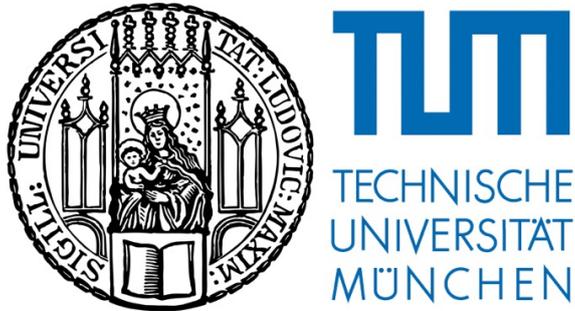


LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
&
TECHNISCHE UNIVERSITÄT MÜNCHEN

Elite Master Course in Theoretical and Mathematical Physics



Master Thesis

**Novel simulation techniques
for a sterile neutrino search with KATRIN**

Supervisor: Prof. Dr. Susanne Mertens

Student: Federico Roccati

W.S. 2017/2018

To all the people who inspire me.

*"According to Marxism and Leninism everything is matter,
even the left hand side of Einstein's equation"*

V. M.

*"La vita si svolge felicemente
in molteplici attività"*

giorgio

"I know that I know nothing"

Aristotle

Socrates

"Voglio fare l'università all'estero e divenire un scienziato"

Sossoldi - Maccio Capatonda

Declaration of authorship

I, Federico Roccati, hereby declare that I am the sole author of this Master Thesis and that I have not used any sources other than those listed in the bibliography and identified as references. I further declare that I have not submitted this thesis at any other institution in order to obtain a degree.

Date:

Signature:

Acknowledgments

I want to list here all the people that I want to thank for various reasons. I am sure I am forgetting somebody because I am getting old, so feel offended or included if your name is not in this list.

- Dio, perché non si stanca mai di raggiungermi come solo Lui sa fare,
- la mamma e la zia, per il bene che mi volete e il coraggio che mi dimostrate assecondando le mie scelte,
- my supervisors Prof. Dr. Susanne Mertens and Dr. Martin Slezàk for your great help and support during the last year and all the members of the TRISTAN group, Thierry, Alexey, Tobias, Tim, Anna, Frank, Pablo, Dominik, Lisa, Madlen, Christian, Martin, Daniel and, for a short period, Paul and Aude, you have been a great everyday company,
- Elena, finalmente una milanese simpatica, grazie per la compagnia, per quel tocco di italianità che dai alle giornate nel container, ma soprattutto per il supporto negli ultimi giorni di tesi,
- Ciccio, Mali, Albo, Bruno, Raskòl'nikolov, non saprei come fare senza di voi,
- i membri del gruppo FermiLà, siete stati la migliore compagnia che si potesse avere al Fermilab, mi avete fatto spaccare di risate ed è stato bellissimo condividere con voi un'esperienza così bella,
- Wes, because you reminded me to not undersell myself,
- la Banda della Tartina & Co. : Ceci, Romano, Rizzuto, More, Donato, Batti, Mathilde, Sanghi, Gshesh, Pablo Esala, Sasha, Thomas, Daniel, Giovanni, Benny, Leo and Mazzu, you are and have been great friends and I could always count on you,
- Tiziana e Giulia, so che mi pensate sempre e non so come ringraziarvi per il vostro sostegno,
- Frau Gritta, because... you know,

- Davide, ti taggerai volentieri nella mia tesi, ma non so come si fa,
- Azu, come farei senza i tuoi consigli?
- Michelangelo, il nonno Vander, Massimo, Pamela e bimbe, e papà, per il bene che ci vogliamo a modo nostro,
- la prof.ssa Ferro, perchè mi ha dato la possibilità di fare da tutor per i suoi corsi,
- Robert Helling, because thanks to him I had the chance to study what I really wanted,

Federico Roccati

Abstract

The KATRIN experiment major task is to probe the effective electron anti-neutrino mass with a sensitivity of 200 meV at 90% confidence level. The KATRIN setup, along with an upgraded detector and readout system, is suitable for keV-scale sterile neutrino search. The signature of a sterile neutrino in the tritium beta decay spectrum is a minuscule kink-like distortion. To enable a sensitive search for this characteristic feature, an ultra-precise modeling of the entire tritium beta spectrum is necessary. For this reason, a novel semi-analytical, multi-dimensional convolution technique has been developed. It tracks both the energy and angular distribution of the electrons as they leave the windowless gaseous tritium source of KATRIN.

In this thesis the idea and first results obtained with this new technique will be deeply discussed. Furthermore, we will present the concept of how to integrate the results in a more general simulation framework for a keV-scale sterile neutrino search with KATRIN.

Contents

1	Neutrino Physics	1
1.1	Discovery of neutrino and some history	1
1.1.1	Postulation	2
1.1.2	Description	2
1.1.3	Electron antineutrino detection	3
1.1.4	Discovery of ν_μ	3
1.1.5	Discovery of ν_τ	5
1.1.6	Are there more?	5
1.1.7	Is this the final picture?	5
1.2	Neutrino Oscillations	7
1.2.1	Solar neutrino problem	7
1.2.2	Theoretical description of neutrino oscillations in the vacuum	8
1.3	Neutrino Mass theory	11
1.3.1	Higgs mechanism in the SM	11
1.3.2	Neutrino mass term	12
1.4	Neutrino Mass: limits and determination	14
1.4.1	Cosmology	14
1.4.2	Neutrinoless double β -decay	16
1.4.3	Single β -decay	18
2	The KATRIN experiment	26
2.1	Measurement principle: MAC-E filter	26
2.2	Experimental setup	38
2.2.1	Tritium source section: features and implementation	38
2.2.2	Rear section	40
2.2.3	Transport section	41
2.2.4	Spectrometer section	43
2.2.5	Detector section	44
2.3	Sensitivity of KATRIN experiment	45
2.3.1	Expected statistical uncertainty	45
2.3.2	Sources of systematic errors	46
2.3.3	Background sources	48

2.3.4	Sensitivity	48
3	Sterile Neutrinos	50
3.1	eV Sterile Neutrinos	50
3.2	GeV Sterile Neutrinos and Dark Matter	51
3.3	keV Sterile Neutrinos and Dark Matter	51
3.3.1	Signature of a keV-sterile neutrino in KATRIN	52
4	Modeling of the full tritium spectrum	58
4.1	SSC	58
4.2	Convolution code	61
4.2.1	Motivation	61
4.2.2	Idea of the Convolution code	63
4.2.3	Structure of the Convolution code	67
4.3	Comparison of SSC and Convolution code in the last 10 eV	80
4.3.1	The transmission condition	81
4.3.2	Differential spectra	82
4.3.3	Differential spectra + Transmission	85
4.3.4	Differential spectra + Scatterings + Transmission	85
4.4	SSC-Sterile	93
4.4.1	Response matrices of the source	94
4.4.2	Conclusion	105
	Bibliography	107

Chapter 1

Neutrino Physics

Despite being one of the most beautiful and powerful theories in physics, the Standard Model (SM) has still some flaws, pointing to a theory beyond it. Baryon asymmetry in the universe, the existence of Dark Matter (DM) and the discovery of neutrino oscillations are some of the major experimental evidences not explained by the SM. Interestingly, all of them are connected to neutrino physics which makes it a very promising research field.

In order to set the ground for our discussion, in this chapter I give a general overview on the present status of neutrino physics. In section 1.1 I will highlight the major steps in neutrino physics history up to their description in the SM. In section 1.2 I will discuss neutrino oscillations and their mathematical description. Finally, in section 1.3 I will talk about the absolute neutrino mass scale, its importance and possible ways to determine it.

1.1 Discovery of neutrino and some history

The history of neutrino physics started alongside with the theory of radioactivity. While the α and γ decay show discrete spectra, the electron energy spectrum for single β decay was found to be continuous [Cha14].

This fact was in contrast with the then assumed form of the β -decay:

$$n \rightarrow p + e^{-}. \quad (1.1)$$

where n stands for *neutron*, p stands for *proton* and e^{-} stands for *electron*. In fact, in a two body decay, the kinematics constrains the final energies of the produced particles to be definite. An other inconsistency of this assumption is the violation of spin conservation.

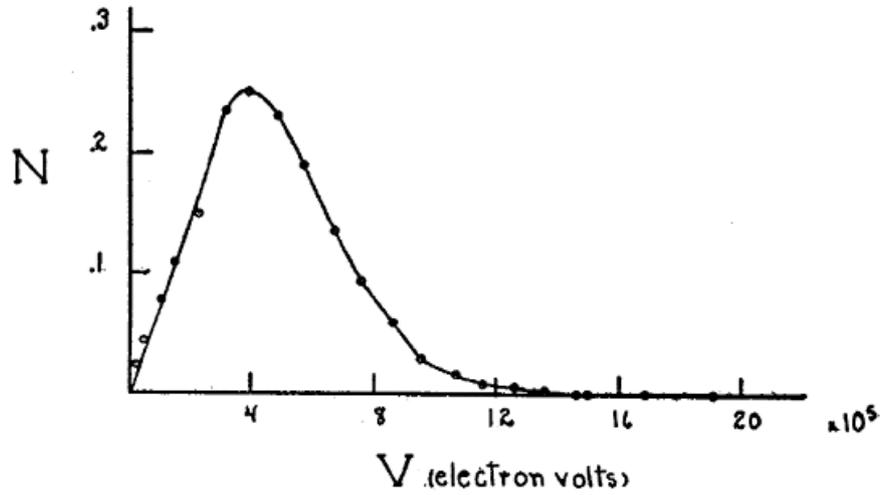


Figure 1.1: Continuous energy spectrum of the β -decay from radium. Source: [Sco35]

1.1.1 Postulation

In order to solve these puzzles, in a famous letter to the nuclear physics community in 1930 [Pau30], Pauli came up with the idea of introducing a new neutral, (almost) massless fermion, which he initially called *neutron*. This particle would then be responsible for taking away some energy from the electron and therefore for the continuous spectrum of the decay.

1.1.2 Description

After the discovery of the neutron in 1932 by Chadwick, which could not be Pauli's particle because of its mass, this still undetected particle was named by Fermi "neutrino" (Italian diminutive of neutron). The correct form of the β -decay was then



where $\bar{\nu}_e$ is the electron antineutrino because of lepton number conservation. After a couple of years, Fermi established the theory of β -decay, giving a theoretical explanation for the β spectrum [Fer34]. Fermi theory, or $V-A$ theory, is now regarded as a low energy effective theory of weak interaction and it is still valid for low energy processes as the tritium β decay taking place in the KATRIN (KARlsruhe TRITium Neutrino) experiment. According to such a theory, the decay rate of the β -decay is given by

$$\Gamma = 2\pi G_F^2 |\langle f | M_{fi} | i \rangle|^2 \frac{dN}{dE}, \quad (1.3)$$

where G_F is the Fermi constant, describing the effective weak interaction coupling, M_{fi} is the transition matrix element between the initial state $|i\rangle$ and final state $|f\rangle$ and $\frac{dN}{dE}$ is the final state density. In 1.4.3 the general form of Fermi's golden rule is used to derive the beta spectrum used in KATRIN.

1.1.3 Electron antineutrino detection

Neutrinos participate only in the weak interaction, which makes them very challenging to detect. This is why the first detection of electron antineutrinos took place more than 20 years after their postulation. In 1956 at the Savannah River Nuclear Power Plant in South Carolina, USA, a team guided by Cowan and Reines measured for the first time a signal which could be explained only by inverse β -decay [Cow56]

$$\bar{\nu}_e + p \rightarrow n + e^+. \quad (1.4)$$

The Savannah River detector consisted of liquid scintillator tanks interspaced with a Cadmium-loaded (Cd) water target (H_2O and $CdCl_2$). The electron antineutrinos coming from a near nuclear reactor, that produced an electron antineutrino flux of the order of $10^{13} \bar{\nu}_e / (s \cdot cm^2)$, interact with the protons in the water target producing a positron and a neutron. The positron immediately annihilates with an electron producing two monoenergetic gammas of $E_\gamma = 511$ keV (the mass of the electron). Instead, the neutron takes some milliseconds to thermalize (i.e. lose energy) until it is captured by the Cadmium. Subsequently, the excited Cadmium decays by emitting a γ -ray. The prompt (from the positron) and the delayed (from the neutron) light signals represent a distinct coincidence signature of an electron antineutrino interacting in the detector 1.2.

1.1.4 Discovery of ν_μ

Just six years later, in 1962, at the Brookhaven Alternating Gradient Synchrotron (AGS), Long Island, NY, Lederman, Schwartz and Steinberger proved the existence of a second neutrino flavor: the muon neutrino. In this experiment, a beam of 15 GeV protons hit a beryllium target producing π 's which subsequently decay through the channels

$$\pi^+ \rightarrow \text{antilepton} + \text{lepton neutrino}, \quad (1.5)$$

$$\pi^- \rightarrow \text{lepton} + \text{lepton antineutrino}. \quad (1.6)$$

To identify the flavor associated with the pion decay, the flavor of the charged lepton had to be detected. Pions and muons were then stopped by

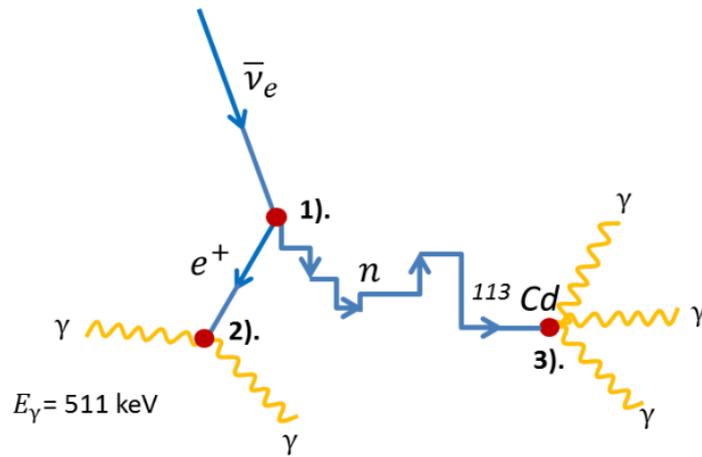


Figure 1.2: An electron anti-neutrino interacts with a proton in what is known as inverse β -decay (point 1) to a positron and a neutron. The positron immediately annihilates with an electron and subsequently decays into two photons with an energy of 511 keV (point 2). A delayed gamma signal is coming from the neutron capture (point 3). Figure from [Hub15].

a steel wall. Muon and electron neutrinos fly instead towards a spark chamber made of 10 t of aluminum where the electron showers produced by the impinging electron neutrinos are quickly stopped by the aluminium plates and muons produced by the impinging muon neutrinos make a track and go through all aluminium plates. As only tracks from muons and no electronic showers were observed, the conclusion was, that the neutrinos produced together with a muon, i.e. muon neutrinos, are intrinsically different from electron (anti-)neutrinos.

1.1.5 Discovery of ν_τ

In the year 2000, 25 years after the discovery of the tau lepton, the tau neutrino ν_τ was detected in the DONUT (Direct Observation of Nu Tau) experiment at Fermilab [Kod01]. A proton beam of 800 GeV onto a tungsten target was used to create a particle shower. A small fraction of the particles decays into τ leptons, which in turn produce ν_τ 's when they decay. After filtering all particles but the ν_τ 's, the ν_τ 's are guided towards an emulsion lead target. There they produce τ leptons, which, due to their short lifetime (3×10^{-13} s), produce a short track in the emulsion. All decay products of the τ , but the neutrinos, leave a track in the emulsion with a different direction than the τ . The famous signature of τ neutrinos is therefore a kink in the particle tracks.

1.1.6 Are there more?

At this point, the picture of leptons in the SM seems rather complete: three charged leptons paired with their respective neutrinos (and antiparticles of both). This leads to the natural question: are there any more neutrinos? The answer is related to the invisible width of the Z boson resonance and the answer was actually known before the ν_τ discovery. In short, the Z boson shows up as a peak of the total cross section of electron positron scattering. Measuring the Z peak one can get the total width Γ_{tot} and the cross section at the peak σ_{peak} . Splitting Γ_{tot} in visible (decay into leptons and quarks) and invisible (decay into neutrinos), $\Gamma_{\text{tot}} = \Gamma_{\text{vis}} + \Gamma_{\text{inv}}$, one then has $\Gamma_{\text{tot}} - \Gamma_{\text{vis}} = \Gamma_{\text{inv}} = N_\nu \Gamma(Z \rightarrow \nu\bar{\nu})$, where N_ν is the number of active light neutrinos. In 1989, the three generation picture with $N_\nu = 3$ was established by the ALEPH experiment at the LEP collider at CERN [ALE06], see figure 1.3.

1.1.7 Is this the final picture?

Up to this point of our discussion, neutrinos are included in the SM as left-handed, neutral, (only) weakly interacting fermions. A right-handed interacting component of the neutrino is excluded as observed in the Goldhaber experiment of 1958 [Gol58]. This implies that the neutrinos in the

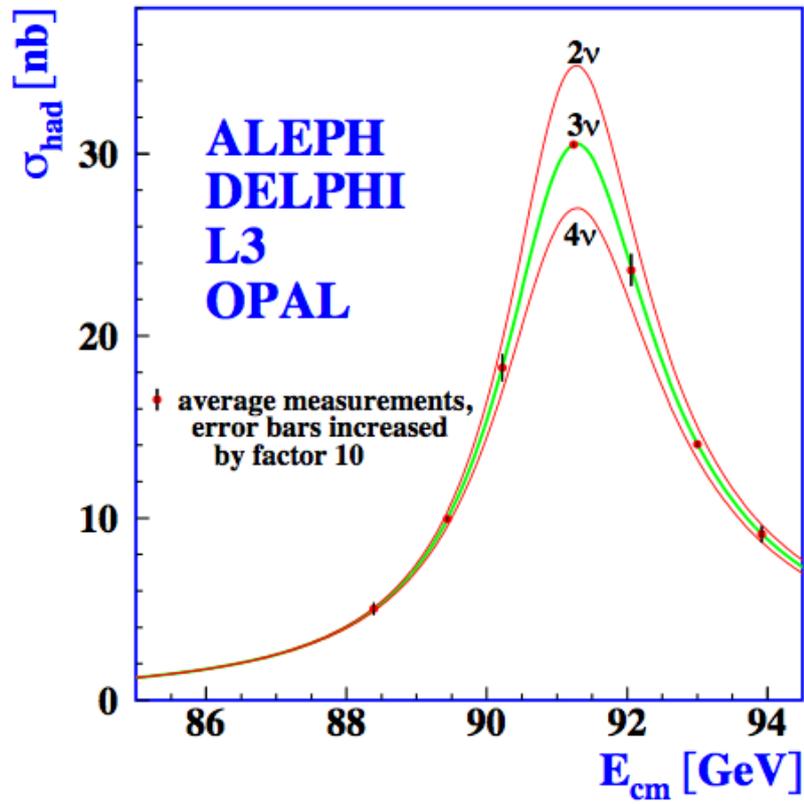


Figure 1.3: Measurement of the hadron production cross section σ_{had} around the Z resonance as a function of COM E_{cm} . The curves indicate the predicted cross section for two, three and four neutrino species with SM couplings and negligible mass. Figure from [ALE06].

SM are massless (fermions get a mass via the Higgs mechanism, coupling their right-handed part with the Higgs doublet and the $SU(2)_L$ -doublet). This means that, as their leptons partners, neutrino flavor (or weak interaction) eigenstates and mass eigenstates coincide. As I will discuss in the next section, this was found not to be the case.

1.2 Neutrino Oscillations

The discovery of neutrino flavor oscillations (NFO) represents a milestone in neutrino physics. It implies that neutrinos are not massless, pointing clearly to physics beyond the SM. Correspondingly to the three neutrino flavor states e , μ and τ , there are three neutrino mass states labelled with 1, 2 and 3. Assuming a vanishing mass for ν_1 , NFO results provide a lower mass bound. In 1.2.1 I will discuss one of the first evidences of NFO and in 1.2.2 I will show their theoretical description.

1.2.1 Solar neutrino problem

The **solar neutrino problem** refers to the observed deficit of electron neutrinos coming from the nuclear reactions taking place in the sun's core. Electron neutrinos in the sun are mainly produced via the pp-chain and CNO-cycle and their flux is roughly $6 \cdot 10^{10}/(\text{cm}^2 \cdot \text{s})$. The first experiment aimed at detecting electron neutrinos from the sun was located in the Homestake Gold Mine in Lead, South Dakota, led by Ray Davis Jr. [Dav94]. Since the neutrinos produced in the sun are **antineutrinos**, inverse β -decay ($\bar{\nu}_e + N(A, Z) \rightarrow N(A, Z + 1) + e^-$, where A is the mass number, Z is the atomic number and $N(A, Z)$ is an atom with Z protons and $A-Z$ neutrons) must be used. For this purpose, Homestake was equipped with a tank of 600 t of perchloroethylene, containing the stable isotope of chlorine ^{37}Cl as neutrino target, producing the argon radioactive isotope ^{37}Ar and electrons:



The produced argon was then extracted from the target. ^{37}Ar decays then via electron capture producing again an electron neutrino and an excited state of ^{37}Cl :



The excited chlorine then emits then X-rays and Auger electrons which are counted. Surprisingly, the observed number of neutrinos was less than the one predicted by the Solar Standard Model (SSM) [Bah05] and this result was confirmed by later experiments, e.g. GALLEX [GAL99], SAGE [SAG02] and Kamiokande [Kam96]. This either pointed towards a new SSM or to NFO, i.e. electron neutrinos change their flavor flying from the sun to the

Earth. However, the Homestake experiment was only sensitive to electron neutrinos. Therefore, a new flavor-sensitive experiment was needed to look for NFO. For this purpose, the Sudbury Neutrino Observatory (SNO), a detector located in the Creighton mine (INCO Ltd.), near Sudbury (Ontario, Canada), was built. It consists of 1000 t of almost pure D_2O contained inside a spherical 12 m diameter acrylic vessel. Cherenkov light is detected by photomultiplier tubes.

The SNO experiment detects solar neutrinos through the three reactions

$$\nu_e + D \rightarrow p + p + e^-, \quad (1.9)$$

$$\nu_\alpha + D \rightarrow n + p + \nu_\alpha, \quad (1.10)$$

$$\nu_\alpha + e^- \rightarrow \nu_\alpha + e^-, \quad (1.11)$$

where α denotes the flavor of the neutrino. The neutral-current (NC) reaction of deuterium in equation 1.10 is extremely important for checking the neutrino oscillation hypothesis of the solar neutrino problem, because it is equally sensitive to all active neutrinos. Hence, the NC reaction can measure the total flux of active neutrinos coming from the Sun.

The results from the first so called D_2O phase (Phase 1) confirmed the deficit of solar electron neutrinos observed previously. The NC measurement of the total flux of active neutrinos demonstrated that about two electron solar neutrinos out of three change their flavor to ν_μ or ν_τ on their way from the center of the Sun to the Earth. These important results have been confirmed by the more precise salt phase data (Phase 2). Hence, the results of the SNO experiment have proved that the solar neutrino problem is due to neutrino flavor transitions [SNO02].

1.2.2 Theoretical description of neutrino oscillations in the vacuum

In this part I will describe NFO in the vacuum following [Giu07]. From a mathematical point of view, NFO are described in the same way one describes quark mixing with the CKM matrix in the quark sector of the SM. This phenomenon is called *mixing* because what one sees at the beginning and at the end are bound states of quarks (e.g. K and B), and neutrino mixing is called *oscillations* because they actually propagate as free states and the probability of changing flavor oscillates as a function of the travelled distance, as I will show in the following. The reason why neutrinos oscillate is based on the fact that the flavor eigenstates, i.e. the ones that are produced in a weak interaction, are not equal to the ones that propagate through space-time, i.e. the mass eigenstates¹. The physical idea one should have

¹Idea: remember that when you write Schrödinger equation, you write an eigenvalue problem $\mathcal{H}\Psi = E\Psi$ where \mathcal{H} is the Hamiltonian, i.e. the energy, but energy is equal to mass.

in mind is that at a vertex of a Feynman diagram representing a weak interaction, the generated neutrino has a definite flavor. As soon as it starts propagating, it has an oscillating probability of changing flavor; so what is actually “propagating” is a superposition of flavor eigenstates, each with its weight, which is the mass eigenstate. As this now massive state flies, due to oscillations, it may participate in another weak interaction with a different flavor than the starting one.

Now I would like to give the mathematical description of NFO in the vacuum. Let us denote the flavor eigenstates with ν_α where $\alpha = e, \mu, \tau$, and the mass eigenstates with ν_i where $i = 1, 2, 3$. I will follow the usual description that is given in literature, i.e. I will treat neutrinos as quantum mechanical plane waves. This treatment is obviously not correct and it can be improved by treating them as localized wave packets or even quantum fields. However, it turns out this is accurate enough. Flavor and mass eigenstates are related by the so called Pontecorvo-Maki-Nakagawa-Sakata (PMNS) unitary matrix $U = (U_{\alpha i})$, which is the analogue of the CKM matrix:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1}^* & U_{e2}^* & U_{e3}^* \\ U_{\mu 1}^* & U_{\mu 2}^* & U_{\mu 3}^* \\ U_{\tau 1}^* & U_{\tau 2}^* & U_{\tau 3}^* \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (1.12)$$

If neutrinos are not Majorana particles², the PMNS matrix is parametrized by three mixing angles θ_{ij} and a Dirac CP violating phase δ_D :

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_D} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta_D} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.13)$$

where $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$. Let us consider a concrete example in which an electron neutrino ν_e is produced at time $t = 0$. According to equation (1.12) one has

$$|\nu_e\rangle = |\nu(t=0)\rangle = U_{e1}^* |\nu_1\rangle + U_{e2}^* |\nu_2\rangle + U_{e3}^* |\nu_3\rangle. \quad (1.14)$$

Now, each of the massive neutrino states $|\nu_i\rangle$ is an eigenstate of the Hamiltonian \mathcal{H} with the energy eigenvalue $E_i = \sqrt{\vec{p}^2 + m_i^2}$, where \vec{p} is the neutrino momentum and m_i its mass:

$$\mathcal{H} |\nu_i\rangle = E_i |\nu_i\rangle. \quad (1.15)$$

The Schrödinger equation ($\hbar = 1$)

$$i \frac{d}{dt} |\nu_i(t)\rangle = \mathcal{H} |\nu_i(t)\rangle \quad (1.16)$$

²that is, they are not their own antiparticles

implies that the massive neutrino states evolve in time as plane waves:

$$|\nu_i(t)\rangle = e^{-iE_i t} |\nu_i\rangle \quad (1.17)$$

Combining equation (1.14) and (1.17) one then has

$$|\nu_e(t > 0)\rangle = U_{e1}^* e^{-iE_1 t} |\nu_1\rangle + U_{e2}^* e^{-iE_2 t} |\nu_2\rangle + U_{e3}^* e^{-iE_3 t} |\nu_3\rangle. \quad (1.18)$$

The neutrino state in equation (1.18) is not necessarily an electron neutrino. It is just the time evolution of what once was an electron neutrino, so the notation $|\nu_e(t > 0)\rangle$ should not confuse the reader. To see clearly how neutrinos can change their flavor, it is better to calculate

$$P_{\alpha\beta}(t) := P(\nu_{\alpha \rightarrow \beta}(t)), \quad (1.19)$$

i.e. the probability that the state at time t $\nu_\alpha(t)$ is found in the state ν_β .

Using equation (1.18) and equation (1.12) two times I can write

$$|\nu_\alpha(t)\rangle = \sum_i U_{\alpha i}^* e^{-iE_i t} |\nu_i\rangle \quad (1.20)$$

$$= \sum_i U_{\alpha i}^* e^{-iE_i t} \left(\sum_{\beta=e,\mu,\tau} U_{i\beta} |\nu_\beta\rangle \right) \quad (1.21)$$

$$= \sum_{\beta=e,\mu,\tau} \left(\sum_i U_{\alpha i}^* e^{-iE_i t} U_{i\beta} \right) |\nu_\beta\rangle \quad (1.22)$$

Now from quantum mechanics one has

$$P_{\alpha\beta} = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2 \quad (1.23)$$

$$= \left| \sum_i U_{\alpha i}^* e^{-iE_i t} U_{i\beta} \right|^2 \quad (1.24)$$

$$= \sum_{i,j} U_{\alpha i}^* U_{i\beta} U_{\alpha j} U_{j\beta}^* \exp[-i(E_i - E_j)t]. \quad (1.25)$$

Using now the ultrarelativistic approximation³ $p_i = p = E$, I get the standard formula for the oscillation probability

$$P_{\alpha\beta}(L/E) = \sum_{i,j} U_{\alpha i}^* U_{i\beta} U_{\alpha j} U_{j\beta}^* \exp \left[-i \Delta m_{ij}^2 \frac{L}{2E} \right], \quad (1.26)$$

³Here is where it gets confusing, because quantum mechanics is not a relativistic theory, I am describing neutrinos as plane waves as if they had the speed of light, despite their mass.. but hold on! we are almost done! (see p.253 [Giu07])

where $\Delta m_{ij}^2 = m_i^2 - m_j^2$, L is the distance between the source and the detector (in ultrarelativistic approximation: $L = t$ if $c = 1$), and E is the neutrino energy. This probability is thus determined by the parameters in the PMNS matrix, the two mass splittings and L

The general three neutrino mixing can be simplified considering that the mixing angle θ_{13} is small compared to the other two. Equation (1.26) reduces then to the most readable two neutrino mixing probability:

$$P(\nu_e \rightarrow \nu_\mu) = \sin^2(2\theta_{12}) \sin^2\left(\frac{\Delta m_{12}^2 L}{4E}\right). \quad (1.27)$$

Equation (1.27) shows more clearly that the amplitude of the oscillation depends on the mixing angle, while the frequency on the squared mass difference.

NFO were a remarkable result in recent years research. However, as the above formulas show, oscillation experiments are only sensible to the squared mass *difference* and not to the actual absolute mass scale. Before diving into the KATRIN experiment, whose goal is investigating the absolute mass scale of the neutrino, let us give an overview of how the neutrino mass can be introduced theoretically, experimental current limits and in which way one can effectively discover it.

1.3 Neutrino Mass theory

NFO alone do not provide any information about the absolute scale of neutrino mass. However, to proceed in our discussion, I must anticipate now that neutrino mass is very small, at most in the eV scale, as I will discuss later.

1.3.1 Higgs mechanism in the SM

In the SM it is a well established fact that all massive particles gain their mass through the Higgs mechanism, which is a mechanism to get mass terms in the lagrangian in a gauge invariant way. I focus here only on fermion mass. At high energy (i.e. in the early universe) the gauge group of the SM is (I do not consider the color charge here) $SU(2)_L \times U(1)_Y$ and the Higgs potential has one minimum, where the Higgs field resides. Introducing in the Lagrangian the so called *Yukawa term* (for simplicity, I write it just for the electron flavor)

$$\lambda_e \bar{e}_R \phi^\dagger \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} + \text{h.c.} \quad (1.28)$$

where λ_e is the Yukawa coupling for electrons, \bar{e}_R is the right-handed part of the electron field (gauge singlet), ϕ is the column Higgs complex $SU(2)$ -

doublet, and $\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}$ is the $SU(2)_L$ -doublet which takes part in the weak interaction. At low energy (early universe transition: electroweak symmetry breaking), the Higgs potential takes its typical Mexican hat shape and is still $SU(2)$ -symmetric. However, the Higgs field, choosing one of the minima, breaks spontaneously the $SU(2)$ -symmetry. Choosing a convenient gauge, one can write this minimum as

$$\phi_0 \propto \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad (1.29)$$

where v is the Higgs vacuum expectation value (VEV), and after some algebra one gets

$$\lambda_e v (\bar{e}_R e_L + \bar{e}_L e_R) = \lambda_e v \bar{e} e. \quad (1.30)$$

A term like $\bar{\psi}\psi$ is a mass term in the Lagrangian of the SM for the field ψ , so $m_e \propto \lambda_e v$. Measuring the Higgs VEV one then fine tunes the Yukawa coupling, which is a parameter in the SM.

Now come the issues with neutrinos. There are two major problems. First, to get a mass term, one needs the right-handed part of a fermion field and, from the Goldhaber experiment in 1958, neutrinos were observed to have only negative helicity which implies that they are only left-handed (because they were assumed massless). Second, all charged fermions have masses in the range from about 0.5 MeV (electron) up to 173 GeV (top quark), so the respective Yukawa couplings span 6 orders of magnitude, from $\lambda \sim 10^{-6}$ (for electron) to $\lambda \sim 1$ (top quark) orders of magnitude. There are many ways to address these issues and I present some of them here. Whatever the solution is, it is obviously beyond the SM physics.

1.3.2 Neutrino mass term

In principle, one can introduce by hand a right handed neutrino component ν_R . However, one can show that all its quantum numbers vanish so it is a singlet under all SM interactions, that is why this is called **sterile neutrino**. Assuming this ν_R exists, even if it is impossible to detect being coupled only to gravity, then one could write the Yukawa coupling for neutrinos as in equation (1.28) (only for the electron neutrino for simplicity) as⁴

$$\lambda_{\nu_e} \bar{\nu}_{eR} \phi^T i\sigma^2 \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} + \text{h.c.}, \quad (1.31)$$

⁴(1.31) looks different from (1.28) because one has to give mass to the upper component of the $SU(2)_L$ -doublet

where σ^2 is the third Pauli matrix.

However, this is not very appealing because giving the smallness of neutrino masses, their Yukawa coupling would be all of the order of $\lambda_\nu \sim 10^{-13}$ leaving a big gap between the heaviest neutrino state and the electron. In physics this is considered fine-tuning of the parameters and it is not very nice. As a consequence, the question arises why the Higgs mechanism is good only for the MeV-GeV range and the \sim eV range. This fact points to other mechanisms viable for neutrino mass. I describe here only the so-called **see-saw mechanism** [Pet13].

See-saw mechanism

The main issue the Higgs mechanism has to address is the smallness of neutrino mass. To do so, let us make the following observation. The mass term one gets from the Yukawa coupling is called **Dirac mass term**, because it involves Dirac fields with left- and right-handed components. However, neutral fermions have the theoretical possibility of being **Majorana** fermions. A fermion represented by the field

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \quad (1.32)$$

is Majorana if $\psi_R = \psi_L^c$, i.e. if it is equal to its own antiparticle. In the SM this is only possible for neutrinos since they are the only neutral fermions. Furthermore, if a fermion is Majorana, one can directly write a mass term in the lagrangian without violating gauge invariance as $M\bar{\psi}_R\psi_R$. In the **see-saw** mechanism one introduces one or more right-handed heavy neutrino field with the corresponding Majorana mass term. After symmetry breaking, the mass terms in the Lagrangian look then like ($\nu = \nu_e$)

$$\mathcal{L}_{\nu, mass} = m_D \bar{\nu}_L \nu_R + \frac{M}{2} \bar{\nu}_R \nu_R, \quad (1.33)$$

where m_D stands for Dirac mass and m_M for Majorana mass, so left- and right-handed parts mix. Thus, the mass eigenstates are those states that diagonalize the matrix

$$\begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix}. \quad (1.34)$$

In the limit of $m_D \ll M$, the two eigenvalues are

$$m_1 \approx m_D, \quad (1.35)$$

$$m_2 \approx \frac{m_D^2}{M}. \quad (1.36)$$

If the heavy neutrino mass is in the GeV scale, then m_2 is in the eV scale and can be interpreted as the SM neutrino mass.

