct hits in the detector. A clarification on the term “distinct hit” is given in the passage about propagation output further below. Opposed to the intermediate states, the output state accumulates what is written into it instead of being overwritten in each iteration. Also, the output state only has a forward component since any backward-facing component re-enters the beamline.

### Iterative convolution algorithm

The iterative convolution algorithm is structured in two stages: initialization and iteration. During initialization, the input states are filled, and their content is transferred to the appropriate intermediate state. In figure 4.6, this is illustrated by green arrows. All other states are set to zero. Also, all response matrices are calculated and stored in memory during initialization.

In the iteration stage, the algorithm moves through the list of states, and for each state, it performs all the necessary convolutions that point to it from adjacent states. In figure 4.6 this is visualized with black arrows. To be precise, *adjacent* states are defined as those that connect to the state of interest only through response matrices but not other states. This way, a part of the electron distribution can only move forward or backward by at most one state per iteration, allowing for easier debugging. In the case of reflection processes, such as backscattering or magnetic reflection, states are also considered adjacent to themselves. For example, the forward-facing part of a state can feed into its backward-facing part during a single iteration step. To make this all possible, each intermediate state exists twice. One instance is for storing the result of the current iteration - the *current state* - while the other serves as a memory for the spectrum of



**Figure 4.6:** Propagation model structure: Response matrices are shown in blue, intermediate states in yellow, the output state in red, and the initialization injectors in green. The field conditions relevant to the model are shown on the right. The magnetic field is mostly irrelevant in the spectrometer section due to how the individual responses are modeled. Hence, the transition between the source and detector magnetic field can already be included in the transition between states 3 and 4.

the previous iteration - the *memory state*. The memory state holds the necessary input for calculating the spectrum written to adjacent current states. After each iteration, the next one is prepared by overwriting the memory states with the content of the current states and setting the current states to zero.

### Break condition

The transfer along the state chain continues until an output state at the detector is reached, where the distribution content is accumulated and leaves the system. With this method, it is possible to model the part of the electron spectrum belonging to those electrons that travel back and forth multiple times before they are removed from the beamline. To converge to a result, the algorithm relies on the fact that over time the total content stored in intermediate states decreases, corresponding to electrons being removed from the beamline. The iteration is stopped as soon as the total state content falls below a specified tolerance threshold, and the propagation result is passed on to the detector response model. As a default value, the tolerance is set to  $10^{-7}$  times the total starting content. This number is roughly oriented at the required  $10^{-6}$  precision on the predicted spectrum, and it seems to be sufficient for the sensitivity studies presented in chapter 6. In the future, the threshold may need to be optimized for a compromise between calculation speed and the bias introduced due to insufficient convergence. An example of the evolution of the state content and its convergence towards 0 over the iteration is shown in figure 4.8.

### Removers, injectors, neutral elements

Apart from the detector, other locations in the network may also decrease or even increase the total state content during the iteration. For clarity, all responses shown in figure 4.6 may be categorized into injectors, removers, and neutral elements. During the iteration process, injectors add to the total content, removers decrease it, and neutral elements are normalized to leave the content unchanged during the response application.

The most notable remover is the rear wall. The overall normalization of the rear wall backscattering response is  $< 1$ . Therefore, the state content decreases in each response application. The exact number for the removal efficiency depends on the input spectrum, the rear wall material, and the magnetic field settings. Depending on the column density and the lower energy threshold defined by the binning, the source scattering response may also function as a remover. Due to the energy loss associated with scattering, part of the electron distribution can migrate below the energy threshold and is therefore removed. However, due to the effect of ionization in source scattering, where the impacting electron removes a shell electron from the tritium molecule, the source responses can also increase the state content. Hence, the source response acts as both an injector and a remover.

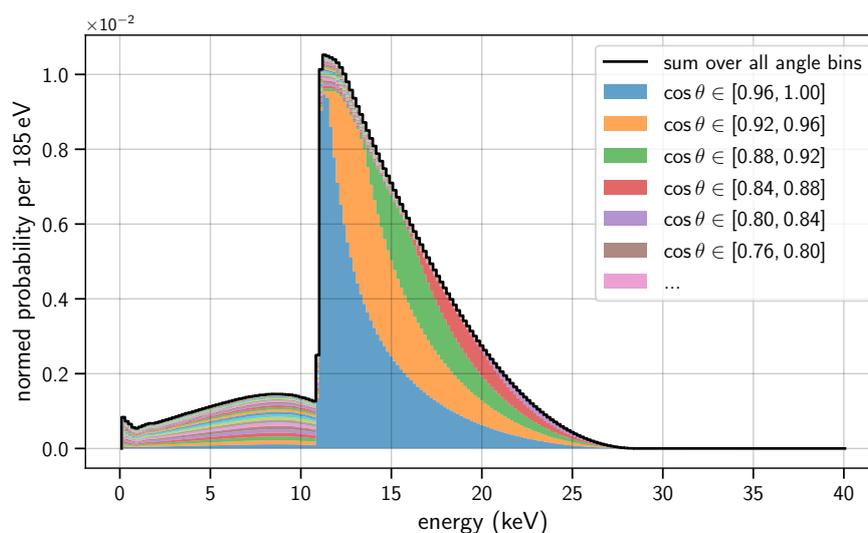
The main injector is the detector back-reflection response, which describes the slow back-reflection effect explained in sections 4.2 and 5.4.1. Some electrons that are backscattered and back-reflected at the detector will cause multiple hits due to the excellent time resolution of the detector, which is necessary to deal with the large event rate.

All remaining responses are related to electromagnetic fields, which transport the electrons along the beamline. Assuming adiabatic transport, no electrons can be lost. Accordingly, the field responses are constructed as neutral elements and conserve the state content. In figure 4.6, this is also the reason why magnetic collimation and reflection responses are used in parallel to each other since a successive application of reflection and collimation would cause minor inaccuracies. For more information about the construction and usage of these responses, see chapter 5.

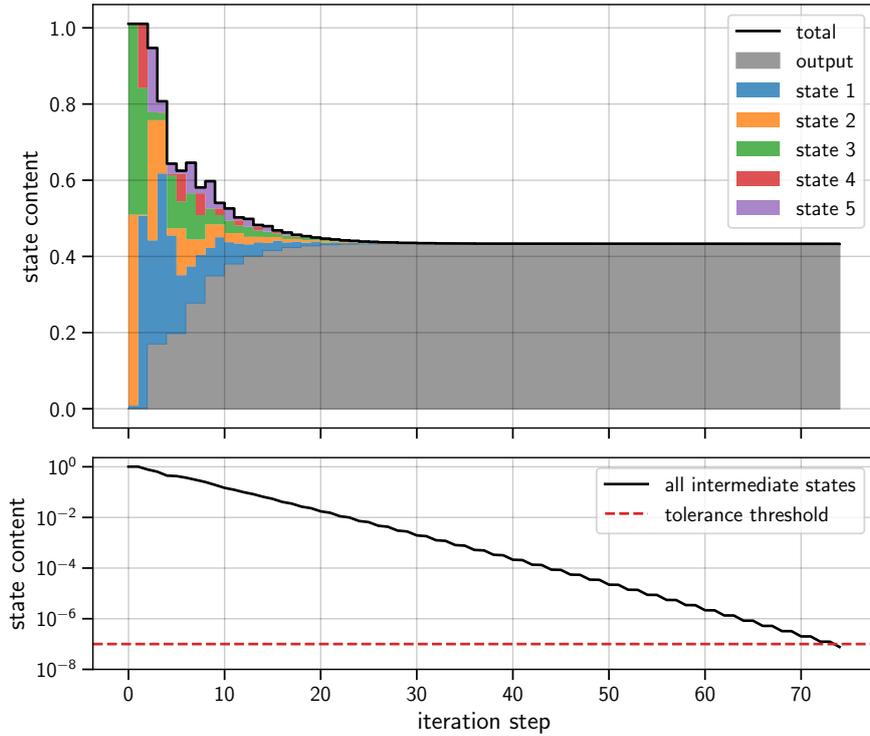
### Propagation output

As mentioned above, the output state represents the distribution of electrons that create distinct hits in the detector. The term “distinct” here refers to the fact that sometimes the electrons that are backscattered and back-reflected toward the detector return so quickly that the trigger cannot resolve them. Their charge deposition thus merges with the deposition of the prior hit. This merging of hits can not be performed with the propagation model and is instead considered during the production of the detector backscattering response.

Since energy and pitch angle are relevant when calculating the proper charge creation in the detector, the propagation output is two-dimensional. Figure 4.7 illustrates the output state by plotting the energy spectrum for each angular bin in a stack plot and the projection of the output matrix where all bins are added together.



**Figure 4.7:** Shown is the energy projection of the 2D output of the propagation model, i. e. the spectrum of distinct electron hits incident on the detector. The colored stack plot shows the contribution of the separate angle bins, with the largest contributions stemming from the angle bins close to the perpendicular incidence.



**Figure 4.8:** Coverage of the propagation model: The top plot shows the content of all states as a function of the iteration step. In step 0, almost all content is contained in states 1 and 2, located at the ends of the WGTS. State 1 at the rear wall also has a small initial contribution due to tritium activity on the rear wall surface, which is here chosen with an amplitude of 1% relative to the source activity. Over the iterations, state content is removed from the propagation network and written into the output or removed by other components, such as the rear wall. The bottom plot demonstrates the break condition. As soon as the total content of all intermediate states falls below the threshold, the iteration procedure is halted.

### 4.4.3 Rebinning

#### Binning interface between propagation and detector

At the time of writing, most response matrices and states of the propagation model are constructed with  $n_E = 100$  bins in the tritium spectrum energy range and  $n_A = 50$  angular bins for  $\cos\theta$  in the interval  $[-1, 1]$ . For the output state and the intermediate state at the detector, the energy range is increased up to about 40 keV and the number of bins is increased to  $n_E = 217$  to account for up to 20 keV of post-acceleration potential. The bin width is kept at 185 eV. However, the detector response is constructed with different energy bins that can also be chosen with smaller widths for increased model accuracy. The bin width can be chosen at either 185 eV, 100 eV, or 10 eV, which means the models will not necessarily interface correctly.

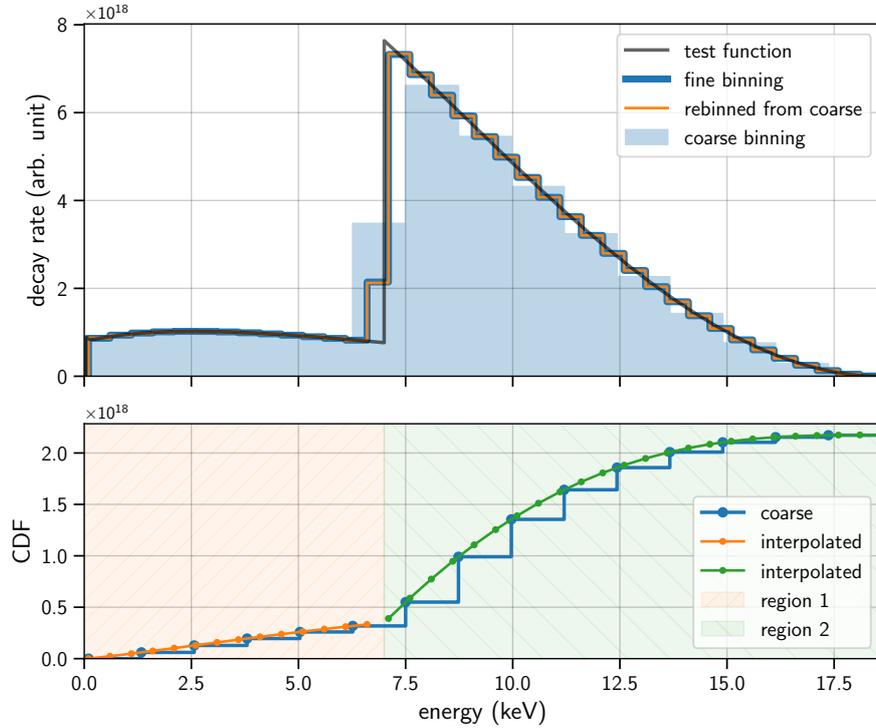
This discrepancy between the detector and propagation model binning primarily stems from the memory requirements of the 4-dimensional response matrices involved in the propagation. The database dimension of the most extensive matrix, namely the side source response, is increased by one to account for the column density parameter. The total memory usage for loading the source database is about 2 GiB. While this size is still manageable and leaves room for increasing the binning, it scales with  $n_E^2$  and  $n_A^2$  and, therefore, quickly will reach file sizes that are more difficult to manage. Response databases with finer binning are in production at the time of writing. A more clever way of loading only part of the database via memory mapping is also an option for optimization that is being investigated. Still, at least a single 4D source response matrix from the database must ultimately be able to fit into memory at a time, and it is yet unclear if the binning can be matched for the entire TRModel.

#### Rebinning via CDF interpolation

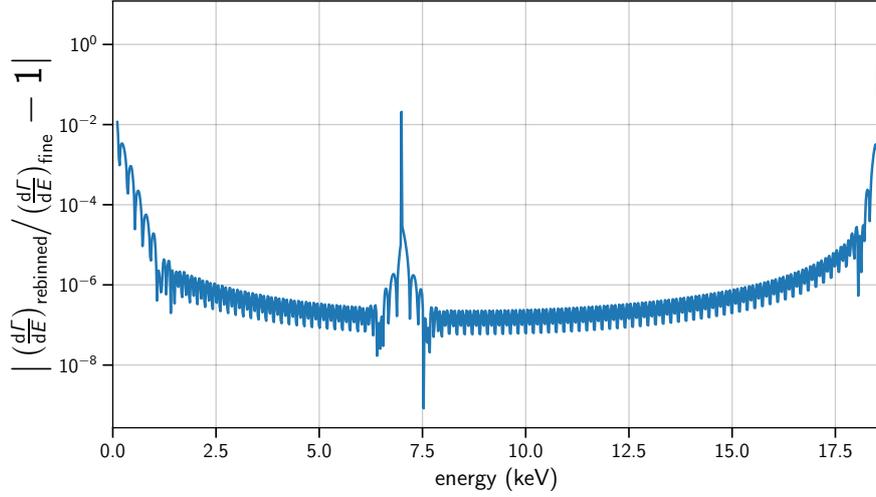
The currently employed method of interfacing propagation output and detector response is rebinning from a coarse to a finer binning. The rebinning method developed for TRModel is an interpolation method that assumes a generic smoothness of the underlying distribution and conserves the content within the intervals of the original bins. Since no method was found in the literature, a new method was constructed based on prior work within the KATRIN collaboration by K. Urban and A. Onillon in earlier versions of TRModel. With this method, the cumulative distribution function (CDF) of the binned spectrum is interpolated with a cubic spline, where the nodes of the CDF are located at the edges of the input bins. The spline is evaluated at the positions of the new bin edges to get the CDF of the rebinned spectrum. The rebinned spectrum is retrieved by calculating the discrete derivative of the CDF, which is the pairwise difference of the rebinned CDF vector.

This method works best for smooth distributions with slight gradients. Since, for the most part, the underlying spectrum follows the shape of the tritium spectrum, the rebinning works well in this application. One caveat is the step introduced by post-acceleration and the retarding potential cut-off seen in figure 4.7. Here, the sharp increase at the step is bound to produce a spline with a large gradient. Due to the spline conditions

for the first and second derivatives at the neighboring nodes, the interpolation creates an oscillation pattern referred to as the Gibbs effect in literature [265]. An attempt to deal with this issue is to exploit that the location of the step occurs at a known position, namely  $E_{\text{pae}} + qU$ . A better result is then achieved by splitting the cubic spline interpolation into two regions: One above and one below the step. This way, the spline is not constrained by a boundary condition at the edge and produces more minor oscillations.



**Figure 4.9:** Demonstration of the rebinning method via CDF interpolation. Larger bin widths for both the coarse input spectrum and the rebinned spectrum are used for increased visibility. The top plot shows the binned spectra, where the fine and the coarse spectra are calculated via numerical integration of the test function. By eye, no noticeable difference is visible between the rebinned and fine spectrum, where the latter serves as a proxy for the ideal outcome. The rebinning bias is quantified and shown with a more realistic binning choice in figure 4.10. The bottom plot shows the underlying principle of cubic spline CDF interpolation with separated interpolation regions left and right of the step in the test function.



**Figure 4.10:** Quantification of the rebinning bias for bin increase from 100 to 1000. The rebinned spectrum is compared to the finely binned truth proxy by calculating the relative difference. The absolute value is taken to display the difference in the logarithmic scale. For this calculation, the step is located at the same position as in figure 4.9, namely at 7 keV. At this position, and towards the edges, the bias increases.

### Bias quantification

Knowledge about the true underlying distribution is needed to quantify the bias introduced by rebinning. Since this is not available for the real propagation model output, a test function is used as a proxy. This test function is constructed from the TRModel tritium spectrum, and a cut-off is introduced with the Heaviside function  $\Theta$ :

$$f_{\text{test}}(E) = \left( \Theta(E - E_{\text{step}}) + \frac{1}{10} \Theta(E_{\text{step}} - E) \right) \cdot \left( \frac{d\Gamma}{dE} \right)_{\text{tritium}} \quad (4.10)$$

Below the cutoff, the back-reflection tail shown in figure 4.7 is emulated by scaling down the tritium spectrum by a factor of 10. This test function is now integrated over two sets of bins: one with narrow bins serving as the ideal rebinning output and one with larger bin width to perform a rebinning. After rebinning the coarse spectrum to the same binning as the truth proxy, they are compared by computing the relative difference. Since the difference is relatively small, the absolute value is plotted with a logarithmic scale in figure 4.10. The plot shows that the bias introduced by rebinning is comparatively small in the central region of interest, where the shape follows the smooth tritium spectrum. The rebinning works less well at the edges and the position of the step because the interpolation oscillations cannot be avoided altogether. Since the bias in the central region of interest is below the  $10^{-6}$  level, it is assumed sufficient for the sensitivity studies presented in this work. For model accuracy improvements and future studies, rebinning should be avoided, or its impact should be propagated to a sensitivity curve. This impact should also be related to the overall bias introduced by binning.

#### 4.4.4 Detector Response

The detector model consists of 3 responses that describe energy deposition, Fano noise, and charge sharing. This section outlines what these responses represent and focuses on how they are employed in TRModel. For more detailed information on the physical effects and the creation of the response matrices, see chapter 5.

##### Energy deposition

The response for energy deposition combines the effects of partial energy deposition due to backscattering and the energy loss in the dead layer. It is calculated based on the method published in [263], where a Geant4 simulation is combined with an empirical depth-dependent collection efficiency for charges in the detector bulk. The simulations and implementation into TRModel are maintained by A. Nava, a member of the group which published the method. With this method, the partial energy deposition due to backscattering directly results from the Geant4 simulation output. Electrons backscattered in the simulation are also stored and used to construct the detector reflection response, which is then included in the propagation model.

The dead layer energy loss is modeled by splitting the simulated silicon absorber into several layers and recording the deposited charge in each. The depositions in each layer are weighted according to a depth-dependent quantum efficiency model in a post-processing step after the simulation. The efficiency model depends on several parameters (see section 5.4.2). The most influential parameter  $\lambda$  is implemented as a free parameter in TRModel. It describes the characteristic length of the exponential efficiency decrease from inside the bulk toward the surface. Having  $\lambda$  as a free parameter is computationally feasible by calculating responses for a range of  $\lambda$  values in advance and storing them in a database.

Since the probability of backscattering and the amount of energy lost in the dead layer depend on the incident angle and the energy of the primary electron, the simulation is performed for several fixed incident angles and energies. The response for each angle bin is linearly interpolated from the simulation results to match the angle bins of the propagation mode. The resulting response matrix has three dimensions: the input angle, the input energy, and the output energy. A demonstration of the application of the response to the output of the propagation model is shown in figure 4.11.

The simulation for fixed incident angles and the linear interpolation can be considered shortcomings of the current implementation since the underlying distribution within the bin is neglected. It is planned to run future simulations with an underlying isotropic distribution within the angle bins and uniform distributions inside energy bins. Furthermore, the angle bins could be matched to those used in the propagation model, eliminating the need for interpolation.

### Fano noise and charge sharing

Fano noise and charge sharing are described analytically, and their responses are computed at runtime according to the methods from Korzeczek [248] and Urban [251]. Both responses are 2D matrices with input and output energy dimensions. The combined effect of charge sharing and Fano noise on the spectrum is demonstrated in figure 4.11. Both effects are also explained further in chapter 5, but their consequences for the spectrum are briefly summarized here.

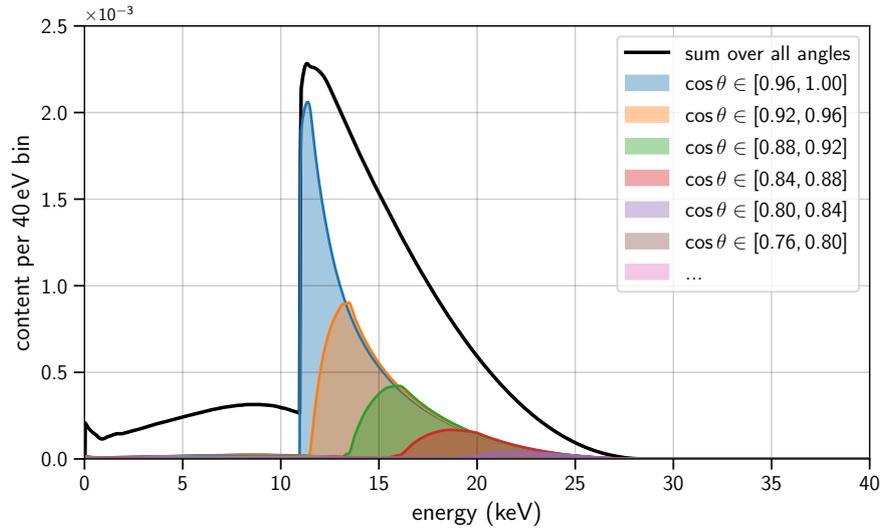
Fano noise introduces an energy-dependent broadening to the spectrum. Its impact is comparatively small and is therefore not shown separately in the figure. The energy-scaling of the broadening width is determined by the energy required for electron-hole pair creation  $\epsilon_{\text{eh}}$  and the dimensionless Fano factor  $F_{\text{fano}}$  [261]. Both are constants for semiconductor detectors, and for silicon, their values are  $\epsilon_{\text{eh}} = 3.65 \text{ eV}$  [278] and  $F = 0.115$  [279]. The broadening width is then calculated with

$$\sigma_{\text{fano}}(E) = \sqrt{E \cdot \epsilon_{\text{eh}} \cdot F_{\text{fano}}} , \quad (4.11)$$

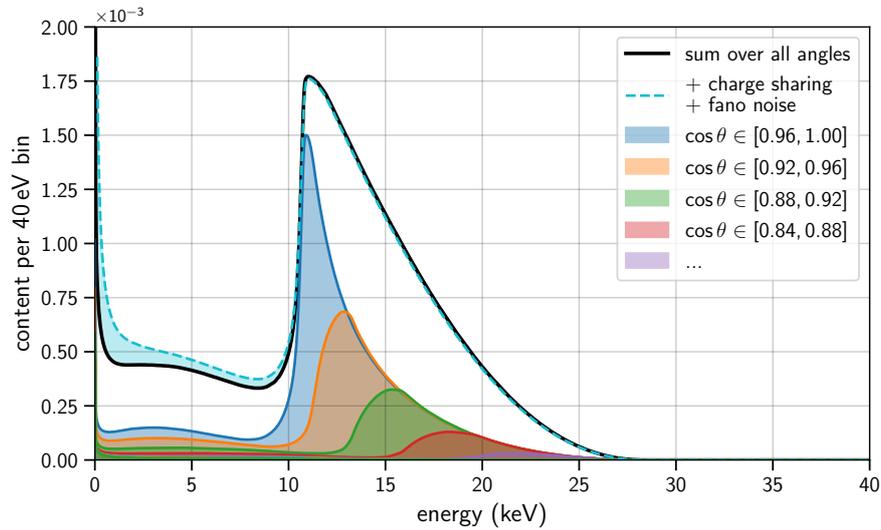
where  $E$  is the deposited energy. For the propagation of experimental uncertainties, TRModel offers the option to apply a scaling factor to the Fano factor.

Charge sharing refers to the effect that the charge created by an incident electron is split between multiple pixels if the hit position is close to a pixel boundary. This effectively redistributes events from high to low energy since some charge is lost to neighboring pixels. And due to the charge received from surrounding pixels, additional low-energy events are recorded if no coincidence rejection is used. The overall amount of charge-sharing events depends on the pixel size and the size of the charge cloud created inside the detector. The charge cloud is assumed to be Gaussian, and its standard deviation is implemented as the  $w_{\text{cc}}$  parameter in TRModel.

A strategy to decrease the impact of charge sharing is searching for temporal coincidences on the order of  $\mu\text{s}$  in neighboring pixels, also called multiplicity detection - see [280] for an in-depth explanation. If such coincidences are caught, the events can be rejected, or the depositions in both pixels can be added to reconstruct the event. Another possible cause for the multiplicity of events is detector backscattering and back reflection. Depending on the cyclotron radius and the point of reflection of a backscattered electron, it can end up in a different pixel in the next hit. Multiplicity from charge sharing has been experimentally tested with TRISTAN detectors [248, 251, 280]. However, a model describing the impact of multiplicity cuts or event reconstruction on the spectrum has yet to be implemented. It is planned to do so for future versions of TRModel and test if the sensitivity for keV sterile neutrinos can be improved with this technique [270].

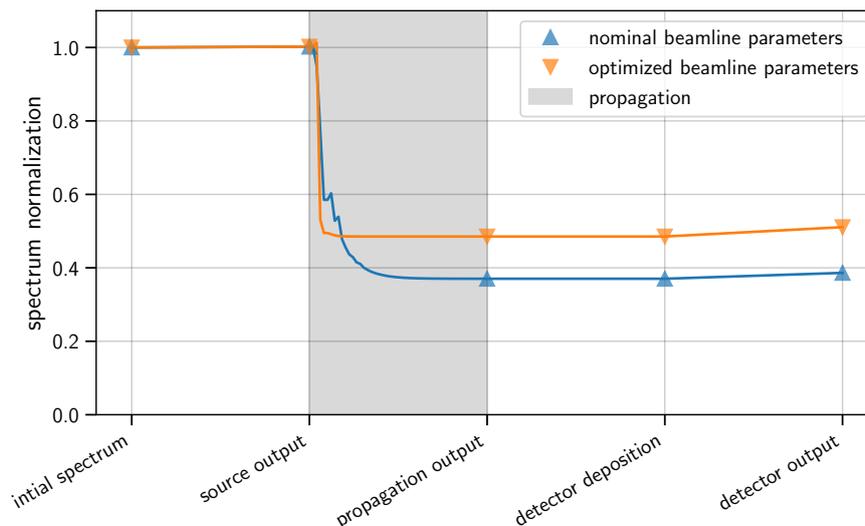


(a) Rebinned input spectrum



(b) After applying detector responses

**Figure 4.11:** Spectrum before and after detector model response application: For the input (a), the 2D propagation output from figure 4.7 is rebinned along the energy dimension to match the detector response. The input can be interpreted as several energy spectra from each angular bin. (b) For this demonstration, the incident angle-dependent detector response is applied to each of these spectra separately and summed up afterward to form the black spectrum. In TRModel, both steps are realized in one convolution of the 2D input with a 3D response. Then the 2D detector responses for Fano noise and charge sharing are applied to the black curve, resulting in the dashed light blue spectrum.



**Figure 4.12:** Evolution of the normalization over the spectrum calculation steps before renormalization. During propagation, the spectrum normalization decreases approximately exponentially, as shown in figure 4.8. The example is calculated for the two sets of parameters used for sensitivity studies in chapter 6. In the optimized case, the increased spectrum normalization is a side effect of the rear wall backscattering mitigation explained in section 5.2.1.

#### 4.4.5 Normalization

Up to this point in the spectrum calculation, the total normalization, i. e. the sum of the spectrum vector, corresponds to the average number of events registered on the detector level per electron emitted in the source. Initially, the spectrum integral is equal to 1 and slightly increased by up to several percent depending on the amount of activity assumed for the rear wall. During propagation, the overall normalization is diminished due to the multi-passage of electrons through the beamline and remover responses, such as the rear wall backscattering response.

The detector responses for energy deposition and charge sharing also have a minor effect on the normalization. Notably, the charge-sharing response will always increase the normalization because charges shared into adjacent pixels are counted as separate events. If coincidence detection were included, that would not be the case. By rejecting multiplicity events, the normalization would decrease instead. Figure 4.12 shows the evolution of the normalization over the model evaluation stages.

Knowing the rate of signals created in each detector pixel for the readout model is crucial. This rate determines the probability of signal pileup and the efficiency loss due to dead time. For this reason, the spectrum is scaled by considering the tritium decay rate, the column density and purity of tritium molecules in the source, and how the source-crosssection maps to one detector pixel. The integral of the spectrum then corresponds to the incident event rate per pixel.

### Incident rate per pixel

The experimentally measured tritium half-life of  $T_{1/2} = (4500 \pm 8) \text{ d}$  is taken from [275]. This translates to a decay constant of

$$\lambda_{\text{T}} = \ln(2)/T_{1/2} = 1.78278596(4) \times 10^{-9} \text{ s}^{-1} . \quad (4.12)$$

This is used to calculate the source activity per magnetic flux unit by multiplying with the number of tritium atoms per  $\text{cm}^2$  of the source and dividing by the source magnetic field  $B_{\text{src}}$ :

$$\frac{a}{\Phi} = \lambda_{\text{T}} \cdot 2 \cdot \rho d \cdot \epsilon_{\text{T}} / B_{\text{src}} \quad (4.13)$$

Here, the number of tritium atoms per  $\text{cm}^2$  is calculated as the product of column density  $\rho d$ , tritium purity  $\epsilon_{\text{T}}$ , and a factor 2 to account for the fact that each tritium molecule consists of two atoms. Isotopic impurities such as HT and DT are considered in the tritium purity.<sup>2</sup> For the implementation,  $\rho d$  is replaced by a new parameter  $\rho d_{\text{glob}}$ , which is introduced to disentangle the normalization from the column density parameter used to describe the magnitude of the source scattering effect. Lastly, the activity per flux is multiplied by the flux coverage of a detector pixel to get the spectrum normalization factor

$$f_{\text{norm}} = \frac{a}{\Phi} \cdot A_{\text{px}} \cdot B_{\text{det}} , \quad (4.14)$$

where  $A_{\text{px}}$  is the area of the pixel and  $B_{\text{det}}$  is the magnetic field at the detector surface.

A number example to conclude: For the TRISTAN detector pixel dimensions, nominal magnetic field settings, and 1% of the nominal column density, the factor is  $f_{\text{norm}} = 1.20 \times 10^6 \text{ s}^{-1}$ . Taking the last spectrum normalization of the blue curve in figure 4.12, this results in a pixel input rate of  $r_{\text{px}} = 46 \text{ kcps}$ . This is close to the target pixel rate of 100 kcps [8] and matches the order of magnitude of other estimations for 1% column density [225].

### Total registered rate

Due to pileup, dead time, and detection threshold, the readout model will also affect the spectrum normalization. Afterward, the spectrum of the entire detector is calculated by multiplying by the number of illuminated pixels. As a default value,  $n_{\text{px}} = 936$  is used, corresponding to the golden pixel selection of TRISTAN Phase-1 as shown in section 3.4.4. This simplified normalization scheme assumes that each pixel's conditions are the same. This is likely not a valid assumption for an accurate model due to spatial inhomogeneities of beamline parameters and pixel areas. An alternate approach for accurately predicting the spectrum is to evaluate the spectrum for each pixel separately and combine all spectra at the end. A method also used for the  $^{83\text{m}}\text{Kr}$  spectrum analysis at KATRIN can be applied to decrease computational effort. Namely, pixels can be grouped into patches with similar conditions, given that the inhomogeneities within one patch are sufficiently small [281, 282].

<sup>2</sup>This matches the tritium purity definition from [222].

#### 4.4.6 Data Acquisition Model

Now that the spectrum is appropriately scaled to represent the rate per pixel, the readout or data acquisition model can be applied. This includes the effects of the signal pileup, electronic noise, the detection threshold, dead time, and the mapping to the final binning, i. e. the calibration. Detailed information about the modeling of these effects is given in section 5.5. The impact of ADC-nonlinearities on the spectrum is studied outside the scope of TRModel with a specialized code called TRAINS [283]. More information on this topic can be found in section 5.5 and also in [252, 284].

##### Pileup

In each readout channel, signal pulses from consecutive events below a specific time difference pile up and are read out as an event with increased energy. Most of these events are discarded with pileup rejection mechanisms, but a fraction is indistinguishable from regular events for the trigger. These so-called unresolved pileup events lead to an additional spectrum component that extends above the endpoint.

The model for the pileup spectrum is based on an earlier work [253]. It is demonstrated that the unresolved pileup can be modeled with a self-convolution of the spectrum. The description is valid for a digitization-based readout system with pileup rejection mechanisms. TRModel realizes this with a numerical convolution of the binned spectrum. Higher pileup orders like triple- or quadruple-pileup, even up to an arbitrarily high order specified by the user, are also included in the calculation. The pileup spectrum is scaled with an analytical factor that depends on the pulse pair time resolution  $\tau_{\min}$  and the incident rate per pixel. Here, the rate is assumed to be Poissonian, i. e. the events are uncorrelated in time. In the future, an extension of the model for a non-Poissonian component may be necessary since the effect of detector back-reflection introduces a time correlation between events.

Apart from the rate-dependent scaling factor, a supplementary scaling factor  $A_{\text{pu}}$  is introduced. Per default, it is equal to 1. In sensitivity studies, it is treated as a free nuisance parameter to ensure the pileup spectrum is always scaled appropriately concerning the null hypothesis. See section 6.1 for more information on the sensitivity estimation procedure.

##### Electronic noise

Electronic noise on the readout waveform causes fluctuations in the extracted pulse height. These fluctuations are non-negligible and decrease the energy resolution. Assuming white noise, the impact on the energy spectrum can be modeled as a Gaussian broadening. TRModel constructs a response matrix where the matrix elements are computed with an analytical integration of a Gaussian kernel over the width of the energy bins. The Gaussian's width  $\sigma_{\text{en}}$  is implemented as a parameter.

### Energy threshold

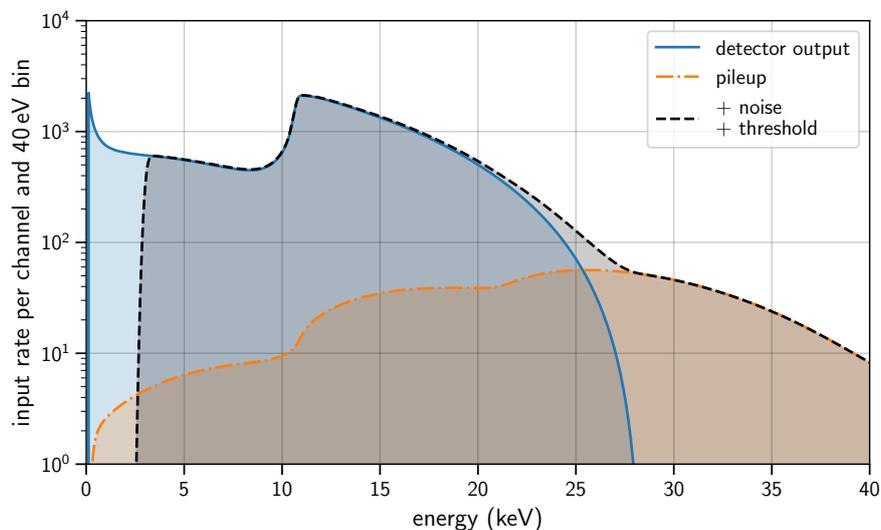
Events with energies below a certain threshold cannot be detected since their signal is not sufficiently distinguishable from noise. This leads to a lower cutoff of the spectrum at approximately 1 to 2 keV [248]. The trigger condition of the DAQ system determines the shape of this cutoff. Notably, the cutoff is not sharp but exhibits a specific width since the trigger condition is decoupled from the energy filter. The noise level and the trigger settings define the width and position. In TRModel, the cutoff is implemented as the cumulative distribution function of a Gaussian, where the width and position of the Gaussian are free parameters.

### Dead time

After applying the threshold response, the spectrum vector sum represents the channel's total rate. This is necessary to calculate the dead time efficiency loss  $\epsilon_{\text{dead}}$ , which depends on the trigger rate. Neglecting the effects from detector back-reflection, the events can be assumed as uncorrelated and the efficiency as energy-independent. Consequently, the dead time response simplifies to a scalar factor

$$1 - \epsilon_{\text{dead}}, \quad (4.15)$$

which normalizes the spectrum such that its integral equals the detected rate in the channel. The calculation of the dead time efficiency is based on a previous work [253] that characterizes the dead time of a readout system used for prototype detectors.



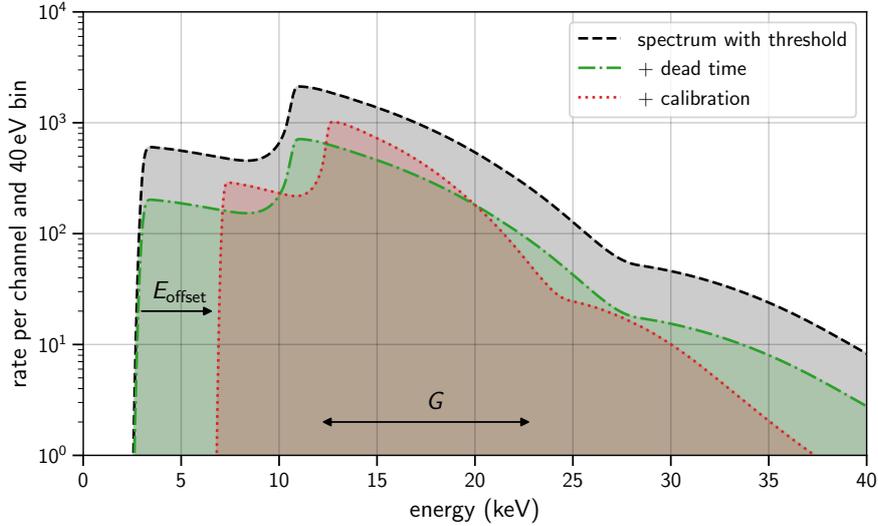
**Figure 4.13:** Demonstration of data acquisition effects: The input spectrum (blue) is convolved with itself to calculate the pileup spectrum (orange). The detection threshold cuts off low energetic events, and the electronic noise broadens the energy resolution (black).

## Calibration

Up to this point, the energy bins used in TRModel do not necessarily correspond to the ones used in measurements. For instance, the latter may be defined by the ADC channels of the data acquisition system, which are translated to energy with a separate calibration measurement. Accordingly, in the final step of the data acquisition model, the spectrum is mapped to a set of output bins specified by the user. Simultaneously, a linear calibration is applied, i. e. the input energy is multiplied with a gain factor  $G$ , and an energy offset  $E_{\text{offset}}$  is added. This stretches and shifts the spectrum along the energy axis. Since the spectrum is binned, it is necessary to calculate how this translates to the migration of contents from input to output bins. For this, each output bin is mapped to the input bins under consideration of gain and offset, and a response matrix is constructed by calculating the overlap of each pair of bins. In sensitivity studies, the parameters  $G$  and  $E_{\text{offset}}$  can express uncertainties on the energy calibration.

This concludes the spectrum calculation procedure for one of the detector pixels. As mentioned in section 4.4.5, the spectrum is now multiplied by the number of pixels to arrive at the final output of the differential spectrum calculation. The spectrum is shown in figure 4.4 at the beginning of section 4.4.

TRModel further offers the options of adding a flat background component to the spectrum and scaling the spectrum with a global amplitude factor  $A_{\text{glob}}$ . The scaling factor is primarily used as a free nuisance parameter in sensitivity studies to account for the uncertainty of the total detected rate.



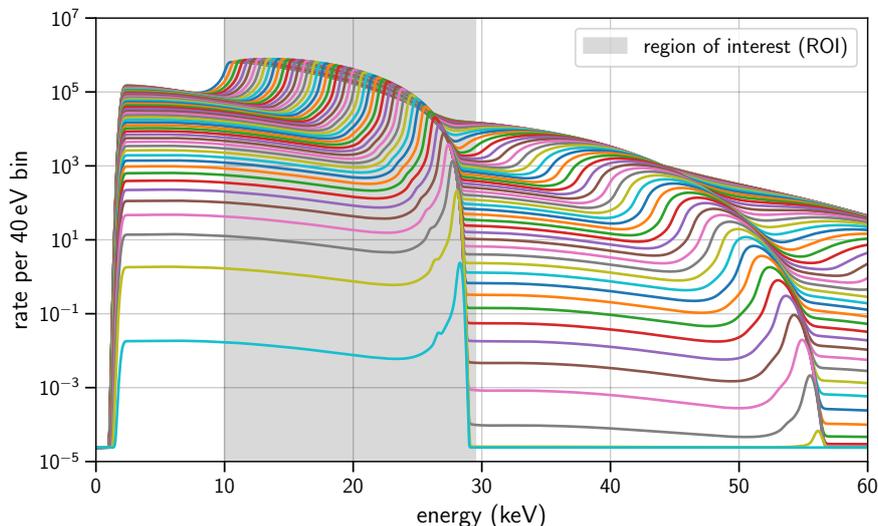
**Figure 4.14:** Impact of dead time and the calibration response on the spectrum. The dashed black curve equals the corresponding curve from figure 4.13. The deadtime efficiency scales the rate from the input to the output rate of the readout system, while the calibration gain factor  $G$  and the offset  $E_{\text{offset}}$  stretch and shift the spectrum.

## 4.5 Integral Spectrum Calculation

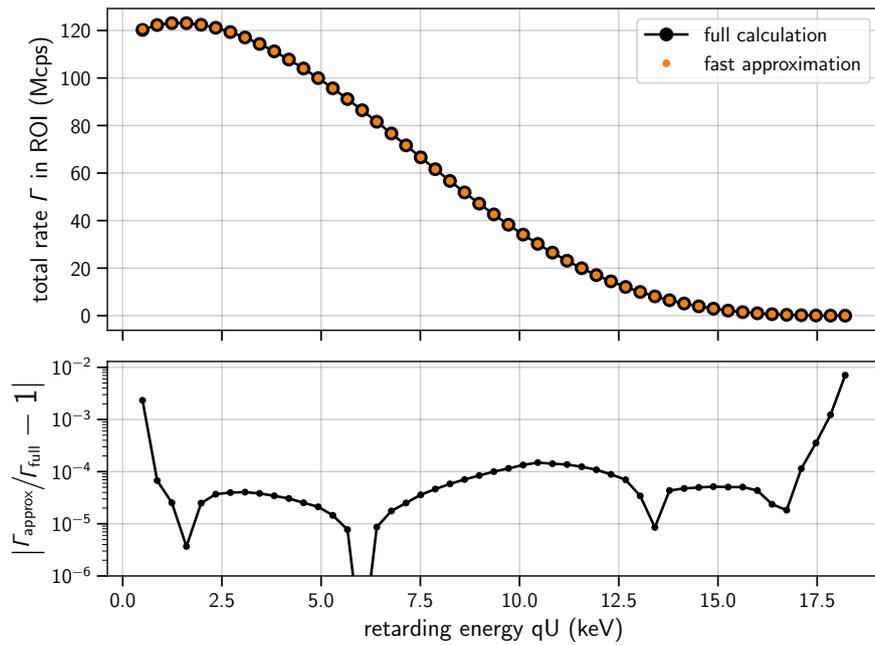
Starting from the differential spectrum, the integral spectrum calculation is comparatively straightforward. Emulating the measurement procedure, the differential spectrum is evaluated for an array of retarding energy setpoints. Each differential spectrum can then be integrated over a specified region of interest, resulting in the measured rate at each retarding potential. Figures 4.15 and 4.16 show an example of this procedure.

A crucial characteristic of the integral model evaluation is that the retarding potential is applied early in the calculation during the time-consuming propagation step. Consequently, almost the complete differential spectrum calculation is performed for each  $qU$  setpoint, and practically no cached intermediate results can be used. For that reason, the integral spectrum calculation is comparatively slow. For example, it can take up to several minutes on the KATRIN server with 80 computing threads.

An approximation is devised to improve this, making the integral model faster and more suited for sensitivity studies. The spectrum is first calculated using the lowest value of all setpoints. For all other setpoints, the retarding potential cut is applied to the output of the propagation model instead of during the propagation. This allows the utilization of the cached result from the first evaluation, thereby increasing the calculation speed significantly. However, the approximation is only reasonable if detector back-reflection can be neglected. As shown in figure 4.7, back-reflection will create a tail below the retarding potential cutoff. This tail is completely removed with the approximation method, which leads to a non-neglectable bias of up to 10%. If the back-reflection effect is disabled, the approximation introduces a more negligible bias on the  $10^{-4}$  level, as shown by the relative difference in figure 4.16.



**Figure 4.15:** Shown are the differential spectra for each retarding potential setpoint of the integral model evaluation. The total count rates are calculated by adding all bin contents in a specified region of interest.



**Figure 4.16:** Comparison of the integral spectrum calculated via the full calculation and the faster approximation. Default beamline settings and 1% column density are used. With these settings, the total rate decreases for low retarding potential due to the increasing input rate per pixel and the associated dead time efficiency loss. The bias of the approximation is quantified with the relative difference between the two spectra. The bias is roughly on the  $10^{-4}$ -level but increases near the ends of the spectrum. The approximation is only useful if back-reflection is neglected. With detector back-reflection, the bias increases to 10%, which is not shown here for easier viewing.

