

Data Analysis Tools and Sterile Neutrino Investigations for the KATRIN Experiment

vorgelegte
Diplomarbeit
von

Sebastian Schams

am Institut für Experimentelle Kernphysik
der Fakultät für Physik

Erstgutachter:	Prof. Dr. G. Drexlin
Zweitgutachter:	Prof. Dr. T. Müller
Betreuender Mitarbeiter:	Dr. M. Steidl

Bearbeitungszeit: 15. Juni 2012 – 15. Juni 2013

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Karlsruhe, den 14. Juni 2013

Contents

1	Introduction	1
2	Neutrino- and Astrophysics	5
2.1	History of Neutrino Physics	5
2.2	Neutrinos and the Standard Model	7
2.3	Neutrino Oscillations	9
2.4	The Measurement of the Neutrino Mass	10
2.4.1	The Neutrino Mass in Cosmology	10
2.4.2	The Neutrino Mass in Astrophysics	11
2.4.3	The Neutrino Mass in Neutrinoless Double- β -Decay	13
2.4.4	Direct Neutrino Mass Measurements with β -Decay	13
2.5	The Influence of the Neutrino Mass on Astrophysics and Cosmology .	14
2.6	Evidences for Sterile Neutrinos	14
3	The Karlsruhe Tritium Neutrino Experiment	17
3.1	Experimental Setup	17
3.1.1	Source	17
3.1.2	Transport Section	18
3.1.3	Spectrometer Section	18
3.1.4	Focal Plane Detector	19
3.2	Measurement Principle	20
3.2.1	Calculation of the Fundamental β -Decay Spectrum	20
3.2.2	MAC-E Filter	21
3.3	KATRIN Simulation	21
3.4	KATRIN Data Analysis	24
4	KATRIN and keV Sterile Neutrinos	27
4.1	Dark Matter	27
4.1.1	Evidences for the Existence of Dark Matter	27
4.1.2	Cold Dark Matter	31
4.1.3	Hot Dark Matter	31
4.1.4	Warm Dark Matter	32
4.2	Sterile Neutrinos	32
4.3	Can KATRIN Measure keV Sterile Neutrinos?	34
4.4	Influence of Systematics	35
4.5	Summary	36

5	Calculations for the β-Spectrum	37
5.1	Fermi Corrections	37
5.2	Radiative Corrections	40
5.3	Full Molecular Final State Distribution	41
5.4	Other Corrections	42
5.5	Implementation	45
5.6	Summary and Results	45
6	Data Analysis for KATRIN	49
6.1	Frequentist and Bayesian Probability	50
6.2	Data Analysis with Maximum Likelihood	51
6.3	Underlying Principle	53
6.3.1	Markov Chain Monte Carlo	54
6.3.2	Metropolis-Hastings Algorithm	55
6.3.3	Proof of Convergence	55
6.4	Gaussian Transition Kernel	57
6.4.1	Implementation and Properties	58
6.4.2	Results	60
6.4.3	Posterior Distribution	67
6.5	Hamiltonian Monte Carlo	69
6.5.1	Concept of Hamiltonian Monte Carlo	70
6.5.2	Implementation	72
6.5.3	Results	73
6.5.4	Influence of the Parameters	79
6.6	Riemannian Manifold Hamiltonian MC	83
6.6.1	Concept of RMHMC	83
6.6.2	Implementation	84
6.6.3	Results	85
6.7	Comparison of the Algorithms	86
6.7.1	Runtime and Effective Sample Size	88
6.7.2	Convergence Rate	89
6.7.3	Sampling Process	89
6.7.4	Autocorrelation	89
6.7.5	Conclusion	89
7	Summary and Outlook	91
A	Implementation of the Corrections for the Tritium β-Spectrum	95
B	Implementation of the MCMC Analysis Methods	97
C	Configuration File for the MCMC Analysis	99
	Deutsche Zusammenfassung	103
	Acknowledgements	109
	Bibliography	111

1. Introduction

The **K**Arlsruhe **T**Ritium **N**eutrino (KATRIN) experiment is a formidable project. Scientists from fifteen institutions in Germany, Russia, Czech Republic, Great Britain and the United States work together on the ambitious goal to answer one of the most important open questions in modern neutrino physics: What is the absolute mass scale of neutrinos?

KATRIN's dimensions are massive: its total length is 70 meters and the spectrometer alone weighs over 200 tons. It is located on the Campus North of the Karlsruhe Institute of Technology (KIT). The ultra-precision experiment requires a mastery of a wide spectrum of science and engineering, involving molecular and nuclear physics, vacuum and cryogenic technology and sophisticated programs for the data analysis. A team of 150 scientists, engineers and students is committed to working hard for the success of this project.

My first involvement with the KATRIN experiment was in the group of Dr. Alan Poon at the Lawrence Berkeley National Laboratory in Berkeley, California (USA). Together with Dr. Markus Hötzel I worked within a student exchange program on the implementation of the Unified Approach by G. Feldman and R. Cousins for the neutrino mass analysis with KATRIN. This method represents an ingenious way to construct correct confidence belts, even in regions close to unphysical values. Though the results were encouraging, it will not be included in this thesis. The results can be found in [Höt12] and [Scha12].

For my diploma thesis that I started roughly a year ago in the summer of 2012, I stayed in the intriguing field of simulation and data analysis. For many years, the existence of a hypothetical particle, the sterile neutrino, has been discussed vividly. It has been suggested in [dVeg⁺11] that an extension of the KATRIN experiment might open the possibility to detect these particles. As a serious dark matter candidate, sterile neutrinos could answer the question what our universe really consists of.

In order to augment KATRIN's abilities towards this search a new, intricate calculation of the tritium β -decay spectrum is necessary. Also, a more sophisticated approach for the analysis of the measured data is required. The fitting routines that are sufficient for the measurement of the mass scale of regular neutrinos cannot deal with too many free parameters. The existence of a sterile neutrino would add a second minimum and further parameters to the likelihood function, so that new analysis methods are in order.

It is the objective of this thesis to deal with both of these endeavors.

Knowing the shape of the tritium β -decay spectrum as accurately as possible is important because the existence of a sterile neutrino would lead to a kink in the run of this curve. For this purpose intricate corrections to the spectrum were calculated and implemented into the KATRIN analysis software.

Until recently, only frequentist methods were used for the evaluation of the measured data of a simulated run. These methods have now been extended by the implementation of Bayesian Markov Chain Monte Carlo (MCMC) methods. These routines offer a variety of advantages for the present-day KATRIN experiment as well as for the future detection of sterile neutrinos. They are very stable and robust, give precise confidence intervals and offer a chance to implement prior information correctly. Their ability to handle a vast amount of free parameters makes them perfect to deal with fits for the sterile neutrino analysis.

This work on MCMC methods was conducted at KIT and at the Massachusetts Institute of Technology, where I stayed in late summer and fall of 2012 in the group of Prof. J. Formaggio. I presented the results at the KATRIN Collaboration meeting 2013 and at the 77th Annual Meeting of the DPG, the German Physical Society [Scha13a] [Scha13b].

This thesis is structured as follows.

The second chapter deals with the fundamental principles of neutrino- and astrophysics that are the motivation for ultra-precision experiments like KATRIN.

In chapter 3 the KATRIN experiment is presented. This discussion is divided into two parts. At first the experimental setup is presented. Then we deal with the measurement principle and the general data analysis.

The fourth chapter extends the scope of this thesis towards sterile neutrinos. The idea of dark matter is outlined and its hypothetical composition is summarized. Subsequent to the introduction of sterile neutrinos as a dark matter candidate, KATRIN's abilities towards their detection are discussed.

A precise knowledge of the spectrum of tritium β -decay is essential to discover these particles. Chapter 5 gives the most important augmentations: Fermi- and Coulomb corrections, radiative corrections and an extended final state distribution for the description of the energy lost due to excitations of the daughter nuclei. A complete spectrum is presented and the different corrections are analyzed.

Chapter 6 deals with the second and most extensive part of this thesis. Three different MCMC methods are presented: MCMC with a Gaussian transition kernel, Hamiltonian Monte Carlo and Riemannian Manifold Hamiltonian Monte Carlo. The pros and cons of all methods are compared and results for simulated KATRIN measurements are demonstrated.

Finally, chapter 7, gives an overview of what we have seen in this diploma thesis, in which the most important results will be summarized.

2. Neutrino- and Astrophysics

This chapter provides some of the basics that are needed to motivate the KATRIN experiment and to understand this thesis. As mentioned in the introduction, the goal of the experiment is to measure the mass of the neutrino. But why do we want to know the mass of one of the lightest particles in the universe?

In order to understand this we first have to comprehend what a neutrino is. Therefore, in the first section of this chapter, the history of the neutrino is illustrated. After that we have a look at their properties and their place in one of the most important theories in modern physics: the standard model (section 2.2). In section 2.3 the scope is extended beyond this theory. It deals with neutrino oscillations that are proof that neutrinos are not massless as described by the standard model. The options to determine its mass are discussed in section 2.4. After that, section 2.5 is dedicated to the effects the precise measurement of the neutrino mass will have on astrophysics. This also answers the opening question about the great importance of such a tiny parameter. To round off this chapter, we have a look at evidences for neutrino-like particles that do not interact weakly and therefore are called sterile neutrinos (section 2.6). These particles will be discussed even further, together with KATRIN's abilities to measure them, in chapter 4.

2.1 History of Neutrino Physics

In the late 1920s and early 1930s there was an energy crisis quite different from what we experience today. Rather than oil, it was about fundamental physics: the law of conservation of energy, one of the very foundations of physics, was shaking. The reasons were observations of the nuclear β -decay. These had shown that the spectrum of the β -electron was continuous, which meant that somehow energy was lost in the decay process. When even Niels Bohr started to lose faith, the Austrian physicist Wolfgang Pauli postulated the existence of a new particle that was created in the β -decay along with the electron. This particle was light, neutral, had spin- $\frac{1}{2}$ and was accountable for the lost energy, which went into its mass and kinetic energy. It was later named *neutrino* (“little neutral object”) by Enrico Fermi. At first it

seemed impossible to verify the existence of such a weakly interacting, light particle¹. Even though its detection did not succeed for another 23 years, it eventually saved the laws of energy and momentum conservation. [Cald01]

The two physicists C. Cowan and F. Reines finally discovered the neutrino in 1953. At a nuclear reactor in Hanford, Washington, they used the ability of protons to capture electron antineutrinos according to the nuclear equation,



which is a kind of “inverse β -decay” with a neutron and a positron in the final state². They measured both, the neutron and the positron, simultaneously and thereby were able to discriminate the signal from background radiation. For their discovery they were awarded the Nobel Prize in Physics in 1995. [GrBu12]

The next big step was the insight that there is more than one neutrino. Danby et al. showed in 1962 that muon neutrinos are different from electron neutrinos. They studied neutrinos that were created in the decay of pions. In interactions, these neutrinos produced μ mesons, unlike the electron neutrinos that produced electrons. For this observation L. Lederman, M. Schwartz and J. Steinberger were rewarded with the Nobel Prize. [Danb⁺62]

Another milestone of neutrino physics are the solar neutrino experiments by R. Davis. In 1970 he observed the chemical reactions in a huge tank full of CCl_4 . Argon atoms were created by the capture of solar electron neutrinos³:



Contrary to expectations, he measured a neutrino flux of only one third of the expected flux. This observation became known as the “solar neutrino problem” and was resolved by the theory of neutrino oscillations (see section 2.3). It states that neutrinos have a mass and can change flavors. [Zube12]

In 1975 Perl et al. found indications for the existence of the last of the three active neutrinos. Observing electron-positron annihilations they reported missing energy that could not be explained by conventional means, but by the production of additional particles [Perl⁺75]. These particles turned out to be the τ -lepton and the corresponding τ -neutrino. The discovery of the lepton was awarded with a Nobel Prize as well. In 2000 the DONUT experiment (Direct Observation of NU Tau) published the first detection of the tau neutrino at Fermilab.

For everyone who was hoping to get another piece of the Nobel Prize cake that was being cut after most of these discoveries, the subsequent LEP experiments were a small disappointment. From 1993 to 2006 the width of the Z^0 -resonance was measured directly at CERN. This led to a value of 2.93 ± 0.05 for the number of light ν types. [Voge10] [Grou07]

In the last thirty years a variety of experiments with atmospheric, solar, reactor and

¹W. Pauli, himself, said: “I have done a terrible thing, I have postulated a particle that cannot be detected.”. [Sutt92]

²Originally, they considered using a nuclear explosion in order to create a neutrino pulse that was stronger than the background radiation. [GrBu12]

³The major fusion processes in the Sun are the proton-proton chain and the CNO cycle. Among these the proton-proton chain is dominant. In all processes only electron neutrinos are created.

accelerator neutrinos have been conducted. They confirmed two major discoveries. For one thing the neutrinos do have a very small but finite mass and, on the other hand, there really is neutrino mixing, as suggested by B. Pontecorvo. In 2012 the last of the three neutrino mixing angles was determined in China at the Daya Bay experiment [Daya12]. The mass, however, is still not known precisely because no model-independent measurement provided the needed sensitivity. With KATRIN this might change.

But not everything that was measured by these experiments could be explained right away. The Gallium and reactor anomalies, for example, are results that, for their proper explanation, require a new particle: the sterile neutrino. This particle is even harder to detect than the regular neutrino and could explain the dark matter phenomenon⁴. These results show once again that neutrino physics still provides plenty of exciting areas of research.

2.2 Neutrinos and the Standard Model

The standard model is one of the very foundations of particle physics. It describes the elementary particles that make up all baryonic matter and the gauge bosons that are the force carriers between them. In 2012 the last particle of the standard model was - in all likelihood - discovered at CERN: the Higgs boson that is responsible for the mass of all massive particles. A conceptual illustration of this theory is depicted in figure 2.1. The gauge bosons are the photon (carrier of the electromagnetic force), the gluon, which is responsible for the strong force, and the force carriers of the weak force, Z^0 and W^\pm . Not included is the carrier of the gravitational force, as gravity is not part of the standard model. The Higgs boson, with an assumed mass of 125 GeV, is the particle that is responsible for the mass of all other elements [CMS-12].

There are six quarks, up and down, charm and strange, top and bottom, and six corresponding antiquarks. The quarks are spin- $\frac{1}{2}$ fermions and carry an electric charge of $-\frac{1}{3}$ or $+\frac{2}{3}$ and a color charge. Up to this point quarks have only been observed in bound states, either as baryons (net quark number of three) or mesons (quark-antiquark pair with a net quark number of zero). The net quark number is assumed to be constant; however, quark-antiquark pairs can be created or destroyed. [CoGr07]

Leptons are the third group of particles and include the neutrinos. They are, like quarks, spin- $\frac{1}{2}$ fermions. There are three charged leptons: electron e^- , muon μ^- and tau τ^- and, as predicted by the Dirac equation, there is an antiparticle with opposite charge for each of them: e^+ , μ^+ and τ^+ . Of these only e^- and e^+ are stable. One observes that the charged leptons are associated with corresponding neutrinos, the electron neutrino ν_e , the muon neutrino ν_μ and the tau neutrino ν_τ . There are also three anti neutrinos: $\bar{\nu}_e$, $\bar{\nu}_\mu$ and $\bar{\nu}_\tau$. [CoGr07]

Just like the charged leptons, the neutrinos are spin- $\frac{1}{2}$ fermions. They carry neither electric, nor color charge and so far only left-handed neutrinos (helicity $\mathcal{H} = -1$) and right-handed antineutrinos ($\mathcal{H} = +1$) have been observed. [Zube12]

Even though the standard model is a very convenient theory, it has its shortcomings. As mentioned above, it does not include gravity, only the three other fundamental

⁴For more information about sterile neutrinos see section 2.6 and chapter 4 of this thesis.

Three Generations
of Matter (Fermions)

	I	II	III		
mass→	2.4 MeV	1.27 GeV	171.2 GeV	0	? GeV/c ²
charge→	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
spin→	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
name→	u up	c charm	t top	γ photon	H Higgs boson
	4.8 MeV	104 MeV	4.2 GeV	0	
	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
Quarks	d down	s strange	b bottom	g gluon	
	<2.2 eV	<0.17 MeV	<15.5 MeV	91.2 GeV	
	0	0	0	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	Z⁰ weak force	
	0.511 MeV	105.7 MeV	1.777 GeV	80.4 GeV	
	-1	-1	-1	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
Leptons	e electron	μ muon	τ tau	W[±] weak force	Bosons (Forces)

Figure 2.1: An illustration of the standard model of particle physics. Not included is the mass of the Higgs boson. This particle is assumed to have been detected at CERN in 2012. Its mass is supposed to be 125 GeV [CMS-12]. [webs12]

forces. Another point of criticism is that there are 18 free parameters that need to be measured and are not predicted by the theory [Zube12]. Important for this thesis is another deficit: standard model neutrinos are massless. However, there are many experimental results that indicate that neutrinos have a mass. One of the first of these observations was the fact that neutrinos oscillate.

2.3 Neutrino Oscillations

The observation of neutrino oscillations was one of the first indicators for physics beyond the standard model. Less than 50 years ago, R. Davis observed solar neutrinos and measured the flux of electron neutrinos to be about one third of the flux predicted by the standard solar model. This became known as the “solar neutrino problem” and caused many questions. Did this mean that it was not completely understood how the sun creates energy? Or did the neutrinos get lost on the way? The answer is startling: the neutrinos do not get lost on the way, they simply change their flavor - from electron neutrinos to muon and tau neutrinos. This change is known as *neutrinos oscillations* and implies that neutrinos do have a mass, because the separate mass eigenstates need to have different masses. The experiments that measured the solar electron neutrino flux were not sensitive to these other flavors. To test the hypothesis of neutrino oscillations the Sudbury Neutrino Observatory (SNO) was built in Canada. Its detector was filled with 1000 tons of heavy water and was sensitive to all three types of neutrino flavors. In the beginning of this millennium the SNO collaboration managed to resolve the solar neutrino problem by the demonstration of neutrino oscillations [SNO-01]. But how do these oscillations work?

Neutrino oscillations are a quantum mechanical interference phenomenon with close analogies to the mixing of left-handed quarks. The origin is the fact that the neutrino mass eigenstates are not equal to the flavor eigenstates of neutrinos. The flavor eigenstates $\nu_{e,\mu,\tau}$ can be written as a superposition of the mass eigenstates $\nu_{1,2,3}$. This is elucidated with the PMNS-matrix that was introduced by Z. Maki, M. Nakagawa and S. Sakata and is depicted in eq. (2.3). The P in PMNS stands for B. Pontecorvo, who predicted the oscillations as early as 1957. [Mach05]

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (2.3)$$

U is a unitary, Hermitian matrix that includes three mixing angles and a CP-violating phase δ . According to [Mach05] it can be written as:

$$\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \cdot \begin{pmatrix} c_{13} & 0 & s_{13}e^{i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \cdot \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.4)$$

where the abbreviations $s_{ij} = \sin \theta_{ij}$ and $c_{ij} = \cos \theta_{ij}$ were used. The first matrix describes the transition between μ and τ neutrinos, the second matrix the transition between e and τ neutrinos and the third matrix the transition between e and μ neutrinos. Possible Majorana CP violating phases can be described by two additional

phase factors, which are not depicted in eq. (2.4).

With the use of this identity the electron neutrino at time $t = 0$ can be described with the wave function

$$\Psi(0) = |\nu_e\rangle = U_{e1} |\nu_1\rangle + U_{e2} |\nu_2\rangle + U_{e3} |\nu_3\rangle. \quad (2.5)$$

The eigenfunctions of the mass operator are the propagating states and after a certain time Δt the evolved state can be written as

$$\Psi(x) = \sum_{j=1}^3 e^{i(p_j x - E_j t)} U_{ej} |\nu_j\rangle, \quad (2.6)$$

where E denotes the energy, p the momentum and $x = ct$. The flavor of the $|\nu_e\rangle$ changes after a certain propagation length with a non-zero probability, which can be calculated. [Mach05]

From solar and atmospheric neutrino oscillation observations the square of the mass differences can be determined to be [Barn⁺96]

$$\begin{aligned} \Delta m_{12}^2 &= (7.59 \pm 0.20) \cdot 10^{-5} \text{ eV}^2, \\ \Delta m_{23}^2 &= (2.43 \pm 0.13) \cdot 10^{-5} \text{ eV}^2. \end{aligned} \quad (2.7)$$

However, this does not allow any conclusions about the absolute mass. Neither does it tell us which neutrino is the heaviest⁵. That means another way to measure the absolute mass scale has to be established. [MoPa04]

2.4 The Measurement of the Neutrino Mass

Neutrinos are the most abundant particles in the universe and as such involved in many different phenomena. That is why there is theoretically a variety of options for the determination of their mass. Due to the smallness of their mass and the ability of neutrinos to penetrate even massive structures without interaction, the measurement is nevertheless a complicated undertaking. In the following the different approaches are listed.