Generalizing to the complete three flavor case, one can fine-tune the three masses of heavy neutrino states to set the scale of light neutrinos.

However, no sterile neutrino has been detected yet. Other neutrino mass generation mechanisms can be found in [Kin04].

1.4 Neutrino Mass: limits and determination

In this section I want to give an overview on the current limits on neutrino masses given by present experimental data and how it will be improved in the future. The current best limit on neutrino mass comes from the Mainz [Bon01] and Troitsk [Lob99] experiments and is given by

$$m_\nu < 2 \text{ eV}. \quad (1.37)$$

An important question regarding neutrino masses is whether they are **hierarchical** or **degenerate**. From oscillation experiments, it is known that the numerical value of the difference of the squared masses is very small ($\Delta m_{12}^2 \sim 10^{-5} \text{ eV}^2$, $\Delta m_{23}^2 \sim 10^{-3} \text{ eV}^2$, so $\Delta m_{13}^2 = \Delta m_{12}^2 + \Delta m_{23}^2$). Writing m_2 and m_3 as functions of m_1

$$m_i = \sqrt{m_1^2 + |\Delta m_{1i}^2|} \quad i = 2, 3, \quad (1.38)$$

one can see that if neutrino masses are in the eV scale, then they are **degenerate** ($m_1 \approx m_2 \approx m_3$) because the m_1 term in 1.38 dominates the Δm_{1i}^2 term; if instead neutrino masses lie in the sub-eV scale, then there is a precise hierarchy among them. However, having the mass splittings and not knowing the sign of Δm_{13}^2 , it is not clear whether the hierarchy is **normal** ($m_1 < m_2 < m_3$) or **inverted** ($m_3 < m_1 < m_2$). Absolute neutrino mass searches are then important to rule out one of the two scenarios, normal or inverted.

If they lie in the eV scale, then they are degenerate. If they lie in the sub-eV scale then they are hierarchical, either in normal or inverted hierarchy.

Unfortunately, it is important to stress, as I will discuss in the following, that from cosmology, neutrinoless double β -decay ($0\nu\beta\beta$) and single β -decay experiments one only gets upper bounds on an effective neutrino mass.

1.4.1 Cosmology

According to the current accepted cosmological model, the Λ CDM-model [Ber03], the history of the Universe is the history of a very hot soup of interacting particles, born at the Big Bang, that cools down while expanding. Every time the expansion rate exceeds the collision rate for particular particle species, a **freeze-out** (or **decoupling**) takes place.

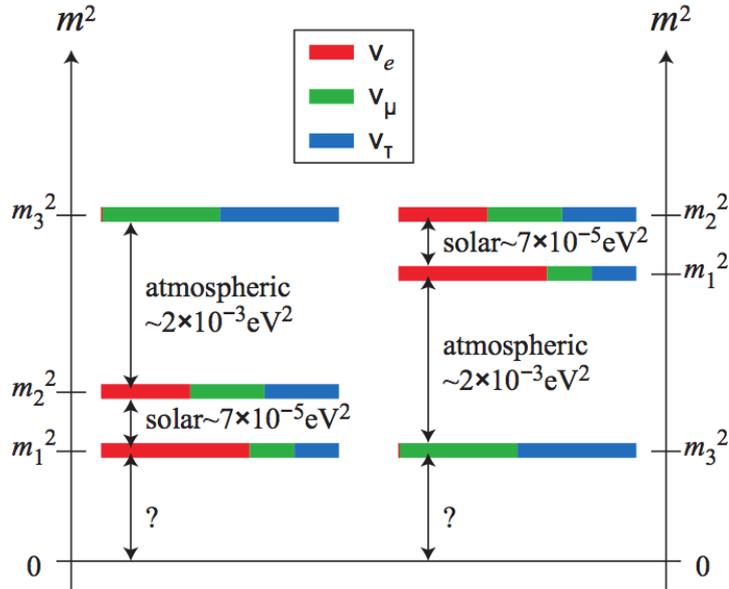


Figure 1.4: Neutrino mass hierarchy and flavor content of mass eigenstates for the two generic scenarios of normal (left) and inverted (right) hierarchy. Figure from [Kin13].

A simple way to picture this scenario is thinking about a box full of very fast interacting particles. If one expands slowly the box, the whole plasma has the time to adapt to the distortion, so no decoupling occurs. If one instead expands the box abruptly, then some of the particles will never reach any other particle they can interact with, i.e. they will be decoupled from the plasma.

The history of the Universe is then the history of subsequent decouplings of all kinds of particles. The most famous decoupling, which is also one of the most outstanding predictions of the Λ CDM-model, is the photon decoupling, better known as **Cosmic Microwave Background (CMB) radiation** [Alp48, Pen65]. It took place 300.000 years after the Big Bang leaving a relic homogeneous (part per thousand) background of ~ 400 photons/cm³ at a temperature of 2.7 K.

Analogously to photon decoupling, the Λ CDM-model predicts neutrino decoupling. When the Universe was only 1 s old ($T \sim 10^9$ K, $E \sim 1$ MeV), these high energy, then relativistic, neutrinos would have frozen out from thermal equilibrium leaving a density of about 336 neutrinos/cm³ in the whole Universe. This means that neutrinos are the most abundant particles in the Universe after photons. It is clear then that the knowledge of their absolute mass scale is of major importance in determining their contribution

to the matter density in the Universe. The contribution of relic neutrinos Ω_ν to the total energy density of the Universe Ω_{tot} is given by

$$\Omega_\nu = \frac{\sum_i m_{\nu_i}}{93.14 h^3 \text{ eV}} \quad (1.39)$$

where h is the dimensionless Hubble constant [Les12]. One sees then that cosmology is sensitive to the effective neutrino mass

$$m_{\text{cosm}} = \sum_i m_{\nu_i}. \quad (1.40)$$

The best bound on this value comes from the Planck experiment [Pla14] which sets

$$m_{\text{cosm}} < 0.23 \text{ eV (95\% C.L.)}. \quad (1.41)$$

Unfortunately, due to their very low energy (sub-eV scale), these relic neutrinos have not been detected yet. Furthermore, cosmological observations strongly rely on the underlying astrophysical models.

1.4.2 Neutrinoless double β -decay

In nature, there are some heavy nuclei (e.g. ^{76}Ge) for which a single β -decay is energetically disfavoured, while a double β -decay

$$2n \rightarrow 2p + 2e^- + 2\bar{\nu}_e \quad (1.42)$$

is allowed. If the neutrino is a Majorana particle, i.e. its own antiparticle, then it is in principle possible to observe the following decay mode, whose Feynman diagram is shown in figure 1.5:

$$2n \rightarrow 2p + 2e^-. \quad (1.43)$$

This rare decay, called **neutrinoless double β -decay** (denoted by $0\nu\beta\beta$ -decay), besides providing information on the neutrino mass, can also tell us something interesting on neutrino nature. It violates lepton number conservation and it would definitely be a proof for physics beyond the SM.

The rate of this very rare event is proportional to another effective (electron anti-)neutrino mass, called **Majorana mass**,

$$\Gamma_{0\nu\beta\beta} \propto m_{\beta\beta} = \left| \sum_i U_{ei}^2 m_i \right|. \quad (1.44)$$

Many experiments, like GERDA and MAJORANA DEMONSTRATOR [GER06, MAJ14], are looking for a signal of this rare event, which would reveal itself as a peak at the endpoint of the double beta decay spectrum. To infer the parameter $m_{\beta\beta}$ from the data, the lifetime $T_{1/2}$ of the process (1.43)

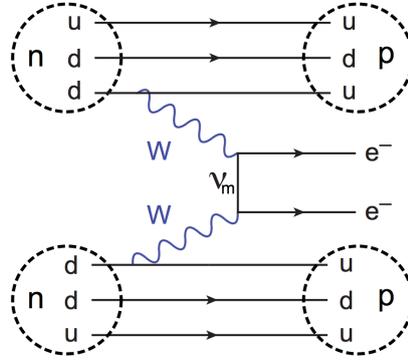


Figure 1.5: Feynman diagram for the neutrinoless double β -decay. Two neutrons decay simultaneously into two protons and two electrons via the W boson by exchanging a virtual Majorana neutrino ν_m . Figure from [Sch13].

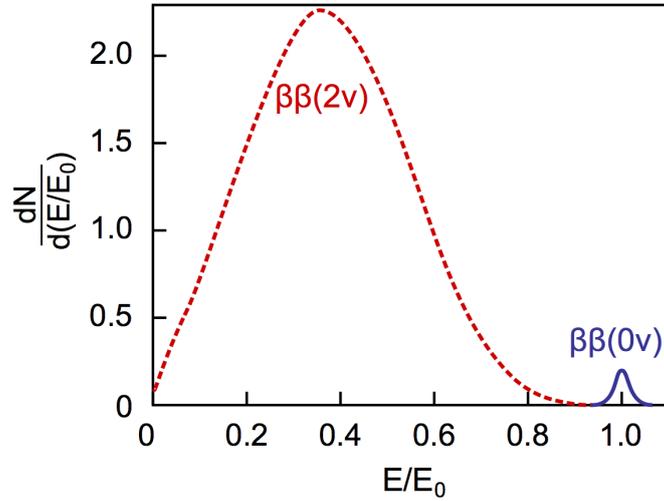


Figure 1.6: Energy spectrum of both electrons is shown for neutrinoless (0ν) as well as for normal (2ν) double β -decay. The $0\nu\beta\beta$ -decay results in a peak at the endpoint E_0 , which is broadened by the energy resolution of the detector of here 5% and increased artificially for the shown example. Figure from [Sch13].

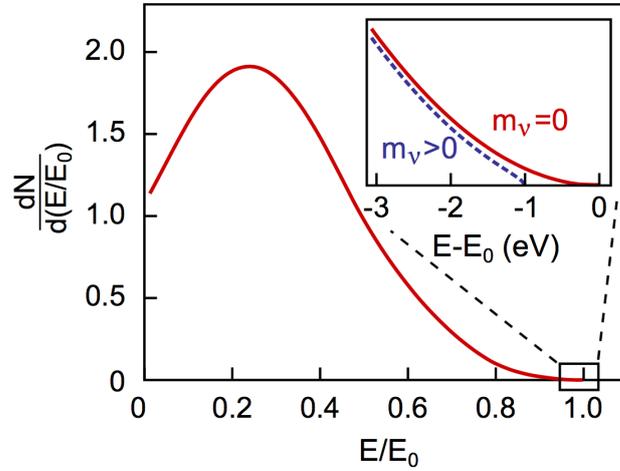


Figure 1.7: Energy spectrum of electrons from tritium β -decay normalized to the endpoint E_0 . A non-zero neutrino mass (blue dashed line) would reduce the maximal kinetic energy of the electron by m_ν . Figure from [Sch13].

is measured. Current results sets a lower bound of $\sim 10^{25}$ years for $T_{t/2}^5$, which in turn implies an upper bound for the Majorana mass of $0.2 - 0.4$ eV [GER13, Ago16].

However, these results rely on the assumption that the neutrino is a Majorana particle and on detailed decay models.

A model-independent way of measuring the neutrino mass is indeed needed and is presented in the following paragraph.

1.4.3 Single β -decay

Single β -decay experiments are the most model-independent way to measure neutrino mass, because they make only use of kinematics and energy-momentum conservation. In such an experiment, one wants to measure the β spectrum of the decay, i.e. the kinetic energy distribution of the produced electrons. For neutrino mass searches, one needs only a high precision measurement of the spectrum near the **endpoint**, that is the maximal kinetic energy the electron can carry away. Being the energy divided between the electron and the neutrino (and the daughter nucleus, but I will be more detailed later), when the electron has almost all the kinetic energy, the neutrino has enough energy only for its rest mass. Therefore, the kink-like signature due to neutrino mass is more visible near the endpoint of the spectrum.

I want to derive now in more details the formula of the β -spectrum.

⁵as I once heard from GERDA people toasting: "To nothing being detected!"

Derivation of the β -spectrum

In this section I derive the differential spectrum of tritium β -decay closely following [Dre13]. According to Fermis Golden Rule, the decay rate for a β -decay is given by

$$\Gamma = 2\pi \sum \int |\mathcal{M}|^2 df \quad (1.45)$$

where \mathcal{M} is the transition matrix element and $\sum \int df$ represents the sum (integral) over all possible discrete (continuous) final states f , respectively⁶.

First let us focus on the df term of formula (1.45). Let us define dn as the number of different final states of outgoing particles inside a normalization volume V into the solid angle $d\Omega$ with momentum in $[p, p + dp]$. One has

$$\begin{aligned} dn &= \frac{V d^3 \vec{p}}{h^3} = \\ &= \frac{V p^2 dp d\Omega}{h^3} = \quad (E_{\text{tot}}^2 = m^2 + p^2 \Rightarrow p dp = E_{\text{tot}} dE_{\text{tot}}) \\ &= \frac{V p E_{\text{tot}} dE_{\text{tot}} d\Omega}{(2\pi)^3}, \end{aligned} \quad (1.46)$$

where h is the Plank constant, which in natural units is equal to 2π , E_{tot} is the total energy (rest and kinetic) of the state n . The state density per energy interval and solid angle will then be

$$\frac{dn}{dE_{\text{tot}} d\Omega} = \frac{V p E_{\text{tot}}}{(2\pi)^3}. \quad (1.47)$$

The mass of the daughter nucleus is much larger than the energies of the two emitted leptons, so one can assume that the nucleus takes nearly no kinetic energy, but balances all momenta (I will consider the recoil energy of the nucleus later). Therefore, one can write only the state density of the electron and the neutrino as

$$\begin{aligned} \rho(E_e, E_\nu, d\Omega_e, d\Omega_\nu) &= \frac{dn_e}{dE_e d\Omega_e} \frac{dn_\nu}{dE_\nu d\Omega_\nu} \\ &= \frac{V^2 p_e E_e p_\nu E_\nu}{(2\pi)^6} \\ &= \frac{V^2 \sqrt{E_e^2 - m_e^2} E_e \sqrt{E_\nu^2 - m_\nu^2} E_\nu}{(2\pi)^6}, \end{aligned} \quad (1.48)$$

where I have used formula (1.47) for the electron and neutrino energy E_e and E_ν , respectively.

⁶2 and π were defined in (1.3)

Let us now focus on the matrix element \mathcal{M} of formula (1.45). It can be divided into leptonic and nuclear part as

$$\mathcal{M} = G_F \cos \theta_C \mathcal{M}_{\text{lep}} \mathcal{M}_{\text{nucl}} \quad (1.49)$$

where θ_C is the Cabibbo angle.

Let us focus here on **allowed or superallowed transitions** like those of tritium. In this kind of decays, none of the leptons carries away angular momentum, both leptons are treated as plane waves and the leptonic matrix element is just the product of the probability of finding the neutrino at the nucleus, which is $1/V$, and the probability of finding the electron at the nucleus, which is instead $1/V \cdot F(E, Z')$:

$$|\mathcal{M}_{\text{lep}}|^2 = \frac{1}{V^2} F(E, Z'). \quad (1.50)$$

$F(E, Z')$ is the **Fermi function** describing the Coulomb interaction of the β -electron and the daughter nucleus with atomic number Z' .

In an allowed or superallowed transition, $\mathcal{M}_{\text{nucl}}$ is independent of the kinetic energy of the β -electron and the coupling of the lepton spins to the nuclear spin is usually contracted into this term. The nuclear matrix element of an allowed or superallowed transition can be divided into a Fermi part ($\Delta I_{\text{nucl}} = 0$) and into an Gamov-Teller part ($\Delta I_{\text{nucl}} = 0, \pm 1$, but no $I_{\text{nucl}} = 0 \rightarrow I_{\text{nucl}} = 0$). In the former case, the spins of electron and neutrino couple to $S = 0$, in the latter case to $S = 1$. What remains is an angular correlation of the two outgoing leptons. Since charge current weak interactions like β -decay maximally violate parity, they prefer, depending on velocity, negative helicities for particles and positive helicities for antiparticles. Therefore, the momenta or directions of the leptons are correlated with respect to their spins and therefore to each other. This implies a (β, ν) angular correlation factor

$$1 + a(\vec{\beta}_e \cdot \vec{\beta}_\nu) \quad (1.51)$$

with the electron velocity $\beta_e = v_e/c$ and the neutrino velocity $\beta_\nu = v_\nu/c$. The angular correlation coefficient a amounts to $a = 1$ for pure Fermi transitions and to $a = -1/3$ for pure Gamov-Teller transitions within the SM [Sev06].

The phase space density (1.48) is distributed over a surface in the two-particle phase space which is defined by a δ -function conserving the decay energy. Every decay has its decay channels, with a certain probability P_i of taking place. As a consequence, one can write the decay rate as

$$\Gamma = \sum_i P_i \Gamma_i \quad (1.52)$$

To start I compute the partial decay rate Γ_0 from equation (1.45):

$$\begin{aligned}
\Gamma_0 &= 2\pi P_0 \int_{E_e, E_\nu, \Omega_e, \Omega_\nu} |G_F \cos \theta_C \mathcal{M}_{\text{lep}} \mathcal{M}_{\text{nucl}}|^2 dn_e dn_\nu \\
&= \frac{P_0}{(2\pi)^5} \int_{E_e, E_\nu, \Omega_e, \Omega_\nu} G_F^2 \cos^2 \theta_C F(E, Z') |\mathcal{M}_{\text{nucl}}|^2 \cdot \\
&\quad \cdot \sqrt{E_e^2 - m_e^2} \cdot E_e \cdot \sqrt{E_\nu^2 - m_\nu^2} \cdot E_\nu \cdot (1 + a(\vec{\beta}_e \cdot \vec{\beta}_\nu)) \cdot \\
&\quad \cdot \delta(Q - (E_e - m_e) - E_\nu - E_{\text{rec}}) dE_e d\Omega_e dE_\nu d\Omega_\nu. \tag{1.53}
\end{aligned}$$

Q is the Q -value of the process, i.e. the energy released in the decay. According to the δ -function in equation (1.53), this energy is distributed into the kinetic energy of the electron $E := E_e - m_e$, the total energy of the neutrino E_ν and the recoil energy of the daughter nucleus E_{rec} . The maximal kinetic energy an electron can take from Q for the case of zero neutrino mass is called **endpoint energy** E_0 which is defined by a vanishing neutrino energy E_ν :

$$E_0 = \max_{m_\nu=0} E. \tag{1.54}$$

A correct integration over the unobserved neutrino variables in (1.53) has to respect the (β, ν) angular correlation factor (1.51), which also has to be considered when calculating the exact recoil energy of the nucleus E_{rec} . If one assumes that the β -electrons of interest have a certain minimal kinetic energy E_{min} , then one can calculate the range of recoil energies of the daughter nucleus of mass m_{daughter} : the recoil energy E_{rec} is bound from above by the case in which the outgoing electron takes the maximum kinetic energy E_0 (to understand this, think of it as the electron and the nucleus recoiling back to back: if the electron takes all the kinetic energy it can, then it will have 3-momentum \vec{p}_e and $\vec{p}_{\text{rec}} = -\vec{p}_e$)

$$\begin{aligned}
E_{\text{rec}} &\leq E_{\text{rec, max}} \\
&:= \frac{p_{\text{max}}^2}{2m_{\text{daughter}}} \\
&= \frac{(m_e + E_0)^2 - m_e^2}{2m_{\text{daughter}}} \\
&= \frac{E_0^2 + 2m_e E_0}{2m_{\text{daughter}}} \tag{1.55}
\end{aligned}$$

and from below by the case, in which the electron of kinetic energy E_{min} is emitted opposite to the direction of the neutrino, which has in this case a momentum $p_\nu = E_\nu = E_0 - E_{\text{min}}$ (neglecting for a moment the nonzero value of the neutrino mass)

$$\begin{aligned}
E_{\text{rec}} &\geq E_{\text{rec},\text{min}} \\
&:= \frac{(p_e - p_\nu)^2}{2m_{\text{daughter}}} \\
&= \frac{(\sqrt{(m_e + E_{\text{min}})^2 - m_e^2} - (E_0 - E_{\text{min}}))^2}{2m_{\text{daughter}}} \\
&= \frac{\left(\sqrt{E_{\text{min}}^2 + 2m_e E_{\text{min}}} - (E_0 - E_{\text{min}})\right)^2}{2m_{\text{daughter}}}. \tag{1.56}
\end{aligned}$$

Being m_{daughter} very large for every electron energy below the endpoint E_0 , according to (1.55) and (1.56) the recoil energy E_{rec} does not change much. A fortiori, in the region of interest below the endpoint, one can assume $E_{\text{rec}} = \text{const}$ and (1.54) becomes

$$E_0 = Q - E_{\text{rec}}. \tag{1.57}$$

Integrating then over the angles yields through (1.51) an averaged nuclear matrix element. Furthermore, one has to sum over all final states. For the β -decay of an atom or a molecule it is a double sum: one summation runs over all neutrino mass eigenstates m_i with probabilities $|U_{ei}|^2$ which are kinematically accessible ($m_i \leq E_0$). The second summation runs over all electronic final states of the daughter system each of which has probability P_j and excitation energy V_j . These include excitations of the electron shell, but also, in the case of β -decaying molecules, which is the case for KATRIN, rotational and vibrational excitations. The latter are caused by the sudden change of the nuclear charge from Z to $Z + 1$ which requests a rearrangement of the electronic orbitals of the daughter atom or molecule and the interatomic distances in case of a molecule. They give rise to shifted endpoint energies. Defining

$$\varepsilon := (E - E_0), \tag{1.58}$$

the total neutrino energy amounts to $E_{\nu,j} = \varepsilon - V_j$ if the excitation energy is V_j .

I am interested in writing a formula for the **differential** energy spectrum

$$\frac{dN}{dt dE} = \frac{d\Gamma}{dE} \tag{1.59}$$

which gives the number of counts per second per energy ⁷.

⁷However KATRIN measures the **integral** energy spectrum, i.e. the number of counts per fixed retarding potential. I will go into details later.

One can read the differential spectrum directly from (1.53) without performing the second integration over E . Using (1.58) and summing over the final states I get

$$\begin{aligned} \frac{d\Gamma}{dE} &= C \cdot F(E, Z') \cdot p_e \cdot (E + m_e) \cdot \sqrt{(E + m_e)^2 - m_e^2} \cdot \\ &\quad \cdot \sum_{i,j} |U_{ei}|^2 \cdot P_j \cdot (\varepsilon - V_j) \cdot \sqrt{(\varepsilon - V_j)^2 - m_i^2} \end{aligned} \quad (1.60)$$

where $C = \frac{G_F^2 \cos^2 \theta_C |\langle \mathcal{M}_{\text{nucl}} \rangle_{\text{angles}}|^2}{2\pi^3}$ and a Θ -function $\Theta(\varepsilon - V_j - m_i)$ to confine the spectral component to the physical sector is assumed.

Considering now the fact that up to now one has not enough sensitivity at our disposal to resolve the three neutrino mass eigenstates, the unitarity of the PMNS matrix and that $m_i^2 \ll (\varepsilon - V_j)^2$, equation (1.60) becomes

$$\begin{aligned} \frac{d\Gamma}{dE} &= C \cdot F(E, Z') \cdot p_e \cdot (E + m_e) \cdot \sqrt{(E + m_e)^2 - m_e^2} \cdot \\ &\quad \cdot \sum_j P_j \cdot (\varepsilon - V_j) \cdot \sqrt{(\varepsilon - V_j)^2 - m_\beta^2} \end{aligned} \quad (1.61)$$

where

$$m_\beta^2 = \sum_i |U_{ei}|^2 m_i^2 \quad (1.62)$$

is the so called **effective electron anti-neutrino mass**, an incoherent sum of neutrino masses, where, in contrast to $m_{\beta\beta}$, no cancellations can occur. One can immediately see that, contrarily to $0\nu\beta\beta$ experiments, single β -decay experiments are not sensitive to the Majorana nature of the neutrino, because any eventual Majorana phase in the PMNS matrix is wiped out by the modulus over U_{ei} . This fact stresses particularly the complementarity of the approaches, from cosmology to laboratory experiments.

1.4.3.0.1 Tritium β -decay Let us now highlight how tritium as a β emitter impacts the spectrum (1.61). In the next chapter, I will see additional reasons to use tritium in the KATRIN experiment.

Tritium β -decay is the following process



Tritium has a half-life of 12.3 y. Tritium and Helium-3 are mirror nuclei of the same isospin doublet; therefore, the decay is superallowed. Consequently, the nuclear matrix element for tritium is close to that of the β -decay of the free neutron and amounts to $|\mathcal{M}_{\text{nucl}}^2(\text{tritium})|^2 = 5.55$ [Rob88].

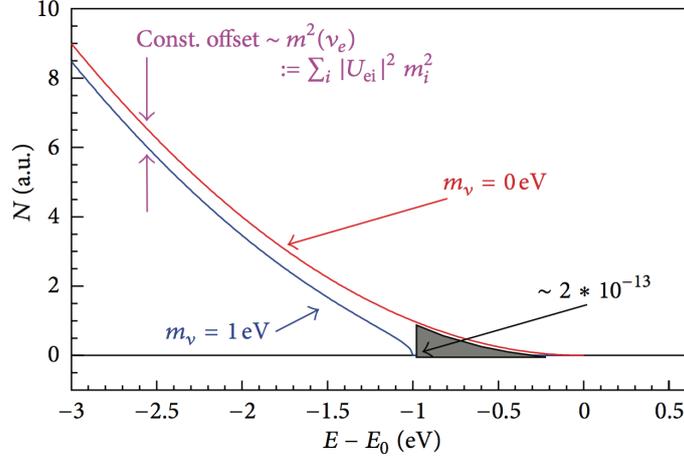
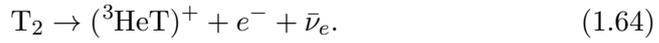


Figure 1.8: Expanded β -spectrum of an allowed or superallowed β -decay around its endpoint E_0 for $m_\nu = 0$ (red line) and for an arbitrarily chosen neutrino mass of $m_\nu = 1$ eV (blue line). In the case of tritium, the gray-shaded area corresponds to a fraction of $2 \cdot 10^{-13}$ of all tritium β -decays. Figure from [Dre13].

Tritium is a good compromise between a long half-life and a low endpoint (18.575 keV). Its simple electronic structure allows us to measure the tritium β spectrum with small systematic uncertainties. The recoil correction for tritium is not an issue. Up to now all tritium β -decay experiments used molecular tritium, which gives a maximal recoil energy to the daughter molecular ion of $E_{\text{rec,max}} = 1.72$ eV. Even for the most sensitive tritium β -decay experiment, the KATRIN experiment (see next chapter), the maximum variation of E_{rec} over the energy interval of investigation (the last 30 eV below the endpoint) only amounts to $E_{\text{rec}} = 3.5$ meV. It was checked [Mas07] that this variation can be neglected and the recoil energy can be replaced by a constant value of $E_{\text{rec}} = 1.72$ eV, yielding a fixed endpoint according to (1.57).

Concerning the calculation of the electronic final states with energies V_j , one has to take into account that in reality one works with molecular tritium decay



Its wave functions are much more complicated, since in addition to two identical electrons, they comprise also the description of rotational and vibrational states, which may be excited during the decay. The first group of excited electronic states starts at around $V_j = 25$ eV. Therefore, excited states play almost no role for the energy interval considered for the KATRIN experiment: only the decay to the ground state of the $({}^3\text{HeT})^+$ daugh-

ter molecule, which is populated with about 57% probability, has to be taken into account. Due to the nuclear recoil, however, a large number of rotational-vibrational states with a mean excitation energy of 1.7 eV and a standard deviation of 0.4 eV are populated. These values hold for a pure T₂ source without contamination by other hydrogen isotopes. Fortunately, a contamination of the T₂ molecules by DT or HT molecules does not matter in first order⁸.

In conclusion, for a sufficiently precise formula for the decay of molecular tritium near the endpoint, one can assume $P_0 = 1$, $P_j = 0 \quad \forall j \geq 1$ and $V_j = 0 \quad \forall j$ yielding

$$\begin{aligned} \frac{d\Gamma}{dE} = & C \cdot F(E, Z') \cdot p_e \cdot (E + m_e) \cdot \sqrt{(E + m_e)^2 - m_e^2} \cdot \\ & \cdot (E_0 - E) \cdot \sqrt{(E_0 - E)^2 - m_\beta^2} \end{aligned} \quad (1.65)$$

In the next chapter I will describe the measuring principle and the experimental setup of the next generation KATRIN experiment for neutrino mass determination.

⁸Shift of the mean rotational-vibrational excitation of HT with respect to T₂ is compensated by a corresponding change of the nuclear recoil energy of HT with respect to the 1.5 times heavier T₂ molecule.

Chapter 2

The KATRIN experiment

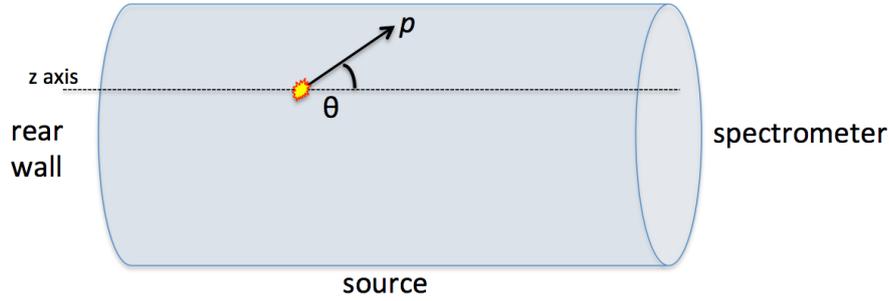
The **K**ARlsruhe **T**RITium **N**eutrino experiment (KATRIN) is a large-scale tritium β -decay experiment, whose main goal is to determine the effective electron anti-neutrino mass, as defined in (1.62) in a model-independent way. It uses high precision spectroscopy of the β electrons produced in the decay close to the endpoint E_0 . After 3 years of data taking, KATRIN will reach a neutrino mass sensitivity of 200 meV (90% C.L.), improving current laboratory limits (1.37) by one order of magnitude.

In this chapter, I will give an introduction to the measurement principle of KATRIN and an overview on the whole experimental setup. I will conclude with a discussion about sources of statistical and systematic uncertainties. More details about KATRIN, complementing the outline given in this chapter, can be found in the original KATRIN design report [KAT05].

2.1 Measurement principle: MAC-E filter

A finite effective electron antineutrino mass reduces the endpoint of the spectrum by a tiny amount (\sim eV). In order to detect this tiny distortion, the spectrometer has to act as an electrostatic filter transmitting only the electrons lying in the region of interest. The filtering technique used in KATRIN is the so called **Magnetic Adiabatic Collimation with Electrostatic filtering**, shortly **MAC-E filtering**. The basis of a MAC-E filter is an axially symmetric magnetic guidance field reaching from the source to the detector, created by multiple sequentially arranged superconducting magnets. In this section, I want to describe the physics of a MAC-E filter. I will give a more quantitative description in the next section.

The MAC-E filter is for the β electrons produced in the source what a hill is for rolling balls: a barrier they overcome only if they have enough kinetic energy. This comparison is helpful, but obviously not quite accurate.

Figure 2.1: Pictorial view of a β -decay in the source.

Magnetic guidance

A β electron is produced at a random point in the source, which is a 10 m long tube, with a random polar angle θ and a random kinetic energy¹ E . I will denote these quantities without any subscript s , assuming that the notation E, θ means E, θ in the source at generation time.

The polar angle θ is defined as the angle between the momentum of the electron \mathbf{p} and the tube axis², which I set as the z axis of our coordinate system

$$\cos \theta = \frac{\mathbf{p} \cdot \hat{\mathbf{z}}}{|\mathbf{p}|} \quad (\hat{\mathbf{z}} \text{ is the unit vector generating the } z \text{ axis}). \quad (2.1)$$

The kinetic energy is related to the momentum via the non-relativistic³ relation $E = \frac{\mathbf{p}^2}{2m_e}$.

Splitting the momentum into a parallel and a transversal component $\mathbf{p} = \mathbf{p}_{\parallel} + \mathbf{p}_{\perp}$ ⁴, I can also formally split the kinetic energy into parallel and transversal kinetic energy⁵:

$$E = E_{\parallel}(\theta) + E_{\perp}(\theta) = E \cos^2 \theta + E \sin^2 \theta = \frac{\mathbf{p}_{\parallel}^2}{2m_e} + \frac{\mathbf{p}_{\perp}^2}{2m_e} \quad (2.2)$$

The source is set to a uniform magnetic field B_s . Therefore, a β electron experiences a Lorentz force $\mathbf{F} = q(\mathbf{v}_{\perp} \times \mathbf{B}_s)$, where q is the charge of the electron and \mathbf{v}_{\perp} is the transversal component of its velocity $\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}$, that guides it through the source towards the spectrometer⁶ in cyclotron

¹Following the distribution given by the β spectrum.

²or the magnetic field lines, which in this case are parallel to the tube axis

³We are in a non-relativistic regime since the electron mass (~ 500 keV) is much higher than the highest allowed kinetic energy E_0

⁴ $\mathbf{p}_{\parallel} = |\mathbf{p}| \cos \theta \hat{\mathbf{z}}$

⁵energy is a scalar, but there is no problem in doing this

⁶or towards the rear wall

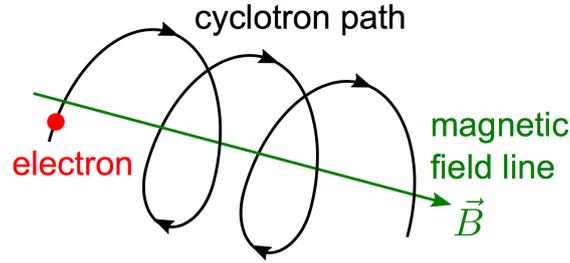


Figure 2.2: Cyclotron motion of a β electron in the source. Figure from [Gro15]

motion with radius⁷

$$r = \frac{m|\mathbf{v}_\perp|}{qB_s} = \frac{\sqrt{2m_e E_\perp}}{qB_s}. \quad (2.3)$$

Electrostatic filtering

The β electrons that make their way out of the source will eventually enter the spectrometer. Since we are interested in the endpoint region of the spectrum, only the electrons with kinetic energy in this region have to be selected. This is achieved by setting the spectrometer vessel on negative high voltage whose value is about -18.6 kV at the center of the spectrometer. This creates a symmetric potential barrier with its maximum in the center of the spectrometer, the so called **analyzing plane**. As the electric field is aligned parallel to the magnetic field lines, only the longitudinal energy E_\parallel of the electrons can be filtered. Considering that only $\sim 10^{-13}$ of the β electrons lie in the region $[E_0 - 1\text{eV}, E_0]$ and that nearly all of them have a significant fraction of their kinetic energy in the transversal component, one has to find a way to transform as much transversal energy as possible into parallel energy, i.e. reduce the polar angle in the analyzing plane as much as possible, to have the highest possible statistics. The reason one uses the electrostatic filtering it to perform an integral measurement of the spectrum, that is, a measure of counts per second at various voltages.

Magnetic Adiabatic Collimation

The solution to this problem is provided by the magnetic adiabatic collimation technique, which is implemented in the KATRIN setup.