---

## Experimental effects

---

Systematic experimental effects will distort and modify the shape of the tritium  $\beta$ -spectrum, thereby obscuring the imprint of possible new physics processes hidden in the spectrum. The previous chapter shows how descriptions of these effects are combined into a holistic model that predicts the measured spectrum. Details about the modeling of each effect, such as the response matrix construction method, systematic parameters, as well as the discussion of validity, are presented in this chapter. While chapter 4 already provides abridged explanations of some effects, this chapter aims to go deeper and, in some aspects, will go beyond what is implemented in TRModel at the time of writing. Only effects that are implemented into TRModel are presented in this chapter. Effects covered by external studies are referenced in this chapter's appropriate locations.

The chapter is structured into sections for source effects, rear wall effects, transport effects, detector effects, and readout effects. This is oriented at the categories given in the systematics overview of section 4.2. The tritium decay spectrum, including atomic and molecular corrections, is explained in section 3.1.

Other influential works discussing the systematic effects of the keV sterile neutrino measurement at KATRIN are the dissertations by Korzeczek [248], Huber [243]. The former details of detector systematics and many methods chosen for TRModel are inspired by or based on this work. The latter provides an overview of the systematics for the full beamline for the integral measurement mode. This work combines the detector and beamline aspects to achieve a differential spectrum description for the KATRIN apparatus with the TRISTAN Phase-1 detector.

## 5.1 Source Effects

### 5.1.1 Electron Scattering in the Source

Inside the volume of the source,  $\beta$ -electrons can scatter off tritium molecules. In the process, the  $\beta$ -electron will lose energy and be deflected in its angle. To calculate the impact on the 2-dimensional energy and angle distribution used in the deep tritium spectrum model, response matrices are constructed with the *TRISTAN Source Simulation* Software [277]. This software was developed within the KATRIN collaboration by M. Slezak, A. Lokhov, and N. Trost based on electron scattering descriptions as they are implemented in the Kassiopeia particle tracking simulation [268]. The approach used for the response matrix construction is a combination of numerical integration of cross-section models to compute probabilities for energy loss and angle changes in a single scattering process. With this information, an iterative convolution emulates repeated scattering along the length of the WGTS.

The software is written in C++ and uses general-purpose GPU computation with the Nvidia CUDA Toolkit [285] for enhanced calculation speed. Specifics about the implementation are not documented in a publication or otherwise. Instead, this section outlines the procedure based on personal communication with A. Lokhov, the interpretation of the source code, and in part on the information provided in Trost [286].

#### 5.1.1.1 General Concepts

The gas inside the WGTS primarily consists of  $T_2$ , and the other hydrogen isotopologues DT,  $D_2$ , HT, and  $H_2$  are also present in trace amounts [5]. However, the source scattering simulation assumes that the gas consists of pure  $H_2$  based on the availability of cross-section data and models from Kassiopeia, solely adjusting the models to account for the increased mass of  $T_2$  compared to  $H_2$ .

Three types of electron impact scattering processes are relevant for molecular tritium and electrons in the keV energy range: Elastic scattering, electronic excitation, and ionization. On energy scales below 1 eV, rotational and vibrational excitations are also possible [287], but these are irrelevant for the deep spectrum model. Ionization is the dominant effect for the deep spectrum model because of two factors. For one, it exhibits the highest total cross-section in the deep model energy range. But even more importantly, it results in the most significant energy and angle changes of all three effects.

#### Scattering probability

The probability for scattering depends on the total cross-section  $\sigma_{\text{tot}}(E)$  of each particular process, where  $E$  is the kinetic energy of the incoming electron. Figure 5.1 shows the cross-sections used in the TRISTAN Source Simulation. For a longitudinal slice of the source, where the length corresponds to a fraction  $\Delta\rho d$  of the total column density, the scattering probability is also determined by the effective column density that the electron experiences on its cyclotron path in the source magnetic field. The effective

column density is calculated by dividing by the cosine of the pitch angle  $\theta$  [221]:

$$(\Delta\rho d)_{\text{eff}} = \frac{\Delta\rho d}{\cos\theta} \quad (5.1)$$

Thus, the probability for scattering in a particular source slice depends on both  $E$  and  $\theta$ . Assuming a fixed scattering rate, the probability of  $n$ -fold scatterings in one slice follows a Poisson distribution

$$P_n(\lambda) = \frac{\lambda^n}{n!} \cdot e^{-\lambda} \quad (5.2)$$

with the expected value

$$\lambda(E, \theta) = \sigma_{\text{tot}}(E) \cdot \Delta\rho d / \cos\theta . \quad (5.3)$$

Consequently, the probability for any number of scatterings  $n > 0$  is

$$P_{\text{scatt}}(E, \theta) = 1 - P_0 = 1 - e^{-\sigma_{\text{tot}}(E) \cdot \Delta\rho d / \cos\theta} . \quad (5.4)$$

It should be noted that the assumption of a Poisson process is only reasonable as long as  $\lambda$  is sufficiently constant between successive scatterings. Hence, energy and angle changes must be sufficiently low, or the source slice must be adequately thin to neglect two-fold and more scatterings.

### Energy loss and scattering angle

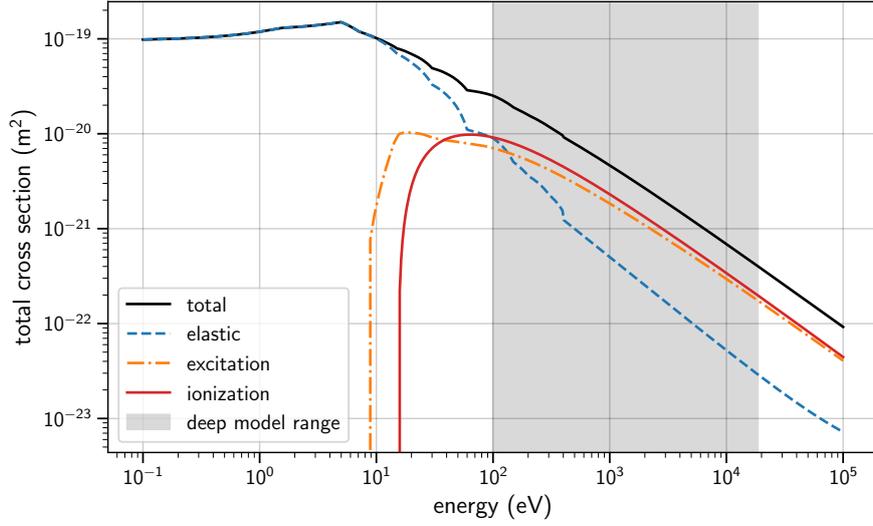
After a scattering process, the electron energy is diminished by the energy loss  $\Delta E$ , and its direction of travel is altered by the scattering angle  $\varphi$ . The probability for the electron to fall into a specific energy and angle interval after scattering is determined by the double differential cross-section

$$\sigma(E, E', \varphi) = \frac{d\sigma}{d\varphi dE'} , \quad (5.5)$$

where  $E' = E - \Delta E$  is the energy after scattering. The integration over the second component of the solid angle is neglected due to the axial symmetry.

The scattering angle  $\varphi$  should not be confused with the azimuthal angle  $\phi$  and the polar angle  $\theta$  of the spherical coordinate system, which is oriented such that the z-axis is the center axis of the source. With the magnetic field also being directed along the z-Axis,  $\theta$  is equal to the pitch angle  $\angle(\vec{p}_e, \vec{B})$ .

Furthermore, the magnetic flux density is assumed to be homogenous along the length of the source. Inhomogeneities in the magnetic field will lead to the trapping of electrons in subsections of the source volume, where they can scatter much more often before escaping the WGTS. The effect of magnetic traps on the WGTS output spectrum is studied in section 5.1.2.



**Figure 5.1:** Total cross-sections used in the TRISTAN Source Simulation, which are adopted from Kassiopeia [268]. Originally, the inputs are based on [288, 289] for elastic scattering, data from [290] further processed by Trost [286] for electronic excitation, and based on [291] for ionization.

### 5.1.1.2 Relevant Scattering Types

#### Elastic scattering

In elastic scattering, the energy loss of the electron only depends on the masses of the collision partners, and the scattering angle [221]:

$$\Delta E = 2 \frac{m_e}{m_{T_2}} E \cdot (1 - \cos \varphi) \quad (5.6)$$

Consequently, the energy loss and scattering angle are fully correlated, which can be reflected with a delta function in the double differential cross-section (5.5). At the highest possible scattering angle  $\cos \varphi = -1$ , this relation yields the maximal energy loss of 6.77 eV, which is much smaller than the bin width of 185 eV used in the deep model (see section 4.4.2). This leads to a bias since the elastic component of the scattering response can only change the angular distribution while virtually neglecting energy loss.

As stated above, the total and double differential cross-section calculations in the TRISTAN Source Simulation are adopted from Kassiopeia. Tracking back, for  $E \geq 400$  eV the total elastic scattering cross-section is taken from Liu [289], who provides a theoretical description using the first Born approximation. With Rydberg atomic units it is given by

$$\sigma_{\text{tot}}^{\text{el}} = \frac{\pi}{k_0^2} (4.2106 - 2/k_0^2), \quad (5.7)$$

with  $k_0^2$  as the electron energy in Rydbergs. For  $E < 400$  eV, the total cross-section is interpolated from experimental data compiled by Trajmar *et al.* [288].

The differential cross-section is also implemented based on a calculation in the first Born approximation by Liu [292]. At low momentum transfers, corresponding to roughly  $E < 300$  eV, the theoretical calculation is corrected with experimental data from [293].

### Electronic excitation

In the case of inelastic scattering, a shell electron of the tritium molecule can be excited to a higher energy level. In a single scattering, the energy loss equals one of the excitation energies. Consequently, the energy loss distribution from excitation is a collection of individual lines, with widths much smaller than can be resolved at KATRIN [286]. In the Source Simulation, the excitation energy loss can be calculated individually or as a weighted sum over all states.

Due to the fixed energy transfer, kinematics also will determine the scattering angle. Similar to elastic scattering, the double differential cross-section (5.5) can be reduced to a singly differential cross-section via integration over a delta function. The differential cross-section is calculated according to the methods from Arrighini *et al.* [294].

The total excitation cross-section is described with the model implemented into Kasiopeia by Trost [286]. It is based on data from [290], which contains 90 states with energies  $E_i \in [10, 16.516]$  eV each with their total cross-section in the range  $[E_i, 1$  keV]. The Trost model extrapolates all total cross-sections to energies above 1 keV. The sum of all extrapolated cross-sections is included in figure 5.1.

### Ionization

For electron energies above the ionization threshold  $E_{\text{ion}}(\text{T}_2) = 15.486$  eV [295], a shell electron can receive sufficient energy in the scattering process to escape from the tritium molecule:



Unlike in an excitation or elastic scattering process, any combination of energy loss and angular change is possible since the final state has three bodies of motion. This leads to a continuum contribution in the KATRIN energy loss function [257]. In addition to ionization exhibiting the highest total cross-section, the possibility for substantial energy loss makes ionization the crucial source scattering effect for the deep tritium model.

The total and double differential ionization cross-sections are calculated according to the semi-empirical model by Rudd [291]. In his publication, Rudd argues that the exchange of scattered and secondary electrons must be allowed. A primary outgoing electron of any particular energy is indistinguishable from a secondary electron of the same energy. Thus, the energy loss extends from zero to the incident energy in a symmetrical manner around  $(E - E_{\text{ion}})/2$ , as visible in figure 5.2.

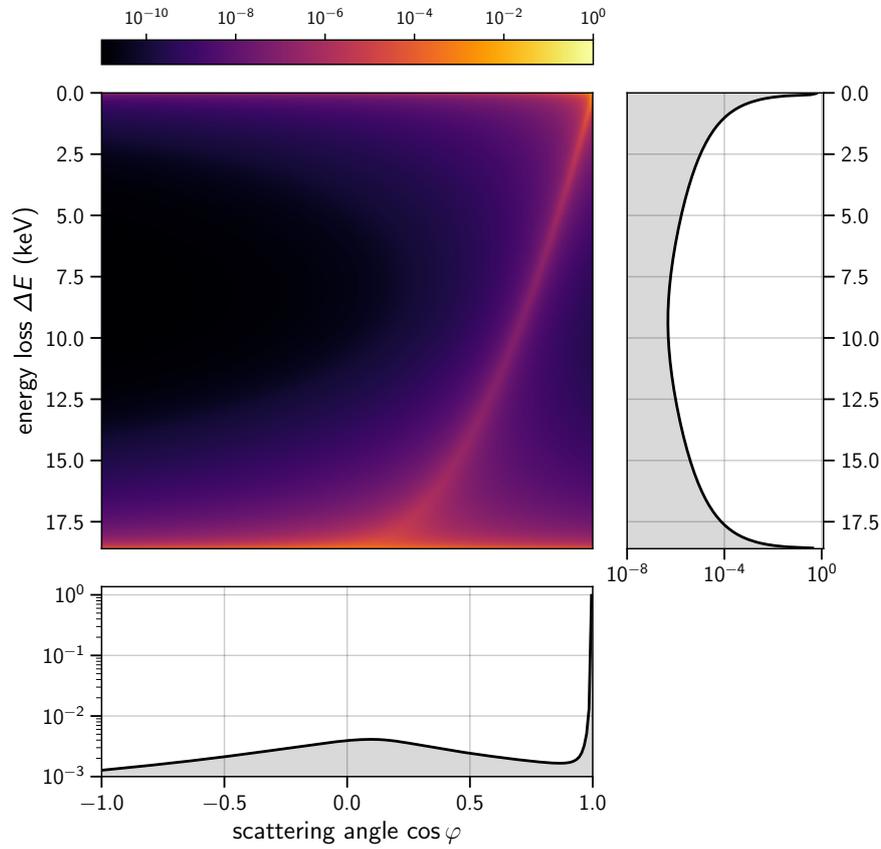
### 5.1.1.3 Response Matrix Construction

The TRISTAN Source Simulation calculates response matrices that describe the impact of source scattering on the electron spectrum using a binned, iterative convolution. For that purpose, the WGTS is subdivided into a series of  $n_z$  longitudinal slices, where each slice contains the fraction  $\Delta\rho d = \rho d/n_z$  of the total column density. According to the reasoning in section 5.1.1.1,  $n_z$  should be chosen sufficiently large, such that the probability for more than one scattering per slice can be neglected. The boundaries at the ends and between slices are associated with electron spectra, represented by two-dimensional matrices with energy and pitch angle axes. The angle is equidistantly binned in  $\cos\theta$  for the reasons given in section 4.4.2. Further, the spectra are split into up- and downstream-propagating components with  $\cos\theta < 0$  and  $\cos\theta > 0$ , respectively.

The first step of the source simulation computes the scattering probability for a transition through one source slice. For this, equation (5.4) is averaged between the bin edges of each energy-angle bin  $(E_{\text{in}}, \cos\theta_{\text{in}})_{ij}$ . The result is a 2D matrix, denoted as  $P_{ij}$ . Likewise, the double differential cross-sections are integrated across all energy-angle bins and averaged over each input energy bin. The result, the so-called double differential map, is stored in a 3D matrix on disk. It encodes the binned energy loss and scattering angle information for each input energy bin. The matrix is denoted as  $D_{ikl}$ , where  $i$  is the input energy,  $k$  is the energy loss, and  $l$  is the scattering angle bin. To combine all three scattering effects in one matrix, their contributions are weighted according to their fraction of the total cross-section sum. Figure 5.2 shows the features of the double differential map for the highest input energy bin.

Using these objects, a spectrum at the boundary of a source slice can be convolved to acquire the scattered spectrum at both ends of the slice - including the possibility for reflection. For this, the spectrum is split into a scattered and an unscattered part according to the  $P_{ij}$  matrix. The unscattered part is left unchanged, and the scattered part is convolved using the double differential map. In this process, it must be taken into account that the scattering angle  $\varphi$  changes both the polar angle  $\theta$  and the azimuthal angle  $\phi$  of the electron momentum. This is considered by multiplying the  $D_{ikl}$  matrix elements with factors that express an integration over  $\phi$ . These factors are also pre-computed for any input, output, and scattering pitch angle combination.

The next step is the iterative convolution, which functions analogously to the propagation model from section 4.4.2. First, the boundary spectra in between the source slices are initialized. Then, the spectra are propagated between the slices, convolving them in each transition. The output spectra are collected at both ends of the WGTS until the remaining content falls below a specified tolerance threshold. The iterative convolution is accelerated with general-purpose GPU computation using the Nvidia CUDA Toolkit [285]. This dramatically improves calculation speed since every convolution is performed via matrix multiplication, a highly optimized operation in graphical processing. On the KATRIN computing cluster, the speed is improved by roughly a factor of 10 when performed with an Nvidia A100 GPU card, compared to a purely CPU-based computation.



**Figure 5.2:** Shown is the double differential cross-section map computed by the TRISTAN Source Simulation [277]. It contains the binned probabilities of scattering angle and energy loss combinations for a traverse of an electron through a specified amount of tritium column density. For this calculation, a column density  $\rho d = 5 \times 10^{15} \text{ cm}^{-2}$  is chosen, which corresponds to 1% of the nominal KATRIN setting and is the anticipated setpoint for the keV sterile search. A binning of 200 bins equidistant in  $\cos \varphi$ , and 50 eV energy bins in the range  $[0, 18.6 \text{ keV}]$  are used. One such map exists for any of the input energy bins. Here, only the map corresponding to the highest input energy bin is shown, so the map represents a single scattering with initial energy  $\approx 18575 \text{ eV}$  and angle  $\cos \theta \approx 1 \Leftrightarrow \theta \approx 0^\circ$ .

For the convolution's initialization, two cases are considered:

- **Init** response: Electrons originating in the source

All slice boundaries are initialized with a monoenergetic spectrum. The spectrum is flat in  $\cos\theta$  in the angular dimension, representing isotropic emission (see appendix A.1). The convolution is repeated for each input energy bin to construct the so-called init source response. This three-dimensional matrix maps each input energy to a 2D output spectrum.

- **Side** response: Electrons entering the source from the side

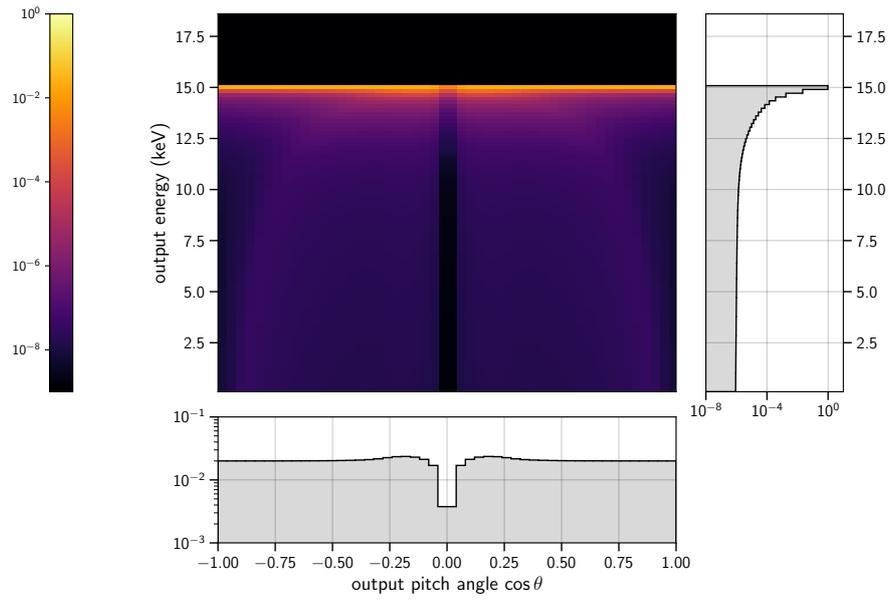
In this case, only the source-facing component of the boundary at the upstream end of the source is initialized. This time the spectrum is monoenergetic and monoangular since no prior assumption about the angular distribution can be made. The convolution is repeated for each input energy and downstream facing angular bin, which results in a four-dimensional matrix that maps each input bin to a 2D output spectrum. This matrix is called the side source response.

If  $n_E$  and  $n_A$  denote the numbers of energy and angular bins, the resulting init source response corresponds to a matrix of dimension  $n_E \times n_E \times n_A$ , where the first dimension is the input energy. The side response has an additional angular input dimension, for which only the forward-facing bins with  $\cos\theta > 0$  are needed. For the output angular binning, the full range is required due to the possibility of reflection. Hence it has the shape  $n_E \times \frac{n_A}{2} \times n_E \times n_A$ .

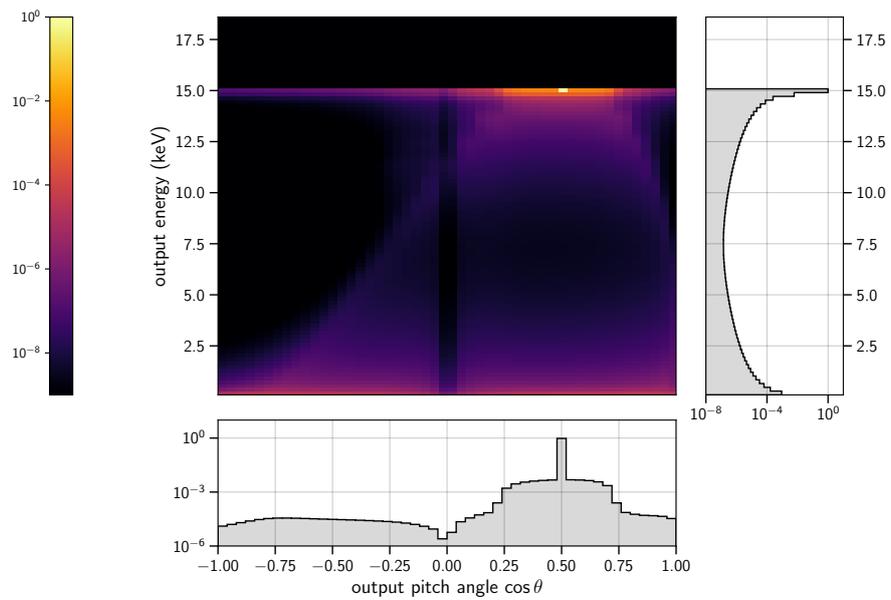
Figure 5.3 demonstrates the general features of the response matrices by showing their output spectrum for a single input bin. In both cases, matrix elements above the input energy are zero because the scattering of fast electrons can only lead to energy loss.

For the initial scattering, the resulting distribution is approximately isotropic near  $\cos\theta = -1, 1$ , i. e. for axially aligned electron momentum. These electrons are subject to less scattering since they encounter fewer tritium molecules on their cyclotron path, corresponding to a lower effective column density (see equation (5.1)). Electrons near  $\cos\theta = 0$ , however, experience a higher effective column density, and the resulting angular distribution deviates from the uniform shape. Furthermore, an electron with a perpendicular pitch angle cannot escape the source volume as its cyclotron path would describe a perfect circle. Consequently, the histogram's underlying distribution must equal zero at  $\cos\theta = 0$ .

The image shows a clear peak at the input bin in the side scattering case, which can be interpreted as predominantly unscattered electrons. Scattering with energy loss and angular change below the bin width, primarily for excitation and elastic scattering, cannot be resolved. Notably, the projection on the energy axis resembles the energy loss from figure 5.2. This matches the expectation since it is analogous to KATRIN's energy loss function measurement method [257], where electrons from the rear section electron gun are sent through the WGTS. As for the init response, the figure shows a drop in the bins around  $\cos\theta = 0$  because scattered electrons with pitch angle  $\theta = 90^\circ$  cannot escape the source volume until their angle is changed through another scattering.



(a) Init source scattering response



(b) Side source scattering response

**Figure 5.3:** Demonstration of the source scattering response matrices for  $n_z = 20$  and a column density of  $\rho d = 5 \times 10^{15} \text{ cm}^{-2}$ , corresponding to 1% of the nominal KATRIN setting. Shown are the two-dimensional outputs of the source responses for a fixed input bin and their projections. In the case of the init response (a), there is only the input energy ( $E_{\text{in}} \approx 15 \text{ keV}$ ) since the distribution along the angular axis is assumed as isotropic. For the side source response (b), input energy and angle bin are fixed ( $E_{\text{in}} \approx 15 \text{ keV}$ ,  $\cos \theta \approx 0.5$ ).

#### 5.1.1.4 Integration into TRModel

As described in chapter 4, the TRISTAN Source Simulation is used to create response databases that can be loaded in TRModel. The impact of source scattering on the tritium spectrum shape is shown in figure 5.5.

#### Column density as free parameter

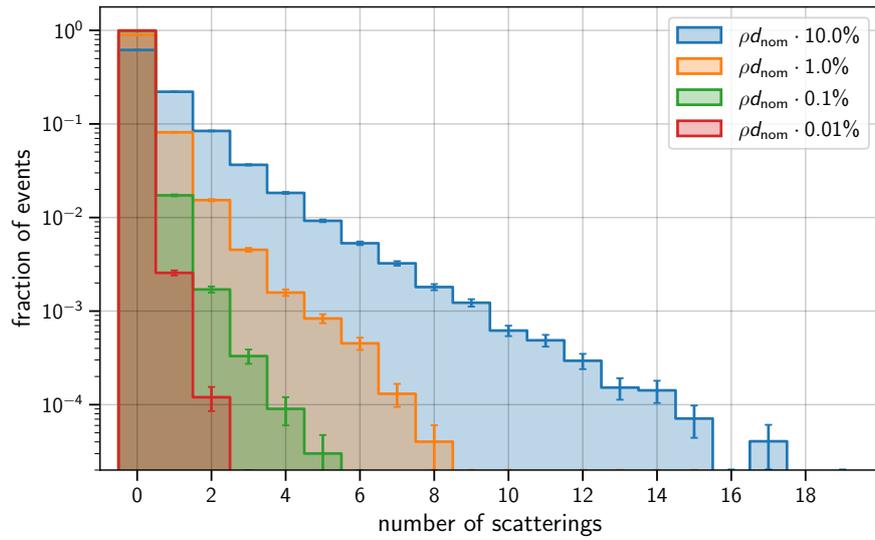
To enable sensitivity studies concerning the impact of scattering, parameters that carry systematic uncertainties must be variable. At the time of writing, only the column density  $\rho d$  is adjustable in the Source Simulation software. Other significant sources of uncertainties are the total and double differential cross-sections that the Source Simulation is based on. However, these are not adjustable, and further software development is required before the corresponding systematic parameters can be included in TRModel.

For the TRModel source database, the Source Simulation is performed multiple times with the settings listed in table 5.1. The concept here is to calculate the source scattering responses for a number  $n_{\rho d}$  of different column densities around a reference column density setpoint  $\rho d_{\text{ref}}$ . The response matrices belonging to one reference value are combined into one large matrix and stored in a Python-compatible binary file representing a single database. A total of eight databases are created to account for init and side responses for the listed  $\rho d$  setpoints.

During the initialization of TRModel, specified databases are loaded into memory. Interpolation is utilized to provide response matrices for  $\rho d$  values between the calculated points in the database. To be exact, the entire database matrix is interpolated element-wise along the  $\rho d$  axis with cubic splines. Tests were conducted by comparing interpolated matrix elements to those from a direct calculation, which showed that the bias from interpolation is well below the  $10^{-7}$  level when cubic interpolation is used. Linear interpolation, on the other hand, would result in a stronger bias at the given spacing of  $\rho d$  values.

**Table 5.1:** TRISTAN Source Simulation settings used to create the TRModel source databases. The column density is given in percentage values of the nominal column density  $5 \times 10^{17} \text{ cm}^{-2}$ .

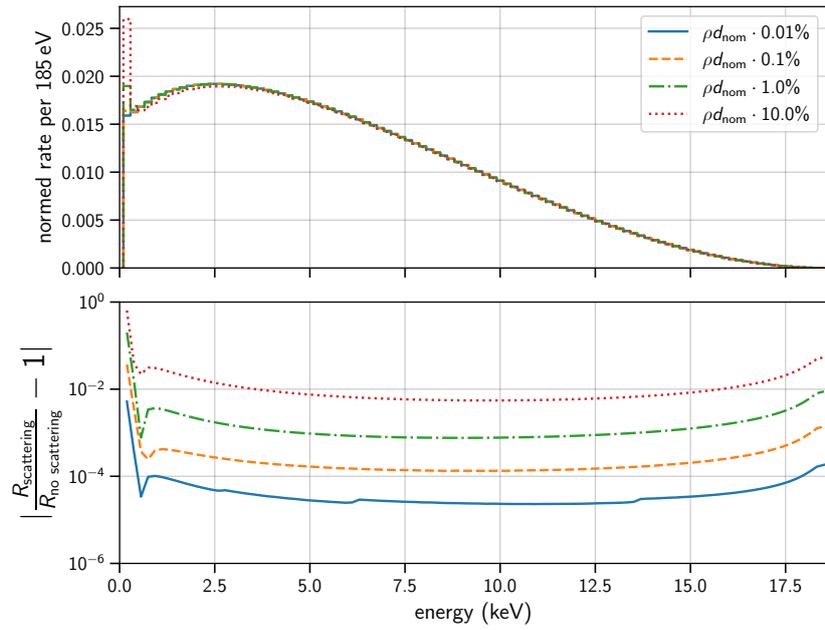
Energy range		100 eV to 18 600 eV
Number of energy bins	$n_E$	100
Number of angle bins	$n_A$	50
Number of source slices	$n_z$	20
Column density setpoints	$\rho d_{\text{ref}}$	0.01%, 0.1%, 1%, 10%
Range around setpoint		$\pm 20\%$ of $\rho d_{\text{ref}}$
$\rho d$ values per setpoint	$n_{\rho d}$	21
Tolerance threshold		$10^{-5}$



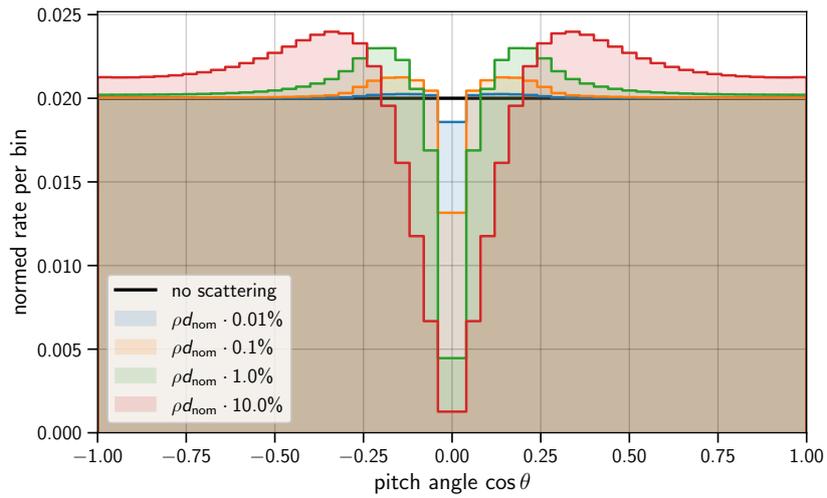
**Figure 5.4:** Number of scatterings from a Monte Carlo simulation for different column density setpoints, where  $\rho d_{\text{nom}} = 5 \times 10^{17} \text{ cm}^{-2}$ . The simulation is provided in the TRISTAN Source Simulation Software [277] and relies on the same double differential map also used in the convolution method. For each setpoint,  $10^5$  electrons are initialized throughout the source.

### Precision limitation

One obstacle to TRModel’s precision is the number of energy and angle bins. Together with the number  $n_{\rho d}$  of values per setpoint, they determine the file size and, consequently, the required memory. The number of elements in the databases is given by  $n_{\rho d} \times n_E \times n_E \times n_A$  for the init response, and by the product of  $n_{\rho d} \times n_E \times \frac{n_A}{2} \times n_E \times n_A$  for the side response. Multiplying this by 8 Bytes yields the file size when double precision floats are used. For the settings used here, the init and side source databases have sizes of 80 MiB and 2.0 GiB, respectively. This is a significant amount of data, considering that memory size increases by a factor of 4 to account for the cubic spline coefficients. There is room to increase the precision since the machines of the KATRIN server offer up to several 100 GiB memory. Still, the memory requirement scales quadratically in  $n_E$  and  $n_A$ , so increases by factors  $< 10$  may already not be feasible. In the future, this issue could be addressed with smarter database handling. For example, only a smaller subset of the databases containing the requested  $\rho d$  value could be dynamically loaded. With a tighter placement of  $\rho d$  interpolation nodes, linear interpolation may also become feasible. This would increase the requirement for disk storage space but alleviate the memory requirement since fewer interpolation coefficients are required.



(a) Energy spectrum



(b) Angular distribution

**Figure 5.5:** Source Simulation result for the tritium spectrum at different column densities. Here,  $\rho d_{\text{nom}}$  refers to the nominal KATRIN column density of  $5 \times 10^{17} \text{ cm}^{-2}$ , and 1% of this value is the anticipated setpoint for the deep spectrum measurement. For the calculation, the binned tritium spectrum is convolved with the init source response and integrated over the angular or energy dimensions for (a) and (b), respectively. In (a), the lower plot demonstrates the impact of scattering compared to the “no scattering” case. Overall, the spectral distortion is low compared to other systematics. However, the angular distribution is affected much more severely, as shown in (b).

### 5.1.2 Magnetic traps

The magnetic field along the KATRIN WGTS is not perfectly homogeneous but exhibits minima in the regions between magnet coils. This is shown in figure 5.6 with a Kassiopeia simulation of the magnetic field along the central axis.  $\beta$ -electrons originating in these regions are magnetically trapped if they are born with large pitch angles (see section 5.3.1). They oscillate back and forth until they eventually escape by altering their pitch angle through scattering or non-adiabatic motion. Due to their long storage times, trapped electrons are characterized by an increased chance of scattering, thereby affecting their energy spectrum and angular distribution at both exits of the source region.

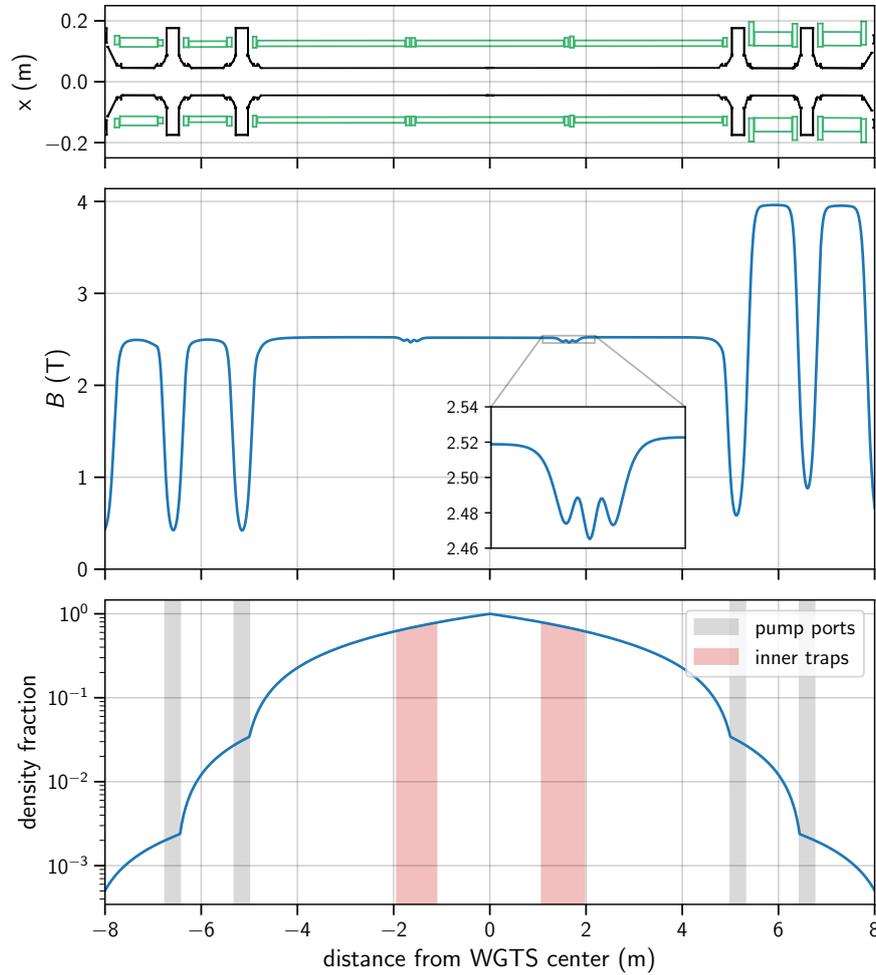
The traps have been characterized in a previous study by Knapp [296] based on Kassiopeia simulations. Similar simulations were used in [243] to quantify the impact on the sterile neutrino search with the KATRIN FPD. In the latter work, the simulation is used to determine energy-dependent correction factors for the integral  $\beta$ -spectrum, and fully uncorrelated uncertainties are assumed on these factors.

This work aims to estimate the spectral component from traps for TRISTAN Phase-1, including the dependence on model parameters. To do so, the source scattering responses from section 5.1.1 and magnetic field transition responses from sections 5.3.1 - 5.3.2 are used. The traps are described with a simplified *bucket model* (see figure 5.7), where the trap is imagined to be of constant depth. The model parameters are the trap's average magnetic field and the fraction of the column density in the trap region. This approach, while less accurate than a Kassiopeia simulation, allows to express an uncertainty via input parameter variation instead of assuming an uncorrelated uncertainty on the trap output spectrum.

#### Trap characterization

The traps can be grouped into two categories: pump port traps and inner traps. The pump port traps have a significant drop in the magnetic field, but the column density fraction within their regions is comparatively small. The inner traps are located at the borders between the 3.3 m long superconducting coils in the WGTS. Compensation coils are installed at these locations to reduce the drop in magnetic field, but some deviation from the otherwise homogeneous value of 2.52 T remains. Only the inner traps are considered in the following since they contain most of the trapped column density fraction as demonstrated in [296].

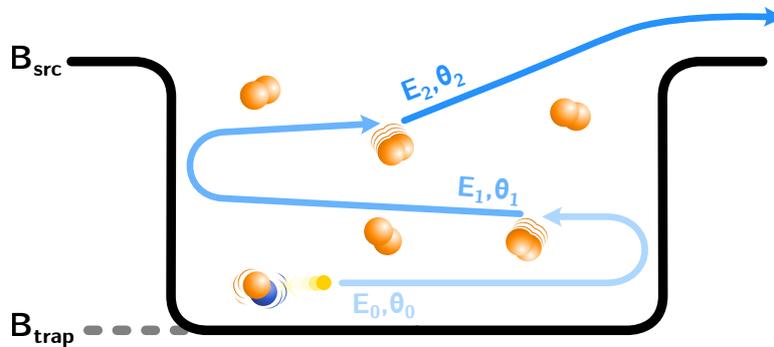
The trap locations and sizes in the longitudinal direction are calculated according to the work by Knapp [296], which provides a procedure for calculating the regions with non-zero trapping probability. For the column density within the traps, a model of the longitudinal density profile is required. Close to the nominal column density, advanced density calculations by Kuckert et al. [229] can be used. However, the anticipated setpoint for the deep spectrum measurement is 1 % of the nominal value, for which no advanced calculations are available. Instead, the method by Reiling [287] and Kellerer [297] is adopted, where the density profile is modeled using linear segments in between the gas inlet and the pump ports. The density fractions at the pump ports are taken from [229].



**Figure 5.6:** Magnetic traps in the WGTS: The top plot shows a cross-section of the beamline elements of the simulation with beam tube surfaces in black and magnet coils in green. The center plot shows a Kassiopia simulation of the central magnetic field in the WGTS. An estimate of the tritium density fraction is shown in the bottom plot. It is modeled with six linear segments based on simulated reduction factors from [229], which is a method adopted from [287, 297].

Integration of the density profile over the length of the inner traps yields fractions of 11.1 % and 11.9 % of the total WGTS column density for the upstream and downstream trap, respectively. To reduce the inherent complexity of the model, only one trap with the average value  $\rho d_{\text{trap}}/\rho d = 11.5\%$  is included in the calculation. Accordingly, the contribution to the source output is scaled by a factor of two.

To obtain a representative value for the magnetic field in the traps, the simulated central magnetic field is averaged over the length of the traps. The longitudinal density profile is used as an averaging weight to account for the fact that the density is not constant along the trap length. The weighted standard deviation is also calculated as a measure of uncertainty, even though a Gaussian does not represent the underlying distribution very well. This yields values of  $(2.5014 \pm 0.0024)$  T and  $(2.4991 \pm 0.0033)$  T for the two inner traps, which again is averaged to  $B_{\text{trap}} = (2.5003 \pm 0.0029)$  T. This translates to a value of 0.78 % reduction when compared to the overall source magnetic field  $B_{\text{src}} = 2.52$  T, and a relative uncertainty of roughly 0.01 %.

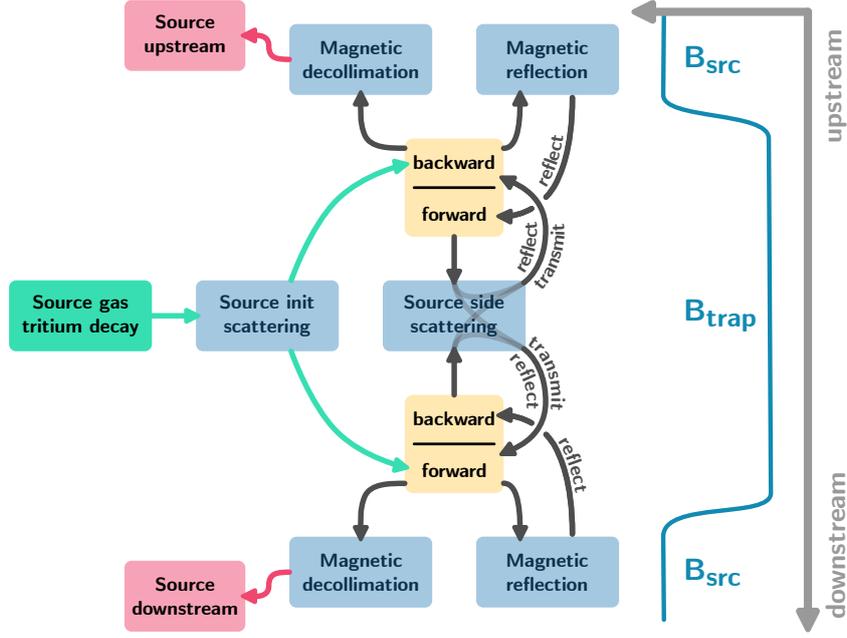


**Figure 5.7:** Bucket model for magnetic traps where the magnetic field inside the trap is approximated as constant. A trapped electron traverses back and forth, and its pitch angle is altered through scattering until it fulfills  $\theta < \theta_{\text{max}}(B_{\text{trap}}, B_{\text{src}})$ .

### Bucket model implementation

The inner traps are modeled with a simplified magnetic field shape over the longitudinal direction. Instead of the “wiggly” shape from the Kassiopeia simulation (see figure 5.6), a “bucket” shape with fixed magnetic fields inside and outside of the trap is taken as an approximation. This is schematically visualized in figure 5.7.

Through the assumption of constant magnetic fields, the source response matrix from the previous section becomes a valid description for the traversal of an electron distribution from one edge of the trap to the other. The source responses and the magnetic field responses from sections 5.3.1 - 5.3.2 are combined in the manner detailed in figure 5.8, which is similar to the propagation model described in section 4.4.2. In the first step, the electron distribution at the edges of the trap is calculated using the init source response. Adopting the terminology from the propagation model, the distribution is stored in intermediate states.



**Figure 5.8:** Response structure for the magnetic trap bucket model. The color coding is the same as in section 4.4.2 with response matrices in blue, intermediate states in yellow, the output states in red, and the initialization injector in green.

The part of the angular distribution below the acceptance angle  $\theta_{\max}(B_{\text{trap}}, B_{\text{src}})$  immediately escapes and is stored in output states. Here, it must be considered that the angular distribution is de-collimated due to the step up in the magnetic field. The other part of the spectrum outside the acceptance is reflected and convolved with the side source response. The latter operation slightly alters the angular distribution, shifting part of the distribution below the acceptance angle and allowing for further escape from the trap to the output states. This process is continued over many iterations until the content of the intermediate states falls below a pre-specified threshold.

### Interface to the propagation model

The calculation time for this procedure will depend on the column density inside the trap. The higher the column density, the faster the calculation since, in each iteration, a larger change to the angular distribution is possible. To achieve a better overall performance, the result of the trap spectrum calculation is performed separately from the propagation model, and its result is cached.

The trap spectrum is then injected into the propagation model at the source output states, where it is added to the regular source output. To conserve the overall normalization of the source output, the regular source output is rescaled with the factor  $(\rho d_{\text{init}} - \rho d_{\text{trap}})/\rho d_{\text{init}}$ , and the initial content of the trap is scaled with the factor  $\rho d_{\text{trap}}/\rho d_{\text{init}}$ .