2.4.1 The Neutrino Mass in Cosmology

Cosmological studies that are of relevance for neutrinos include cosmic microwave background radiation (CMBR) surveys and large scale structure (LSS) experiments, like Lyman- α -forest observations and galaxy redshift surveys (GRS). The CMBR is studied with satellites like WMAP [Spot03] and Planck [Plan13] and the biggest galaxy redshift surveys are 2dF [O.El⁺02] and SDSS [SDSS09]. The combined results of these experiments can be used to determine the contribution Ω_ν of all neutrinos to the total mass in the universe. Together with the calculated neutrino density of $336 \frac{1}{\text{cm}^3}$ this number can be used to calculate the neutrino mass.

⁵There are two possible scenarios: normal hierarchy ($m_1 > m_2 > m_3$) and inverted hierarchy ($m_2 > m_1 > m_3$).

V. Barger et al. combined, for example, the results from SDSS, 2dFGRS and WMAP to determine the sum of all three neutrino masses to be [BaMT04]

$$\sum m_\nu < 0.75 \text{ eV}. \quad (2.8)$$

Other groups have reported results that are in the same range, but nevertheless different⁶. This reveals one of the major drawbacks of the cosmological neutrino mass studies: Different values for important input parameters can be used to describe the same measured data equally well, but with different implications for the neutrino mass. In order to get rid of these degeneracies further assumptions are needed. Thus the neutrino mass we obtain from cosmology is not model-independent. [KATR04]

2.4.2 The Neutrino Mass in Astrophysics

There are several ways to use astrophysical observations for the measurement of the neutrino mass. Two of the most promising approaches are time-of-flight measurements of supernovae neutrinos and estimations with ultra high energy cosmic rays (UHECR).

In a supernova, which is the explosion of a massive star at the end of its lifetime, an order of 10^{58} neutrinos are emitted [Beac99]. In figure 2.2 the remnants of the last exploding supernova in our Milky Way galaxy are shown. The ν -burst can be detected on earth with neutrino telescopes like IceCube and Super-Kamiokande. Neutrinos have a mass, so there should be a difference in the time-of-flight of these neutrinos compared to notional massless neutrinos, because some of the energy went into its mass. This time difference can be measured and as a reference either black holes or data from gravitational wave experiments can be used additionally [KATR04]. A sensitivity of

$$m_\nu = 0.75 \text{ eV (95 \% C.L.)} \quad (2.9)$$

is expected. However, the results can be influenced by neutrino oscillations or the like. [KATR04]

The cosmic ray approach towards the neutrino mass measurement uses estimations based on the Z-model. This model predicts that UHE neutrinos annihilate with massive relic neutrinos into Z-bosons. With the use of a relation between Z-mass, neutrino mass and the resonance energy, simulations can be made to compute the ν -mass. Drawbacks of this method are that the Z-burst theory is not entirely proved and that the statistics of the experimental data (e.g. from the Fly's Eye experiment) is limited and the systematic errors not fully understood. Results were given by [FoKR02]:

$$m_\nu = 2.75_{-0.97}^{+1.28} \text{ eV for Galactic halo origin,} \quad (2.10)$$

$$m_\nu = 0.26_{-0.14}^{+0.20} \text{ eV for extragalactic origin.} \quad (2.11)$$

Independent data from KATRIN can be used to test this theory. [KATR04]

⁶This is demonstrated in Table 1 of [KATR04].

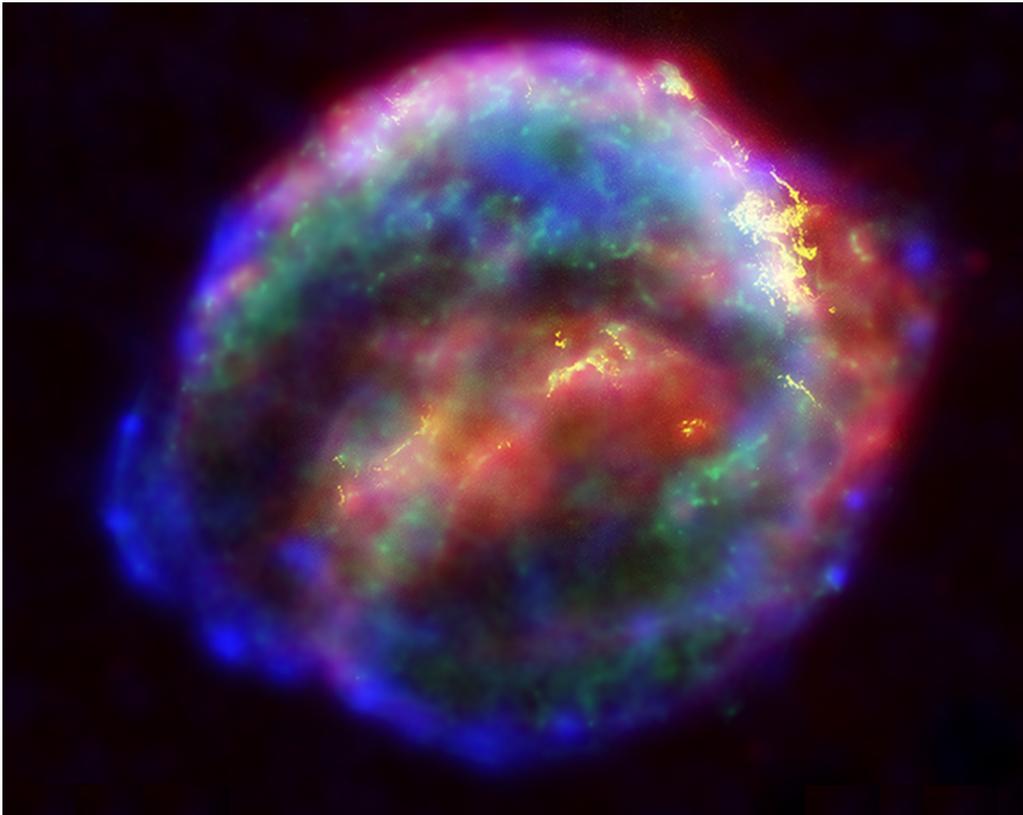


Figure 2.2: The latest supernova in the Milky Way galaxy. It is called Kepler's supernova because it was observed by the famous astronomer in 1604. Observations from the Chandra X-ray Observatory, the Hubble Space Telescope and the Spitzer Space Telescope were combined for this photomontage [NASA04].

2.4.3 The Neutrino Mass in Neutrinoless Double- β -Decay

Regular, two neutrino double- β -decay ($2\nu\beta\beta$) is a weak decay process of second order that is explained within the Standard Model. Neutrinoless double- β -decay ($0\nu\beta\beta$), on the other hand, is beyond the Standard Model. It is the decay of two neutrons,

$$2n \rightarrow 2p + 2e^-, \quad (2.12)$$

into two protons and two electrons and works only if the neutrinos are Majorana particles, i.e. their own antiparticles. The $\bar{\nu}_e$ emitted at the first decay is then absorbed as a ν_e in the other decay. Hence no neutrinos can be detected.

Theoretically, the $0\nu\beta\beta$ -decay can take place in any β -emitter, however, it is much less likely than $2\nu\beta\beta$.

Experiments that look for $0\nu\beta\beta$ measure the decay products and determine the half-life $T_{\frac{1}{2}}^{0\nu\beta\beta}$ of this decay. From $T_{\frac{1}{2}}^{0\nu\beta\beta}$ the effective Majorana neutrino mass can be calculated.

A famous example for an experiment looking for this rare decay is the Heidelberg-Moscow experiment. It yielded an upper limit for the effective Majorana neutrino mass [BoUr01]

$$m_{\text{eff.}} < 0.35 \text{ eV (90 \% C.L.)}. \quad (2.13)$$

The data analysis of this experiment was very controversial [FeSV03]. Nevertheless, more recent results from the Enriched Xenon Observatory (EXO) give a similar outcome. This experiment used liquid ^{136}Xe as both, source and detector. Depending on the matrix element calculation the calculated effective Majorana mass is

$$m_{\text{eff.}} < (0.14 - 0.38) \text{ eV (90 \% C.L.)}. \quad (2.14)$$

according to [EXO-12].

2.4.4 Direct Neutrino Mass Measurements with β -Decay

The neutrinoless double- β -decay only works if the neutrinos are in fact Majorana particles. Direct measurements have the advantage that they do not require this assumption. Next to time-of-flight investigations, which were presented in section 2.4.2, β -decay studies are the most suited for the direct neutrino mass measurement. This investigation makes use of the relativistic energy momentum relation $E^2 = p^2c^2 + m^2c^4$ and the laws of energy and momentum conservation.

Quite generally, the functional principle is to measure the energy of all decay products, but the neutrino, which cannot be detected. With this data and sufficient statistics one can infer the neutrino mass.

For the masses of ν_μ and ν_τ the decay of pions into muons was studied [K.As⁺96]. The upper limits obtained thereby are in the range of keV or even MeV. This is much higher than the cosmological constraints for the neutrino mass. β -decay experiments, on the other hand, are sensitive in the eV-range and therefore promising for ν_e mass measurements. [KATR04]

There are several experiments that study this decay. The Microcalorimeter Arrays for a Rhenium Experiment (MARE) uses ^{187}Re and is currently running. The Mainz and Troitsk experiments are already finished and yielded the results [KATR04]:

$$m_{\nu, \text{Mainz}}^2 = (-0.6 \pm 2.2 \pm 2.1) \text{ eV}^2 \quad (2.15)$$

$$m_{\nu, \text{Troitsk}}^2 = (-2.3 \pm 2.5 \pm 2.0) \text{ eV}^2. \quad (2.16)$$

Like KATRIN these experiments studied the decay of tritium,



Which will be discussed even further when we have a look at KATRIN's measurement principle in section 3.2. Tritium has the advantage of a simple nuclear structure and a suitable lifetime of $T_{\frac{1}{2}}^T = 12.3$ years.

2.5 The Influence of the Neutrino Mass on Astrophysics and Cosmology

As shown in the previous section, there are many different ways to measure the neutrino mass. But this still does not answer the question why one is interested in the mass of such a tiny particle in the first place.

The neutrino is an elementary particle and it should be in every experimental physicist's interest to measure its mass to complete mankind's understanding of the universe. Neutrinos have a significant influence on the total matter distribution. It is known that there are on average $336 \frac{1}{\text{cm}^3}$ neutrinos in the universe and so far there is only an upper limit from model-independent measurements on their mass. Neutrinos could be accountable for a part of the mysterious substance that is dark matter. As very fast moving particles they are a candidate for *hot dark matter*⁷.

Another important use of the neutrino mass is that it can be used as an input parameter for the CMBR and LSS data analysis. The interpretation of this data with the use of a precise neutrino mass will increase our understanding of structure formation and Big Bang Nucleosynthesis. The higher accuracy will have positive effects on the whole field of cosmology and ensure precise future measurements.

It is particularly important to have a model-independent measurement of m_ν^2 , which can be used to verify other results and to cross-check experimental data. This is why KATRIN is a vital addition to the other experiments - model-independent, precise and without any presumptions. [KATR04]

2.6 Evidences for Sterile Neutrinos

So far mainly regular - also called active - neutrinos have been discussed. Neutrino experiments, however, did not only find these, but also evidence for particles that are even harder to detect. These particles are quite similar to neutrinos, however, they do not interact weakly. This lack of weak interaction gave them the prefix *sterile*. In this section only evidences for sterile neutrinos are listed. Their properties and especially their connection to the KATRIN experiment will be presented later, in chapter 4.

Among others, the gallium anomaly as well as the reactor anomaly are indicators for their existence.

⁷See also section 4.1.

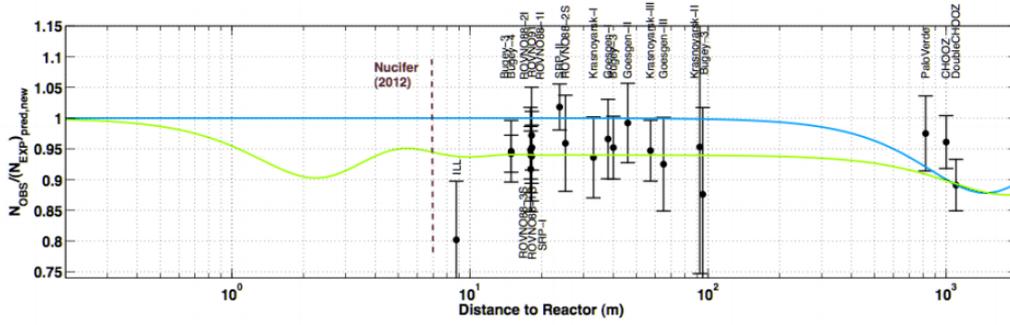


Figure 2.3: This figure from [Ment⁺11] shows the reactor anomaly. On the y-axis the ratio of observed to predicted events is plotted as a function of the distance from the reactor. The blue line shows a solution with the assumption that there are only three active neutrinos. For the green line a sterile neutrino was taken into account as well. The latter is able to reproduce the measurements better [Schw12].

Gallium anomaly is the term for observations made with the GALLEX and SAGE solar neutrino detectors. They are operated with Chromium and Argon sources and the neutrinos are detected with the neutrino capture of Gallium,



Both detectors measure values that are significantly (3σ) lower than the expected rates. An anomaly that could be explained with the existence of at least one sterile neutrino that mixes with the active ones [AcGL07] [GiLa10].

The reactor anomaly describes a similar phenomenon. Nuclear reactors are observed and the neutrino flux is measured. Then the values are compared with a theoretical spectrum, which was calculated by Schreckenbach et al. (in [Schr⁺85] and further publications) and improved recently by Mueller et al. [Muel⁺11]. Again, one observes a deficit of $\bar{\nu}_e$, as is shown in figure 2.3. According to [Schw12] a sterile neutrino can account for the reduction [Ment⁺11].

Besides these two evidences there are more experimental results that lead to similar conclusions. Short baseline appearance data and global data on ν_e disappearance suggest the existence of sterile neutrinos, as well as the results of the Liquid Scintillator Neutrino Detector (LSND) experiment. More information about these can be found in [Schw12].

3. The Karlsruhe Tritium Neutrino Experiment

The previous chapter was dedicated to neutrinos. Their basic properties were shown and different approaches to measure their mass were presented. Also, the KATRIN experiment was introduced as a direct, model-independent measurement, which, after three years of data taking, will reach a sensitivity of less than $200 \frac{\text{meV}}{c^2}$ at 90 % C.L. [KATR04].

In the following chapter the experiment, which is currently set up on the KIT Campus North site, is discussed. In section 3.1 the experimental setup is explained. After that, section 3.2 deals with the precise way how KATRIN will measure m_ν^2 . The subsequent sections are about simulation and analysis. Section 3.3 shows how KATRIN measurements can be simulated. It is explained in section 3.4 how either the simulated, or real measured data can be analyzed.

3.1 Experimental Setup

As illustrated in figure 3.1, KATRIN consist of several components. The most important parts are explained in the following subsections.

3.1.1 Source

The source, also called WGTS (Windowless Gaseous Tritium Source), is where the tritium decays into helium, an electron and an electron antineutrino. There are two major demands that need to be met by the WGTS. Most importantly it has to ensure a stable flow of β -electrons towards the spectrometer. And second, it needs to be controllable and monitorable [Babu⁺12].

The WGTS is 10 meters long, has an inner diameter of 90 millimeters and it is fed with molecular tritium gas from the Tritium Labor Karlsruhe, which is one of the few places with sufficient tritium to run the experiment. Thanks to a two-phase liquid Neon thermosiphon cooling system, the tritium gas is kept at 27 K with fluctuations of only a few mK [Groh⁺13]. The T₂ purity of 95 % is closely monitored [Schl⁺11].

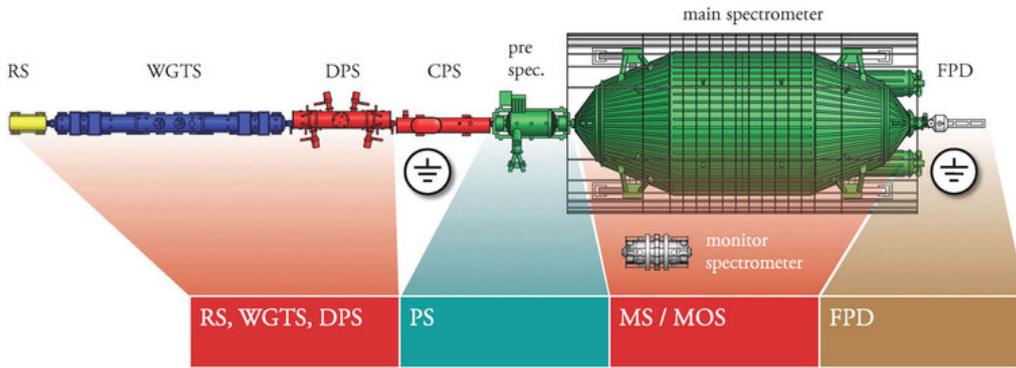


Figure 3.1: Overview of the setup of the 70m long KATRIN experiment. The different parts are explained in the text. [KATR13]

β -electrons from the other components, mainly DT and HT, have different energies that need to be taken into account for the analysis. The WGTS also ensures a stable pumping pressure [Fisc⁺11] that keeps the amount of gas inside the source – and consequently the number of decays – stable.

The WGTS is operated as follows. Strong magnetic fields with 3.6 T guide the tritium β -decay electrons towards the spectrometer, while the gas density is kept stable to guarantee a uniform flow of electrons. The total flow is $1.7 \cdot 10^{11}$ electrons per second. [KATR13]

3.1.2 Transport Section

The next step in the journey of the β -electrons is the transport section that guides them towards the spectrometers. Because of the limited tritium inventory of a few grams, it is mandatory to lead all of the T_2 -gas that enters the transport section back into the WGTS. Also, inside the spectrometer tritium and other ions would cause unwanted background. For these reasons a Differential Pumping Section (DPS2F) and a Cryogenic Pumping Section (CPS) are installed between the WGTS and the spectrometers.

The DPS consists of beam tubes within a superconducting solenoid and is equipped with four turbo-molecular pumps that can pump more than 2000 liters per second [Luki⁺12].

As the name of the CPS suggests, it uses cryo-sorption to trap the remaining tritium. The beam tube is cooled down to about 3 K. Its inside is covered with Argon frost, which binds the remaining tritium [Gil⁺10].

The electrons are not affected by these elaborate systems. The tritium pressure, however, is decreased by twelve orders of magnitude at the end of the transport section. [KATR13]

3.1.3 Spectrometer Section

After they passed the transport section the β -electrons arrive at the spectrometer section. There are three spectrometers: Pre-Spectrometer, Main Spectrometer¹ and

¹The main spectrometer is one of the most impressive parts about KATRIN. It is the first ultra-high vacuum vessel of this size and manufacturing and the transport to Karlsruhe were a



Figure 3.2: The BBC picture of the year 2006 shows the KATRIN Main Spectrometer during the transport through Leopoldshafen. [BBC13]

the Monitor Spectrometer that is used to surveil the voltage stability of the main spectrometer.

The Pre-Spectrometer suppresses the β -electrons with low energies, because they do not carry any information about the endpoint of the spectrum and therefore no information about m_ν^2 . Only the β -electrons with high energies pass the Pre-Spectrometer and enter the 200 t Main Spectrometer.

The Main Spectrometer only allows electrons with an energy higher than a variable threshold to pass towards the detector. Both spectrometers are operated as MAC-E filters, a measuring principle that is explained in 3.2.2. The electrons are guided magnetically through them.