Keeping small the magnetic and electric field gradients within a cyclotron length of the electron

⁷this radius is small enough. It is at most $\sim 10^{-4}$ m, while the diameter of the source tube is 90 mm

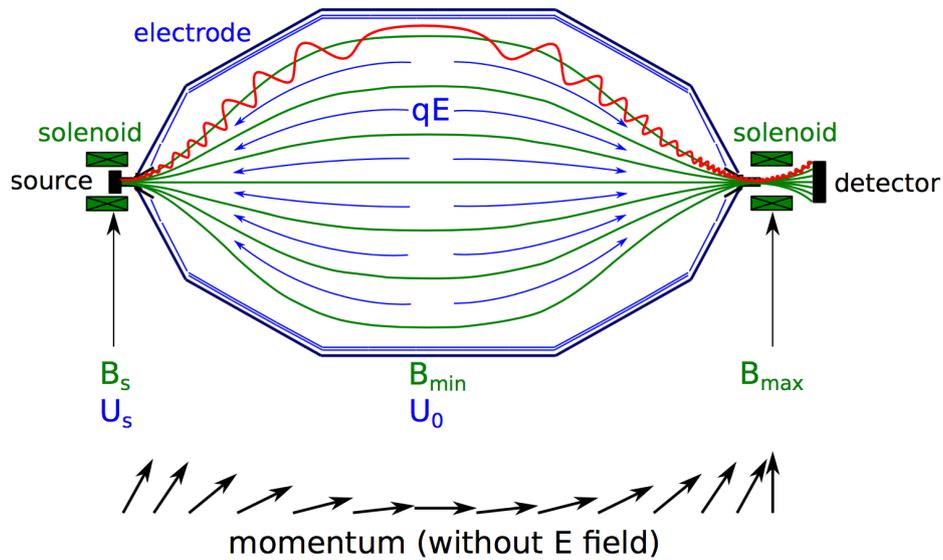


Figure 2.3: MAC-E filter principle. Superconducting magnets (green) produce a guiding field which is minimal in the center of the spectrometer. The magnetic field gradient transforms the electrons (red) transversal momentum into longitudinal momentum as indicated by the momentum arrows in the lower part of the figure, as the magnetic moment remains constant along the trajectory that is shown in exaggerated size. The electric field (blue) acts on the longitudinal energy E_{\parallel} only, filtering out those electrons with $E_{\parallel} < qU_0$. Figure from [Wan13].

$$\frac{\Delta B}{B} \ll 1, \quad \frac{\Delta E}{E} \ll 1, \quad (2.4)$$

one can be sure that the magnetic moment of the electron is constant and given by

$$\begin{aligned} \mu &:= I \cdot S \\ &= \frac{qv_{\perp}}{2\pi r} \cdot \pi r^2 \\ &= \frac{qv_{\perp}r}{2} \quad [\text{using eq. (2.3)}]. \\ &= \frac{mv_{\perp}^2}{2B} \\ &= \frac{E_{\perp}}{B} = \text{const} \end{aligned} \quad (2.5)$$

in non-relativistic approximation. The constraint in equation (2.4) is called **adiabatic hypothesis**. It ensures that there is no abrupt change in the magnetic (or electric) field within a cyclotron length, which would lead to a chaotic motion of the electron in the spectrometer. The requirement of adiabatic motion of the electrons constrains the length of the spectrometer.

According to equation (2.5), dropping the magnitude of the magnetic field, the transversal kinetic energy will be transformed into parallel kinetic energy, because of energy conservation. As this conversion takes place at the same time as the electrostatic potential reduces the longitudinal kinetic energy component E_{\parallel} , the interplay between magnetic field and electric potential needs to be adjusted and optimized so that electrons confront the electrostatic barrier after the conversion $E_{\perp} \rightarrow E_{\parallel}$ has already occurred. All the electrons that have not enough kinetic energy in their longitudinal component are back reflected in the direction of the rear wall.

Magnetic mirroring

If on one hand β electrons are reflected back by the retarding potential in the spectrometer, on the other hand they can be reflected magnetically even before in what is known as **magnetic mirroring**. Let us briefly describe it. In the KATRIN setup between the source and the spectrometer, there is a magnet, called **pinch magnet**, which generates a magnetic field of magnitude $B_p \gtrsim B_s$, where B_s is the magnitude of the magnetic field in the source. According to equation (2.5) one has

$$\frac{E_{\perp,p}(\theta_p)}{B_p} = \frac{E_{\perp}(\theta)}{B_s}, \quad (2.6)$$

or

$$\frac{\sin^2(\theta_p)}{B_p} = \frac{\sin^2(\theta)}{B_s}, \quad (2.7)$$

where p stands for **pinch magnet** (or any **point**). An electron is back reflected at the point p , if $\theta_p = 90^\circ$ and transmitted if $\theta_p < 90^\circ$, regardless of the initial kinetic energy E . Thus, using equation (2.7), an electron is transmitted beyond the pinch magnet if

$$\theta < \theta_{max} := \arcsin \sqrt{\frac{B_s}{B_p}}. \quad (2.8)$$

The angle θ_{max} is called **acceptance angle**. In nominal KATRIN mode, one has $B_s = 3.6$ T and $B_p = 6$ T at the pinch magnet so that $\theta_{max} = 51^\circ$. The magnetic field at the pinch magnet is set such that one gets this value. It might seem contradictory that one lowers the statistics of two thirds. The reason to do this in KATRIN is because electrons starting with a large polar angle travel a longer path (due to cyclotron motion) through the source, increasing their synchrotron losses and the scattering probability with the gaseous source, enhancing then the systematic uncertainties.⁸

Energy resolution

Ideally, one would set B_a to zero to transform E_\perp completely into E_\parallel at the analyzing plane. However, this is not possible for a simple technical reason. The conservation of magnetic flux imposes the further constraint $B_s \cdot A_s = B_a \cdot A_a$, where A_i is the cross sectional area at point $i = s, a$, and A_a cannot be infinitely large. The conservation of magnetic flux sets then the radius of our spectrometer. This also implies that, unless a β electron is generated with no polar angle, there is always a part of its kinetic energy left in the transversal component at the analyzing plane, which cannot be analyzed. Therefore, some electrons, which were born with a kinetic energy in our region of interest, need a surplus energy, which depends on the polar angle, to overcome the potential barrier. In other words, one cannot *resolve* all the interesting electrons, because some of them are back reflected due to their non-zero transversal energy component at the analyzing plane. The technical challenge is then to have the best $E_\perp \rightarrow E_\parallel$ conversion possible, or, better, to have the highest possible **energy resolution**.

Let us illustrate how this works.

After the pinch magnet, if $\theta < \theta_{max}$, a β electron enters the spectrometer with kinetic energy E and polar angle θ_p . Here the electrical potential U is

⁸ Another reason is that, letting only the electrons with polar angles less than 51° entering the spectrometer, in the worst case scenario one has $E_\perp = E \sin \theta_{max}$ at the entrance of the spectrometer and it is easier to convert this energy into longitudinal, instead of the whole E .

non zero⁹ and it decreases from 0 at the entrance down to $U_a \lesssim -18.6$ kV at the analyzing plane. The variable U_a is very important in KATRIN because one performs an integral measurement of the spectrum at different retarding potential U_a .

This electron is transmitted through the analyzing plane if

$$E_{\parallel,a} \geq 0. \quad (2.9)$$

Considering the conservation of total energy $E = E_a + qU_a$ one can write

$$E_{\parallel,a} \geq 0 \Leftrightarrow E \geq qU_a + E_{\perp,a} \quad (2.10)$$

From this equation one can see that the surplus energy our electron needs is exactly $E_{\perp,a}$. Using the conservation of magnetic moment for adiabatic motion one can express it as

$$E_{\perp,a} = E \sin^2 \theta_a = E \frac{B_a}{B_p} \sin^2 \theta_p = E \frac{B_a}{B_s} \sin^2 \theta. \quad (2.11)$$

This surplus depends then on the initial kinetic energy and polar angle. In the worst case, an electron needs the highest surplus energy ΔE , called **energy resolution**, when it has the maximal kinetic energy E_0 all distributed in the transversal component ($\theta_p = 90^\circ$)

$$\Delta E := \max_{E,\theta} E_{\perp,a} = E_0 \frac{B_a}{B_p}. \quad (2.12)$$

In normal KATRIN mode one has $\Delta E = 0.93$ eV.

Combining the transmission condition (2.10) with equation (2.11), one can derive useful information:

- if one fixes the kinetic energy E , then the transmission condition turns into a constrain on the initial polar angle, namely

$$\sin^2 \theta \leq \frac{B_s}{B_a} \left(\frac{E - qU_a}{E} \right). \quad (2.13)$$

or, when it makes sense¹⁰,

$$\theta \leq \theta_{\text{tr}} := \arcsin \sqrt{\frac{B_s}{B_a} \left(\frac{E - qU_a}{E} \right)}. \quad (2.14)$$

An electron starting with energy E is transmitted at the analyzing plane only if $\theta \leq \theta_{\text{tr}}$.

⁹Whereas it is set to zero in the source and at the pinch magnet.

¹⁰basically always in the case of KATRIN, since $E \gtrsim qU_a$

- if one fixes the initial polar angle θ (or equivalently θ_p), then the transmission condition turns into a constrain on the initial kinetic energy, namely

$$E \geq E_{\text{tr}} := \frac{qU_a}{\left(1 - \frac{B_a}{B_p} \sin^2 \theta_p\right)} = \frac{qU_a}{\left(1 - \frac{B_a}{B_s} \sin^2 \theta\right)} \quad (2.15)$$

This formula shows precisely the meaning of the energy resolution. In fact, for an electron generated with maximal polar angle ($\theta = \theta_{\text{max}}$, i.e. $\theta_p = 90^\circ$), one has that

$$E_{\text{tr}} = \frac{qU_a}{\left(1 - \frac{B_a}{B_p}\right)} \approx \left(1 + \frac{B_a}{B_p}\right) qU_a \approx qU_a + \Delta E, \quad (2.16)$$

which shows that the energy resolution is exactly the surplus energy needed to overcome the potential barrier.

Transmission function

Following [Gro15] I want to compute here the **transmission function** which I will use in the following and in section 2.3. It is defined as the probability that an electron, starting with a certain initial kinetic energy E and polar angle θ , passes the analyzing plane set at a potential U_a ¹¹.

If one fixes E and θ , then it is just

$$T(U_a) := P(E > E_{\text{tr}}(U_a)) = \Theta(E - E_{\text{tr}}(U_a)) = \begin{cases} 1 & E > E_{\text{tr}}(U_a) \\ 0 & E < E_{\text{tr}}(U_a) \end{cases}. \quad (2.17)$$

If instead the initial kinetic energy follows a distribution whose density is $g(E)$, then one has

$$T(U_a) = \int_{E_{\text{tr}}(U_a)}^{\infty} g(E) dE. \quad (2.18)$$

Finally, if also the polar angle is randomly distributed with density $\omega(\theta)$, one has

$$T(U_a) = \int_{E_{\text{tr}}(U_a)}^{\infty} \underbrace{\left(\int_0^{\theta_{\text{tr}}(E, qU_a)} \omega(\theta) d\theta \right)}_{=: t(E, qU_a)} g(E) dE. \quad (2.19)$$

¹¹or equivalently as the fraction of electrons starting with a certain initial kinetic energy E and polar angle θ , that passes the analyzing plane set at a potential U_a

The inner integral can then be seen as a weighting factor for each single energy E of the energy distribution $g(E)$. The function $t(E, qU_a)^{12}$ is also referred to as transmission function, but depending on the initial energy¹³.

Let us consider now our concrete case. The angular distribution in the source is assumed to be isotropic, which implies

$$\omega(\theta)d\theta = \sin\theta d\theta. \quad (2.20)$$

The transmission function becomes then

$$\begin{aligned} t(E, qU_a) &= 1 - \cos(\theta_{\text{tr}}(E, qU_a)) \\ &\quad [\text{using eq.(2.14) and } \cos(\arcsin \sqrt{x}) = \sqrt{1-x}] \\ &= 1 - \sqrt{1 - \left(\frac{E - qU_a}{E}\right) \frac{B_s}{B_a}}. \end{aligned} \quad (2.21)$$

As the term in the square root can be negative or larger than one, both implying unphysical results, three cases must be distinguished:

$$t(E, qU_a) = \begin{cases} 0 & E - qU_a < 0 \\ 1 - \sqrt{1 - \left(\frac{E - qU_a}{E}\right) \frac{B_s}{B_a}} & 0 < E - qU_a < E \frac{B_a}{B_s} \\ 1 & E - qU_a > E \frac{B_a}{B_s} \end{cases}. \quad (2.22)$$

It is zero, if the surplus energy of the electrons is negative, as one might expect. As soon as the starting energy is as large as the retarding potential, electrons with zero polar angles are transmitted. Electrons with larger polar angles are transmitted when the surplus energy increases, until finally all electrons are transmitted. This will happen once the surplus energy is larger than the energy resolution (at energy E), which is $\Delta E(E) = E \frac{B_a}{B_p}$. Considering that $B_p > B_s$, electrons starting with a large polar angle in the source will not reach the detector, as they are reflected magnetically. not transmitted. Therefore, the transmission function stops rising after all electrons with polar angles below θ_{max} are transmitted, which happens at a surplus energy of $\Delta E(E)$. A normalization can then be applied to make sure that the transmission function is a cumulative distribution function (i.e. a probability), resulting in

¹²note that actually $t(E, qU_a) \equiv t(E - qU_a)$ and $\theta_{\text{tr}}(E, qU_a) \equiv \theta_{\text{tr}}(E - qU_a)$

¹³in other KATRIN theses it is called still T

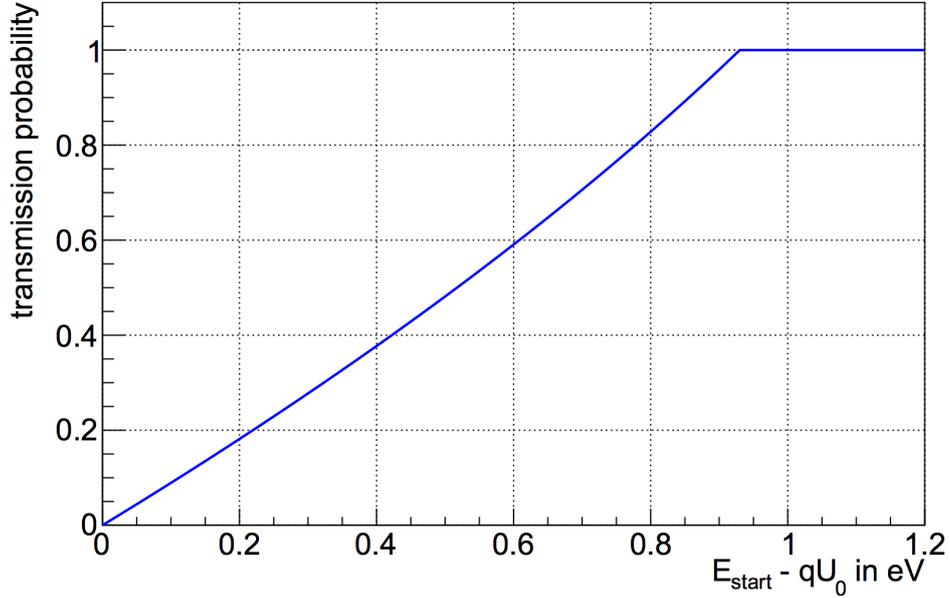


Figure 2.4: Transmission function for an isotropic source and the KATRIN design values of $U_a = 18.6$ kV, $B_s = 3.6$ T, $B_a = 3 \cdot 10^{-4}$ T and $B_p = 6$ T. The transmission starts at a surplus energy of $E_{\text{start}} - qU_a = 0$ eV. For larger surplus energies accordingly electrons with larger starting angles are transmitted until the full transmission is reached at a surplus energy of $E = 0.93$ eV. Figure from [Gro15]

$$t(E, qU_a) = \begin{cases} 0 & E - qU_a < 0 \\ \frac{1 - \sqrt{1 - \left(\frac{E - qU_a}{E}\right) \frac{B_s}{B_a}}}{1 - \sqrt{1 - \frac{B_s}{B_p}}} & 0 < E - qU_a < \Delta E(E) \\ 1 & E - qU_a > \Delta E(E) \end{cases} \quad (2.23)$$

For more details about the transmission function see [Gro15].

Response function

The transmission probability as a function of starting energy for a given retarding potential is given by $t(E, qU_a)$. However, this simple description of is only valid to first order, as multiple effects may change the transmission probability of electrons.

First, the electric potential has a radial dependency at the analyzing plane. Second, the magnetic field in the source is not constant. Consequently, the maximal polar starting angle is also not constant over the source region, which leads to the necessity of summing up multiple transmission functions and weighting them according to the corresponding tritium density. Third, the acceleration of charged particles in electromagnetic fields leads to energy loss due to synchrotron radiation. Although the maximal radiative energy losses are less than 100 meV (~ 12 meV in the source, ~ 130 meV in the transport section), the shape of the transmission function gets modified significantly. All these modifications and their impacts are deeply discussed in [Gro15].

Finally, and most importantly, the β electrons scatter off tritium molecules in the source, losing energy at each collision and only about 40% of all electrons reach the analyzing plane without any inelastic scattering (predominantly in the source). The scattered electrons consequently need a higher starting energy E to be transmitted.

A better approximation of the transmission probability is then the so called **response function**, which takes into account possible collisions in the source. In order to compute it precisely, one needs to know with high precision the average probability of scattering i times P_i and the energy-loss function (i.e. the probability density of losing energy ϵ in a collision), given by

$$f(\epsilon) = \frac{1}{\sigma_{\text{tot,inel}}} \cdot \frac{d\sigma_{\text{inel}}}{d\epsilon} \quad (2.24)$$

The response function is then given by

$$\begin{aligned} R(E, qU_a) &= P_0 \cdot t(E, qU_a) + \\ &P_1 \cdot (t * f)(E, qU_a) + \\ &P_2 \cdot (t * [f * f])(E, qU_a) + \\ &\dots \end{aligned} \quad (2.25)$$

where¹⁴

$$(t * G)(E, qU_a) = \int_0^E t(E - \epsilon, qU_a) G(\epsilon) d\epsilon, \quad (2.26)$$

for any function G , and

$$(f * f)(\epsilon) = \int f(\epsilon - \eta) f(\eta) d\eta. \quad (2.27)$$

¹⁴note also that $R(E, qU_a) \equiv R(E - qU_a)$

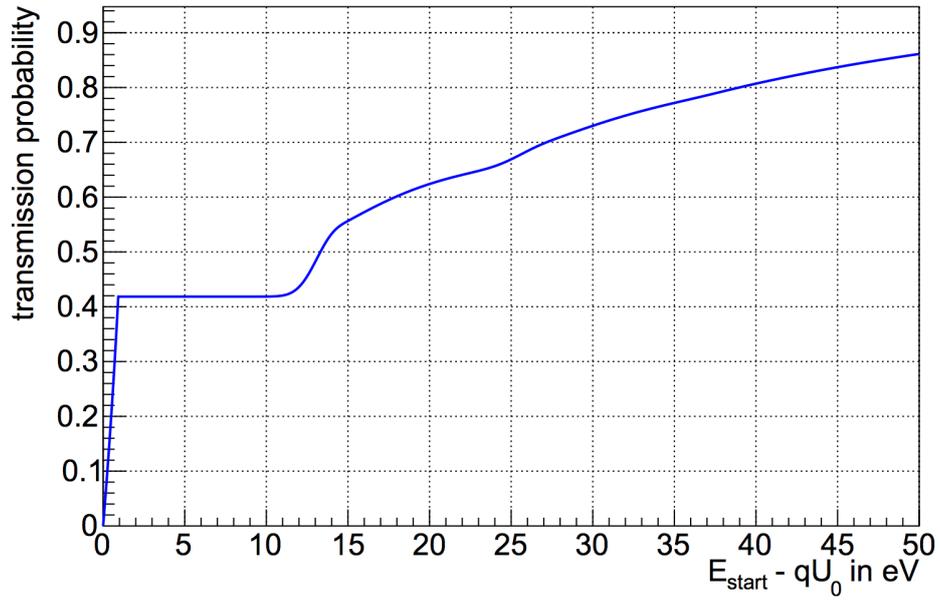


Figure 2.5: Response function as function of the surplus energy $E - qU_0$ for an isotropic source with $\theta_{max} = 50.77^\circ$ and a fixed spectrometer potential of $U_0 = 18.55$ kV. An inelastic cross section of $\sigma_{\text{tot,inel}} = 3.4 \cdot 10^{-22}$ m² and a column density of $\rho d = 5 \cdot 10^{21}$ m⁻² has been used for the calculation. As only $P_0 = 0.418$ of all electrons leave the source without experiencing any inelastic scattering on tritium molecules, the response function rises only to that level and stays at that plateau until after about 10 eV surplus energy the first scattered electrons get transmitted, as this is the minimal energy loss in inelastic scattering. Figure from [Gro15]

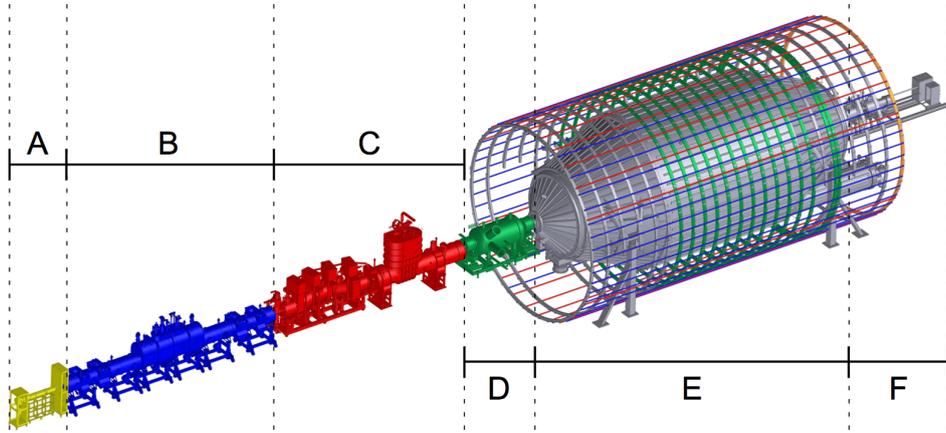


Figure 2.6: The 70 m long beamline of the KATRIN experiment. (A) The rear section, used for calibration of the experiment and monitoring of the source; (B) the WGTS, a windowless gaseous tritium source; (C) the transport section, used for differential and cryogenic pumping of tritium and magnetic guidance of the signal electrons to the spectrometers; (D) the pre-spectrometer, (possibly) used for rejection of low energy electrons; (E) the main spectrometer used for precise spectroscopy of the signal electrons; (F) detector section. Figure from [Gro15]

I will use all these tools in section 2.3 where I will discuss KATRIN sensitivity.

For a detailed calculation of the response function see [Gro15]

2.2 Experimental setup

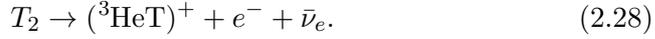
In order to detect the tiny distortion near the endpoint of the tritium spectrum due to a non-zero effective electron antineutrino mass, the KATRIN experiment implements a high luminosity (i.e. high count rate) stable molecular tritium source combined with a variable retarding potential that act as a high pass filter. Measuring the count rates at different retarding potentials the shape of the integral spectrum ($\Gamma(t) = \dot{N}(t)$) near the endpoint is determined.

In the following I will describe in some detail the main components and features of the KATRIN setup. A detailed discussion can be found in [KAT05].

2.2.1 Tritium source section: features and implementation

In the KATRIN experiment the isotope ${}^3\text{H}$ (denoted also as T) of hydrogen, known as **tritium**, is used as a β emitter in its molecular form T_2 .

The decay that takes place in the source is then



I highlight here the main reasons why molecular tritium is used as a β emitter:

- Tritium has a rather low endpoint energy of $E_0 = 18.6$ keV. Although the total number of counts per second (cps), usually called just *rate*, increases with E_0 ($\Gamma \propto E_0^4$ for low E_0 , $\Gamma \propto E_0^5$ for high E_0), the relative number of cps near the endpoint decreases with E_0 ($\propto E_0^{-3}$). Therefore a low endpoint energy is preferred [Giu07].
- Tritium β -decay is a superallowed transition between mirror nuclei¹⁵ with a relatively short half-life of about 12.3 years, which implies a high statistics with rather low source density during the experiment lifetime¹⁶ [Giu07]. Furthermore, the nuclear matrix element is energy independent near the endpoint and easily computable.
- In tritium β -decay the molecular structure is less complicated than those of heavier atoms, leading to a more accurate calculation of atomic effects. In addition, Fermi function of the daughter nucleus is also almost energy independent near the endpoint.
- Molecular tritium at low temperature can be used in gaseous form, which lower the systematic effects as compared to the usage of a condensed source. However, rotational and vibrational excitations of the daughter molecule need to be taken into account.

In KATRIN, tritium of high isotopic purity (>95%) is injected at the center into a 10 m long tube, called **Windowless Gaseous Tritium Source (WGTS)**. The tritium molecules then diffuse towards both ends of the WGTS and are pumped out at both ends by turbo-molecular pumps (TMP). The pumped-out tritium will be collected and re-injected, thus forming a closed tritium cycle. The total length of the WGTS is 16 m considering the two pumping sections at both ends that reduce the gas flow of a factor of 10^2 . The WGTS beam tube is situated in a nearly homogeneous magnetic field of $B_s = 3.6$ T, oriented in beam direction, that guides the electron towards the spectrometer (or towards the rear wall, depending on the angle).

¹⁵Superallowed transitions are allowed transitions between nuclei belonging to the same isospin multiplet. Mirror nuclei are pairs of nuclei which have equal numbers of protons and neutrons plus an extra proton in one case and an extra neutron in the other. In this case, the overlap of the initial and final nuclear wave functions is close to one, leading to a large nuclear matrix element.

¹⁶remember that $N(t) \propto e^{-t/t_{1/2}}$, so if the half-life is small, a small amount of time is needed to have a lot of decays

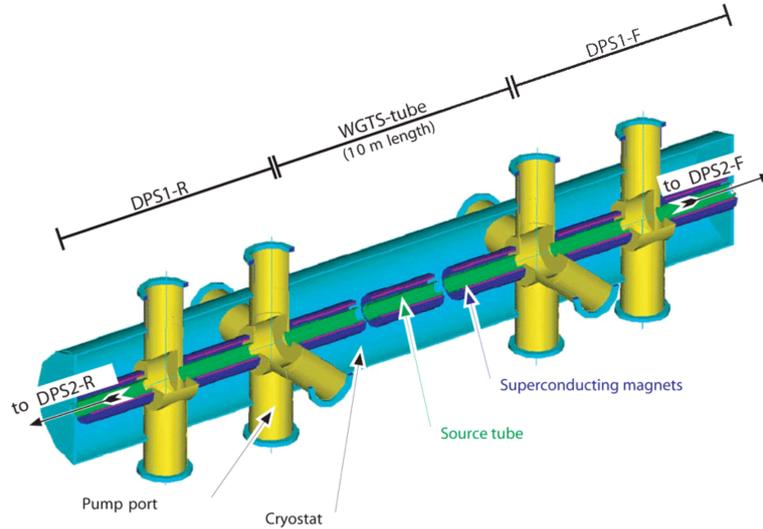


Figure 2.7: Schematic view of the WGTS, with central beam tube and subsequent pumping stations. Figure from [KAT18]

The WGTS tube is made of stainless steel and has a diameter of 90 mm. It is kept at a temperature $T = 27$ K. The low temperature assure a high column density $\rho d = 5 \cdot 10^{17} \text{ cm}^{-2}$ with a rate injection of $5 \cdot 10^{19}$ molecules/s and a smaller Doppler broadening. The main systematics of the WGTS come from the stability of the column density. Its stability depends mainly on the injection pressure (10^{-3} mbar) and the temperature, which therefore need to be kept stable with high precision. To this end, a two phase Neon cooling system is used. It was shown that the temperature variations are much smaller than 30 mK, which is necessary for a stability of the column density at the per mill level. A column density of $\rho d = 5 \cdot 10^{17} \text{ cm}^{-2}$ ensures a source activity of 10^{11} cps.

2.2.2 Rear section

At least half of all electrons from the source will leave the WGTS in backwards direction because their starting polar angle is uniformly distributed. Moreover, most of the electron emitted in forward direction will be reflected either at a magnetic field larger than the source magnetic field (e.g. pinch magnet) or at the analyzing plane. Therefore, almost all created electrons will hit the rear wall, whose task, among others, is then to monitor the tritium activity.

Further information about the rear section can be found in [Bab14].

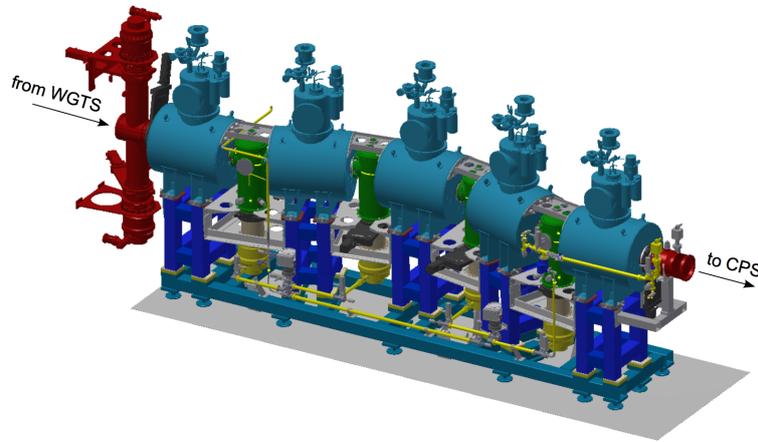


Figure 2.8: Differential pumping section. Four turbo molecular pumps (yellow) and the chicane geometry reduce the gas flow by 5 orders of magnitude, while 5 super conducting coils (turquoise) create a magnetic field of up to 5.5 T to guide the signal electrons through the beam tube (red). Figure from [Gro15]

2.2.3 Transport section

The transport section connects the WGTS and the spectrometer section. Its main task is to reduce drastically the tritium flow and guide adiabatically the β electrons to the spectrometer. First, one wants to avoid energy losses of β electrons in the spectrometer due to scattering with remaining tritiated molecules. Second, one wants to avoid background due to tritium decay inside the spectrometer. To this end the tritium flow is reduced by 12-14 orders of magnitude to achieve a high vacuum of 10^{-11} mbar in the spectrometer¹⁷.

The transport section consists of two fundamental units, the **differential pumping section (DPS)** and the **cryogenic pumping section (CPS)**, both being described in the following.

DPS

As I anticipated in 2.7, Differential Pumping Sections (DPSs) sit both at the rear section of the WGTS (DPS1-R) and at the front side (DPS1-F, DPS2-F). To block all the neutral tritium molecules, the beam lines DPS2-F cryostat is not straight, but has chicanes of 20° (see figure 2.8).

In this way the charged products are guided through the beamline, whereas neutral products hit the walls of the beam tube and are pumped out by **turbomolecular pumps (TMPs)**. This process reduces the tritium

¹⁷this is the same vacuum there is on the moon

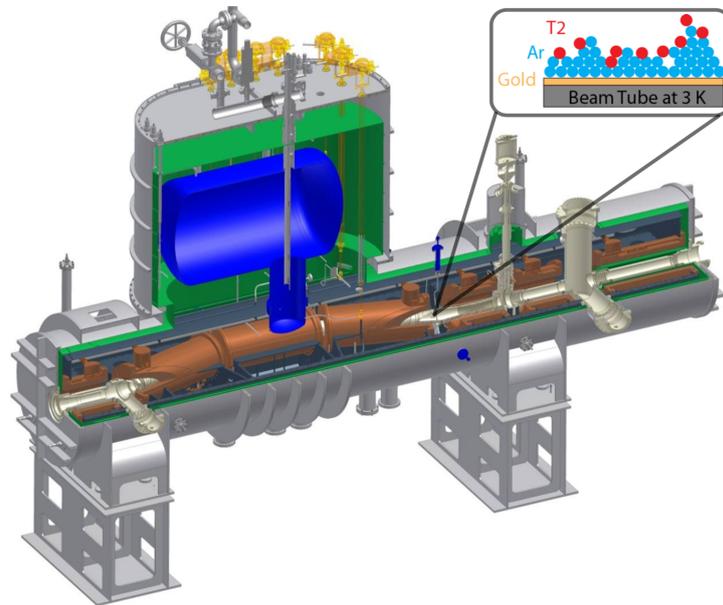


Figure 2.9: The cryogenic pumping section reduces the gas flow of tritiated molecules by 7 orders of magnitude. Tritiated molecules hit the cold beam tube surfaces covered by argon snow and are adsorbed there. The signal electrons are guided by 7 super conducting coils, producing a magnetic field of up to 5.6 T. Figure from [Gro15]

flow of about 7 order of magnitudes.

However, TMPs are not efficient for charged products of the decay (such as $({}^3\text{HeT})^+$). For this reason, a slightly more positive potential is set at the end of the DPS2-F to block the positive ions.

The DPS-F is composed out of 5 beam tubes tilted by 20° against each other to avoid a direct line-of-sight for propagating tritium molecules. Between the beam tubes 4 pump ports are situated, each containing a large turbo molecular pump.

CPS

Following the beamline, after the DPS a cryogenic pumping section is implemented. Its task is to block as much as tritiated molecules as possible. It is made of 7 beams tubes tilted in parts by 15° , to make sure that tritiated molecules will hit the inner surface multiple times. The CPS does not employ TMPs. The pumping of the CPS is based on the principle of **cryo-sorption**: a 3 K cold layer of argon frost on the inner surface of several beam tube elements to adsorb tritium molecules. This process reduces the tritium flow of 7 orders of magnitude.

2.2.4 Spectrometer section

In the spectrometer section ideally only electrons are left and analyzed according to the MAC-E filter technique explained in 2.1 by the pre- and main spectrometer.

Pre-spectrometer

The pre-spectrometer is a 3.4 m long vessel with a diameter of 1.7 m. It acts as a MAC-E filter with retarding potential of -18.3 kV and a magnetic field of 15.6 mT at its analyzing plane and a magnetic field of 4.5 T at both ends. The task of the pre-spectrometer is to offer the option to be operated as pre-filter to discard β electrons with a kinetic energy below about 18.3 keV, since these electrons do not contribute to the experiments sensitivity on the effective antineutrino mass. The flux of signal electrons entering the sensitive main spectrometer could be reduced by up to 7 orders of magnitude. However, it was shown that with this setup a large Penning trap¹⁸ would form between pre- and main spectrometer, which can be avoided by operating the pre-spectrometer at vanishing potential [Pra11].

Main Spectrometer

The main spectrometer (MS) is a 23.3 m long vessel with a diameter of 10 m operating as a MAC-E filter. Its huge dimensions are due to the energy resolution KATRIN wants to achieve: considering that the magnetic flux along the whole setup is $191 \text{ T}\cdot\text{cm}^2$, the best compromise between a good energy resolution, adiabaticity and conservation of magnetic flux is given by setting $B_a = 3 \cdot 10^{-4}$ [KAT05], which yields an energy resolution of $\Delta E = 0.93 \text{ eV}$ (see 2.1).