## 5.2 Rear Wall Effects

### 5.2.1 Backscattering

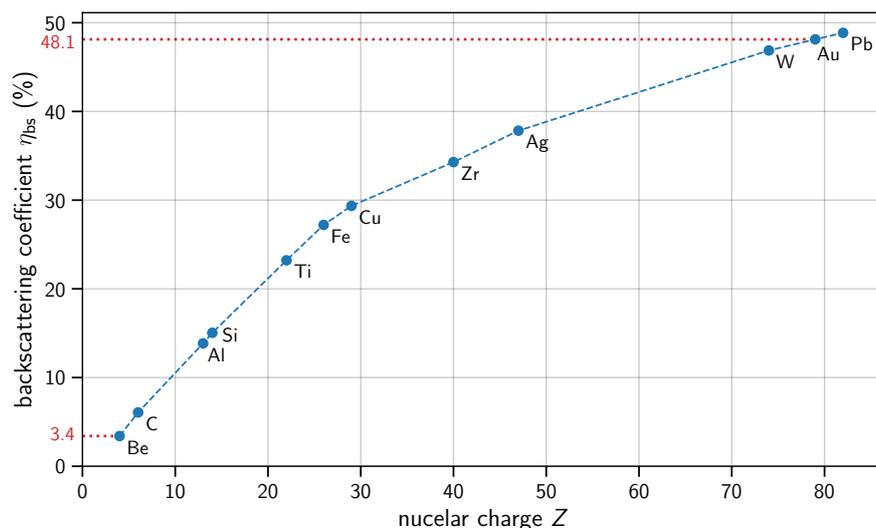
Half of all electrons emitted in the WGTS travel upstream and hit the rear wall. Furthermore, depending on their energy and pitch angle, many downstream emitted source electrons are reflected at the main spectrometer and the pinch magnet. These electrons eventually also make their way to the rear wall. Most electrons hitting the rear wall are absorbed, but in some cases, the electron re-emerges after a single or multiple scattering. Secondary electrons can also be produced, for example, through ionization or Auger de-excitation of an atom. The backscattered electrons re-enter the beamline with modified energy and angle, where they can reach the detector or be reflected on the way. In this manner, a significant fraction of electrons travels back and forth along the beamline multiple times until being absorbed. The rear wall backscattering energy spectrum of  $\beta$ -electrons carries virtually no information on the signature of a sterile neutrino. Due to the effect's magnitude and significant uncertainties on the backscattering simulation, it represents a major systematic effect for the deep spectrum measurement [243, 298]. For TRModel, rear wall backscattering is modeled with Geant4 simulations [266], which were performed by D. Batzler and M. Röllig.

#### Pitch angle and incidence angle

In the propagation stage of TRModel from section 4.4.2, the electron distribution is two-dimensional in energy and cosine of the pitch angle. Assuming the magnetic field lines to be perpendicular to the rear wall surface, the pitch angle equals the angle of incidence. Figure 5.12 shows that this is not necessarily the case, as the field line angle generally will depend on the radial distance to the central axis. At the time of writing, the discrepancy between pitch angle and angle of incidence is neglected in TRModel. Depending on the final experimental realization of the rear wall, a transformation response matrix for the transition between pitch angle and incidence angle would be required. This would enable a spectrum calculation for each radial position and could then be used in pixel- or ring-wise data analysis, similar to the neutrino mass analysis [7]. Apart from the rear wall, this must also be considered for the detector response should the inhomogeneity of the angle of incidence lead to significant changes in the measured spectrum.

#### Backscattering coefficient

The fraction of backscattered and incident electrons is called the backscattering coefficient  $\eta_{bs}$ . For the deep spectrum measurement, the rear wall ideally should have a low backscattering coefficient and thus remove a large fraction of incident electrons from the beamline. However, the KATRIN rear wall is gold-plated for optimal electrical coupling to the source plasma, and gold features a large probability of backscattering due to its sizeable nuclear charge  $Z = 79$ . The dependence of  $\eta_{bs}$  on the nuclear charge is shown in figure 5.9 based on Geant4 simulations. The simulation results in a backscattering coefficient of 48 % for gold, whereas the optimal material beryllium with  $Z = 4$  features a much smaller backscattering coefficient of only 3.4 %.



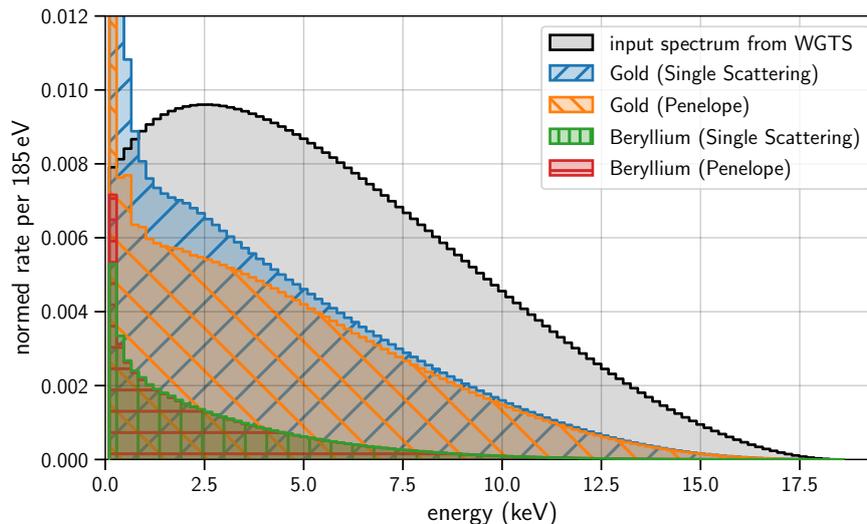
**Figure 5.9:** Backscattering coefficient for different materials as a function of their nuclear charge. Based on Geant4 simulations of 18.5 keV electrons with  $0^\circ$  angle of incidence. Data kindly provided by D. Batzler.

In the earlier works of Förstner and Huber, it was thus proposed to exchange the present gold rear wall with a new one coated with beryllium. While the impact on the measured spectrum was already investigated to some degree in those works, a full differential spectrum and an estimation of the sensitivity impact for TRISTAN Phase-1 were not yet realized. Therefore, one of this work’s central goals is to compare the sensitivity impact for gold and beryllium-based rear walls.

### Response matrix construction

Compared to source scattering, rear wall backscattering is much more complicated due to the number of interactions. This renders the construction of the response by numerical convolution infeasible. So instead, a Geant4 monte carlo simulation has to be used. As mentioned, the backscattering coefficient depends on both the energy and angle of incidence. And since backscattered electrons re-enter the beamline with a modified angle and energy, the corresponding response matrix must be 4-dimensional, mapping each input energy and angle bin combination to a 2D output distribution.

The response matrix is constructed by simulating  $10^6$  incident electrons for each energy and angle bin combination, with the binning matched to the source response. An underlying flat distribution is used within the bins for each primary electron’s energy and  $\cos \theta$ . The simulation geometry is set up in a pure vacuum without electromagnetic fields, with a disk of either gold or beryllium as a target. For each simulation, the outgoing particles are collected in a 2D histogram and normalized by dividing the counts by the number of incident particles. Finally, all simulations are combined into one very large matrix of dimension  $n_E \times \frac{n_A}{2} \times n_E \times \frac{n_A}{2}$  with  $n_E$  as the number of energy bins and  $n_A$



**Figure 5.10:** Backscattering energy spectra directly at the rear wall for the responses available in TRModel (“Penelope” (EM Pen) and “Single Scattering” (EM SS) [299]). The input spectrum is the WGTS output, including the collimation of its angular axis due to the drop in the magnetic field at the rear wall.

as the number of angular bins in  $\cos\theta \in [-1, 1]$ . Here, the number of input and output angle bins is divided by two since the incident and backscattered electrons are restricted to  $\cos\theta < 0$  and  $\cos\theta > 0$ , respectively.

In TRModel, the response matrix is scaled with an amplitude factor  $A_{\text{rwbs}}$  that is used to express the systematic uncertainty on the total amount of backscattering. On the other hand, shape uncertainties of the backscattered energy spectrum are addressed by constructing response matrices with different Geant4 physics lists. The resulting backscattering probabilities and output spectra differ between the lists, as shown in figure 5.10. Responses from all physics lists can be loaded simultaneously into TRModel with independent scaling factors  $A_{\text{rwbs},i}$ . This way, the rear wall response can be chosen as a weighted sum from all physics lists. Uncertainty can be expressed by randomly varying the scaling factors while keeping their sum equal to 1. At the time of writing, responses from simulations with the electromagnetic standard lists “Penelope” (EM Pen) and “Single Scattering” (EM SS) [299] are available.

### Mitigation of rear wall backscattering

The magnitude of the rear wall backscattering component in the differential spectrum measured at the detector has a crucial impact on the sterile neutrino sensitivity. With the TRModel propagation package (see section 4.4.2), it is possible to determine this magnitude since electrons that are backscattered at the rear wall several times are included. This allows for exploring mitigation techniques as proposed in previous studies [243, 298] and uncovering new ones.

Four mitigation aspects are listed in the following, and figure 5.11 demonstrates their reduction of the rear wall backscattering component in the detector spectrum. This work's considerations include modified magnetic field settings and minimal hardware modifications accessible with the current version of TRModel. Exotic ideas involving significant changes to the vacuum chamber, electromagnetic design, or complete redesigns of the rear section are currently discussed within the collaboration. However, these will require further studies and extensions of TRModel. Any modifications must be compatible with other experimental constraints, such as vacuum and magnetic field compatibility. The design criteria for the current rear wall are discussed in [300], although the requirements should be adapted to the new mode of operation and physics goals.

- **Beryllium rear wall:**

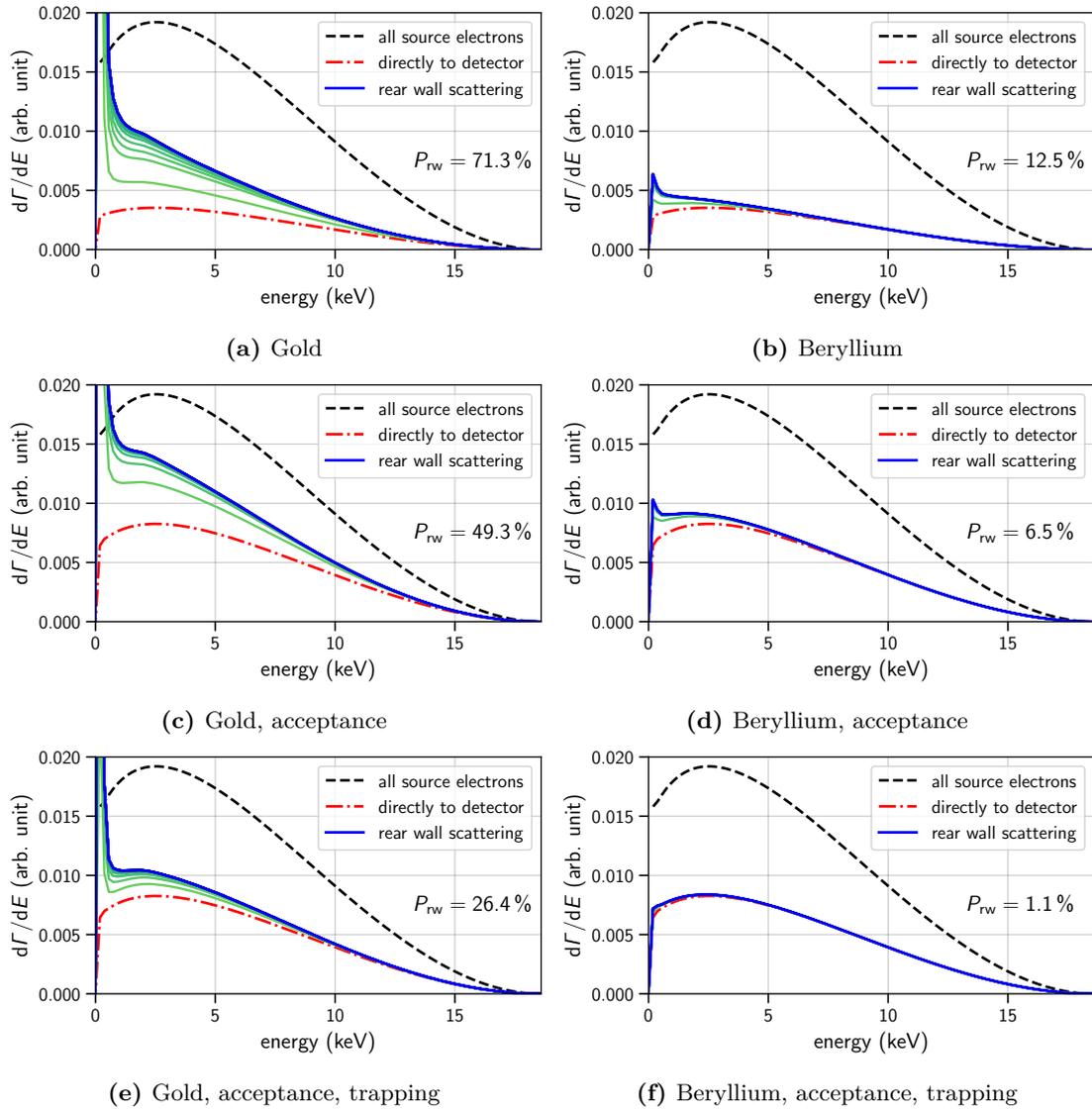
As mentioned above, replacing the rear wall material with one with a smaller nuclear charge can significantly decrease the backscattering coefficient. First proposed in [298], the most optimal choice for backscattering would be Beryllium with  $Z = 4$ . Other important aspects for material choice are vacuum compatibility, electrical conductivity, and adsorption properties concerning tritium molecules and ions. Beryllium, in particular, forms an insulating oxide layer [301] that has to be considered in detail in further investigations.

- **Decreased maximum beamline magnetic field:**

Apart from lowering the probability of backscattering, the relative contribution of rear wall electrons in the detector spectrum can also be reduced by increasing the fraction of events that directly make it to the detector from the source. This can be achieved by lowering the maximum magnetic field along the beamline. This increases the maximal magnetic reflection pitch angle, which raises the acceptance. Ideally, the pinch magnetic field is lowered from 4.2 T to  $B_{\text{pch}} = 2.57$  T, which is sufficiently above  $B_{\text{src}} = 2.52$  T, when accounting for uncertainties (see section 6.2.1). The fields in DPS, CPS, and the pre-spectrometer must also be decreased accordingly while ensuring the fluxtube visible to the detector is still transported.

- **Larger rear wall at magnetic field minimum:**

The magnetic field at the rear wall surface amounts to  $B_{\text{rw}} = 1.24$  T, according to a Kassiopeia simulation. As this is lower than  $B_{\text{src}}$ , backscattered electrons with large pitch angles are magnetically back-reflected to the rear wall. This traps a portion of the rear wall electrons and gives them more chances to be removed from the beamline. Lowering the magnetic field at the rear wall would improve this effect. However, it would also increase the flux tube diameter, mapping the outer field lines to the vacuum chamber surface. This is also proposed as a mitigation option in [298] and [243], but it further complicates the modeling of the backscattering spectrum. Alternatively, as shown in figure 5.12b, a larger rear wall could be installed further downstream within the chamber at the location of the minimum magnetic field, where the surface magnetic field would approximately equal 0.4 T. Furthermore, the field lines would be perpendicular to the rear wall, simplifying modeling since the radial inhomogeneity of the electron incidence angle could be neglected.



**Figure 5.11:** Mitigation of rear wall backscattering in the detector spectrum. Plots on the left/right show the results for a rear wall made of gold/beryllium. The red curve shows the part of the spectrum that falls into the source-to-detector acceptance and travels directly to the detector within the first propagation model iteration. At most, this contribution can be half of all electrons since the other half is emitted toward the rear wall anyway. Green and blue curves add the part of the spectrum that reaches the detector in later iterations, where the dark blue curve is the converged propagation result. The electrons have backscattered at the rear wall at least once for the added contribution above the red curve, and  $P_{rw}$  denotes the integral amount. The first mitigation technique (c,d) is to lower the pinch magnetic field, which increases the source-to-pinch acceptance. The second method (e,f) is to lower the rear wall magnetic field, which improves backscattered electron trapping.

- **Simultaneous reduction of detector and rear section magnetic field:**

Lowering the field in the rear wall chamber only is problematic. However, if the magnetic field at the detector is also decreased, both could be reduced in tandem such that the detector still fully maps to the rear wall.<sup>1</sup> As indicated in figure 5.12b, the outer portion of the WGTS flux tube would then extend past the rear wall diameter. In past studies [145, 302], it was shown that it is beneficial to reduce the detector magnetic field from 2.52 T to approximately 1.0 T to reduce detector systematics. Going further below this value would make modeling more difficult due to increased electron gyroradii and pixel migration in detector back-reflection (see section 5.4.1). With a 1.0 T detector setting, a small decrease of the rear wall magnetic field seems possible, although the gain is much smaller than for the previous bullet point. According to the simulation shown in figure 5.12b, the rear wall surface field can be reduced to 0.32 T with both methods combined.

To further illustrate all the field arguments, it is helpful to disentangle the acceptance into two components: The source-to-detector acceptance and the rear wall-to-detector acceptance. Assuming an isotropic starting angular distribution  $f(\cos\theta) = 1$  in the interval  $\cos\theta \in [0, 1]$  at a magnetic field  $B_1 < B_2$ , the fraction of electrons making it past the field  $B_2$  can be computed as

$$\epsilon = \frac{\int_{(\cos\theta)_{\min}}^1 f \, d\cos\theta}{\int_0^1 f \, d\cos\theta} = 1 - \sqrt{1 - \frac{B_1}{B_2}}. \quad (5.9)$$

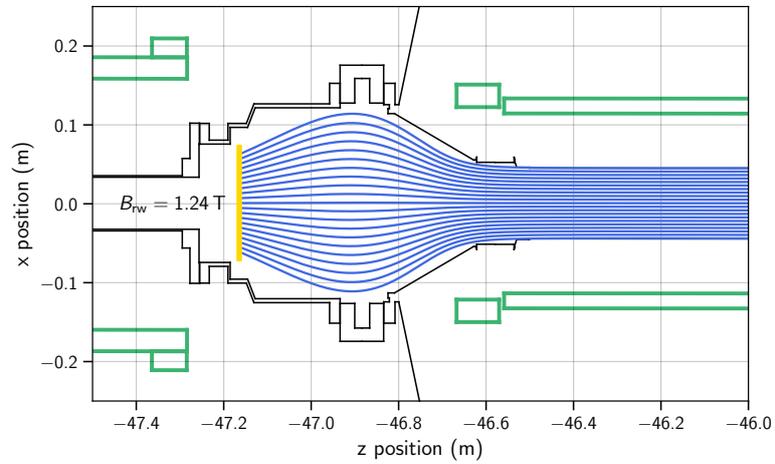
Table 5.2 shows the results of this simplified model for the optimization steps from above. The outcome of this calculation does not represent the real fraction of rear wall backscattering events at the detector because the angular distribution and multiple rear wall hits are neglected. Still, it can be a useful demonstration and a quick way to assess the rough impact of particular field scenarios.

The sensitivity study in chapter 6 combines all optimization steps into one improved beamline operation scenario to reduce the study’s complexity. As shown in figure 5.11, the fraction of rear wall electrons in the detector spectrum is improved by about a factor of 60 when comparing the default scenario and the fully optimized one.

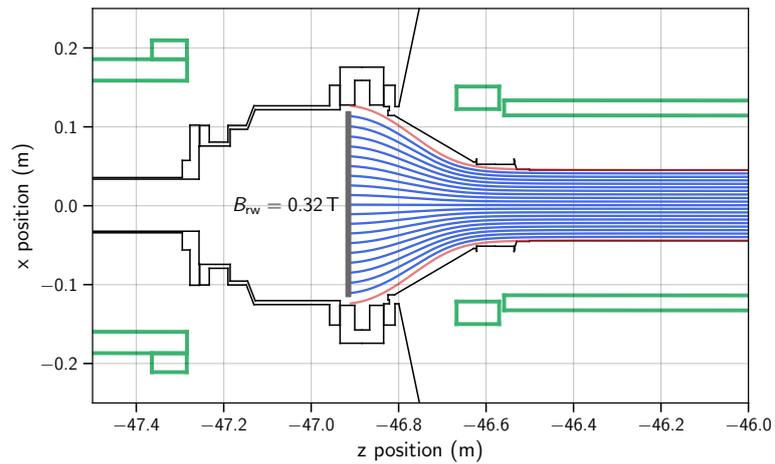
**Table 5.2:** Rear wall and source acceptance fractions for the field optimization steps described above. The numbers are calculated according to equation (5.9).

$B_{\text{rw}}$ (T)	$B_{\text{src}}$ (T)	$B_{\text{max}}$ (T)	$\epsilon_{\text{rw}}$ (%)	$\epsilon_{\text{src}}$ (%)	$\epsilon_{\text{rw}}/\epsilon_{\text{src}}$ (%)
1.24	2.52	4.20	16.0	36.8	43.7
1.24	2.52	2.57	28.1	86.1	32.6
0.32	2.52	4.20	3.9	36.8	10.6
0.32	2.52	2.57	6.4	86.1	7.5

<sup>1</sup>Having the flux seen by the detector map to different materials at different angles would be very challenging to model and could presumably only be treated with a full MC simulation. A way of circumventing this problem would be the construction of a bigger rear wall chamber.



(a) Default settings



(b) Optimized setting

**Figure 5.12:** Rear wall and flux tube optimization: Shown are images of the rear wall chamber and the magnetic field lines from Kassiopeia simulations. Blue field lines map to the detector, while red lines map to the source but not the detector. Magnet coil positions are indicated in green. (a) is the default setting with the gold rear wall currently used at KATRIN. (b) shows an optimized setting where three backscattering mitigation methods are combined: Beryllium rear wall, increased diameter, and lowered magnet currents in the detector and rear sections. Combining the latter two methods reduces the rear wall magnetic field from 1.24 T to 0.32 T, greatly improving the trapping of backscattered electrons.

### **5.2.2 Tritium activity**

There are two aspects leading to the adsorption of tritium on the rear wall: Firstly, ions from the source plasma are guided to the rear wall by the magnetic field. Depending on their charge and the rear wall voltage, they are even attracted to it. Secondly, the residual pressure of neutral tritium molecules is not negligible at the rear wall. Molecular flow simulations by Raspopin [303] show that the pressure at the rear wall surface is expected to be about  $4 \times 10^{-6}$  mbar at nominal source column density.

As tritium is continuously pumped through the WGTS, the amount of deposited tritium accumulates over time as a function of the integral flow rate [304]. The decay of deposited tritium leads to a distinct component in the detector spectrum, which differs from the source gas spectrum shape for several reasons: The electrons can immediately scatter off the rear wall depending on their angle of emission and deposition depth. Also, the final state distribution (FSD) may differ depending on the chemical composition of the deposited tritium, and its decay recoil may be altered. Furthermore, the electromagnetic starting field conditions are different.

#### **State of knowledge on the sorption process**

It is highly desirable to understand how the tritium molecules are bound on the gold surface to model the shape of the  $\beta$ -spectrum emitted from the rear wall. Past studies with the TRIADE setup [305, 306] imply that chemisorption, as opposed to physisorption, is the favored sorption process from the neutral tritium gas component. However, the deposited tritium's true location and molecular composition are unknown. For example, energetic tritium ions could be implanted inside the gold bulk. Tritium could also reside on the surface in molecular or atomic form, or be embedded in larger molecules on the surface, such as hydrocarbon contaminants [259]. If tritium is deposited inside the gold bulk, decay electrons can scatter off the material on their way out. This effect was studied in [307] by modeling the depth profile of tritium concentration based on linear diffusion. However, the depth profile inside the gold surface is challenging to assess, and it is associated with significant uncertainties. Since then, a new rear wall cleaning method has been developed. It was shown that ozone production inside the rear wall chamber effectively removes the accumulated activity [234, 304]. To produce the ozone, the pressure inside the chamber is increased with air, and the rear wall is illuminated with UV light [308]. During cleaning, the observed decay rate follows an exponential decay law, and 99.9% of the activity can be removed over a few days [234, 308]. This has led to the hypothesis that most of the tritium is bound to hydrocarbon contaminants on the surface of the rear wall [234, 259]. For this reason, the model presented here assumes the tritium to sit on the gold surface, while migration into the bulk is neglected.

#### **Magnitude and shape of the rear wall spectrum**

The rear wall spectrum's magnitude is measured between neutrino mass measurement campaigns with integral scans extending down to 125 eV below the endpoint. These measurements yield an amplitude on the sub-percent to percent level - compared to the

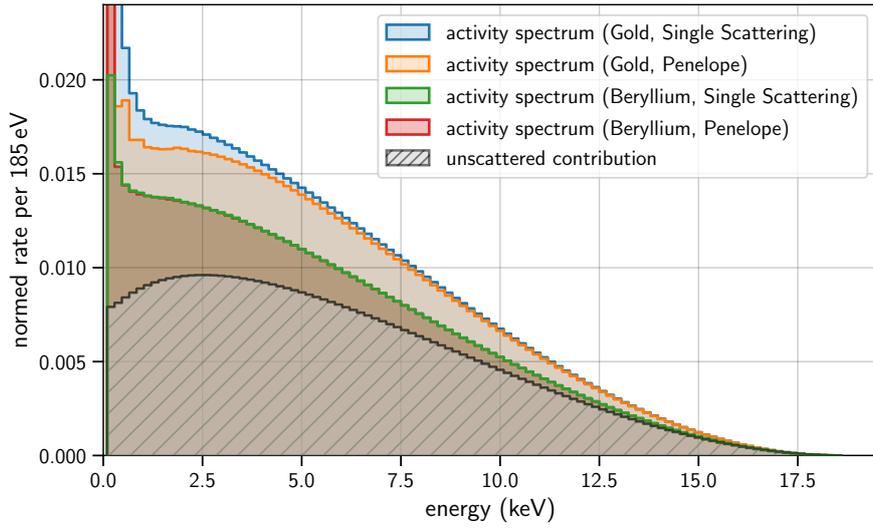
WGTS signal rate - the size of which depends on the time since the last rear wall cleaning [304]. Due to the limited statistics and scan range, information about the shape deep below the endpoint is currently unavailable and must be modeled from first principles instead. To account for the chemical composition of the bound tritium, the spectrum can be convolved with the corresponding FSD [259]. However, the correct FSD for tritiated hydrocarbons is currently unavailable. Furthermore, an energy dependence cannot be neglected for the deep spectrum description, as is the case for the molecular tritium FSD relevant to the WGTS decay spectrum. Due to the unavailability of this energy-dependent FSD, it is neglected at the present state.

### Activity spectrum in TRModel

As stated above, any depth dependence of the deposited tritium has been neglected so far. With the tritium sitting on the surface and decaying isotropically, half of the  $\beta$ -electrons are emitted forward into the beamline, and half will proceed towards the gold. This implies that the total rear wall activity spectrum can be calculated as a superposition of an unscattered and a smaller backscattered contribution. For the backscattered part, the response from the previous section can be convolved with a tritium spectrum that is initialized with an isotropic angular distribution. This results in a slightly different backscattering spectrum than for electrons originating in the source because the source electrons are magnetically collimated toward smaller pitch angles on their way to the rear wall. After adding the unscattered half, the rear wall activity spectrum is injected in the propagation model (see section 4.4.2). At this point, the injected spectrum is scaled with the corresponding amplitude factor  $A_{\text{rwact}}$ , which is the ratio of the rear wall activity and the source activity. The accumulation over time is neglected since the activity can be assumed to be constant when analyzing single measurement runs, which are typically only several hours long. The injected energy spectrum and the final detector spectrum after propagation are shown in figure 5.13. As a future improvement to the model, the initial tritium spectrum can be convolved with an energy-dependent FSD as soon as available.

### Mitigation techniques

Due to the similarity to backscattering, all the field-related mitigation methods from section 5.2.1 can also be applied. Regarding the replacement of the material, the new material should be tested for its tritium adsorption behavior, as was done previously with gold [305]. Ozone cleaning of the rear wall may also be a crucial mitigation tool independent of the substrate. During neutrino mass measurements, ozone cleaning is only performed between major measurement campaigns since it affects the rear wall work function on the sub-eV scale [308]. Such a small energy shift can completely be neglected for the differential spectrum measurement, and a more frequent cleaning procedure could be considered. Lastly, the decreased flow rate in the anticipated 1% column density setting for the deep spectrum measurement (see section 6.2.2) could lead to a slower accumulation of tritium. However, this assumes that the scaling of tritium adsorption with the integral flow rate holds for low column density settings.



**Figure 5.13:** Model for the rear wall activity spectrum. The spectrum is constructed as an unscattered and a backscattered component by relying on the backscattering response matrix. The result is summed over all angular bins for this demonstration.

### 5.3 Transport Effects

This section covers the construction of all response matrices necessary for the electron distribution transport along the beamline's electromagnetic field configuration. All descriptions are based on the adiabatic principle, i.e. a constant magnetic moment is assumed. This allows for a fully analytical description and thus a fast response calculation at runtime. Non-adiabatic transport effects and synchrotron radiation energy loss are so far neglected. For a detailed discussion of these effects, see [243, 260].

The magnetic moment is the first adiabatic invariant in plasma physics, and it is given by [237]

$$\mu = \frac{p_{\perp}^2}{2m_e B} = \text{const.} \rightarrow \frac{p_{\perp}^2}{B} = \text{const.} , \quad (5.10)$$

where  $p_{\perp}$  is the component of the electron momentum perpendicular to the magnetic field  $B$ . With the definition of the pitch angle  $\theta = \angle(\vec{p}, \vec{B})$ , the perpendicular momentum can be written as  $p_{\perp} = p \cdot \sin \theta$ . Further, one can define the parallel and perpendicular kinetic energy components as

$$E_{\parallel} = E \cdot \cos^2 \theta \quad \text{and} \quad E_{\perp} = E \cdot \sin^2 \theta . \quad (5.11)$$

Using these definitions, the squared perpendicular momentum can be expressed as  $p_{\perp}^2 = E_{\perp}(\gamma + 1)m_e$  with the relativistic gamma factor  $\gamma = (\frac{E}{m} + 1)$ . From equation (5.10) then follows

$$\frac{E_{\perp} \cdot (\gamma + 1)}{B} = \text{const.} , \quad (5.12)$$

which is a useful formulation of the invariant. Assuming constant kinetic energy, i. e. when there is no change in the electric potential, the equation can be reduced to

$$\frac{\sin^2 \theta}{B} = \text{const.} \Rightarrow \frac{\sin^2 \theta_1}{B_1} = \frac{\sin^2 \theta_2}{B_2}. \quad (5.13)$$

This implies that an electron will change its pitch angle when transitioning from one magnetic field region to another. The pitch angle will decrease when transitioning to a lower field, which means the angle is collimated. Reversely, the angle is de-collimated when moving to a higher field. Due to the energy independence, all magnetic field responses need to be constructed only for the angular distribution. The responses can then be applied separately to all angular distributions of the 2D electron spectrum, each belonging to one energy bin.

### 5.3.1 Magnetic reflection

When transitioning from a specific magnetic field  $B_1$  to a higher magnetic field  $B_2 > B_1$ , an electron is reflected if it is de-collimated to the point that it changes direction. This is commonly referred to as the magnetic mirror effect. The critical initial pitch angle above which an electron is reflected can be inferred from equation (5.13) by requiring  $\theta_2 = 90^\circ$ . This yields the maximum angle of transmission  $\theta_{\max}$  or alternatively the minimum cosine of the pitch angle:

$$\theta_{\max} = \arcsin \sqrt{\frac{B_1}{B_2}} \Leftrightarrow (\cos \theta)_{\min} = \sqrt{1 - \frac{B_1}{B_2}} \quad (5.14)$$

From this relation, the construction of a response matrix is comparatively simple. Taking a binning equidistant in  $\cos \theta \in [0, 1]$ , i. e. for forward-moving electrons, the response can be represented as a vector  $\vec{R}_T$  that encodes the probability for transmission for each bin. If the bin is fully below/above  $(\cos \theta)_{\min}$ , the corresponding vector element is 0/1, respectively. For the bin that contains  $(\cos \theta)_{\min}$ , the probability for transmission can be computed as the fraction of the bin width above and below the cutoff value. This relies on the assumption that the underlying angular distribution within a bin is isotropic, i. e. flat in  $\cos \theta$ . A reflection vector can be defined by taking  $\vec{R}_R = \vec{1} - \vec{R}_T$ . By flipping the response vectors, they can be used for backward- or upstream-moving electrons with  $\cos \theta \in [-1, 0]$ . In TRModel, this is necessary to describe backscattered electrons from the detector that are then back-reflected again at the pinch magnet.

To obtain the distribution after reflection, the corresponding portion of the angular distribution can just be flipped since the electrons remain in the region of the magnetic field  $B_1$ . However, the transmitted portion will be de-collimated at the elevated magnetic field  $B_2$ . For example, for electrons coming from the low field of the rear wall to the higher field in the source region, some part of the angular distribution will be reflected. Concerning the transmitted part, the angular distribution at the elevated source field will be de-collimated. This requires the use of an additional response, which is covered in the next section.

The reason collimation and reflection are treated separately is to provide more flexibility when assembling the responses into a model of the beamline. For instance, the angular distribution at the pinch magnet position is irrelevant so long as it is known which portion of an incoming angular distribution is transmitted. So the application of a de-collimation response can be skipped, which reduces the complexity and saves computation time. In the rear wall-to-source transition example from above, on the other hand, the information on the angular distribution at the elevated source field is crucial, since the source scattering effect is dependent on the angle.

### 5.3.2 Magnetic collimation

The transformation of the pitch angle for magnetic collimation and de-collimation is described by equation (5.13). It can be rewritten in terms of  $\cos \theta$  as

$$\cos \theta_2 = \sqrt{1 - \frac{B_2}{B_1}(1 - \cos^2 \theta_1)} , \quad (5.15)$$

where  $\theta_{1/2}$  is the pitch angle at  $B_{1/2}$ . This equation is used to determine the transformed pitch angle of single electrons. To apply the transformation to an entire distribution, the concept of random variable transformation [265] is adapted for binned distributions. The application of the method is introduced here in broader terms since it is also utilized for the description of the post-acceleration response and the readout calibration response (see sections 5.3.4 and 5.5.5).

#### Binned random variable transformation

The cosines of the pitch angles are interpreted as random variables  $x = \cos \theta_1$  and  $y = \cos \theta_2$  that follow the probability density functions (PDFs)  $f(x)$  and  $g(y)$ . Equation (5.15) is the function  $y(x)$  that relates the two random variables, and it can also be inverted to a function  $x(y)$ . In the case of equation (5.15), the inversion can be accomplished by simply exchanging  $B_1$  and  $B_2$ . From the conservation of probability in any matching interval of both distributions, it follows that  $g(y)$  can be calculated via [265]

$$g(y) = f(x(y)) \left| \frac{dx}{dy} \right| . \quad (5.16)$$

However, for using this transformation with binned distributions, the method must be adjusted since there is no information on the underlying PDF. The binned distributions can be written as vectors  $\vec{f}$  and  $\vec{g}$ , which are not necessarily of the same length. The elements  $f_i$  and  $g_j$  are the contents of the bins intervals  $[x_i, x_{i+1}]$  and  $[y_j, y_{j+1}]$ , respectively.

A reasonable assumption is needed for the underlying PDF within a bin. Here, a flat distribution is chosen for simplicity, and since it matches the visual implication from a histogram. For the calculation of a response matrix for the transformation  $\vec{g} = R\vec{f}$ , the following algorithm was developed as part of this work:

1. Loop over all combinations of input and output bins  $i, j$
2. Map the output bin edges to the corresponding locations in the input range by using the inverted mapping function:

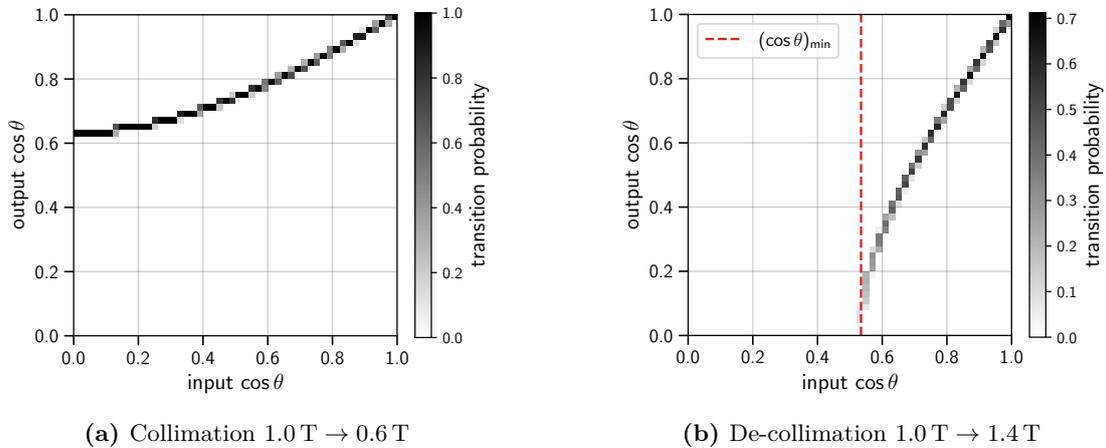
$$[\tilde{x}_j, \tilde{x}_{j+1}] = x([y_j, y_{j+1}])$$

3. Calculate the overlap  $\Delta w_{ij}$  of the input bin interval  $[x_i, x_{i+1}]$  and the mapped output bin interval  $[\tilde{x}_j, \tilde{x}_{j+1}]$
4. Obtain the response matrix element for the transition from bin  $i$  to  $j$  by normalizing with the input bin width  $w$ :

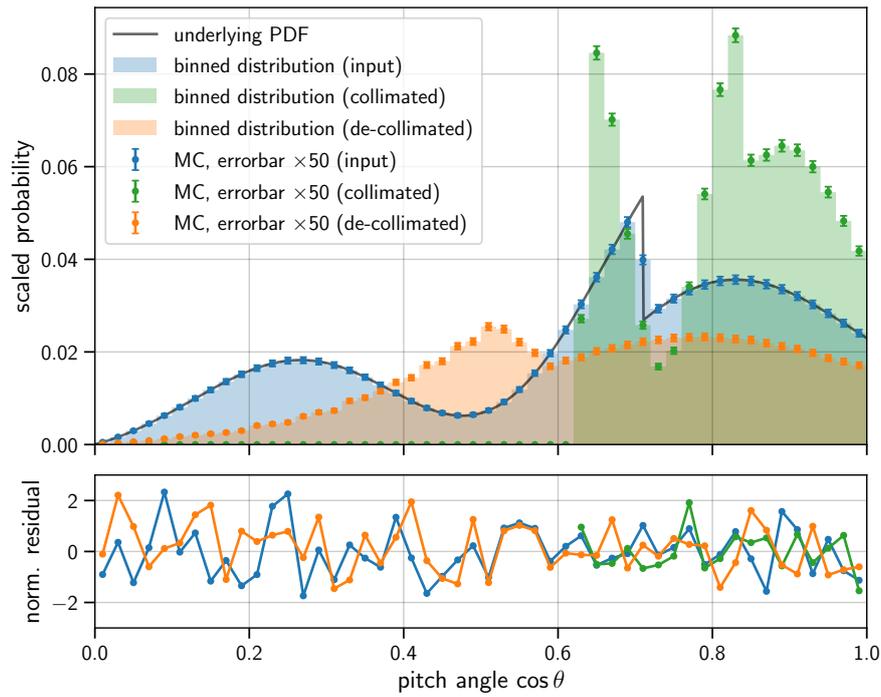
$$R_{ij} = \Delta w_{ij}/w$$

Examples of the resulting response matrix are shown in figure 5.14, and their effect on an arbitrary angular distribution is demonstrated in figure 5.15. Also, a simple MC simulation is used to test whether the response matrices produce the desired result and adhere to the expectation from equation (5.15). For this, a total of  $10^8$  random numbers are sampled from a binned input distribution and transformed with equation (5.15). The simulation result matches the convolution result within statistical uncertainty, as shown by the residuals in figure 5.15.

The magnetic de-/ collimation and reflection responses are utilized for the field transitions between the rear wall and WGTS, the detector and WGTS, and the pinch magnet reflection. For specifics, see section 4.4.2.



**Figure 5.14:** Response matrices for magnetic collimation and de-collimation. In the case of de-collimation, all matrix elements with  $(\cos \theta)_{\text{in}} < (\cos \theta)_{\text{min}}$  are zero since electrons with these pitch angles are reflected.



**Figure 5.15:** Effect of magnetic de- / collimation on an arbitrary angular distribution for transitions from 1.0 T to 0.6 T (collimation) and to 1.4 T (de-collimation). The result from the response matrix convolution is shown with the shaded bar plots. Each result is compared to a monte carlo simulation where  $10^8$  random samples are drawn from the binned input distribution and transformed with equation (5.15). The bottom plot compares the convolution result to the monte carlo simulation by taking the residuals normalized to the statistical uncertainty.

### 5.3.3 Spectrometer transmission

As mentioned in chapter 3, the KATRIN spectrometer is a MAC-E type high-pass filter allowing only electrons with sufficient kinetic energy to be transmitted. A sharp energy cutoff with a width on the eV-scale is achieved [5] by combining extreme magnetic collimation with an electrostatic retardation potential. The electric potential has its maximum in the analyzing plane in the center of the spectrometer.<sup>2</sup> The maximal retarding potential  $qU$  coincides with the magnetic field minimum  $B_{\text{ana}}$ , where the electron pitch angle is maximally collimated. Given the large diameter of the KATRIN spectrometer, the magnetic flux tube can be widened to the extent that the magnetic field can be reduced by up to a factor of  $10^4$  [4]. Through the resulting magnetic collimation in the analyzing plane, most of the electron energy is in the parallel component  $E_{\parallel}$  (see equation (5.11)). Electrons with insufficient parallel energy are reflected.

As in the case of the pinch magnet, it is only relevant whether electrons are transmitted or reflected at the spectrometer. The collimation response from the previous section is thus not required.

The transmission behavior of the MAC-E filter is well-understood [141, 256]. As a function of initial kinetic energy and pitch angle, the transmission condition is given by [221]

$$\mathcal{T}(E, \cos \theta, qU) = \begin{cases} 1 & \text{if } E \cdot \left(1 - (1 - \cos^2 \theta) \cdot \frac{B_{\text{ana}}}{B_{\text{ini}}} \cdot \frac{\gamma + 1}{2}\right) > qU \\ 0 & \text{else} \end{cases} . \quad (5.17)$$

It can be computed based on the adiabatic principles outlined at the beginning of section 5.3. Here,  $B_{\text{ini}}$  refers to the initial magnetic field of the electron. It will equal  $B_{\text{src}}$  if the electron arrives from the source side, and  $B_{\text{det}}$  if it comes from the detector side via backscattering.

The energy resolution, which is quantified by the so-called filter width  $\Delta E$ , can be calculated from the transmission condition by assuming an initial angular distribution and integrating over  $\cos \theta$ . For an isotropic distribution and accounting for the reflection at the maximal magnetic field  $B_{\text{max}}$  of the pinch magnet, the calculation yields [221]

$$\Delta E = \frac{B_{\text{ana}}}{B_{\text{max}}} \cdot E \cdot \frac{\gamma + 1}{2} . \quad (5.18)$$

Figure 5.16 demonstrates the transmission condition and the filter width of the transmission function for the maximal magnetic field that can currently be achieved in the analyzing plane. A value of  $B_{\text{ana}} = 20.5$  G is determined with a Kassiopeia simulation. The high field offers the best conditions for adiabatic transport through the spectrometer, which is a crucial aspect of the deep spectrum measurement [243]. This results in a filter width of  $\Delta E = 9.2$  eV, and  $\Delta E = 15.4$  eV on condition that the pinch reflection is neglected. The latter value represents an upper limit for the filter width during the deep spectrum measurement, regardless of the angular distribution and the choice for

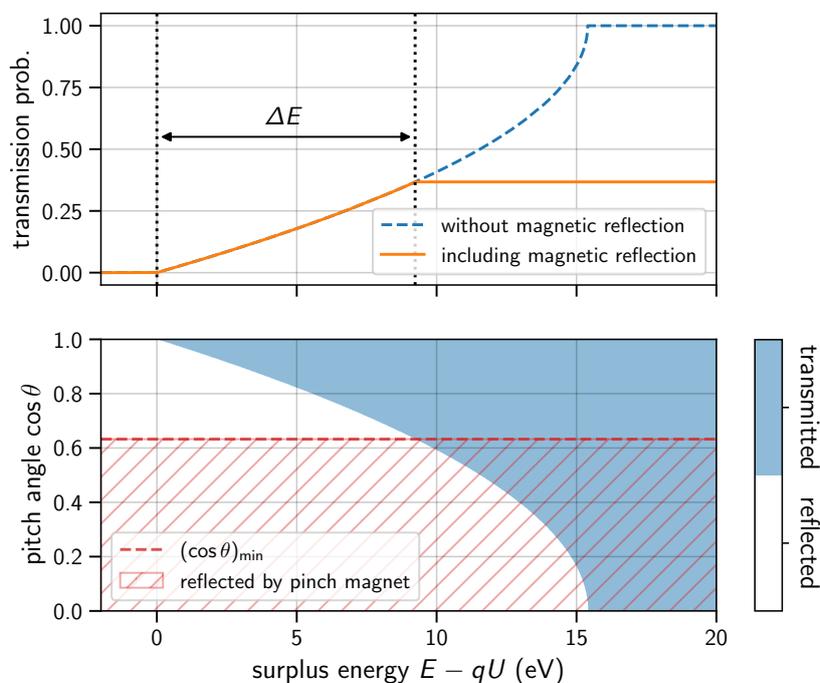
<sup>2</sup>Except in cases where an asymmetric field configuration is used [309].

the maximal magnetic field. Due to the linear energy dependence, the width is much smaller at low retarding potentials.