To guarantee the high precision of the experiment all background events and collisions have to be minimized. For this purpose the vessel is operated with ultra-high vacuum and a variety of methods to suppress backgrounds have been installed. A photo of the main spectrometer can be seen in figure 3.2.

3.1.4 Focal Plane Detector

After passing the spectrometer section, where they have been retarded and reaccelerated, the remaining β -electrons arrive at the focal plane detector (FPD). The FPD sits inside a superconducting magnet that is shielded against backgrounds of all kind. It is a multi-pixel silicon semiconductor detector with an energy resolution of 1.4 keV (FWHM) per pixel. In order to improve the analysis and subtract

major challenge. Interesting videos about the scientific background, manufacturing and transport can be found on the “KITVideoclips”-channel on youtube [KATR06].

background events the position of the electrons is measured as well. One obtains an integrated spectrum of the flux of the tritium β -decay electrons with energies near the endpoint of the spectrum.

3.2 Measurement Principle

Now that we have seen how the experiment is set up, we can have a look at how this instrument can determine the mass of one of the smallest particles in the universe. The total energy that is released in the tritium β -decay is branched out to:

- the mass and the kinetic energy of the β -electron,
- the mass and the kinetic energy of the electron antineutrino $\bar{\nu}_e$ and to
- the recoil energy and excitations of the daughter nuclei HeT^+ .

In order to determine the rest mass of the electron antineutrino $\bar{\nu}_e$ we need to understand how the energy is distributed to the different components. The excitations of HeT^+ are a statistical process that was calculated by Doss et al. specifically for the KATRIN experiment. This final state distribution is discussed later, in section 3.3. The mass of the electron is known and its kinetic energy is measured. Close to the endpoint of the spectrum the kinetic energy of $\bar{\nu}_e$ can assumed to be vanishing. Therefore it is this region that gives the wanted information about the neutrino mass that can be optimized with the use of complex data analysis algorithms (see section 3.4).

3.2.1 Calculation of the Fundamental β -Decay Spectrum

In order to carry out a proper analysis of the endpoint region, one needs to know the shape of the spectrum very well. The tritium β -decay spectrum can be calculated using Fermi's golden rule,

$$\lambda = \frac{2\pi}{\hbar} |M_{fi}|^2 \rho_E, \quad (3.1)$$

where λ is the decay constant, M_{fi} is the matrix element of the perturbation between the final and the initial state and ρ_E is the density of states [Blec10]. This equation was first formulated by Wolfgang Pauli in 1926, but Fermi named it 14 years later and the name stuck [Blec10]. It gives the transition probability of one quantum-mechanic state to another. Beginning with this equation and taking into account the Fermi function $F(Z, E)$ one obtains the energy spectrum,

$$\frac{dN}{dE} = C \times F(Z, E) p(E + m_e c^2) (E_0 - E) [(E_0 - E)^2 - m_\nu^2]^{\frac{1}{2}} \Theta(E_0 - E - M_\nu), \quad (3.2)$$

with E and m_ν denoting energy, respectively mass of the electron and E_0 corresponding to the endpoint energy. The Θ -function takes care of the conservation of energy. [KATR04]

The difference of a neutrino with mass compared to a massless neutrino is illustrated in figure 3.3. Because of the sharp increase in the decay rate at lower energies, only the endpoint region shows a statistically relevant difference.

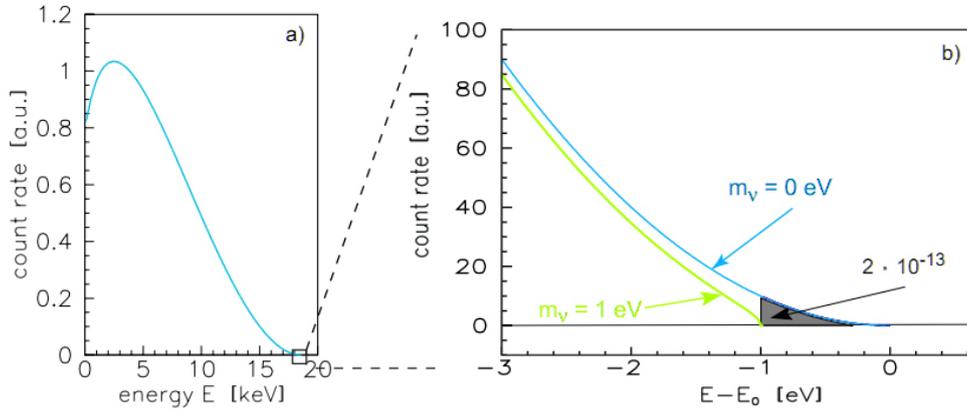


Figure 3.3: “The electron energy spectrum of tritium β -decay: (a) complete and (b) narrow region around endpoint E_0 . The β -spectrum is shown for neutrino masses of 0 and 1 eV.” [KATR04]

3.2.2 MAC-E Filter

The ultra-high precision of the KATRIN experiment is reached thanks to the use of MAC-E filters (Magnetic Adiabatic Collimation combined with an Electrostatic Filter) [BePT80]. This filter basically acts as a high-energy pass filter. All β -electrons with an energy that is lower than a variable electrostatic potential are reflected, the others are measured.

At the WGTS the electrons are emitted with an isotropic momentum. Once they are in the spectrometer they move along the magnetic field lines in a cyclotron motion. With the use of a massive magnetic gradient (in the center of the spectrometer the magnetic field is decreased by several orders of magnitude), their momentum is aligned, so that all of their energy is converted into the longitudinal motion. In the middle of the spectrometer the retarding potential reaches its maximum. Afterwards, all electrons that are not rejected are reaccelerated and guided towards the detector. The process is illustrated in figure 3.4. [KATR04] [KATR13]

3.3 KATRIN Simulation

A considerable amount of work is dedicated to the simulation of the KATRIN experiment. This is essential to comprehend the different processes and to determine their influence on the neutrino mass measurements even before they have started. For this purpose the simulation package KASSIOPEIA was created. It contains several different modules, e.g. for the simulation of the tritium β -decay spectrum or of electron collisions in the main spectrometer. The former is of relevance to this thesis, which is why the focus will be on the Source Spectrum Calculation (SSC) package. Other packages are implemented for particle tracking and detection (KTRACK and KESS) and the field calculation (KAFCA and KEMFIELD).

For the simulation in SSC the WGTS is described using small volume elements, called voxels. Each one can be assigned different physical values. This allows taking

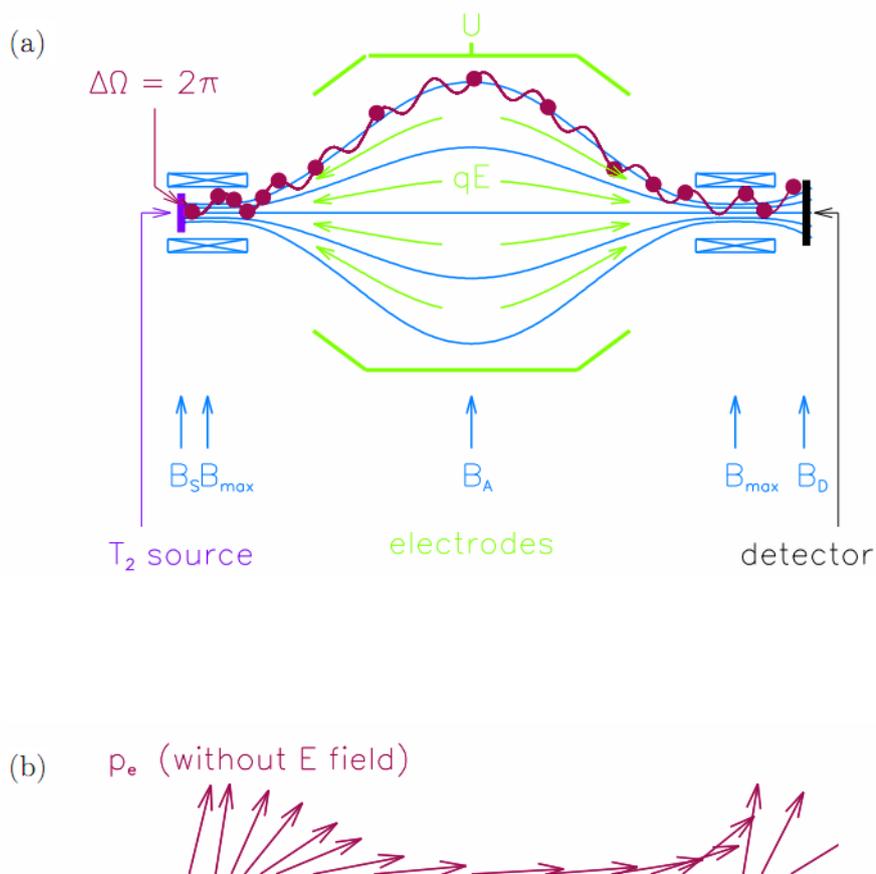


Figure 3.4: This figure shows the general principle of a MAC-E filter. In part (a) the experimental setup is depicted, while (b) shows “the momentum transformation due to the adiabatic invariance of the magnetic orbit momentum μ in the inhomogeneous magnetic field.” [KATR04]

the prevailing magnetic field or the temperature distribution into account. Scattering of the electrons with tritium molecules is considered, too. To describe the β -electrons on their further way through the main spectrometer the response and transmission functions come into play. The former takes care of energy losses and the latter incorporates the properties of the MAC-E filter that means it provides the probability with which electrons pass the potential barrier. [Höt12] SSC also incorporates the description of the β -decay spectrum. It was shown earlier, in section 3.2.1, that, with the use of Fermi's golden rule (eq. (3.1)), the decay spectrum can be calculated (eq. 3.2). This function is implemented together with the following corrections.

- The *Fermi function* takes care of the Coulomb interactions between the β -electron e^- and the daughter nuclei HeT^+ . Both are electrically charged and a non-relativistic approximation for the endpoint region was implemented according to [Simp81]:

$$F(Z, E) = \frac{x}{1 - e^{-x}} \left(a_0 + a_1 \cdot \frac{v_e}{c} \right). \quad (3.3)$$

Here $x = \frac{2\pi Z\alpha}{v_e}$, $\alpha = \frac{1}{137.036}$ and the empirical values for the constants are $a_0 = 1.002037$ and $a_1 = -0.001427$.

- In order to take the influence of the emission of virtual and real photons on the decay spectrum into account, *radiative corrections* were implemented according to [ReWu83].
- Another correction accounts for the energy that is lost due to the *nuclear recoil* of the HeT^+ -molecule. The nuclear recoil energy E_{rec} was calculated in [KATR04] to be

$$E_{rec} = \frac{p_{rec}^2}{2m_{\text{HeT}}} = \frac{p_e^2}{2m_{\text{HeT}}} = E_e \frac{m_e}{m_{\text{HeT}}} + \frac{E_e^2}{2m_{\text{HeT}}} \approx E_e \frac{m_e}{m_{\text{HeT}}}, \quad (3.4)$$

where p and E denote the momentum and energy of either the electron (index e) or the daughter nuclei (index rec) and m represents the mass.

- As mentioned in the beginning of chapter 3.2, a part of the energy of the decay is used for the excitation of the HeT^+ -molecule. This statistical energy loss needs to be taken into account, because otherwise it would limit the accuracy of the m_ν^2 measurement. This is why an elaborate final state distribution from [Doss07] was implemented into SSC. The probabilities for the excitations of the different states² are depicted in figure 3.5.

These corrections are only valid at the endpoint region of the spectrum, which is sufficient for the measurement of m_ν^2 . An extension was carried out in the scope of this thesis and is presented in chapter 5. [Höt12]

²Rotational, vibrational and electronic excitations are possible because HeT^+ is a two-atom molecule.

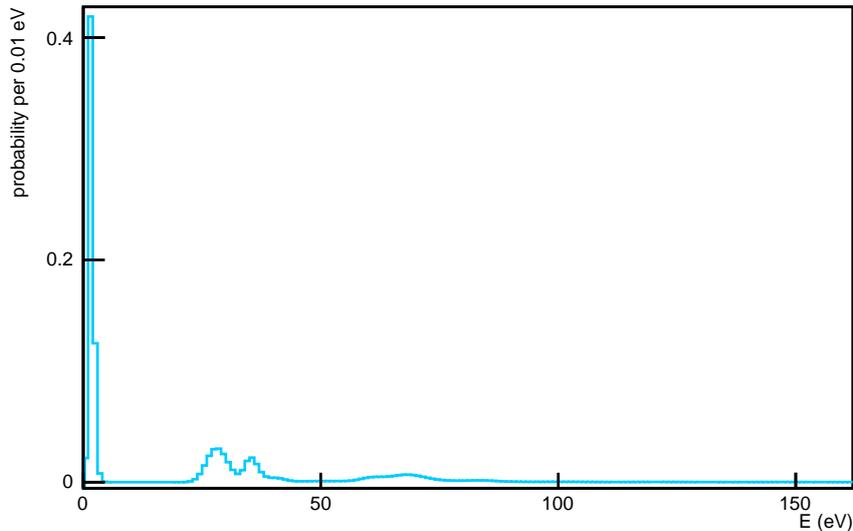


Figure 3.5: The final state distribution shows the probability for the excitation as a function of the energy. The first big peak is due to rovibrational excitations whereas the two smaller peaks are due to inelastic scattering.

All electrons above a certain energy level pass the MAC-E filter and are recorded in the detector. Therefore the consequent spectrum is a differential spectrum.

These spectrum simulations can also be used to simulate a measurement. For this purpose the expected value is randomly smeared with the use of a Gaussian distribution. Such a simulation is shown in figure 3.6.

Further information about KASSIOPEIA can be found in [Babu⁺11], [KATR04] or the detailed report in [Höt12].

3.4 KATRIN Data Analysis

The KASSIOPEIA framework is embedded in an even bigger software package, the KATRIN analysis toolbox KASPER.

This analysis framework is crucial and very important for the determination of the neutrino mass. As we have seen in figure 3.6, KATRIN does not weight the neutrino mass like a regular bathroom scale and returns its value right away, but it measures the number of β -electrons at the detector. The task of the sophisticated analysis is to extract the value of m_ν^2 from these results.

Therefore curves with at least four free parameters (endpoint energy E_0 , neutrino mass squared m_ν^2 , signal amplitude R_s , background strength R_b and possibly the tritium purity and high voltage fluctuations) are plotted to the measured values³.

³Until KATRIN is running and real data is obtained, simulated results can be used to test different fitting methods.

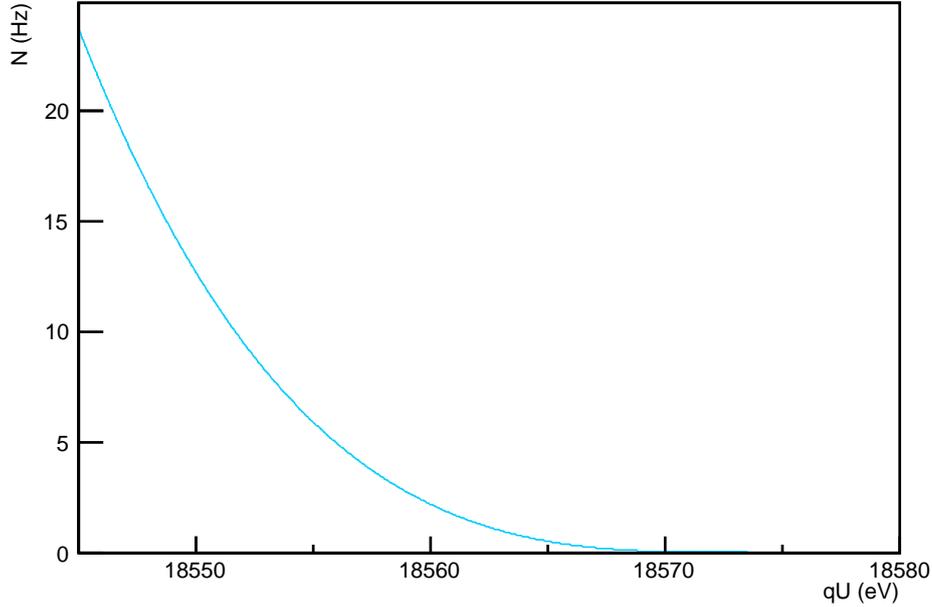


Figure 3.6: The simulated KATRIN measurement shows the typical shape of an integral spectrum.

The curve that describes the measured values the best can be obtained by minimizing the χ^2 -function [KATR04],

$$\chi^2(E_0, m_\nu^2, R_s, R_b) = \sum_i \left(\frac{N_{\text{exp}}(qU_i) - N_{\text{theo}}(qU, E_0, m_\nu^2, R_s, R_b)}{\sigma_{\text{theo}}(U_i)} \right)^2. \quad (3.5)$$

N_{exp} is the measured count rate, N_{theo} the theoretical count rate that is depicted in eq. (3.6) and σ_{theo} are the theoretical uncertainties.

$$N_{\text{theo}}(qU, E_0, m_\nu, R_s, R_b) = R_s \cdot N(qU, E_0, M_\nu) + R_b \cdot N_b, \quad (3.6)$$

where N_b denotes the background count rate and $N(qU, E_0, M_\nu)$ the integrated spectrum [KATR04].

This process of minimizing the χ^2 -function to determine the best values of the parameters is called *parameter estimation*. The minimization of the χ^2 -function can be accomplished by frequentist minimizers like MINUIT. Besides the χ^2 method also the maximum likelihood method can be used to estimate the best values of the parameters. Instead of minimizing the squared sum of the distances from the fitted curve to the measured one, the maximum likelihood tries to maximize the value of the likelihood function at the point of the measurement. Both methods are explained further in section 6.2.

Complementary and additional methods for the analysis and determination of m_ν^2 are also presented in chapter 6.

4. KATRIN and keV Sterile Neutrinos

In the previous chapter the KATRIN experiment was presented. It was shown how the mass squared of the electron antineutrino m_ν^2 can be measured. Additional to this search KATRIN may offer a chance to verify the existence of another particle and measure its mass. This particle is the so-called sterile neutrino. It is a serious dark matter candidate and its detection would cast light upon the question what the universe really consists of.

The following chapter discusses KATRIN's abilities towards the detection of this particle. At first, in section 4.1, the nature of dark matter is discussed. Then, in section 4.2, the sterile neutrino is introduced. Section 4.3 elaborates on the question whether KATRIN can measure such a particle. Systematic effects are discussed in section 4.4 and section 4.5 gives a brief summary.

4.1 Dark Matter

Today we know that baryonic matter only makes up about 4.6 % of the whole universe. The majority consists of dark (vacuum) energy and dark matter [KKVi02]. This is visualized in figure 4.1. Even though dark matter was already noticed in 1933 by Fritz Zwicky and described as a non-luminous, gravitationally interacting type of matter, its nature is still unknown. But where did the idea come from that most of the energy and matter in the universe do not consist of the same baryonic particles that we know from everyday experience?

4.1.1 Evidences for the Existence of Dark Matter

There were several observations that gave rise to the unique idea of dark matter. In the following subsection a brief introduction to this topic is given. A more elaborate discussion took place in [Drex11].

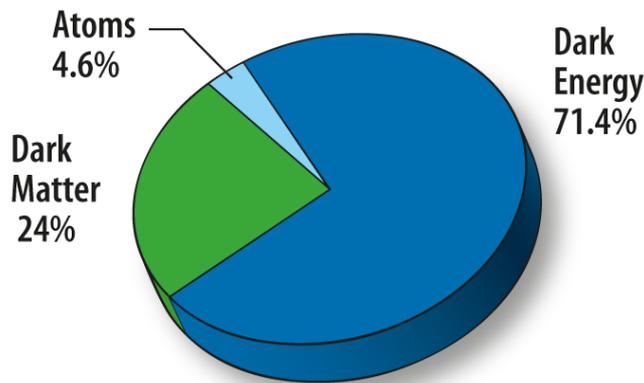


Figure 4.1: The distribution of matter and energy in today's universe [Nasa11].