A 4.5 T solenoid at the entrance and a 6 T solenoid at the end, provide the magnetic guidance through the spectrometer section. These coils produce a magnetic field at the analyzing plane of 0.179 mT. This value is however too low, since it would bring the flux tube out of the MS, and corrections due to Earth magnetic field need to be taken into account. Another problem is the asymmetry of the flux tube given by the two different values of the the magnetic field at the entrance and end of the MS.

Both these problems are significantly reduced and remedied by the **air coil system**, consisting of two units: the earth magnetic field compensation

¹⁸In a Penning trap charged particles can be stored by a certain superposition of an static electric and magnetic field. The magnetic field forces the particle to move in cyclotron paths and thus prevents the particles from leaving the trap radially. The electric field confines the particle axially. The cyclotron frequency is mass dependent. Most of todays high-precision mass measurements of charged particles come from Penning trap measurements. At KATRIN however, Penning traps are a background causing effect.

system (EMCS) and the low-field coil system (LFCS). The former compensates the vertical and horizontal, non-axially symmetric components of the earth's magnetic field. The latter produces an axially symmetric magnetic guiding field, enhancing the stray field of the superconducting solenoids in the spectrometer to shape and fine-tune the flux tube.

At the center of the MS the retarding potential is varied in steps of 0.5 - 1 V close to the endpoint to perform an integral measurement of the tritium spectrum.

Monitor Spectrometer

In parallel to the main KATRIN beam line, the former Mainz spectrometer is set up in a second beam line, now acting as a monitor spectrometer. It has a length of about 4.0 m, a diameter of about 1 m, but its MAC-E filter has the same energy resolution as the main spectrometer. It is coupled with a mono-energetic conversion electron source based on $^{83\text{m}}\text{Kr}$ source as a nuclear standard. The voltage of the monitor spectrometer is directly fed by the high voltage on the main spectrometer. Thus, by scanning the mono-energetic, narrow 17.8 keV $^{83\text{m}}\text{Kr}$ (K-32) line, even small drifts of the high voltage on the ppm or even sub-ppm scale can be detected. Therefore, the stability of the retarding potential is continuously being monitored [Gou10].

2.2.5 Detector section

Once electrons overcome the analyzing plane they are guided through the pinch magnet to **the focal plane detector** (FPD), a semi-conductor based silicon PIN diode. The detector is divided into 148 pixels in order to resolve radial and angular inhomogeneities of the retarding potential at the analyzing plane and each pixel measures an independent spectrum. The detector is situated in a magnetic field of 3.6 T.

An active veto system made of plastic scintillators reduces the background coming from cosmic rays.

The detector consists of a monolithic silicon wafer of 9 mm diameter and an effective thickness of $500\ \mu\text{m}$ which counts the signal electrons with an energy resolution of about 2 keV. This resolution is sufficient for KATRIN, as the energy resolution is provided by the spectrometer section, and the detector only counts the electrons. Its value ensures that the electrons are coming from the tritium decay in the source [Har12].

An active veto system made of plastic scintillators reduces the background coming from cosmic rays.

As I will discuss later, this will not be enough for sterile neutrino search. Therefore a new detector system is needed.

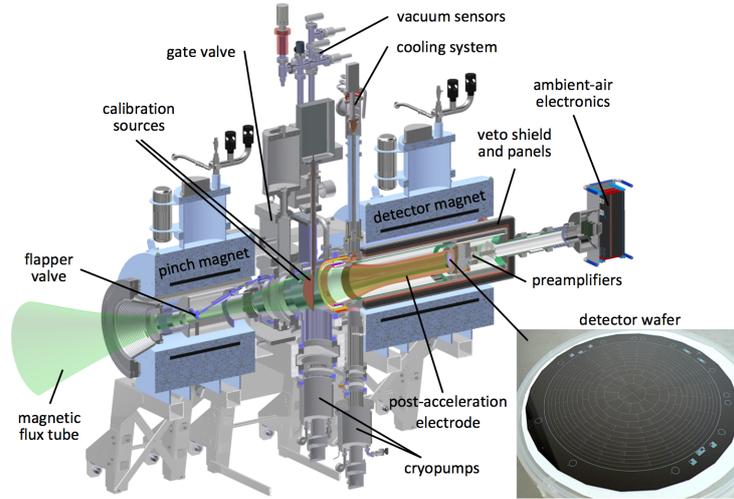


Figure 2.10: Focal plane detector system, with the silicon PIN-diode segmented into 148 pixels at the end. Figure from [Gro15]

2.3 Sensitivity of KATRIN experiment

In this section I give an overview on statistical and systematic uncertainties in KATRIN.

2.3.1 Expected statistical uncertainty

The KATRIN experiment performs an integral measurement of the tritium β spectrum measuring the convolution of the differential spectrum with the response function¹⁹

$$\begin{aligned} N_S(qU; E_0, m_\beta^2) &= N_{\text{tot}} t_U \Gamma_{\text{tot}}(qU; E_0, m_\beta^2) \\ &= N_{\text{tot}} t_U \int_{qU}^{E_0} \frac{d\Gamma}{dE}(E; E_0, m_\beta^2) R(E - qU) dE. \end{aligned} \quad (2.30)$$

$N_S(qU; E_0, m_\beta^2)$ is the number of signal electrons at retarding potential qU . N_{tot} and t_U denote the total number of tritium nuclei and the measurement time at retarding potential qU . The differential tritium spectrum

¹⁹Equation (2.30) can be written, calling $R'(x) = R(x)$ as

$$N(qU; E_0, m_\beta^2) = \left(\frac{d\Gamma}{dE} * R' \right) (qU; E_0, m_\beta^2) \quad (2.29)$$

It does not really matter whether one writes $R(E - qU)$ or $R(qU - E)$, but to stick with my convention I introduce R'

$\frac{d\Gamma}{dE}(E; E_0, m_\beta^2)$ represents the number of decays per second, per energy, per nucleus. Variables and parameters are separated by semicolons.

In addition to the signal electrons also background events will be measured. A constant background is assumed, for the fitting of the spectral shape. An overview on background sources will be given in the next section.

The number of theoretically predicted signal electrons is then

$$N_{\text{th}}(qU; E_0, m_\beta^2, R_s, R_b) = R_s \cdot N_S(qU; E_0, m_\beta^2) + R_b \cdot N_b, \quad (2.31)$$

where R_s and R_b are the relative fraction of signal and background respectively, and $N_b = \Gamma_b \cdot t_U$ is the number of background electrons, where Γ_b is the total background rate.

For the simulation of a KATRIN-like spectrum, the expected value N_{th} is randomly smeared according to a Gaussian distribution with $\sigma_{\text{th}}(qU) = \sqrt{N_S(qU) + N_b}$. The number $\sigma_{\text{th}}(qU)$ is called **statistical uncertainty** (note that it depends on the measuring time through N_b , so in general it depends on the total measuring time. The single measuring times at each retarding potential have to be optimized to minimize this statistical uncertainty). The number of expected (or experimental) electron is then

$$N_{\text{exp}}(qU) = N_S(qU; E_0, m_\beta^2) + N_b + \text{RndGauss}(\sigma_{\text{th}}(qU)). \quad (2.32)$$

The analysis of simulated spectra is based on a minimization of the function

$$\chi^2(E_0, m_\beta^2, R_s, R_b) = \sum_i \left(\frac{N_{\text{th}}(qU_i; E_0, m_\beta^2, R_s, R_b) - N_{\text{exp}}(qU_i)}{\sigma_{\text{th}}(U_i)} \right)^2, \quad (2.33)$$

by varying E_0, m_β^2, R_s and R_b independently. To avoid any bias, the fitting procedure allows for negative, i.e. unphysical values of m_β^2 . One deduces limits of confidence on m_β^2 in terms of frequency of occurrence by repeating large samples of simulated experiment-like integral β spectra.

2.3.2 Sources of systematic errors

Looking at equations (2.30) and (2.31) one can identify the main sources of systematic uncertainties:

- **Theoretical corrections to the spectrum** $\frac{d\Gamma}{dE}(E; E_0, m_\beta^2)$.

First, in the decay of molecular tritium, electronic and molecular excitations of the daughter molecule can occur. Considering that the lowest electronic excitation energy in the final state occurs at about

10 eV, this effect plays a role only for measurement intervals more than 10 eV below the endpoint (this will then be an issue for keV sterile neutrino search). The vibrational and rotational excitation energies of the daughter nucleus fall in the sub-eV regime, thus impacting the exploration of neutrino mass. A theoretical computation of the final state distribution can be found in [Dos06, Dos08].

Second, thermal motion of tritium molecules in the source induces a Doppler shift of the β electron energies. A typical value of this shift $\Delta E_{\text{doppler}} = 100$ meV. [Hoet09, Mer12].

Finally, moving away from the endpoint E_0 the Fermi function F and the nuclear matrix element both become energy dependent. Uncertainties in their theoretical description would lead to a systematic errors on the neutrino mass. This is not an issue for KATRIN, but it is for a keV sterile neutrino search.

- **Uncertainty on experimental parameters**

A major source of systematic uncertainties is the uncertainty on experimental parameters, such as the retarding potential, the column density and the magnetic field in the source.

The limit for relative allowed variation of the retarding potential is $\Delta V/V < 3$ ppm, while the variation of the potential at the WGTS can be in the per mill level.

The stability of the column density, which fixes the count rate, will be measured with an electron gun on a regular basis by measuring the response function at different surplus energies.

Finally, the magnetic field in the source has to be stable within $\Delta B_s/B_s < 2 \cdot 10^{-3}$. The reason is twofold: the transmission function and consequently the response function depend on the magnetic field. In fact, the scattering probabilities depend on the effective path of the electrons in the source, which itself depends on the cyclotron radius and hence on the magnetic field see equation (2.3).

- **Uncertainties in the response function**

The response function is one of the main source of systematic uncertainties. As I show, it encodes the energy loss probability $f(\epsilon)$. β electrons can experience energy losses in their way to the detector due to:

1. inelastic scattering in the WGTS: $\epsilon_{\text{inel}} > 13.6$ eV,
2. elastic scattering in the WGTS: $\langle \epsilon_{\text{el}} \rangle = 20$ meV ($\sigma_{\text{el}} \approx 10^{-1} \cdot \sigma_{\text{inel}}$),
3. synchrotron losses due to their cyclotron motion, which in turn depends on the magnetic field. One has [Mer12] $\epsilon_{\text{syn, max}} = 130$ meV in the transport section and $\epsilon_{\text{syn, max}} = 12$ meV in the WGTS.

2.3.3 Background sources

I give an overview on background sources in a different section because they represent a topic on their own. Possible countermeasures to these are given in [Mer12].

The main source of background originates are the spectrometer section and partly the detector section. A very low background level is essential to minimize the systematic errors. The higher the background rate, the further away from the endpoint one has to measure the spectrum in order to optimize the signal to background ratio N/N_b . However, further away from the endpoint systematics on the spectral shape take over.

To achieve the desired sensitivity (see next section) a background level of 0.01 cps is aimed for. Such a low background is needed for KATRIN because the background should be at most as much as the count rate of signal electrons (total decays per second $\sim 10^{11}$, decays per second in the region of interest $\sim 10^{-13}$).

Detector background

KATRIN detector has an energy resolution of about 2 keV, therefore all non-signal electrons in the energy range 17 - 19 keV contribute to background. These electrons are produced by: cosmic muons (and subsequent neutrons and gammas), high energetic gammas of environmental radioactivity in the surrounding area and decays of radio-nuclei in the detector material. Therefore, the detector is surrounded by a muon veto and post acceleration is used to better discriminate signal electrons.

Spectrometer background

All low energy electrons being created by several processes in the spectrometer are accelerated on their way to the FPD to tank potential and, therefore, lie in the energy region of interest, contributing significantly to background.

2.3.4 Sensitivity

In normal KATRIN mode, the quadratic sum of all known systematic uncertainties is expected to be $\sigma_{\text{sys,tot}} = 0.017 \text{ eV}^2$, where the largest contribution comes from systematic uncertainties in final states distribution. For determining the statistical errors experimental data are simulated, which take into account all relevant processes. By fitting the theoretical curve to the simulated spectrum, the neutrino mass parameter can be inferred. Repeating this procedure many times, one generates a neutrino mass distribution. The width of this distribution then corresponds to the statistical error. The measurement time of KATRIN is chosen such that the statistical

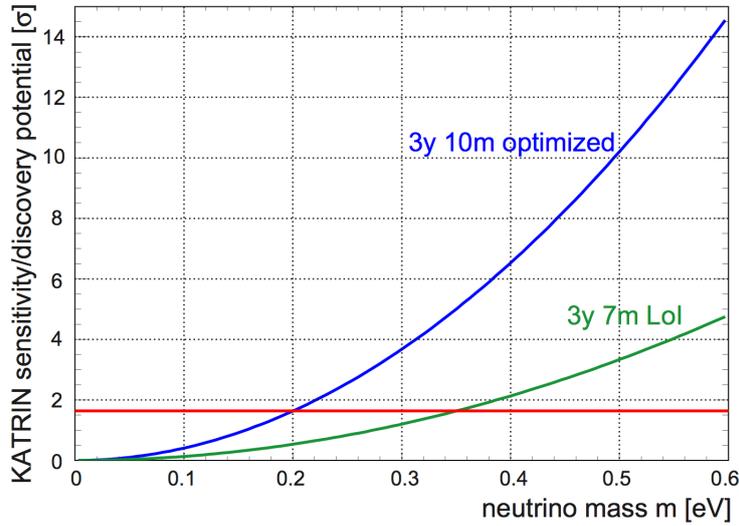


Figure 2.11: Discovery potential of the KATRIN experiment. The horizontal line represents the 1.64σ or 90% C.L. Figure from [KAT05]

error is of the same order as the systematic error. This is reached after three *full beam* years measurement time. Adding systematic and statistical error quadratically one gets a number known as **sensitivity**, i.e. a total error whose value is $\sigma_{\text{tot}} \approx 0.025 \text{ eV}^2$.

Assuming a vanishing neutrino mass, this uncertainty translates into an upper limit, which one calls $L(90\% \text{C.L.})$, which is connected to the error on m_β^2 via $L(90\% \text{ C.L.}) = \sqrt{1.64 \cdot \sigma_{\text{tot}}}$. After three years of measuring time, this limit becomes

$$m_\beta < 0.2 \text{ eV} \quad (90\% \text{C.L.}) \quad (2.34)$$

with no finite neutrino mass being observed. Within a few weeks of effective β spectrum scanning, the KATRIN statistical accuracy will exceed the accuracy of the Mainz and Troitsk experiments by an order of magnitude, see 2.11.

Figure 2.11 demonstrates the discovery potential of KATRIN as significance in units of σ_{tot} for an effective electron anti-neutrino mass in the sub-eV range. For example, a potential neutrino mass $m_\beta = 0.4 \text{ eV}$ could be seen by KATRIN with a $6.5\sigma_{\text{tot}}$ significance.

Chapter 3

Sterile Neutrinos

The SM includes both left- and right-handed part of massive fermions in order to include a generation mechanism for their mass. As I showed in the first chapter, a minimal extension of the SM that could explain the smallness of neutrino mass was given by the introduction of a (heavy) sterile right-handed neutrino state, labelled ν_s . This is a minimal extension of the SM considering that neutrinos are the only fermions lacking of a right-handed component. Such a sterile neutrino would be a singlet under all interaction in the SM and then interacting only via gravity. The state ν_s can be considered as a new "flavor" eigenstate (even if it has no flavor) which corresponds then, through oscillations, to a fourth mass eigenstate ν_4 with mass m_4 . This ν_4 is therefore a mixture of ν_e, ν_μ, ν_τ and ν_s which I will call *sterile* meaning *mostly sterile*. Therefore, when one discusses sterile neutrino detection, one actually refers to the measurement of its mass and mixing angle, given the fact that by definition one cannot detect a non-interacting particle. It is not known yet whether such a particle exists. However, there are several hypothesis on its mass scale, all motivated by experimental measurements and observation.

In the following I will illustrate the most popular hypothesis on the mass scale of a sterile neutrino, their motivation and the relation to dark matter.

3.1 eV Sterile Neutrinos

The existence of one or more eV sterile neutrinos is still today a hot topic of discussion [Aba12]. They are widely disfavored as dark matter candidates: due to the smallness of their mass they would make up *hot* dark matter (HDM) with a free streaming length of ~ 1 Mpc. In a Universe dominated by such DM¹, galaxy clusters form at an earlier stage, while galaxies are washed out and form at a later stage. This is the so called

¹as a reference, the size of a galaxy is of the order of kpc, while the size of a galaxy cluster is of the order of Mpc

top-down scenario. However, there are galaxies which are older than the cluster they sit in, therefore the top-down scenario is basically ruled out: a HDM dominated universe is not realized in nature.

The major motivations for an eV scale sterile neutrino come from short baseline oscillation experiments. For example, MiniBooNE experiment results show a low energy excess in the energy spectrum of electron neutrinos [Min07]. This anomaly can be accommodated with the addition of a light sterile neutrino that mixes through oscillations with the active neutrinos [Kop11, Men11].

The Short Baseline Neutrino program at Fermilab has among its goals a followup on this MiniBooNE anomaly and will in the near future shed light on this puzzle [Bas17].

3.2 GeV Sterile Neutrinos and Dark Matter

Particles with mass in the GeV scale and free streaming lengths of ~ 1 kpc are good candidates for the so called **cold dark matter (CDM)**. In a Universe dominated by CDM, galaxies form at an earlier stage, while galaxy clusters form at a later stage. This is the so called **bottom-up** scenario.

GeV right-handed neutrinos as described in the see-saw mechanism would be a good candidate for CDM. Other famous candidates are the hypothetical Weakly Interacting Massive ParticleS (WIMPS). The most promising WIMPS are the LSP (lightest supersymmetric particles), which are degrees of freedom of the supersymmetric extension of the SM that cannot decay into SM particle because of the conservation of R-parity [Jun96].

However, neither WIMPS nor GeV-scale right handed neutrinos have been detected yet, which leaves this scenario wide open.

3.3 keV Sterile Neutrinos and Dark Matter

The **warm dark matter** scenario is, as the name suggests, a scenario between hot and cold dark matter. Warm dark matter would be composed of particles in the keV scale and $l_{fs} \sim 10$ kpc. This mass scale is a good compromise between hot and cold dark matter and can solve issues affecting both scenarios. A keV sterile neutrino is therefore good warm dark matter candidate²: it is in the right mass range and it is "mostly" sterile. The parameters of interest are its mass m_4 and its active-sterile mixing amplitude $\sin^2 \theta_s$ (I will be more specific in the next paragraph). The main effort of cosmological observation and laboratory experiments is to constrain the allowed region of the parameter space (the plane $(m_4, \sin^2 \theta_s)$ where these two "free" parameters live (one can actually already guess that $\sin^2 \theta_s$ is very small and m_4 is in the keV range).

²it could also be cold dark matter, depending on its production mechanism.

There are at the moment very good reasons to look for a dark matter particle in the keV range. First of all, the phase-space density evolution of dwarf spheroidal satellites in the Milky Way provides a model-independent lower bound to fermionic dark matter and respectively for the sterile neutrino mass of $m_4 > 2$ keV [Hub15].

DM should be stable. However, if it is made of keV sterile neutrinos, it can theoretically decay via the channel

$$\nu_s \rightarrow \nu_\alpha + \gamma, \quad (\alpha = e, \mu, \tau). \quad (3.1)$$

This decay is loop-suppressed, so its branching ratio is very low. This decay would take place in relativistic regime, therefore the decay energy is divided equally into the two products, giving the photon an energy equal to half m_4 . Such radiation could in principle be detected as X-rays. There are two main observations which allow for such a process. First, the results of XMM-Newton and Chandra telescopes gave an upper bound for m_4 and constrained $\sin^2 2\theta_s$ [Boy12]

$$m_4 < 50 \text{ keV} \quad (3.2)$$

$$10^{-13} < \sin^2 2\theta_s < 10^{-7} \quad (3.3)$$

Second, the XMM-Newton telescope observed a weak X-ray emission line from stacked galaxy clusters, which could be the first evidence for relic sterile neutrinos. The results set the sterile mass to 7.1 keV and the active-sterile mixing amplitude to $7 \cdot 10^{-11}$ [Bul14]. The same emission line has been observed also from the Andromeda galaxy and the Perseus galaxy cluster [Boy14].

However, these astrophysical observations are strongly model-dependent and have to be firmly supported by laboratory experiments.

In conclusion, it is possible to construct a model with three sterile neutrinos, one per flavor, solving a large number of puzzles. The one with the lowest mass in the keV range (which I called m_4) would solve the dark matter puzzle. If one then allows the remaining two neutrino massive state to be in the GeV region, then one would potentially solve the open question of matter-antimatter asymmetry in the Universe and give mass to active neutrinos via see-saw mechanism.

As I said, these are interesting speculations that have to be supported by observations and experiments.

In the next section I will show how the KATRIN experiment has the potential to further constrain the current experimental bounds.

3.3.1 Signature of a keV-sterile neutrino in KATRIN

An observation before going on. When I refer to *keV sterile neutrino* I actually mean *keV sterile neutrino (the mass state) which is mostly sterile*,



Figure 3.1: Neutrino mixing angles in the 3+1 model. Figure from IceCube Collaboration [Ice18]

as one can see in figure 3.1.

With an endpoint of $E_0 = 18.6$ keV, the tritium β spectrum is a perfect tool to probe the impact of an hypothetical sterile neutrino with mass up to E_0 . Let us now show how a fourth neutrino state would impact the shape of the β spectrum. The mixing of four neutrino states is described by an extended PMNS matrix

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \\ \nu_s \end{pmatrix} = \begin{pmatrix} U_{e1}^* & U_{e2}^* & U_{e3}^* & U_{e4}^* \\ U_{\mu 1}^* & U_{\mu 2}^* & U_{\mu 3}^* & U_{\mu 4}^* \\ U_{\tau 1}^* & U_{\tau 2}^* & U_{\tau 3}^* & U_{\tau 4}^* \\ U_{s1}^* & U_{s2}^* & U_{s3}^* & U_{s4}^* \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}. \quad (3.4)$$

One can then write

$$1 = \sum_{i=1}^3 |U_{ei}|^2 + |U_{e4}|^2 := \cos^2 \theta_s + \sin^2 \theta_s, \quad (3.5)$$

where θ_s is the **active-sterile mixing angle**. It can be considered as an effecting mixing of an active neutrino with mass m_{light} and a sterile one with mass m_4 , since one is not able to resolve the three light massive states yet³.

Starting from formula (1.60) and considering a 4 flavor mixing I get

³ m_{light} is not the same as m_β in KATRIN. One might think that they are the same thing, but this is not the case because the elements of a 4-flavor mixing matrix are different from those of a 3 flavor mixing one. The confusion rises because their elements are both called $U_{\alpha i}$

$$\frac{d\Gamma}{dE} = \cos^2 \theta_s \frac{d\Gamma}{dE}(m_{\text{light}}^2) + \sin^2 \theta_s \frac{d\Gamma}{dE}(m_4^2), \quad (3.6)$$

where

$$\begin{aligned} \frac{d\Gamma}{dE}(m_{\text{light}}^2) &= C \cdot F(E, Z') \cdot p_e \cdot (E + m_e) \cdot \sqrt{(E + m_e)^2 - m_e^2} \cdot \\ &\quad \cdot (E_0 - E) \cdot \sqrt{(E_0 - E)^2 - m_{\text{light}}^2} \end{aligned} \quad (3.7)$$

(which has the same form of equation (1.65)) and

$$\begin{aligned} \frac{d\Gamma}{dE}(m_4^2) &= C \cdot F(E, Z') \cdot p_e \cdot (E + m_e) \cdot \sqrt{(E + m_e)^2 - m_e^2} \cdot \\ &\quad \cdot (E_0 - E) \cdot \sqrt{(E_0 - E)^2 - m_4^2}. \end{aligned} \quad (3.8)$$

The spectrum of a decay into the sterile neutrino (3.8) would have a lower endpoint, precisely $E_0 - m_4$ and a really tiny amplitude $\sin^2 \theta_s$.

Therefore, the imprint of a sterile neutrino on the entire β spectrum shows up as a kink-like signature at $E_0 - m_4$ and a tiny distortion of the amplitude in the region $[0, E_0 - m_4]$ as one can see in figure 3.2.

3.3.1.1 KATRIN *as is*

In contrast to the nominal KATRIN mode, the entire tritium β -decay spectrum has to be measured to search for a keV sterile neutrino, as discussed in the previous section. To enable a full spectrum measurement with KATRIN major parameter and design changes are required.

The first approach to search for keV sterile neutrino is to use the KATRIN setup *as is* tuning some parameter in a convenient way. I refer to the measurement phase using the current KATRIN setup as **Phase-0**.

The Phase-0 measurement, which will most likely take place before the neutrino mass measurement, is characterized by a low statistics because the present KATRIN detector cannot handle the total decay rate from the WGTS.

Here I will discuss the reduction of signal count rate and optimization of the magnetic field configuration.

Reduction of signal count rate

The first idea would be to just turn off the retarding potential, such that all β electrons get to the detector. However, the KATRIN FPD, as well as the readout system, are only designed to handle a maximum rate of 10^6 cps, which is 4 orders of magnitude lower than the source activity in normal KATRIN mode.

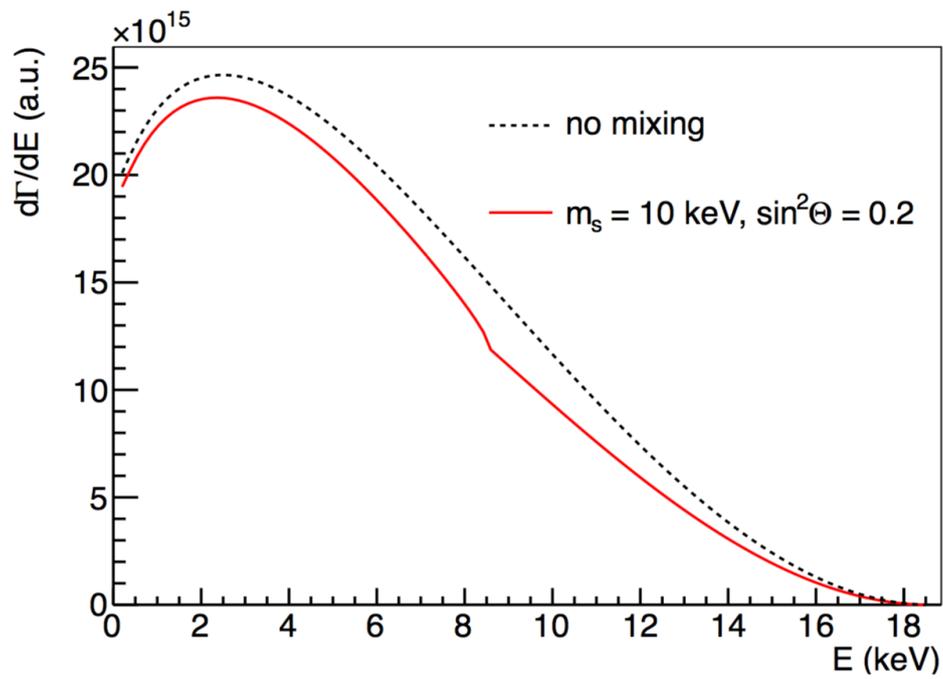


Figure 3.2: The differential tritium β spectrum without mixing with a fourth neutrino state (dashed curve) and with a 10keV sterile neutrino with unphysical mixing angle of $\sin^2\theta_s = 0.2$ (red curve). Figure from [Mer14].

The reduction of the count rate is achieved in two ways.

First, the column density in the WGTS can be lowered. A lower gas density reduces the number of scatterings of β electrons in the source. However, at very low gas densities the source stability can no longer be guaranteed. A good compromise is a reduction of 2 orders of magnitude of the column density which ensures the same stability performance as for the full column density.

Second, the acceptance angle given by

$$\theta_{\max} = \arcsin \sqrt{\frac{B_s}{B_p}}. \quad (3.9)$$

can be reduced.

In this way the electrons that start with a pitch angle smaller than θ_{\max} reach the detector while the remaining electrons (in this case almost all!) are magnetically backreflected and fly backwards to the rear wall. However, the source magnetic field is constrained by the requirement of adiabatic guidance through the entire KATRIN setup. A viable value, used as a benchmark point for the studies of this work is $B_s = 0.045$ T which yields an acceptance angle θ_{\max} of $\sim 5^\circ$.

parameter	benchmark value
B_{source}	0.045 T
θ_{\max}	4.97°
ρd	$3 \cdot 10^{15} \text{ cm}^{-2}$
ρd reduction factor	167
rate at detector	10^6 cps

Table 3.1: Benchmark values for the Phase-0 measurement.

Optimization of magnetic field configuration

The magnetic field layout needs to be optimized along the whole KATRIN setup with respect to the following aspects:

- a tiny magnetic field strength is needed at the rear wall of the experiment, in order to prevent electron backscattering *from* the rear wall to reach the detector, which would then increase background;
- as already discussed, a reduced magnetic field in the source, reduces the acceptance angle and hence the total count rate;
- a *small* source magnetic field and increased spectrometer magnetic field assure adiabatic transport **through the main spectrometer**; a *large enough* source magnetic field, on the other hand, is needed to assure adiabatic transport of electrons **through the WGTS**;

- the detector magnetic field is chosen to match source magnetic field, in order to **fully map the source onto the detector**.

On the road to Phase-0 measurement

Besides the study of the optimal magnetic field configuration, extensive studies have been carried out on the major systematic errors and on the possible measurement modes (differential, exploiting the KATRIN detector energy resolution, and integral, as in normal KATRIN mode) [Mer14].

The major steps towards a Phase-0 measurement will be the investigation of the source stability and adiabaticity of the electron transport along the beamline during the next STSIIIa measurement campaign (\sim first months of 2018).

On the software side the major challenge is the development of a new model capable of simulating the entire tritium spectrum, including as many experimental effects as possible, that will be used for the fitting of the data.

The development and description of such a software are the major goals of this thesis.

3.3.1.2 The TRISTAN project

As the name suggests, Phase-0 will be followed by a **Phase-1** measurement. By Phase-1 I refer to the realization of a new detector and readout system capable of handling the full rate coming from the WGTS in normal KATRIN mode.

The measurement campaign with the **TRISTAN** (**TRITium Beta Decay to Search for STerile (A?) Neutrinos**) detector is planned to take place after the neutrino mass measurement.

The final TRISTAN detector is designed to have ~ 20 cm diameter and 21 modules. Each module should have 168 pixel and be ~ 4 cm. Each pixel has ~ 3 mm diameter.

At the present moment, a number of 7-pixel silicon drift detector system have been realized by the Halbleiterlabor of the Max Planck Society. Currently, they provide two running systems, one equipped with ASIC from CEA Saclay and one with ASIC from XGLab.

The performance and features of these prototypes are currently under investigation.

Chapter 4

Modeling of the full tritium spectrum

The goal of this final chapter is to present a new simulation software for the modelling of the entire tritium spectrum. In section 4.1 I will shortly describe the current model used by the KATRIN collaboration to compute the integral tritium spectrum after the source, *in the last 30 eV*. In section 4.2 I will describe in detail the new simulation software called **Convolution code** developed by Martin Slezák and Dr. David Radford. The goal of this software is the implementation of source related effects for the precise modelling of the *entire* tritium β -spectrum. In section 4.3 I will discuss the comparison of SSC and the Convolution code in the last eV's of the spectrum: this comparison is essential to test the validity of the new model. In section 4.4 I will present the concept of how to integrate the results in a more general simulation framework, which makes use of the Convolution code, for a keV-scale sterile neutrino search with KATRIN.

4.1 SSC

The software framework developed and utilized in KATRIN for most simulations and analysis studies is called KASPER. It is a software written in C++ and it is made up of several modules that can be used individually, nevertheless, thanks to the cmake build system, they are forced to remain compatible and link correctly together.

KASPER has been initially developed by Kaefer and M. Hoetzel and later widely upgraded and extended by Groh and Kleesiek over the past ten years [Gro15, Kle14]. The only relevant KASPER module that is used in this thesis is the **Source Spectrum Calculation** module, shortly **SSC**.

The SSC module of KASPER allows the computation of the integral tritium spectrum **in the vicinity of its endpoint**¹.

¹it is very important to stress this fact, because it implies that many parameters (recoil

```

<include base="fields.xml" />
<ssc>
  <TemperatureAsymmetric Name="myTemperatureAsymmetric"
    T0="29.0"
    DeltaTRear="0.5"
    DeltaTFront="-0.5"
    LengthRear="5.0745"
    LengthFront="5.0075" />

  <TemperatureAsymmetric Name="myTemperatureAsymmetricHighTemp"
    T0="110.0"
    DeltaTRear="0.5"
    DeltaTFront="-0.5"
    LengthRear="5.0745"
    LengthFront="5.0075" />

  <DensityAsymmetric Name="myDensityAsymmetric"
    InjectionPressure="0.3368"
    RatioPRear="0.05"
    RatioPFront="0.05"
    LengthRear="5.0745"
    LengthFront="5.0075" />

  <DensityAsymmetric Name="myDensityAsymmetricFixedCD"
    ColumnDensity="[cd:5E21]"
    RatioPRear="0.05"
    RatioPFront="0.05"
    LengthRear="5.0745"
    LengthFront="5.0075" />

  <DensityAsymmetric Name="myDensityAsymmetricHighTemp"
    ColumnDensity="[cd:5E21]"
    RatioPRear="0.0"
    RatioPFront="0.0"
    LengthRear="5.0745"
    LengthFront="5.0075"
    Viscosity="4.033E-6" />

```

Figure 4.1: Structure of the configuration file for SSC.

A simulation run with SSC is completely defined by a configuration file where all simulation input data are defined and created. This includes, among others, the magnetic field configuration of the source, gas dynamics properties and the spatial resolution of the flux tube. The configuration file for SSC is based on the Extensible Markup Language (XML) 4.1.

In chapter 2 I described *analytically* the transmission and response function, providing an exact formula for both. However, they are only an (accurate) approximation: non-uniformity of the column density and magnetic field's inhomogeneities are, for example, key factors that change the shape of transmission and response function significantly [Gro15]. More sophisticated versions of the transmission and response function can be implemented through the xml configuration file used in SSC.

It is important to stress that the integral spectrum computed with SSC

energy of the nucleus, cross sections, Fermi function, radiative corrections, etc...) are energy independent. This will not be the case for the full tritium spectrum calculation.

is **not** the result of a Monte Carlo simulation. Its instead the "exact" computation of the convolution of the differential spectrum with the response function:

$$\int_{qU}^{E_0} \frac{d\Gamma}{dE} R(E; qU) dE. \quad (4.1)$$

The differential spectrum in SSC is computed exactly according to the formula (1.65). In order to have a more precise description in the endpoint region further corrections have to be taken into account:

- **Fermi function**

It describes the electric interaction between the daughter nucleus and the outgoing electron with energy E . In the decay, the outgoing electron is attracted by the nucleus, thus shifting the endpoint to slightly lower energies.

- **Radiative corrections**

An electron emitted within the Coulomb field of a nucleus experiences energy losses due to their interaction with virtual and real photons.

- **Nuclear recoil**

The daughter molecule has not an infinite mass. Therefore, after the decay it will gain a small amount of kinetic energy, subtracting it from the beta electron. In the vicinity of the endpoint, this energy is nearly constant (~ 1.7 eV).