### Spectrometer response in TRModel

In principle, the transmission condition cannot be fully integrated over the angular axis to describe the transmission probability for the 2D electron spectrum used in TRModel. Instead, it must be integrated for each energy-angle bin, which can be done with analytical or numerical integration. However, since the upper limit for the filter width is much smaller than the currently used bin width of 185 eV, the cutoff energy is practically constant for all  $\cos \theta$ , and the angular dependence can be neglected.

For this reason, the energy cutoff is modeled as a zero-width step function. The response can then be represented by one-dimensional transmission and reflection vectors that encode the corresponding probabilities for each energy bin. These vectors are computed analogously to the magnetic reflection response (see section 5.3.1), the only modification being that the cutoff is in the energy dimension instead of the angular one. If the energy bin width is significantly decreased in the future, this response can be revisited.



**Figure 5.16:** Transmission function and transmission condition of the spectrometer section as a function of  $\cos \theta$  and the surplus energy. The bottom plot shows the transmission condition of the spectrometer in blue. Integrating over the pitch angle yields the transmission function shown in the top plot. The parameters for the calculation are  $B_{\text{src}} = 2.52$  T,  $B_{\text{ana}} = 20.5$  G,  $B_{\text{max}} = 4.20$  T, and  $qU = 18575$  eV.

### 5.3.4 Post acceleration

Electrons transmitted through the spectrometer will arrive at the detector. But before this will happen, they are accelerated by the positive electric potential of the post-acceleration electrode (PAE). The post-acceleration voltage can be adjusted, but it is usually operated at 10 kV and limited to 12 kV [240]. The boost of the electrons' kinetic energy reduces the magnitude of detector and readout systematics, such as energy loss in the dead layer, charge sharing, and pileup (see sections 5.4 and 5.5.1). Since the electric acceleration affects only the parallel energy of the electron, its pitch angle is collimated in the process, thereby reducing the backscattering probability (see section 5.4.1).

Due to backscattering, electrons also approach the electric potential gradient of the post-acceleration from the other side, where they can be reflected back toward the detector. As shown in section 5.4.1, the reflection can be slower than the time resolution of the detector, and electrons can thus produce several events. Consequently, the partial back-reflection of backscattering electrons needs to be included in the propagation stage of TRModel. Backscattered electrons transmitted past the post-acceleration electrode are de-collimated and decelerated before entering the spectrometer.

To summarize, three effects need to be described with response matrices: a) the reflection of electrons approaching the PAE from the detector side, b) the angular de-collimation, and c) the energy shift from acceleration/deceleration.

#### Approach

Electrons when transitioning from the electric potential  $U$  to  $U'$ , will change their kinetic energy by  $E_{\text{pae}} = e(U - U')$ . Depending on the side that the electron is coming from,  $E_{\text{pae}} > 0$  implies an acceleration and  $E_{\text{pae}} < 0$  a deceleration. Assuming a constant detector magnetic field with field lines parallel to the electric field, the energy  $E_{\text{pae}}$  is added only to the parallel component. So the energy transforms as follows:

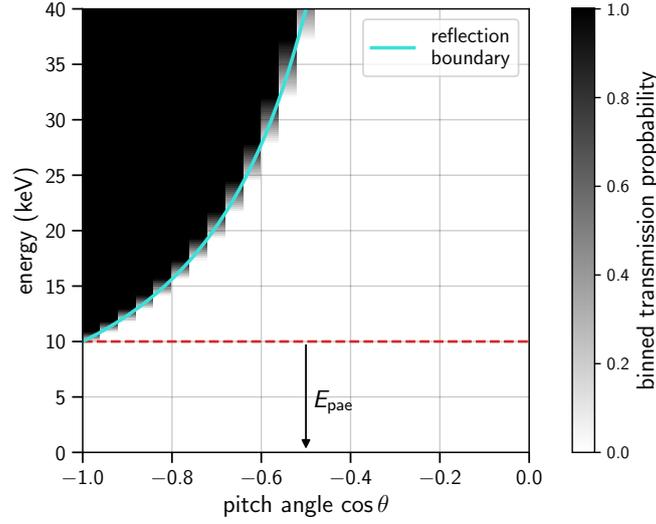
$$E'_{\parallel} = E_{\parallel} + E_{\text{pae}} \quad \text{and} \quad E'_{\perp} = E_{\perp} \quad (5.19)$$

Looking at the simulation shown in figure 5.20, it should be noted that the assumption of a constant magnetic field may not be valid in the potential gradient region, and the model may need to be supplemented with the process of magnetic collimation in the future. The parallelity of the electric and magnetic field lines is indicated in [240]. The applicability of these assumptions should be tested as soon as the model accuracy gradually becomes more critical.

#### Reflection

Based on equation (5.19), the following transmission condition can be formulated for the electrons coming from the detector side:

$$\mathcal{T}_{\text{pae}}(E, \cos \theta, E_{\text{pae}}) = \begin{cases} 1 & \text{if } E \cdot \cos^2 \theta + E_{\text{pae}} > 0 \\ 0 & \text{if } E \cdot \cos^2 \theta + E_{\text{pae}} \leq 0 . \end{cases} \quad (5.20)$$



**Figure 5.17:** Transmission response for electrons approaching the post-acceleration electrode from the detector side. The matrix elements are calculated from an analytical integration of the area above the reflection boundary. In this example, a potential difference of  $|E_{\text{pae}}| = 10$  keV is assumed. Any electrons below this energy are reflected. Above  $|E_{\text{pae}}|$ , only electrons with sufficiently low pitch angles can overcome the electrode potential.

A two-dimensional response matrix for the transmission probability in the  $(\cos \theta, E)$ -plane can be calculated by integrating the transmission condition (5.17) for each energy-angle bin. When rearranging the condition  $E \cdot \cos^2 \theta + E_{\text{pae}} = 0$  this results in a hyperbola for the reflection boundary:

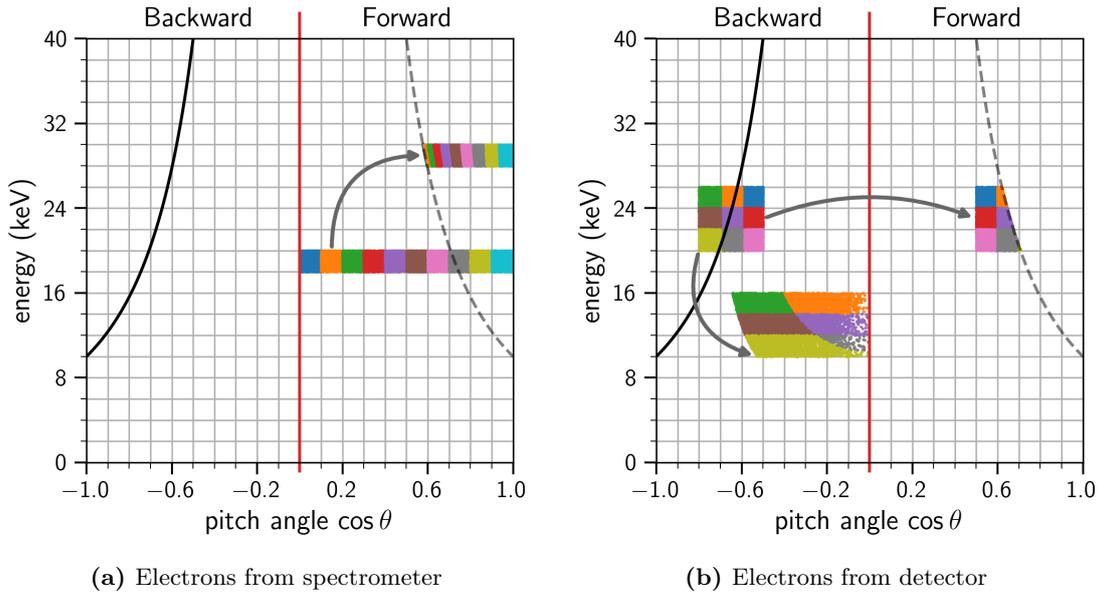
$$E_{\text{ref}}(\cos \theta) = \frac{-E_{\text{pae}}}{\cos^2 \theta} \quad (5.21)$$

Given the simple form of this boundary, the integration over the bins can be calculated analytically for maximal computational efficiency. The resulting response matrix is visualized in figure 5.17.

### Collimation and energy shift

Electrons transmitted through the PAE are accelerated and collimated when they approach from the spectrometer side and decelerated and de-collimated when they approach from the detector side. The energy  $E'$  and  $\cos \theta'$  after transmission through the PAE, in either case, can be determined from equation (5.19) and the definition of the parallel energy (5.11). The calculation yields:

$$\begin{pmatrix} E' \\ \cos \theta' \end{pmatrix} = \begin{pmatrix} E + E_{\text{pae}} \\ \sqrt{1 - \frac{E}{E + E_{\text{pae}}} \cdot (1 - \cos^2 \theta)} \end{pmatrix}. \quad (5.22)$$



**Figure 5.18:** Qualitative demonstration of the post-acceleration on the electron distribution in the  $(\cos \theta, E)$ -plane. The reflection boundary (see equation (5.21)) is indicated in black. Reflection can only happen for backward-propagating electrons. The mirror image of the reflection boundary for  $\cos \theta > 0$  is shown in dashed gray. Random samples are drawn from uniform distributions for several grid squares representing bins. Each sample is transformed according to equations (5.20) and (5.22) with  $E_{\text{pae}} = 10 \text{ keV}$  (a), and  $E_{\text{pae}} = -10 \text{ keV}$  (b). In case (a), the electrons are shifted upward in energy and collimated to the range inside the reflection boundary. In case (b), they are shifted downward and de-collimated if they are above the reflection boundary. Otherwise, they are reflected, and their angle is flipped accordingly.

This is a two-dimensional non-linear transformation and its effect on the electron distribution in the  $(\cos \theta, E)$ -plane is shown qualitatively in figure 5.18 using random sampling. Compared to the process of magnetic collimation, which is energy-independent, this transformation is much more difficult to translate to a binned response. To simplify the problem, a constant energy within each bin is assumed, reducing it to a set of one-dimensional angular transformations for each energy bin. This way, a response matrix can be constructed for only the de/-collimation. The energy shift is applied afterward with a separate response.

The construction of the collimation response proceeds as follows: For each energy bin, the angle transformation from equation (5.22) is taken with  $E$  being fixed to the central energy of the bin. This function is used as a mapping function for the binned random variable transformation from section 5.3.2. For one energy bin, this will result in a matrix of the shape  $n_A \times n_A$ , where  $n_A$  is the number of angular bins in  $\cos \theta \in [0, 1]$ . This matrix is similar to the ones for magnetic collimation (see figure 5.14) since the mapping function is almost equivalent (see equation (5.15)). Repeating this process for all energy bins, the matrices can be concatenated into the collimation response of the

shape  $n_E \times n_A \times n_A$ .

Concerning the energy shift, the mapping function  $E' = E + E_{\text{pae}}$  is independent of the angle. Using the binned random variable transformation again, a response of the shape  $n_E \times n'_E$  can be constructed, which maps each input energy bin to the corresponding output energy bins. The matrix is in general not square since the energy range needs to be extended for acceleration and decreased in the case of deceleration. To obtain an equal bin width for input and output, the response has the shape of a shifted bidiagonal matrix because each input energy bin maps to two output bins.

For a more accurate description of the problem, the book by Blobel [265] outlines a general solution for nonlinear 2D transformations of smooth PDFs involving the calculation of the Jacobi determinant. However, the solution must be adapted to binned distributions. For future versions of TRModel, performing this approach should be considered once the assumptions on the magnetic and electric field conditions from above are verified.

## 5.4 Detector Effects

In-depth descriptions of detector systematics are indispensable in a differential measurement with the TRISTAN silicon drift detector (SDD). Consequently, the TRISTAN SDD's response has been subject to many studies [248, 251, 310] and the main characteristics of prototype detectors, such as the energy resolution and linearity, have been published [8, 145]. The relevant systematics can be grouped into three categories:

- Energy deposition losses
  - Backscattering
  - Dead layer
  - X-ray escape
- Fano noise
- Charge sharing

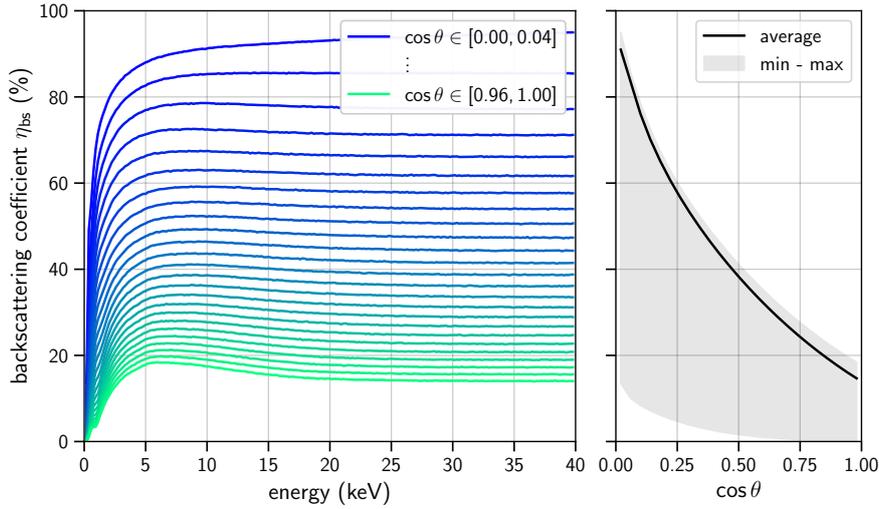
Each of the three categories corresponds to one response matrix in TRModel. Furthermore, the interplay with the electromagnetic guiding field must be considered when the detector is integrated into the KATRIN beamline, leading to the important effect of back-reflection of backscattered electrons. Therefore, TRModel also includes a back-reflection response matrix that re-injects electrons into the propagation model, similar to the rear wall backscattering response (see sections 4.4.2 and 5.2.1). However, in contrast to the rear wall, the effect is now more difficult to model due to the possibility of electrons returning to the detector faster than its time resolution.

### 5.4.1 Backscattering and back-reflection

Backscattered electrons will re-enter the beamline after partially depositing their energy in the detector. The chance of backscattering generally depends on the incident electron's energy and pitch angle. It is quantified by the backscattering coefficient  $\eta_{\text{bs}}$ , which is visualized in figure 5.19.

Backscattered electrons can be reflected again by the magnetic or electric fields before they can pass the spectrometer, depending on their energy and pitch angle. In the vicinity of the detector, four reflection points need to be considered: the detector magnet, the post-acceleration electrode, the pinch magnet, and the retarding potential. For these distinct positions, the flight time  $\Delta t$ , until electrons return to the detector, can be shorter than the detector's time resolution  $\tau_{\text{min}} \approx 100$  ns, as demonstrated by [302].

The back-reflected electrons will hit the same or a different pixel depending on the initial hit position, the pixel diameter, their gyroradius, and the guiding center drift. Per design, they will likely hit the same pixel or migrate at most to adjacent pixels within one back-reflection cycle [145]. If the same pixel is hit within the time resolution, the deposited energies of the initial and second hit will add up. Otherwise, the second hit will be registered as a separate event. Hence, it must be distinguished between fast and slow back-reflection for  $\Delta t < \tau_{\text{min}}$  and  $\Delta t > \tau_{\text{min}}$ , respectively.



**Figure 5.19:** Backscattering coefficient for electrons on silicon from the Geant4 simulation also used to construct the detector reflection response. For the simulation,  $10^6$  electrons are initialized uniformly in every  $(\cos \theta, E)$ -bin.

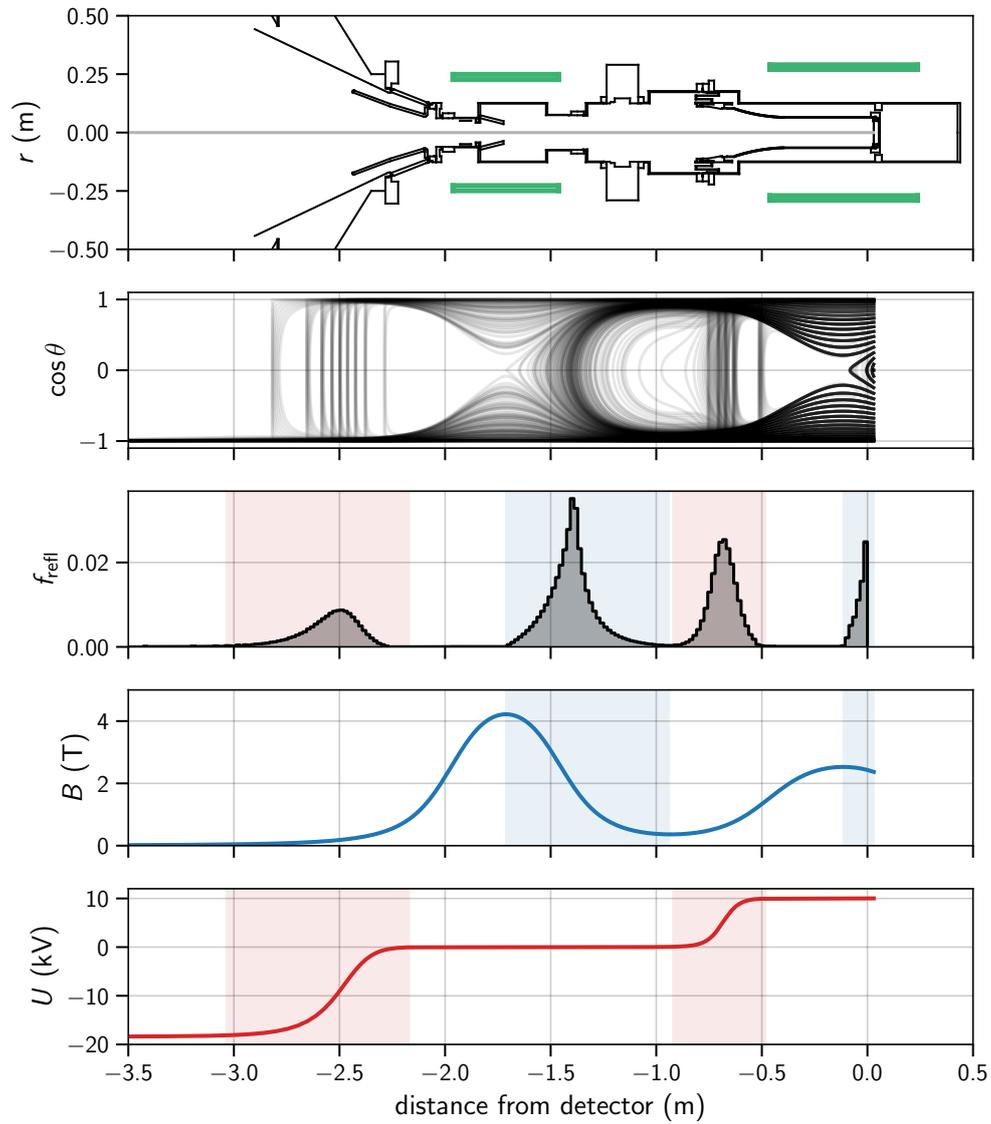
In both cases, a Geant4 simulation has been used to construct a 4D response matrix that maps each input bin in the  $(\cos \theta, E)$ -plane to each output bin. A total of  $10^6$  electrons are simulated for every bin with underlying uniform distributions. The simulation stores the energy and angle of backscattered electrons as output, which is needed for the slow back-reflection response. Furthermore, the deposited energy in the detector for each event is recorded, which is then used to construct the response matrix for energy depositions losses, as described in section 5.4.2.

### Fast back-reflection

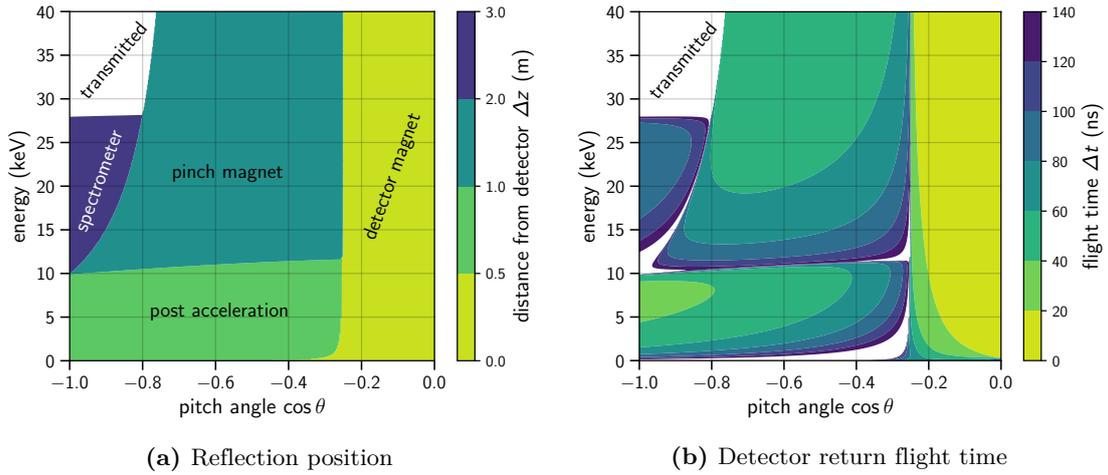
Additional information about the return flight time as a function  $\Delta t(E, \cos \theta)$  over the backscattering phase space is mandatory for fast back-reflection. This is calculated with the Kassiopia simulation shown in figure 5.20, where electrons are initialized in the center of the detector with  $E$  and  $\cos \theta$  spanning a  $500 \times 500$  grid. The resulting points of reflection and the return flight time over the phase space are shown in figure 5.21.

The return flight time map is interpolated to arrive at the function  $\Delta t(E, \cos \theta)$ , which can be used in a novel Geant4 algorithm developed by A. Nava [311]. This algorithm checks an electron backscattered from the detector during Geant4 tracking for its expected return flight time. If it is lower than a specified detector time resolution, the electron is re-injected into the simulation with flipped pitch angle, adding another energy deposition toward the initial hit. Through this procedure, the loss of deposited energy from multiple passes through the dead layer is taken into account.

As visible in figure 5.21b, even multiple backscattering processes are possible within  $\tau_{\min} \approx 100$  ns. Consequently, the re-injection procedure is continued iteratively until the



**Figure 5.20:** Back-reflection Kassiopeia simulation. Electrons are generated in the center of the detector with initial pitch angles  $90^\circ$  to  $180^\circ$ , i. e.  $\cos\theta$  in  $[0, -1]$ , and energies 0 keV to 40 keV chosen in a uniform  $500 \times 500$  grid. The top plot shows the simulation geometry's electrodes (black) and magnet coils (green). The second plot shows the evolution of the pitch angle of individual electrons. Each crossing from negative to positive  $\cos\theta$  represents a reflection. The fraction of reflections depending on the distance from the detector is shown in the center plot. The shaded reflection regions correspond to the regions where the electric potential and the magnetic field exhibit gradients for the incoming electron.



**Figure 5.21:** Results of the Kassiopeia simulation from figure 5.20 as functions of  $E$  and  $\cos \theta$ .

cumulative flight time exceeds  $\tau_{\min}$ .<sup>3</sup> All electrons surviving the re-injection stage are considered back-reflected slowly and thus stored in the corresponding output. The impact of this on the spectrum re-injected into the propagation model is demonstrated in figure 5.22. While the number of backscattered electrons is reduced with fast back-reflection, the resulting spectral shape is much less smooth.

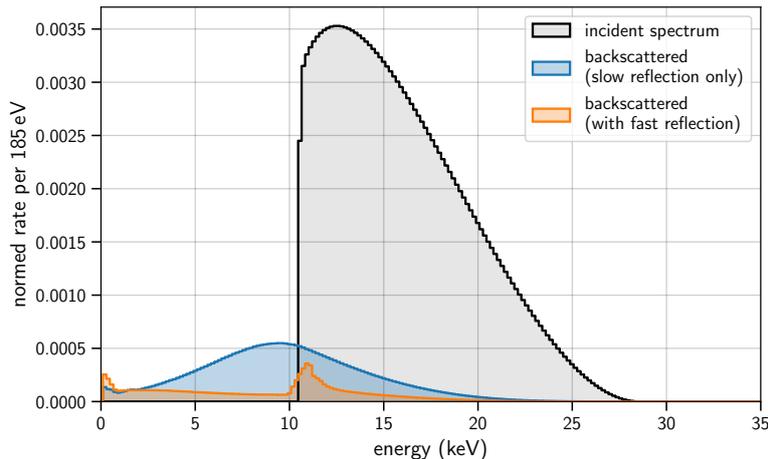
### Current status

At the time of writing, the possibility of pixel changes for back-reflected electrons and the interplay with the pileup rejection of the readout system are still being neglected. Furthermore, the model must be variable concerning its systematic parameters to study the sensitivity impact of fast back-reflection. Here, the uncertainty of the electromagnetic field and the detector time resolution must be considered. Since each systematic parameter will increase the dimensionality of the response, this is a computationally expensive and challenging task that has yet to be realized. Lastly, the interplay with the pileup rejection mechanism of the readout system must be accounted for. The development of a combined Kassiopeia, Geant4, and trace-level readout simulation is still ongoing. For these reasons, the studies in chapter 6 will only include slow back-reflection in the spectrum calculation.

### Mitigation

Previous studies [145, 302] assume that the reduction of the probability for escape past the spectrometer, i. e. a higher fraction of fast back-reflection events, will benefit the sensitivity since the reconstructed event energy is, on average, closer to the initial energy. On the other hand, fast back-reflection is challenging to model and may introduce more

<sup>3</sup>Recent developments have shown that the drift time inside the SDD, which depends on the precise hit location within the pixel, must also be considered when comparing the time between hits with the time resolution.



**Figure 5.22:** Expected backscattering spectrum associated with slow back-reflection. The backscattering spectrum is re-injected into the TRModel propagation. Shown are the two cases of only slow back-reflection, corresponding to an ideal detector time resolution  $\tau_{\min} = 0$  ns, and including fast back-reflection with a realistic estimate  $\tau_{\min} = 112$  ns.

considerable systematic uncertainties since the spectral impact is less smooth (see figure 5.22). In a paper by Mertens et al. [225], it was shown that systematics with short energy correlation lengths are expected to have a more significant impact on the sensitivity.

Depending on the outcome of further studies with an improved model, the fast and slow back-reflection ratio could be optimized by adjusting the electromagnetic field and tuning the detector time resolution. For example, the time resolution can be decreased to favor fast back-reflection, or instead, more advanced event reconstruction could be attempted on the readout level. Also, the pinch magnetic field could be lowered, and the detector could be relocated past the maximum of the detector magnetic field, which will remove the possibility of fast back-reflection at the magnetic fields and would thus reduce the model complexity.

In all scenarios, the post-acceleration voltage can be increased, and the detector magnetic field be decreased to reduce the initial backscattering probability through pitch angle collimation. Furthermore, the tagging of fast back-reflection with intra- and inter-pixel coincidence detection, i. e. multiplicity analysis, can be investigated further.

#### 5.4.2 Energy deposition losses

An incident electron will deposit its energy in the SDD through multiple elastic and inelastic scattering [267]. However, three key effects can lead to an incomplete deposition of the energy [263]:

- **Backscattering:** Over its path through the silicon lattice, the electron can reverse its direction and exit the material again. If it is not externally redirected, the energy it still carries is lost. This leads to a characteristic tail in the measured spectrum.

- **X-ray escape:** The electron can also excite the emission of X-rays when traveling through the silicon. This includes Bremsstrahlung and the emission of characteristic X-rays with an energy of  $E_{\text{Si}} = 1.74 \text{ eV}$  from the de-excitation of silicon lattice atoms [248, 263]. All X-rays escaping the detector will constitute an energy loss.
- **Dead layer:** Semiconductor detectors exhibit a thin surface layer from which the charge created via energy deposition is not fully collected [262, 312]. It consists of a silicon oxide layer of about 10 nm thickness, followed by a transition region of several 100 nm thickness where the charge collection efficiency is incomplete while approaching full efficiency [310].

### Simulation method

In TRmodel, a response covering all these effects is constructed with the simulation method from Biassoni et al. [263] that combines a Geant4 simulation with a sophisticated dead layer model. Geant4 inherently includes backscattering and X-ray escape. The simulation of the dead layer is realized by recording the spatial position of each energy deposition from individual scatterings. All depositions are weighted with a depth-dependent signal conversion efficiency and summed up for the overall energy deposition. The weighting function representing the dead layer model is chosen as:

$$f_{\text{QE}}(z; t, p_0, p_1, \lambda) = \begin{cases} p_0 & \text{if } z < t \\ 1 + (p_1 - 1) \cdot \exp\left(\frac{t - z}{\lambda}\right) & \text{if } z \geq t \end{cases} \quad (5.23)$$

Here, the parameter  $t$  specifies the thickness of the passivation oxide layer and  $p_0$  any remaining efficiency in that layer. Past the oxide layer, the model exponentially approaches the maximal efficiency with a characteristic length  $\lambda$ , starting from an efficiency  $p_1$  at  $z = t$ . Biassoni et al. fit the model parameters to data taken with a silicon drift detector and an electron gun, achieving remarkable agreement.

For the TRModel calculation,  $p_0$  is set to zero, as the creators of the method did. Furthermore, since the model seems to be insensitive to the parameter  $p_1$ , it is also set to zero to reduce complexity. The oxide thickness  $t$  is fixed to the value reported by [310], leaving  $\lambda$  as a free parameter to express systematic uncertainty. From data taken with a TRISTAN detector and an electron gun, it was determined that a value  $\lambda = (58 \pm 1) \text{ nm}$  results in the best match.

### Response construction

The likelihood of backscattering and the energy loss in the dead layer depend on the electron's energy and, due to geometric effects, also on the incident angle. The higher the energy, the deeper the electron will penetrate the silicon lattice, thereby reducing the energy fraction deposited in the dead layer [267]. At non-zero angle, the effective path length through the dead layer is increased, leading to the relation [312]

$$\Delta E(\theta) = \frac{\Delta E_0}{\cos \theta}, \quad (5.24)$$

where  $E_0$  is the energy loss at normal incidence. To account for this, the Geant4 simulation is performed for all  $(\cos\theta, E)$ -bins, as stated in the previous section. Hence, the response has the dimensions  $n_E \times n_A \times n_E$ , where  $n_E$  and  $n_A$  are the numbers of energy and angle bins. After applying this response, the dimensionality of the spectrum is reduced to only the energy dimension, allowing for an increased number of energy bins. The rebinning scheme from 4.4.3 is employed to match the input spectrum binning to the response.

After the Geant4 simulation, the dead layer model is applied for a range of  $\lambda$  values to build a response database, similar to the case of the source response. The response for a particular  $\lambda$  value is retrieved during the model evaluation via linear interpolation. All parameters used to create the detector response database are listed in table 5.3.

### 5.4.3 Fano Noise

In semiconductor detectors, the deposited energy is converted to electron-hole pairs, which are drifted to their respective electrodes to produce a measurable signal. The number of created charge carriers is proportional to the deposited energy  $E$  and is given by  $n_{\text{eh}} = E/\epsilon_{\text{eh}}$ , where  $\epsilon_{\text{eh}} = 3.65 \text{ eV}$  is the energy required to produce an electron-hole pair [278]. As the creation of charge carriers is subject to statistical fluctuations, the number of charges created for a particular energy deposition will vary.

This gives rise to a broadening of the measured energy spectrum. However, the number of charges does not fluctuate according to a pure Poisson distribution. Instead, the variance is much smaller than expected due to the correlation of the scattering events along the path of the electron [312]. The improvement of the observed variance in comparison to the expected variance from Poisson statistics is the so-called Fano factor [261, 279]:

$$F_{\text{fano}} = \frac{\text{observed variance}}{E/\epsilon_{\text{eh}}} = 0.115 \quad (5.25)$$

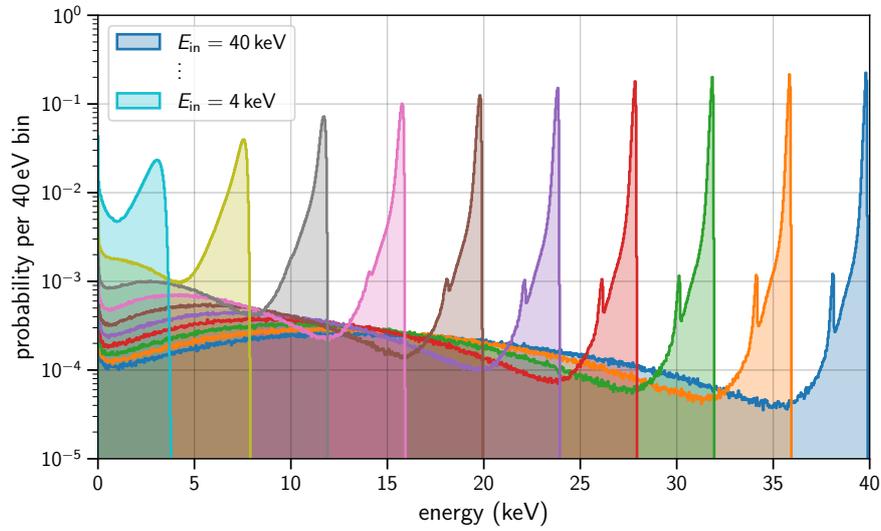
Thus, the impact on the measured spectrum can be accounted for with an energy-dependent Gaussian broadening with a standard deviation

$$\sigma_{\text{fano}}(E) = \sqrt{E \cdot \epsilon_{\text{eh}} \cdot F_{\text{fano}}} . \quad (5.26)$$

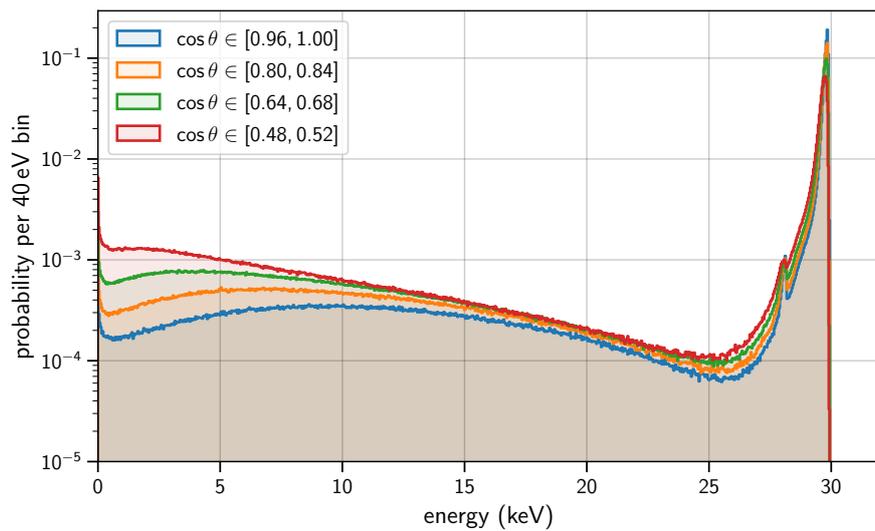
The Fano factor is temperature dependent [278] and could carry some uncertainty [312]. To minimize the uncertainty, a determination of the Fano broadening can be attempted

**Table 5.3:** Settings used to create the TRModel detector response database. The number of output bins and the required  $\lambda$  range can be adjusted depending on the application.

Energy range		0 keV to 40 keV
Number of energy bins	$n_E$	216, 400, 1000, 4000
Number of angle bins	$n_A$	25
Dead layer parameter ranges	$\lambda$	(57 to 59) nm, (56 to 60) nm, (53 to 63) nm, (48 to 68) nm
Number of $\lambda$ values	$n_\lambda$	3, 5, 11, 21

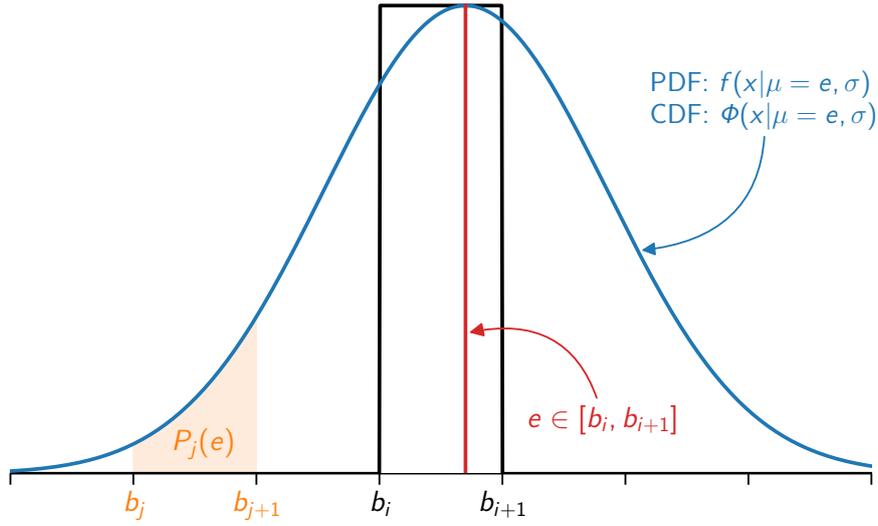


(a) Energy dependence



(b) Angular dependence

**Figure 5.23:** Demonstration of the detector response: Shown is the detector spectrum for different input energies (a), and several of the input angle bins (b). The output energy spectrum exhibits a peak at the input energy, a broadening of the peak to the left from dead layer energy loss, a silicon escape peak 1.74 keV below the main peak, and the backscattering tail extending to low energies.



**Figure 5.24:** Illustration of the method used to construct the noise response via integration of a Gaussian. The probability  $P_j$  for a transition to bin  $j$ , represented by the orange-shaded region, is averaged over all positions  $\mu = e$  of the Gaussian within the input bin  $i$ .

via an in-situ measurement according to the method described in [312]. Therein, the full broadening is measured under conditions where all other broadenings can be determined with different means, e. g. by inserting test pulses. Subtracting all contributions will leave only the Fano broadening, assuming that no further non-statistical effects remain.

### Response construction

The width of the energy bins cannot be neglected compared to the width of the Gaussian smearing. For example, at 10 keV Fano noise contributes a broadening with  $\sigma \approx 65$  eV, while the bin width of the spectrum is 10 to 185 eV depending on the choice of the detector response. Therefore, the Gaussian is integrated and averaged over all possible energies within a particular bin to account for its width, as illustrated in figure 5.24.

The response matrix is constructed with an analytical integration method. Each matrix element  $M_{ij}$  expresses the transition probability from an input bin interval  $[b_i, b_{i+1}]$  to an output interval  $[b_j, b_{j+1}]$ . For each particular energy  $e$  in the input bin range, a Gaussian with the probability density  $f(x|\mu = e, \sigma)$  describes the redistribution of events from  $e$  to all other energies. The probability  $P_j$  for a transition to the target bin  $j$  is calculated by integrating the Gaussian over  $[b_j, b_{j+1}]$ , which can be achieved by taking the difference of the cumulative distribution  $\Phi(x|\mu = e, \sigma)$  evaluated at  $b_{j+1}$  and  $b_j$ :

$$P_j(e) = \Phi(b_{j+1}|e, \sigma) - \Phi(b_j|e, \sigma) \quad (5.27)$$

To obtain the matrix element,  $P_j$  is averaged over all energies  $e$  within the input bin  $i$ :

$$M_{ij} = \frac{1}{b_{i+1} - b_i} \int_{b_i}^{b_{i+1}} P_j(e) de \quad (5.28)$$

In practice,  $\Phi$  is expressed with the error function

$$\Phi(x|\mu, \sigma) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{x - \mu}{\sqrt{2}\sigma} \right) \right] , \quad (5.29)$$

allowing for an analytical integration. The result of the integration, which is not explicitly written here for reasons of length, only depends on exponentials and error functions. Hence the calculation of the response matrix is very efficient.

Since the Gaussian width is a function of  $e$  for Fano noise, averaging over  $\sigma$  would also be necessary to be exact. But the variation of  $\sigma$  over the bin width can be considered to be negligible. So instead, it is evaluated for the central energy of the input bin.

#### 5.4.4 Charge sharing

The TRISTAN SDD is segmented into 166 hexagonal pixels per module. When incident electrons hit close to a boundary between pixels, the charge created by the event is split between them. The fraction of charge shared into the neighboring pixel will depend on the distance of the hit position from the boundary and the lateral width of the charge cloud. The likelihood of charge-sharing is related to the pixel side length to area fraction, so more charge-sharing is expected for smaller pixels [302].

#### Simplified analytical model

A model describing the impact of charge sharing on the measured energy spectrum for TRISTAN detectors was developed in previous studies by M. Korzeczek [248] and K. Urban [251]. Both follow the same approach and assumptions: The charge cloud is assumed as Gaussian with the width  $\sigma =: w_{cc}$ , and the possible hit positions are reduced to one dimension, i. e. a distance from the boundary, neglecting the shape of the pixel.

The first assumption is well-motivated by the description of the dynamics of charge carriers in the SDD by [313]. In both studies, estimates of the charge cloud width  $w_{cc}$  result in values near 20  $\mu\text{m}$ , which they also validate with X-ray and electron measurements. The current knowledge about the charge cloud properties is summarized below in the following paragraph.

The second assumption is also reasonable since the pixel radius  $r_{\text{px}} = 1.5 \text{ mm}$  of the final detector is much larger than  $w_{cc}$ . Naturally, the behavior near the pixel corners and the possibility of triple charge sharing are neglected. TRModel also contains two additional descriptions with a 2D pixel geometry: an analytical model assuming circular pixels and an MC model using a hexagonal shape. However, both models require further testing and validation with data. Hence, they are not detailed here.

For TRModel, the method from [251] is adopted, which is briefly outlined as follows: Given the charges  $Q_A$  and  $Q_B$  deposited in the neighboring pixels from one incident electron, the charge sharing fraction is defined as

$$\eta = \frac{Q_A}{Q_A + Q_B} . \quad (5.30)$$

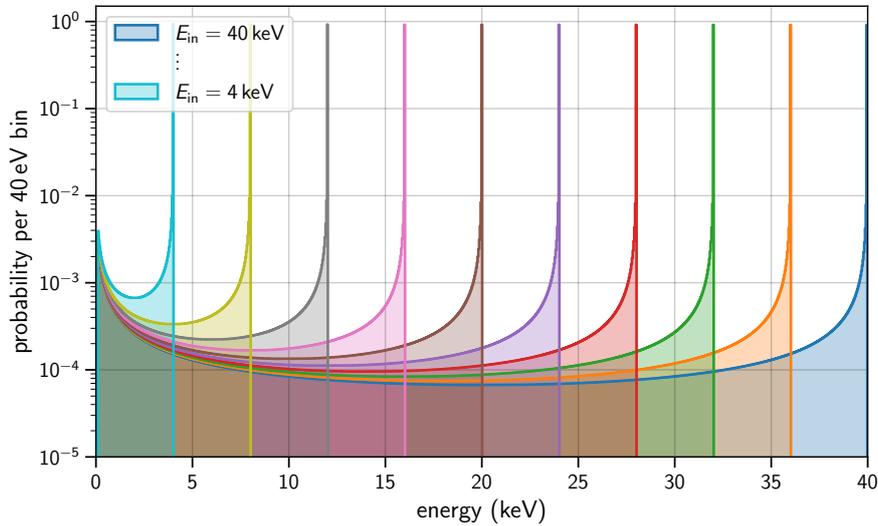
With the assumption of a Gaussian charge cloud with  $\sigma = w_{\text{cc}}$  and its center being located at a distance  $\mu = x$  from the boundary,  $\eta$  can be calculated as a function of  $x$  by computing the cumulative distribution function. Next, homogeneous illumination of the detector is assumed, i. e. the hit position  $x$  will follow a uniform distribution  $P(x) = \frac{1}{\Delta x}$  in the approximation  $w_{\text{cc}} \ll r_{\text{px}}$ . Considering that the hit position can be located anywhere inside the pixel,  $\Delta x$  is identified with the pixel area-circumference ratio, which equals  $\frac{r_{\text{px}}}{2}$  for a circular pixel. The distribution  $P(\eta)$  is then calculated via random variable transformation. In the implementation, this transformation is performed numerically. However,  $P(\eta)$  can also be calculated analytically:

$$P(\eta) = \frac{w_{\text{cc}}}{r_{\text{px}}} \cdot 2\sqrt{2\pi} \cdot e^{(\text{erf}^{-1}(1-2\eta))^2} \quad (5.31)$$

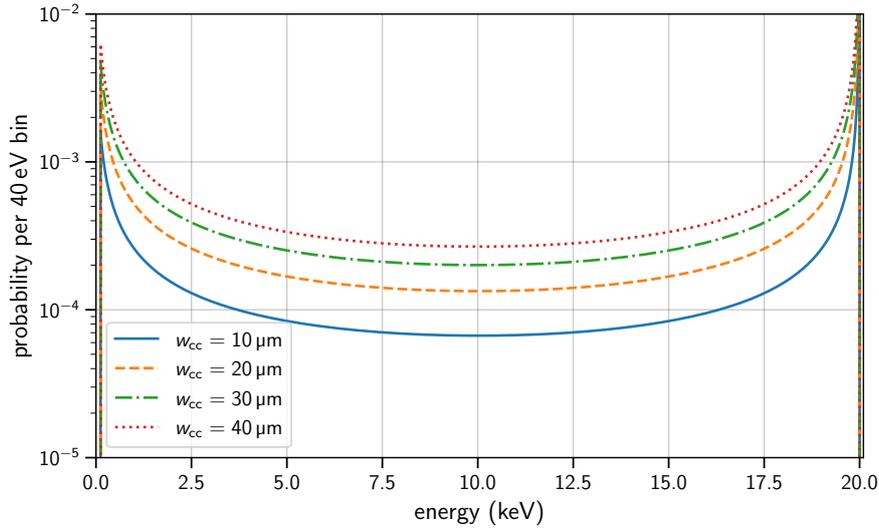
Here,  $\text{erf}^{-1}$  is the inverse error function. The detected energy is related to  $\eta$  by  $E_{\text{det}} = \eta \cdot E_{\text{in}}$ . Consequently,  $P(\eta)$  can be integrated over the corresponding output energy intervals to compute the response matrix elements for each input energy. Taking the center of a bin  $\bar{E}_i$  for  $E_{\text{in}}$  and a target interval  $[E_j, E_{j+1}]$  for  $E_{\text{det}}$ , the integration yields:

$$R_{ij} = \int_{E_j/\bar{E}_i}^{E_{j+1}/\bar{E}_i} P(\eta) d\eta = \frac{w_{\text{cc}}}{r_{\text{px}}} \cdot 2\sqrt{2\pi} \cdot \left( \text{erf}^{-1}\left(1 - 2\frac{E_j}{\bar{E}_i}\right) - \text{erf}^{-1}\left(1 - 2\frac{E_{j+1}}{\bar{E}_i}\right) \right) \quad (5.32)$$

As  $P(\eta)$  approaches infinity for  $\eta \rightarrow 0, 1$ , the integration is not possible for the particular bin intervals containing 0 and those where  $i = j$ . To deal with this, a lower energy threshold of 100 eV is applied, and the elements  $R_{ii}$  are chosen such that the normalization for  $\eta > 0.5$  equals 1 for each input energy (see [251] for details). The response is shown for several energies in figure 5.25, and its scaling with  $w_{\text{cc}}$  in figure 5.26.