Coma Galaxy Cluster

One of the first observations that indicates an existence of dark matter was performed by Fritz Zwicky. In his paper from 1933 [Zwic33] he describes his observations of the Coma nebula cluster. Under the assumption that the Coma-system is in a stationary condition, he applied the virial theorem,

$$\bar{E}_k = -\frac{1}{2}\bar{E}_p, \quad (4.1)$$

where \bar{E}_k is the mean kinetic and \bar{E}_p the mean potential energy. With a radius of about a million light years and an approximate total mass of $8 \cdot 10^{11}$ solar masses he calculated the peculiar velocity of single galaxies. He came up with $\sqrt{v^2} = 80 \frac{\text{km}}{\text{s}}$, which was much less than the observed values. He concluded that, in order to explain the observed velocities, the average density in the Coma system needed to be at least 400 times larger than the one derived from the observed matter. An instance that could be explained by the extensive existence of dark matter.

Rotational Velocity of Galaxies

In 1970 the American astronomer Vera Rubin studied the rotation of the Andromeda Nebula [RuFo70]. Observations of rotation curves of spiral galaxies, like the Andromeda Nebula, provide one of the major tools for determining the mass distribution in such galaxies. A rotation curve describes the velocity of the galaxy as a function of the distance to its center. Similar to the movement of planets in our solar system, one expects the rotational velocity to decrease as the distance to the center increases. Vera Rubin, however, discovered that the stars at the periphery of the galaxy rotated faster than it was expected for the observed mass. Figure 4.2, that was taken from [CoSa99], illustrates this matter. It shows the rotational curve of the M33 galaxy. The visible matter, the stellar disk and gas distributions alone, cannot explain its behavior. With the existence of a dark matter halo, on the other hand, theory can explain the measurements.

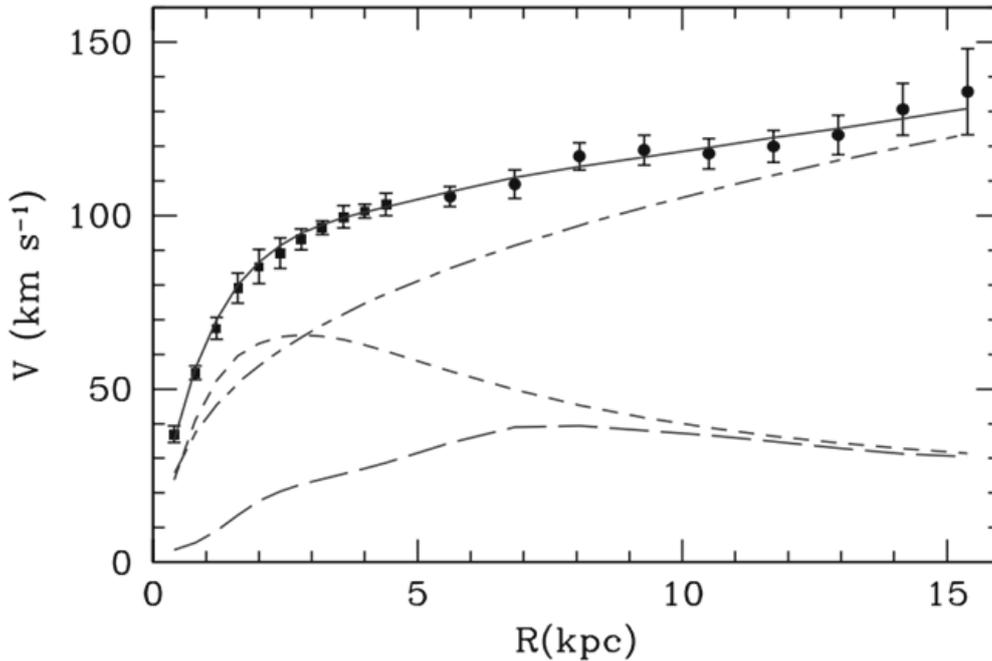


Figure 4.2: “M33 rotation curve (points) compared with the best fit model (continuous line). Also shown are the halo contribution (dashed-dotted line), the stellar disk (short dashed line) and the gas contribution (long dashed line)” [CoSa99]

Gravitational Lensing

According to Einstein’s famous formula, $E = mc^2$, mass and energy are proportional [Eins05]. Therefore a mass can be assigned to moving photons. That is why they are subjected to gravitational interaction. If there is a distribution of matter between a distant star and an observer, the trajectory of the light will be bent by the gravitational potential. This phenomenon is called gravitational lensing. One can distinguish three different types: strong lensing, weak lensing and micro-lensing. They are distinguished by their different behavior when they interact as a lens between a distant object (usually a bright star or a galaxy) and the observer.

The influence of strong lensing results in multiple images of galaxies or quasars, arcs or even Einstein rings.

Weak lensing does not have such a strong effect. It results in a statistical distortion of remote galaxies, due to the (dark) matter distribution between the galaxies and the telescope.

Micro lensing is the effect when the luminosity of a stellar objects is increased by a punctiform lens, like a star or a planet [Drex11].

Scientists have observed a lot of these phenomena, for example the gravitational lens magnification of the distant cluster Abell 1689 [Tayl⁺98]. Stars, or other baryonic objects, often cannot be observed in the areas where the gravitational lenses are expected. Therefore, these studies indicate a massive existence of dark matter to be responsible for these effects.

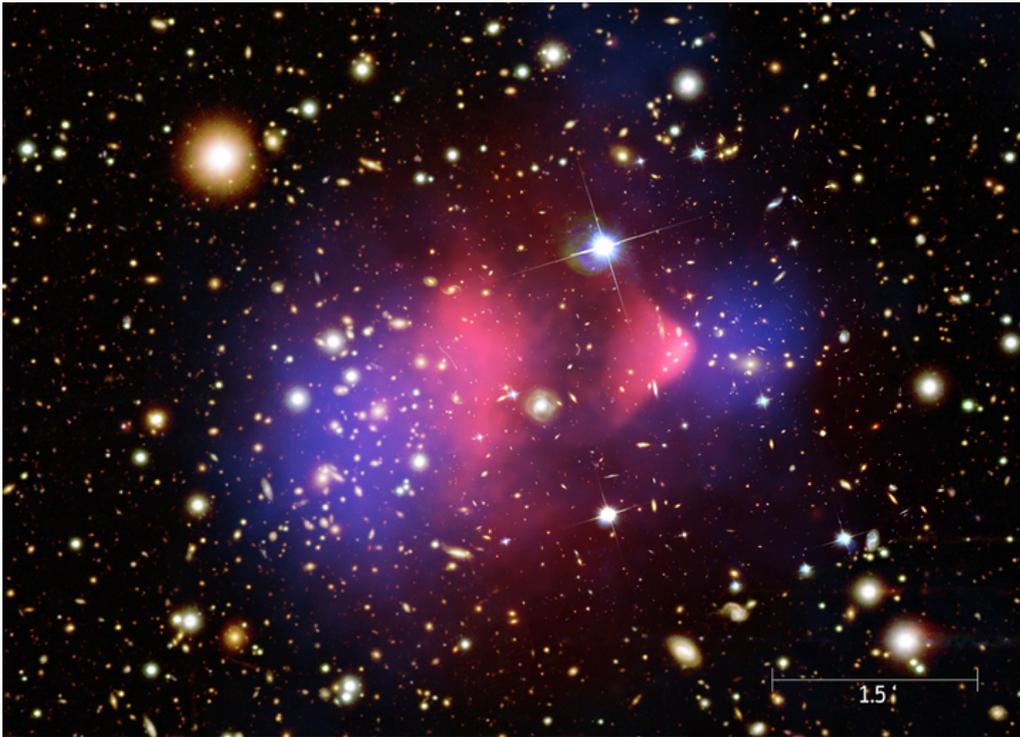


Figure 4.3: Photomontage of the bullet cluster. The galaxies can be seen in the background, the dark matter distribution is depicted in blue and the clouds of interstellar gas are colored in hot pink [NAWe06]. Their arrangement is strong indicator for the existence of dark matter.

Bullet Cluster

As a last example for dark matter evidences the bullet cluster observations need to be mentioned. It actually consists of two galaxy clusters that collided about 100 million years ago. According to [Clow⁺06] this merger enables a direct detection of dark matter with the use of gravitational lensing. As illustrated in figure 4.3 the different components of the bullet cluster can be observed separately: The galaxies are observed with the Magellan and Hubble space telescopes in the wavelength of visible light. The interstellar plasma can be detected with x-ray observations using the Chandra-telescope. Finally the distribution of matter is visualized with a statistical analysis of gravitational lensing (as described in the previous subsection). It is evident that the interstellar plasma interacted heavily during the collision and got separated from the rest. On the other hand the galaxies themselves passed through each other almost without any interaction, due to the vast distances between them. Also, the dark matter shows no interaction because it moves without dissipation. This spatial segregation of baryonic matter and dark matter is one of the biggest evidences for the existence of a substance that is called dark matter. But this leaves one question still unanswered: what does this mysterious type of matter consist of?

4.1.2 Cold Dark Matter

Even though the true nature of dark matter is still unknown, there are a lot of speculations about its composition. The possible dark matter candidates can be roughly divided into three categories: cold dark matter (CDM), hot dark matter (HDM) and warm dark matter (WDM). In this and the following two subsections these different types are discussed.

Cold dark matter is the label for all possible dark matter particles with a mass in the area of 100 GeV. The existence of cold dark matter would explain different things nicely, for example the formation of large scale structures. However, there are several problems that contradict the hypothesis that dark matter is composed of these slow moving, massive particles. For one thing, the corresponding free-streaming length l_{fs} for particles with a mass of about 100 GeV is $l_{\text{fs}} \approx 0.1 \text{ pc}$. The free-streaming length is the range that dark matter particles can travel freely. Therefore it is the connection between the scale of formed dark matter structures in the universe and the mass of these particles [KoTu94]. It can be expected that any structure smaller than l_{fs} is washed out by free-streaming [dVeg⁺11]. A free-streaming length in the area of 0.1 pc means that only smaller structures are erased, but structures in the size of 0.1 pc or bigger exist. But this is not what astronomers observe, since the smallest dark matter structures are in the area of 100 kpc¹ [dVeg⁺11].

Furthermore, CDM models for the surface density in galaxies yield a problem. Simulations have shown that the values obtained with CDM are about three orders of magnitude higher than the observations [Hoff⁺07].

Still another astronomical observation questions the existence of CDM. Simulations that were carried out under the assumption that CDM is the dominant form of dark matter indicate that dark matter halos have cusped profiles. However, astronomical observations show a different result: dark matter halos have cored profiles [dVSS10]. The distinction is visualized in figure 4.4. These reasons are already sufficient enough to disregard CDM alone as a serious candidate for dark matter, but for the sake of completeness another problem is listed: the satellite anomaly. Only about a third of the small dwarf galaxies that are predicted in simulations are actually observed in the milky way [Klyp⁺99].

4.1.3 Hot Dark Matter

Light particles - with a mass in the eV scale - are called hot dark matter. Because of their small mass these particles travel almost at the speed of light. Often discussed candidates for HDM particles are neutrinos. However, the combined mass of all active neutrinos is too small to explain the whole dark matter phenomenon.

Another argument against the hypothesis that dark matter only consists of HDM is the free streaming length. For particles in the eV range the free-streaming length is $l_{\text{fs}} \approx 1 \text{ Mpc}$. That means one should not be able to observe smaller structures. Astronomical observations have shown that this is not the case.

There would also be consequences for the way structure has formed in the universe. The majority of dark matter being HDM would result in a top-down-scenario in which smaller structures like galaxies and single stars formed after larger structures

¹Other structures or stellar objects that are smaller than 100 kpc are made of baryons.

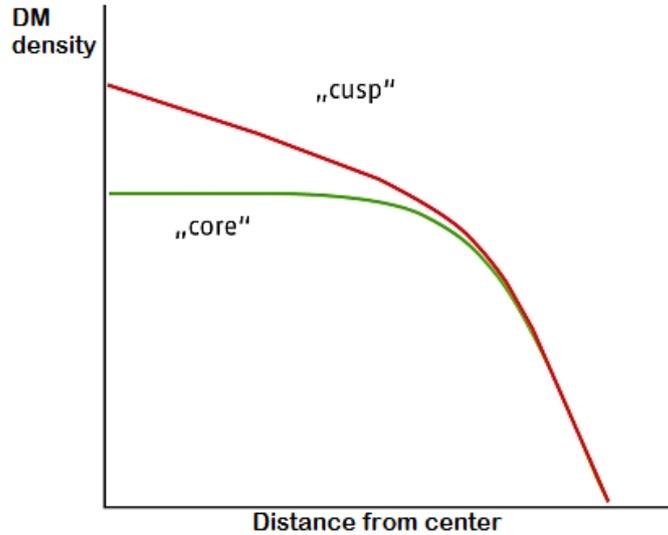


Figure 4.4: Illustration of cusped (red) and cored (green) profiles in dark matter halos [Bart12]. Astronomical observations indicate that dark matter halos have cored profiles, whereas CDM simulations lead to cusped profiles.

like galaxy clusters. However, age determination has shown that some galaxies are much older than the clusters they make up [dVeg⁺11] [Dode03]. These arguments show that, from a cosmological perspective, HDM alone needs to be discarded as well.

Mixed scenarios of both, cold and hot dark matter, as described in [Tran94], are not excluded by these observations

4.1.4 Warm Dark Matter

Warm dark matter is the term for particles with properties in between those of hot and cold dark matter. The mass of these particles lies between 1 keV and 10 keV. The corresponding free-streaming length is about $l_{\text{fs}} \approx 100$ kpc. Structures in this area exist, so WDM agrees well with the astronomical observations. Models that are calculated with warm dark matter predict cored profiles (see subsection 4.1.2) in agreement with the observations. Also, the right surface density of dark matter dominated galaxies can be reproduced with WDM simulations. Additionally, radio observations are in agreement with warm dark matter [dVeg⁺11].

That means in all the areas, where HDM and CDM alone cannot explain the astronomical observations, WDM succeeds. Therefore it is very reasonable to look for possible dark matter candidates in the mass range between 1 keV and 10 keV. Among others, these candidates are gravitinos, the light neutralino and majorons. According to [dVeg⁺11] the most promising candidate is the sterile neutrino.

4.2 Sterile Neutrinos

Sterile neutrinos are possible candidates for a dark matter particle. Opposed to regular neutrinos, they are right handed and do not take part in weak interactions.

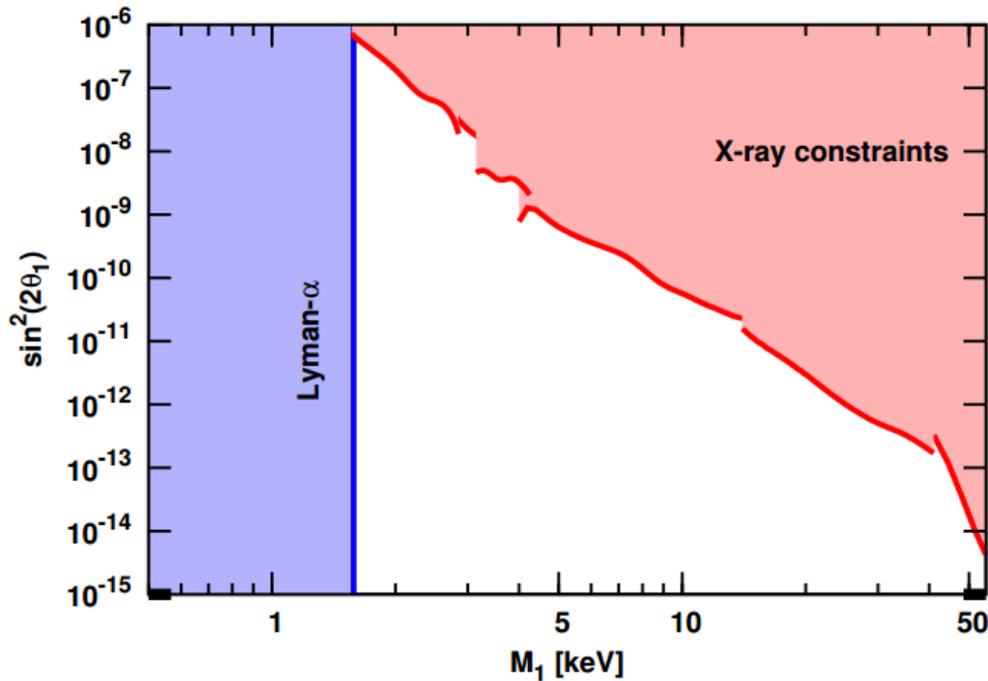


Figure 4.5: Constraints on the mass and the mixing angle of keV sterile neutrinos according to Dr. F. Bezrukov [Bezr11]. The forbidden region from the Lyman- α bound is depicted in blue, whereas the constraints from X-ray observations of the sterile neutrino decay are depicted in red. The latter depend on the mixing angle, which is plotted on the y-axis.

Due to their mass, however, they are subjected to gravity. They are singlets of color, weak SU(2) and weak hypercharge and their electric charge and isospin are zero. The lack of weak interactions makes their detection extremely difficult and gave them the name *sterile neutrinos*. One can think of light sterile neutrinos in the range of eV², or heavier ones in the range of keV. The implications of the former on the KATRIN experiment were discussed in a paper by Joseph Formaggio and John Barrett [FoBa11]. The latter are candidates for warm dark matter and are dealt with in chapters 4 and 5 of this diploma thesis.

The possible mass range of WDM sterile neutrinos and the allowed mixing angles can be narrowed down. For one thing there are restrictions from the Lyman- α bound. These exclude sterile neutrino masses below 1.6 keV [Bezr11]. The upper limit is given by X-ray constraints and is dependent on the mixing angle (see figure 4.5). These constraints are due to the fact that X-ray observations should be compatible with the model of the decay of the sterile neutrino:

$$\nu_{\text{sterile}} \rightarrow \nu_{\text{active}} + \gamma. \quad (4.2)$$

The observed quantity of X-rays ought to be in correspondence with the sterile neutrino lifetime in this radiative decay [Bezr11]. Another important requirement is met by sterile neutrinos. They only interact so weakly that they have not been discovered

²As we have seen in section 2.6.

so far. This sounds like a handicap rather than an advantage, but it is important insofar that only particles, that are to date unknown, are possible candidates for dark matter – which has not been discovered either. Regular, weak interacting particles in the mass range described by figure 4.5 would have been detected already.

But where does the idea come from that there are more than the known neutrinos? From a particle-physicist’s point of view an obvious analogy comes into mind: the existence of right and left handed quarks, whereas there are only right handed neutrinos. This gave birth to the idea that the same symmetry can be found in neutrinos. In [dVeg⁺11] this argument is described as “quark-lepton similarity”. [dVSS10]

4.3 Can KATRIN Measure keV Sterile Neutrinos?

In section 4.2 it was mentioned that sterile neutrinos are extremely difficult to detect. There is, however, a property that gives decay experiments like MARE or KATRIN a chance to verify their existence: sterile neutrinos mix with active neutrinos. This is similar to the e^- , μ^- and τ^- neutrino oscillations (see section 2.3) and it is described by the before-mentioned mixing angle. The neutrino mass matrix is given in [dVeg⁺11], as

$$(\bar{\nu}_L \quad \bar{\nu}_R) \begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix} \begin{pmatrix} \nu_L \\ \nu_R \end{pmatrix}, \quad (4.3)$$

where the indices L and R describe left- respectively right-handed neutrinos, M is the mass of ν_R and $m_D = yv$ (y is the Yukawa coupling and $v = 174 \text{ GeV}$). The eigenvalue for the mass of the active neutrino in this model is $m_{\text{active}} = \frac{m_D^2}{M}$ and the sterile neutrino mass is $m_{\text{sterile}} = M$. Using this mechanism with reasonable values for the masses it is possible to reproduce the observations³. Different models are discussed in [BeHL09].

If there are sterile neutrinos in the range of keV, then, thanks to their mixing with active neutrinos, there need to be detectable traces in β -decay experiments. The sterile neutrinos that are created in the decay would carry away the energy needed to make up their mass and momentum. This leads to a kink in the β -spectrum and should therefore be visible in decay experiments like MARE and KATRIN. Figure 4.6 shows such a kink for an excessively large mixing angle. However, due to the smallness of the mixing angle, which could be in the area of 10^{-4} or even smaller, sterile neutrinos are very difficult to detect.