- **Final State Distribution**

Contrarily to an atomic decay, in the molecular decay of tritium the daughter molecule can find itself being in a rotational and/or vibrational excited states. Each of these final states come with its probability and will accordingly reduce the maximum kinetic energy of the emitted electron, modifying then the spectrum. The resulting spectrum is thus a superposition of a large number of single branches, each one with a lower endpoint, weighted by the final states probability. From the practical point of view, the source will also incorporate small amounts of the tritiated hydrogen isotopologues DT and HT, whose corresponding final state distributions will have to be taken into account.

For a more detailed description of SSC see [Gro15, Kle14].

4.2 Convolution code

The KASPER framework along with the SSC module provides a very sophisticated tool for simulations of the whole KATRIN apparatus, and in particular of the integral tritium spectrum, **near the endpoint of the beta spectrum**. This is motivated by the fact that, as discussed, the spectral distortion due to neutrino mass is maximal near the endpoint.

Tritium decay in the KATRIN setup can be exploited for keV sterile neutrino search, as discussed in chapter 3. Such a sterile neutrino would impact the shape of the spectrum far from the endpoint region, therefore the simulation software requires substantial modifications.

In this section I will discuss the new simulation tool developed by Martin Pleštil, Slezák and Dr. David Radford for keV sterile neutrino search with KATRIN along with the modifications I made during the work that led to this thesis.

4.2.1 Motivation

The main reason why a new simulation software is needed is because it describes a different physics. The physical effects taken into account in SSC are very specific to the endpoint region of the spectrum, that is the region of interest for neutrino mass measurement. On the contrary, the Convolution code, among other features I will describe in the following paragraphs, takes into account energy dependent effects which are relevant far from the endpoint and track the angular distribution of electrons.

The physics differences

The most important feature to take into account regarding β -electrons from the source is that they scatter off tritium molecules before entering the spectrometer. These electrons can scatter inelastically (leading to ionization or excitation of the tritium molecule) or elastically. Corresponding to each scattering, elastic or inelastic, an energy loss and a polar angle change take place, leading to a substantial modification of the energy spectrum *and* of the angular distribution. These effects need to be taken into account in a neutrino mass measurement and in a keV sterile neutrino search.

In the first case, where one focuses on the endpoint region, some simplifications can be made. First, at ~ 18 keV the elastic cross section is more than one order of magnitude smaller than the inelastic one ($\langle \varepsilon_{\text{loss,inel}} \rangle \sim 13.6$ eV and for all purposes < 30 eV, [Gro15]). One can then include only the inelastic cross section in the model in a first approximation. Second, for a neutrino mass measurement, one wants to detect only the most energetic electrons. Therefore, the electrons that make their way towards the detector are those that scatter (inelastically) few times (≤ 2) and therefore do not

undergo significant angle changes (due to inelastic scattering)².

In the case of a model for a sterile neutrino search, these simplifications are not applicable and a more accurate description is needed. First, far from the endpoint the total inelastic cross section is energy dependent. Second, considering that one wants to scan the whole spectrum and the kink-like signature of a sterile neutrino is very small, the entire energy spectrum of the electrons after the source has to be precisely modelled, taking into account all energy losses (inelastic or elastic) and polar angle change regardless of their size.

Missing pieces in SSC

In SSC, the source related effects are included in the response function (2.25), or more precisely into the energy loss function $f(\varepsilon)$ and in the *average* probability of scattering i times in the source P_i . For neutrino mass measurement, the most prominent energy loss β -electrons can undergo is due to inelastic scattering off tritium molecules. The energy loss function is given by

$$f(\varepsilon) = \frac{1}{\sigma_{\text{tot,inel}}} \cdot \frac{d\sigma_{\text{inel}}}{d\varepsilon}. \quad (4.2)$$

In SSC a constant value (in energy) is used by default for the total inelastic cross section, that is $\sigma_{\text{tot,inel}} = (3.40 \pm 0.07) \cdot 10^{-18} \text{ cm}^2$ at $\sim 18 \text{ keV}$ [Ase00]. In the configuration file one can also include the elastic cross section, which anyway has a minor impact.

Besides including energy-independent parameter, the major limitations of SSC are mainly two: first, the maximum number of scatterings is fixed at the beginning of the simulation³; second, the angular related effects that change the shape of the spectrum are not taken into account and when they are, as I will discuss in the next paragraph, they are taken wrongly into account.

Nevertheless, in [Gro15] substantial changes have been applied to SSC to make it a more powerful simulation tool, as I will discuss in the next paragraph.

Inclusion of angular related effects in SSC

In [Gro15], among other effects, it is investigated how one can include angular related effects into SSC.

²The mean angular change due to inelastic scattering for energy losses below 30 eV is only 0.61° [Gro15].

³the response function is a finite sum, not a series, therefore it needs to be truncated at some point.

On one hand, the transmission function is modified. Considering that an electron generated with polar angle θ has a probability $P_i(\theta)$ of leaving the source after i inelastic scatterings, one can define a transmission function for each i substituting

$$\omega(\theta)d\theta = \sin\theta d\theta \longrightarrow \omega_i(\theta)d\theta = \sin\theta \cdot P_i(\theta)d\theta \quad (4.3)$$

in equation (2.19). This modification leads to the implementation of a *detailed transmission function for i scatterings* $T_i^*(E, qU - a)$. As [Gro15] shows, the major impact of this modification is due to T_0^* .

On the other hand, the angular change due to inelastic scattering and energy loss and angular change due to elastic scattering are included.

However, even when angular related effects are taken into account, it still made the assumption that after each scattering the polar angle of the electron is uniformly distributed. This assumption can still make sense in the endpoint region of the spectrum, but leads to wrong results in the modelling of the entire spectrum. I will show the modification I brought to the model in 4.2.3.

4.2.2 Idea of the Convolution code

In this section I want to give a general overview on how the Convolution code models the spectrum after the source, while in the next section I will go into the details of the code. The main goal of this new tool is the modelling of the *entire* tritium spectrum after source related effects. To this end, as I already mentioned, one has to take into account the fact that electrons can scatter many times (opposite to the fixed number of times in SSC) and that the angular change after scattering has a fundamental role. Therefore, instead of using a one-dimensional distribution (i.e. the differential energy spectrum) as in SSC, in the Convolution code one works with a two-dimensional distribution in energy and (cosine of the) polar angle.

At the beginning of the calculation the initial spectrum is isotropic in the angular variable and shaped as the differential spectrum in the energy variable. The result of the calculation will be a two-dimensional distribution with a modified shape, which will not be isotropic in the angular variable.

The basic idea of the simulation is that one starts binning the source (in the z direction) and fills each bin with the initial distribution⁴. At each step of the calculation, one redistributes the electrons born in a certain z bin in the neighbouring (or in the same) bins according to their initial polar angle and scattering probability⁵. In this way, as the simulation goes along, the z bins are emptied and two two-dimensional distributions are

⁴one can think of this as filling the source with electrons.

⁵which depends on the fraction of column density in the z bins, which itself is defined to be uniform.

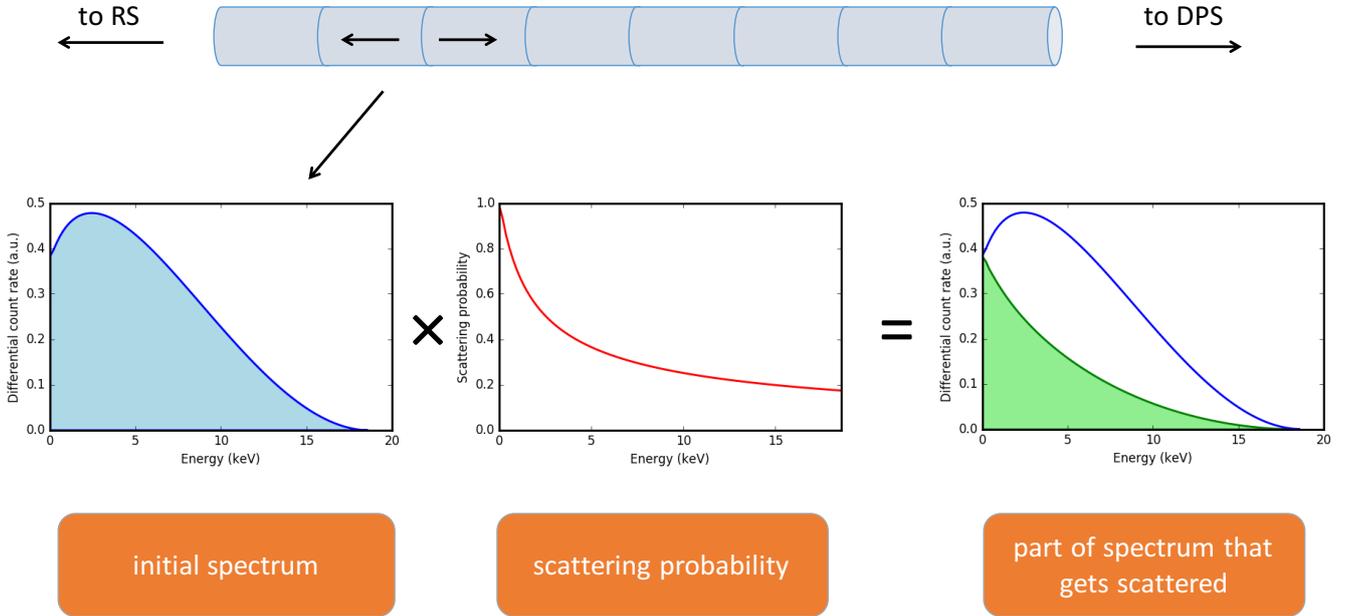


Figure 4.2: In the first step of the Convolution code one calculates for each energy bin the scattered (green) and unscattered (blue - green) part of the spectrum multiplying the initial distribution by the absolute scattering probability. Figure adapted from [Sle17]

generated: one for the *lost* electrons, which *exit* the source from the left⁶, and one for the electrons that exit the source from the right, heading to the detector. The fact that one has a two-dimensional distribution is very important here, because the transmission probability at the analyzing plane (and the detector response for instance) depends on *both* energy and angle.

At each step of the simulation, for each z bin, one multiplies the two-dimensional distribution in that z bin by the energy-dependent scattering probability, which is the probability of scattering at all. In this way one divides the distribution (or spectrum) in two parts, the part that gets scattered and the part that does not, see figure 4.2.

The part of the spectrum that get scattered is then convoluted with the double differential cross section, which describes the energy loss and angular change. In this way we get the part of the spectrum that gets scattered after scattering, see figure 4.3.

At this point, calling f the probability of scattering in the same z bin,

⁶i.e. they hit the rear wall.

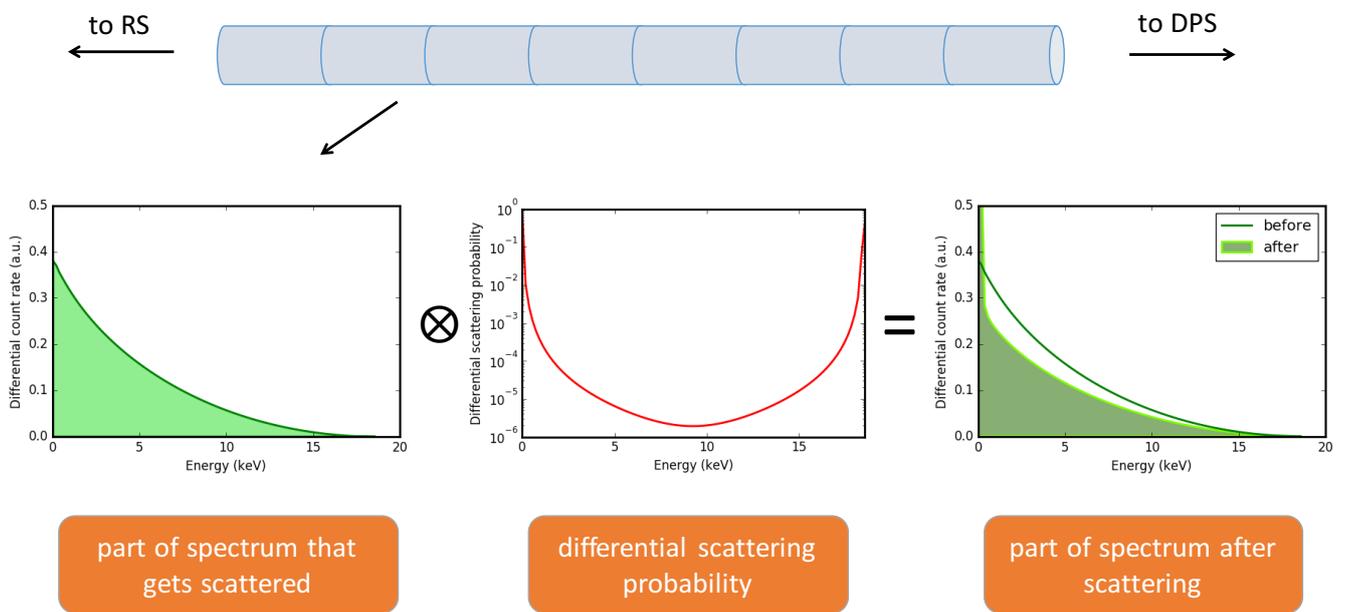


Figure 4.3: As second step one takes the part of the spectrum that scatters (green) and convolutes it with the double differential cross section to get the scattered part of the spectrum after scattering. Figure adapted from [Sle17]

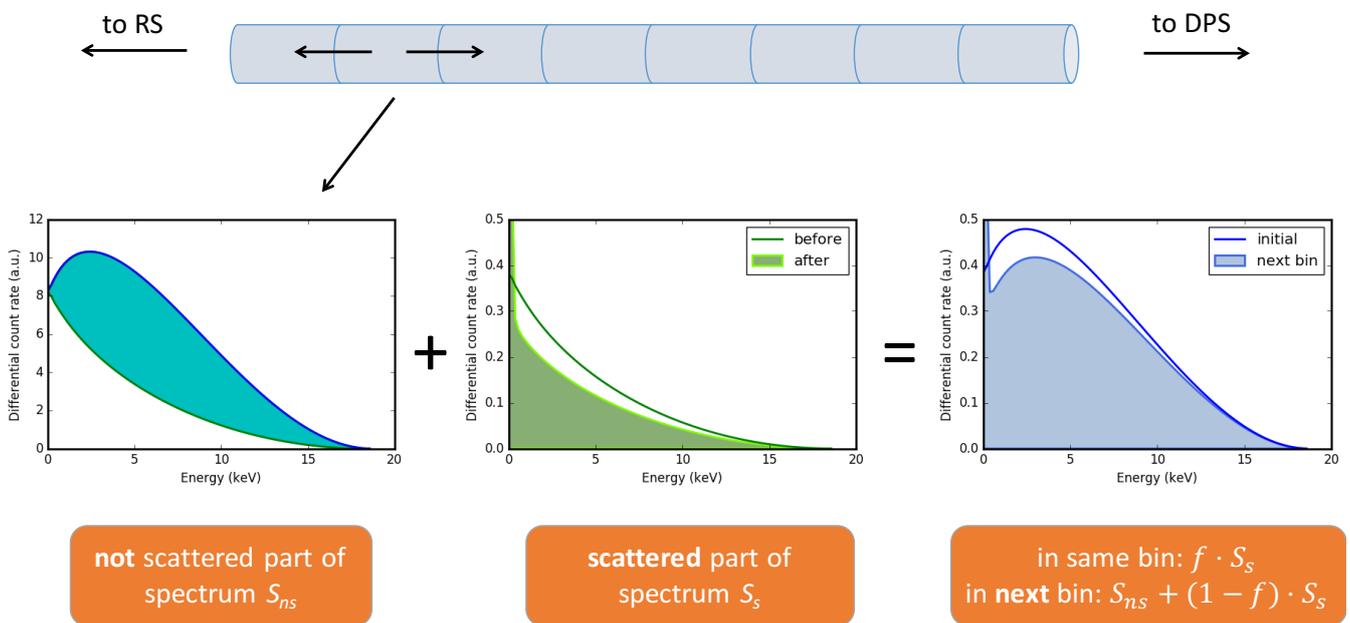


Figure 4.4: The part of the spectrum left in the same bin will be a fraction f of the scattered part of the spectrum S_s . In the next bin there will be the unscattered part of the spectrum S_{ns} and the fraction $(1 - f)$ of electrons that scatter S_s . Figure adapted from [Sle17]

the fraction of electrons left in the same bin will be $f \cdot S_s$, where S_s is the scattered part of the spectrum, while the fraction of electrons that goes into the neighbouring bin (left or right) will be $S_{ns} + (1 - f) \cdot S_s$, where S_{ns} is the not-scattered part of the spectrum, see figure 4.4. f is calculated as P/\bar{P} where P is the probability to scatter after passing the distance of one z bin and \bar{P} is the probability averaged over the bin width to scatter in the same bin. More precisely, $P = 1 - \exp(-\rho d_{\text{eff}} \cdot \sigma)$ and $\bar{P} = \int_0^1 1 - \exp(-\rho d_{\text{eff}} \cdot \sigma x) dx = 1 - P/(\rho d_{\text{eff}} \cdot \sigma)$, where ρd_{eff} is the effective column density traversed by an electron starting with polar angle θ given by $\rho d_{\text{eff}} = \rho d / \cos \theta$ and σ is the total cross section.

All these steps take place until the source gets empty, or better until the fraction of electrons in the z bins gets smaller than a fixed tolerance. The two-dimensional distributions that comes out of the last z bin is the result of the Convolution code simulation.

The model provided by the outlined procedures exhibits several advantages: one can keep track of the energy and angular distribution, there is no randomization, therefore no statistical uncertainties, the transmission function becomes a transmission *condition*, i.e. a yes or no question for each energy and angle and last, but not least, the angular change is taken into account at each scattering (even if it can be refined, as I will show in 4.2.3). On the downside, the complexity of the calculation grows exponentially with each new dimension [Sle17].

4.2.3 Structure of the Convolution code

Before going into the details of the code, I want to give here the outline of the code so that it can be used as a reference. One starts the simulation setting the output distribution $D_1(E, \theta, z)$ to zero and creating an initial distribution $D_0(E, \theta, z)$ flat in z , isotropic in θ and distributed as $\frac{d\Gamma}{dE}$ in energy. After that, the main loops of the calculations have the following structure:

Loop over angular values θ_i

 Loop over energy values E_i

 Calculate total scattering probability (probability of scattering at all) $P_1(E_i, \theta_i)$

 Calculate probability P_0 that a β electron gets to the next z bin before scattering

 Loop over angular change $d\theta$ ($\theta_f = \theta_i + d\theta$)

 Loop over energy loss $dE = 0 \dots E_i$ ($E_f = E_i - dE$)

 Calculate scattering probability $P_2(E_i, dE, d\theta)$

Loop over z bins

Increment $D_1(E_f, \theta_f, z)$ by $(1 - P_0) \cdot P_1 \cdot P_2 \cdot D_0(E_i, \theta_i, z)$

If the electron goes right ($\cos \theta_i > 0$), then increment $D_1(E_f, \theta_f, z + 1)$ by $P_0 \cdot P_1 \cdot P_2 \cdot D_0(E_i, \theta_i, z)$

If the electron goes left ($\cos \theta_i < 0$), then increment $D_1(E_f, \theta_f, z - 1)$ by $P_0 \cdot P_1 \cdot P_2 \cdot D_0(E_i, \theta_i, z)$

End loop on z

End loop on energy loss

End loop on angular change

If the electron goes right ($\cos \theta_i > 0$), then increment $D_1(E_i, \theta_i, z + 1)$ by $(1 - P_1) \cdot D_0(E_i, \theta_i, z)$

If the electron goes left ($\cos \theta_i < 0$), then increment $D_1(E_i, \theta_i, z - 1)$ by $(1 - P_1) \cdot D_0(E_i, \theta_i, z)$

End loop on E_i

End loop on θ_i

If the source is empty, quit.

Set $D_0 = D_1$.

The two variables of interest for us are the kinetic energy E of the betas and their polar angle θ . I will work in a reference frame where the z axis overlaps with beam axis and points to the detector. The information about the (E, θ) distribution needs to be kept until the detector section where angular related effects still matter.

Considering that in this reference frame the polar angle lies in the range $[0, \pi]$, I will rather use the cosine of the polar angle $\cos \theta \in [-1, 1]$. The reason to use $\cos \theta$ instead of θ is that all the cross sections become easier and the isotropic distribution is just flat in $\cos \theta$. **From now on, when I will say the word “angle” I will instead mean “cosine of angle”, or use them interchangeably.**

It is perfectly fine to assume that when a beta electron is produced in a decay, its angular distribution is isotropic. The goal of the Convolution code can be summarized in the following problem:

$$p(E, \cos \theta) = \frac{d\Gamma}{dE} \times Unif[-1, 1] \xrightarrow{\text{scattering effects}} p'(E, \cos \theta) = ? \quad (4.4)$$

Or in words:

If an electron in the source starts isotropically with energy distribution $\frac{d\Gamma}{dE}$, what is the probability that after the source it has energy E and polar angle θ ?

```

1. Initialize eMin = 0, eMax = E0;
2. Initialize EB, TB, zB;
3. Initialize tolerance ε and column density of the source ρd;
4. Initialize spectrum (dΓ/dE);
5. Initialize the vector scattprobs = (Elastic,Excit,Ion);
6. Initialize the object scattEff(scattprobs,EB,TB,eMin,eMax)
7. Initialize the sourceSim object: sourceSim(zB,scattEff);
8. Initialize the Distribution spec_after_source(EB,TB);
9. (caveat: for mono-energetic simulations initialize it as
10.vector of Distributions)
11.If(FULL SPECTRUM)
12.   spec_after_source = sourceSim.DoFullSpecSim(dΓ/dE, ε, ρd);
13.If(MONO)
14.   spec_after_source = sourceSim.DoMonoESim(ε, ρd);

```

Figure 4.5: Pseudo-code of the main function of the convolution code.

In this section I will explain how the Convolution code computes the probability $p'(E, \cos\theta)$. I will discuss how the main functions of the code work and what is the idea behind them. Not all the small details of the code will be discussed due to its complexity and no source code will be shown for the sake of readability. However, I will make use of *pseudo-code* to guide the reader through the steps and to provide a documentation of the actual source code.

I hope that using this reference one could easily navigate through the actual code, if needed.

tristanSourceSim

The Convolution code is a software written in C++ and it is made up of several modules that can be used individually, that, thanks to the cmake build system, are forced to remain compatible and link correctly together.

The main function lies in a file called `tristanSourceSim.cpp`. I will here show the pseudo-code of the main function and subsequently explain the functionality step by step.

In line 1 one sets the energy range of our interest, which will either be $[E_0 - 10 \text{ eV}, E_0]$ or $[0, E_0]$. I chose 10 eV to make the comparison near the endpoint with SSC.

In line 2 one declares the number of energy (EB), angular (TB) and z (zB) bins used in the simulation. EB and TB will be used to bin the initial distribution $\frac{d\Gamma}{dE} \times Unif[-1, 1]$, as I will soon explain, while zB will be used for the binning of the source along the beam axis. Usual values are $EB = 100$, $TB = 100$ and $zB = 40$, but I will specify their values again for each simulation.

In line 3 the tolerance ε and the column density of the source ρd are set. The use of the tolerance will be clear in the next parts. The column density

is in cm^{-2} units and typical values are 5×10^{17} for nominal KATRIN mode and 3×10^{15} for the Phase-0 measurement.

In line 4 one declares the spectrum object of type `KTSpectrumBeta`. In its default constructor the formula (3.6) is implemented.

In line 5 one initializes a vector whose entries are the effects that take place in the source. These effects are pushed back one by one into the scattering probabilities (`scattprobs`) object, so that if one is interested only in one effect, for example Ionization, one can just push back this effect and forget about the others. Although not displayed in the pseudo-code, each effect is constructed using the column density in one z bin ($= \rho d / zB$). In the WGTS electrons can undergo elastic and inelastic collisions with tritium. The latter can lead to either ionization of the tritium molecule or to excitation of electrons in the tritium shells. Therefore the three possible effects are elastic scattering (*Elastic*) and inelastic scatterings (*Excit, Ion*).

From the `scattprobs` vector one then initializes the scattering effects (`scattEff`) object (line 6). This object is initialized also using the number of angular and energy bin and our energy range of interest.

In line 7 one finally initializes our simulation tool `sourceSim`. It is constructed using the number of bins the source is divided into and the scattering effects. The type of the object `sourceSim` is the class template `KTSourcSim<template>`. The template argument can either be the class `KTDistribution` or `KTDistributionPhi`. One can think of `KTDistribution` as a prototype for `KTDistributionPhi` considering that it was developed earlier and it did not include the right formula to compute the angular change after each scattering (I will come back to this problem in 4.2.3).

In line 8 one initializes the spectrum after the source (`spec_after_source`) object whose type is `KTDistributionPhi` (or `KTDistribution`). The data member of this object is a $EB \times TB$ matrix that whose elements will be the binned content of the initial distribution. At the moment of its declaration, `spec_after_source` is null matrix. As the *caveat* in 4.5 lines 9-10 says, if one wants to simulate mono-energetic electrons, one would instead initialize `spec_after_source` as a vector of distributions (one for each starting energy).

Finally (lines 11-on), if one want to simulate starting from the tritium differential spectrum one calls the `DoFullSpecSim` method of the `sourceSim` object. This method takes the differential spectrum $\frac{d\Gamma}{dE}$, the tolerance ε and the column density ρd as parameters, returns a `KTDistributionPhi` (or `KTDistribution`) which is a binned version of $p'(E, \cos\theta)$ (see (4.4)) and assigns it to `spec_after_source`. If instead one wants to simulate mono-energetic electrons one calls the `DoMonoESim` method of the `sourceSim` object that does not need the tritium spectrum and will instead return a vector of `KTDistributionPhi` (or `KTDistribution`).

In the next part I will describe how the function `DoFullSpecSim` performs the simulation. The `DoMonoESim` function works in a similar fashion,

```

1. DoFullSpecSim(dΓ/dE, ε, ρd)
2. {
3.   Initialize an isotropic Distribution (which is
   essentially a matrix) A storing in it the binned
   differential spectrum dΓ/dE;
4.   Return DoSimulation(A, ε, ρd);
5. }

```

Figure 4.6: Pseudo-code of the DoFullSpecSim function of the convolution code.

therefore I will not discuss it here.

DoFullSpecSim

The main goal of the DoFullSpecSim is to transform the nice theoretical spectrum $\frac{d\Gamma}{dE}(\times Unif[-1, 1])$ into a matrix A that one can actually use for the simulation through the DoSimulation function.

I will describe now the binning of the differential spectrum and how one initializes an isotropic distribution. The binning procedure introduces an intrinsic approximation in our calculation, whose impact will be discussed in 4.3.

The first step is to represent the continuous spectrum $\frac{d\Gamma}{dE}$ with a discrete version of it:

$$\left(E, \frac{d\Gamma}{dE}(E)\right) \rightarrow (E_i, s_i) \quad i = 0, \dots, EB - 1. \quad (4.5)$$

In order to do so, the energy range $[e_m, e_M]$ is divided into EB intervals whose width is $e_w = \frac{e_M - e_m}{EB}$. One then has EB intervals numbered from 0 to $EB - 1$ and $EB + 1$ points $e_j = e_m + j \cdot e_w$, $j = 0, \dots, EB$ ($e_0 = e_m$, $e_{EB} = e_M$). One then choose the midpoints of each interval as the points where one wants to approximate the spectrum. I will denote the midpoint of the i^{th} interval as $E_i = e_m + i \cdot e_w + e_w/2$. The exact value of the spectrum at E_i is then $\frac{d\Gamma}{dE}(E_i)$. One approximates this value with the bin content of the i^{th} interval, which is given by:

$$s_i \equiv \int_{e_i}^{e_{i+1}} \frac{d\Gamma}{dE}(E) dE \quad i = 0, \dots, EB - 1. \quad (4.6)$$

It is worth noticing that while $\frac{d\Gamma}{dE}(E_i)$ has units of counts per energy (per second), s_i has units of counts (per second). Therefore s_i should be thought as number of beta electrons (per second) in the energy range $[e_i, e_{i+1}]$.

At this point a vector (s_0, \dots, s_{EB-1}) that approximates our theoretical spectrum is created.

Considering the initial isotropic angular distribution, the matrix A is simply given by

$$A = \begin{pmatrix} s_0 & s_0 & \dots & s_0 \\ s_1 & s_1 & \dots & s_1 \\ \vdots & \vdots & \ddots & \vdots \\ s_{EB-1} & s_{EB-1} & \dots & s_{EB-1} \end{pmatrix} \quad (4.7)$$

up to a normalization constant. This matrix has EB rows and TB columns reflecting the energy and angular binning respectively. The angular range $[-1, 1]$ is discretized similarly to the energy range: it is divided into TB intervals whose width is $t_w = \frac{2}{TB}$. In total one has TB intervals numbered from 0 to $TB - 1$. The midpoints of each interval are the points where one wants to approximate the angular distribution: the midpoint of the i^{th} interval is $\cos \theta_i = -1 + i \cdot t_w + t_w/2$. This procedure provides a discrete uniform distribution of the values $\cos \theta_0, \dots, \cos \theta_{TB-1}$.

The structure of A ($A_{ij} = s_i \forall j$) reflects exactly the isotropy of the angular distribution: for a particular energy, the probability of having a certain angle is the same for all angles.

After the matrix A is built, one can give it as a parameter to the `DoSimulation` function.

DoSimulation

The `DoSimulation` function is the function that takes the initial binned spectrum A as an input and returns a final binned spectrum (a matrix) that I will denote by B . The calculation is performed in the `KTDistribution(Phi)` method `Convolution` that I will discuss in the next section.

As one can see from figure 4.7, in the `DoSimulation` function one first initializes the object `specPre`, which is a vector of `KTDistribution(Phi)`'s (matrices $EB \times TB$), with A :

$$\text{specPre} = \begin{array}{c|c|c|c|c} \textit{index} & 0 & 1 & \dots & zB - 1 \\ \textit{value} & A & A & \dots & A \end{array}$$

`specPre` represent the starting object of the simulation. One can think of this initialization as filling each z bin of the source with the right amount of beta electrons⁷.

The main idea of the simulation is that at each step, these electrons will move right or left into the source changing the number of electrons per z bin. As eventually all these electrons will be guided out of the source, the

⁷Remember that the `sourceSim` object is initialized with the `scattEff` object, which itself is initialized with the `scattprobs` object, which itself (almost there!) is initialized with the column density in **one** z bin! So it all makes sense!

```

1. DoSimulation(A, ε, ρd)
2. {
3.   initialize the vector of distributions (EB×TB) specPre with zB entries: specPre = (A,...,A);
4.   initialize the vector of distributions (EB×TB) specPost with zB entries: specPost = (0,...,0);
5.   initialize the distributions (EB×TB) specFinal and specLost;
6.   initialize sumAll summing up all the elements of specPre;
7.   initialize sumFinal, sumLost to 0;
8.   while(there are still more than "enough" electrons in the source)
9.   {
10.    sumAll=0;
11.
12.    For each zBin do
13.    {
14.      For internal bins (i=1,...,zB-2) do
15.        specPre[i].Convolution(scattEff, specPost[i-1], specPost[i], specPost[i+1]);
16.      For the first bin (i=0) do
17.      {
18.        specPre[0].Convolution(scattEff, specLost, specPost[0], specPost[1], false, true);
19.        sumLost = sum up all the elements of specLost;
20.      }
21.      For the last bin (i=zB-1) do
22.      {
23.        specPre[zB-1].Convolution(scattEff, specPost[zB-2], specPost[zB-1], specFinal, true, false);
24.        sumFinal = sum up all the elements of specFinal;
25.      }
26.    }
27.  }
28.  // update infos:
29.  For each zBin (i=0,...,zB-1)
30.  {
31.    specPre[i] = specPost[i];
32.    reset specPost[i];
33.    increment sumAll by the sum of all elements in specPre[i];
34.  }
35.  increment sumAll by sumLost and sumFinal;
36.  return spectrumFinal;
37.}

```

Figure 4.7: Pseudo-code of the DoSimulation function of the convolution code.

simulation stops when the number of remaining betas is under the fixed tolerance (that depends on ε).

From `specPre` one builds the `specPost` and in the end sets `specPre = specPost` (lines 29 - 31 figure 4.7).

The `specPost` object is initialized as a null vector of distributions:

$$\text{specPost} = \begin{array}{c|c|c|c|c} \text{index} & 0 & 1 & \dots & zB - 1 \\ \hline \text{value} & 0 & 0 & 0 & 0 \end{array}$$

where every 0 is a $EB \times TB$ matrix.

In line 5 one initializes the `specFinal` and `specLost` `KTDistribution(Phi)` objects. `specFinal` will be the final output of the DoSimulation: it represent the electron distribution that comes out of the source and goes towards the spectrometer section. `specLost` represents the electrons that at each iteration leave the source from the left. `sumAll` is the number of electrons in the source at each iteration step. It is initialized summing up all elements of all matrices of `specPre`. `sumFinal` and `sumLost` keep track of the electrons that at each step leave the source from the right and left, respectively. They are both initialized at 0.

The condition to enter in the while loop (line 8) depends on `sumAll`, `sumLost`, `sumFinal` and the tolerance ε , and it is easily read as *while there are still more than "enough" electrons in the source do ...*

Inside the while loop the first thing is setting the total number to 0 as it will be incremented at each step summing up what is left in `specPre` (lines 33,35).

At this point for each z bin one calls the `KTDistribution(Phi)` method `Convolution`. Its task is to distribute the electrons of a z bin into itself or into the neighbouring z bins according to the (scattering) effects that take place.

Three cases need to be distinguished:

line 14 if the current z bin is an internal bin ($i = 1, \dots, zB - 2$) then one calls the `Convolution` method from the `specPre[i]` distribution object. The result of this operation will be splitting the distribution `specPre[i]` into `specPost[i - 1]`, `specPost[i]` and `specPost[i + 1]`, which means distributing the electrons that are in the i^{th} z bin into the $(i - 1)^{\text{th}}$, i^{th} and $(i + 1)^{\text{th}}$ z bins.

line 16 if the current z bin is the first bin ($i = 0$) then one calls the `Convolution` method from the `specPre[0]` distribution object. The result of this operation will be splitting the distribution `specPre[0]` into `specLost` (because the electrons that go left are *lost*), `specPost[0]` and `specPost[1]`, which means distributing the electrons that are in the first z bin into *electrons that go out from the left*, first and second z bins.

Here one also keeps track of the number of lost electrons saving it into the `sumLost` variable.

line 21 if the current z bin is the last bin ($i = zB - 1$) then one calls the `Convolution` method from the `specPre[zB - 1]` distribution object. The result of this operation will be splitting the distribution `specPre[zB - 1]` into `specPost[zB - 2]`, `specPost[zB - 1]` and `specFinal` (because after the last bin they just go out of the source from the right side), which means distributing the electrons that are in the last z bin into the second-last and last bin, and into "electrons that go out from the right".

Here one also keeps track of the number of electrons that exit the source from the right side storing it into the `sumFinal` variable.