**Figure 5.25:** Demonstration of the charge sharing response for different input energies.



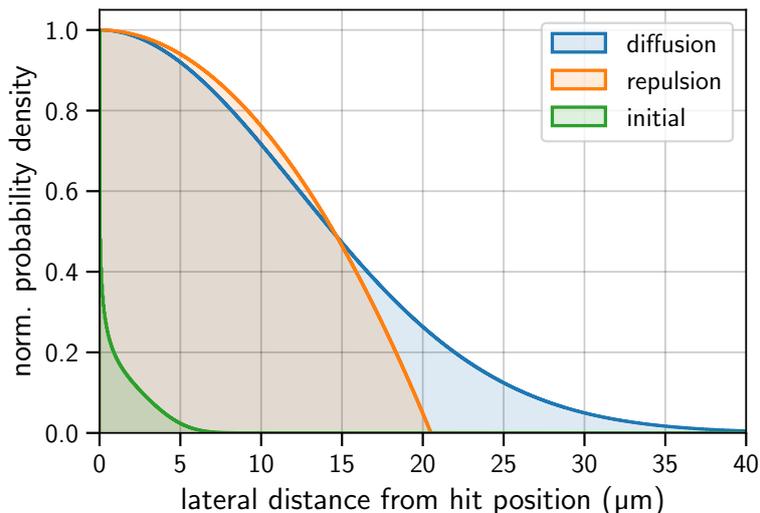
**Figure 5.26:** Scaling of the charge sharing response with the charge cloud width parameter  $w_{cc}$ .

### Charge cloud properties

The lateral shape and width of the charge cloud are essential input parameters to the charge-sharing model. The Gaussian shape is physically motivated by thermal diffusion during the drift time  $t_{\text{drift}} \approx 20$  ns until the charges reach the saddle point of the SDD's electric potential between pixels [251]. Yet, other effects may lead to a deviation from the Gaussian shape.

A publication by Gatti et al. [313] describes the general dynamics of charges in an SDD, which include thermal diffusion and electrostatic repulsion. It is shown that diffusion is the primary component governing the lateral distribution of charge clouds with  $< \mathcal{O}(10^4)$  electrons. Electrostatic repulsion becomes dominant for higher electron count but may still play a role for TRISTAN according to estimations by [248] and measurements by [251]. The initial lateral profile from scattering also plays a minor role since it seeds the other effects [248]. The following list summarizes key literature findings:

- **Initial distribution:** The spatial distribution of the energy deposition from the scattering is studied in [248] based on KESS simulations [267]. Several parameterizations are provided, including the lateral distribution at an input energy of 30 keV and formulas for the energy scaling of distribution quantiles.
- **Thermal diffusion:** Diffusion by itself leads to an isotropic Gaussian spread of the distribution with a width  $\sigma = \sqrt{2Dt_{\text{drift}}}$  [313], where  $D \approx 3.7$  cm<sup>2</sup>/s [278] represents the diffusion constant for electrons in silicon at room temperature.
- **Electrostatic repulsion:** The charge cloud from just repulsion is approximately shaped like an inverted parabola. A procedure for calculating the charge cloud shape is provided in [313]. It is energy-dependent since the repulsion scales with the number of charges in the cloud.



**Figure 5.27:** Estimated lateral charge cloud distributions from thermal diffusion, electronic repulsion, and initial energy deposition. The latter two are energy-dependent and shown for  $E = 30$  keV. Diffusion and repulsion are computed for  $t_{\text{drift}} = 20$  ns. The calculations of diffusion and repulsion are based on [313], and the initial distribution is taken from [248]. Each calculation neglects the other effects, i. e. no interplay is considered.

To compare the magnitude of these effects, the  $1\sigma$  coverage radii of the lateral distributions for a 30 keV incident electron can be calculated:

$$r_{\text{initial}} \leq 2.1 \mu\text{m}, \quad r_{\text{diffusion}} \leq 12.2 \mu\text{m}, \quad r_{\text{repulsion}} \leq 11.6 \mu\text{m} \quad (5.33)$$

This indicates the primary role of diffusion and the possible significance of repulsion. The shape of the distributions is shown in figure 5.27. However, it is difficult to assess the true impact of each effect on the charge cloud due to the difference in the underlying distribution shapes and the interplay during the drift time. A more accurate model could be realized by solving the partial differential equation from [313], including diffusion and repulsion, and initializing it with the distribution from a scattering simulation. The charge cloud size can also be computed using Monte Carlo simulations. This method has been successfully employed in recent work by C. Forstner [314].

### Past measurements

In work by K. Urban [251], the charge cloud was studied with a measurement of conversion electrons and X-rays from a  $^{83\text{m}}\text{Kr}$  source. Charge-sharing events were selected via their multiplicity in neighboring pixels with a time coincidence cut of 200 ns, and the event energy was reconstructed. Using the Gaussian model, a charge cloud width of  $w_{\text{cc}} \approx 20 \mu\text{m}$  for electrons was extracted from the data. A slight trend of the charge cloud size regarding the incident electron energy could be seen, which is attributed to electronic repulsion.

The results are compatible with measurements from [248], where the electron response was measured with a scanning electron microscope. Here, a charge cloud width of  $w_{\text{cc}} =$

$(18.1 \pm 1.2) \mu\text{m}$  for 20 keV electrons was measured with a similar method that included the detection of coincidences from triple charge-sharing. The triple charge-sharing events lead to a discrepancy between the 1D model and the measured response in the low-energy region. Furthermore, the X-ray response was studied with an  $^{55}\text{Fe}$  source, leading to consistent results.

In the recent work by C. Forstner [314], charge-sharing was studied using a collimated monochromatic laser beam, where the laser spot could be positioned with high precision, and each laser pulse amounted to an energy deposition of 10.6 keV. The drift time and charge cloud width were determined from the measured signal shape as a function of the beam position. The further this position was from the center of the pixel, the higher the drift time and the larger the charge cloud. Overall, the data was found to be in  $> 99\%$  agreement with thermal diffusion, which can be expected at this energy. By scanning the laser over the edge between two pixels and analyzing the profile of energy depositions as a function of the position, the charge-cloud width relevant for modeling was determined as  $w_{\text{cc}} = (16.3 \pm 0.2) \mu\text{m}$ .

From these results, the following conclusions for the model are obtained. The Gaussian charge cloud model describes the data well, and below 20 keV, the dominance of thermal diffusion can be assumed due to the consistency with X-ray measurements and laser characterizations. Above 20 keV, the influence of repulsion becomes more relevant, but the impact on the scaling and shape of the charge cloud will require further studies. For the rather small detector pixels used in the measurements ( $r_{\text{px}} = 0.125 \text{ mm}$  and  $1 \text{ mm}$ ), triple-charge sharing cannot be neglected. Hence, its impact should be addressed in future models that respect the pixel shape to evaluate its relevance for the final TRISTAN module with  $r_{\text{px}} = 1.5 \text{ mm}$ . Furthermore, the impact of multiplicity cuts on the spectrum can be modeled to assess the merit as a mitigation technique.

## 5.5 Readout Effects

Each energy deposition in a detector pixel will cause a charge signal, which will then be integrated by the feedback capacitance of a charge-sensitive preamplifier located on the front-end board. The ETTORE preamplifier [249] used for the TRISTAN modules is operated with a pulsed reset, discharging the accumulated charge from events and leakage current with a fixed frequency. A second stage amplification adds an exponential decay with a time constant of  $15\ \mu\text{s}$  to each signal pulse. The amplified signal is carried outside the vacuum chamber and eventually fed into a digitizer. Here, an ADC circuit samples the waveform. Pulse heights and event time stamps are extracted using a combination of digital filters. Understanding how the filters accomplish their task is crucial to understand and model the associated systematic effects.

### Digital filter setup

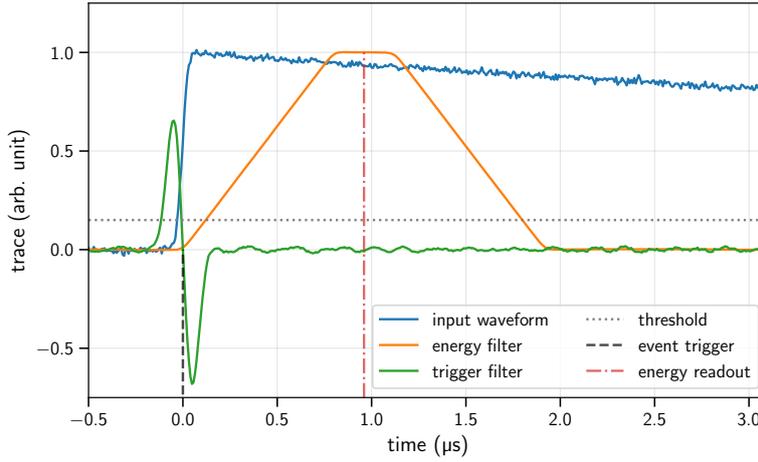
The data acquisition system envisioned for TRISTAN uses a combination of two parallel filters, a method employed by digitizers used for TRISTAN prototypes such as the DANTE DPP [315] or the CAEN V1782 card [316]. A fast trigger filter detects events and marks their time stamp. In the meantime, a trapezoidal filter [317] smoothens the waveform and transforms the pulses into a trapezoid, allowing for precise height extraction. The position of the energy readout at the peak of the trapezoid is determined by the trigger timestamp, accounting for the peaking time and flat top length of the trapezoidal filter. Figure 5.28 shows an example of this filter combination.

Using an independent fast filter for triggering allows for the detection of pileup events, where the signals of consecutive events in the trapezoidal filter are overlapping to such a degree that the individual energies cannot be extracted. These pileup events can then be flagged or rejected by invoking a dead time window after each trigger. The required trigger dead time window length depends on the energy filter's peaking time and flat top length. Like the energy filter, the trigger filter may also be implemented as a trapezoidal filter with a short peaking time, e. g. the DANTE DPP makes use of this approach. The CAEN card, on the other hand, uses a bipolar triangular trigger filter instead, with the advantage of an energy-independent trigger location defined by the zero crossing of the shaped pulse.

### Systematic effects

The following effects can be identified as relevant for a description of the readout system:

- ADC-nonlinearity
- Pileup
- Electronic noise
- Detection threshold
- Dead time
- Calibration



**Figure 5.28:** Simulation of the signal processing inside the digitizer. The simulation is oriented at the filter setup used in the CAEN V1782 card [316] and performed with the trace simulation software described in [253]. The input waveform is generated with an event at  $0 \mu\text{s}$  and a pulse height of 1. The second stage of the preamplifier adds an exponential tail with a  $15 \mu\text{s}$  time constant, which is accounted for in the calculation of the trapezoidal filter via pole-zero cancellation [318]. An energy readout is triggered by the zero crossing of the bipolar trigger filter, which is only armed when its signal first rises above a specified trigger threshold.

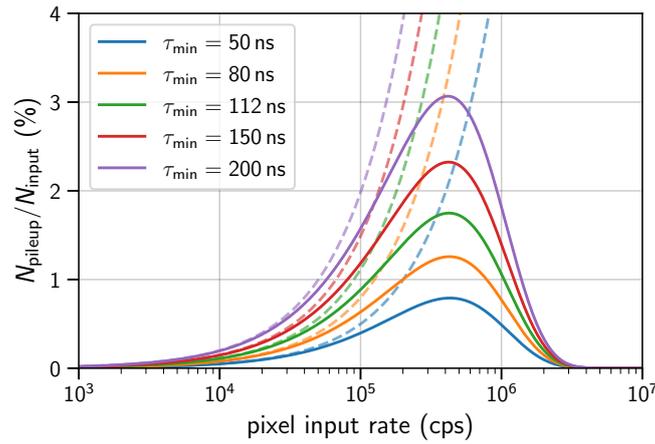
Of these effects, ADC-nonlinearity is not covered here since the impact on the spectrum is similar to statistical fluctuations and cannot be predicted with sufficient accuracy. Instead, it is planned to mitigate ADC-nonlinearity with experimental techniques to a level where it can be neglected. Previous work evaluating several mitigation methods can be found under [252, 284]. The models for the other effects are presented in the following subsections. The pileup and dead time descriptions are adopted from [253].

The digitizer may also be tasked with detecting inter-pixel coincidences to gain multiplicity information for treating charge-sharing and back-reflection [8]. As these features are still in development and the back-reflection model does not yet address pixel changes, they are neglected in the readout model presented here.

### 5.5.1 Pileup

Independent of the exact realization of the trigger filter, each has a limited time resolution above which consecutive pulses can be resolved - the pulse pair resolution  $\tau_{\text{min}}$ . For high input rates, there is an increasing likelihood for events to occur with time differences below  $\tau_{\text{min}}$ . In this case, their signals also overlap in the energy filter, and a so-called unresolved pileup event with increased energy is read out. Assuming events to be uncorrelated, the probability of unresolved pileup can be calculated from the Poisson distribution. The calculation yields [253]

$$P_{\text{pu}} = 1 - e^{-r \cdot \tau_{\text{min}}} , \quad (5.34)$$



**Figure 5.29:** Fraction of pileup events as a function of pixel rate and pulse pair resolution. For the expected time resolution of  $\approx 112$  ns<sup>4</sup> and  $10^5$  cps pixel rate, about 1 % of the input events create unresolved pileup. For extremely high rates, the dead time efficiency loss dominates. The pileup probability without accounting for dead time is indicated with dashed lines.

where  $r$  is the incident event rate of the channel. In TRModel, the rate is calculated from the integral of the differential spectrum, which is normalized according to the procedure from section 4.4.5.  $\tau_{\min}$  is implemented as an adjustable parameter and is utilized as a systematic input for sensitivity studies. In the following, a central value  $\tau_{\min} = 112$  ns is adopted<sup>4</sup>. The fraction of pileup events compared to the number of input events will also depend on the detection efficiency  $\epsilon_{\text{det}} = 1 - \epsilon_{\text{dead}}$  since pileup events can themselves fall into a dead time window of a previous pulse. Figure 5.29 shows the resulting pileup fraction.

### Pileup energy spectrum

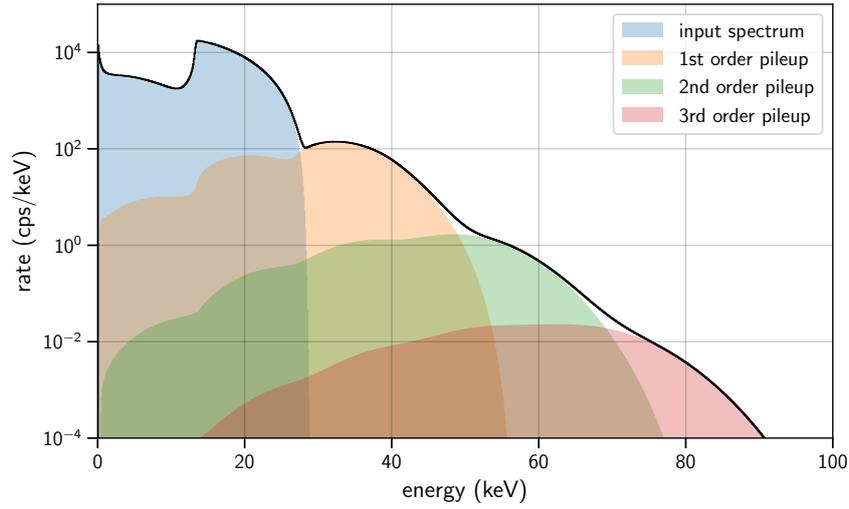
In [253] it is demonstrated that the pileup energy spectrum can be calculated via convolution of the input energy spectrum with itself and scaled with the pileup probability:

$$\left(\frac{d\Gamma}{dE}\right)_{\text{pu}} = P_{\text{pu}} \cdot \int \frac{d\Gamma}{dE}(E - \varepsilon) \cdot \frac{d\Gamma}{dE}(E) d\varepsilon \quad (5.35)$$

For higher order pileup, e. g. triple event pileup, the pileup spectrum is iteratively convolved with the input spectrum again. In TRModel, the convolution is performed up to an arbitrary order and considers the non-zero bin width. The resulting spectra are shown in figure 5.30.

The condition for this description is a complete overlap of the pileup pulses such that the readout energy is the sum of the individual energies  $E_{\text{pu}} = E_1 + E_2$ . This is not necessarily the case for all readout systems or settings. In general, the height of a pileup pulse at the readout position will be a function of the event time difference  $\Delta t$  [319].

<sup>4</sup>K. Urban (personal communication, Feb., 21, 2022) conversation about data acquisition settings



**Figure 5.30:** Pileup spectra up to the third order calculated with the self-convolution method during a complete TRModel tritium spectrum evaluation with all systematics enabled.

For the trigger system shown in figure 5.28, the pileup energy computes to [253]

$$E_{\text{pu}}(\Delta t) = \begin{cases} E_1 + E_2 & \text{if } \Delta t < F/2 \\ E_1 + E_2 \cdot \left(1 - \frac{\Delta t - F/2}{P}\right) & \text{if } F/2 \leq \Delta t \leq P + F/2 \\ E_1 & \text{if } \Delta t > P + F/2 \end{cases}, \quad (5.36)$$

where  $P$  and  $F$  are the energy filter's peaking time and flat top length. So the condition for the convolution from equation (5.35) is met as long as the trigger pulse pair resolution  $\tau_{\text{min}}$  is below  $F/2$  and pileup rejection is used for the transition region. Given the expected value for resolution  $\tau_{\text{min}}$  stated above, this assumption is reasonable, considering that the flat top length can be increased if necessary. For a detailed discussion of the topic, including possible underestimation from the pileup energy due to signal rise time, see [253]. It is also shown that most of the pileup spectrum is shifted to energies above the main region of interest using post-acceleration, as illustrated in figure 5.31, among other beneficial mitigation aspects of using post-acceleration.

### 5.5.2 Electronic Noise

The waveform received by the ADC circuit is affected by electronic noise stemming from various sources [251], inducing fluctuations in the energy filter readout. This will lead to an energy-independent Gaussian broadening of the measured spectrum. The response matrix for electronic noise is calculated with the method described for Fano noise in section 5.4.3, which respects the non-zero width of the energy bins. The broadening from electronic noise is  $\sigma_{\text{en}} \approx 45 \text{ eV}^4$ . Quadratically adding the Fano noise contribution leads to an overall energy resolution of 240 eV FWHM at  $E = 20 \text{ keV}$ , which is in accord with measurements using prototype detectors [145].

### 5.5.3 Detection Threshold

Like the energy filter, the trigger filter is affected by noise. Hence, a threshold must be applied for the trigger filter to avoid too many accidental noise triggers. The corresponding cutoff in the energy spectrum is at the threshold energy  $E_{\text{thr}} \approx 1$  to 2 keV. Since the trigger filter operates independently from the energy filter, the cutoff is not sharp but has a specific width  $\sigma_{\text{thr}} \approx 150$  eV, depending on the filter settings.

Assuming Gaussian fluctuations of the trigger filter due to white noise, the probability of a signal with energy  $E$  fluctuating from below to above the threshold, or the other way around, is given by the integral of the normal distribution:

$$\int_{-\infty}^E \mathcal{N}(x|\mu = E_{\text{thr}}, \sigma_{\text{thr}}) dx \quad (5.37)$$

This can be identified as the cumulative Gaussian distribution  $\Phi(E|\mu = E_{\text{thr}}, \sigma_{\text{thr}})$ . Hence, a vector equaling  $\Phi$  evaluated at the center of each energy bin represents the threshold response. The same approach was successfully employed to model experimental data in [248]. Since the threshold is far below the region of interest for the deep spectrum measurement, this effect does not significantly influence the sensitivity. However, it is necessary to factor in the threshold when estimating the trigger rate  $r_{\text{tr}}$ , which is an essential input for the dead time efficiency model.

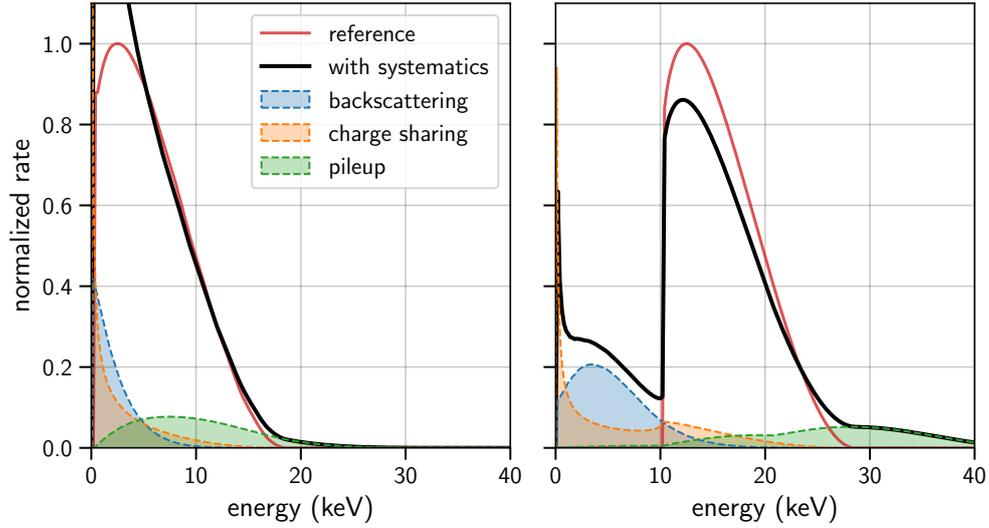
### 5.5.4 Dead time

Two sources of dead time are considered. The primary component is the trigger dead time  $\tau_{\text{dead}}$ , invoked after each trigger for pileup rejection. The time window length is dictated by the energy filter's peaking time  $P$  and flat top length  $F$ . For undistorted event energies, it must be at least  $P + F/2$ . The corresponding dead time efficiency loss increases exponentially with the trigger rate. The secondary component is a rate-independent reset dead time that accounts for the time window  $t_{\text{inhibit}}$  in which the readout system is inhibited to account for the periodic preamplifier reset. The reset rate must be sufficiently high to avoid waveform clipping at the upper limit of the dynamic range. Since all events are uncorrelated under the assumption of a Poisson process, the response for the detection efficiency is an energy-independent scalar. According to investigations with the DANTE digitizer [253], it is given by:

$$\epsilon_{\text{det}} = 1 - \epsilon_{\text{dead}} = (1 - f_{\text{reset}} \cdot t_{\text{inhibit}}) \cdot \exp(-2 \cdot r_{\text{tr}} \cdot \tau_{\text{dead}}) \quad (5.38)$$

The following values are assumed for the studies of chapter 6:  $t_{\text{inhibit}} = 51 \mu\text{s}$  and  $f_{\text{reset}} = 1$  kHz [251], and  $\tau_{\text{dead}} = 1.15 \mu\text{s}$  from the assumption of  $P = 1 \mu\text{s}$  and  $F = 0.3 \mu\text{s}$  [145].

Since dead time causes just an energy-independent scaling of the spectrum, and as the spectral amplitude is always treated as a free parameter in sensitivity studies, the effect is not associated with a systematic uncertainty. However, when a correct estimate of the measured output rate is required, which determines the statistics that can be accumulated in a given timeframe, this parameter will affect the statistical sensitivity.



**Figure 5.31:** Demonstration of the mitigation capabilities of employing post-acceleration (left:  $E_{\text{pae}} = 0$  keV, right:  $E_{\text{pae}} = 10$  keV). Due to the energy shift, a large fraction of charge-sharing and backscattering events falls below the main region of interest. Also, most of the pileup spectrum is shifted above the main spectrum.

### 5.5.5 Calibration

The pulse height spectra from the readout are provided in units of the ADC least significant bit (LSB). These can be related to energies using calibration measurements with monoenergetic electron and photon sources. The result of the calibration for each readout channel is a linear relation

$$E = G \cdot E_{\text{LSB}} + E_{\text{offset}} \quad (5.39)$$

with a gain factor  $G$  and an offset energy  $E_{\text{offset}}$ . The linearity must be experimentally verified, as is done in [145].

Since the calibration measurements themselves are subject to statistical and systematic uncertainties, the model must be capable of adjusting the energy scaling. The input bins are mapped to a set of output bins with a calibration response matrix to achieve this. The matrix is constructed with the binned random variable transformation method described in section 5.3.2 using the mapping function (5.39) from above. Since the number and range of input and output binning of this operation are not required to be equal, the response can be configured to match the bin edges from the measurement precisely.

---

## Sensitivity studies

---

The model described in the previous chapters is geared towards evaluating the keV sterile neutrino sensitivity of the tritium spectrum measurement with the TRISTAN Phase-1 detector. In the future, the model can also be used to search for other imprints of new physics in the differential tritium spectrum at KATRIN.

By modeling a large number of experimental effects, including systematic input parameters and estimation of the measured differential rate through a detailed normalization procedure (see sections 4.4.5 and 5.5), the model is capable of assessing the statistical and systematic uncertainties of the predicted spectrum. This allows to estimate sensitivity exclusion contours in the phase space of the parameters of interest, namely the mass  $m_s$  of a fourth neutrino eigenstate and its mixing amplitude  $\sin^2 \theta$  to the electron neutrino flavor.

This chapter demonstrates the model's present capacities to perform statistical and systematic uncertainty sensitivity studies. The primary goal is to showcase TRModel's capabilities with all available experimental effects being enabled. Subordinate purposes of this work are to provide preliminary estimates of the achievable statistical sensitivity limits and to determine which systematic effects will require special attention with respect to modeling and mitigation in the future. Several mitigation techniques from the previous chapter are tested by running all studies for two cases: a base scenario and an optimized scenario with improved beamline settings. Furthermore, it is investigated if the covariance matrix approach is appropriate to propagate systematic uncertainties.

TRModel does not yet include all effects discussed in chapter 4, and several systematic uncertainty inputs are still arbitrary (see table 6.2). Accordingly, the sensitivity contours shown here are preliminary. The methods and settings used for this work are included in the TRModel repository [254] and available to all users.

## 6.1 Sensitivity Calculation Method

The sterile neutrino sensitivity refers to the experiment’s capability to exclude the existence of a sterile neutrino imprint of a given magnitude in the measured spectrum. Since the measured spectrum and the fit model it is compared to are both subject to uncertainties, only a statistical statement with a specified confidence level is possible. As in previous studies [145, 243, 248, 252], the  $\chi^2$  test is used. Since the expected statistical error on the counts in each energy bin of the spectrum is small, the distribution of counts in any particular bin can be assumed as Gaussian. Then, the  $\chi^2$  test is equivalent to the likelihood-ratio test [265].

### Parameter types

The magnitude of the sterile neutrino signature is expressed by the mixing amplitude  $\sin^2 \theta$ . In addition, the mass  $m_s$  is unknown, so the statistical test contains two free parameters of interest. Moreover, the shape of the spectrum depends on a set of nuisance parameters  $\vec{p}$ , which can be grouped into three categories:

- Fixed parameters  $\vec{p}_{\text{fix}}$  carry neglectable uncertainties or are unrelated to spectral distortions that could mimic the sterile neutrino signature. All fixed parameters are listed in table 6.1.
- Free parameters  $\vec{p}_{\text{free}}$ , are considered to be entirely uncertain. These parameters can compensate for the sterile neutrino signature freely and are determined with a fit. For this work, two parameters fall into this category: The spectral amplitude and the amplitude of the pileup spectrum.
- Systematic parameters  $\vec{p}_{\text{syst}}$  that carry small uncertainties  $\vec{\sigma}_{\text{syst}}$  but have significant impact on the spectral shape. Their values are constrained through external measurements or generic assumptions, and their uncertainties are propagated to uncertainties on the bin contents of the energy spectrum. All systematic parameters and the assumed uncertainties are listed in table 6.2.

### Test statistic

The  $\chi^2$  test statistic measures the compatibility between data and a model prediction, where the latter depends on a hypothesis, i. e. the existence of a sterile neutrino with a particular mixing amplitude  $\sin^2 \theta$  and corresponding new mass state  $m_s$ . The observable is the binned differential (or integral) tritium spectrum, represented by a vector  $\vec{\Gamma}(\sin^2 \theta, m_s | \vec{p})$  that specifies the detected rate in each energy bin (for each retarding potential). For sensitivity studies, the model evaluated under the “no-mixing” hypothesis with  $\sin^2 \theta = 0$  and a fixed set of parameters  $\vec{p}_{\text{ref}}$  serves as a proxy for the data. This reference spectrum is denoted here as

$$\vec{\Gamma}_{\text{ref}} = \vec{\Gamma}(\sin^2 \theta = 0, m_s | \vec{p}_{\text{ref}}) . \quad (6.1)$$

The scalar  $\chi^2$  test statistic is calculated as follows [225, 265]:

$$\chi^2(\sin^2 \theta, m_s | \vec{p}) = \left( \vec{\Gamma}(\sin^2 \theta, m_s | \vec{p}) - \vec{\Gamma}_{\text{ref}} \right)^T V^{-1} \left( \vec{\Gamma}(\sin^2 \theta, m_s | \vec{p}) - \vec{\Gamma}_{\text{ref}} \right) \quad (6.2)$$

Here,  $V^{-1}$  is the inverse covariance matrix expressing uncertainties of bin contents and their correlations. For any randomly generated spectrum with the same underlying hypothesis as the reference, this test statistic will follow the  $\chi^2$  distribution with two degrees of freedom corresponding to the two parameters of interest.

### Covariance construction

The covariance matrix can be sub-divided into statistical and systematic contributions:

$$V = V_{\text{stat}} + V_{\text{syst}} \quad (6.3)$$

$V_{\text{stat}}$  is a diagonal matrix containing only the variances from Poissonian statistical uncertainty. Its elements are given by  $V_{ii} = (\Gamma_{\text{ref}})_i / t$  where  $t$  is the measurement time.

Monte Carlo sampling is the method of choice to obtain the systematics covariance. In this context, a large number  $N$  of spectra is generated, where the systematic parameter values are drawn from Gaussian distributions centered around their reference values, each with the corresponding standard deviation from  $\vec{\sigma}_{\text{syst}}$ . The elements of  $V_{\text{syst}}$  are constructed with the sample covariance estimator. By arranging the simulated data into a matrix  $D = (\vec{\Gamma}_1, \dots, \vec{\Gamma}_N)$  with the average spectrum  $\langle \vec{\Gamma} \rangle$ , the estimator can be written as [320]:

$$(V_{\text{syst}})_{jk} = \frac{1}{N-1} \sum_{i=1}^N \left( D_{ji} - \langle \vec{\Gamma} \rangle_j \right) \cdot \left( D_{ki} - \langle \vec{\Gamma} \rangle_k \right) \quad (6.4)$$

The calculation can either be performed for each systematic parameter individually, or by fluctuating all parameters simultaneously for a combined study. For sufficiently small uncertainties, the summation of all individual covariances is expected to yield equivalent results as in the case of the combined construction. In this work, both approaches are tested and compared with respect to their effect on the sensitivity. An alternative method of incorporating systematic uncertainties is adding pull terms to equation (6.2) and fitting the corresponding systematic parameters [225]. This approach is, however, more computationally expensive and thus not used in this work.

### Sensitivity contour

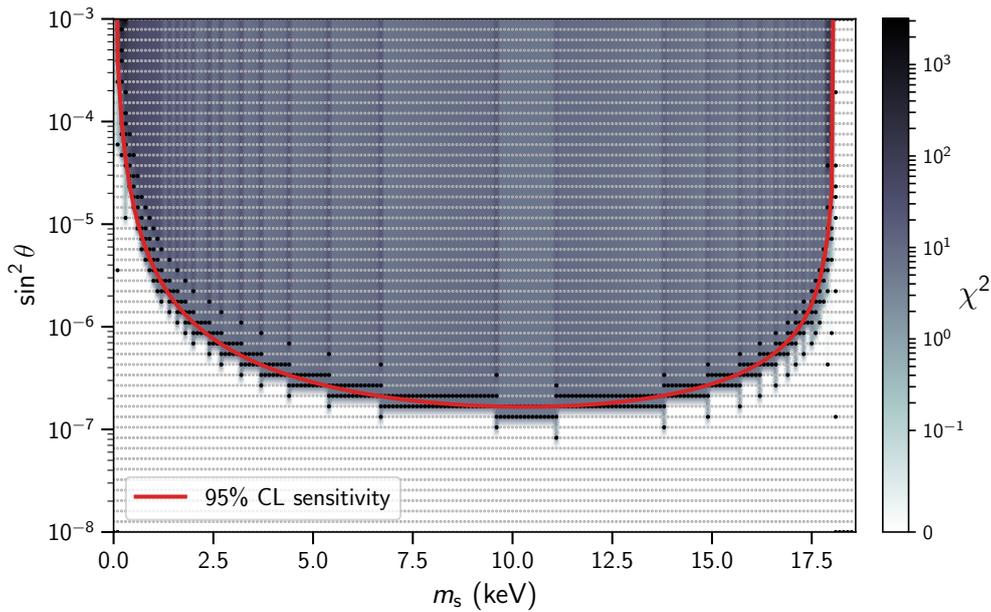
The  $\chi^2$  value can now be calculated for any given pair of  $\sin^2 \theta$  and  $m_s$  constituting a hypothesis. For increasing  $\sin^2 \theta$ , the deviation between the reference and the hypothesis spectrum will grow, resulting in a larger  $\chi^2$ . The hypothesis is rejected at a 95 % confidence level (CL) if  $\chi^2$  is larger than the critical value  $\chi_{\text{crit}}^2 = 5.99$ , which is calculated from the corresponding quantile of the  $\chi^2$  distribution.

To construct sensitivity contours in the  $(m_s, \sin^2 \theta)$ -plane where  $\chi^2 = \chi_{\text{crit}}^2$ , the values can be scanned in a grid, resulting in a two-dimensional  $\chi^2$  map. A grid spacing of 100 eV

in the  $m_s$  range, and a total of 50 points for  $\sin^2 \theta$  logarithmically spaced in the range  $10^{-7}$  to  $10^{-3}$  are chosen for the calculations presented here. In practice, only one contour is required, foregoing the need to scan the entire phase space. A contour-seeking algorithm was developed for this work, which saves a substantial amount of computation time. It scans the mass range in the positive direction, and for each fixed  $m_s$ , it determines the  $\sin^2 \theta$  interval that contains  $\chi^2_{\text{crit}}$ . Advancing to the subsequent  $m_s$  value, it is first tested if the interval is the same. If it does not contain  $\chi^2_{\text{crit}}$ , the interval is broadened, and a binary search algorithm is used to find the new interval efficiently. The result of the algorithm is shown in figure 6.1 for the 95 % CL contour.

To further increase the precision of the contour, the scaling properties of  $\chi^2$  regarding  $\sin^2 \theta$  are exploited. Since the sterile contribution of the spectrum is linear in  $\sin^2 \theta$ , the  $\chi^2$  values will depend quadratically on  $\sin^2 \theta$  according to equation (6.2). Like any power law, this leads to a linear relationship of their logarithms. Therefore, linear log-log interpolation is used to pinpoint the position where  $\chi^2$  crosses the critical value. In the following, the sensitivity contour for the 95 % CL is denoted by  $C_{95\%}$ :

$$C_{95\%}(m_s) = \sin^2 \theta(m_s) \Big|_{\chi^2=5.99} \quad (6.5)$$



**Figure 6.1:** Demonstration of the sensitivity grid scan and construction of the 95 % CL contour using the contour-seeking algorithm. Grid points are shown in gray, and points where the  $\chi^2$  is evaluated, are shown in black. The scan direction is from left to right. If the critical  $\chi^2$  is not contained in the same  $\sin^2 \theta$  interval as for the previous  $m_s$ -value, the window is broadened, and a binary search is employed to find the correct interval. Log-log-interpolation is used to find the precise location of the critical  $\chi^2$  within the interval.

## 6.2 Model Parameters and Settings

This section describes the criteria by which the parameter values for this study are selected. All parameters are summarized in tables 6.1 and 6.2. The detector model parameters are chosen according to the TRISTAN Phase-1 specifications (see chapter 3.4.4). Regarding the rest of the beamline, two scenarios are defined to demonstrate the sensitivity and how it may be improved:

- **Base scenario:** The beamline settings generally match the current operation mode, the sole adjustments being the column density and the retarding potential.
- **Optimized scenario:** Here, the electromagnetic field settings are adjusted to mitigate rear wall backscattering and detector systematics. Furthermore, the rear wall material is changed to beryllium. Specifics concerning the optimization process are given in section 6.2.2.

### 6.2.1 Parameters and Uncertainties

#### Magnetic and electric fields

As a baseline, the electromagnetic field parameters are chosen to match the standard operation settings of the KATRIN beamline [5], except for the retarding potential, which is decreased. The lowest possible retarding potential with a fully adiabatic transport through the main spectrometer is under investigation at the time of writing [243, 244]. Therefore, several  $qU$  setpoints are used for the statistical sensitivity: 0.6, 3.6, 8.6 and 13.6 keV. Since the sterile neutrino kink is located at  $18.6 \text{ keV} - m_s$ , the corresponding accessible mass ranges are 18, 15, 10 and 5 keV. For the study of the systematic impact, the broad 18 keV mass range is used to show for which  $m_s$  region the impact is most significant.

Uncertainty values for the source and pinch magnetic fields are at hand since these also represent systematic parameters in the neutrino mass analysis. The pinch field uncertainty of 0.1 % is taken from [7]. The reason for this comparatively small uncertainty is that it has been measured with a Hall probe during commissioning [321]. An uncertainty value of 0.25 % is used for the source magnetic field, which is retrieved with a novel method described in [282], where the precision of the pinch magnetic field is leveraged to determine the WGTS field in a comparative measurement. A similar method could be employed for the rear wall magnetic field<sup>1</sup>. Here, a benchmark uncertainty of 0.2 % is assumed. The same value is taken for the detector magnetic field. However, this could be improved with the same method employed at the pinch magnet<sup>1</sup>. The field of the inner WGTS magnetic traps and their uncertainty are detailed in section 5.1.2.

The retarding potential of the main spectrometer is stable and reproducible on the sub-ppm level [7, 322]. Nevertheless, an arbitrarily large uncertainty of 1 eV is used here to demonstrate that this effect can be neglected for differential measurements. Similarly, a sizeable uncertainty of 10 eV is assumed on the post-acceleration potential, where no documented value is available.

<sup>1</sup>F. Block (personal communication, Mar., 13, 2023) conversation about field uncertainties

### Source column density

In case of a decreased retarding potential, the WGTS tritium column density has to be lowered to avoid too large rates at the detector. Therefore, the column density normalization parameter  $\rho d_{\text{glob}}$  is adjusted for each setting, such that the output rate of the detector amounts to  $10^5$  cps per pixel. For the largest mass range, this results in setpoints of  $< 1\%$  of the nominal column density  $5 \times 10^{17} \text{ cm}^{-2}$  [5] (see table 6.1). In 2022, corresponding hardware tests showed that the WGTS can be operated with setpoints as low as  $0.01\%$ , and with permille stability at  $0.1\%$  [270]. The tritium purity is always assumed as  $95\%$  [5].

As described in section 5.1.1, the source scattering response is only available for specific column density ranges, which causes a discrepancy between the scattering parameter and the normalization parameter. Erring on the conservative side, the scattering parameter  $\rho d = \rho d_{\text{init}} = \rho d_{\text{side}}$  is chosen at the next larger setpoint, which is  $5 \times 10^{15} \text{ cm}^{-2}$  corresponding to  $1\%$  of the nominal value.

In the neutrino mass analysis, an uncertainty of  $0.25\%$  is used for the column density [7]. However, since it is unclear whether this value also applies to the operation at decreased column density in deep spectrum measurements, a considerably larger uncertainty of  $2\%$  is used here. The uncertainty on  $\rho d_{\text{trap}}$  is assumed to be as  $1\%$ . By construction, this parameter affects the relative fraction of trapped to untrapped portions of the source gas and does not add or subtract from the overall column density (see section 5.1.2).

### Rear wall parameters

In the base scenario, the rear wall backscattering response from the Geant4 Simulation with gold as a target is used, while for the optimized scenario it is switched to beryllium. In both cases, the response was simulated using two different Geant4 physics lists to obtain a measure for the shape uncertainty of the backscattering spectrum (see section 5.2.1). In the model implementation, the two responses are assigned scaling factors  $A_{\text{rws},1}$  and  $A_{\text{rws},2}$ , and the following reparametrization allows to disentangle the sensitivity impact of the overall backscattering amplitude and the shape uncertainty:

$$A_{\text{rws},1} = \alpha_{\text{bs}} \cdot (1 - \delta) \quad \text{and} \quad A_{\text{rws},2} = \alpha_{\text{bs}} \cdot \delta \quad (6.6)$$

Here,  $\delta$  is the shape parameter, and  $\alpha_{\text{bs}}$  defines the overall combined amplitude. The rear wall backscattering response can then be written as

$$R_{\text{rws}} = \alpha_{\text{bs}} \cdot [(1 - \delta)R_1 + \delta R_2] \quad (6.7)$$

For the reference value, the two responses are mixed in equal amounts with  $\delta = 0.5$  and  $\alpha_{\text{bs}} = 1$ . Then, Gaussian uncertainties are applied to the two parameters so as to construct the covariance.

A similar construction is used for the rear wall activity spectrum since its calculation is based on the backscattering simulation (see section 5.2.2). Consequently, the activity shape parameter  $\delta$  is assumed to equal the backscattering shape parameter. The scaling factor, i. e. the ratio of the rear wall and WGTS source activities, is chosen as  $\alpha_{\text{act}} = 10^{-2}$ .

**Table 6.1:** Fixed model parameters for the sensitivity studies.

Category	Parameter	Symbol	Base value	Optimized
Normalization	Column density	$\rho_{d_{\text{glob}}}$	$1.53 \times 10^{15} \text{ cm}^{-2}$	$3.10 \times 10^{15} \text{ cm}^{-2}$
	Tritium purity	$\epsilon_{\text{T}}$	95 %	-
	Pixel count	$n_{\text{px}}$	936	-
	Pixel efficiency	$\epsilon_{\text{px}}$	100 %	-
Detector	Pixel radius	$R_{\text{px}}$	1.5 mm	-
Readout	Reset frequency	$f_{\text{reset}}$	1 kHz	-
	Inhibit duration	$t_{\text{inhibit}}$	51 $\mu\text{s}$	-
	Trigger threshold	$E_{\text{thr}}$	2 keV	-
	Threshold width	$\sigma_{\text{thr}}$	150 eV	-
	Trigger dead time	$\tau_{\text{dead}}$	1.15 $\mu\text{s}$	-
Background	Differential rate	$r_{\text{bkg}}$	$6 \times 10^{-7} \text{ cps/keV}$	-

This conservative estimate is based on measurements, which extend to 1.6 keV below the endpoint, where the rear wall activity was found to be well below the 1 % level throughout several neutrino mass measurement campaigns [304].

A sizeable uncertainty of 10 % is attached to all rear wall parameters due to the fact that measurements of the backscattering coefficient show significant discrepancies when compared to simulations [243]. In the future, this can potentially be improved via measurements at dedicated test setups.

### Detector and readout parameters

The detector “golden” pixel count and pixel diameter are chosen according to the specifications of the TRISTAN Phase-1 detector with 9 modules (see section 3.4.4). The adopted values for all other detector and readout-related parameters are given in the corresponding sections of chapter 5.

The uncertainties of 2 nm for the dead layer parameter and of 1  $\mu\text{m}$  for the charge cloud width are based on measurements, as mentioned in sections 5.4.2 and 5.4.4. For back-reflection, a sizeable uncertainty of 20 % is assumed for identical reasons as in the case of rear wall backscattering. Regarding noise, a significant uncertainty of 20 % is used for  $\sigma_{\text{en}}$  and  $F_{\text{fano}}$  to demonstrate that their impact on the sensitivity can be neglected. The same is done for the pulse pair resolution determining the pileup spectrum magnitude. However, the shape uncertainty on the pileup spectrum is currently being neglected. If taken into account, the impact of the time resolution uncertainty may be more dramatic, as shown in prior work [253]. Readout gain and offset uncertainties are roughly motivated by the characterization measurements from [145].

**Table 6.2:** Systematic parameter values for the base/optimized scenario, assumed uncertainties, and uncertainty sources.