The tritium decay observed in the KATRIN experiment has an endpoint energy of about $E_0 \approx 18.6 \text{ keV}$. This is significantly higher than the decay energy of MARE, which is about 2.5 keV , and covers most of the area in which a sterile neutrino is expected (see figure 4.5). In the first run of KATRIN only the area near the endpoint E_0 is observed (see chapter 3). In a second run, however, it might be possible to extend KATRIN’s range to map the whole spectrum. Then KATRIN could be able to measure a kink that indicates the existence of the new particle. In order for such a task to be successful it is necessary to know the shape of the tritium β -spectrum

³ $M \approx 1 \text{ keV}$ and $m_D \approx 0.1 \text{ keV}$ yields a mixing angle of about 10^{-4} , which is in concord with the amount of observed dark matter in the universe [dVeg⁺11].

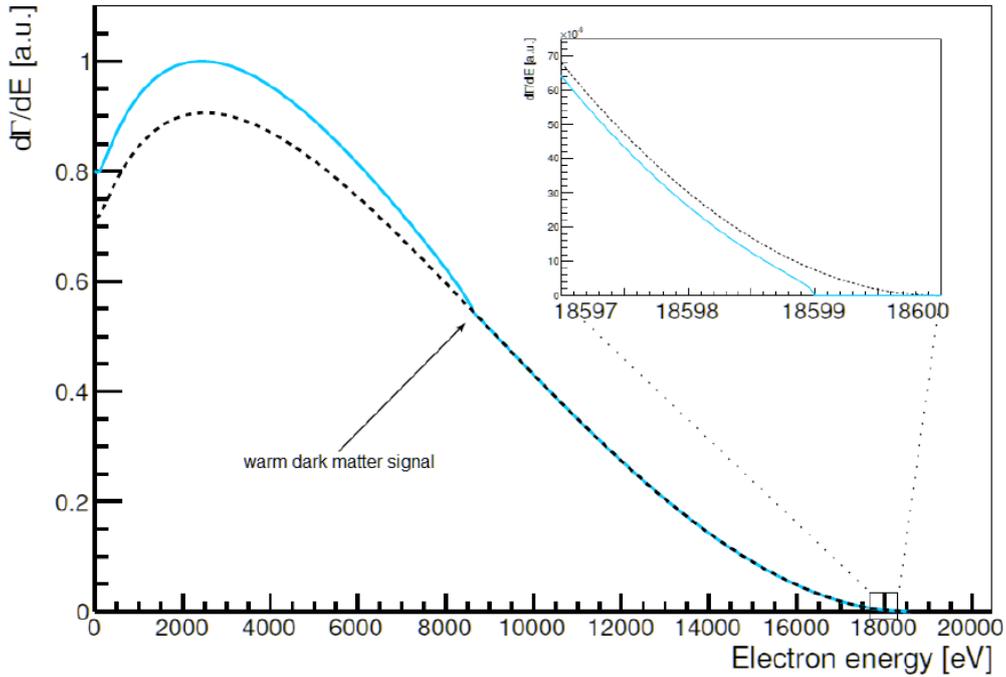


Figure 4.6: Visualization of a kink in the tritium decay spectrum that is caused by the existence of a sterile neutrino. The mass of the sterile neutrino is $M_s = 10$ keV and the mixing angle is $\Theta = 20^\circ$. The endpoint region, which is important for the standard KATRIN measurements, is also separately depicted [Mert13].

as accurately as possible⁴. Therefore calculations for the full spectrum have been conducted and can be found in chapter 5. [EsPe12] [Mert13]

4.4 Influence of Systematics

We have seen that there should be a kink in the tritium β -decay spectrum, provided that sterile neutrinos exist, their mass lies within the reach of KATRIN and the mixing angle is not too small. But can one be certain that such a kink is really due to the existence of sterile neutrinos and not owed to some other circumstances? A quick look at the contemplable systematic effects will cast light upon this question. One of these effects is the influence of electrons that are back-scattered from the detector. Another systematic effect is the stability of the tritium source and its composition. These effects, just like possible variations in the high voltage potential, can have an effect on the measured tritium spectrum. Nevertheless, none of these effects should lead to a knee in the spectrum. The same is true for the corrections for the tritium β -spectrum that are presented in chapter 5. Therefore it is very likely that the origin of a kink in the measured spectrum is a new particle, most likely the sterile neutrino [Mert13].

⁴We do not want a fiasco, like the “detection” of the 17 keV neutrino by J.J.Simpson in 1985, to happen again [Simp85].

4.5 Summary

In this chapter it was shown that there are serious evidences for the existence of a type of matter that is not baryonic, but nevertheless interacts gravitationally. If there is one particle that makes up the majority of this dark matter, then it is very likely that this particle is in the keV-scale. Because the movements of such a particle are neither very fast (such particles are called hot dark matter), nor very slow (they are called cold dark matter), this type of matter is named warm dark matter. As was shown in subsection 4.1.4, this type of dark matter is in nice agreement with astronomical observations. A candidate for a warm dark matter particle is the hypothetical keV sterile neutrino. It only interacts gravitationally and does not take part in weak interactions. Thanks to its mixing with light active neutrinos, it could leave determinable traces, for example a kink in the tritium β -decay spectrum. Provided the tritium spectrum is known well, it is possible that KATRIN is sensible enough to look for such a knee. The necessary calculations for a complete β -spectrum are carried out in the following chapter 5.

5. Calculations for the β -Spectrum

In order to be able to determine a small kink in the spectrum, we need to know the shape of the spectrum very well. The present-day KATRIN experiment is only interested in the region near the endpoint. That is why the standard data analysis uses calculations and corrections that are correct in the endpoint region and not over the whole range of the spectrum. Therefore further calculations are needed.

It was shown in section 3.2.1 how the basic shape of the tritium β -decay spectrum is calculated. There are several physical effects that have an influence on its form and need to be calculated separately. The dominant effects are Coulomb and Fermi corrections, radiative corrections and the effect of electron excitations in the final state of the daughter nuclei HeT^+ . These are discussed in sections 5.1, 5.2 and 5.3. Further corrections is dealt with in section 5.4. Comments on the implementation of the corrections can be found in section 5.5. Finally, section 5.6 compares the results and gives a summary.

5.1 Fermi Corrections

The tritium β -decay is described by the following nuclear equation:



It is a three-body-decay because there are three decay products, an electron, an electron antineutrino and a helium-3 isotope. Because the electron and the helium isotope are both electrically charged, they interact electromagnetically, while the electron moves in the field of daughter nuclei. This interaction is described with the Fermi function.

The implementation of the full Fermi function (which includes all Coulomb corrections) is therefore important in order to take the electromagnetic interaction between the electron and the daughter nuclei into account. In the case of KATRIN the daughter nuclei is generally not ${}^3\text{He}^+$, but molecular HeT^+ because the starting product is mainly molecular T_2 .¹

¹That means the nuclear equation for KATRIN is $\text{T}_2 \rightarrow {}^3\text{HeT}^+ + e^- + \bar{\nu}_e$.

There is a variety of papers dealing with the general topic. However, none of them give the full Fermi function for the tritium decay. Luckily, it is possible to look at the neutron β -decay and extend these considerations.

In [Wilk82] D. Wilkinson deals explicitly with this decay. He states the Fermi function that can be calculated with the use of the Dirac equation for an electron moving under the influence of a spatially finite nuclear charge distribution.

The Fermi function is quite long and, especially because of the complex $\Gamma(a)$ -function, intricate to calculate. That is why different approximations (for example from [Glü93]) were implemented into the KATRIN analysis software at first. However, it was not possible to obtain a sufficient accuracy, which is why the full Fermi function (eq. (5.2)) is now used. The equation is taken from [Wilk82] and is stated in natural units ($\hbar = m_e = c = 1$):

$$F(Z, E) = 2(1 + \gamma)(2pR)^{-2(1-\gamma)} |\Gamma(\gamma + iy)|^2 [\Gamma(2\gamma + 1)]^{-2} e^{\pi y}. \quad (5.2)$$

The following notations were used [Wilk82]:

- Z is the atomic number of the daughter nuclei (in our case $Z = 2$),
- E is the total energy of the electron, with E_0 being the endpoint energy,
- $\gamma = [1 - (\alpha Z)^2]^{\frac{1}{2}}$,
- $\alpha = \frac{e^2}{\hbar c} = \frac{1}{137.036}$,
- $p = (E^2 - 1)^{\frac{1}{2}}$,
- $R = \frac{\text{nuclearradius}}{\text{nucleonradius}}$,
- $y = \frac{\alpha Z E}{p}$.
- The $\Gamma(a)$ -function was introduced by L. Euler as [Domb06]

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx, \text{ with } a > 0. \quad (5.3)$$

For the Fermi function the complex version of this equation is needed. C. Gauß introduced this convoluted formula. It is discussed in [Domb06] and [Arti64] and tables for the values of the $\Gamma(a)$ -function with complex arguments are given in [oSta54]. For the implementation into KASPER the GNU scientific library was used.

The influence of this complete Fermi function on the shape of the decay spectrum can be seen in figure 5.1. Its impact is massive and raises the height of the spectrum at the maximum by about 40 %. Also, the rate at lower energies is increased stronger than at higher energy values.

Details on the implementation of these corrections are provided in section 5.5 and in appendix A.

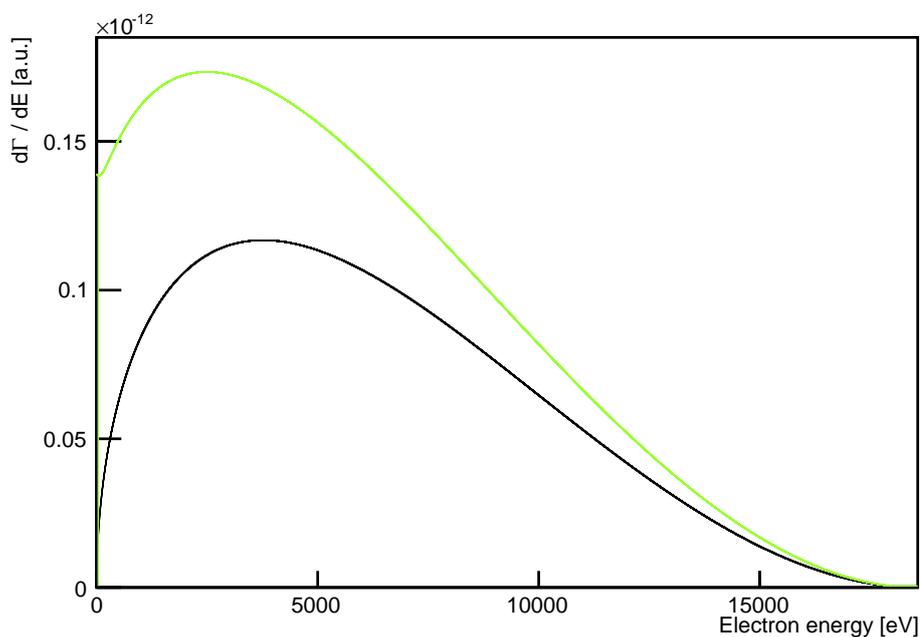


Figure 5.1: The Fermi corrections have a huge influence on the spectrum. The spectrum without corrections is depicted in black, with the corrections the spectrum is colored green. It is raised significantly. Also, the maximum is shifted towards lower energies. The decay rate on the y-axis is in arbitrary units. During this chapter the same presumptions were used for all spectrum calculations, so that the presented figures can be compared easily.

5.2 Radiative Corrections

During the tritium β -decay the emission of virtual and real photons can have an influence on the energy of the β -electron and therefore on the shape of the decay spectrum. This circumstance can be described by implementing terms for the radiative corrections that characterize these photon emissions [Step10].

For the standard KATRIN experiment the calculations by W.W.Repko and C.Wu (as given in [ReWu83]) are used [Höt09]. For the simulation of the whole spectrum, these have to be replaced with calculations that are valid over its whole range. A. Sirlin conducted these calculations in 1967 for β -decays in general [Sirl67]. In the following they have been adapted for the tritium β -decay.

Eq. (5.4) gives these model independent radiative corrections that, for the implementation, have to be multiplied with the formula for the decay spectrum. As the equation is quite elaborate it is divided into several parts. These are given in eq. (5.5), eq. (5.6), eq. (5.7) and eq. (5.8) and are stated in natural units ($\hbar = m_e = c = 1$) [Glü93]:

$$\text{Correction factor} = 1 + 0.01 \cdot r_e(E), \quad (5.4)$$

$$r_e(E) = 100 \frac{\alpha}{2\pi} g(E), \quad (5.5)$$

$$\begin{aligned} g(E) = & 3 \ln \left(\frac{m_f}{m_{e^-}} \right) - \frac{3}{4} \\ & + 4 \left(\frac{N}{\beta} - 1 \right) \left[\frac{E_0 - E_{\text{kin}}}{3E_{\text{kin}}} - \frac{3}{2} + \ln \left[\frac{2(E_0 - E_{\text{kin}})}{m_{e^-}} \right] \right] \\ & + \frac{4}{\beta} L \left(\frac{2\beta}{1+\beta} \right) + \frac{N}{\beta} \left[2(1+\beta^2) + \frac{(E_0 - E_{\text{kin}})^2}{6E_{\text{kin}}^2} - 4N \right], \end{aligned} \quad (5.6)$$

$$N = \frac{1}{2} \ln \left(\frac{1+\beta}{1-\beta} \right), \quad (5.7)$$

$$L(z) = \int_0^z \frac{\ln|1-t|}{t} dt. \quad (5.8)$$

The following abbreviations were used:

- $\alpha = \frac{e^2}{\hbar c} = \frac{1}{137.036}$,
- E_0 is the electron endpoint energy,
- E_{kin} is the kinetic energy of the electron e^- ,
- m_f, m_i are the masses of the final respectively the initial state of the decay,
- β is calculated as follows: $\beta = \frac{p_{e^-}}{E_{\text{kin}} + m_{e^-}}$, with p_{e^-} being the electron momentum and

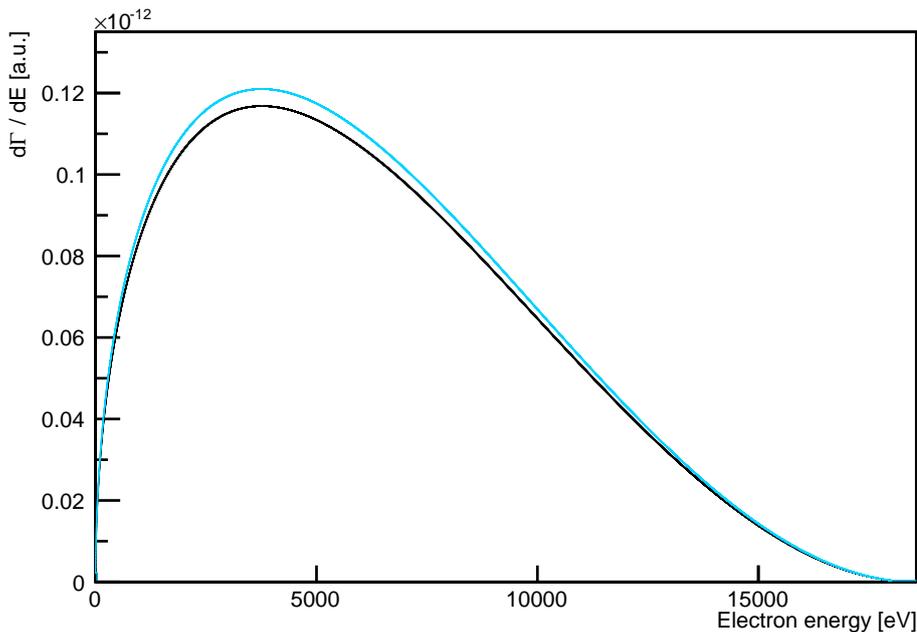


Figure 5.2: The Influence of the radiative corrections is not as massive as those due to the complete Fermi function, but it is still very apparent. The spectrum without corrections is depicted in black, with the corrections in blue.

- $L(z)$, finally, is the Spence function [Muta10].

The influence of these radiative corrections is shown in figure 5.2. The spectrum is raised distinctly compared to a curve that is calculated without any corrections. A comparison of the different corrections can be found in section 5.6.

Further details on the implementation are given in section 5.5 and in appendix A.

5.3 Full Molecular Final State Distribution

As discussed in chapter 3.3, the final state distribution takes the excitation of the daughter ion after the decay into account. It gives a total transition probability into the different excited states. The molecule ${}^3\text{HeT}^+$ consists of two atoms and as such has not only electronic excited states, but additionally rotational and vibrational excitations are possible. Its shape was shown in figure 3.5.

Like the other corrections for the KATRIN experiment, the final state energies were only calculated in the final region of the electron spectrum. That means excited states with energies higher than 240 eV were not taken into account. For the description of the full spectrum, however, it is important to have calculations that extend over the whole spectrum. Therefore additional corrections were implemented based on papers by Saenz et al. [SaFr97] [SaJF00].

For their calculations they used the sudden approximation and discarded the effect of electron excitations due to recoil, as well as the influence of the coupling of

electronic and nuclear motions. Nevertheless, a myriad of other factors were taken into account, like the electronic wave functions, Born-Oppenheimer potential curves, rovibrational wave functions and the relativistic recoil².

The final state distribution can be divided into two parts. The part below the dissociation threshold of HeT^+ is discrete, whereas the one above is continuous. The transition probabilities above 240 eV can be approximated with the following equation [SaJF00]:

$$P(E) \approx 14.7 \left(\frac{8e^{-\frac{(4 \arctan \kappa)}{\kappa}}}{\sqrt{1 - e^{-\frac{4\pi}{\kappa}} (1 + \kappa^2)^2}} \right)^2 \frac{dE}{\text{eV}}, \quad (5.9)$$

with $\kappa = \sqrt{\frac{E - 45 \text{ eV}}{13.606 \text{ eV}}}$.

For the calculation of this tail of the final state distribution, the β -electron energy was set to a fixed value in the endpoint region of the spectrum (at about 18 keV). If the β -electron has an energy in the middle of the spectrum - say, for example, 9 keV - then, strictly speaking, these calculations may not be valid. It is possible that the final state distribution is dependent on this β -energy and therefore different from what is predicted by eq. (5.9). However, according to Prof. J. Tennyson, it is very complicated to carry out these new calculations and presumably a task for a whole dissertation. This is why in the course of this thesis, only the already existing additions were implemented.

Their influence can be observed by comparing figures 5.3 and 5.4. In the former the sudden break at 240 eV is distinct. In figure 5.4 this abrupt end is replaced with a smooth exponential drop. Compared to the height of the final state distribution up to this point, that is occasionally in the area of percent, this small atomic tail, with a maximum height of 10^{-3} percent, is not likely to have a huge impact.

Figure 5.5 confirms this assumption. Opposed to the corrections presented in the previous two sections, the full final state distribution has a very small influence on the tritium β -decay spectrum. The height of the curve is only lifted slightly, but for the greater part both lines overlap. By zooming into the region, where the spectrum reaches its maximum, the distinction becomes obvious. This is depicted in figure 5.6.

The implementation of the extension to the final state distribution is discussed in section 5.5 and more explicitly in appendix A.

5.4 Other Corrections

The three corrections, discussed in the previous sections, have the largest influence on the shape of the spectrum. For the sake of completeness some other effects shall be mentioned, even though their influence is expected to be very small.

For one thing this is the influence of forbidden transitions that is not accounted for so far. The quark-substructure of the daughter nuclei is another effect that might have a small influence. Nevertheless, it is unknown how to deal with this effect [Wilk82]. Some other corrections deal with the way the β -spectrum is calculated in

²For a very detailed description please have a look at [SaFr97] [SaJF00].