Finally, one updates `specPre` with `specPost` (line 31), resets `specPost` to zero (line 32), increases `sumAll` with the number of electrons left in the source (line 33) and returns `specFinal` (which is still a matrix that I will call B).

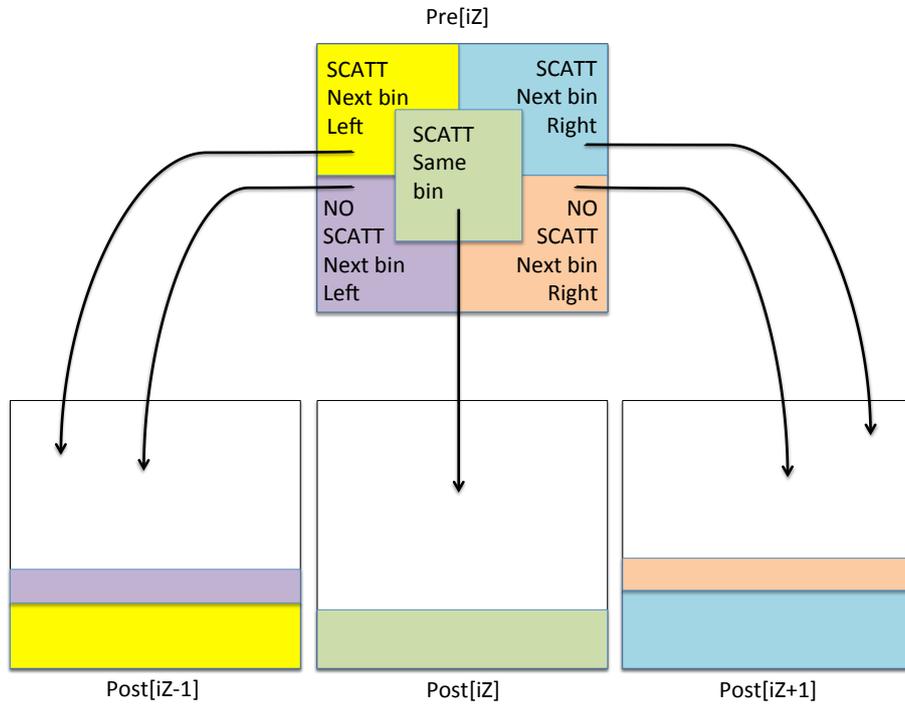


Figure 4.8: Pictorial representation of how the Convolution function distributes the electrons in one z bin into the same and neighbouring z bins.

Convolution

In this section I will sketch how the Convolution method works. This is the most technical part of the code and, instead of showing a pseudo-code, I will show the underlying logic and the formulas implemented. A pictorial description of this function is shown in figure 4.8.

The Convolution method, called from the `specPre[i]` object, takes `scattEff`, `specPost[i - 1]`, `specPost[i]` and `specPost[i + 1]` as parameters. From now on I will refer to them as Pre, left, middle and right for simplicity.

The idea of this method is that each beta will eventually scatter somewhere in the source (that is, in some z bin). One can then divide the electrons in two classes (see figure 4.9):

SCATT Electrons that scatter either in the same z bin they are born or in the next z bin (either left or right).

NO SCATT Electrons that scatter neither in the same z bin they are born nor in the next z bin. Since they will eventually scatter in some other z bin, at each step of the iteration these electrons will be move left or right according to they polar angle.

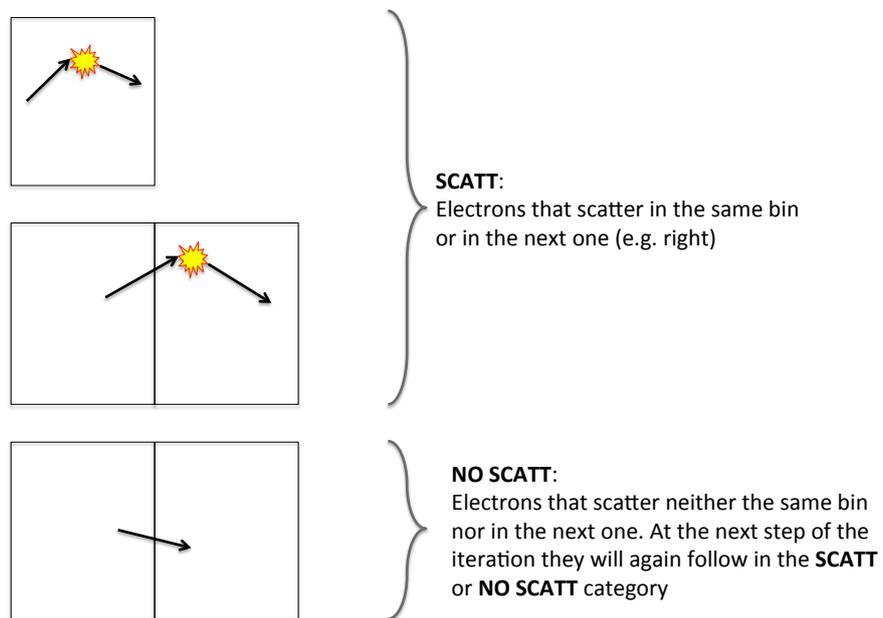


Figure 4.9: Pictorial representation of how the **SCATT** and **NO SCATT** category.

For each energy bin iE and angular bin iT , the code distinguishes two cases:

- if(**SCATT**)

if the electron goes right (or left) then one declares the pointer to double $value$ and makes it point to $right[iE][iT]$ (or $left[iE][iT]$). Furthermore, **if we are in the last (or first) z bin, then one sets the double variable fac to 0.**

At this point, looping over all allowed angular and energy changes, one computes the convolution of the Pre distribution with the scattering cross section and assign it to $value$. This is mathematically given by the formula⁸:

$$\begin{aligned} *value = \sum_{iT} \sum_{idE} & Pre[iE - idE][iT + idT] \cdot P(iE - idE, iT + idT) \\ & \cdot DDP(iE - idE; idE, idT) \\ & \cdot FracNextBin(iE - idE, iT + idT)(\cdot fac), \end{aligned} \quad (4.8)$$

where \sum_{iT} and \sum_{idE} represent the sum over all allowed energy and angular changes, $P(E, T)$ is the absolute probability that an electron in the energy bin E and angular bin T scatters at all, $DDP(E, dE, dT)$ is the Double Differential Probability (which encodes the double differential cross section $\frac{d^2\sigma}{dEd(\cos\theta)}$), that is the probability that an electron in the energy bin E undergoes an energy change of dE and an angular change of dT due to scattering, and finally $FracNextBin(E, T)$ is the fraction of electrons with energy E and angle T that scatter in the next z bin. The factor fac makes sure that if the electron is going right (or left) and we are at the last (or first) z bin then $*value = 0$.

The electrons that scatter in the same z bin will keep being in the same z bin, therefore one computes $middle[iE][iT]$ as

$$\begin{aligned} middle[iE][iT] = \sum_{iT} \sum_{idE} & Pre[iE - idE][iT + idT] \cdot P(iE - idE, iT + idT) \\ & \cdot DDP(iE - idE; idE, idT) \\ & \cdot FracSameBin(iE - idE, iT + idT), \end{aligned} \quad (4.9)$$

where $FracSameBin(E, T)$ is the fraction of electrons with energy E and angle T that scatter in the same z bin.

⁸the star operator in front of $value$ is there because it is a pointer

- if(**NO SCATT**)

Each electron goes either right or left. Let us assume for the sake of simplicity that it goes right. Then one updates the value of the right distribution in the following way:

$$\begin{aligned}
 right[iE][iT] = \sum_{iT} \sum_{iE} & \quad Pre[iE][iT] \cdot \left[(1 - P(iE, iT)) + \right. \\
 & \quad \left. + P(iE, iT) \cdot \underbrace{FracNextBin(iE, iT)}_{\text{if we are in the last } z \text{ bin}} \right] \quad (4.10)
 \end{aligned}$$

this formula represents the fact that, in the **NO SCATT** case, then one convolutes the current distribution with the absolute probability of not scattering. If furthermore we are in the last z bin, then one needs to add also the leftovers, which are the electrons that will scatter in the *would-be* next z bin.

This completes our discussion on how the most important components of the Convolution code work. From now on I will only discuss modifications of it and use it to produce useful results.

Convolution 2.0: The φ -problem

It has now come the moment to discuss the so called φ -*problem*. As sophisticated as the convolution code is, up to now there is still a missing ingredient. It is still making the wrong assumption, as in SSC, that at each scattering the new polar angle *is uniformly distributed*.

I will now show how to get the right formula for the polar angle of electrons after scattering and its distribution. Let us work in a frame of reference where the z axis is parallel to the experiment axis and points to the detector. Given the cylindrical symmetry of the source it does not matter how the x and y axis are oriented as long as the system is right-handed. Let us assume that an electron is generated at some point O , which I set as the origin of our reference frame, with a polar angle θ_0 ($\cos \theta_0 \sim \text{Unif}[-1, 1]$) and an azimuthal angle φ_0 ($\varphi_0 \sim \text{Unif}[0, 2\pi]$). This electron will then scatter at a point O' and go to the point P . At O' a new uniform polar angle θ_p and azimuthal angle φ_p will be generated. However, they refer to the reference frame of the electron, where $\overrightarrow{OO'}$ defines the z axis. Therefore the problem is the following: *what is the polar angle of the electron with respect to the beam axis?*

As it turns out, this is a simple geometry exercise.

Let us set up a bit of notation referring to figure 4.10. Let $\vec{v} \equiv \overrightarrow{OO'}$ and $v = |\vec{v}|$. I define also $\hat{u} = \frac{\hat{z} \times \vec{v}}{|\hat{z} \times \vec{v}|}$ ⁹.

⁹for any vector $\vec{v} = (v_1, v_2, v_3)$ one defines $\hat{v} = \frac{\vec{v}}{|\vec{v}|}$ and $v = |\vec{v}|$

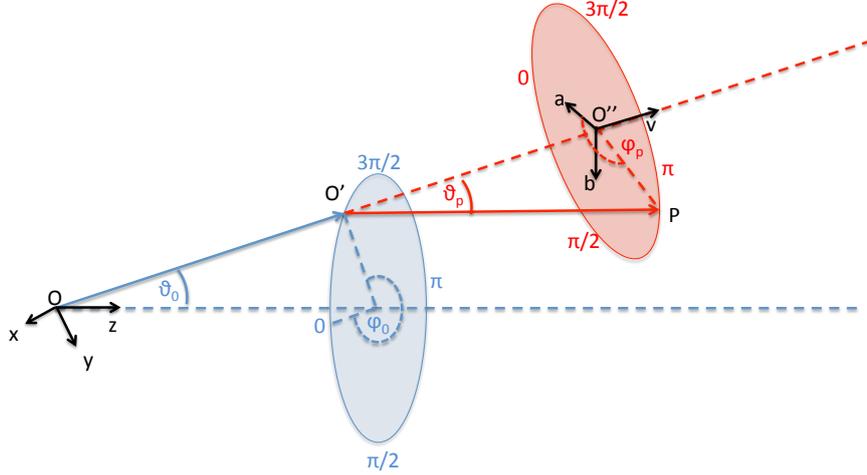


Figure 4.10: Geometry of the scattering.

The coordinates of O' in the Oz reference frame are

$$O' \sim \overrightarrow{OO'} = (R_0 \cos \varphi_0, R_0 \sin \varphi_0, v \cos \theta_0) \quad (4.11)$$

where $R_0 = v \sin \theta_0$. One then has that $\hat{v} = (\sin \theta_0 \cos \varphi_0, \sin \theta_0 \sin \varphi_0, \cos \theta_0)$.

Let us now define $R_u(\theta)$ as the (right-handed) rotation around \hat{u} of an angle θ . From its definition one has that $\hat{u} = (-\sin \varphi_0, \cos \varphi_0, 0)$. $R_u(\theta_0)$ is the rotation that brings \hat{z} on \hat{v} and is given by

$$R_u(\theta_0) = \begin{pmatrix} \cos \theta_0 + \sin^2 \varphi_0 (1 - \cos \theta_0) & -\sin \varphi_0 \cos \varphi_0 (1 - \cos \theta_0) & \cos \varphi_0 \sin \theta_0 \\ -\sin \varphi_0 \cos \varphi_0 (1 - \cos \theta_0) & \cos \theta_0 + \cos^2 \varphi_0 (1 - \cos \theta_0) & \sin \varphi_0 \sin \theta_0 \\ -\cos \varphi_0 \sin \theta_0 & -\sin \varphi_0 \sin \theta_0 & \cos \theta_0 \end{pmatrix}$$

Calling the first and second column \hat{a} and \hat{b} one has that $\hat{a} = R_u(\theta_0)\hat{x}$ and $\hat{b} = R_u(\theta_0)\hat{y}$, and the third column is $\hat{v} = R_u(\theta_0)\hat{z}$ by construction.

Setting up this machinery, our initial problem has turned in the following simple question: *what is the angle Θ_p between $\overrightarrow{O'P}$ and \hat{z} ?*

Let us set $\vec{w} \equiv \overrightarrow{O'P}$ and $R_p = w \sin \theta_p$. One has that

$$P \sim \overrightarrow{OP} = \underbrace{\overbrace{(v + w \cos \theta_p)}^{|\overrightarrow{OO''}|}}_{\text{vector } \overrightarrow{OO''}} \hat{v} + R_p \cos \varphi_p \hat{a} + R_p \sin \varphi_p \hat{b}. \quad (4.12)$$

Considering that $\vec{w} = \overrightarrow{O'P} = P - O'$ and that $\cos \Theta_p = \hat{w} \cdot \hat{z} = w_3$, then one gets

$$\boxed{\cos \Theta_p = \cos \theta_p \cos \theta_0 - \sin \theta_p \sin \theta_0 \cos(\Delta\varphi)} \quad (4.13)$$

where $\Delta\varphi = \varphi_p - \varphi_0$. The angle φ_0 in formula (4.13) has no particular meaning, since it can always be set to 0 rotating the x and y axes. Therefore the variable $\Delta\varphi$ is uniformly distributed in $[0, 2\pi]$.

It is true in general that given a uniformly distributed variable x and a function f , the distribution of $y = f(x)$ is given by $\frac{d}{dy}f^{-1}(y)$. Therefore the distribution of $\cos \Theta_p$, given that $\Delta\varphi \sim \text{Unif}[0, 2\pi]$, is

$$\frac{d(\Delta\varphi)}{d(\cos \Theta_p)} = \frac{1}{\sin \theta_p \sin \theta_0 \sqrt{1 - \frac{\cos \theta_p \cos \theta_0 - \cos \Theta_p}{\sin \theta_p \sin \theta_0}}} \quad (4.14)$$

To include the distribution $\frac{d\Delta\varphi}{d(\cos \theta)}$ in the convolution code the additional library `KTDistributionPhi.h` is implemented. It implements a method also called Convolution, whose structure is exactly the same as the one discussed in 4.2.3. The only difference is that one replaces the double differential cross section with its convolution with the $\frac{d\Delta\varphi}{d(\cos \theta)}$ distribution in formula (4.8):

$$\frac{d^2\sigma}{dEd(\cos \theta)} \longrightarrow \frac{d^2\sigma}{dEd(\cos \theta)} \otimes \frac{d\Delta\varphi}{d(\cos \theta)}. \quad (4.15)$$

From the theoretical point of view there is no issue in doing this. However, from the point of view of computational time this procedure lengthens the runtime of the Convolution code considerably¹⁰.

4.3 Comparison of SSC and Convolution code in the last 10 eV

Before performing any meaningful calculation with our Convolution code, I tested its validity. I start from the assumption that SSC is a model of the integral spectrum sophisticated enough to test the accuracy of the Convolution code. The idea is then to compare the integral and differential spectra computed with SSC and the Convolution code in the last 10 eV of the spectrum with various parameter configurations.

If they give the same result in this energy range, one can conclude that the Convolution code can be used to model the entire tritium spectrum.

¹⁰because we are adding a 5th for loop inside 4 nested for loops

```

1. AnalyzPlane(spec_after_source, Bs, Bp, Ba, Emin)
2. {
3.   For retarding potential qU from E0-Emin up to E0
4.   {
5.     Initialize temporary distribution temp;
6.     For angular bin iT from 0 to TB-1
7.     {
8.       PM = electrons in this angular bin pass the pinch magnet;
9.       For energy bin iE from 0 to EB-1
10.      {
11.        AP = electrons in this energy and angular bin pass the analyzing plane;
12.        if (PM and AP) temp[iE][iT] = spec_after_source[iE][iT];
13.        else temp[iE][iT] = 0;
14.      }
15.    }
16.    write into a file qU and temp.Sum();
17.  }
18. }
19.

```

Figure 4.11: Pseudo-code of the `AnalyzPlane` function of the convolution code.

4.3.1 The transmission condition

Up to now I did not showed any integral spectrum with the Convolution code. According the figure 4.5, after the call of the `DoFullSpecSim` method in line 12, the `KTDistribution(Phi)` object `spec_after_source` has the $EB \times TB$ matrix B as data member (see 4.2.3). Each element B_{ij} is the probability¹¹ that an electron after the source has energy E_i and cosine of the polar angle $\cos \theta_j$.

As already mentioned, in the Convolution code, source and transmission effects are implemented separately. The matrix B is computed only taking into account scattering effects in the source. I will show now how to get an integral spectrum from it, that is, how I implemented implement a *transmission condition*.

I defined a new method called `AnalyzPlane` in the `KTDistribution(Phi)` class. Here is its pseudo-code:

The idea of this method is fairly simple and it makes use of the transmission condition at the pinch magnet (where the magnetic field is maximal) and at the analyzing plane¹²:

$$\cos \theta > \cos \left(\arcsin \sqrt{B_s/B_p} \right) \quad \text{transmission at the pinch magnet,} \quad (4.16)$$

$$\left(1 - \frac{B_a}{B_s} \sin^2 \theta \right) \cdot E > qU \quad \text{transmission at the analyzing plane.} \quad (4.17)$$

¹¹or the number of electrons if not normalized

¹²see (2.8) and (2.15)

For each retarding potential qU ranging in $[E_0 - E_{min}, E_0]$, one loops over all angular and energy bin and sets the bin value to zero if that bin does not fulfil either (4.16) or (4.17). One then sums the remaining elements in our distribution to get the number of electrons that make their way to the detector at the fixed retarding potential. One then saves the values of the integral spectrum into a text file.

I will now proceed with the comparison of the spectra produced with SSC and the Convolution code under the following assumptions:

- energy range = $[E_0 - 10 \text{ eV}, E_0]$ with $E_0 = 18575 \text{ eV}$,
- $m_\beta^2 = 0$,
- no sterile mixing,
- $F(E, Z') = 1$ (because in the Convolution code it is not implemented yet),
- no final state corrections.

4.3.2 Differential spectra

At first one wants to make sure that both software have the same starting point. Although the theoretical formula for the differential spectrum should be the same, tiny differences can depend on the implementation. A minor difference is the way the codes implement the physical constants (\hbar , Cabibbo angle, etc...). The major difference comes instead from the binning of the Convolution code.

In SSC the differential spectrum is implemented choosing n points in the energy range and then applying (1.65).

In the Convolution code the differential spectrum is computed setting $eMin = E_0 - 10 \text{ eV}$ and $TB = 1$ ¹³ in `tristanSourceSim.cpp`.

Finally, in the `DoFullSpecSim` function, instead of returning `DoSimulation(A, ε , ρd)`, one returns A . This is equivalent to just binning the spectrum given by (1.65).

Given that the only calculation involved in the Convolution code is the integration over the energy bin width, I chose to use $EB = 10000$ and therefore to compute the SSC differential spectrum in 10000 points.

The plot of the relative difference and ratio are shown in figure 4.12 and 4.13. The relative difference is more than 0.01% only in the last 0.03 eV and it is therefore due to the binning of the theoretical spectrum.

¹³using one angular bin is the same thing of not resolving the angular distribution.

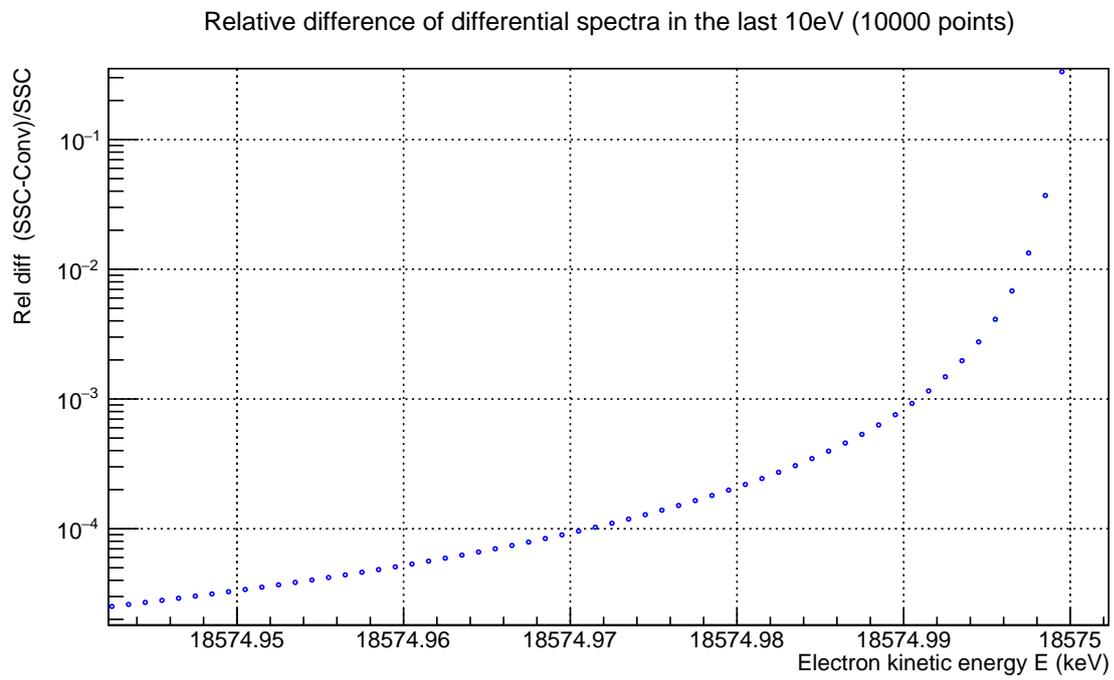


Figure 4.12: Relative difference of SSC and Convolution code's differential spectra in log scale.

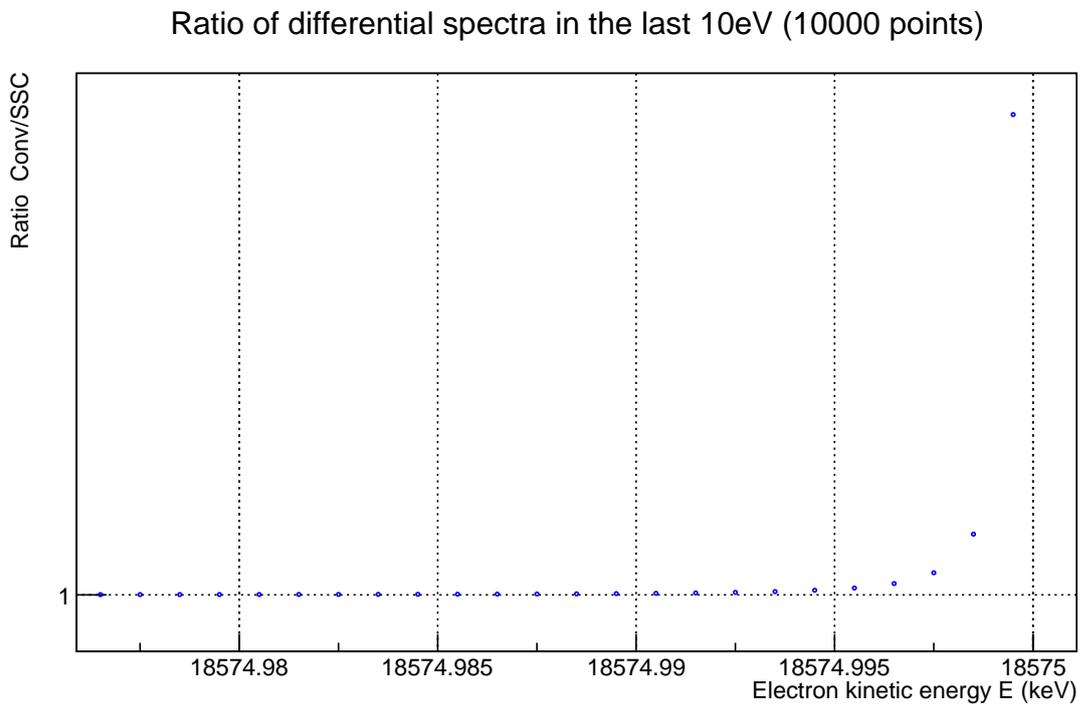


Figure 4.13: Ratio of SSC and Convolution code's differential spectra ($Conv/SSC$) in log scale.

4.3.3 Differential spectra + Transmission

The previous results are promising. I now want to add an ingredient to the pure differential spectrum and compare the results again.

I previously said that in SSC the scattering effects in the source and the transmission probability through the spectrometer are tied together in the response function (see 2.1) and cannot be treated separately. However, this is not quite true.

One can introduce these two ingredients into the response function in steps. In fact, if one set the column density $\rho d = 0$ (that is, if the source is empty) in SSC, then the response function reduces to the transmission function. Setting then the column density to a non zero value one gets back the whole response function (see next section).

In this section I will then compare the following spectra:

- SSC differential spectrum convoluted with the transmission function ($\rho d = 0$),
- the result of `AnalyzPlane`($A, B_s, B_p, B_A, E_0 - 10 \text{ eV}$) where A is just the binned differential spectrum (that is, scatterings are OFF).

The values of the magnetic field for both simulations are $B_s = 3.6 \text{ T}$ (source), $B_p = 6 \text{ T}$ (pinch magnet, maximal B field) and $B_A = 3 \cdot 10^{-4} \text{ T}$ (analyzing plane). The plot of the relative difference and ratio are shown in figure 4.14 and 4.15.

One can see from 4.14 that the difference is still less than 1%.

4.3.4 Differential spectra + Scatterings + Transmission

I can finally add the final ingredient to the comparison: scattering effects. In this section I will then compare the following spectra:

- SSC integral spectrum,
- the result of `AnalyzPlane`($B, B_s, B_p, B_A, E_0 - 10 \text{ eV}$) where B is result of `DoSimulation`($A, \varepsilon, \rho d$).

The parameter settings are shown in table 4.1. Let us discuss shortly the parameters D and A .

D stands for *detailed transmission for D scatterings*. This is a parameter in the configuration file of SSC and it can take values from -1 up to the max number of scatterings¹⁴. For $D = -1$ the usual analytical transmission function as in (2.21) is used. For $D = i$ ($i = 0, 1, 2, \dots$) the formula (2.19) is modified in the following way:

¹⁴in the configuration file of SSC one fixes a priori the maximal number of times an electron can scatter in the source, so that the sum in (2.25) has a finite number of terms.

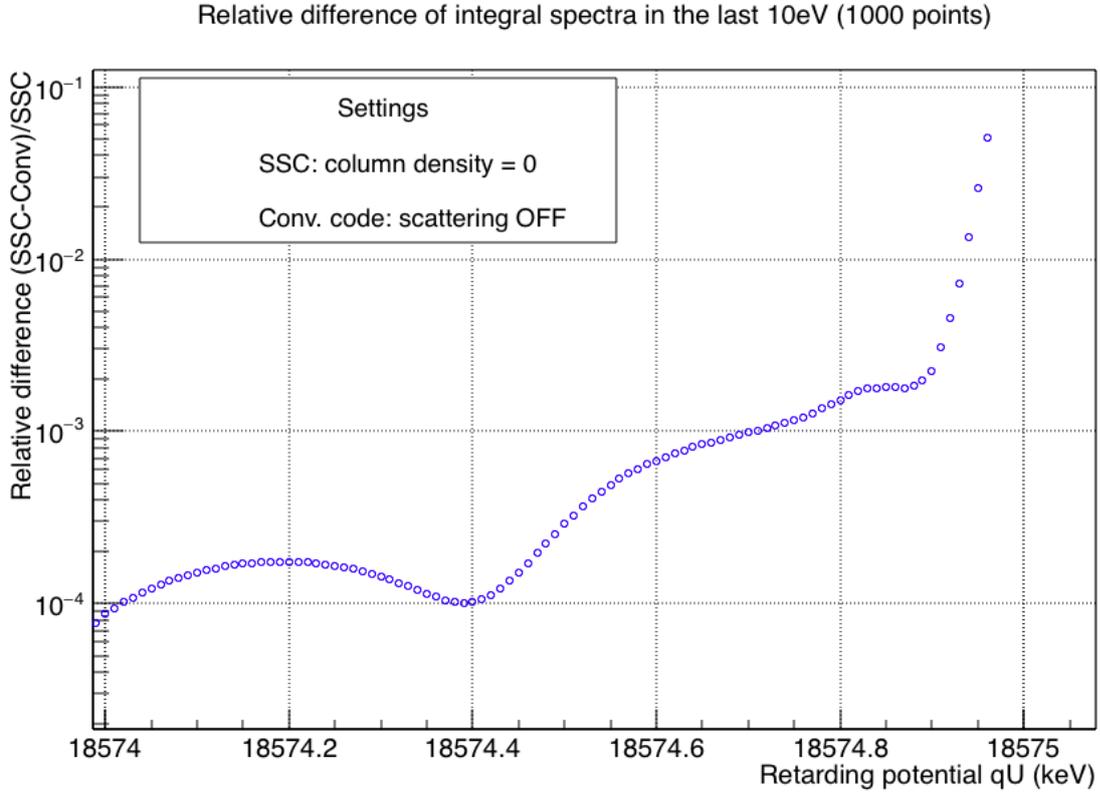


Figure 4.14: Relative difference of SSC and Convolution code's differential spectra plus transmission in log scale. SSC parameters: $\rho d = 0$, 1000 points. Convolution parameters: $EB = TB = 1000$, $zB = 100$, $B_s = 3.6$ T, $B_p = 6$ T and $B_A = 3 \times 10^{-4}$ T.

Model	ρd (cm^{-2})	B_s (T)	B_p (T)	B_A (T)	$\frac{EB}{TB}$ zB	Elastic scatterings	Inelastic scattering	# of points	D	A	φ -convolution
SSC	$5 \cdot 10^{17}$	3.6	6	$3 \cdot 10^{-4}$	X	ON	ON	1000	-1,1 or 5	True or False	X
Conv.					1000 200 100			X	X	X	OFF

Table 4.1: Parameter values for the comparison of the complete integral spectra.

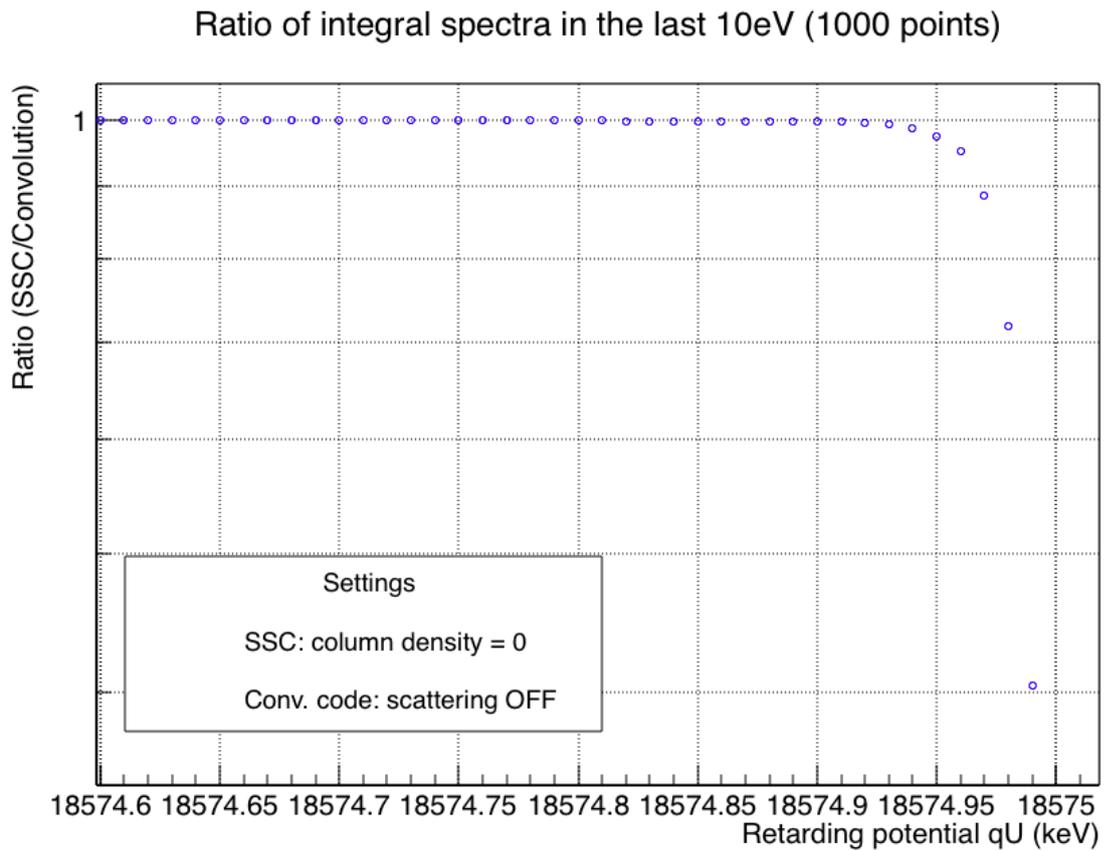


Figure 4.15: Ratio of SSC and Convolution code's differential spectra plus transmission in log scale. SSC parameters: $\rho d = 0$, 1000 points. Convolution parameters: $EB = TB = 1000$, $zB = 100$, $B_s = 3.6$ T, $B_p = 6$ T and $B_A = 3 \times 10^{-4}$ T.

$$t(E, qU_a) \longrightarrow t_i^*(E, qU_a) = \int_0^{\theta_{\text{tr}}(E, qU_a)} \omega_i(\theta) d\theta \quad (4.18)$$

where now $\omega_i(\theta) = P_i(\theta) \sin \theta$ and $P_i(\theta)$ is the probability that an electron that is born with polar angle θ leaves the source after i (inelastic) scatterings [Gro15]. Therefore, setting D to a non-negative value is equivalent to resolve an angular dependent effect for electrons that scatter D times. A transmission function computed with non negative D is called *detailed transmission function*. I chose the representative values of $D = -1$, $D = 1$ and $D = 5$ for my simulation.

A stands for *angular change* which is a boolean parameter in the configuration file of SSC. If $A = \text{true}$ one includes the angular change due to elastic scattering; if $A = \text{false}$ one does not.

The main goal of this comparison is to see if, including a detailed transmission for $D > 0$ scatterings, one gets a better agreement with the convolution code which includes angular effects automatically.

In figure 4.16, 4.17, 4.18 and 4.19 the plots of the comparison are displayed.