Category	Parameter	Symbol	Base value	Optimized	Uncertainty	Source
Source	Column density total	$\rho d$	$5 \times 10^{15} \text{ cm}^{-2}$	-	2%	arbitrary
	Column density trap	$\rho d_{\text{trap}}$	$5 \times 10^{14} \text{ cm}^{-2}$	-	1%	arbitrary
Rear wall	Trap magnetic field	$B_{\text{trap}}$	$B_{\text{src}} - 20 \text{ mT}$	-	3 mT	motivated
	Backscattering amplitude	$\alpha_{\text{bs}}$	1 (Au)	(Be)	10%	motivated
	Activity amplitude	$\alpha_{\text{act}}$	$10^{-2}$	-	10%	arbitrary
Transport & Spectrometer	Backscattering shape	$\delta$	0.5	-	10%	arbitrary
	Magnetic field rear wall	$B_{\text{rw}}$	1.24 T	0.32 T	0.2%	arbitrary
Spectrometer	Magnetic field source	$B_{\text{src}}$	2.52 T	-	0.25%	data
	Magnetic field pinch	$B_{\text{pinch}}$	4.20 T	2.57 T	0.1%	data
	Magnetic field detector	$B_{\text{det}}$	2.52 T	1.00 T	0.2%	motivated
Detector	Retarding potential	$qU$	500 eV	-	1 eV	arbitrary
	Post acceleration energy	$E_{\text{pae}}$	10 keV	20 keV	10 eV	arbitrary
	Back reflection amplitude	$A_{\text{ref}}$	1	-	20%	arbitrary
Readout	Dead layer parameter	$\lambda$	58 nm	-	2 nm	data
	Charge cloud width	$w_{\text{cc}}$	20 $\mu\text{m}$	-	1 $\mu\text{m}$	data
	Fano factor	$F_{\text{fano}}$	0.115	-	20%	arbitrary
Calibration gain	Pulse pair resolution	$\tau_{\text{min}}$	112 ns	-	20%	arbitrary
	Electronic noise	$\sigma_{\text{en}}$	44 eV	-	20%	arbitrary
	Calibration gain	$G$	1	-	0.01%	motivated
Calibration offset	$E_{\text{offset}}$	0 eV	-	10 eV	motivated	

## 6.2.2 Beamline Optimization

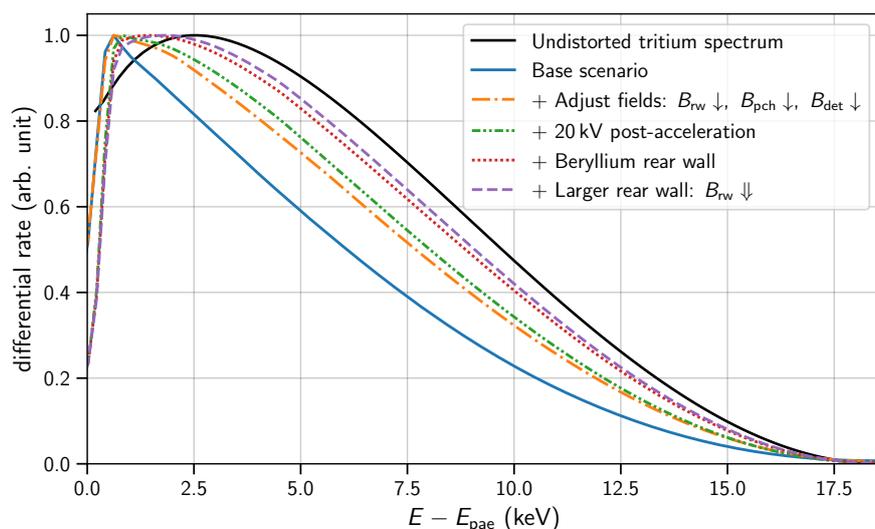
The beamline operation can be adjusted in several ways to mitigate the impact of dominant systematics, such as rear wall and detector effects. Many of these techniques are discussed in detail in the corresponding sections of chapter 5. The key aspects are briefly summarized here and categorized into invasive and noninvasive methods, i. e. adjustments that can be performed with the current beamline hardware, and ones that would require hardware improvements:

- Noninvasive:
  - Decrease the rear wall magnetic field to improve the trapping of electrons emitted or backscattered from the rear wall. In this case, the detector field must also be lowered to map all detector field lines to the rear wall.
  - Decrease the pinch magnetic field to improve the rear wall-detector to source-detector acceptance ratio.
  - Decrease the detector magnetic field to enhance the collimation of the angular distribution, thus reducing backscattering.
- Invasive:
  - Increase the post-acceleration potential from 10 kV to 20 kV, to further mitigate charge sharing, energy deposition losses, back-reflection, and signal pileup. At the time of writing, the post-acceleration voltage is limited to 12 kV [240].
  - Exchange of the rear wall for one made of a low- $Z$  material, ideally beryllium.
  - Increase the diameter of the rear wall to allow for a significantly decreased rear wall magnetic field and relocate it such that the field lines are perpendicular to the surface.

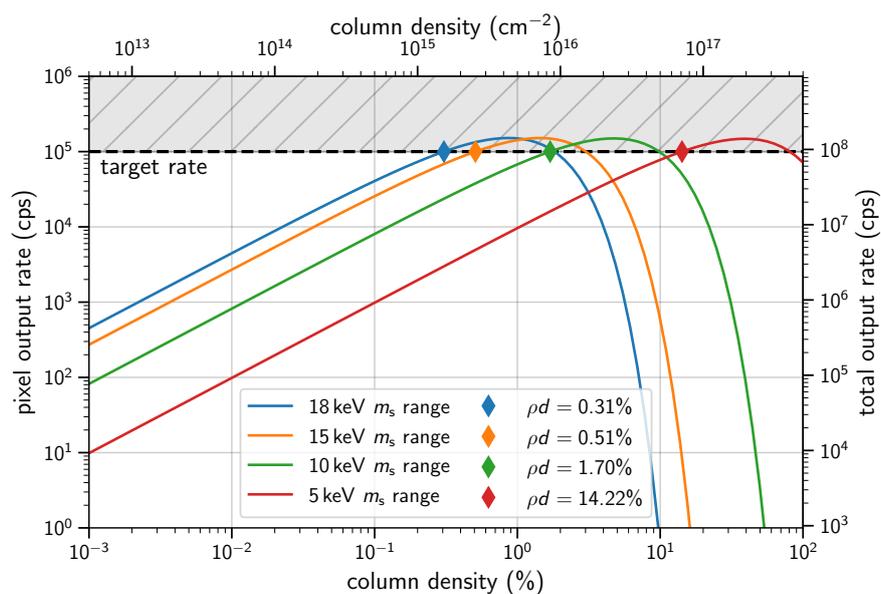
For further explanations concerning rear wall-related mitigation methods and a quantification of their impact, see section 5.11. As an outcome of all these techniques, the distortion of the measured tritium spectrum will be significantly reduced, as shown in figure 6.2. To make the sensitivity study presented in this work concise, all mitigation methods are applied simultaneously in the *optimized* scenario.

### Rate optimization

The output rate of the detector will depend on the column density, the retarding potential, and the acceptance defined by the magnetic fields. For very high input rates, the dead time induced by the trigger of the readout system for pileup rejection will lead to a substantial decrease in detection efficiency. Thanks to the normalization procedure detailed in sections 4.4.5 and 5.5, the model is capable of estimating the detected rate. For optimal statistics, the output rate can be maximized for each field configuration by adjusting the column density, as shown in figure 6.3. As a safety margin, a limit for the pixel output rate of  $10^5$  cps is used according to the design goal from [145]. At this rate, the detection efficiency is  $\epsilon_{\text{det}} = 65.8\%$  based on the formula from section 5.5.4.



**Figure 6.2:** The distortion of the measured differential tritium spectrum from experimental effects can be alleviated with several beamline optimizations. Adjusting the magnetic fields can be done with the beamline in its current state. All other optimizations require hardware changes.



**Figure 6.3:** Determination of the appropriate column density for the target output rate per pixel of  $10^5$  cps. Above this threshold, dead time becomes a significant issue and decreases the measured rate at high column density. The rate and the optimal column density setpoint depend on the retarding potential, which determines the accessible sterile mass range. The optimal values are given in percent of the nominal column density  $5 \times 10^{17} \text{ cm}^{-2}$ . Shown is the result for the base scenario parameter set. For the optimized setting values, see appendix A.2.

## 6.3 Statistical Sensitivity

### 6.3.1 Differential Mode

The statistical sensitivity can be calculated using the methods presented in section 6.1. As stated in the previous section, four distinct retarding potential setpoints are considered: 0.6, 3.6, 8.6 and 13.6 keV. These settings correspond to accessible mass ranges of 18, 15, 10 and 5 keV. Using the method outlined in figure 6.3, the appropriate value for the normalization parameter  $\rho d_{\text{glob}}$  is determined for both scenarios and all selected retarding potentials. The measurement time is assumed as 1 full year, not accounting for calibration intervals or other potential downtimes. With the output count rate limit of  $10^5$  cps per pixel and the “golden” pixel selection  $n_{\text{px}} = 936$  this will result in a total of

$$N_{\text{evt}} = 10^5 \text{ cps} \cdot n_{\text{px}} \cdot 1 \text{ y} = 2.954 \times 10^{15} \quad (6.8)$$

registered events, regardless of the field setting, due to the rate adjustment. When running the sensitivity scans with these numbers, the 95% CL contours shown in figure 6.4 are obtained. The scans show that a statistical sensitivity minimum of  $\sin^2 \theta \approx 2 \times 10^{-7}$  can be achieved. Furthermore, the optimized scenario exhibits a slightly better sensitivity which can be attributed to a less distorted shape of the tritium spectrum (see figure 6.2).

While the statistical sensitivity minimum is unaffected by the mass range setting, narrowing the range is beneficial to increase the sensitivity in the lower mass region. This could be leveraged to achieve specific physics goals, such as focusing the search on testing potential signal indications from other experiments within the lower mass range.

### Time evolution

To facilitate decisions on the measurement strategy, it is very helpful to consider the sensitivity evolution as a function of the collected statistics or, equivalently, as a function of measuring time at a given detection rate. Notably, the general shape of the sensitivity contour is independent of the statistics. Therefore, it is sufficient to look at the sensitivity time evolution of a single mass. Here a mass value of  $m_s = 10$  keV is chosen, which is an ideal test case given the 18 keV phase space. Performing sensitivity scans for this fixed mass at different statistics yields the sensitivity

$$C_{95\%}^{\text{stat}}(N, m_s = 10 \text{ keV}) = 10^{-\log_{10}(N)/2+k} = \frac{10^k}{\sqrt{N}} \quad \text{with } k \approx 1. \quad (6.9)$$

The constant  $k$  slightly deviates from unity, depending on the scenario:  $k = 0.998$  for the base scenario and  $k = 1.088$  for the optimized one. The result is also shown in figure 6.5. As expected, the evolution with  $N$  follows the scaling of the Poissonian relative statistical uncertainty  $1/\sqrt{N}$ , from which the statistics covariance is constructed. This exact scaling no longer holds when systematic uncertainties are considered, which is further investigated in section 6.4.3.

### 6.3.2 Integral Mode

Another strategic decision could be to make use of the integral mode for complementary measurements since the systematics there are much different (see table 4.2 in chapter 4). For example, should a signal indication be found with the differential mode (potentially false positive), it could be further tested with the integral mode. Therefore it is very helpful to compare the statistical sensitivity of both methods.

The integral mode offers the freedom to choose the so-called measurement time distribution (MTD), which defines the fraction of measurement time spent at each retarding potential setpoint. Here, a uniform (flat) MTD with 50 retarding potential setpoints spread over the 18 keV mass range is used as a basis. To investigate whether the integral search can be specialized to search for one specific mass value, three more MTDs are defined for an arbitrarily chosen mass of interest  $m_s = 13$  keV. These are a window MTD of  $\approx 5$  keV width centered on the expected kink position, a Gaussian MTD with 1 keV standard deviation around the kink position, and an MTD with the shape of the expected residual from the sterile neutrino signature. All MTDs are shown in figure 6.6a.

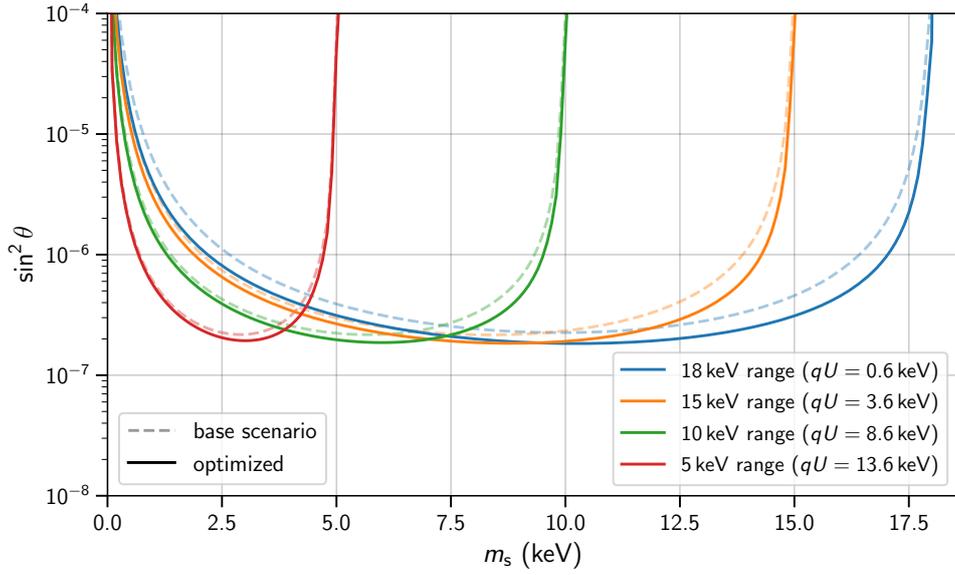
Another choice in the integral mode is the region of interest where events from the differential detector spectrum are integrated. Here, it is opted to integrate the entire energy spectrum.

Since the integral spectrum calculation is slower than the differential one, only the global amplitude parameter is considered free in the fit to speed up the computation. The pileup amplitude, which was fit in the differential study, is neglected here, and pileup is deactivated in the spectrum calculation accordingly. Furthermore, detector back-reflection is deactivated to ensure the validity of the fast approximation method for the integral spectrum from section 4.5.

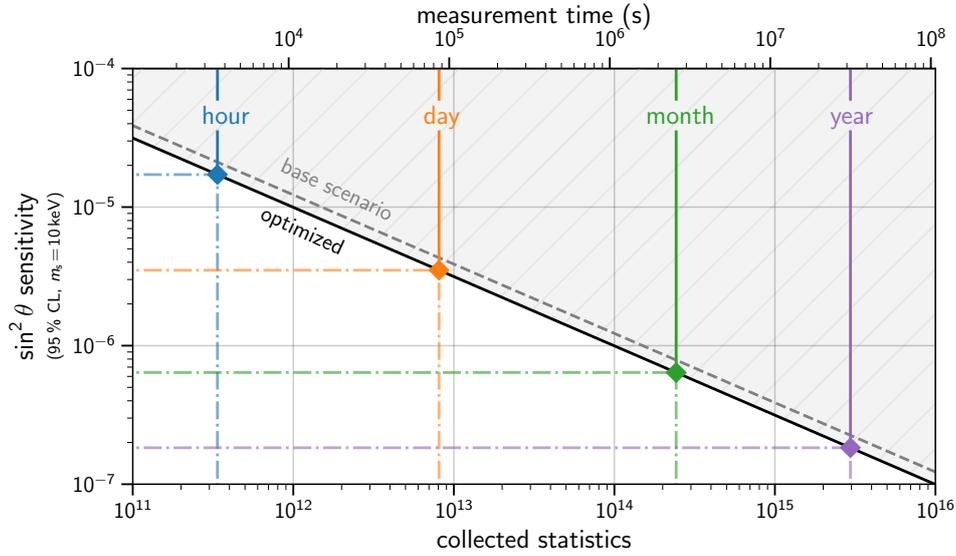
Figure 6.6 shows the result of all scans. Compared to the statistical sensitivity in the differential mode, the sensitivity in the integral mode is reduced by almost an order of magnitude for higher sterile masses, while the lower masses are affected less. The overall loss in sensitivity is expected due to the measurement time spent at high retarding potentials where the rate is low. This will reduce the overall statistics. When comparing the reference spectra from the integral scan (flat MTD) and the differential spectrum, the detected rate and total statistics are reduced to  $\approx 30\%$ .

When one attempts to optimize the MTD for one particular mass, none of the MTD choices will yield a significant improvement. The window and Gaussian MTDs even reduce sensitivity across the entire mass range. Only the MTD shaped after the signature leads to a slight increase in sensitivity at the high-mass region while sacrificing some in the low-mass region.

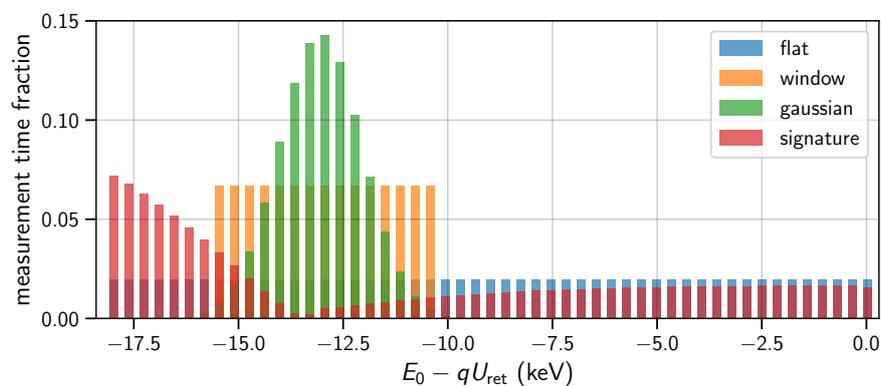
Nevertheless, the statistical sensitivity of the integral mode surpasses the  $10^{-6}$  level, so can be considered as a complementary method. However, further studies and model improvements are mandatory to investigate the impact of systematic effects on the integral deep spectrum measurement.



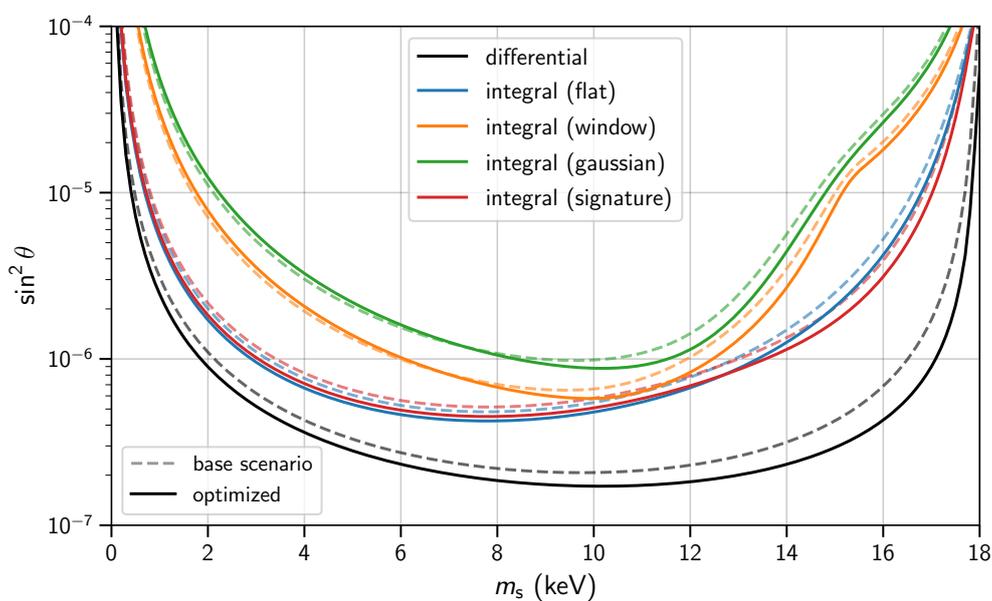
**Figure 6.4:** Statistical sensitivity curves for exclusion of a sterile neutrino at 95% confidence level for one year of data taking, without accounting for downtime. All curves correspond to statistics of  $N_{\text{evt}} \approx 2.95 \times 10^{15}$  registered events. For the calculation, the normalization parameter  $\rho d_{\text{glob}}$  is chosen according to the values from figure 6.3. The statistical sensitivity in the base scenario is slightly worse, even without accounting for systematic uncertainty, since the spectrum is much more distorted (see figure 6.2).



**Figure 6.5:** Evolution of the statistical sensitivity in differential mode as a function of collected statistics at a 95% confidence level for a fixed mass  $m_s = 10$  keV. For the translation to measurement time, an output rate per pixel of  $10^5$  cps and 936 illuminated pixels corresponding to the TRISTAN Phase-1 golden pixel selection are assumed (see section 3.4.4)



(a) Measurement time distributions



(b) Sensitivity countours

**Figure 6.6:** Comparison of statistical sensitivity contours (95% confidence level) for the differential and integral measurement modes (b). Several measurement time distributions (a) are tested for the integral measurement mode to investigate whether an optimization concerning a particular mass value is possible. The measurement time distribution labeled “signature” takes the form of the integral spectrum ratio for  $m_s = 13$  keV and the no-mixing case.

## 6.4 Systematic Uncertainty Impact

Systematic uncertainties can be included in the  $\chi^2$  calculation via the covariance matrix method described in section 6.1. As explained there, the covariance matrix is obtained by propagating the uncertainties of systematic parameters (see section 6.2) to the covariance of the energy bin contents of the spectrum via Monte Carlo sampling of the parameter with an underlying Gaussian distribution, calculating the spectrum for each, and employing the covariance estimator. A total of  $N_{\text{ind}} = 10^3$  spectra are generated for each individual parameter study. For a combined impact of all 20 parameters, the covariance is constructed in two ways by

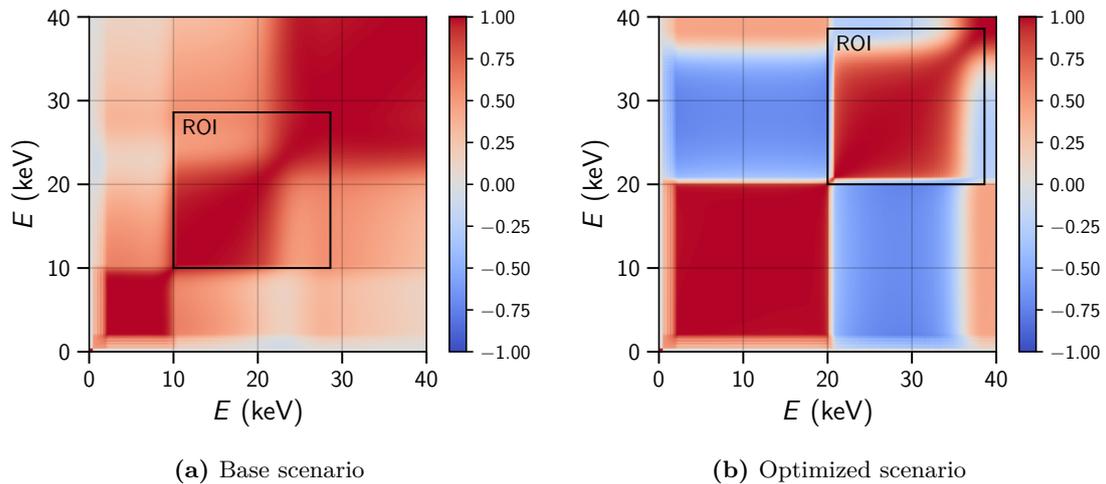
1. taking the sum of all individual covariance matrices.
2. sampling all parameters simultaneously for a combined covariance estimation.

For the second method, the number of generated spectra is increased to  $N_{\text{comb}} = 10^4$ . As a visual example, the covariance matrices from the combined construction are normalized by their diagonal elements, yielding the correlation matrices in figure 6.7.

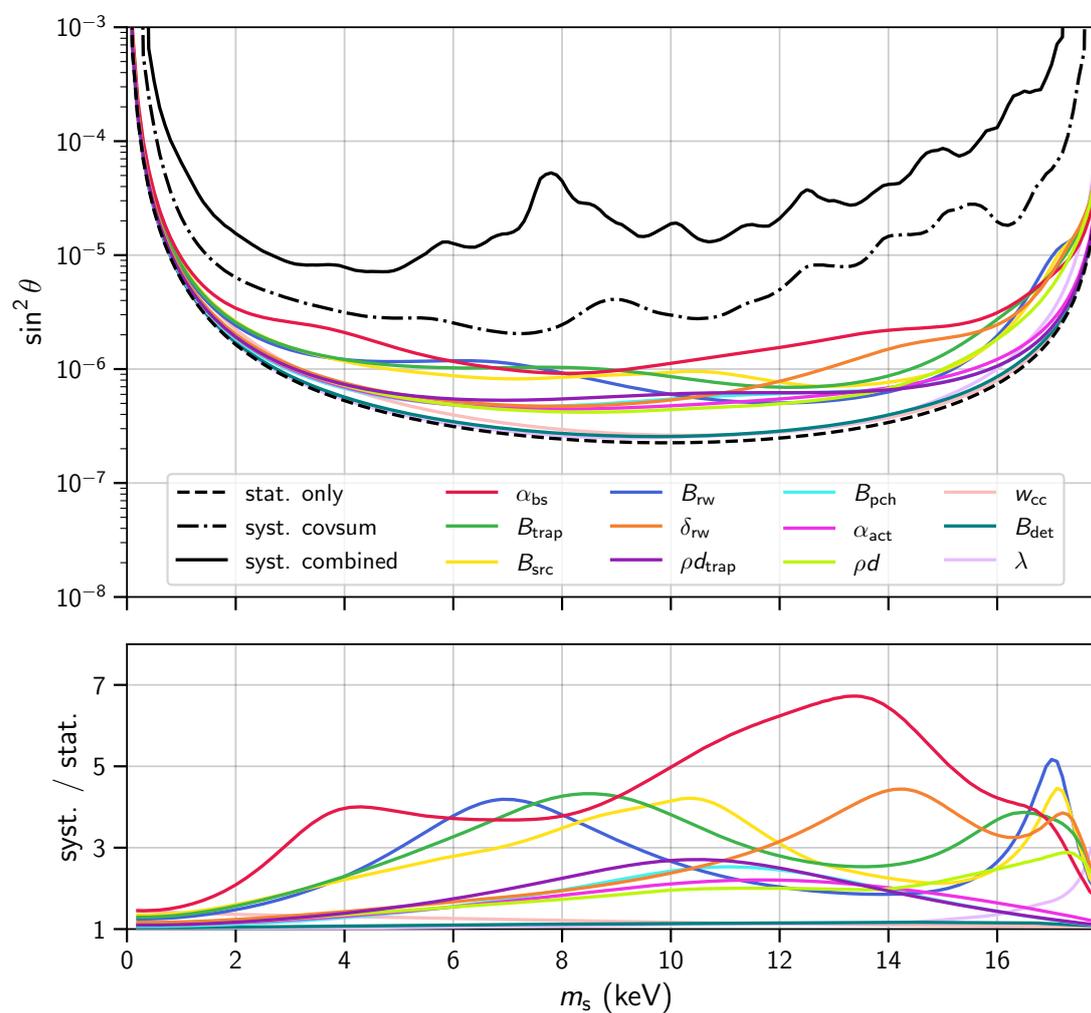
Next, all scans are performed with the inclusion of the covariance matrices using the contour-seeking method described in section 6.1. The resulting 95% CL contours are shown in figure 6.8. A decomposition of the plots into smaller groups of contours can be found in appendix A.3. As a rough scalar measure of the significance of individual effects, one can define the average weighted sensitivity loss factor as:

$$\langle f_s \rangle = \frac{1}{\sum_i w_i} \cdot \sum_i w_i \frac{(C_{95\%}^{\text{stat}})_i}{(C_{95\%}^{\text{syst}})_i} \quad \text{where} \quad w_i = \frac{1}{(C_{95\%}^{\text{stat}})_i} \quad (6.10)$$

A breakdown of all factors from individual parameter scans is shown in figure 6.9.

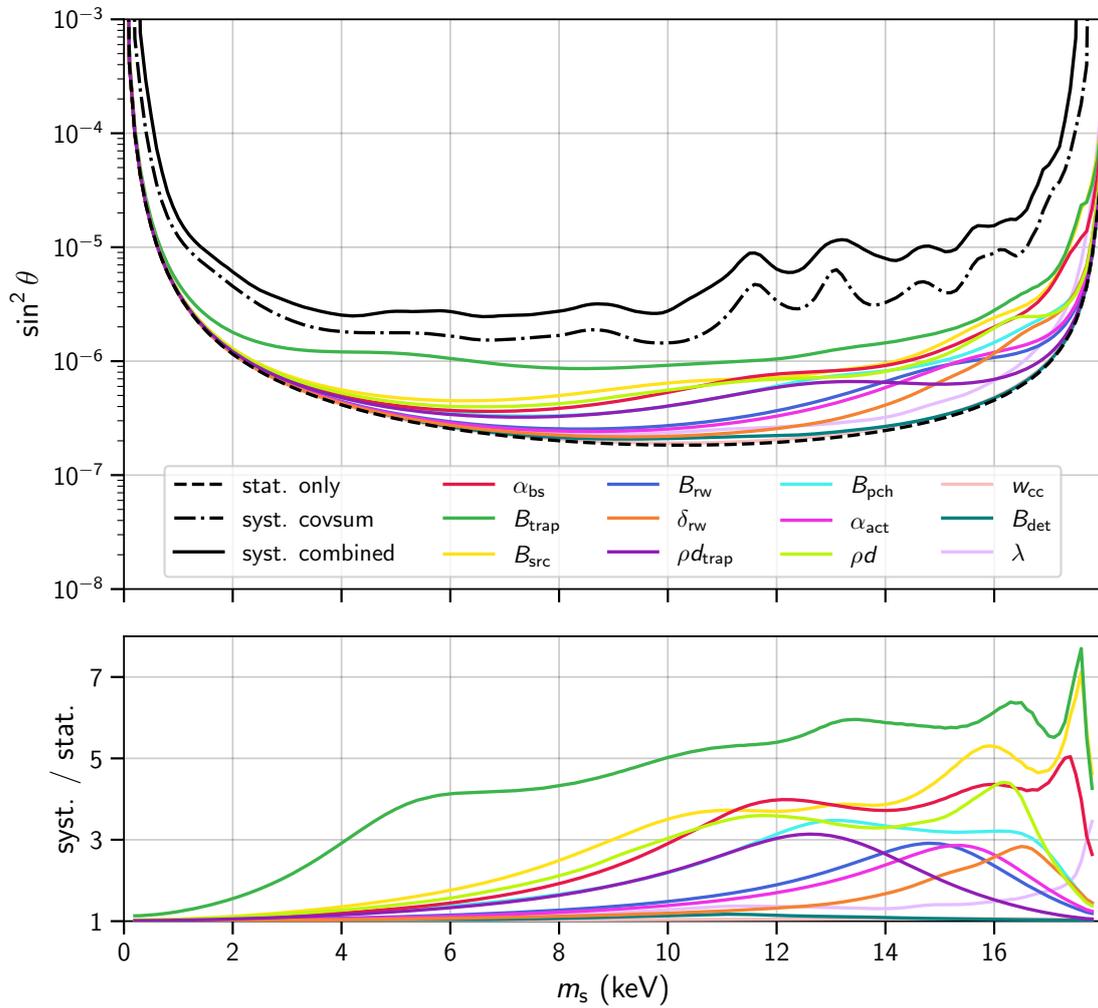


**Figure 6.7:** Correlation matrices obtained via simultaneous fluctuation of all parameters.

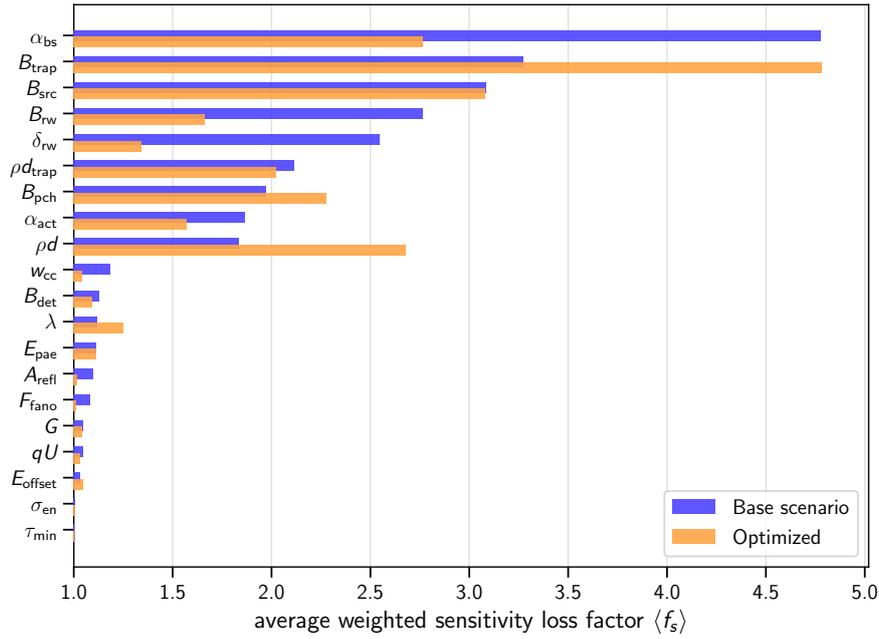


(a) Base scenario

**Figure 6.8:** Impact of systematic uncertainties on the sensitivity (95% confidence level) for both scenarios (a), (b). The top plot shows the resulting sensitivity contours for the individual parameter uncertainties and their combined impact. Effects with  $\langle f_s \rangle < 1.1$  are not shown for easier viewing. The bottom plot shows the ratio of the individual contours and the statistics-only contour. For the combined study, the covariance is estimated in two ways: Summation of all individual covariances (labeled “covsum”) and simultaneous fluctuation of all parameters (labeled “combined”).



(b) Optimized scenario



**Figure 6.9:** Breakdown of systematic parameter impacts on the sensitivity for both scenarios, as quantified by the average weighted sensitivity loss factor  $\langle f_s \rangle$  from equation (6.10). See table 6.2 for information regarding each parameter.

#### 6.4.1 Individual Effect impacts

The results of this study evidently are of preliminary nature since the model still has to incorporate essential effects such as fast back-reflection. Furthermore, most uncertainty estimates have to be refined. Several models describing important processes, such as the magnetic trap and the rear wall backscattering model, at present lack sound experimental validation. However, an interpretation of the results is very helpful, keeping in mind that a future interpretation in the future based on an improved model and refined inputs is vital. Furthermore, additional settings could be constructed to pinpoint the effect of each adjustment.

The individual parameter scans and the breakdown from figure 6.9 show that the rear wall backscattering amplitude dominates in the baseline scenario, followed by the magnetic trap field depth and the source magnetic field. Other rear wall and source-related effects also lead to a considerable impact. The optimized scenario will improve rear wall systematics drastically, as was the aim. At the same time, however, the impact of the trap depth, source scattering, and pinch magnetic field will increase. This may be a side effect of the decreased pinch magnetic field since, at an increased acceptance, the fraction of source electrons with large pitch angles propagating to the detector will rise. These electrons accordingly will have a longer effective path through the source and thus are more likely to scatter, as explained in section 5.1.1. Furthermore, the fraction of electrons from magnetic trap electrons advancing to the detector will increase, especially since their

angular distribution is de-collimated when exiting the trap field. Overall, this suggests that there may be an optimal pinch magnetic field between the two scenarios, where the combined sensitivity impact from the rear wall and source-related effects is minimized. From the modeling perspective, this underlines the importance of the interplay between different systematics when considering beamline optimizations, and it demonstrates the core advantage of the holistic TRModel approach.

Overall, the magnetic fields will have a considerable impact on the result. This is to be expected as they govern the transport along the beamline, thus exponentially amplifying other systematics' influence. For instance, when the pinch magnetic field fluctuates upward, an additional fraction of electrons is reflected toward the rear wall, thus increasing backscattering. This hypothesis is supported by the fact that the magnetic field at the detector has a much lower impact since it only collimates electrons and has no reflection component in the propagation model (see section 4.4.2).

An overall trend is observed when moving from the baseline to the optimized scenario: The sensitivity is mainly restored in the low-mass region. Figure 6.8b shows that the trap magnetic field represents a bottleneck with its impact being largely unaffected in this parameter space. In future studies, it can be attempted to decrease the source magnetic field as a mitigation technique, as proposed in [243].

Another general outcome is the apparent divide between dominant and sub-dominant systematics visible in the breakdown. The sub-dominant effects include the retarding potential, all fields downstream of the pinch magnet, and all detector and readout effects. Of these effects, charge-sharing is significantly reduced in the optimized scenario, which can be attributed to the increased post-acceleration energy. However, the impact of the dead layer parameter  $\lambda$  worsens, possibly due to the broadened angular distribution caused by the increase of the acceptance angle. Back-reflection, on the other hand, is much improved. As of now, it is still a minor effect. But with the inclusion of fast back-reflection, the impact and its mitigation by increased post-acceleration may become much more influential.

Readout effects are almost entirely negligible in this study. However, minute shape uncertainties of the pileup spectrum are currently not included. These can be expected to produce a more significant impact based on what was demonstrated in a previous work [253]. Also, the effect of the calibration parameters will require further exploration in terms of systematic uncertainty input and drifts over time.

### 6.4.2 Combined Impact

When combining all systematics the sensitivity will be reduced by roughly one to two orders of magnitude depending on the scenario, method, and mass value. Comparing the sensitivity curves from the two approaches to construct the covariance, it is evident that the methodologies are not interchangeable. The baseline scenario mainly exhibits a significant discrepancy in the central mass region. In the optimized scenario, the curves are much more similar. But in both scenarios, the construction via simultaneous fluctuation of parameters will lead to a heightened impact on the sensitivity. Nevertheless, the combined

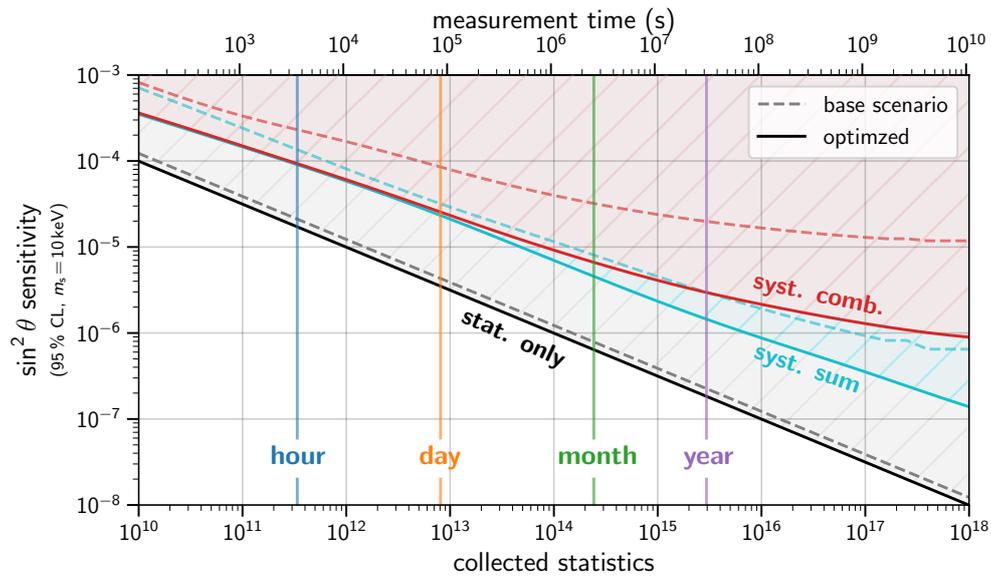
sensitivity is much improved for the optimized scenario. In general, the results presented here agree with those from an earlier study also performed with TRModel [323].

A potential reason for the dissimilarity of the covariance construction methods could be the magnitude of some uncertainties and the poor linearity of the model concerning systematic parameters. Since the covariance method is based on a linear expansion of the spectrum model [265], sufficient linearity is a prerequisite for its application. The other preliminary study with TRModel shows that considering uncertainties by fitting will lead to improved results and more stability of the acquired contours [323]. Hence, the inclusion of systematics via pull terms should be explored. Ultimately, a means of assessing the accuracy of each error propagation method is required. For this, the Monte Carlo propagation method [324] adopted for the KATRIN neutrino mass analysis [222] can be considered.

### 6.4.3 Sensitivity Time Evolution

In section 6.3.1, it was shown that the statistical sensitivity scales with the relative statistical error  $1/\sqrt{N}$ . The same calculation is repeated here, including all effects using the summed and the combined covariance matrices. Apart from that, the sensitivity calculation as a function of the collected statistics is performed in the same manner as before for the fixed mass value  $m_s = 10$  keV.

The result of the calculation is shown in figure 6.10. It shows that the sensitivity with systematics no longer strictly follows the  $N^{-0.5}$  power law and gradually levels out in case of high statistics. However, in the time domain of up to several years of data taking with TRISTAN Phase-1, the sensitivity still scales well with statistics. The ultimate sensitivity that can be reached strongly depends on the scenario and the description of systematics. Thus, further investigations into uncertainty propagation techniques are required to obtain a reliable sensitivity estimate.



**Figure 6.10:** Evolution of the sterile neutrino sensitivity as a function of collected statistics at 95% CL for a fixed mass  $m_s = 10$  keV. For the translation to measurement time, an output rate per pixel of  $10^5$  cps and 936 illuminated pixels corresponding to the TRISTAN Phase-1 golden pixel selection are assumed (see section 3.4.4). Systematic uncertainties are included via the covariances from the summation of individual effects and the combined construction.



---

## Conclusion and Outlook

---

The extension of the SM with sterile neutrinos is a natural way of introducing neutrino mass terms to the theory. Depending on their mass, sterile neutrinos may explain many open questions in particle physics and cosmology. For instance, a sterile neutrino on the keV scale would be an excellent candidate for Warm Dark Matter. Several vital aspects concerning keV sterile neutrinos are provided in section 2.6.2.

By performing a differential measurement of the tritium  $\beta$ -spectrum to energies deep below the endpoint, the KATRIN experiment can be sensitive to sterile neutrinos with masses  $m_s \lesssim 18$  keV. KATRIN's first deep spectrum measurement will be performed with the TRISTAN Phase-1 detector, which will be integrated into the KATRIN beamline in 2026. In section 6.3, the statistical sensitivity for a 1 y live-time measurement with the Phase-1 detector was calculated using a new model described in detail throughout chapters 4 and 5. It was shown that with an optimized detector rate, approximately  $2.95 \times 10^{15}$  registered events could be expected, which translates to a statistical sensitivity with a minimum of  $\sin^2 \theta < 2 \times 10^{-7}$  at 95 % CL.

In section 2.6.2, the Phase-1 reach was put into perspective in the broader field. Concerning laboratory experiments, the Phase-1 statistical sensitivity surpasses existing limits by more than two orders of magnitude. An overview of cosmological keV-scale sterile neutrino searches has also been provided. These searches usually assume that all Dark Matter is composed of sterile neutrinos, which results in constraints on the parameter space that strongly disfavor the range accessible by present and future laboratory limits. In the KATRIN mass range, thermal overproduction and X-ray bounds exclude the region above  $\sin^2 \theta \approx 10^{-8} - 10^{-12}$ . Nevertheless, laboratory experiments have significant merit due to their model independence. Section 2.6.2 also summarizes an argument by another study, showing that cosmological bounds can be realistically weakened, which re-establishes KATRIN's importance for cosmology cross-checks.

The central result of this thesis is the successful development of a new differential spectrum model for KATRIN named TRModel, which is geared towards holistically assessing the sensitivity impact of systematic effects. It uses a forward-convolution approach in which the differential tritium spectrum is convolved with response matrices to predict the energy spectrum of registered events. The response matrices representing the influence of individual systematic effects on the spectrum are obtained from MC

simulations, numerical convolution, and analytical computation. Therein, all responses are constructed so systematic parameters are freely adjustable, which is mandatory for uncertainty propagation. For several systematic effects, methods from previous studies could be adapted. Yet, many procedures, such as the electromagnetic field transition responses, have been newly developed. At the current state, all systematic effects assumed to be dominant are included in TRModel. This allows for investigations concerning the interplay of experimental effects. Yet, several components require further improvements or validation with calibration or MC data. Each section of chapter 5 provides an account of the necessary next steps and discusses potential mitigation techniques.

Another essential aspect of TRModel is its treatment of higher-order electron reflections along their passage through the KATRIN beamline. This is realized with an intricate iterative convolution algorithm. It requires the description of the tritium  $\beta$ -spectrum as a 2D distribution over energy and pitch angle, adding complexity. Using this model, it was demonstrated that higher-order effects are certainly not neglectable. This is illustrated, for example, in section 5.2.1 where it is established that rear wall backscattering events from higher order reflections can account for more than half of all registered events in a deep spectrum measurement when an unmodified beamline setting is used. It is also shown that the effect can be mitigated to the percent level by optimizing the magnetic field settings and replacing the rear wall material. Similar optimizations are demonstrated for other effects, such as charge-sharing and rear wall activity.

In chapter 6, TRModel is employed to assess the sensitivity impact of all implemented systematics using the covariance matrix method, the common method of choice in previous studies. The sensitivity impact has been quantified for 20 systematic parameters in two scenarios: a base scenario using the default KATRIN settings and an optimized scenario featuring field adjustments and minimal hardware modifications. The results of chapter 6 successfully demonstrate the model's capability to perform studies of this scale efficiently.