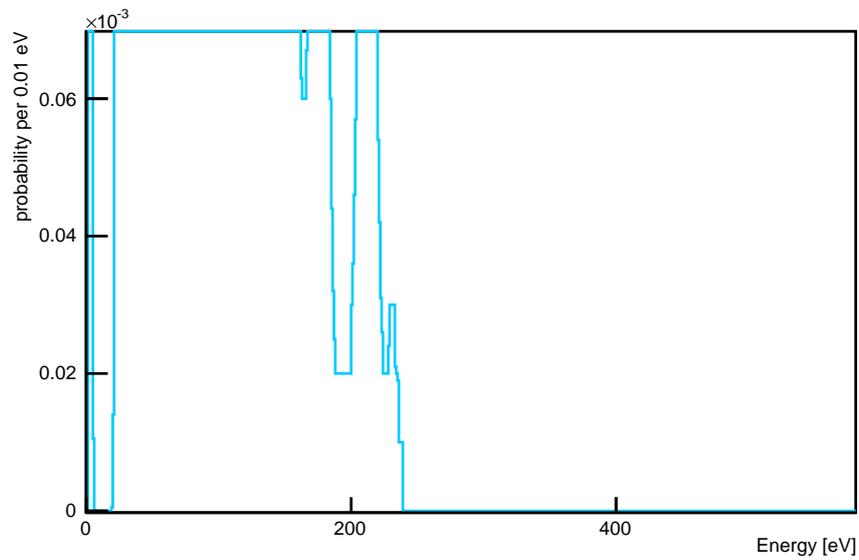


Figure 5.3: The incomplete corrections for the final state excitation spectrum of HeT^+ stop suddenly at 240 eV. Higher excitations are not taken into account - and for the standard version of the KATRIN experiment do not need to be.

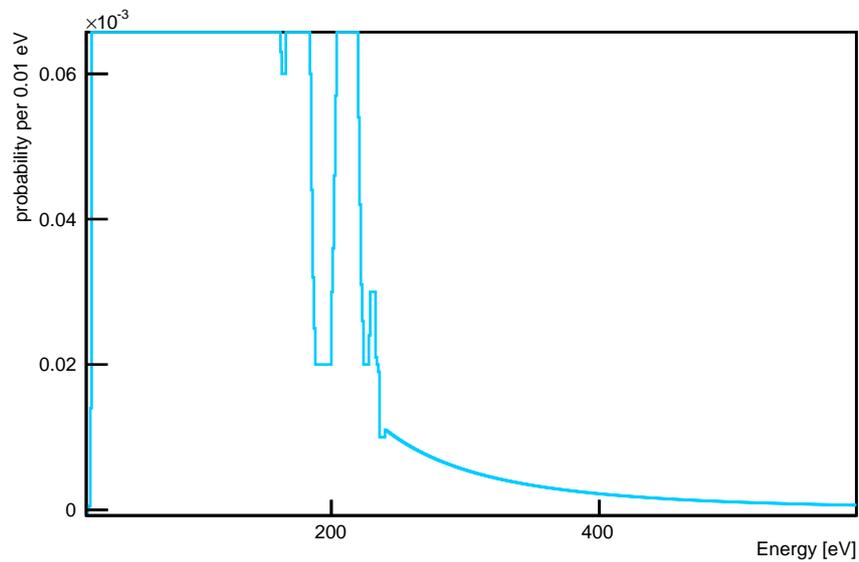


Figure 5.4: The corrections for the final state excitation spectrum of HeT^+ for the whole spectrum. The abrupt stop at 240 eV is replaced with the calculated tail that takes higher energy states into account.

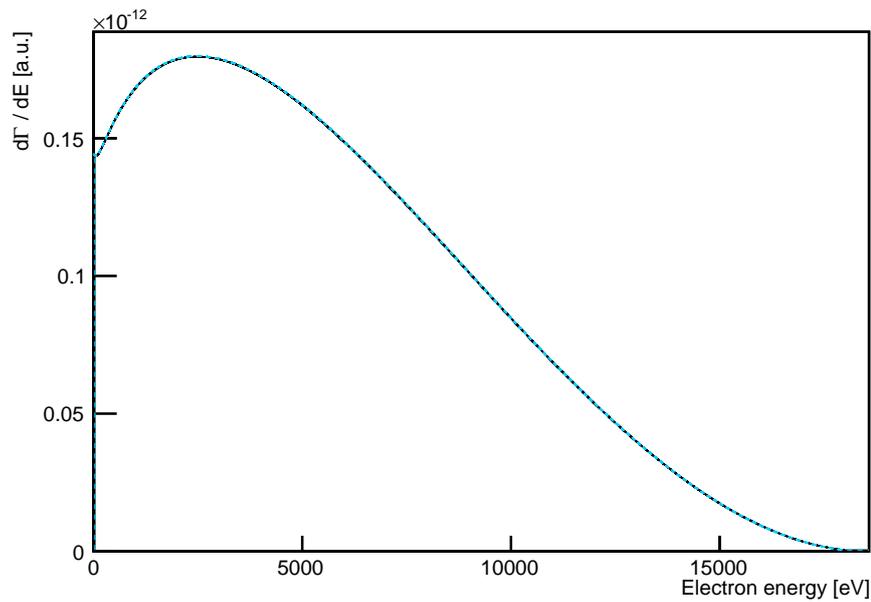


Figure 5.5: The spectrum is plotted with all previous corrections and with the incomplete (black line) and the complete (dashed blue line) final state distribution. The difference between both curves is very little and on this scale hardly visible. Therefore figure 5.6 shows a magnification.

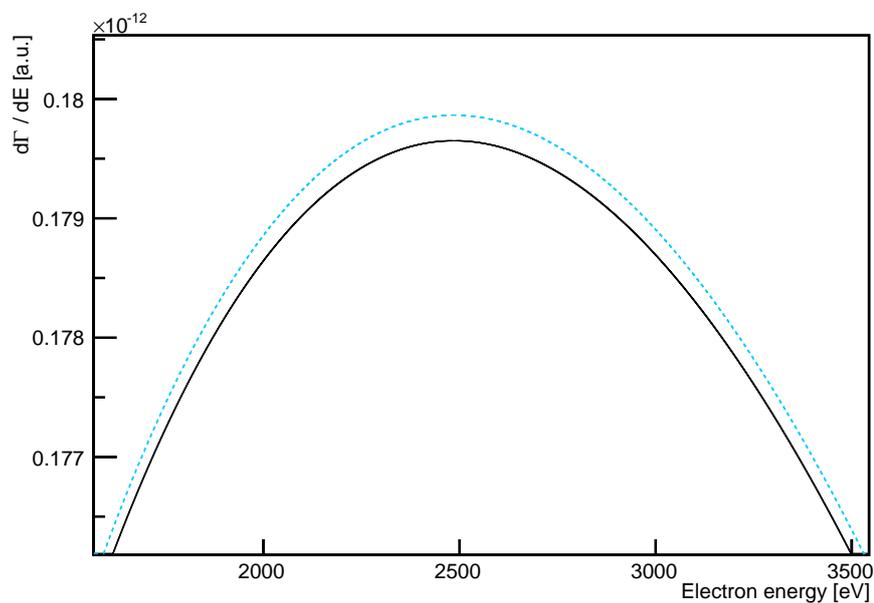


Figure 5.6: This figure shows a magnification of the spectrum, which is plotted with all previous corrections and with the incomplete (black line) and the complete (dashed blue line) final state distribution. Only the area around the maximum of the spectrum is shown. The corrections lift the spectrum slightly.

the first place. The β -decay can be described using the coupling constants g_V and g_A for axial and vector weak-coupling. Interference terms, for example $g_A g_V$, and terms that involve only one of the currents, like $g_V g_{WM}$ (g_{WM} describes the weak magnetism) also have a small influence. However, they are not implemented in the current calculations.

In this thesis only corrections that have a significant influence on the spectrum are regarded. This is why all miniscule effects have been discarded. However, if the mixing angle, of the sterile neutrino with the active neutrinos, turns out to be very small - e.g. in the area of 10^{-7} , as some cosmological observations suggest [RuIv12] - it might be important to implement these other corrections, as well.

Most of these additional effects are discussed in greater detail in [Wilk82].

5.5 Implementation

For the current version of the KATRIN experiment there are already Fermi and radiative corrections implemented into the simulation software³. These corrections are only valid near the final point of the decay spectrum, which is everything that is needed for the standard version of KATRIN. The corrections that were presented in this chapter are valid over the whole range of the spectrum. They are not compatible with the already implemented corrections. That means there needs to be an option to disable and substitute them when a whole spectrum is calculated. This task was accomplished with the program written for this diploma thesis. The user of the program can choose whether he wants to use the final region corrections or the ones for a complete spectrum. This decision depends on the task, whether a measurement for regular, active neutrinos or a full-range measurement for a keV sterile neutrino is conducted. The whole code was written in C++. Further details on the implementation and the program are given in appendix A.

5.6 Summary and Results

In this chapter the three corrections with the highest influence on the tritium β -decay spectrum were presented. In figure 5.7 the influence of the combined corrections is depicted. The largest change is due to Fermi corrections. Radiative corrections are quite significant as well and cause a further enhancement of the spectrum. The influence of the implemented tail for the final state distribution almost vanishes compared to these corrections.

With the help of this complete decay spectrum a future version of the KATRIN experiment hopefully has the chance to search for sterile neutrinos.

³See chapter 3 for an overview over the KATRIN simulation and data analysis.

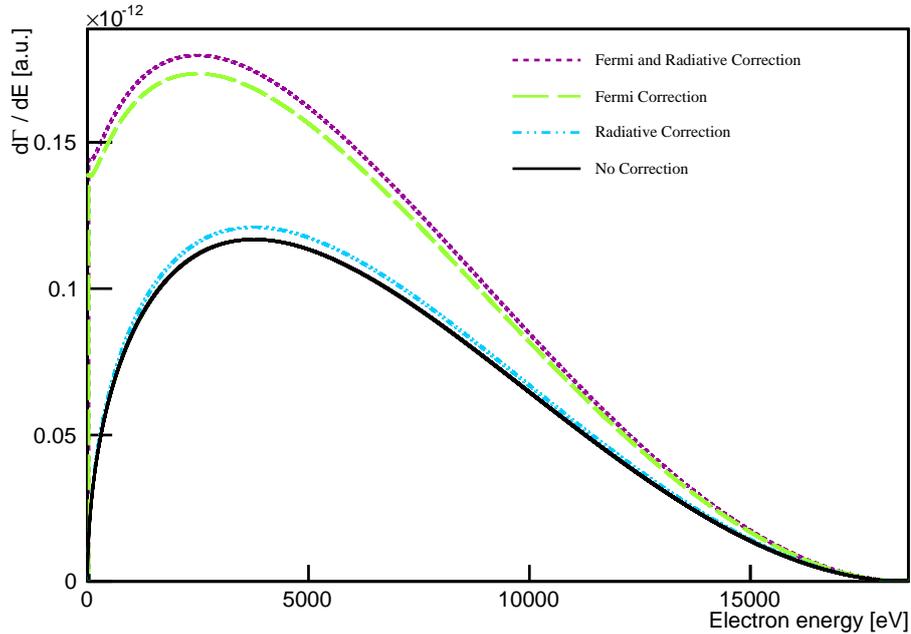


Figure 5.7: Combinations of the implemented corrections for the full spectrum. The influence of the final state distribution far from the final point is so small that it cannot be seen next to the other corrections.

Another interesting result is the comparison of the new and the “old” corrections⁴. Figure 5.8 makes their differences apparent. The fact that they do not lie on top of each other, but rather far apart, emphasizes the importance of the newly implemented corrections. As expected, one cannot obtain a valid complete spectrum with the “old” corrections.

The difference of both spectra is plotted in figure 5.9. It shows that the absolute difference between both calculations is relatively large at low energies and much smaller at the endpoint region, where they are both supposed to be valid.

The implementation of the different corrections for the tritium β -decay spectrum gives us another interesting opportunity. We can test the fitting routines implemented so far and thereby evaluate these programs. For the data analysis of a simulated KATRIN measurement usually the spectrum is fitted with the same assumptions that are used to simulate the measurement. Now it is possible to simulate the decay spectrum with all of the new corrections activated. That means the spectrum is computed with the sophisticated and intricate corrections presented in this chapter. This spectrum is then used to simulate a measurement. The simulated electron rates are then fitted without taking the new corrections into account - only

⁴The corrections that are used for the standard version of the KATRIN experiment are older than the corrections that were implemented to simulate the whole spectrum shape. That is why they are called “old” within the scope of this thesis. This does not mean that they are outdated - they are simply not valid anywhere but in the final region of the spectrum. But they can still be used for the standard KATRIN m_ν^2 measurements. The quotation marks are used to emphasize this fact.

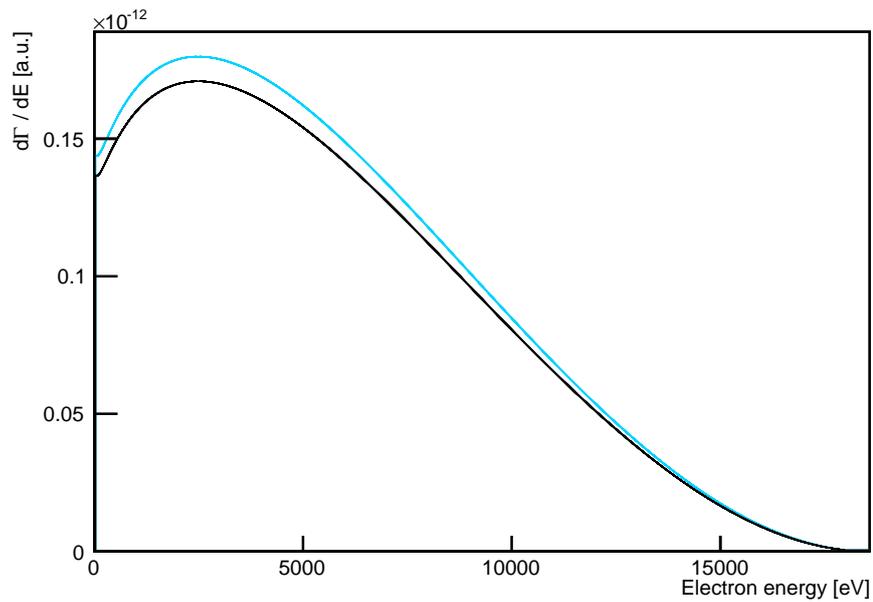


Figure 5.8: The “old” corrections that were implemented to approximate the shape at the final region also deliver a result for the full spectrum. These corrections are depicted in black. The new corrections that are valid over the whole range of the spectrum are depicted in blue. A significant difference is visible.

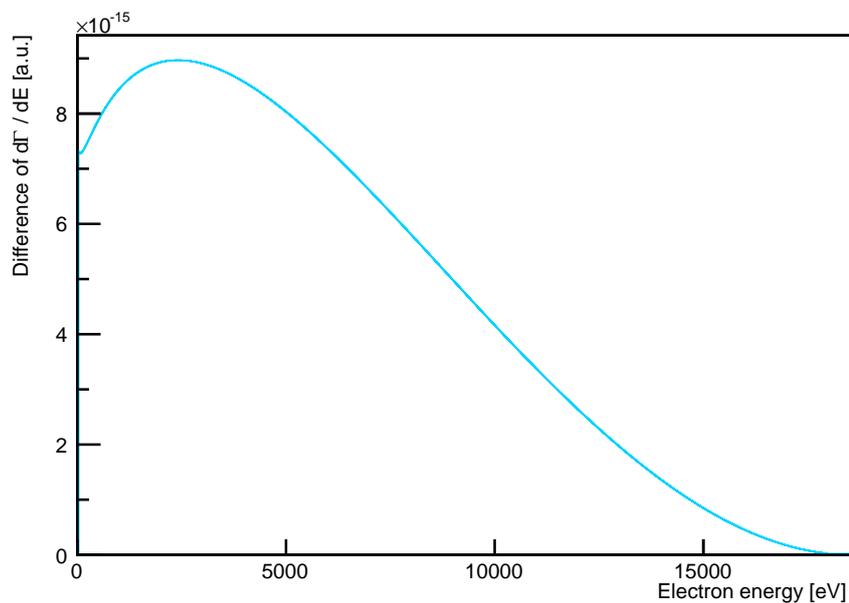


Figure 5.9: The spectrum that was calculated with the “old” corrections is subtracted from the spectrum with the corrections for the full range. At the area of the endpoint the difference becomes very small. The difference was calculated in the same units used in figure 5.8.

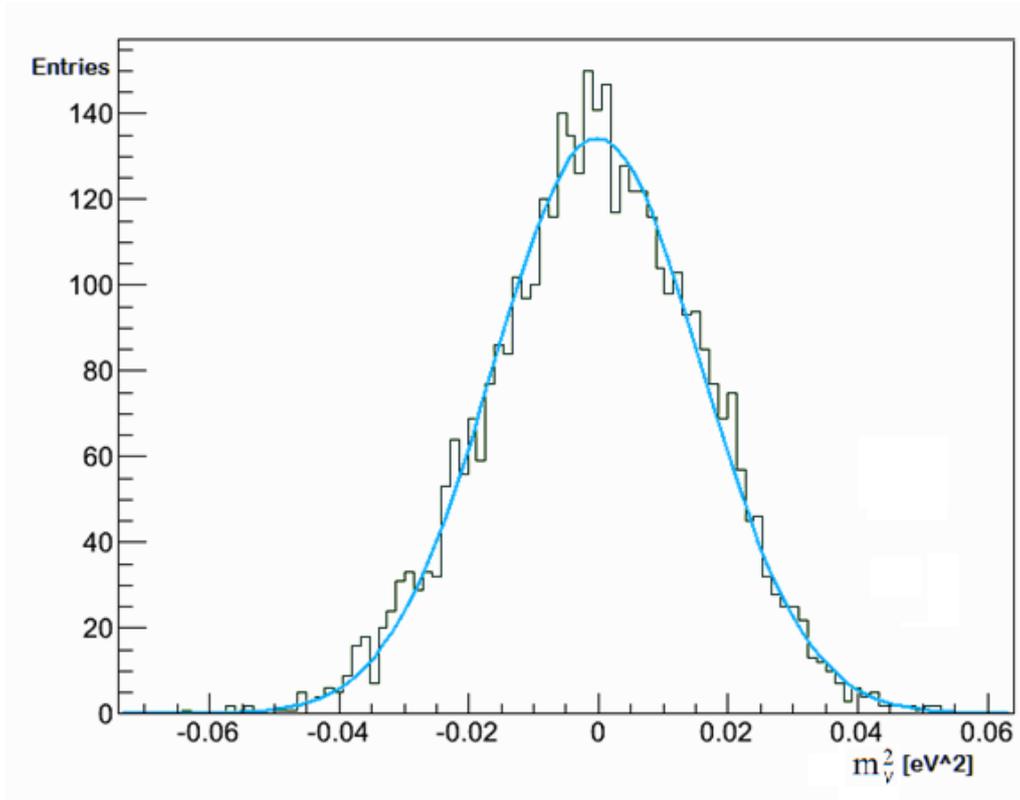


Figure 5.10: The measurement was simulated with the new corrections, while the fit was performed with the “old” corrections. The result for the neutrino mass squared is $m_\nu^2 = -3.2 \cdot 10^{-4}$. This is only a very small shift and in nice correspondence with the input value, which was zero.

the simpler approximations are used. This is a measure to benchmark the fit’s abilities to deal with measurements that are not like what we expect them to be. This is important because statistical effects and fluctuations may be accountable for such little differences in a real KATRIN measurement. The result is very satisfying and can be seen in figure 5.10. The value of the neutrino mass squared m_ν^2 was evaluated and plotted. Within the scope of this simulation, the result is correct and shows how well the KATRIN data analysis already works. Also, it justifies the use of the simpler and less computing-time consuming approximations for the standard version of the KATRIN experiment.

As we have seen, the simulation for the KATRIN experiments works well - for the standard neutrino mass measurements as well as for the simulation of β -spectra for the search for sterile neutrinos. But what about the data analysis? All questions about this topic will be answered in the following chapter.

6. Data Analysis for KATRIN

The standard procedure of KATRIN data analysis was presented in chapter 3.4. In order to determine our parameter of interest, the mass squared of the electron antineutrino m_ν^2 , an expected spectrum needs to be fitted to the measured electron count rates (see figure 3.6). The exemplary way to do this is, as we have seen, a χ^2 -fit using the minimizing routines that are implemented in ROOT: MINUIT and its subroutines MINIMIZE and MINOS [CERN13]. These algorithms, however, may be restricted by limitations if we implement more free parameters, the mass of a sterile neutrino and its mixing angle, for example. These lead to a more complicated likelihood function and the line-search-algorithm or minimizers like Migrad or Simplex could fail or get stuck in local minima.