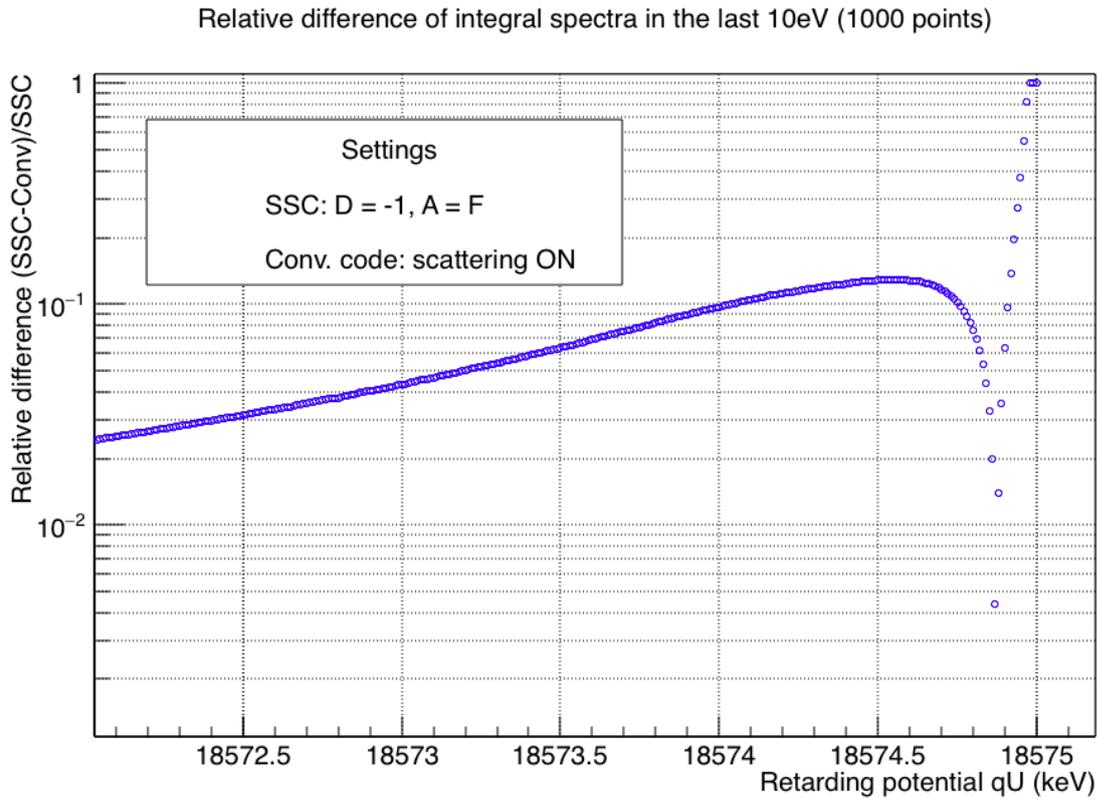
In plot (d) I chose to use $D = 1$ instead of 5 for two reasons. First, the major impact on the shape of the transmission function comes from setting $D = 1$ and increasing its value does not increase the precision since scattering effects takes over (see [Gro15]). Second, the simulation time of the integral spectrum in SSC is already huge including $A = \text{true}$ (\sim a day), therefore a detail transmission for 5 scattering does not add any useful information.

Looking at the plots, one can see that the angular change parameter A has no sizable impact on the comparison. The number of scatterings for which one uses the detailed version of the transmission function really matters. From $D = -1$ to $D = 1$ (or 5) one can see an improvement of 10%. For $D = -1$ the relative difference is greater than 10% in the last eV regardless of angular change. For $D = 1$ (or 5) the relative difference is greater than 1% in the last 0.4 eV regardless of angular change¹⁵.

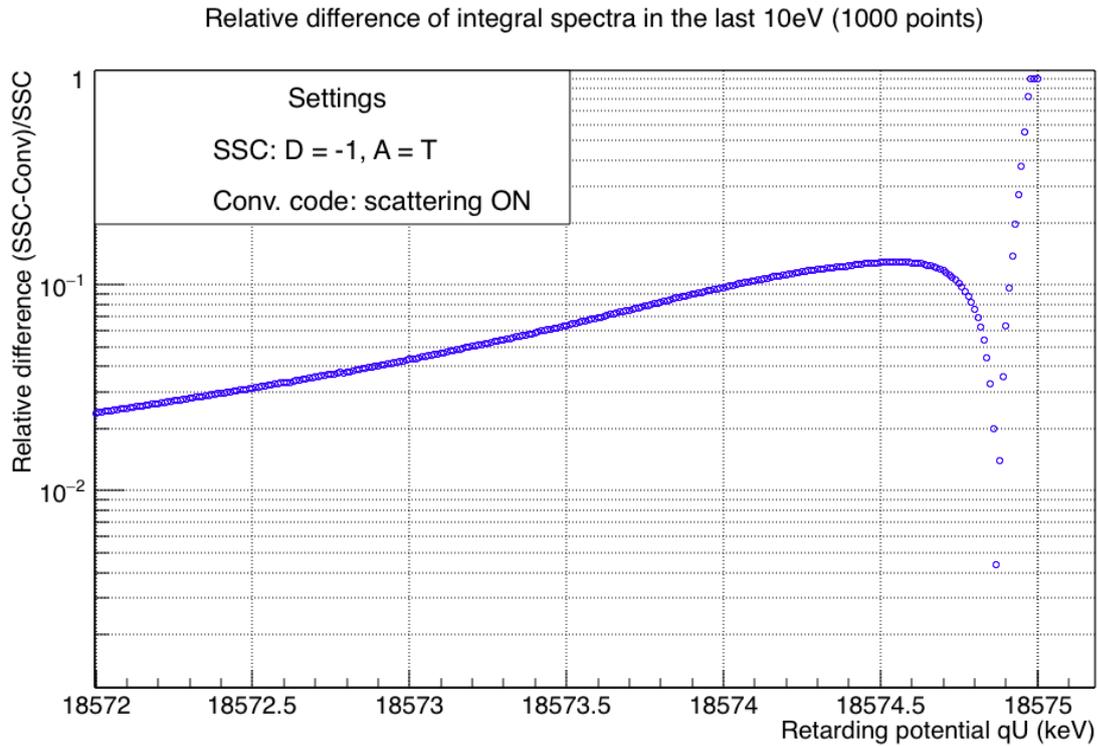
All these results are very promising and suggest that the Convolution code provides a good model for the entire tritium spectrum. However, some issues arose in this study.

First, in SSC there is no such thing as a φ -convolution. Therefore it did not make sense to include it in our comparison. Although it is clear (as we will see in the next section) that the convolution code provides meaningless results without the φ -convolution, which modifies the final angular spectrum only for large polar angles θ (that is, small $\cos \theta$), our comparison was meaningful. That is because, in normal KATRIN mode, only electrons

¹⁵for a sterile neutrino search, the discrepancy very very close to the endpoint is not an issue. Remember that the Convolution code is designed to model the entire spectrum. Far from the endpoint it is precise enough.

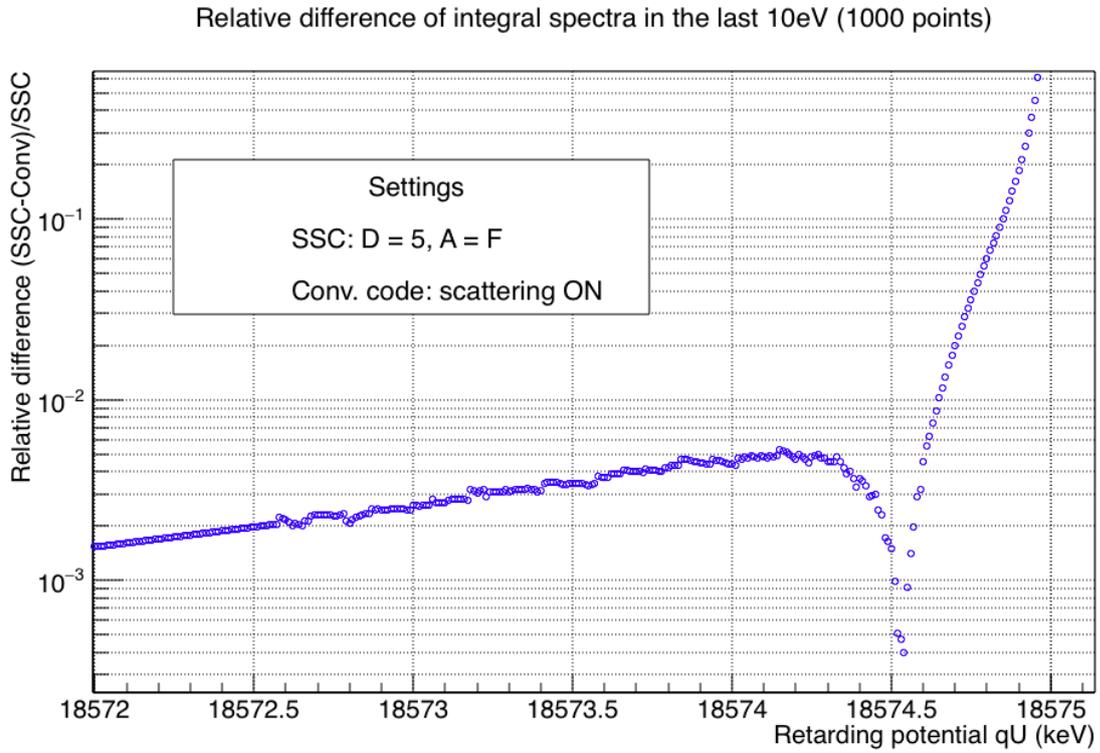


(a)

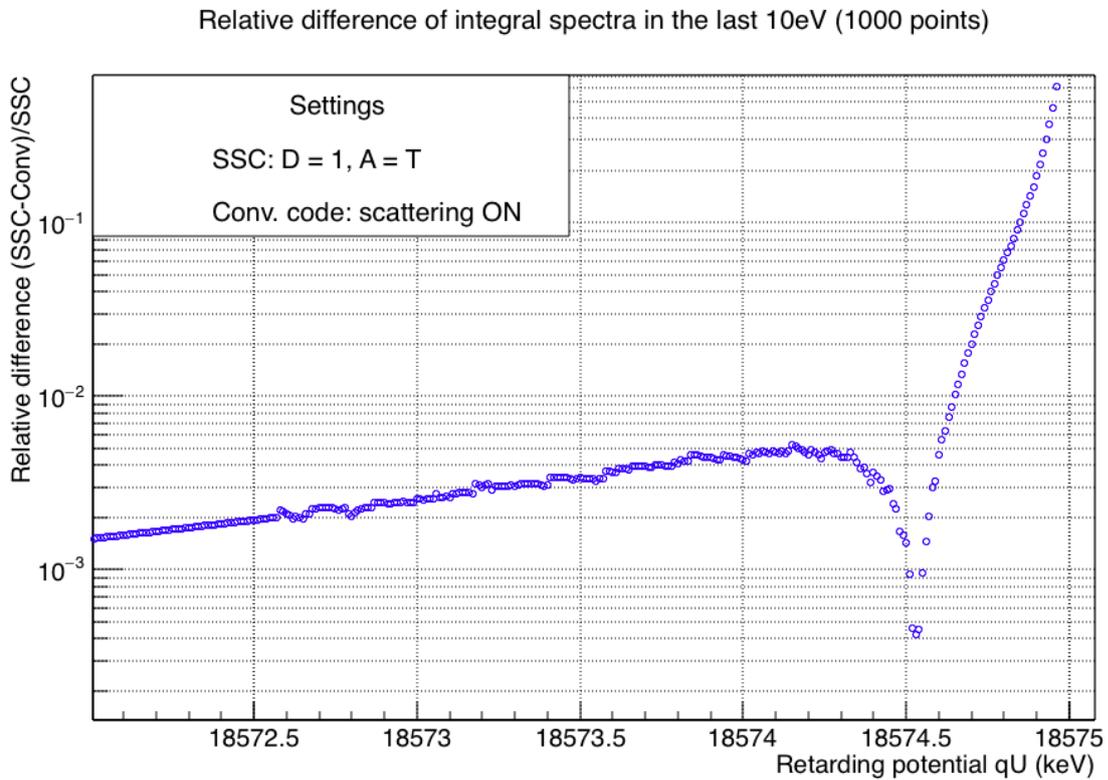


(b)

Figure 4.16: Relative difference of SSC and Convolution code's integral spectra using the analytical transmission function in log scale. (a): no angular change in SSC. (b): including angular change in SSC.



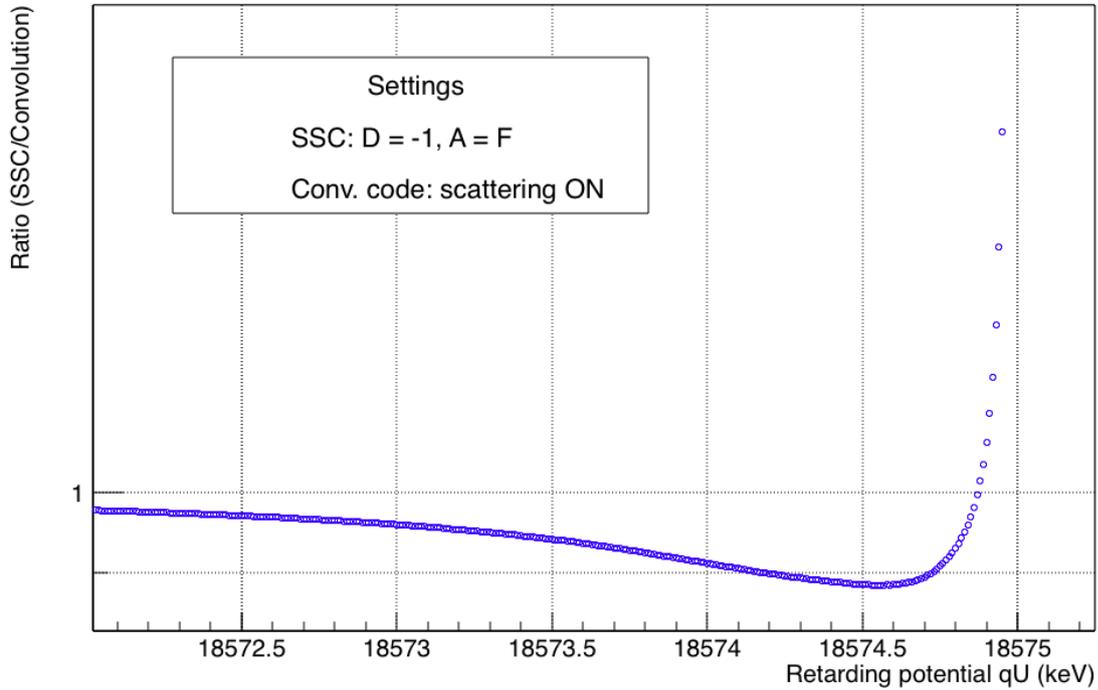
(a)



(b)

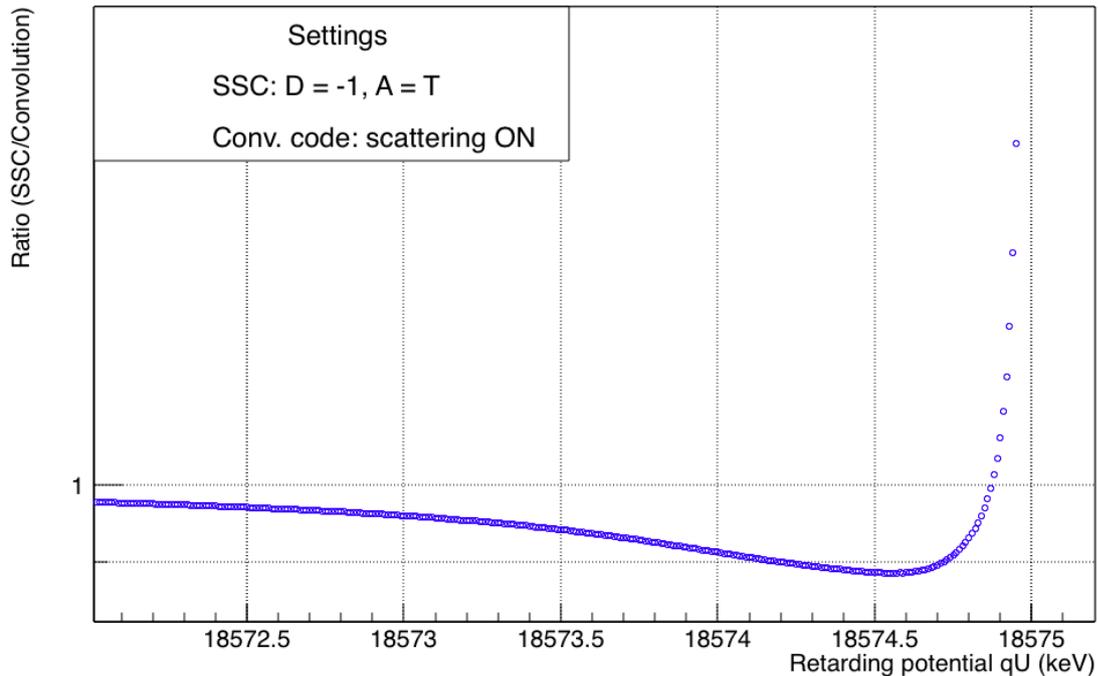
Figure 4.17: Relative difference of SSC and Convolution code's integral spectra using the detailed transmission function in log scale. (a): detailed transmission for 5 scatterings, no angular change in SSC. (b): detailed transmission for 1 scattering, including angular change in SSC.

Ratio of integral spectra in the last 10eV (1000 points)



(a)

Ratio of integral spectra in the last 10eV (1000 points)



(b)

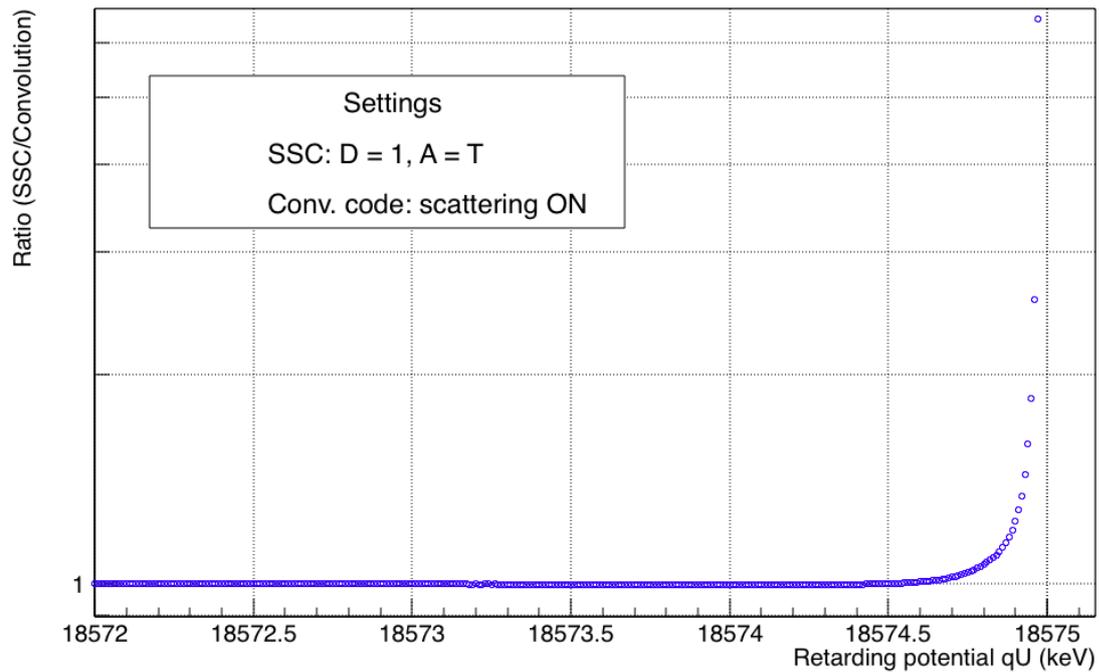
Figure 4.18: Ratio of SSC and Convolution code's integral spectra using the analytical transmission function in log scale. (a): no angular change in SSC. (b): including angular change in SSC.

Ratio of integral spectra in the last 10eV (1000 points)



(a)

Ratio of integral spectra in the last 10eV (1000 points)



(b)

Figure 4.19: Ratio of SSC and Convolution code's integral spectra using the detailed transmission function in log scale. (a): detailed transmission for 5 scatterings, no angular change in SSC. (b): detailed transmission for 1 scattering, including angular change in SSC.

born with $\theta < 51^\circ$ make their way out of the source, and the φ -convolution modifies the angular distribution only for $\theta \gtrsim 65^\circ$ (see figure 4.2).

Second, the whole simulation with the convolution code with a reasonable number of angular and energy bins (~ 100 for both) including the φ -convolution takes days. This represents a major problem for fitting because one would have to run a simulation for each parameter setting¹⁶.

These issues led to the idea of using the convolution code as a generator of response matrices of the source (and eventually the spectrometer) that can be used as a database to later calculate the whole tritium spectra.

This new framework/idea is called **SSC-Sterile** and I am going to discuss it in the next section.

4.4 SSC-Sterile

With the term **SSC-Sterile** I refer to a new way of calculating the model of the tritium spectrum for sterile neutrino searches.

In this section I want to discuss the general idea and the role of the Convolution code in SSC-Sterile.

The underlying idea is that one can separate the KATRIN setup in different sections: source, spectrometer, detector, etc... Each component comes with its own *response* which is just a property of the component and has nothing to do with the tritium spectrum. The response is always given in terms of response "to something", for example, response to electrons with definite energy and/or angle.

Here I will focus on the response to mono-energetic (and subsequently mono-angular) electrons. The response of a component, for example the source, to mono-energetic electrons is given by the energy distribution after the source. It describes how a mono-energetic line is spread after the scatterings in the source. The structure of the response depends on what one is looking for. For example, if one just focuses on the energy and forget about the angles, then for each energy E_I one will get a vector of probabilities $(s_0, s_1, \dots, s_{I-1}, s_I, 0, \dots, 0)$, where s_i = probability that the electron starting with E_I has energy E_i after the source $i = 0, \dots, I$ ¹⁷. The more information one wants to keep track of, the bigger the response becomes (matrix, tensor,...)

Theoretically, having the responses of all components to all variables of interest, starting with a binned differential tritium spectrum, as in the Convolution code, one can apply on it the responses of each component one by one:

$$M = \dots R_{\text{detector}} \times \dots \times R_{\text{spectrometer}} \times R_{\text{source}} \times S \quad (4.19)$$

¹⁶neutrino mass, endpoint, etc ...

¹⁷the i index runs up to I because the electron can only lose energy in the source

where S is the binned initial differential spectrum, R_i 's are the response matrices of each component and M is our final model for the tritium spectrum used for data fitting.

Equation (4.19) is at the moment very vague because neither the dimensions of these objects nor the meaning of " \times " is specified. One can rather think of equation (4.19) in this way: applying all the responses one by one to the initial spectrum one gets the final model for the integral spectrum.

4.4.1 Response matrices of the source

In this section I show how to compute the response matrices of the source and how they are used to build a *spectrum after the source*. This can be used as a guideline on how to add the responses of other components of the experiment.

Response of the source to mono-energetic electrons

Let us discuss first the easier case of mono-energetic electrons. Our goal is to find the response of the source to each energy and build a tritium spectrum after the source¹⁸.

To this goal one can make use of our Convolution code. I am interested in how the source responds to a single energy, therefore I will make use of the method `DoMonoESim` instead of `DoFullSpecSim` in `tristanSourceSim`. Remember that the only differences are that `DoMonoESim` does not have the tritium spectrum as a parameter and it works with a vector of distributions (one per each initial energy).

I will not write pseudo-code here again because once the `DoFullSpecSim` method is understood, it is straightforward to understand `DoMonoESim`. I will just show the mathematical approach which reflects what goes on in the code.

Starting with mono-energetic electrons with energy E_I ¹⁹, the input matrix A for the convolution code will be

¹⁸the REAL goal is to build the response. I show how to build the spectrum for illustration, as this will be the last thing one does after all the responses are provided (see (4.19))

¹⁹I will use capital indices for starting energy

$$A^I = \begin{matrix} E_0 \\ \vdots \\ E_{I-1} \\ E_I \\ E_{I+1} \\ \vdots \\ E_{EB-1} \end{matrix} \begin{pmatrix} \cos \theta_0 & \cos \theta_1 & \cdots & \cos \theta_{TB-1} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ s_I & s_I & \cdots & s_I \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

where s_I is the probability of starting with energy E_I ²⁰ and I have displayed on the side of the matrix the energies and angles the elements correspond to. The matrix A^I represent the binned probability distribution of mono-energetic electrons starting isotropically. Using this matrix as an input to the Convolution code one will output what I called the B matrix which will have the form

$$B^I = \begin{matrix} E_0 \\ \vdots \\ E_I \\ \vdots \\ \vdots \\ E_{EB-1} \end{matrix} \begin{pmatrix} \cos \theta_0 & \cdots & \cos \theta_{TB-1} \\ B_{0,0}^I & \cdots & B_{0,TB-1}^I \\ \vdots & \ddots & \vdots \\ B_{I,0}^I & \cdots & B_{I,TB-1}^I \\ \vdots & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

where B_{ij}^I is the probability that an electron, starting isotropically with energy E_I , has energy E_i and cosine of polar angle $\cos \theta_j$ after the source.

The matrix B^I represent the response of the source to mono-energetic isotropically-born electrons. For each starting energy (a number) there is a matrix:

$$s_I \longrightarrow B^I. \quad (4.20)$$

How can one build the tritium spectrum from this information?

Binning the initial tritium spectrum as I did in 4.2.3, one gets a vector (s_0, \dots, s_{EB-1}) where s_i is the bin content of the i^{th} energy bin, that is, the probability that an electron starts with energy E_i . Repeating the just outlined procedure, one can build a response matrix *for each* starting energy:

²⁰it should be 1 being a mono-energetic distribution. However, its actual value changes if one normalizes the two-dimensional distribution including the angles. Therefore I will keep calling it s_I .

$$\begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{EB-1} \end{pmatrix} \longrightarrow \begin{pmatrix} B^0 \\ B^1 \\ \vdots \\ B^{EB-1} \end{pmatrix} \quad (4.21)$$

If one now weights each response B^I with its probability s_I and sum over all I 's one will get the same result one would get if one used the whole tritium spectrum as an input to the Convolution code²¹. Symbolically:

$$B = \text{ConvolutionCode}(A) \iff \sum_{I=0}^{EB-1} s_I B^I \quad (4.22)$$

where B is the same as in section 4.2.3. Although equivalent, the response matrices method has several advantages. Calculating all these B^I matrices provides a database that can be easily reused. If for instance one would like to fit the spectrum after the source with some measured spectrum, in the case of the Convolution code one would have to *simulate the whole spectrum* for each parameter set (neutrino mass, final state distribution, etc. . .), while using the response matrix one only needs to *bin the differential spectrum* for each parameter set, because all the response matrices are already calculated. This looks like a great advantage!

To show that these two methods are indeed equivalent, I plot the result of both simulations in figures 4.21 and 4.22 and their relative difference in figure 4.23.

In the next sections I will show some plots of the responses to mono-energetic electrons and outline how one can further extend the response matrices method.

Plots of the response functions

In this section I want to show how a typical response function to mono-energetic isotropically-generated electrons looks like. In the previous section I showed how one generates a response matrix B^I from the probability s_I . B^I represents also a binned two-dimensional probability distribution²².

Marginalizing either the angles or the energies, one gains information on the final energy or angular distribution, in fact:

- the vector $(B_0^I, \dots, B_{I-1}^I, B_I^I, 0, \dots, 0)$ with $B_i^I = \sum_{j=0}^{TB-1} B_{ij}^I$ for $i = 0, \dots, I$ and $B_i^I = 0$ for $i = I + 1, \dots, EB - 1$ is the energy

²¹this has been checked and it is true.

²²the key fact to keep in mind is that for mono-energetic electrons the concepts of *response* and *distribution after the source* are actually the same object

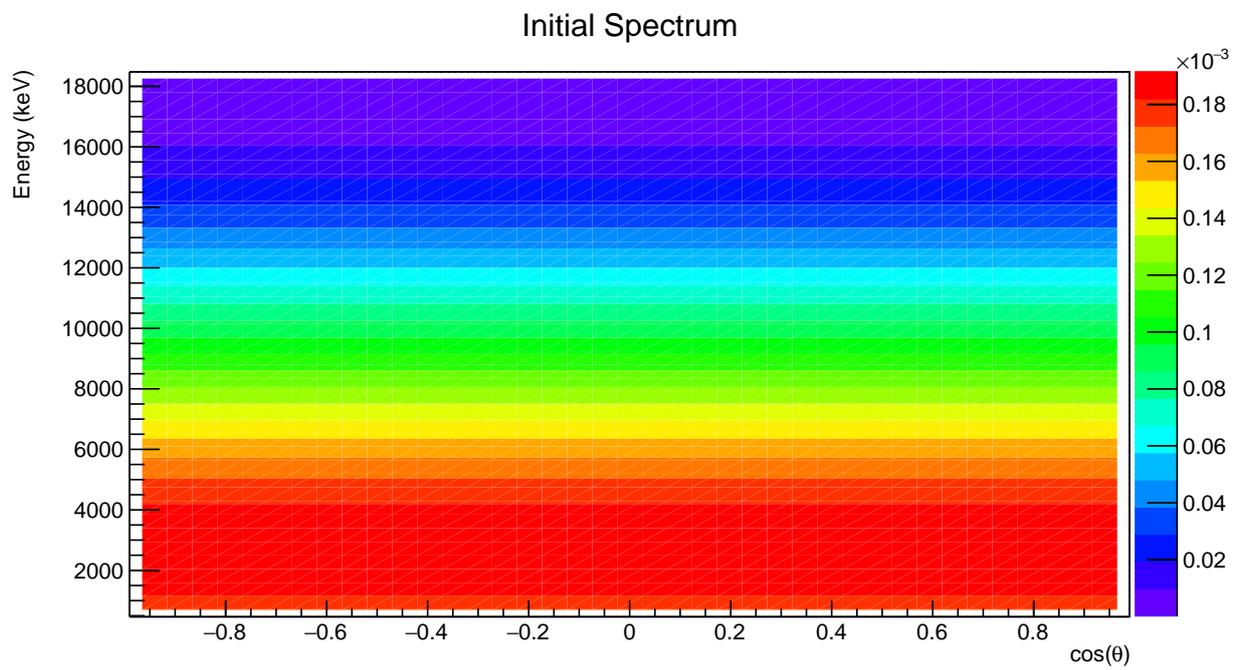


Figure 4.20: Initial spectrum *A*. $EB = TB = 100$. From this plot we can see that the initial spectrum is isotropic and distributed as $d\Gamma/dE$ along the energies.

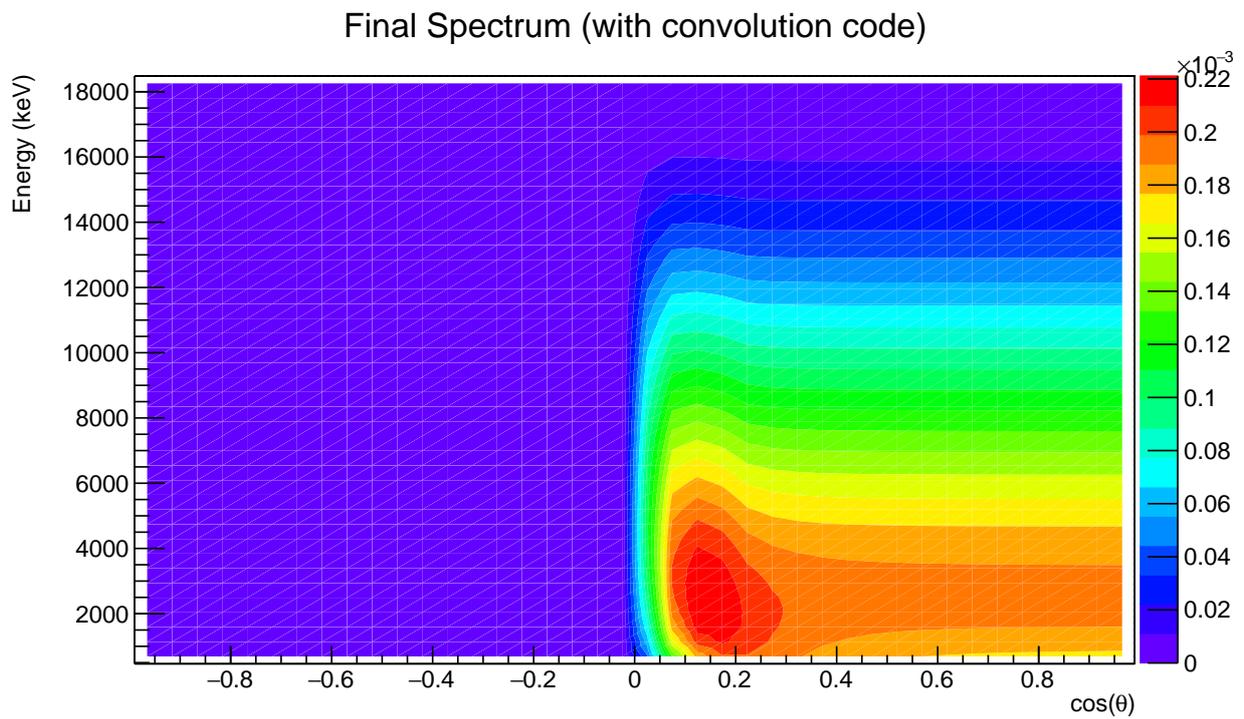


Figure 4.21: Distribution $\text{ConvolutionCode}(A)$ from eq (4.22) where A is displayed in figure 4.20. $EB = TB = 100$. I use a small number of energy and angular bins because the goal is just to show the equivalence of the two methods.

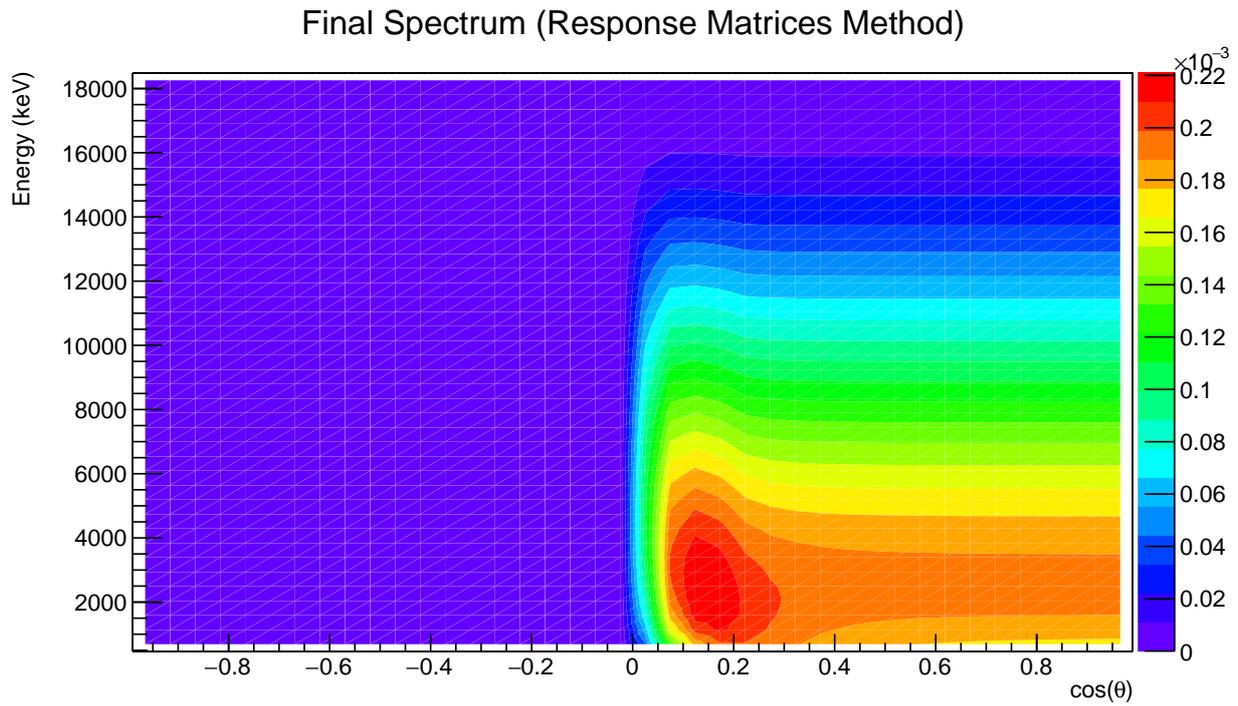


Figure 4.22: Distribution $\sum_{I=0}^{EB-1} s_I B^I$ from eq (4.22) built using the response matrices of the source to mono-energetic electrons. $EB = TB = 10$. No secondary electrons are included. I use a small number of energy and angular bins because the goal is just to show the equivalency of the two methods.