It should be mentioned that the provided sensitivity curves cannot be considered final estimates for Phase-1 since several systematic inputs are currently chosen based on assumptions due to insufficient availability of calibration data. Naturally, the inputs are picked to be on the conservative side. Due to their size, the uncertainty values may not be fully compatible with the covariance matrix method of uncertainty propagation, which is only appropriate for sufficiently small values. This is indicated in the results by the sensitivity discrepancy that emerges when comparing two ways of obtaining the covariance. More robust estimates could thus be obtained with other uncertainty propagation methods, such as pull term addition or MC propagation.

Still, the results allow for an approximate sensitivity estimate and the identification of dominant effects. Furthermore, the direct comparison of both scenarios demonstrates that the suspected optimizations, especially concerning rear wall backscattering and activity, indeed significantly recover sensitivity, pushing it towards  $\sin^2 \theta \sim 10^{-6}$  in the lower half of the accessible mass region. With the optimization, magnetic trapping in the source emerges as the dominant systematic effect, which limits the sensitivity in the low-mass region. This underlines the need for further studies on magnetic trapping and source-related effects in general.

---

## Appendix

---

### A.1 Isotropic pitch angle distribution

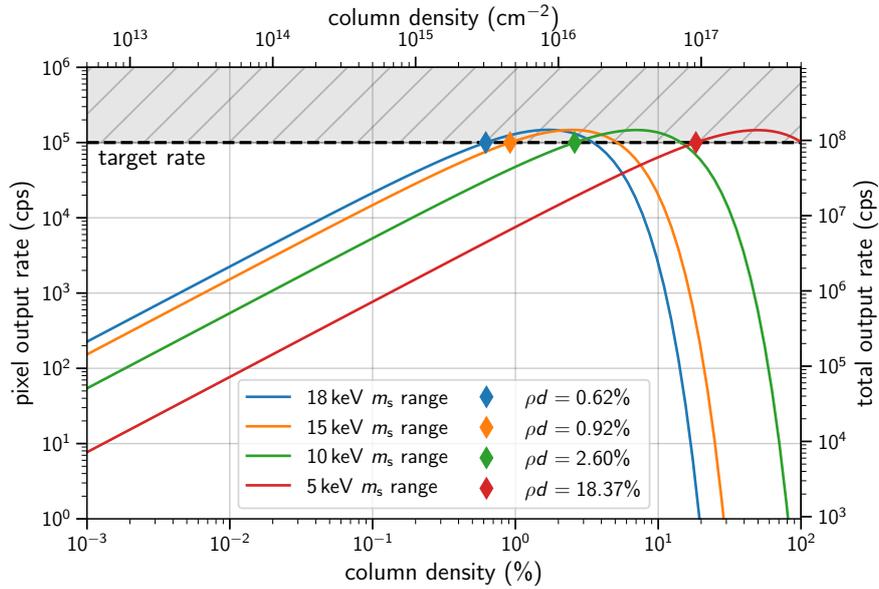
The isotropic pitch angle distribution is uniform in the interval  $[-1, 1]$  when represented as a function of  $\cos \theta$ , and will take the shape of a sine curve in the interval  $[0, \pi]$  when displayed as a distribution of  $\theta$ . One might assume that for isotropic emission, i. e. any combination of azimuthal angle  $\phi$  and polar angle  $\theta$  being equally likely, the pitch angle distribution would be uniform in  $\theta$ . However, this is not the case since for any given  $\theta$  value, the probability density is proportional to the corresponding ring element on the solid angle sphere where  $\phi$  is anywhere between 0 and  $2\pi$ . The ring element surface is maximal at  $\theta = 90^\circ$  and equal to zero at  $\theta = 0^\circ$ . This leads to a sine-shaped pitch angle distribution when represented as a distribution of  $\theta$ .

This also implies that in the WGTS very few electrons are being emitted in approximate alignment with the magnetic field, while most electrons are emitted perpendicular to the field lines. Transforming to a distribution in  $\cos \theta$  essentially narrows the intervals of the distributions near  $\theta = 90^\circ$ , and widens those near  $0^\circ$  in such a way that the sine-shaped distribution is flattened to a uniform distribution. The choice of  $\cos \theta$  as a basis has several advantages but also some drawbacks:

- Advantages:
  - Tritium decay itself is isotropic, so the initial angular distribution is precisely described by a uniform binned distribution, whereas a binned distribution following a sine shape is only an approximation.
  - No kernel is needed for any pitch angle integration during response construction. For example, this applies to the integration over the intervals of angular bins in the source response or magnetic reflection response calculation, where isotropy of the underlying distribution within bins is assumed.
  - For angular distributions similar to an isotropic distribution, which is approximately the case for the output spectrum of the tritium source, the  $\cos \theta$  binning can be considered to be the optimum since the bin width roughly matches the expected bin content. In this case, most information about the shape of the underlying distribution can be retained.

- Drawbacks:
  - At some locations along the beamline, such as at the rear wall and especially at the detector surface, the pitch angle is collimated due to a decrease in magnetic field strength or from post-acceleration. This means the angular distribution will cluster near both boundaries at  $\cos \theta = 1$  or  $\cos \theta = -1$ , corresponding to alignment with and against the z-axis. Then, only very few bins will exhibit most of the distribution content, limiting the information about the shape of the underlying distribution.

## A.2 Rate optimization for sensitivity studies



**Figure A.1:** Determination of the appropriate column density for the target output rate per pixel of  $10^5$  cps. Above this threshold, dead time becomes a significant issue and decreases the measured rate at high column density. The rate and the optimal column density setpoint depend on the retarding potential, which determines the accessible sterile mass range. The optimal values are given in percent of the nominal column density  $5 \times 10^{17} \text{ cm}^{-2}$ . Shown is the result for the optimized scenario parameter set. See figure 6.3 in section 6.2.2 for the baseline setting values.

### A.3 Sensitivity impact of individual parameters

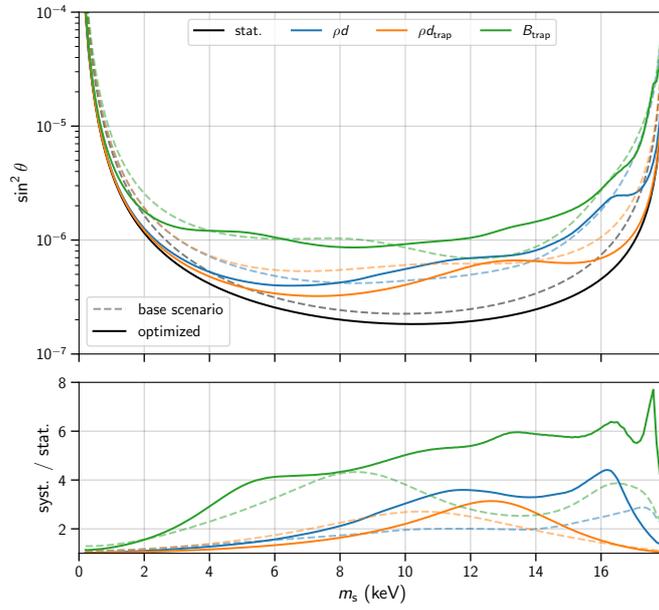


Figure A.2: Impact of source systematics.

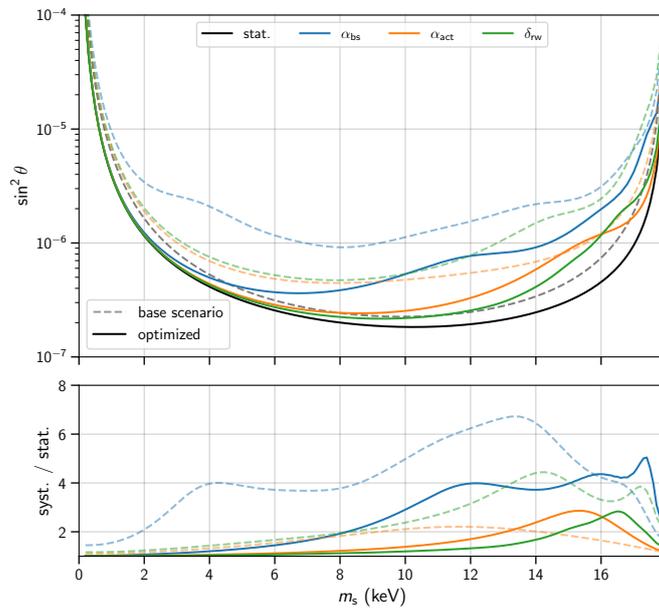


Figure A.3: Impact of rear wall systematics.

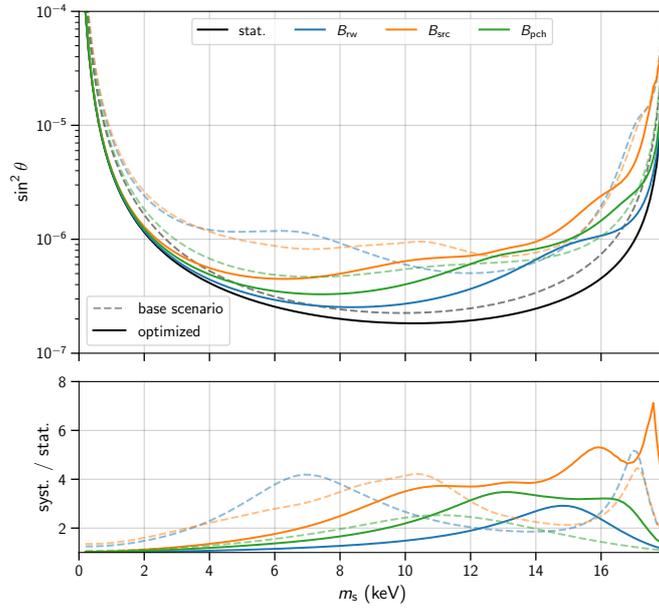


Figure A.4: Impact of transport systematics.

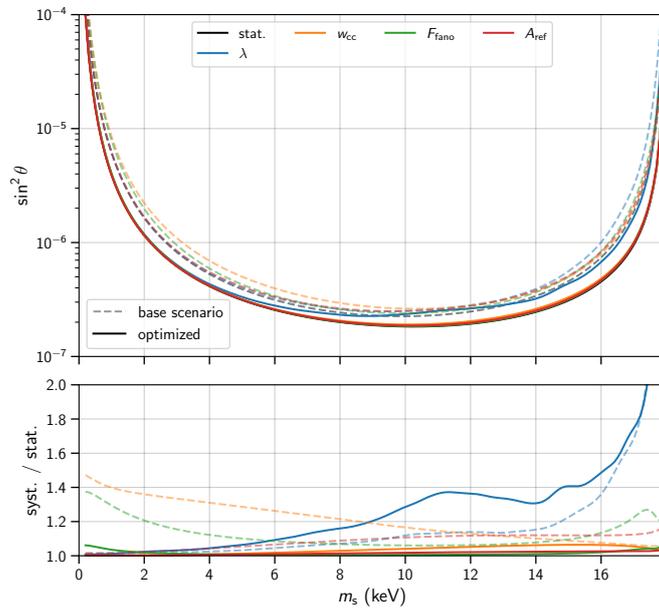


Figure A.5: Impact of detector systematics.

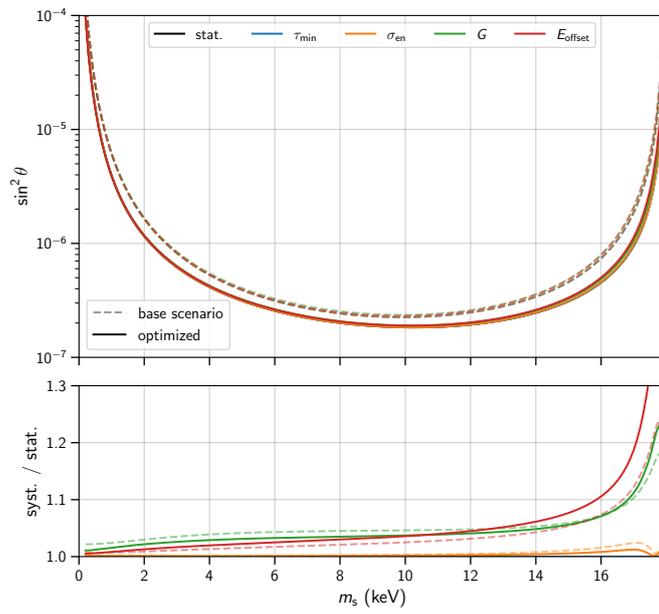


Figure A.6: Impact of readout systematics.



---

## Bibliography

---

- [1] W. Pauli. *Offener Brief an die Gruppe der Radioaktiven bei der Gauvereins-Tagung zu Tübingen*. Letter. 1930.  
URL: <https://cds.cern.ch/record/83282>
- [2] Super-Kamiokande Collaboration. „Evidence for Oscillation of Atmospheric Neutrinos“. *Phys. Rev. Lett.* 81.8 (1998), pp. 1562–1567.  
DOI: 10.1103/PhysRevLett.81.1562
- [3] SNO Collaboration. „Measurement of the Rate of  $\nu_e + d \rightarrow p + p + e^-$  Interactions Produced by  $^8B$  Solar Neutrinos at the Sudbury Neutrino Observatory“. *Phys. Rev. Lett.* 87.7 (2001), p. 071301.  
DOI: 10.1103/PhysRevLett.87.071301
- [4] KATRIN Collaboration. „KATRIN design report 2004“ (2005).  
DOI: 10.5445/IR/270060419
- [5] KATRIN Collaboration. „The design, construction, and commissioning of the KATRIN experiment“. *JINST* 16.08 (2021), T08015.  
DOI: 10.1088/1748-0221/16/08/T08015
- [6] E. Fermi. „Versuch einer Theorie der  $\beta$ -Strahlen. I“. *Z. Phys.* 88.3 (1934), pp. 161–177.  
DOI: 10.1007/BF01351864
- [7] KATRIN Collaboration. „Direct neutrino-mass measurement with sub-electronvolt sensitivity“. *Nature Phys.* 18.2 (2022), pp. 160–166.  
DOI: 10.1038/s41567-021-01463-1
- [8] T. Group. „Conceptual Design Report: KATRIN with TRISTAN modules“. 2021.  
URL: [https://www.katrin.kit.edu/downloads/TRISTAN\\_\\_Technical\\_Design\\_Report%20\(10\).pdf](https://www.katrin.kit.edu/downloads/TRISTAN__Technical_Design_Report%20(10).pdf)
- [9] D. H. Perkins. „Particle astrophysics“. 2nd ed. Oxford Master Series in Physics. Oxford University Press USA, 2009.
- [10] K. Zuber. „Neutrino Physics“. 3rd ed. CRC Press, 2020.
- [11] J. Chadwick. „Intensitätsverteilung im magnetischen Spectrum der  $\beta$ -Strahlen von radium B + C“. *Verh. Phys. Gesell.* 16 (1914), p. 383.

- [12] C. D. Ellis and W. A. Wooster. „The average energy of disintegration of radium E“. *Proc. Roy. Soc. Lond. A* 117.776 (1927), pp. 109–123.  
DOI: 10.1098/rspa.1927.0168
- [13] L. Meitner and W. Orthmann. „Über eine absolute Bestimmung der Energie der primären  $\beta$ -Strahlen von Radium E“. *Z. Physik* 60.3 (1930), pp. 143–155.  
DOI: 10.1007/BF01339819
- [14] J. Chadwick. „Possible Existence of a Neutron“. *Nature* 129.3252 (1932), p. 312.  
DOI: 10.1038/129312a0
- [15] D. Griffiths. „Introduction to Elementary Particles“. 2nd ed. Wiley-VCH, 2008.
- [16] F. L. Wilson. „Fermi’s Theory of Beta Decay“. *Am. J. Phys.* 36.12 (1968), pp. 1150–1160.  
DOI: 10.1119/1.1974382
- [17] G. Drexlin et al. „Current Direct Neutrino Mass Experiments“. *Adv. High Energy Phys.* 2013 (2013), pp. 1–39.  
DOI: 10.1155/2013/293986
- [18] G. W. Rodeback and J. S. Allen. „Neutrino Recoils Following the Capture of Orbital Electrons in  $A^{37}$ “. *Phys. Rev.* 86.4 (1952), pp. 446–450.  
DOI: 10.1103/PhysRev.86.446
- [19] O. Philipsen. „Quantenfeldtheorie und das Standardmodell der Teilchenphysik: Eine Einführung“. Berlin, Heidelberg: Springer, 2018.  
DOI: 10.1007/978-3-662-57820-9
- [20] F. Reines and C. L. Cowan. „Detection of the free neutrino“. *Phys. Rev.* 92 (1953), pp. 830–831.  
DOI: 10.1103/PhysRev.92.830
- [21] C. L. Cowan et al. „Detection of the free neutrino: A Confirmation“. *Science* 124 (1956), pp. 103–104.  
DOI: 10.1126/science.124.3212.103
- [22] P. A. M. Dirac. „The quantum theory of the electron“. *Proceedings of the Royal Society of London. Series A* 117.778 (1928), pp. 610–624.  
DOI: 10.1098/rspa.1928.0023
- [23] P. A. M. Dirac. „Quantised singularities in the electromagnetic field“. *Proceedings of the Royal Society of London. Series A* 133.821 (1931), pp. 60–72.  
DOI: 10.1098/rspa.1931.0130
- [24] C. D. Anderson. „The Positive Electron“. *Phys. Rev.* 43.6 (1933), pp. 491–494.  
DOI: 10.1103/PhysRev.43.491
- [25] E. Majorana. „Teoria simmetrica dell’elettrone e del positrone“. *Nuovo Cim* 14.4 (1937), pp. 171–184.  
DOI: 10.1007/BF02961314

- 
- [26] E. Majorana and L. Maiani. „A symmetric theory of electrons and positrons“. *Ettore Majorana Scientific Papers: On occasion of the centenary of his birth*. Ed. by G. F. Bassani. Berlin, Heidelberg: Springer, 2006, pp. 201–233.  
DOI: 10.1007/978-3-540-48095-2\_10
- [27] M. Gell-Mann and M. L. Goldberger. „Scattering of Low-Energy Photons by Particles of Spin  $\frac{1}{2}$ “. *Phys. Rev.* 96.5 (1954), pp. 1433–1438.  
DOI: 10.1103/PhysRev.96.1433
- [28] R. Davis. „Attempt to Detect the Antineutrinos from a Nuclear Reactor by the  $\text{Cl}^{37}(\bar{\nu}, e^-)\text{A}^{37}$  Reaction“. *Phys. Rev.* 97.3 (1955), pp. 766–769.  
DOI: 10.1103/PhysRev.97.766
- [29] E. J. Konopinski and H. M. Mahmoud. „The Universal Fermi Interaction“. *Phys. Rev.* 92.4 (1953), pp. 1045–1049.  
DOI: 10.1103/PhysRev.92.1045 (visited on Apr. 24, 2023).
- [30] T. D. Lee and C. N. Yang. „Question of Parity Conservation in Weak Interactions“. *Phys. Rev.* 104.1 (1956), pp. 254–258.  
DOI: 10.1103/PhysRev.104.254
- [31] C. S. Wu et al. „Experimental Test of Parity Conservation in Beta Decay“. *Phys. Rev.* 105.4 (1957), pp. 1413–1415.  
DOI: 10.1103/PhysRev.105.1413
- [32] R. L. Garwin, L. M. Lederman, and M. Weinrich. „Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: the Magnetic Moment of the Free Muon“. *Phys. Rev.* 105.4 (1957), pp. 1415–1417.  
DOI: 10.1103/PhysRev.105.1415
- [33] M. Goldhaber, L. Grodzins, and A. W. Sunyar. „Helicity of Neutrinos“. *Phys. Rev.* 109.3 (1958), pp. 1015–1017.  
DOI: 10.1103/PhysRev.109.1015
- [34] E. C. G. Sudarshan and R. E. Marshak. „Chirality Invariance and the Universal Fermi Interaction“. *Phys. Rev.* 109.5 (1958), pp. 1860–1862.  
DOI: 10.1103/PhysRev.109.1860.2
- [35] R. P. Feynman and M. Gell-Mann. „Theory of the Fermi Interaction“. *Phys. Rev.* 109.1 (1958), pp. 193–198.  
DOI: 10.1103/PhysRev.109.193
- [36] S. L. Glashow. „The renormalizability of vector meson interactions“. *Nuclear Physics* 10 (1959), pp. 107–117.  
DOI: 10.1016/0029-5582(59)90196-8
- [37] A. Salam and J. C. Ward. „Weak and electromagnetic interactions“. *Nuovo Cim* 11.4 (1959), pp. 568–577.  
DOI: 10.1007/BF02726525
- [38] S. Weinberg. „A Model of Leptons“. *Phys. Rev. Lett.* 19.21 (1967), pp. 1264–1266.  
DOI: 10.1103/PhysRevLett.19.1264

- [39] F. J. Hasert et al. „Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment“. *Phys. Lett. B* 46.1 (1973), pp. 138–140.  
DOI: 10.1016/0370-2693(73)90499-1
- [40] B. Pontecorvo. „Electron and Muon Neutrinos“. *Sov. Phys. JETP* 10 (1960), pp. 1236–1240.
- [41] M. Schwartz. „Feasibility of Using High-Energy Neutrinos to Study the Weak Interactions“. *Phys. Rev. Lett.* 4.6 (1960), pp. 306–307.  
DOI: 10.1103/PhysRevLett.4.306
- [42] G. Danby et al. „Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos“. *Phys. Rev. Lett.* 9.1 (1962), pp. 36–44.  
DOI: 10.1103/PhysRevLett.9.36
- [43] M. L. Perl et al. „Evidence for Anomalous Lepton Production in  $e^+ - e^-$  Annihilation“. *Phys. Rev. Lett.* 35.22 (1975), pp. 1489–1492.  
DOI: 10.1103/PhysRevLett.35.1489
- [44] D. DeCamp et al. „Determination of the number of light neutrino species“. *Physics Letters B* 231.4 (1989), pp. 519–529.  
DOI: 10.1016/0370-2693(89)90704-1
- [45] K. Kodama et al. „Observation of tau neutrino interactions“. *Phys. Lett. B* 504.3 (2001), pp. 218–224.  
DOI: 10.1016/S0370-2693(01)00307-0
- [46] H. A. Bethe and C. L. Critchfield. „The Formation of Deuterons by Proton Combination“. *Phys. Rev.* 54.4 (1938), pp. 248–254.  
DOI: 10.1103/PhysRev.54.248
- [47] J. N. Bahcall. „Solar Neutrino Cross Sections and Nuclear Beta Decay“. *Phys. Rev.* 135 (1B 1964), B137–B146.  
DOI: 10.1103/PhysRev.135.B137
- [48] H. A. Bethe. „Energy Production in Stars“. *Phys. Rev.* 55.5 (1939), pp. 434–456.  
DOI: 10.1103/PhysRev.55.434
- [49] R. Davis, D. S. Harmer, and K. C. Hoffman. „Search for Neutrinos from the Sun“. *Phys. Rev. Lett.* 20.21 (1968), pp. 1205–1209.  
DOI: 10.1103/PhysRevLett.20.1205
- [50] N. N. D. Center. „NuDat 3“. URL: <https://www.nndc.bnl.gov/nudat3/> (visited on May 4, 2023).
- [51] J. N. Bahcall, N. A. Bahcall, and G. Shaviv. „Present Status of the Theoretical Predictions for the  $^{37}\text{Cl}$  Solar-Neutrino Experiment“. *Phys. Rev. Lett.* 20.21 (1968), pp. 1209–1212.  
DOI: 10.1103/PhysRevLett.20.1209

- [52] K. S. Hirata et al. „Observation of  $^8\text{B}$  solar neutrinos in the Kamiokande-II detector“. *Phys. Rev. Lett.* 63.1 (1989), pp. 16–19.  
DOI: 10.1103/PhysRevLett.63.16
- [53] W. Hampel et al. „GALLEX solar neutrino observations: results for GALLEX IV“. *Phys. Lett. B* 447.1 (1999), pp. 127–133.  
DOI: 10.1016/S0370-2693(98)01579-2
- [54] J. N. Abdurashitov et al. „Measurement of the solar neutrino capture rate with gallium metal. III. Results for the 2002–2007 data-taking period“. *Phys. Rev. C* 80.1 (2009), p. 015807.  
DOI: 10.1103/PhysRevC.80.015807
- [55] R. Davis Jr., M. Koshiba, and R. Giacconi. „The Nobel Prize in Physics 2002“. NobelPrize.org, 2002.  
URL: <https://www.nobelprize.org/prizes/physics/2002/summary/>
- [56] V. Gribov and B. Pontecorvo. „Neutrino astronomy and lepton charge“. *Phys. Lett. B* 28.7 (1969), pp. 493–496.  
DOI: 10.1016/0370-2693(69)90525-5
- [57] B. Pontecorvo. „Mesonium and anti-mesonium“. *Sov. Phys. JETP* 6 (1957), p. 429.  
URL: <http://jetp.ras.ru/cgi-bin/e/index/e/6/2/p429?a=list>  
Russian original: *Zh. Eksp. Teor. Fiz.* 33.2 (1958), p. 549.
- [58] SNO Collaboration. „The Sudbury Neutrino Observatory“. *Nucl. Instrum. Meth. A* 449.1 (2000), pp. 172–207.  
DOI: 10.1016/S0168-9002(99)01469-2
- [59] L. Wolfenstein. „Neutrino oscillations in matter“. *Phys. Rev. D* 17.9 (1978), pp. 2369–2374.  
DOI: 10.1103/PhysRevD.17.2369
- [60] S. P. Mikheyev and A. Yu. Smirnov. „Resonant amplification of  $\nu$  oscillations in matter and solar-neutrino spectroscopy“. *Nuovo Cim. C* 9.1 (1986), pp. 17–26.  
DOI: 10.1007/BF02508049
- [61] T. Kajita and A. B. McDonald. „The Nobel Prize in Physics 2015“. NobelPrize.org, 2015.  
URL: <https://www.nobelprize.org/prizes/physics/2015/summary/>
- [62] Particle Data Group et al. „Review of Particle Physics“. *PTEP* 2022.8 (2022), p. 083C01.  
DOI: 10.1093/ptep/ptac097
- [63] E. Kh. Akhmedov and A. Yu. Smirnov. „Paradoxes of neutrino oscillations“. *Phys. Atom. Nuclei* 72.8 (2009), pp. 1363–1381.  
DOI: 10.1134/S1063778809080122
- [64] Z. Maki, M. Nakagawa, and S. Sakata. „Remarks on the Unified Model of Elementary Particles“. *Prog. Theor. Phys.* 28.5 (1962), pp. 870–880.  
DOI: 10.1143/PTP.28.870

- [65] B. Pontecorvo. „Neutrino Experiments and the Problem of Conservation of Leptonic Charge“. *Sov. Phys. JETP* 26.5 (1968), p. 984.  
URL: <http://jetp.ras.ru/cgi-bin/e/index/e/26/5/p984?a=list>  
Russian original: *Zh. Eksp. Teor. Fiz.* 53.5 (1968), p. 1717.
- [66] R. Adhikari et al. „A White Paper on keV sterile neutrino Dark Matter“. *JCAP* 2017.01 (2017), pp. 025–025.  
DOI: 10.1088/1475-7516/2017/01/025
- [67] KamLAND Collaboration. „Reactor on-off antineutrino measurement with KamLAND“. *Phys. Rev. D* 88.3 (2013), p. 033001.  
DOI: 10.1103/PhysRevD.88.033001
- [68] Super-Kamiokande Collaboration. „Solar neutrino measurements in Super-Kamiokande-IV“. *Phys. Rev. D* 94.5 (2016), p. 052010.  
DOI: 10.1103/PhysRevD.94.052010
- [69] IceCube Collaboration. „Measurement of Atmospheric Neutrino Oscillations at 6–56 GeV with IceCube DeepCore“. *Phys. Rev. Lett.* 120.7 (2018), p. 071801.  
DOI: 10.1103/PhysRevLett.120.071801
- [70] Super-Kamiokande Collaboration. „Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV“. *Phys. Rev. D* 97.7 (2018), p. 072001.  
DOI: 10.1103/PhysRevD.97.072001
- [71] T2K Collaboration. „Search for  $CP$  Violation in Neutrino and Antineutrino Oscillations by the T2K Experiment with  $2.2 \times 10^{21}$  Protons on Target“. *Phys. Rev. Lett.* 121.17 (2018), p. 171802.  
DOI: 10.1103/PhysRevLett.121.171802
- [72] NOvA Collaboration. „New constraints on oscillation parameters from  $\nu_e$  appearance and  $\nu_\mu$  disappearance in the NOvA experiment“. *Phys. Rev. D* 98.3 (2018), p. 032012.  
DOI: 10.1103/PhysRevD.98.032012
- [73] MINOS Collaboration. „Combined Analysis of  $\nu_\mu$  Disappearance and  $\nu_\mu \rightarrow \nu_e$  Appearance in MINOS Using Accelerator and Atmospheric Neutrinos“. *Phys. Rev. Lett.* 112.19 (2014), p. 191801.  
DOI: 10.1103/PhysRevLett.112.191801
- [74] Daya Bay Collaboration. „Measurement of the Electron Antineutrino Oscillation with 1958 Days of Operation at Daya Bay“. *Phys. Rev. Lett.* 121.24 (2018), p. 241805.  
DOI: 10.1103/PhysRevLett.121.241805
- [75] RENO Collaboration. „Measurement of Reactor Antineutrino Oscillation Amplitude and Frequency at RENO“. *Phys. Rev. Lett.* 121.20 (2018), p. 201801.  
DOI: 10.1103/PhysRevLett.121.201801

- 
- [76] Double Chooz Collaboration. „Measurement of  $\theta_{13}$  in Double Chooz using neutron captures on hydrogen with novel background rejection techniques“. *JHEP* 2016.1 (2016), p. 163.  
DOI: 10.1007/JHEP01(2016)163
- [77] I. Esteban et al. „The fate of hints: updated global analysis of three-flavor neutrino oscillations“. *JHEP* 2020.9 (2020), p. 178.  
DOI: 10.1007/JHEP09(2020)178
- [78] I. Esteban et al. „NuFIT 5.2 (2022)“. 2022.  
URL: <http://www.nu-fit.org>
- [79] T2K Collaboration. „Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations“. *Nature* 580.7803 (2020), pp. 339–344.  
DOI: 10.1038/s41586-020-2177-0
- [80] MissMJ and Cush. „Standard model of elementary particles“. 2019.  
URL: [https://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg) (visited on May 8, 2023). Public domain, via Wikimedia Commons.
- [81] M. E. Peskin and D. V. Schroeder. „An Introduction To Quantum Field Theory“. 1st ed. Frontiers in Physics. Westview Press, 1995.
- [82] W. Pauli. „Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren“. *Z. Physik* 31.1 (1925), pp. 765–783.  
DOI: 10.1007/BF02980631
- [83] S. Weinberg. „The Search for Unity: Notes for a History of Quantum Field Theory“. *Daedalus* 106.4 (1977), pp. 17–35. JSTOR: 20024506.  
URL: <https://www.jstor.org/stable/20024506>
- [84] C. N. Yang and R. L. Mills. „Conservation of Isotopic Spin and Isotopic Gauge Invariance“. *Phys. Rev.* 96.1 (1954), pp. 191–195.  
DOI: 10.1103/PhysRev.96.191
- [85] P. W. Higgs. „Broken symmetries, massless particles and gauge fields“. *Phys. Lett.* 12.2 (1964), pp. 132–133.  
DOI: 10.1016/0031-9163(64)91136-9
- [86] P. W. Higgs. „Broken Symmetries and the Masses of Gauge Bosons“. *Phys. Rev. Lett.* 13.16 (1964), pp. 508–509.  
DOI: 10.1103/PhysRevLett.13.508
- [87] F. Englert and R. Brout. „Broken Symmetry and the Mass of Gauge Vector Mesons“. *Phys. Rev. Lett.* 13.9 (1964), pp. 321–323.  
DOI: 10.1103/PhysRevLett.13.321
- [88] T. W. B. Kibble. „Symmetry Breaking in Non-Abelian Gauge Theories“. *Phys. Rev.* 155.5 (1967), pp. 1554–1561.  
DOI: 10.1103/PhysRev.155.1554

- [89] E. Noether. „Invariante Variationsprobleme“. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), p. 235.  
URL: <https://eudml.org/doc/59024>
- [90] M. D’Onofrio and K. Rummukainen. „Standard model cross-over on the lattice“. *Phys. Rev. D* 93.2 (2016), p. 025003.  
DOI: 10.1103/PhysRevD.93.025003
- [91] J. Goldstone. „Field theories with « Superconductor » solutions“. *Nuovo Cim* 19.1 (1961), pp. 154–164.  
DOI: 10.1007/BF02812722
- [92] H. Yukawa. „On the Interaction of Elementary Particles. I“. *Prog. Theor. Phys. Suppl.* 1 (1955), pp. 1–10.  
DOI: 10.1143/PTPS.1.1
- [93] S. Ordek. „Measurements of Higgs couplings to fermions and bosons at the LHC“. *PoS PANIC2021* (2022), p. 418.  
DOI: 10.22323/1.380.0418
- [94] M. A. Acero et al. „White Paper on Light Sterile Neutrino Searches and Related Phenomenology“. 2022.  
DOI: 10.48550/arXiv.2203.07323 arXiv: 2203.07323. preprint.
- [95] A. Boyarsky et al. „Sterile neutrino Dark Matter“. *Prog. Part. Nucl. Phys.* 104 (2019), pp. 1–45.  
DOI: 10.1016/j.pnpnp.2018.07.004
- [96] P. Minkowski. „ $\mu \rightarrow e\gamma$  at a rate of one out of 10<sup>9</sup> muon decays?“ *Phys. Lett. B* 67.4 (1977), pp. 421–428.  
DOI: 10.1016/0370-2693(77)90435-X
- [97] M. Gell-Mann, P. Ramond, and R. Slansky. „Complex Spinors and Unified Theories“. *Conf. Proc. C* 790927 (1979), pp. 315–321. arXiv: 1306.4669.
- [98] R. N. Mohapatra and G. Senjanović. „Neutrino Mass and Spontaneous Parity Nonconservation“. *Phys. Rev. Lett.* 44.14 (1980), pp. 912–915.  
DOI: 10.1103/PhysRevLett.44.912
- [99] T. Yanagida. „Horizontal gauge symmetry and masses of neutrinos“. *Conf. Proc. C* 7902131 (1979). Ed. by O. Sawada and A. Sugamoto, pp. 95–99.
- [100] J. Schechter and J. W. F. Valle. „Neutrino masses in  $SU(2) \otimes U(1)$  theories“. *Phys. Rev. D* 22.9 (1980), pp. 2227–2235.  
DOI: 10.1103/PhysRevD.22.2227
- [101] M. Magg and Ch. Wetterich. „Neutrino mass problem and gauge hierarchy“. *Phys. Lett. B* 94.1 (1980), pp. 61–64.  
DOI: 10.1016/0370-2693(80)90825-4

- 
- [102] R. N. Mohapatra and G. Senjanović. „Neutrino masses and mixings in gauge models with spontaneous parity violation“. *Phys. Rev. D* 23.1 (1981), pp. 165–180.  
DOI: 10.1103/PhysRevD.23.165
- [103] R. Foot et al. „See-saw neutrino masses induced by a triplet of leptons“. *Z. Phys. C* 44.3 (1989), pp. 441–444.  
DOI: 10.1007/BF01415558
- [104] E. Ma and D. P. Roy. „Heavy triplet leptons and new gauge boson“. *Nucl. Phys. B* 644.1 (2002), pp. 290–302.  
DOI: 10.1016/S0550-3213(02)00815-5
- [105] E. Ma. „Pathways to Naturally Small Neutrino Masses“. *Phys. Rev. Lett.* 81.6 (1998), pp. 1171–1174.  
DOI: 10.1103/PhysRevLett.81.1171
- [106] A. Zee. „A theory of lepton number violation and neutrino Majorana masses“. *Phys. Lett. B* 93.4 (1980), pp. 389–393.  
DOI: 10.1016/0370-2693(80)90349-4
- [107] K. S. Babu. „Model of “calculable” Majorana neutrino masses“. *Phys. Lett. B* 203.1 (1988), pp. 132–136.  
DOI: 10.1016/0370-2693(88)91584-5
- [108] E. Ma. „Verifiable radiative seesaw mechanism of neutrino mass and dark matter“. *Phys. Rev. D* 73.7 (2006), p. 077301.  
DOI: 10.1103/PhysRevD.73.077301
- [109] R. N. Mohapatra and J. W. F. Valle. „Neutrino mass and baryon-number non-conservation in superstring models“. *Phys. Rev. D* 34.5 (1986), pp. 1642–1645.  
DOI: 10.1103/PhysRevD.34.1642
- [110] A. Merle. „keV Neutrino Model Building“. *Int. J. Mod. Phys. D* 22.10 (2013), p. 1330020.  
DOI: 10.1142/S0218271813300206
- [111] A. de Gouvêa. „Neutrino Mass Models“. *Ann. Rev. Nucl. Part. Sci.* 66.1 (2016), pp. 197–217.  
DOI: 10.1146/annurev-nucl-102115-044600
- [112] LSND Collaboration. „Evidence for neutrino oscillations from the observation of  $\bar{\nu}_e$  appearance in a  $\bar{\nu}_\mu$  beam“. *Phys. Rev. D* 64.11 (2001), p. 112007.  
DOI: 10.1103/PhysRevD.64.112007
- [113] MiniBooNE Collaboration. „Updated MiniBooNE neutrino oscillation results with increased data and new background studies“. *Phys. Rev. D* 103.5 (2021), p. 052002.  
DOI: 10.1103/PhysRevD.103.052002
- [114] G. Mention et al. „Reactor antineutrino anomaly“. *Phys. Rev. D* 83.7 (2011), p. 073006.  
DOI: 10.1103/PhysRevD.83.073006

- [115] C. Giunti et al. „Gallium Anomaly: critical view from the global picture of  $\nu_e$  and  $\bar{\nu}_e$  disappearance“. *JHEP* 2022.10 (2022), p. 164.  
DOI: 10.1007/JHEP10(2022)164
- [116] K. N. Abazajian et al. „Light Sterile Neutrinos: A White Paper“. 2012.  
DOI: 10.48550/arXiv.1204.5379 arXiv: 1204.5379. preprint.
- [117] S. Dodelson and L. M. Widrow. „Sterile neutrinos as dark matter“. *Phys. Rev. Lett.* 72.1 (1994), pp. 17–20.  
DOI: 10.1103/PhysRevLett.72.17
- [118] X. Shi and G. M. Fuller. „New Dark Matter Candidate: Nonthermal Sterile Neutrinos“. *Phys. Rev. Lett.* 82.14 (1999), pp. 2832–2835.  
DOI: 10.1103/PhysRevLett.82.2832
- [119] A. D. Dolgov and S. H. Hansen. „Massive sterile neutrinos as warm dark matter“. *Astropart. Phys.* 16.3 (2002), pp. 339–344.  
DOI: 10.1016/S0927-6505(01)00115-3
- [120] K. Abazajian, G. M. Fuller, and W. H. Tucker. „Direct Detection of Warm Dark Matter in the X-Ray“. *Astrophys. J.* 562.2 (2001), p. 593.  
DOI: 10.1086/323867
- [121] K. Abazajian, G. M. Fuller, and M. Patel. „Sterile neutrino hot, warm, and cold dark matter“. *Phys. Rev. D* 64.2 (2001), p. 023501.  
DOI: 10.1103/PhysRevD.64.023501
- [122] T. Asaka and M. Shaposhnikov. „The  $\nu$ MSM, dark matter and baryon asymmetry of the universe“. *Phys. Lett. B* 620.1 (2005), pp. 17–26.  
DOI: 10.1016/j.physletb.2005.06.020
- [123] T. Asaka, S. Blanchet, and M. Shaposhnikov. „The  $\nu$ MSM, dark matter and neutrino masses“. *Phys. Lett. B* 631.4 (2005), pp. 151–156.  
DOI: 10.1016/j.physletb.2005.09.070
- [124] Planck Collaboration. „Planck 2018 results - I. Overview and the cosmological legacy of Planck“. *Astron. Astrophys.* 641 (2020), A1.  
DOI: 10.1051/0004-6361/201833880
- [125] M. R. Blanton et al. „Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe“. *Astron. J.* 154 (2017), p. 28.  
DOI: 10.3847/1538-3881/aa7567
- [126] A. Slosar. „Detecting neutrino mass difference with cosmology“. *Phys. Rev. D* 73.12 (2006), p. 123501.  
DOI: 10.1103/PhysRevD.73.123501
- [127] J. Lesgourgues and S. Pastor. „Neutrino Mass from Cosmology“. *Adv. High Energy Phys.* 2012 (2012), e608515.  
DOI: 10.1155/2012/608515

- 
- [128] Planck Collaboration. „Planck 2018 results - VIII. Gravitational lensing“. *Astron. Astrophys.* 641 (2020), A8.  
DOI: 10.1051/0004-6361/201833886
- [129] Dark Energy Survey Collaboration. „The Dark Energy Survey: more than dark energy – an overview“. *Mon. Not. Roy. Astron. Soc.* 460.2 (2016), pp. 1270–1299.  
DOI: 10.1093/mnras/stw641
- [130] Planck Collaboration. „Planck 2018 results - VI. Cosmological parameters“. *Astron. Astrophys.* 641 (2020), A6.  
DOI: 10.1051/0004-6361/201833910
- [131] eBOSS Collaboration. „Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory“. *Phys. Rev. D* 103.8 (2021), p. 083533.  
DOI: 10.1103/PhysRevD.103.083533
- [132] N. Palanque-Delabrouille et al. „Hints, neutrino bounds, and WDM constraints from SDSS DR14 Lyman- $\alpha$  and Planck full-survey data“. *JCAP* 2020.04 (2020), p. 038.  
DOI: 10.1088/1475-7516/2020/04/038
- [133] E. D. Valentino, A. Melchiorri, and J. Silk. „Cosmological constraints in extended parameter space from the Planck 2018 Legacy release“. *JCAP* 2020.01 (2020), p. 013.  
DOI: 10.1088/1475-7516/2020/01/013
- [134] DES Collaboration. „Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing“. *Phys. Rev. D* 105.2 (2022), p. 023520.  
DOI: 10.1103/PhysRevD.105.023520
- [135] J. Schechter and J. W. F. Valle. „Neutrinoless double- $\beta$  decay in  $SU(2)\times U(1)$  theories“. *Phys. Rev. D* 25.11 (1982), pp. 2951–2954.  
DOI: 10.1103/PhysRevD.25.2951
- [136] GERDA Collaboration. „Final Results of GERDA on the Search for Neutrinoless Double- $\beta$  Decay“. *Phys. Rev. Lett.* 125.25 (2020), p. 252502.  
DOI: 10.1103/PhysRevLett.125.252502
- [137] KamLAND-Zen Collaboration. „Search for Majorana Neutrinos Near the Inverted Mass Hierarchy Region with KamLAND-Zen“. *Phys. Rev. Lett.* 117.8 (2016), p. 082503.  
DOI: 10.1103/PhysRevLett.117.082503
- [138] J. Engel and J. Menéndez. „Status and future of nuclear matrix elements for neutrinoless double-beta decay: a review“. *Rep. Prog. Phys.* 80.4 (2017), p. 046301.  
DOI: 10.1088/1361-6633/aa5bc5
- [139] T. J. Loredo and D. Q. Lamb. „Bayesian analysis of neutrinos observed from supernova SN 1987A“. *Phys. Rev. D* 65.6 (2002), p. 063002.  
DOI: 10.1103/PhysRevD.65.063002

- [140] E. Fermi. „The Nobel Prize in Physics 1938“. NobelPrize.org. 1938.  
URL: <https://www.nobelprize.org/prizes/physics/1938/summary/> (visited on Apr. 27, 2023).
- [141] E. W. Otten and C. Weinheimer. „Neutrino mass limit from tritium  $\beta$  decay“. *Rept. Prog. Phys.* 71.8 (2008), p. 086201.  
DOI: 10.1088/0034-4885/71/8/086201
- [142] C. Kraus et al. „Final results from phase II of the Mainz neutrino mass search in tritium  $\beta$  decay“. *Eur. Phys. J. C* 40.4 (2005), pp. 447–468.  
DOI: 10.1140/epjc/s2005-02139-7
- [143] V. N. Aseev et al. „Upper limit on the electron antineutrino mass from the Troitsk experiment“. *Phys. Rev. D* 84.11 (2011), p. 112003.  
DOI: 10.1103/PhysRevD.84.112003
- [144] Project 8 Collaboration. „The Project 8 Neutrino Mass Experiment“. 2022.  
DOI: 10.48550/arXiv.2203.07349 arXiv: 2203.07349. preprint.
- [145] S. Mertens et al. „A novel detector system for KATRIN to search for keV-scale sterile neutrinos“. *J. Phys. G: Nucl. Part. Phys.* 46.6 (2019), p. 065203.  
DOI: 10.1088/1361-6471/ab12fe
- [146] L. Gastaldo et al. „The electron capture in  $^{163}\text{Ho}$  experiment – ECHo“. *Eur. Phys. J. Spec. Top.* 226.8 (2017), pp. 1623–1694.  
DOI: 10.1140/epjst/e2017-70071-y
- [147] M. Faverrani et al. „The HOLMES Experiment“. *J. Low Temp. Phys.* 184.3 (2016), pp. 922–929.  
DOI: 10.1007/s10909-016-1540-x
- [148] M. P. Croce et al. „Development of Holmium-163 Electron-Capture Spectroscopy with Transition-Edge Sensors“. *J. Low Temp. Phys.* 184.3 (2016), pp. 958–968.  
DOI: 10.1007/s10909-015-1451-2
- [149] KARMEN Collaboration. „Upper limits for neutrino oscillations  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  from muon decay at rest“. *Phys. Rev. D* 65.11 (2002), p. 112001.  
DOI: 10.1103/PhysRevD.65.112001
- [150] MiniBooNE Collaboration. „Neutrino flux prediction at MiniBooNE“. *Phys. Rev. D* 79.7 (2009), p. 072002.  
DOI: 10.1103/PhysRevD.79.072002
- [151] MiniBooNE Collaboration. „The MiniBooNE detector“. *Nucl. Instrum. Meth. A* 599.1 (2009), pp. 28–46.  
DOI: 10.1016/j.nima.2008.10.028
- [152] R. Acciarri et al. „Design and construction of the MicroBooNE detector“. *JINST* 12.02 (2017), P02017.  
DOI: 10.1088/1748-0221/12/02/P02017

- 
- [153] C. A. Argüelles et al. „MicroBooNE and the  $\nu_e$  Interpretation of the MiniBooNE Low-Energy Excess“. *Phys. Rev. Lett.* 128.24 (2022), p. 241802.  
DOI: 10.1103/PhysRevLett.128.241802
- [154] MicroBooNE Collaboration. „First Constraints on Light Sterile Neutrino Oscillations from Combined Appearance and Disappearance Searches with the MicroBooNE Detector“. *Phys. Rev. Lett.* 130.1 (2023), p. 011801.  
DOI: 10.1103/PhysRevLett.130.011801
- [155] Th. A. Mueller et al. „Improved predictions of reactor antineutrino spectra“. *Phys. Rev. C* 83.5 (2011), p. 054615.  
DOI: 10.1103/PhysRevC.83.054615
- [156] P. Huber. „Determination of antineutrino spectra from nuclear reactors“. *Phys. Rev. C* 84.2 (2011), p. 024617.  
DOI: 10.1103/PhysRevC.84.024617
- [157] C. Giunti et al. „Reactor antineutrino anomaly in light of recent flux model refinements“. *Phys. Lett. B* 829 (2022), p. 137054.  
DOI: 10.1016/j.physletb.2022.137054
- [158] Daya Bay Collaboration. „Evolution of the Reactor Antineutrino Flux and Spectrum at Daya Bay“. *Phys. Rev. Lett.* 118.25 (2017), p. 251801.  
DOI: 10.1103/PhysRevLett.118.251801
- [159] RENO Collaboration. „Fuel-Composition Dependent Reactor Antineutrino Yield at RENO“. *Phys. Rev. Lett.* 122.23 (2019).  
DOI: 10.1103/PhysRevLett.122.232501
- [160] O. A. Akindele et al. „High Energy Physics Opportunities Using Reactor Antineutrinos“. 2022.  
DOI: 10.48550/arXiv.2203.07214 preprint.
- [161] J. M. Berryman et al. „Statistical significance of the sterile-neutrino hypothesis in the context of reactor and gallium data“. *JHEP* 2022.2 (2022), p. 55.  
DOI: 10.1007/JHEP02(2022)055
- [162] F. Kaether et al. „Reanalysis of the Gallex solar neutrino flux and source experiments“. *Physics Letters B* 685.1 (2010), pp. 47–54.  
DOI: 10.1016/j.physletb.2010.01.030 (visited on May 25, 2023).
- [163] SAGE Collaboration. „Measurement of the response of a gallium metal solar neutrino experiment to neutrinos from a  $^{51}\text{Cr}$  source“. *Phys. Rev. C* 59.4 (1999), pp. 2246–2263.  
DOI: 10.1103/PhysRevC.59.2246
- [164] J. N. Abdurashitov et al. „Measurement of the response of a Ga solar neutrino experiment to neutrinos from a  $^{37}\text{Ar}$  source“. *Phys. Rev. C* 73.4 (2006), p. 045805.  
DOI: 10.1103/PhysRevC.73.045805

- [165] V. V. Barinov et al. „Results from the Baksan Experiment on Sterile Transitions (BEST)“. *Phys. Rev. Lett.* 128.23 (2022), p. 232501.  
DOI: 10.1103/PhysRevLett.128.232501
- [166] V. V. Barinov et al. „Search for electron-neutrino transitions to sterile states in the BEST experiment“. *Phys. Rev. C* 105.6 (2022), p. 065502.  
DOI: 10.1103/PhysRevC.105.065502
- [167] V. Barinov and D. Gorbunov. „BEST impact on sterile neutrino hypothesis“. *Phys. Rev. D* 105.5 (2022), p. L051703.  
DOI: 10.1103/PhysRevD.105.L051703
- [168] Daya Bay and MINOS+ Collaboration. „Improved Constraints on Sterile Neutrino Mixing from Disappearance Searches in the MINOS, MINOS+, Daya Bay, and Bugey-3 Experiments“. *Phys. Rev. Lett.* 125.7 (2020), p. 071801.  
DOI: 10.1103/PhysRevLett.125.071801
- [169] K. Goldhagen et al. „Testing sterile neutrino mixing with present and future solar neutrino data“. *Eur. Phys. J. C* 82.2 (2022), p. 116.  
DOI: 10.1140/epjc/s10052-022-10052-2
- [170] KATRIN Collaboration. „Improved eV-scale sterile-neutrino constraints from the second KATRIN measurement campaign“. *Phys. Rev. D* 105.7 (2022), p. 072004.  
DOI: 10.1103/PhysRevD.105.072004
- [171] A. P. Serebrov et al. „First Observation of the Oscillation Effect in the Neutrino-4 Experiment on the Search for the Sterile Neutrino“. *JETP Lett.* 109.4 (2019), pp. 213–221.  
DOI: 10.1134/S0021364019040040
- [172] A. P. Serebrov et al. „Search for sterile neutrinos with the Neutrino-4 experiment and measurement results“. *Phys. Rev. D* 104.3 (2021), p. 032003.  
DOI: 10.1103/PhysRevD.104.032003
- [173] STEREO Collaboration. „STEREO neutrino spectrum of  $^{235}\text{U}$  fission rejects sterile neutrino hypothesis“. *Nature* 613.7943 (2023), pp. 257–261.  
DOI: 10.1038/s41586-022-05568-2
- [174] PROSPECT Collaboration. „Improved short-baseline neutrino oscillation search and energy spectrum measurement with the PROSPECT experiment at HFIR“. *Phys. Rev. D* 103.3 (2021), p. 032001.  
DOI: 10.1103/PhysRevD.103.032001
- [175] H. Almazán et al. „Note on arXiv:2005.05301, 'Preparation of the Neutrino-4 experiment on search for sterile neutrino and the obtained results of measurements'“. 2020.  
DOI: 10.48550/arXiv.2006.13147 arXiv: 2006.13147. preprint.