Therefore one needs to find an algorithm that can handle both, multiple parameters as well as more complicated functions to minimize. A powerful tool to tackle these tasks is represented by Markov Chain Monte Carlo methods. These methods are discussed in the following.

Markov Chain Monte Carlo is a method of Bayesian inference. This is opposed to the frequentist approach that is used in MINUIT. In section 6.1, both ways of thinking are summarized to make their distinctions apparent.

In section 6.2, the Maximum Likelihood method is depicted as a more general method than the χ^2 -minimization routine.

After that, the fundamental principle that underlies all implemented Markov Chain Monte Carlo methods is presented in 6.3: the Metropolis-Hastings algorithm. This algorithm is the basis for the three specific methods, *Markov Chain Monte Carlo with Gaussian transition kernel*, *Hamiltonian Monte Carlo* and *Riemannian Manifold Hamiltonian Monte Carlo*, which are discussed in sections 6.4, 6.5, 6.6.

Finally, in 6.7, the three algorithms are compared and their advantages and shortcomings are assessed.

6.1 Frequentist and Bayesian Probability

There are two major ways to evaluate results of experiments in statistics: either with frequentist or Bayesian probability. The frequentist way of thinking is the classical approach. With the use of this technique the probability of a certain event is either determined by the symmetry of a problem or by a limit value consideration. To exemplify this, let us assume we perform an experiment whose outcome can either be A or B ($B \neq A$). Then the probability $P(A)$ for event A happening, is given by eq. (6.1), where n is the number of times the experiment is repeated and n_A is how often event A was the outcome

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}. \quad (6.1)$$

For large numbers of n the value of the fraction converges towards the true value of the probability, which is defined as relative frequency. This is pretty straightforward and can be easily visualized for simple experiments, like the throwing of a dice. According to [Jame08] one does not even need to let n go towards ∞ . It is sufficient to have an experiment that can be repeated often enough to achieve the required accuracy. But this is also one of the shortcomings of this method because it does not state what happens if the experiment cannot be repeated, e.g. if one is interested in the probability that Germany wins the soccer world cup in 2014. There is only one such event in that year and the next time it occurs the underlying circumstances will be very different.

Questions like this one can only be handled with Bayesian probability. By implementing prior information (or just called “prior”) one can obtain the probability of an event as a degree of personal belief. At first sight this does not sound very scientific because the result may change with the input of different knowledge, but it actually suffices the Kolmogorov axioms. These axioms have been stated by A. Kolmogorov in 1933 and present the basis for every theory that handles statistics properly¹. In order to understand Bayesian statistics one has to look at Bayes’ theorem (eq. (6.2)), which is named after the English minister and mathematician Thomas Bayes,

$$P(\theta_i|x) = \frac{P(x|\theta_i) \cdot P(\theta_i)}{P(x)}. \quad (6.2)$$

This universal formula can be clarified with an easy example: If someone measures the neutrino mass m_ν^2 , he obtains a value x , which may or may not be equal to the true value of the neutrino mass, for which we have different hypotheses θ_i . On the left side of eq. (6.2) we find the posterior probability $P(\theta_i|x)$ and on the right side there is the likelihood $P(x|\theta_i)$ together with the prior probability $P(\theta_i)$ and a normalizing constant $P(x)$.

- The posterior $P(\theta_i|x)$ gives the probability that hypothesis θ_i is correct for the measured data - i.e. that the true value for the neutrino mass is equal to θ_i .

¹The axioms and the quantitative rules arising from them are discussed in great detail in the second chapter and appendix A of [Jayn12].

- The likelihood probability $P(x|\theta_i)$ works the other way around. It basically gives the probability that x is measured under the assumption that θ_i is the true parameter for m_ν^2 . This parameter is known.
- $P(\theta_i)$ is called the prior probability and gives the initial degree of belief in the hypothesis. In this parameter everything is included that is known about the measured value before the experiment. In the case of the KATRIN experiment the prior could be a Heaviside step function, hereby excluding negative, unphysical neutrino mass values.
- $P(x)$ is a normalization constant.

The posterior probability $P(\theta_i|x)$ is what Bayesians are really interested in. It represents another distinction to the frequentist approach. With the use of the latter, one only obtains information about what happens if the experiment is repeated multiple times. For example, if one conducts an experiment to determine the parameter x one hundred times and defines a 1σ confidence belt for each one, then 68 of these 100 experiments will deliver a confidence interval that contains the true value. But one does not get any information about the true value itself. This is okay and not wanted differently by frequentists, as everything else requires prior information. Bayes' methods on the other hand will give a probability for the hypothesis of the true value, according to eq. (6.2). The problem is that it is generally not possible to calculate this posterior distribution. Luckily the posterior can be obtained in another way: through sampling of the KATRIN likelihood. What this means and how it is done will be described in section 6.3. But before we have a look at the specifics, it is important that we remind ourselves of the ways KATRIN data analysis is done in general and clarify some fundamental terms. This discussion takes place in the next section².

6.2 Data Analysis with Maximum Likelihood

The fit that is performed to the measured electron rate in the KATRIN experiment yields the parameters of interest, like the neutrino mass m_ν^2 . In order to achieve a good result for these parameters the fitted spectrum needs to describe the measured values to a high degree. There are two general methods to achieve this. One way is the χ^2 - or least-squares-method. A program, which uses this approach, tries to minimize the sum of the squares of the distances (red) between the fitted curve (black) and the measured values (blue), as it is exemplified in figure 6.1. If it performs a lousy fit, the value for χ^2 is big. If, however, the curve is fitted well to the measured points, then the distances are small and we obtain a small value for χ^2 . In order to get a good result one needs to minimize the value of χ^2 by adjusting the fitted curve in the multivariate parameter space. This method works very well for the standard frequentist implementations in KATRIN.

The Markov Chain Monte Carlo methods rely on another approach: the maximum

²For more information on Bayesian and frequentist probability see, i.e. [Jame08].

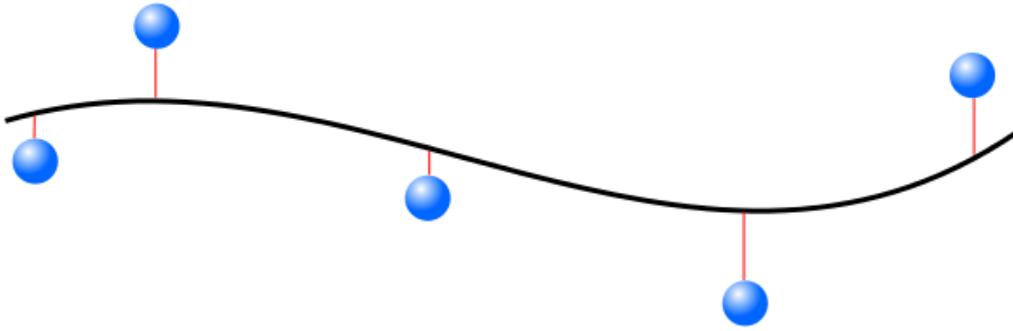


Figure 6.1: A curve (black) is fitted to measured points (blue) by minimizing the sum of the square of the differences (red).

likelihood fit.³ This method utilizes the likelihood that was introduced in the previous section. Following [Jame08] the likelihood can be written as

$$L(\vec{x}|\theta) = \prod_{i=1}^N f(x_i, \theta), \quad (6.3)$$

where x_i are the observed values, θ is the parameter whose true value θ_0 we are interested in and $f(x_i, \theta)$ is the probability density function of the value x_i . Its differences to the least-squares fit is illustrated in figure 6.2. In this graphic a measurement at $x = 2$ is illustrated by a thick red line and three different Gaussian distributions (green, blue and purple) represent different probability density functions, that are to be evaluated. One is interested to know which of these distributions describes the measurement the best. The least squares method compares the distance (dashed orange lines) between the measurement and the expectation value of the fitted curves. The blue and green curve have the same distance $\Delta = 2$ from the measured value, whereas the purple distribution has a better value of $\Delta = 1$. The maximum likelihood method on the other hand bears a different result. Here not the distance is minimized but the height of the probability density function is maximized. This is elucidated by the vertical orange dotted lines on the right of figure 6.2. The purple distribution is still the best fit for our measurement, as it has the highest value at $x = 2$. But now there is a difference between the two bell curves centered at $x = 0$. The green one has a bigger tail and therefore a higher value at the point of the measurement. This makes it a better fit than the blue curve. If the measurement had revealed a value $x \geq 3$, then the green curve would have been the best fit of all. We see that there is generally a difference between the χ^2 - and the maximum likelihood method. For large numbers, however, they can even become the same. The KATRIN likelihood for example is a product of probability density functions (compare eq. (6.3)) that are Poisson distributions. These can be approximated with Gaussian distributions for large numbers and therefore make the likelihood method

³The maximum likelihood method is not limited to the use in Bayesian inference. It can also be utilized in frequentist techniques, i.e. for the Feldman-Cousins Unified Approach implemented for KATRIN in [Höt12].

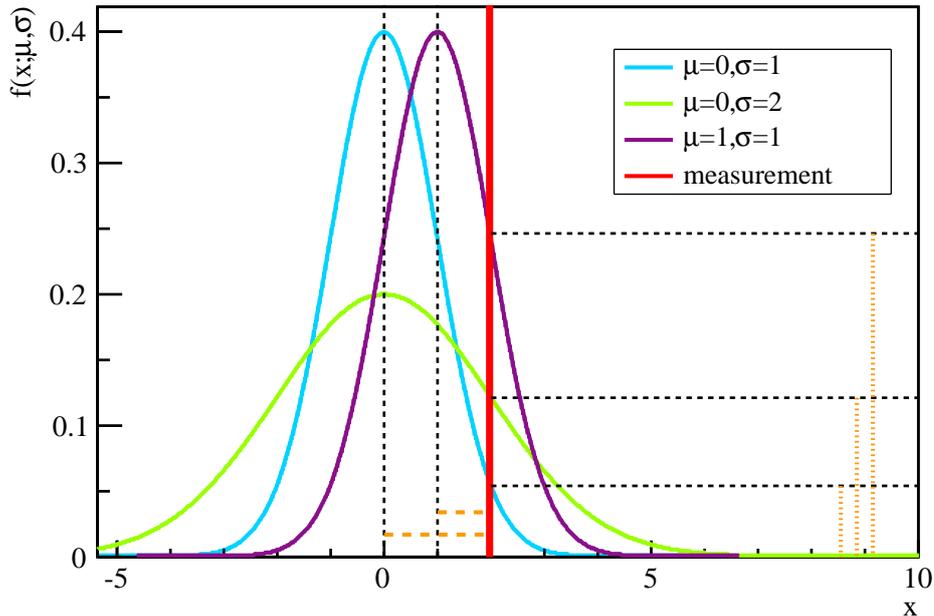


Figure 6.2: Three Gaussian likelihoods are plotted to match a measurement. The orange lines symbolize the difference between the χ^2 (dashed) and the maximum likelihood (dotted) method.

equal to the one of least squares.

For the further analysis it is now important to find the maximum of the likelihood. To do so, the derivative, which is not trivial for functions that consist of a lot of products, is needed. In order to solve this problem, the fact, that the logarithm is a monotone function, can be used. Therefore

$$\ln L(\vec{x}|\theta) = \sum_{i=1}^N \ln f(x_i, \theta) \quad (6.4)$$

has the same extreme values as eq. (6.3). For historical reasons and due to the fact that there are more efficient minimizers than maximizers eq. (6.4) is usually multiplied by -1 . Because of this the full name is *negative log likelihood function*. This minimization can now be achieved using Markov Chain Monte Carlo methods whose basics are shown in the next section.

Please keep in mind that this section only offers a brief introduction into the workings of maximum likelihood. For further reading I recommend [Barl89] or the slides of Prof. Dr. G. Quast's lecture on data analysis that can be found in [Quas11].

6.3 Underlying Principle

Markov Chain Monte Carlo methods present an excellent way to

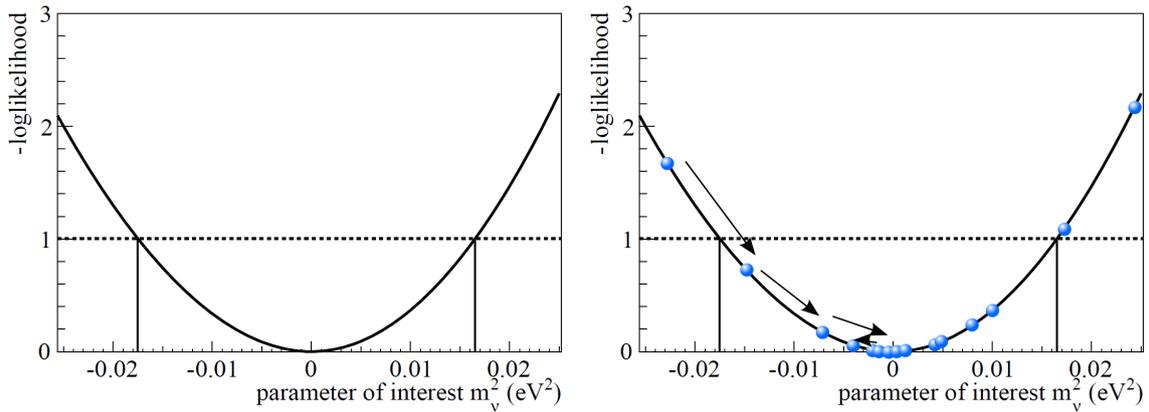


Figure 6.3: Image of a one dimensional KATRIN likelihood [Höt12]. The blue dots in the second graphic represent a location that has been sampled by the algorithm.

- minimize the logarithm of the likelihood function and reveal the most likely value of the parameters of interest,
- sample the likelihood distribution in order to obtain a sample of the posterior distribution. Sampling means in this context that the algorithm “walks” along the likelihood distribution. The states it visits are recorded in a histogram that, after a sufficient period of time, displays a sample of the target distribution. In order to do a good job, the time the sampler spends at a certain region needs to be proportional to the posterior density at the same region.

The latter offers a variety of advantages. Besides the obvious Bayesian posterior probability, we get a visualization of parameter correlations and we obtain error estimates - even if the parameters are not Gaussian - as well as confidence intervals.

6.3.1 Markov Chain Monte Carlo

The basic sampling is done using a random walk algorithm (hence “Monte Carlo”). “Markov Chain” means in this context that the position of the sampler \vec{x}_{t+1} at the time $t + 1$ depends only on the position of the sampler \vec{x}_t one step before. It does not matter where it was two, three or a hundred steps before.⁴ The transition from \vec{x}_{t+1} to \vec{x}_t is performed by the so called transition kernel $p(\vec{x}_{t+1}|\vec{x}_t)$. In order to visualize this sampling process it is helpful to have a look at figure 6.3. This graphic depicts a simulated KATRIN likelihood in the m_ν^2 dimension. Its minimum is at $m_\nu^2 = 0.0 \text{ eV}^2$. The different positions of the sampler are represented by the blue dots in the right hand side figure. It starts somewhat left of 0.02 eV^2 and then moves along the arrows towards the minimum. Once it arrives there it does not rest, but “walks” with a certain probability away from the minimum (and back again). The new positions are marked with the other blue dots. The more interesting area near the minimum is sampled more thoroughly resulting in a higher “blue-dot-density”. Please keep in mind that this graphic is only for illustration, it is not a quantitative model - we will come to those later on.

⁴Markov Chains are used in a variety of fields, including physics, chemistry, economics, Google’s web-search algorithms and many more [Wiki13].

6.3.2 Metropolis-Hastings Algorithm

So far the algorithm behaves completely randomly. It may run towards the minimum but it is equally possible that it moves away from it. In order to minimize this random behavior and steer the sampler in the right direction we use the *Metropolis-Hastings algorithm*. Starting from the current position \vec{x}_t in the multidimensional KATRIN likelihood distribution, the algorithm can be divided in two parts.

- A new point \vec{y} is suggested by a proposal function $q(\vec{y}|\vec{x}_t)$. Different ways to propose new states will be discussed in 6.4, 6.5 and 6.6.
- After a new position has been proposed it is evaluated using the Metropolis ratio (eq. (6.5)), which decides whether \vec{y} is accepted as new position \vec{x}_{t+1} or if its discarded. In the latter case the sampler stays at \vec{x}_t .

$$r = \frac{p(\vec{y}|D, I)}{p(\vec{x}_t|D, I)} \cdot \frac{q(\vec{x}_t|\vec{y})}{q(\vec{y}|\vec{x}_t)} \quad (6.5)$$

Unlike in [Greg10], vectors are used for the different states in eq. (6.5) to clarify that the sampled likelihood is not one, but at least four-dimensional.

The right fraction is for normalization purposes. If the proposal distribution is symmetric, then its value is equal to one. To understand the meaning of the first factor it is helpful to have a look at figure 6.4, which is a magnification of figure 6.3. The dot that is marked green represents the current position \vec{x}_t , for instance, and the one framed in red is the proposed state \vec{y} . The fraction is therefore the ratio of the target distribution (the KATRIN likelihood) at these points. If \vec{y} is closer to the minimum and therefore at a region with a higher probability than at \vec{x}_t , then $r > 1$ and the proposed state is accepted as new state: $\vec{x}_{t+1} = \vec{y}$. The same happens for $r = 1$. But r may also have a value between 0 and 1. In this case r becomes the probability with which we accept \vec{y} as new state. This makes it possible for the sampler to move away from the extremum with a certain possibility. Looking at figure 6.4, this means that the sampler may also come out at the other side of the minimum, or that it may turn around and walk back. This allows the sampling to take place and also - to some extent - to overcome local minima. The evaluation process can be summarized with a term called acceptance probability $\alpha(\vec{x}_t, \vec{y})$, as it is done in [Greg10]

$$\alpha(\vec{x}_t, \vec{y}) = \min(1, r) = \min\left(1, \frac{p(\vec{y}|D, I)}{p(\vec{x}_t|D, I)} \frac{q(\vec{x}_t|\vec{y})}{q(\vec{y}|\vec{x}_t)}\right). \quad (6.6)$$

It gives the probability that the new state is accepted.

6.3.3 Proof of Convergence

It was said earlier that the values obtained by the sampler represent the posterior distribution. But how do we know that this is actually the case?

[Greg10] states three properties a Markov Chain needs to fulfill in order to converge to a stationary distribution:

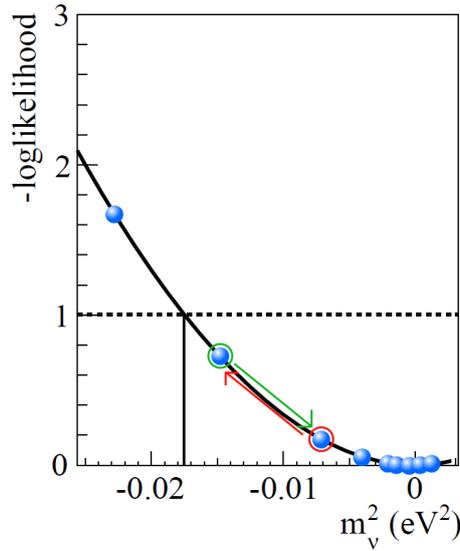


Figure 6.4: Enlargement of the one dimensional KATRIN likelihood to explain the Metropolis-Hastings algorithm.

- The Markov Chain must be irreducible. That means that no matter where it starts sampling the likelihood distribution, the sampler can end up at every point of the distribution - given enough time. The acceptance probability eq. (6.6), implemented in the Metropolis Hastings algorithm, is for the KATRIN likelihood always > 0 . So this criterion is satisfied.
- Furthermore the Markov Chain needs to be aperiodic. This means that there must not be a point from which it can only jump back to the previous state and back again. Basically all deterministic periodic behavior has to be forbidden, as it would contradict criterion one. The random nature of the proposal function takes care of this criterion.
- Finally, it needs to be positive recurrent: If one value \vec{x} is from the posterior distribution, then all sampled values after that are from the same distribution. To show that our Markov Chain suffices this criterion I will present a proof according to [Pres07].