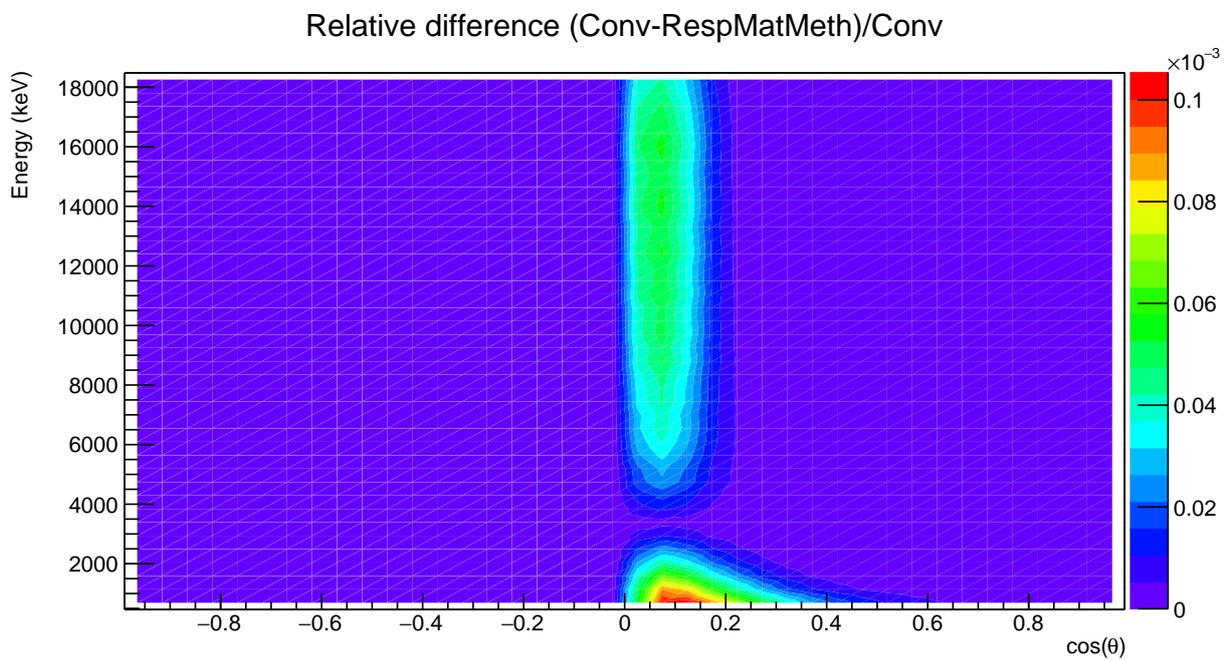


Figure 4.23: Relative difference of the final spectra built with the convolution code and with the response matrices method. $EB = TB = 10$. The relative error is comparable to the tolerance of the simulation ($\varepsilon \sim 10^{-5}$).

distribution after the source marginalizing the angle. More specifically, B_i^I is the probability that an electron, isotropically-generated and with initial energy E_I , has energy E_i after the source.

- the vector $(\Theta_0^I, \dots, \Theta_{EB-1}^I)$ with $\Theta_j^I = \sum_{i=0}^{EB-1} B_{ij}^I$ for $j = 0, \dots, TB - 1$ is the angular distribution after the source marginalizing the energy. More specifically, Θ_j^I is the probability that an electron, isotropically-generated and with initial energy E_I , has cosine of the polar angle $\cos \theta_j$ after the source.

I plot here the marginalized energy and angular distribution after the source for electrons starting with energies $E_1 \sim 6.2$ keV, $E_2 \sim 11.6$ keV and $E_3 \sim 17$ keV. These are just representative values chosen to show how a response function looks like.

Table 4.2: Parameter values used in the simulation of response matrices to mono-energetic electrons.

ρd (cm^{-2})	EB, TB, zB	Effects included	φ -convolution
$3 \cdot 10^{15}$ (Phase-0)	100, 100, 40	Elastic, Excitation, Ionization	ON

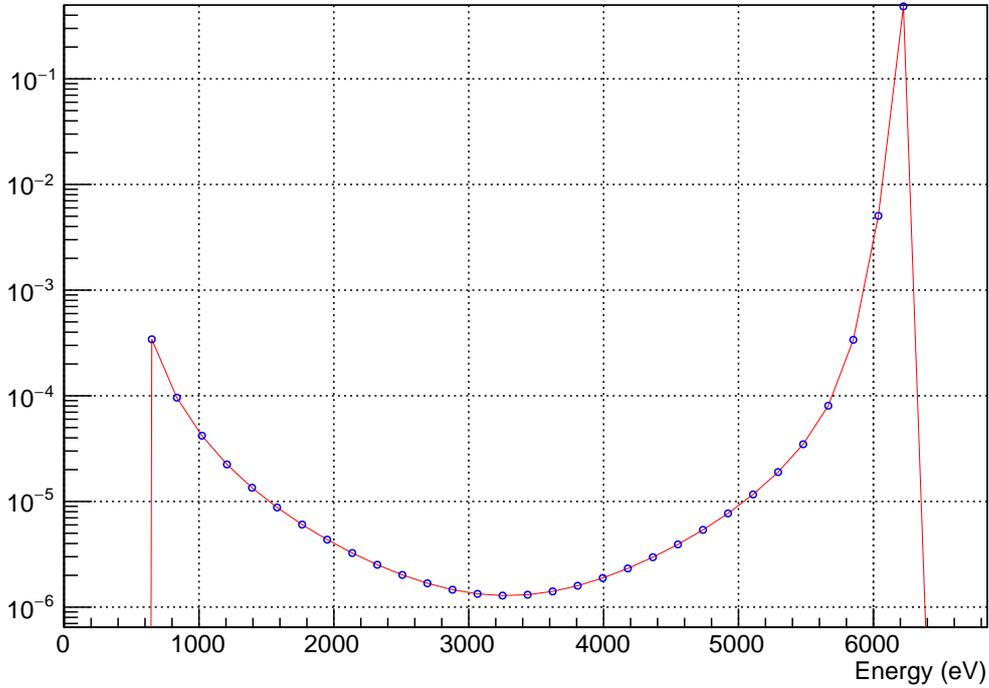
In table 4.2 I display the values I used for our simulation of the response matrices. In figures 4.24, 4.25 and 4.26 one can see the results of the simulation.

Future steps

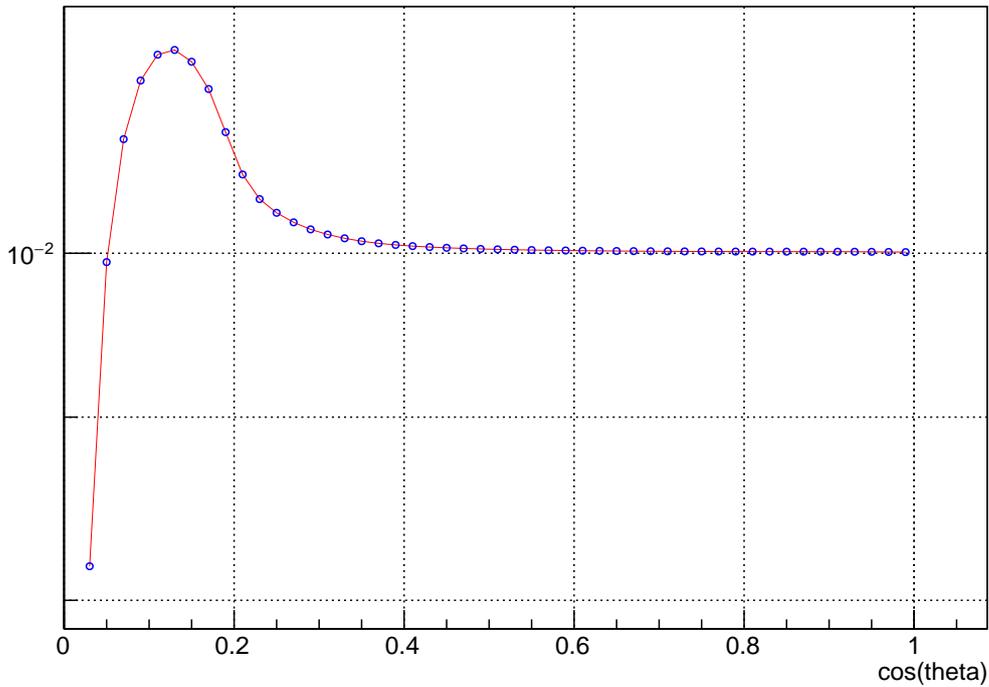
Finally I want to outline how to further extend the response matrices method. Until now this method makes strong use of the assumption on the initial angular distribution of electrons in the source: isotropic distribution. However, electrons back scattering from the rear wall, which then also pass through the source, are not generated isotropically simply because they can just go right. Therefore, it would be useful to create a database of response matrices of the source to mono-energetic *mono-angular* electrons. The extension to an other dimension is theoretically straightforward. Starting from a **non**-isotropic two-dimensional binned distribution

$$A_{\text{non-iso}} = \begin{pmatrix} s_{00} & \dots & s_{0,TB-1} \\ \vdots & \ddots & \vdots \\ s_{EB-1,0} & \dots & s_{EB-1,TB-1} \end{pmatrix} \quad (4.23)$$

where s_{ij} is the bin content of bin (i, j) , one can define the matrix $A_{\text{non-iso}}^{IJ}$ in the following way

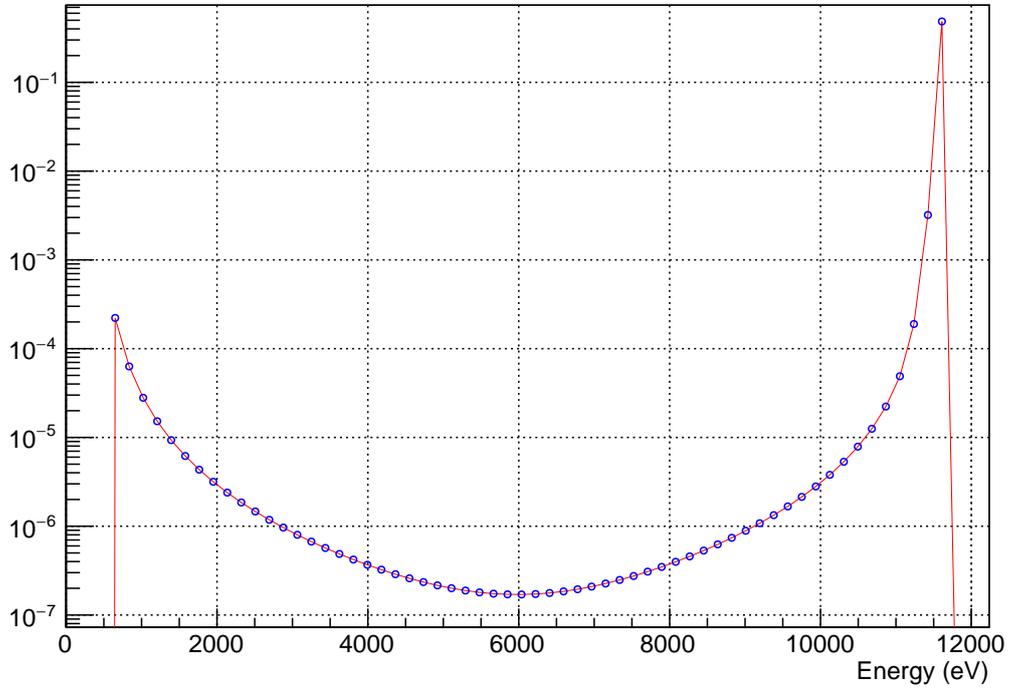


(a)

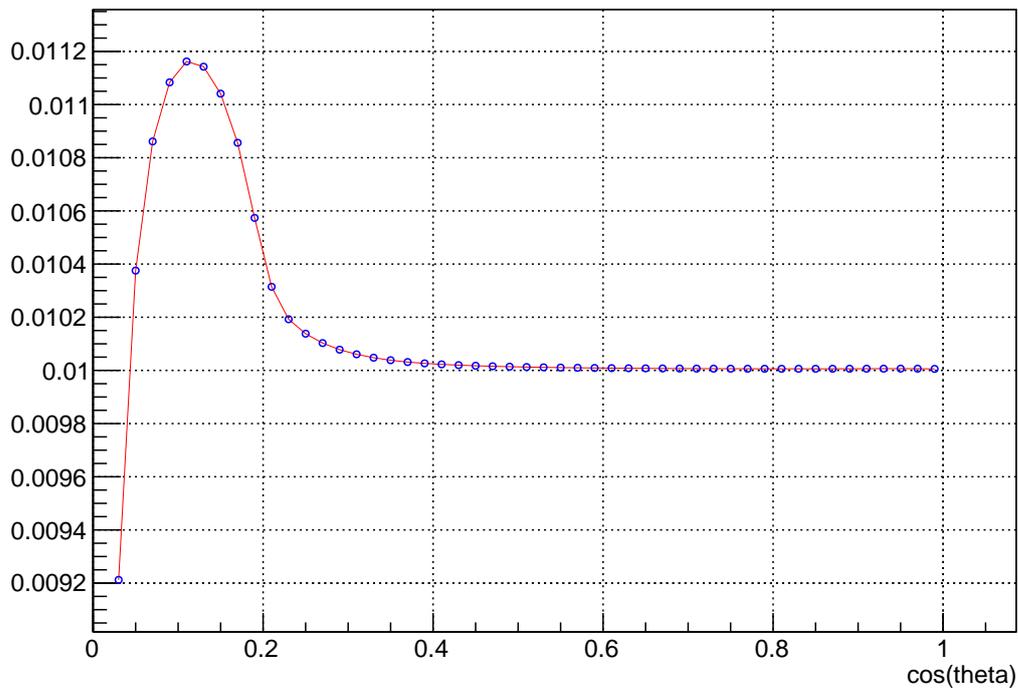


(b)

Figure 4.24: Simulation of the response to mono-energetic isotropically-generated electrons with starting energy $E_I \sim 6.2$ keV. φ -convolution: included. (a): energy distribution (response) marginalizing the angle variable in log scale. (b): angular distribution (response) marginalizing the energy variable in log scale.

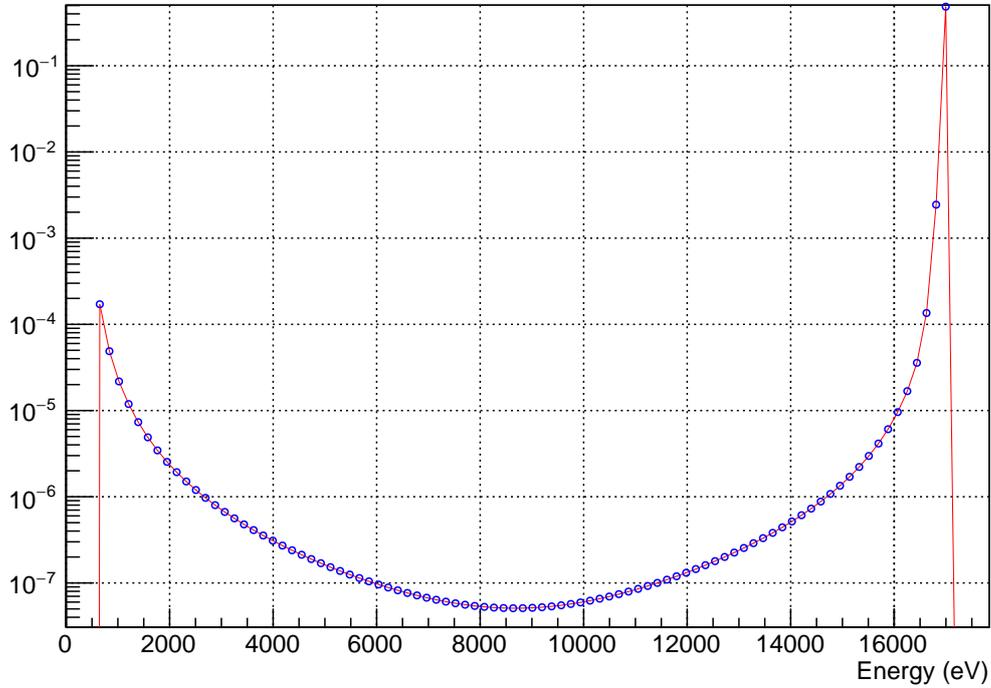


(a)

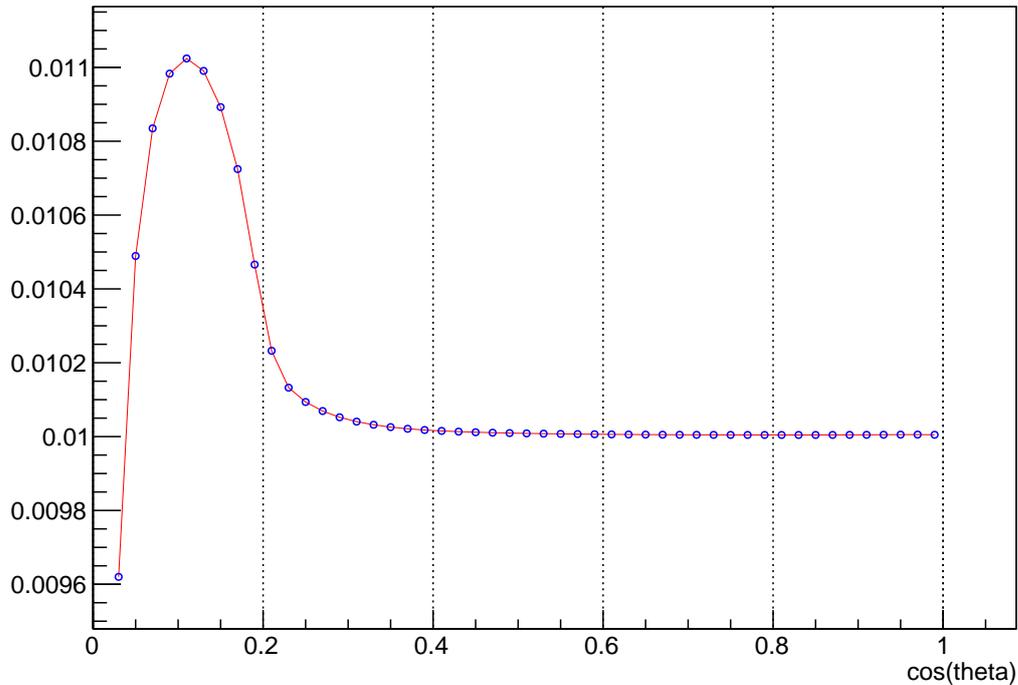


(b)

Figure 4.25: Simulation of the response to mono-energetic isotropically-generated electrons with starting energy $E_I \sim 11.6$ keV. φ -convolution: included. (a): energy distribution (response) marginalizing the angle variable in log scale. (b): angular distribution (response) marginalizing the energy variable.



(a)



(b)

Figure 4.26: Simulation of the response to mono-energetic isotropically-generated electrons with starting energy $E_I \sim 17$ keV. φ -convolution: included. (a): energy distribution (response) marginalizing the angle variable in log scale. (b): angular distribution (response) marginalizing the energy variable.

$$A_{\text{non-iso}, ij}^{IJ} = \begin{cases} s_{ij} & \text{if } i = I \text{ and } j = J \\ 0 & \text{otherwise} \end{cases}, \quad (4.24)$$

that is, $A_{\text{non-iso}}^{IJ}$ has only one non-zero element and that is s_{IJ} . $A_{\text{non-iso}}^{IJ}$ is a mono-energetic and mono-angular two-dimensional binned distribution. Giving it as an input to the Convolution code one will get a matrix B^{IJ} which is the response and corresponds to s_{IJ} . Doing this for all I 's and J 's one gets a matrix of matrices:

$$\begin{pmatrix} s_{00} & \dots & s_{0, TB-1} \\ \vdots & \ddots & \vdots \\ s_{EB-1,0} & \dots & s_{EB-1, TB-1} \end{pmatrix} \rightarrow \begin{pmatrix} B^{00} & \dots & B^{0, TB-1} \\ \vdots & \ddots & \vdots \\ B^{EB-1,0} & \dots & B^{EB-1, TB-1} \end{pmatrix} \quad (4.25)$$

If one now, as previously done, weights each response B^{IJ} with its probability s_{IJ} and sums over all I 's and J 's one gets the final two-dimensional binned spectrum after the source:

$$\boxed{\sum_{I=0}^{EB-1} \sum_{J=0}^{TB-1} s_{IJ} B^{IJ}} \quad (4.26)$$

Although the extension to mono-energetic and mono-angular electrons was straightforward, from the computational time point of view this is really an issue. To put that into perspective one would now have to launch the Convolution code $EB \times TB$ times to calculate all the B^{IJ} matrices. If one uses the usual values for EB and TB (~ 100) this might take months, if not years. Also the size of the data grows considerably.

Perhaps one could bypass this issue reducing the number of angular bins. As one can see from figures 4.24, 4.25 and 4.26, the angular distribution is flat after $\cos \theta \sim 0.5$, therefore one might not want to bin that part. One could reduce the total number of angular bins and have a finer binning for $\cos \theta \in [0, 0.5]$. All these hypotheses are currently under investigation and object of discussion.

4.4.2 Conclusion

In this chapter I presented in details the Convolution code and its structure. The major modifications I brought during this work to the code were the inclusion of the φ -convolution and the `AnalyzPlane` method. I compared the integral spectra from SSC and the Convolution code and the result confirmed that the Convolution code can be used to produce a model of the tritium spectrum for sterile neutrino search. It was observed that the inclusion of the φ -convolution into the code slows down considerably

the computation time of the integral spectra. This lead to the idea of using the Convolution code as a *response matrices generator* for the source. I explained how to build a database of response matrices to mono-energetic electrons and how to build a *spectrum after the source* out of them. I also discussed the possibility of calculating response matrices to mono-energetic *and mono-angular* electrons for a more precise modelling of the spectrum, highlighting its advantages and disadvantages.

Bibliography

- [Cha14] J. Chadwick. Intensitätsverteilung im magnetischen Spektrum der Betastrahlen von Radium B+C. *Verhandlungen der deutschen physikalischen Gesellschaft*, vol. 16, p. 383, 1914.
- [Pau30] W. Pauli. Sehr geehrte radioaktive Damen und Herren. *published in R.Kronig and V. Weisskopf (Eds.)*, 1930.
- [Fer34] E. Fermi. Versuch einer Theorie der beta-Strahlen. *Zeitschrift für Physik A*, vol. 88, p. 161, 1934.
- [Cow56] C.L. Cowan, F. Reines, F.B. Harrison, H.W. Kruse, A.D. McGuire. Detection of the free Neutrino: a Confirmation. *Science*, vol. 124, no. 3212, pp. 103-104, 1956.
- [Hub15] A. Huber. *Search for keV-Scale Sterile Neutrinos with KATRIN*. Master thesis, KIT, 2015
- [Kod01] K. Kodama et al. Observation of Tau Neutrino Interactions. *Phys. Rev. Lett. B*, vol. 504, no. 3, pp. 218-224, 2001
- [ALE06] The ALEPH Collaboration et al. Precision electroweak measurements on the Z resonance. *Physics Reports* 427, pp. 257-454, 2006
- [Gol58] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of Neutrinos. *Phys. Rev.* 109, 1015, 1958
- [Dav94] R. Davis. A Review of the Homestake Solar Neutrino Experiment. *Prog. Part. Nucl. Phys.*, vol. 32, pp. 13-32, 1994
- [Bah05] J.N. Bahcall, A.M. Serenelli, S. Basu. New Solar Opacities, Abundances, Helioseismology, and Neutrino Fluxes. *Astrophys. Journal*, vol. 621, pp. 85-88, 2005.
- [GAL99] GALLEX Collaboration. GALLEX Solar Neutrino Observations: Results for GALLEX IV. *Phys. Rev. Lett. B*, vol. 447, no. 1-2, pp. 127-133, 1999.

- [SAG02] SAGE Collaboration. Solar Neutrino Flux Measurements by the Sovjet-American Gallium Experiment (SAGE) for half the 22-Year Solar Cycle. *Exp. Theo. Phys.*, vol. 95, pp. 181-193, 2002.
- [Kam96] Kamiokande Collaboration. Solar Neutrino Data Covering Solar Cycle 22. *Phys. Rev. Lett.*, vol. 77, pp. 1683-1686, 1996.
- [SNO02] SNO Collaboration. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, vol. 89, no. 011301, 2002.
- [Giu07] C. Giunti and C. W. Kim. *Fundamentals of Neutrino Physics and Astrophysics*. Oxford University Press. 2007
- [Sco35] F. A. Scott, Energy Spectrum of the γ -Rays of Radium E, *Phys. Rev.*, vol. 48, pp. 391395, 1935.
- [Pet13] S. T. Petcov. The Nature of Massive Neutrinos. *Adv. High Energy Phys.*, vol. 2013, id.852987, 2013.
- [Kin04] S. F. King. Neutrino Mass Models. *Rept.Prog.Phys.* 67, pp. 107-158, 2004
- [Bon01] J. Bonn et al. The Mainz Neutrino Mass Experiment. *Nucl. Phys. B Proc. Suppl.*, vol. 91, no. 1-3, pp. 273-279, 2001.
- [Lob99] V.M. Lobashev et al. Direct search for Mass of Neutrino and Anomaly in the Tritium beta-Spectrum. *Phys. Rev. Lett. B*, vol. 460, no. 1-2, pp. 227-235, 1999.
- [Kin13] S. F. King and C. Luhn, Neutrino mass and mixing with discrete symmetry, *Rep. Prog. Phys.*, vol. 76, no. 056201, 2013.
- [Alp48] R. Alpher, R. Herman. On the Relative Abundance of the Elements. *Phys. Rev.*, vol. 12, pp. 1737-1742, 1948.
- [Pen65] A.A. Penzias, R.W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophys. Journal*, vol. 142, no. 1, pp. 419-421, 1965.
- [Les12] J. Lesgourgues, S. Pastor. Neutrino Mass from Cosmology. *Adv. in High Energy Phys.*, vol. 2012, no. 608515, 2012.
- [Pla14] Planck Collaboration. Planck 2013 Results. XVI. Cosmological Parameters. *Astronomy and Astrophys.*, vol. 571, p. 66, 2014.
- [Sch13] M. Schlösser, Accurate calibration of the Raman system for the Karlsruhe Tritium Neutrino Experiment, Ph.D. dissertation, KIT, 2013. [Online]. Available: <http://nbn-resolving.org/urn:nbn:de:swb:90-349674>

- [MAJ14] MAJORANA Collaboration. The Majorana Demonstrator neutrinoless Double beta-Decay Experiment. *Adv. in High Energy Phys.*, vol. 2014, pp. 1-18, 2014.
- [GER06] GERDA Collaboration. Status of the GERmanium Detector Array (GERDA) for the Search of neutrinoless Double beta-Decay of ^{76}Ge at LNGS. *Prog. Part. Nucl. Phys.*, vol. 57, no. 1, pp. 241-250, 2006
- [GER13] GERDA Collaboration. Results on Neutrinoless Double beta-Decay of ^{76}Ge from GERDA Phase I. *Phys. Rev. Lett.*, vol. 111, p. 122503, 2013.
- [Dre13] G. Drexlin, V. Hannen, S. Mertens, and C. Weinheimer. Current Direct Neutrino Mass Experiments. *Adv. High Energy Phys.* 2013, 293986 (2013).
- [Sev06] N. Severijns, M. Beck, and O. Naviliat-Cuncic, Tests of the standard electroweak model in nuclear beta decay, *Reviews of Modern Physics*, vol. 78, no. 3, pp. 991-1040, 2006
- [Rob88] R. G. H. Robertson and D. A. Knapp, Direct measurement of neutrino mass, *Annual Review of Nuclear and Particle Science*, vol. 38, p. 185, 1988
- [Mas07] S.S.Masood, S.Nasri, J.Schechter, M.A.Tortola, J.W.F. Valle, and C. Weinheimer, "Exact relativistic β decay endpoint spectrum," *Physical Review C*, vol. 76, no. 4, Article ID 045501, 2007.
- [KAT05] KATRIN Collaboration. KATRIN Design Report 2004. *FZKA Scientific Report 7090*, 2005.
- [Gro15] S. Groh. Modeling of the Response Function and Measurement of Transmission Properties of the KATRIN Experiment. *PhD thesis, Karlsruhe Institute of Technology, IEKP*, 2015.
- [Wan13] N. Wandkowsky. Study of background and transmission properties of the KATRIN spectrometers. *Ph.D. dissertation, KIT*, 2013. [Online]. Available: <http://nbn-resolving.org/urn:nbn:de:swb:90-366316>
- [Bab14] M. Babutzka. Design and development for the Rearsection of the KATRIN experiment. *Ph.D. dissertation, KIT*, 2014
- [Dos08] N. Doss and J. Tennyson. Excitations to the electronic continuum of $^3\text{HeT}^+$ in investigations of the T_2 beta-decay experiments. *J. Phys. B: At. Mol. Opt. Phys.* 41 (2008) 125701, 2008.
- [Hoet09] M. Hötzel. Berechnung von KATRIN Messspektren unter Einbeziehung der fensterlosen gasförmigen Tritiumquelle. *Masters thesis, KIT*, 2009.

- [Mer12] S. Mertens. Study of Background Processes in the Electrostatic Spectrometers of the KATRIN experiment. *Doctoral thesis. KIT*, 2012.
- [Dos06] N. Doss et al. Molecular effects in investigations of tritium molecule β -decay endpoint experiments. *Phys. Rev. C* 73 (2006) 025502, 2006.
- [Kop11] J. Kopp, M. Maltoni and T. Schwetz. Are there sterile neutrinos at the eV scale? *Phys. Rev. Lett.* 107:091801, 2011.
- [Men11] G. Mention, M. Fechner, Th. Lasserre, Th. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau. Reactor antineutrino anomaly. *Phys. Rev. D* 83, 073006, 2011
- [Aba12] K.N. Abazajian et.al. Light Sterile Neutrinos: A White Paper. *arXiv:1204.5379v1*, 2012.
- [Min07] The MiniBooNE Collaboration. A Search for Electron Neutrino Appearance at the $\Delta m^2 \sim 1 \text{ eV}^2$ Scale. *Phys. Rev. Lett.* 98:231801, 2007
- [Bas17] M. Bass. The Short Baseline Neutrino Oscillation Program at Fermilab. *arXiv:1702.00990v1* 2017
- [Jun96] G. Jungman, M. Kamionkowski, K. Griest. Supersymmetric Dark Matter *Phys.Rept.* 267 (1996) 195-373, 1996
- [Boy12] A. Boyarsky, D. Iakubovskiy, O. Ruchayski. Next Decade of Sterile Neutrino Studies. *Phys. of the Dark Univ., vol. 1, no. 1, pp. 136-154*, 2012.
- [Bul14] E. Bulbul et al. Detection of an Unidentified Emission Line in the Stacked X-Ray Spectrum of Galaxy Clusters. *The Astrophys. Journal, vol. 789, pp. 13-23*, 2014.
- [Boy14] A. Boyarsky, O. Ruchayskiy, D. Iakubovskiy, J. Frans. An unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster. *Phys.Rev.Lett., vol. 113, pp. 251-301*, 2014
- [Mer14] S. Mertens, T. Lasserre, S. Groh, A. Huber, G. Drexlin, C. Weinheimer. Sensitivity of Next-Generation Tritium Beta-Decay Experiments for keV-Scale Sterile Neutrinos. *Journal of Cosmo. and Astropart. Phys., vol. 1502, p. 020*, 2014.
- [Kle14] M. Kleesiek. A Data-Analysis and Sensitivity-Optimization Framework for the KATRIN Experiment. *PhD thesis, Karlsruhe Institute of Technology, IEKP*, 2014.

- [Groh15] S. Groh. Modeling of the Response Function and Measurement of Transmission Properties of the KATRIN Experiment. *PhD thesis, Karlsruhe Institute of Technology, IEKP, p. 182*, 2015.
- [Ago16] M. Agostini et al. Search of Neutrinoless Double Beta Decay with the GERDA Experiment. *Nuclear and Particle Physics Proceedings Volumes 273-275, April-June 2016, Pages 1876-1882*, 2016.
- [Ber03] L. Bergström and A. Goobar. Cosmology and particle astrophysics. *2nd ed. Berlin: Springer*, 2003.
- [Pra11] M. Prall. The Katrin experiment and the pre-spectrometer at reduced retarding potential. *Progress in Particle and Nuclear Physics 66.2 (2011). Particle and Nuclear Astrophysics International Workshop on Nuclear Physics, 32nd Course, pp. 418423*.
- [Gou10] J. Goullon. Installation and commissioning of the monitor spectrometer. *Masters thesis, KIT*, 2010.
- [Har12] F. Harms. Assembly and first Results of the KATRIN Focal-Plane Detector System at KIT. *Masters thesis, KIT*, 2012.
- [Ase00] V. Aseev, A. Belesev, A. Berlev, E. Geraskin, O. Kazachenko, Y. Kuznetsov, V. Lobashev, R. Ostroumov, N. Titov, S. Zadorozhny, Y. Zakharov, J. Bonn, B. Bornschein, L. Bornschein, E. Otten, M. Przyrembel, C. Weinheimer, and A. Saenz. Energy loss of 18 keV electrons in gaseous t and quench condensed d films. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics, vol. 10, no. 1, pp. 3952*, 2000. [Online]. Available: <http://dx.doi.org/10.1007/s100530050525>
- [Sle17] M. Slezàk. TRISTAN simulation and analysis. *TRISTAN workshop November 2017*.
- [KAT18] KATRIN experiment. (2018, Jan. 22). <http://www.katrin.kit.edu/766.php>.
- [Ice18] IceCube Collaboration. (2018, Jan. 22). <http://icecube.wisc.edu/gallery/press/view/2119>.