- 
- [176] M. V. Danilov and N. A. Skrobova. „Comment on “Analysis of the Results of the Neutrino-4 Experiment on the Search for the Sterile Neutrino and Comparison with Results of Other Experiments” (JETP Letters 112, 199 (2020))“. *JETP Lett.* 112.7 (2020), pp. 452–454.  
DOI: 10.1134/S0021364020190066
- [177] P. Coloma, P. Huber, and T. Schwetz. „Statistical interpretation of sterile neutrino oscillation searches at reactors“. *Eur. Phys. J. C* 81.1 (2021), p. 2.  
DOI: 10.1140/epjc/s10052-020-08774-2
- [178] KATRIN Collaboration. „Bound on  $3 + 1$  Active-Sterile Neutrino Mixing from the First Four-Week Science Run of KATRIN“. *Phys. Rev. Lett.* 126.9 (2021), p. 091803.  
DOI: 10.1103/PhysRevLett.126.091803
- [179] M. Andriamirado et al. „PROSPECT-II physics opportunities“. *J. Phys. G* 49.7 (2022), p. 070501.  
DOI: 10.1088/1361-6471/ac48a4
- [180] N. Skrobova. „Upgrade of the DANSS detector of reactor antineutrino“. *PoS ICHEP2022* (2022), p. 1171.  
DOI: 10.22323/1.414.1171
- [181] JUNO Collaboration. „TAO Conceptual Design Report: A Precision Measurement of the Reactor Antineutrino Spectrum with Sub-percent Energy Resolution“. 2020.  
DOI: 10.48550/arXiv.2005.08745 arXiv: 2005.08745. preprint.
- [182] B. Dasgupta and J. Kopp. „Sterile neutrinos“. *Phys. Rept.* Sterile neutrinos 928 (2021), pp. 1–63.  
DOI: 10.1016/j.physrep.2021.06.002
- [183] S. Tremaine and J. E. Gunn. „Dynamical Role of Light Neutral Leptons in Cosmology“. *Phys. Rev. Lett.* 42.6 (1979), pp. 407–410.  
DOI: 10.1103/PhysRevLett.42.407
- [184] A. Boyarsky, O. Ruchayskiy, and D. Iakubovskyi. „A lower bound on the mass of dark matter particles“. *JCAP* 2009.03 (2009), p. 5.  
DOI: 10.1088/1475-7516/2009/03/005
- [185] D. Gorbunov, A. Khmelnitsky, and V. Rubakov. „Constraining sterile neutrino dark matter with phase space density observations“. *JCAP* 2008.10 (2008), p. 041.  
DOI: 10.1088/1475-7516/2008/10/041
- [186] C. Di Paolo, F. Nesti, and F. L. Villante. „Phase-space mass bound for fermionic dark matter from dwarf spheroidal galaxies“. *Mon. Not. Roy. Astron. Soc.* 475.4 (2018), pp. 5385–5397.  
DOI: 10.1093/mnras/sty091
- [187] J. Sommer-Larsen and A. Dolgov. „Formation of Disk Galaxies: Warm Dark Matter and the Angular Momentum Problem“. *Astrophys. J.* 551.2 (2001), p. 608.  
DOI: 10.1086/320211

- [188] W. J. G. de Blok et al. „Mass Density Profiles of Low Surface Brightness Galaxies“. *Astrophys. J.* 552.1 (2001), p. L23.  
DOI: 10.1086/320262
- [189] A. Klypin et al. „Where Are the Missing Galactic Satellites?“ *Astrophys. J.* 522.1 (1999), p. 82.  
DOI: 10.1086/307643
- [190] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat. „Too big to fail? The puzzling darkness of massive Milky Way subhaloes“. *Mon. Not. Roy. Astron. Soc. Lett.* 415.1 (2011), pp. L40–L44.  
DOI: 10.1111/j.1745-3933.2011.01074.x
- [191] I. de Martino et al. „Dark Matters on the Scale of Galaxies“. *Universe* 6.8 (2020), p. 107.  
DOI: 10.3390/universe6080107
- [192] R. A. C. Croft et al. „Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Ly $\alpha$  Forest“. *Astrophys. J.* 495.1 (1998), p. 44.  
DOI: 10.1086/305289
- [193] C. Yèche et al. „Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100“. *JCAP* 2017.06 (2017), p. 047.  
DOI: 10.1088/1475-7516/2017/06/047
- [194] A. Schneider. „Astrophysical constraints on resonantly produced sterile neutrino dark matter“. *JCAP* 2016.04 (2016), p. 059.  
DOI: 10.1088/1475-7516/2016/04/059
- [195] G. Kulkarni et al. „Characterizing the Pressure Smoothing Scale of the Intergalactic Medium“. *Astrophys. J.* 812.1 (2015), p. 30.  
DOI: 10.1088/0004-637X/812/1/30
- [196] C. Benso et al. „Prospects for finding sterile neutrino dark matter at KATRIN“. *Phys. Rev. D* 100.11 (2019), p. 115035.  
DOI: 10.1103/PhysRevD.100.115035
- [197] J. F. Cherry and S. Horiuchi. „Closing in on resonantly produced sterile neutrino dark matter“. *Phys. Rev. D* 95.8 (2017), p. 083015.  
DOI: 10.1103/PhysRevD.95.083015
- [198] A. Boyarsky et al. „Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster“. *Phys. Rev. Lett.* 113.25 (2014), p. 251301.  
DOI: 10.1103/PhysRevLett.113.251301
- [199] E. Bulbul et al. „Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters“. *Astrophys. J.* 789.1 (2014), p. 13.  
DOI: 10.1088/0004-637X/789/1/13
- [200] K. N. Abazajian. „Sterile neutrinos in cosmology“. *Phys. Rept.* Sterile neutrinos in cosmology 711–712 (2017), pp. 1–28.  
DOI: 10.1016/j.physrep.2017.10.003

- 
- [201] D. Sicilian et al. „Probing the Milky Way’s Dark Matter Halo for the 3.5 keV Line“. *ApJ* 905.2 (2020), p. 146.  
DOI: 10.3847/1538-4357/abee9
- [202] C. Dessert, N. L. Rodd, and B. R. Safdi. „The dark matter interpretation of the 3.5-keV line is inconsistent with blank-sky observations“. *Science* 367.6485 (2020), pp. 1465–1467.  
DOI: 10.1126/science.aaw3772
- [203] J. W. Foster et al. „Deep Search for Decaying Dark Matter with XMM-Newton Blank-Sky Observations“. *Phys. Rev. Lett.* 127.5 (2021), p. 051101.  
DOI: 10.1103/PhysRevLett.127.051101
- [204] B. M. Roach et al. „Long-exposure NuSTAR constraints on decaying dark matter in the Galactic halo“. *Phys. Rev. D* 107.2 (2023), p. 023009.  
DOI: 10.1103/PhysRevD.107.023009
- [205] K. C. Y. Ng et al. „Improved limits on sterile neutrino dark matter using full-sky Fermi Gamma-ray Burst Monitor data“. *Phys. Rev. D* 92.4 (2015), p. 043503.
- [206] T. Tamura et al. „An X-ray spectroscopic search for dark matter in the Perseus cluster with Suzaku“. *Publ. Astron. Soc. Jap.* 67.2 (2015), p. 23.
- [207] K. C. Y. Ng et al. „New constraints on sterile neutrino dark matter from NuSTAR M31 observations“. *Phys. Rev. D* 99.8 (2019), p. 083005.  
DOI: 10.1103/PhysRevD.99.083005
- [208] B. M. Roach et al. „NuSTAR tests of sterile-neutrino dark matter: New Galactic bulge observations and combined impact“. *Phys. Rev. D* 101.10 (2020), p. 103011.  
DOI: 10.1103/PhysRevD.101.103011
- [209] E. Holzschuh et al. „Search for heavy neutrinos in the  $\beta$ -spectrum of  $^{63}\text{Ni}$ “. *Phys. Lett. B* 451.1 (1999), pp. 247–255.  
DOI: 10.1016/S0370-2693(99)00200-2
- [210] A. I. Belesev et al. „The search for an additional neutrino mass eigenstate in the 2–100 eV region from ‘Troitsk nu-mass’ data: a detailed analysis“. *J. Phys. G* 41.1 (2013), p. 015001.  
DOI: 10.1088/0954-3899/41/1/015001
- [211] A. I. Belesev et al. „Upper limit on additional neutrino mass eigenstate in 2 to 100 eV region from “Troitsk nu-Mass” data“. *JETP Lett.* 97.2 (2013), pp. 67–69.  
DOI: 10.1134/S0021364013020033
- [212] J. J. Simpson. „Limits on the emission of heavy neutrinos in  $^3\text{H}$  decay“. *Phys. Rev. D* 24.11 (1981), pp. 2971–2972.  
DOI: 10.1103/PhysRevD.24.2971
- [213] C. Kraus et al. „Limit on sterile neutrino contribution from the Mainz Neutrino Mass Experiment“. *Eur. Phys. J. C* 73.2 (2013), p. 2323.  
DOI: 10.1140/epjc/s10052-013-2323-z

- [214] K.-H. Hiddemann, H. Daniel, and O. Schwentker. „Limits on neutrino masses from the tritium beta spectrum“. *J. Phys. G* 21.5 (1995), pp. 639–650.  
DOI: 10.1088/0954-3899/21/5/008
- [215] M. Galeazzi et al. „Limits on the Existence of Heavy Neutrinos in the Range 50–1000 eV from the Study of the  $^{187}\text{Re}$  Beta Decay“. *Phys. Rev. Lett.* 86.10 (2001), pp. 1978–1981.  
DOI: 10.1103/PhysRevLett.86.1978
- [216] M. Aker et al. „Search for keV-scale Sterile Neutrinos with first KATRIN Data“. 2022.  
DOI: 10.48550/arXiv.2207.06337 arXiv: 2207.06337. preprint.
- [217] K. G. Leach and S. Friedrich. „The BeEST Experiment: Searching for Beyond Standard Model Neutrinos Using  $^7\text{Be}$  Decay in STJs“. *J Low. Temp. Phys.* 209.5 (2022), pp. 796–803.  
DOI: 10.1007/s10909-022-02759-z
- [218] C. J. Martoff et al. „HUNTER: precision massive-neutrino search based on a laser cooled atomic source“. *Quantum Sci. Technol.* 6.2 (2021), p. 024008.  
DOI: 10.1088/2058-9565/abdb9b
- [219] G.-B. Kim. „MAGNETO- $\nu$ : Searching for keV sterile neutrino dark matter in  $^{241}\text{Pu}$  beta decays“. Presentation at the UCLA Dark Matter 2023 conference. 2023.  
URL: <https://indico.cern.ch/event/1188759/contributions/5244403/> (visited on May 13, 2023).
- [220] X. Zhang. „Searching for a keV sterile neutrino via  $^{241}\text{Pu}$  beta spectrum“. Poster at the Neutrino 2022 Seoul Virtual Meeting. 2022.  
DOI: 10.5281/zenodo.6805550
- [221] M. Kleesiek et al. „ $\beta$ -Decay spectrum, response function and statistical model for neutrino mass measurements with the KATRIN experiment“. *Eur. Phys. J. C* 79.3 (2019), p. 204.  
DOI: 10.1140/epjc/s10052-019-6686-7
- [222] KATRIN Collaboration. „Analysis methods for the first KATRIN neutrino-mass measurement“. *Phys. Rev. D* 104.1 (2021), p. 012005.  
DOI: 10.1103/PhysRevD.104.012005
- [223] D. H. Wilkinson. „Small terms in the beta-decay spectrum of tritium“. *Nucl. Phys. A* 526.1 (1991), pp. 131–140.  
DOI: 10.1016/0375-9474(91)90301-L
- [224] P. Froelich and A. Saenz. „Calculation of the  $\beta$ -Decay Spectrum of the  $T_2$  Molecule beyond the Sudden Impulse Approximation“. *Phys. Rev. Lett.* 77.23 (1996), pp. 4724–4727.  
DOI: 10.1103/PhysRevLett.77.4724

- [225] S. Mertens et al. „Sensitivity of next-generation tritium beta-decay experiments for keV-scale sterile neutrinos“. *JCAP* 2015.02 (2015), p. 020.  
DOI: 10.1088/1475-7516/2015/02/020
- [226] A. M. Turaclar. „Influence of the fractional recoil momentum on the final state distribution“. Presentation at the 41st KATRIN Collaboration Meeting (unpublished). 2021.
- [227] A. Picard et al. „A solenoid retarding spectrometer with high resolution and transmission for keV electrons“. *Nucl. Instrum. Meth. B* 63.3 (1992), pp. 345–358.  
DOI: 10.1016/0168-583X(92)95119-C
- [228] V. M. Lobashev and P. E. Spivak. „A method for measuring the electron antineutrino rest mass“. *Nucl. Instrum. Meth. A* 240.2 (1985), pp. 305–310.  
DOI: 10.1016/0168-9002(85)90640-0
- [229] L. Kuckert et al. „Modelling of gas dynamical properties of the Katrin tritium source and implications for the neutrino mass measurement“. *Vacuum* 158 (2018), pp. 195–205.  
DOI: 10.1016/j.vacuum.2018.09.036
- [230] F. Priester, M. Sturm, and B. Bornschein. „Commissioning and detailed results of KATRIN inner loop tritium processing system at Tritium Laboratory Karlsruhe“. *Vacuum* 116 (2015), pp. 42–47.  
DOI: 10.1016/j.vacuum.2015.02.030
- [231] A. Marsteller et al. „Operation modes of the KATRIN experiment Tritium Loop System using 83mKr“. *JINST* 17.12 (2022), P12010.  
DOI: 10.1088/1748-0221/17/12/P12010
- [232] M. Arenz et al. „The KATRIN superconducting magnets: overview and first performance results“. *JINST* 13.08 (2018), T08005.  
DOI: 10.1088/1748-0221/13/08/T08005
- [233] L. Kuckert. „The Windowless Gaseous Tritium Source of the KATRIN Experiment - Characterisation of Gas Dynamical and Plasma Properties“. PhD thesis. 2016.  
DOI: 10.5445/IR/1000065077 (visited on Feb. 5, 2023).
- [234] R. Sack and L. Hasselmann. „Rear Wall cleaning investigation in a test setup“. Presentation at the 41st KATRIN Collaboration Meeting (unpublished). 2021.
- [235] F. Friedel et al. „Time-dependent simulation of the flow reduction of D2 and T2 in the KATRIN experiment“. *Vacuum* 159 (2019), pp. 161–172.  
DOI: 10.1016/j.vacuum.2018.10.002
- [236] M. Arenz et al. „Commissioning of the vacuum system of the KATRIN Main Spectrometer“. *JINST* 11.04 (2016), P04011.  
DOI: 10.1088/1748-0221/11/04/P04011
- [237] X. Tao, A. A. Chan, and A. J. Brizard. „Hamiltonian theory of adiabatic motion of relativistic charged particles“. *Phys. Plasmas* 14.9 (2007), p. 092107.  
DOI: 10.1063/1.2773702

- [238] O. Rest et al. „A novel ppm-precise absolute calibration method for precision high-voltage dividers“. *Metrologia* 56.4 (2019), p. 045007.  
DOI: 10.1088/1681-7575/ab2997
- [239] KATRIN Collaboration. „KATRIN: status and prospects for the neutrino mass and beyond“. *J. Phys. G: Nucl. Part. Phys.* 49.10 (2022), p. 100501.  
DOI: 10.1088/1361-6471/ac834e
- [240] J. F. Amsbaugh et al. „Focal-plane detector system for the KATRIN experiment“. *Nucl. Instrum. Meth. A* 778 (2015), pp. 40–60.  
DOI: 10.1016/j.nima.2014.12.116
- [241] A. V. Lokhov and F. V. Tkachov. „Confidence intervals with a priori parameter bounds“. *Phys. Part. Nucl.* 46.3 (2015), pp. 347–365.  
DOI: 10.1134/S1063779615030089
- [242] M. Aker et al. „First operation of the KATRIN experiment with tritium“. *Eur. Phys. J. C* 80.3 (2020), p. 264.  
DOI: 10.1140/epjc/s10052-020-7718-z
- [243] A. Huber. „Analysis of first KATRIN data and searches for keV-scale sterile neutrinos“. PhD thesis. Karlsruhe Institute of Technology, 2021.  
DOI: 10.5445/IR/1000128344
- [244] P. Voigt. „Non-Adiabaticity Measurement and Simulation“. Presentation at the 44th KATRIN Collaboration Meeting (unpublished). 2023.
- [245] N. Steinbrink et al. „Neutrino mass sensitivity by MAC-E-Filter based time-of-flight spectroscopy with the example of KATRIN“. *New J. Phys.* 15.11 (2013), p. 113020.  
DOI: 10.1088/1367-2630/15/11/113020
- [246] N. M. N. Steinbrink et al. „keV-Scale Sterile Neutrino Sensitivity Estimation with Time-Of-Flight Spectroscopy in KATRIN using Self-Consistent Approximate Monte Carlo“. *Eur. Phys. J. C* 78.3 (2018), p. 212.  
DOI: 10.1140/epjc/s10052-018-5656-9
- [247] E. Gatti and P. Rehak. „Semiconductor drift chamber — An application of a novel charge transport scheme“. *Nucl. Instrum. Meth.* 225.3 (1984), pp. 608–614.  
DOI: 10.1016/0167-5087(84)90113-3
- [248] M. Korzeczek. „Sterile neutrino search with KATRIN - modeling and design-criteria of a novel detector system“. PhD thesis. Karlsruhe Institute of Technology, 2020.  
DOI: 10.5445/IR/1000120634
- [249] P. Trigilio et al. „ETTORE: a 12-Channel Front-End ASIC for SDDs with Integrated JFET“. *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. 2018, pp. 1–4.  
DOI: 10.1109/NSSMIC.2018.8824675

- 
- [250] D. Siegmann. „TRISTAN 166 Pixel Detector Module Performance“. Presentation at the 44th KATRIN Collaboration Meeting (unpublished). 2023.
- [251] K. Urban. „Application of a TRISTAN Silicon Drift Detector as Forward Beam Monitor in KATRIN“. MA thesis. Technical University of Munich, 2019.  
URL: [https://www.katrin.kit.edu/publikationen/Urban\\_Korbinian\\_19.pdf](https://www.katrin.kit.edu/publikationen/Urban_Korbinian_19.pdf)
- [252] K. Dolde et al. „Impact of ADC non-linearities on the sensitivity to sterile keV neutrinos with a KATRIN-like experiment“. *Nucl. Instrum. Meth. A* 848 (2017), pp. 127–136.  
DOI: 10.1016/j.nima.2016.12.015
- [253] M. Descher. „High rate systematic effects for keV sterile neutrino searches at KATRIN“. MA thesis. Karlsruhe Institute of Technology, 2019.  
URL: <https://www.katrin.kit.edu/publikationen/mth-descher.pdf>
- [254] M. Descher et al. *TRModel*. KATRIN Software, 2023.  
URL: <https://nuserv.uni-muenster.de:8443/katrin-git/tristan/trmodel>
- [255] M. Klein and F. Glück. „Tritium ion blocking and detection in the KATRIN experiment“. *J. Phys. Conf. Ser.* 888.1 (2017), p. 012073.  
DOI: 10.1088/1742-6596/888/1/012073
- [256] S. Groh. „Modeling of the response function and measurement of transmission properties of the KATRIN experiment“. PhD thesis. Karlsruhe Institute of Technology, 2015.  
DOI: 10.5445/IR/1000046546 (visited on Feb. 3, 2023).
- [257] KATRIN Collaboration. „Precision measurement of the electron energy-loss function in tritium and deuterium gas for the KATRIN experiment“. *Eur. Phys. J. C* 81.7 (2021), p. 579.  
DOI: 10.1140/epjc/s10052-021-09325-z
- [258] KATRIN Collaboration. „Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN“. *Phys. Rev. Lett.* 123.22 (2019), p. 221802.  
DOI: 10.1103/PhysRevLett.123.221802
- [259] R. wall working group. „Rear wall residual tritium spectrum wiki page“. Rear wall residual tritium spectrum. 2023.  
URL: [https://ikp-katrin-wiki.ikp.kit.edu/katrin/index.php/RW\\_residual\\_tritium\\_spectrum](https://ikp-katrin-wiki.ikp.kit.edu/katrin/index.php/RW_residual_tritium_spectrum) (visited on Feb. 7, 2023).
- [260] A. Huber. „Search for keV-Scale Sterile Neutrinos with KATRIN“. MA thesis. Karlsruhe Institute of Technology, 2015.  
URL: [https://www.katrin.kit.edu/publikationen/mth\\_huber\\_anton.pdf](https://www.katrin.kit.edu/publikationen/mth_huber_anton.pdf)
- [261] U. Fano. „Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions“. *Phys. Rev.* 72.1 (1947), pp. 26–29.  
DOI: 10.1103/PhysRev.72.26

- [262] B. L. Wall et al. „Dead layer on silicon p–i–n diode charged-particle detectors“. *Nucl. Instrum. Meth. A* 744 (2014), pp. 73–79.  
DOI: 10.1016/j.nima.2013.12.048
- [263] M. Biassoni et al. „Electron spectrometry with Silicon drift detectors: a GEANT4 based method for detector response reconstruction“. *Eur. Phys. J. Plus* 136.1 (2021), p. 125.  
DOI: 10.1140/epjp/s13360-021-01074-y
- [264] F. T. Harms. „Characterization and Minimization of Background Processes in the KATRIN Main Spectrometer“. PhD thesis. Karlsruhe Institute of Technology, 2015.  
DOI: 10.5445/IR/1000050027
- [265] V. Blobel and E. Lohrmann. „Statistische und numerische Methoden der Datenanalyse“. Hamburg: Creative Commons, 2012.  
URL: <https://www.desy.de/~sschmitt/blobel/ebuch.html>
- [266] J. Allison et al. „Recent developments in Geant4“. *Nucl. Instrum. Meth. A* 835 (2016), pp. 186–225.  
DOI: 10.1016/j.nima.2016.06.125
- [267] P. Renschler. „KESS - A new Monte Carlo simulation code for low-energy electron interactions in silicon detectors“. PhD thesis. Karlsruhe Institute of Technology, 2011.  
DOI: 10.5445/IR/1000024959
- [268] D. Furse et al. „Kassiopeia: a modern, extensible C++ particle tracking package“. *New J. Phys.* 19.5 (2017), p. 053012.  
DOI: 10.1088/1367-2630/aa6950 (visited on Feb. 21, 2023).
- [269] S. Enomoto. *DRIPS: Detector readout simulation*. KATRIN Software, 2022.  
URL: <https://nuserv.uni-muenster.de:8443/katrin-git/kasper> (visited on Mar. 1, 2023).
- [270] T. Group. „TRISTAN Internal review series“. Presentation series on sensitivity and model building (unpublished). 2022.  
URL: <https://indico.scc.kit.edu/event/2701/>
- [271] P. S. Foundation. *The Python Language Reference*. Version 3.8.  
URL: <https://docs.python.org/3/reference/> (visited on Mar. 1, 2023).
- [272] C. R. Harris et al. „Array programming with NumPy“. *Nature* 585.7825 (2020), pp. 357–362.  
DOI: 10.1038/s41586-020-2649-2
- [273] lonelyenvoy. *Python Memoization*. Version 0.4.0. 2021.  
URL: <https://github.com/lonelyenvoy/python-memoization>

- [274] S. K. Lam, A. Pitrou, and S. Seibert. „Numba: a LLVM-based Python JIT compiler“. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM '15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 1–6.  
DOI: 10.1145/2833157.2833162
- [275] L. L. Lucas and M. P. Unterweger. „Comprehensive Review and Critical Evaluation of the Half-Life of Tritium“. *J. Res. Natl. Inst. Stand. Tech.* 105.4 (2000), pp. 541–549.  
DOI: 10.6028/jres.105.043
- [276] N. Haußmann. „Simulation and measurement of the Forward Beam Monitor detector signal for the KATRIN experiment“. Bergische Universität Wuppertal, 2020.  
DOI: 10.25926/8q1y-6752
- [277] M. Slezak, A. Lokhov, and N. Trost. *Tristan Source Sim*. KATRIN Software, 2019.  
URL: <https://nuserv.uni-muenster.de:8443/katrin-git/tristan/TristanSourceSim>
- [278] H. Kolanoski and N. Wermes. „Particle Detectors: Fundamentals and Applications“. New York: Oxford University Press, 2020.
- [279] R. C. Alig, S. Bloom, and C. W. Struck. „Scattering by ionization and phonon emission in semiconductors“. *Phys. Rev. B* 22.12 (1980), pp. 5565–5582.  
DOI: 10.1103/PhysRevB.22.5565
- [280] K. M. Altenmüller. „Search for sterile neutrinos in  $\beta$ -decays“. PhD thesis. Technical University of Munich, University Paris-Saclay, 2019.  
URL: <https://theses.hal.science/tel-02397726>
- [281] A. Lokhov and C. Weinheimer. „Transmission function for pixels and patches in SAP configuration“. Internal presentation. 2020.  
URL: [https://ikp-katrin-wiki.ikp.kit.edu/katrin/images/1/15/2020-06-30\\_sap\\_transmission.pdf](https://ikp-katrin-wiki.ikp.kit.edu/katrin/images/1/15/2020-06-30_sap_transmission.pdf)
- [282] F. Block. „Determination of Electromagnetic Fields and Tritium Column Density for Neutrino Mass Analysis with KATRIN“. PhD thesis. Karlsruhe Institute of Technology, 2022.  
DOI: 10.5445/IR/1000145073
- [283] K. Urban and M. Descher. *TRAINS: TRISTAN ADC integral nonlinearity study*. KATRIN Software, 2020.  
URL: <https://nuserv.uni-muenster.de:8443/katrin-git/tristan/trains>
- [284] K. Dolde. „Detector and read-out development to search for sterile neutrinos with KATRIN“. MA thesis. Karlsruhe Institute of Technology, 2016.  
URL: [https://www.katrin.kit.edu/publikationen/mth\\_dolde.compressed.pdf](https://www.katrin.kit.edu/publikationen/mth_dolde.compressed.pdf)
- [285] J. Nickolls et al. „Scalable Parallel Programming with CUDA“. *Queue* 6.2 (2008), pp. 40–53.  
DOI: 10.1145/1365490.1365500

- [286] N. R.-M. Trost. „Modeling and measurement of Rydberg-State mediated Background at the KATRIN Main Spectrometer“. PhD thesis. Karlsruhe Institute of Technology, 2019.  
DOI: 10.5445/IR/1000090450 (visited on Aug. 11, 2022).
- [287] C. Reiling. „A new Monte Carlo model of the detailed electron and ion distribution in the WGTS of KATRIN“. MA thesis. Karlsruhe Institute of Technology, 2019.  
URL: [https://www.katrin.kit.edu/publikationen/mth2019\\_ReilingChristian.pdf](https://www.katrin.kit.edu/publikationen/mth2019_ReilingChristian.pdf)
- [288] S. Trajmar, D. F. Register, and A. Chutjian. „Electron scattering by molecules II. Experimental methods and data“. *Phys. Rept.* 97.5 (1983), pp. 219–356.  
DOI: 10.1016/0370-1573(83)90071-6 (visited on Jan. 21, 2023).
- [289] J. W. Liu. „Total cross sections for high-energy electron scattering by  $\text{H}_2$  ( $^1\Sigma_g^+$ ),  $\text{N}_2$  ( $^1\Sigma_g^+$ ), and  $\text{O}_2$  ( $^3\Sigma_g^-$ )“. *Phys. Rev. A* 35.2 (1987), pp. 591–597.  
DOI: 10.1103/PhysRevA.35.591
- [290] S. Biagi. *Fortran program MAGBOLTZ*. Version 8.97. 2012.  
URL: <https://nl.lxcat.net/Biagi>
- [291] M. E. Rudd. „Differential and total cross sections for ionization of helium and hydrogen by electrons“. *Phys. Rev. A* 44.3 (1991), pp. 1644–1652.  
DOI: 10.1103/PhysRevA.44.1644
- [292] J. W. Liu. „Elastic scattering of fast electrons by  $\text{H}_2$  ( $^1\Sigma_g^+$ ) and  $\text{N}_2$  ( $X^1\Sigma_g^+$ )“. *Phys. Rev. A* 32.3 (1985), pp. 1384–1394.  
DOI: 10.1103/PhysRevA.32.1384
- [293] H. Nishimura, A. Danjo, and H. Sugahara. „Differential Cross Sections of Electron Scattering from Molecular Hydrogen I. Elastic Scattering and Vibrational Excitation ( $X^1\Sigma^+g, v=0\rightarrow 1$ )“. *J. Phys. Soc. Jap.* 54.5 (1985), pp. 1757–1768.  
DOI: 10.1143/JPSJ.54.1757
- [294] G. Arrighini, F. Biondi, and C. Guidotti. „A study of the inelastic scattering of fast electrons from molecular hydrogen“. *Mol. Phys.* 41.6 (1980), pp. 1501–1514.  
DOI: 10.1080/00268978000103701
- [295] P. Weck, B. Joulakian, and P. A. Hervieux. „Fivefold differential cross section of fast ( $e, 2e$ ) ionization of  $\text{H}_2$ ,  $\text{D}_2$ , and  $\text{T}_2$  by a Franck-Condon approach“. *Phys. Rev. A* 60.4 (1999), pp. 3013–3019.  
DOI: 10.1103/PhysRevA.60.3013
- [296] F. Knapp. „Impacts of Magnetic Traps in the WGTS on a keV-scale Sterile Neutrino Search with KATRIN“. BA thesis. Karlsruhe Institute of Technology, 2017.
- [297] J. Kellerer. „Simulation of the KATRIN Source Plasma using Monte Carlo and Particle in Cell Methods“. PhD thesis. Karlsruhe Institute of Technology, 2022.  
DOI: 10.5445/IR/1000143868

- [298] E. Förstner. „Optimization of the KATRIN rear wall for a keV-scale sterile neutrino search“. BA thesis. Karlsruhe Institute of Technology, 2017.
- [299] Geant4 Collaboration. „Guide for Physics Lists (revision 7.0)“. 2022.  
URL: <https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsListGuide/html/> (visited on Mar. 1, 2023).
- [300] M. Babutzka. „Design and development for the Rearsection of the KATRIN experiment“. PhD thesis. 2014.  
DOI: 10.5445/IR/1000045598
- [301] M. D. Hoover et al. „Determination of the Oxide Layer Thickness on Beryllium Metal Particles“. *Am. Ind. Hyg. Assoc. J.* 50.10 (1989), pp. 550–553.  
DOI: 10.1080/15298668991375146
- [302] M. Korzeczek. „eV- & keV-sterile neutrino studies with KATRIN“. MA thesis. Karlsruhe Institute of Technology, 2016.  
URL: <https://www.katrin.kit.edu/publikationen/MKorz-MA-eV-keV%20sterile%20Neutrinos.pdf>
- [303] E. Raspopin. „Systematic effects of the plasma and surfaces in the KATRIN tritium source“. MA thesis. Karlsruhe Institute of Technology, 2021.  
URL: [https://www.katrin.kit.edu/publikationen/mth\\_raspopin.pdf](https://www.katrin.kit.edu/publikationen/mth_raspopin.pdf)
- [304] M. Aker et al. „Rear wall report for the KNM1-5 neutrino mass analysis“. Technical report. 2023.  
URL: <https://nuserv.uni-muenster.de:8443/Stephanie/rearwall-residual-tritium-working-group> (visited on Feb. 7, 2023).
- [305] M. Klein. „Adsorption of gaseous tritium on gold coated beryllium investigated with TRIADE“. MA thesis. Karlsruhe Institute of Technology, 2015.  
URL: [https://www.katrin.kit.edu/publikationen/ManuelKlein\\_math.pdf](https://www.katrin.kit.edu/publikationen/ManuelKlein_math.pdf)
- [306] M. Röllig. „Tritium analytics by beta induced X-ray spectrometry“. PhD thesis. Karlsruhe Institute of Technology, 2015.  
DOI: 10.5445/IR/1000054050
- [307] P. Becker. „Influence of Rear Wall Activity on keV-scale Sterile Neutrino Searches with KATRIN“. BA thesis. Karlsruhe Institute of Technology, 2021.
- [308] T. App. „Experimentelle Analyse eines Reinigungsverfahrens der KATRIN Rear Wall mittels Ozon in einem Teststand“. BA thesis. Karlsruhe Institute of Technology, 2022.
- [309] A. Lokhov et al. „Background reduction at the KATRIN experiment by the shifted analysing plane configuration“. *Eur. Phys. J. C* 82.3 (2022), p. 258.  
DOI: 10.1140/epjc/s10052-022-10220-4
- [310] D. Siegmann. „Investigation of the Detector Response to Electrons of the TRISTAN Prototype Detectors“. MA thesis. Technical University of Munich, 2019.  
URL: [https://www.katrin.kit.edu/publikationen/2019\\_Master\\_Thesis\\_Siegmann\\_TRISTAN\\_Detector\\_Response\\_Final.pdf](https://www.katrin.kit.edu/publikationen/2019_Master_Thesis_Siegmann_TRISTAN_Detector_Response_Final.pdf)

- [311] A. Nava. „TRISTAN detector backscattering systematics“. Presentation at the 43rd KATRIN Collaboration Meeting (unpublished). 2022.
- [312] G. F. Knoll. „Radiation Detection and Measurement“. 4th ed. Wiley, 2010.
- [313] E. Gatti et al. „Dynamics of electrons in drift detectors“. *Nucl. Instrum. Meth. A* 253.3 (1987), pp. 393–399.  
DOI: 10.1016/0168-9002(87)90522-5
- [314] C. Forstner. „Characterization of a TRISTAN Silicon Drift Detector Array with a Laser System“. MA thesis. Technical University of Munich, 2023.
- [315] XGLab. „DANTE | The Digital Pulse Processor for X-Ray Spectroscopy“. URL: <https://www.xglab.it/products/dante/> (visited on Feb. 27, 2023).
- [316] CAEN. „Octal Digital Multi Channel Analyzer V1782“. URL: <https://www.caen.it/products/v1782/> (visited on Feb. 27, 2023).
- [317] V. T. Jordanov and G. F. Knoll. „Digital synthesis of pulse shapes in real time for high resolution radiation spectroscopy“. *Nucl. Instrum. Meth. A* 345.2 (1994), pp. 337–345.  
DOI: 10.1016/0168-9002(94)91011-1
- [318] Z. Guzik and T. Krakowski. „Algorithms for digital  $\gamma$ -ray spectroscopy“. *Nukleonika* Vol. 58, No. 2 (2013).
- [319] S. Enomoto. „Simple Analytical Models of Signal Pile-up“. Technical report. 2014.  
URL: [https://ikp-katrin-wiki.ikp.kit.edu/katrin/images/2/24/PileupAnalyticalModels\\_20140424.pdf](https://ikp-katrin-wiki.ikp.kit.edu/katrin/images/2/24/PileupAnalyticalModels_20140424.pdf) (visited on Feb. 28, 2023).
- [320] R. A. Johnson and D. W. Wichern. „Applied multivariate statistical analysis“. 6th ed. Pearson Prentice Hall, 2007.
- [321] C. Schönfeld. „Commissioning of the KATRIN Pinch Magnet“. BA thesis. Karlsruhe Institute of Technology, 2015.
- [322] C. Rodenbeck et al. „Wideband precision stabilization of the -18.6kV retarding voltage for the KATRIN spectrometer“. *JINST* 17.06 (2022), P06003.  
DOI: 10.1088/1748-0221/17/06/P06003
- [323] A. Onillon. „Sensitivity study“. Internal presentation. 2022.  
URL: [https://ikp-katrin-wiki.ikp.kit.edu/katrin/images/6/6a/2022.06.13\\_A0\\_Sensitivity\\_investigation.pdf](https://ikp-katrin-wiki.ikp.kit.edu/katrin/images/6/6a/2022.06.13_A0_Sensitivity_investigation.pdf)
- [324] P. M. Harris and M. G. Cox. „On a Monte Carlo method for measurement uncertainty evaluation and its implementation“. *Metrologia* 51.4 (2014), S176.  
DOI: 10.1088/0026-1394/51/4/S176

---

## Acknowledgements

---

Zum Schluss möchte ich mich zutiefst bei all den wunderbaren Menschen bedanken, die zum Gelingen dieser Arbeit beigetragen haben. Darüber hinaus geht mein Dank an alle, die mich in diesem Lebensabschnitt begleitet haben, sowohl bei KATRIN als auch im Privaten. Bei einigen Personen möchte ich mich insbesondere bedanken:

Zuallererst möchte ich Guido Drexlin danken, ohne dessen Vorlesungen ich vermutlich nicht den Weg in Neutrino-Physik und zur Dunklen Materie eingeschlagen hätte. Bezüglich der Doktorarbeit möchte ich danke sagen für das anhaltende Vertrauen, den Rückenwind am Ende Promotionsphase, und für die Freiheit eigene Ansätze langfristig zu verfolgen.

Im gleichen Maße möchte ich Susanne Mertens danken, die das Projekt, welches schließlich Thema meiner Arbeit wurde, vorangetrieben hat wie sonst niemand. Zum einen durch das Zusammenführen der Deep-Model Gruppe, sowie die Organisation des Reviews, welcher die Inhalte meiner Doktorarbeit maßgeblich geformt hat. Zum anderen, weil Susanne immer im täglichen Geschehen mit dabei ist und ich von ihren Herangehensweisen, Nachfragen, und Ratschlägen vieles gelernt habe.

My deepest thanks also go out to all the other Model Deep Divers. I wouldn't be where I am now without your support. Special thanks go to Korbinian Urban, Anthony Onillon, Andrea Nava, Dominic Batzler, Marco Röllig, Alexey Lokhov, Thibaut Houdy, Anton Huber, Ferenc Glück, and many more who supported the group. Thanks also to all the people who proofread part of my thesis!

Der KATRIN Gemeinschaft, sei es aus Karlsruhe, München oder auch Milano, möchte ich danken für die harmonische Zusammenarbeit, die Gemeinschaft, und die Freundschaften, die ich unter euch finden durfte. Die gemeinsam verbrachte Zeit mit euch ist für mich eine Bereicherung. An dieser Stelle möchte ich mich auch besonders bei Stephanie, Leo, und Sanshiro für die großartigen Erlebnisse am Felsen bedanken, sowie bei Daniel und Korbinian für die spaßigen Abende beim bouldern und all den anderen Aktivitäten.

Meinen Eltern Grit und Frank, sowie meiner Schwester Greta, und auch der größeren Familie möchte ich für den stetigen Rückhalt danken. Außerdem möchte ich auch meine Studien- und Schulfreunde Sebastian, Julian, Mirko, und Wieland erwähnen. Die Lebensumstände haben sich zwar stark verändert, aber für die anhaltende Freundschaft durch diesen Lebensabschnitt und darüber hinaus bin ich ewig dankbar! Und zum Schluss möchte ich mich bei meiner Partnerin Pauline bedanken, die mich in dieser anstrengenden Zeit stets mit viel Verständnis unterstützt hat.