Let us begin with the detailed balance equation

$$p(\vec{x}_t|D, I)p(\vec{x}_{t+1}|\vec{x}_t) = p(\vec{x}_{t+1}|D, I)p(\vec{x}_t|\vec{x}_{t+1}). \quad (6.7)$$

The left side of eq. (6.7) describes a process consisting of two steps. First, a point of the likelihood is chosen - e.g. the green-framed one in figure 6.4 - with probability $p(\vec{x}_t|D, I)$. Then, the sampler jumps to a random other point - for example the red one in figure 6.4. This happens with the transition probability $p(\vec{x}_{t+1}|\vec{x}_t)$. This product equals the probability of choosing the red point to begin with, and then jumping towards the green state. This is similar to a thermal equilibrium in statistical mechanics, for instance. But does the Metropolis Hastings proposal function

really satisfy eq. (6.7)? One can easily show that this is the case, by multiplying eq. (6.6) with the denominator of the second term in the brackets

$$\begin{aligned}
 p(\vec{x}_t|D, I)q(\vec{x}_{t+1}|\vec{x}_t)\alpha(\vec{x}_t, \vec{x}_{t+1}) &= \min[p(\vec{x}_t|D, I)q(\vec{x}_{t+1}|\vec{x}_t), p(\vec{x}_{t+1}|D, I)q(\vec{x}_t|\vec{x}_{t+1})] \\
 &= \min[p(\vec{x}_{t+1}|D, I)q(\vec{x}_t|\vec{x}_{t+1}), p(\vec{x}_t|D, I)q(\vec{x}_{t+1}|\vec{x}_t)] \\
 &= p(\vec{x}_{t+1}|D, I)q(\vec{x}_t|\vec{x}_{t+1})\alpha(\vec{x}_{t+1}, \vec{x}_t).
 \end{aligned}
 \tag{6.8}$$

The second step is obtained by writing the same equation with \vec{x} and \vec{x}_{t+1} exchanged and also multiplying with the denominator. Using the identity of the transition probability $p(\vec{x}_{t+1}|\vec{x}_t) = q(\vec{x}_t|\vec{x}_{t+1})\alpha(\vec{x}_t, \vec{x}_{t+1})$, this is equal to the detailed balance equation (6.7).

What is left is that we need to show that a Markov Chain, satisfying the detailed balance equation, is positive recurrent. This is done by integrating over eq. (6.7). The result is the marginal probability distribution of \vec{x}_{t+1} .

$$\begin{aligned}
 \int p(\vec{x}_t|D, I)p(\vec{x}_{t+1}|\vec{x}_t)d\vec{x}_t &= \int p(\vec{x}_{t+1}|D, I)p(\vec{x}_t|\vec{x}_{t+1})d\vec{x}_t \\
 &= p(\vec{x}_{t+1}|D, I) \int p(\vec{x}_t|\vec{x}_{t+1})d\vec{x}_t \\
 &= p(\vec{x}_{t+1}|D, I).
 \end{aligned}
 \tag{6.9}$$

This is exactly the definition of a stationary distribution. Therefore we have shown that once we obtained a sample, all of the following points will be from the same stationary distribution⁵ [Greg10].

6.4 Gaussian Transition Kernel

Now that we know the basic principles of Markov Chain Monte Carlo (subsection 6.3.1), how it works (subsection 6.3.2) and that it works (subsection 6.3.3), it is time to look at a concrete implementation of the Metropolis-Hastings algorithm. The first step of this algorithm is to propose a new state in the multidimensional KATRIN likelihood space. If this state is accepted, it becomes the next member of the Markov Chain. A common and easy way to propose such a new state is to use a multivariate Gaussian distribution as a transition kernel. Such a function - which is depicted in figure 6.5 for two dimensions - will choose a state that is close to the current position with a higher probability than a state that is further away. The width of the Gaussian transition kernel has a high influence on the process of sampling. If it is too small, the movement in the likelihood space is very slow. The sampling process takes a lot of time and we have to wait longer until a sample of the posterior distribution is obtained. If the width of the proposal function is too big, however, it is likely that new states \vec{y} are suggested that are very far away. Because of their large distance to the current state many of them may happen to have a worse value of the likelihood and therefore a small Metropolis ratio (eq. (6.5)).

⁵For a lecture-like introduction to this topic I recommend the online courses by Stanford's professor D. Koller that are available through OpenCourseOnline [Koll12].

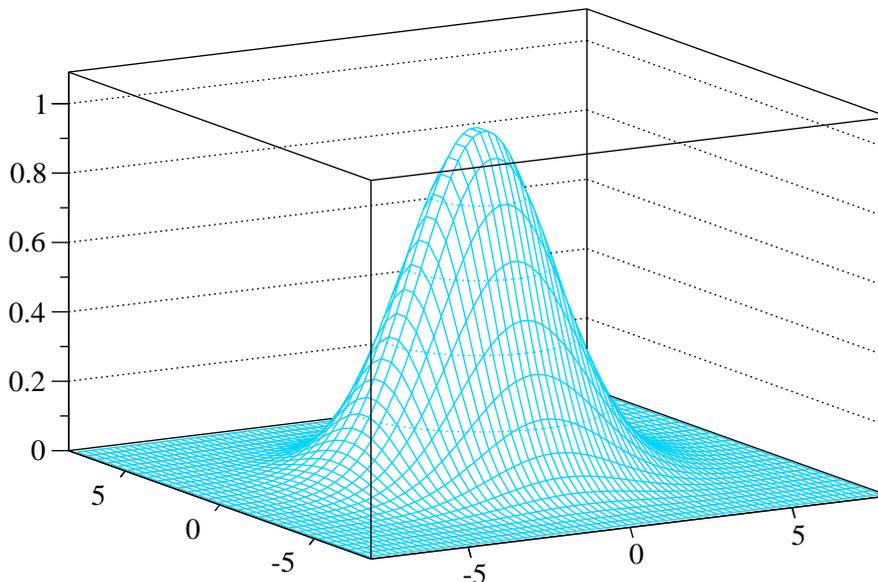


Figure 6.5: Two-dimensional Gaussian distribution that can be used as a transition kernel to propose the next state of the Markov Chain.

These suggestions for the new parameter are usually rejected. If this happens too often, our algorithm is slowed down significantly. Therefore it is important to tweak the width considerably⁶.

6.4.1 Implementation and Properties

This program was implemented into the KATRIN analysis framework KASPER using the programming language C++. It can be used for simulations as well as for the future analysis of real KATRIN data. Further details on the implementation and instructions on how to use the code can be found in appendix B. The code also provides the foundation for the Hamiltonian Monte Carlo and Riemannian Manifold Hamiltonian Monte Carlo method (see sections 6.5 and 6.6). Before we have a look at the results of some simulations it is important to discuss some of the properties. Some basic terms are explained as well. Most of them are also used in the description of the other routines.

- *Burn-in period*: This is the time in the beginning that the sampler needs to converge towards the minimum. These entries need to be deleted from the Markov Chain because the sampler has not converged yet and therefore does not represent the posterior distribution.

⁶For this purpose an algorithm called “error control” was implemented. It is explained in section 6.4.1.

- *Smearing*: This is a method implemented to make the simulation similar to a real measurement. After the expected electron count rates at KATRIN are simulated and multiplied with the measurement time at that point, we have the number of expected counts. Smearing is accomplished by adding a random value of a Poisson distribution with standard derivation equal to the square root of the number of counts. This blurring simulates the β -decay because its random numbers are Poisson distributed.
- *Error Control*: As mentioned earlier the stepsize - for this method it is the width of the Gaussian proposal function - is crucial for an efficient sampling process. Steps that are too large lead to a high rate of discarded entries. If the stepsize is small, so that almost every new state is accepted, the movement in the multidimensional space is slow. A. Gelman, W. R. Gilks and G. O. Roberts calculated in [GeGR97] that it is optimal for a MCMC method with a Gaussian transition kernel to discard 76.6 % of all proposed new states \vec{y} . That means that the acceptance rate is 23.4 %. This is only true for three or more dimensions. If we were to sample a one or two dimensional likelihood, the optimal acceptance rate is about 50 % [RoCa10]. Error Control is a routine in the code that compares the current acceptance rate with the ideal one and adjusts the step-size accordingly. This is only done in the very beginning of the program until the optimal step-size is determined. These entries to the Markov Chain should be discarded.
- *Acceptance rate*: As mentioned in the last item, the acceptance rate needs to be adjusted for each problem. This can be done in the configuration file. An exemplary configuration file is shown in appendix C.
- *Parallel tempering*: Parallel tempering is a procedure that gives MCMC methods a huge advantage over other minimizing routines because it permits to leave local minima. This is important because usually - e.g. if using a simplex minimizer - one cannot tell if one arrived at the global minimum or got stuck in a local one.

This is accomplished by introducing a new parameter, the so called temperature T . The underlying idea is to allow the system to switch from one state in the likelihood space to another. Therefore a series of Markov Chains with a different value of T are run in parallel. T equals one for the desired target distribution, but has a higher value for the other chains. This higher value leads to a flatter posterior distribution - therefore making it easier for the sampler to leave local minima. Obviously the flatter distribution is not what we want to sample from. But this does not matter because only the states that are produced by the $T = 1$ Markov Chain are saved. After a certain number of transitions two chains are chosen randomly and are interchanged with a certain probability r that is shown in eq. (6.10),

$$r = \min \left(1, \frac{\pi(\vec{x}_{t,i+1}|D, \beta_i, I)\pi(\vec{x}_{t,i}|D, \beta_{i+1}, I)}{\pi(\vec{x}_{t,i}|D, \beta_i, I)\pi(\vec{x}_{t,i+1}|D, \beta_{i+1}, I)} \right), \quad (6.10)$$

where i is the index for the chains with different T -values, β equals $\frac{1}{T}$ and π is given by eq. (6.11) [Greg10],

$$\begin{aligned}\pi(\vec{x}|D, \beta, I) &= p(\vec{x}|I) p(D|\vec{x}, I)^\beta \\ &= p(\vec{x}|I) \exp[\beta \ln[p(D|\vec{x}, I)]], \text{ for } 0 < \beta < 1.\end{aligned}\tag{6.11}$$

Using this method it is possible to leave local minima and to get the wanted result [Greg10], [ChQL01].

- *Autocorrelation*: The new state is proposed within a certain range of the recent state. Therefore one can expect subsequent parameter values within the whole Markov Chain to be correlated. On larger scales, however, the Markov Chain should not display a deterministic, but a random behavior. According to [Hans00] this correlation can be quantified by the autocorrelation function,

$$\rho(l) = \frac{1}{N} \sum_{i=1}^N y(i)y(i-l).\tag{6.12}$$

In eq. (6.12) $y(\vec{x})$ represents the sequence and l the lag.

- *ESS*: This is an abbreviation for effective sample size and is described by the following formula:

$$\text{ESS} = \frac{l_{\max} - i_{\text{start}}}{t_{\text{ac}}}.\tag{6.13}$$

The numerator is the difference of the maximum length l_{\max} and the start index i_{start} , which is divided by the autocorrelation time t_{ac} . It provides a good measure for the efficiency of an algorithm. The larger this value, the more uncorrelated and therefore usable samples are obtained [GiCa11], [Lan⁺12].

- $\frac{\text{ESS}}{t}$: This is simply the time-normalized effective sample size. It provides a reasonable measure to compare different algorithms that usually have very different runtimes.

6.4.2 Results

In this section some results for a simulated KATRIN likelihood are presented. Different runs for various starting conditions are made in order to enable an easy comparison with the other methods. Besides the obvious parameters that are to be compared - like the convergence time - the algorithms are also evaluated according to ESS and $\frac{\text{ESS}}{t}$ that were introduced in the previous section. A detailed comparison takes place in section 6.7. The simulations were carried out with a simplified final state distribution.

First example: Sampling around the minimum

For this first run the starting values were set to those of the minimum of the simulated spectrum:

- Endpoint energy: $E_0 = 18,575.0 \text{ eV}$
- Electron neutrino mass squared: $m_\nu^2 = 0.0 \text{ eV}^2$
- Signal amplitude: $R_{\text{sig}} = 1.0$
- Background rate: $R_{\text{back}} = 0.01 \text{ Hz}$

This choice of starting conditions assures that the sampling starts right away because the four parameters that are usually minimized are already at their minimal values. Therefore no burn-in period will be visible. The following settings were used:

- Seed: 12345
- Length: 50000
- Acceptance Rate: 23.4 %

The computation on the KATRIN computing cluster TESLA-cluster took about six minutes⁷. It yielded the following results (median):

- Endpoint energy: $E_0 = 18575.0(-0.00148|0.00139) \text{ eV}$
- Electron neutrino mass squared: $m_\nu^2 = -1.3(-816|826) \cdot 10^{-5} \text{ eV}^2$
- Signal amplitude: $R_{\text{sig}} = 99999.50(-1.72|1.71) \cdot 10^{-5}$
- Background rate: $R_{\text{back}} = 1000.01(-1.94|1.96) \cdot 10^{-5} \text{ Hz}$
- Parameter convergence time: 400, 1100, 400, 250
- ESS: 423
- $\frac{\text{ESS}}{t}$: $1.20 \frac{1}{\text{s}}$

The values in brackets are the 1σ -confidence intervals. All results are consistent with the true values of the simulation. The three figures 6.6, 6.7 and 6.8 illustrate the sampling process, the visualization of the correlation between two parameters and the autocorrelation function. These figures only give a brief insight in what is possible to visualize. The other parameters or three dimensional correlation plots could have been plotted as well. It is important to keep in mind that even though i.e. fig. 6.6 just shows m_ν^2 over time, the sampler always moves in the four dimensional parameter space. The other three parameters are sampled simultaneously.

⁷For comments about the runtime please see subsection 6.7.1.

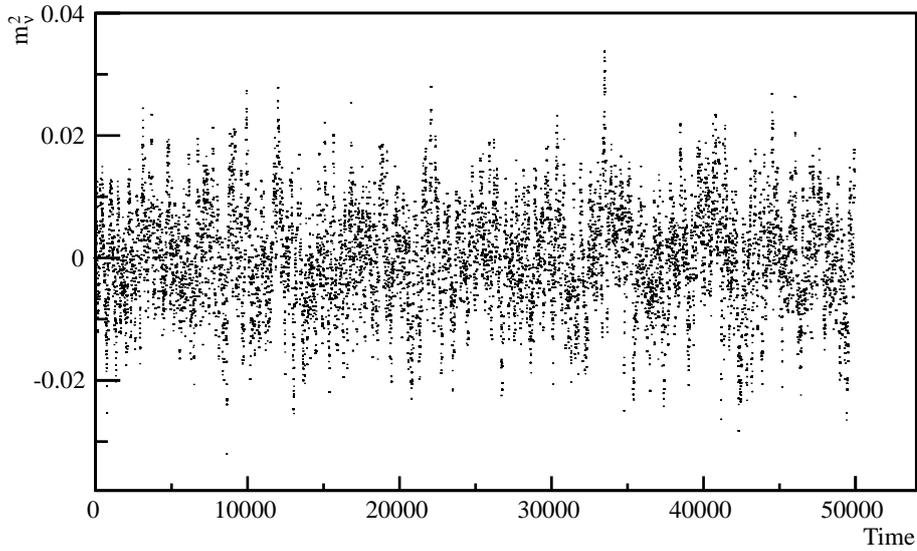


Figure 6.6: **Gaussian transition kernel, example 1:** Sampling of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and a Gaussian transition kernel around the minimum. The sampler walks nicely around zero which is the expected value.

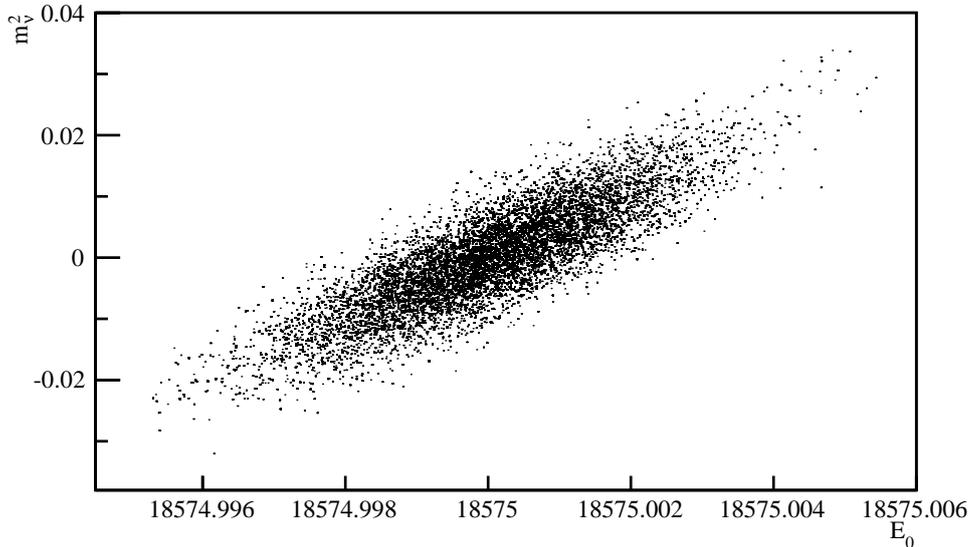


Figure 6.7: **Gaussian transition kernel, example 1:** Correlation of the neutrino mass m_ν^2 with the endpoint energy E_0 as obtained by sampling with the Metropolis Hastings algorithm and a Gaussian transition kernel. One observes a high correlation of 86.3 %

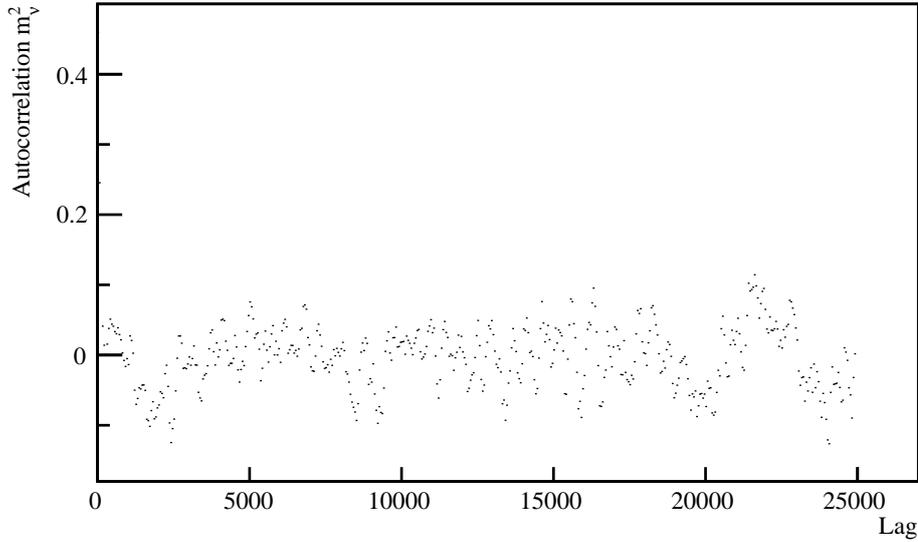


Figure 6.8: **Gaussian transition kernel, example 1:** Autocorrelation function for the neutrino mass m_ν^2 as obtained by sampling with the Metropolis Hastings algorithm and a Gaussian transition kernel.

Second example: Sampling away from the minimum

Now that we saw the behavior of the Metropolis Hastings algorithm when it starts at the minimum, it is time to look at something more realistic. What happens, for example, if we do not feed the program with the right values but start at a random position? The approach will test the ability of the algorithm to find the minimum. This is very important with real life data because usually the true value is not known. The following start values were chosen. These can be compared with the true input values from subsection 6.4.2.

- Endpoint energy: $E_0 = 18575.5 \text{ eV}$
- Electron neutrino mass squared: $m_\nu^2 = 0.2 \text{ eV}^2$
- Signal amplitude: $R_{\text{sig}} = 1.2$
- Background rate: $R_{\text{back}} = 0.013 \text{ Hz}$

Seed, length and acceptance rate were the same as in the previous run. By looking at the results we can see clearly that the Metropolis Hastings algorithm with the Gaussian transition kernel had no real problems with the intricate initial values. The true values lie within the confidence intervals of all the parameters.

- Endpoint energy: $E_0 = 18575.0(-0.00147|0.00143) \text{ eV}$
- Electron neutrino mass squared: $m_\nu^2 = -3.2(-86.8|86.0) \cdot 10^{-4} \text{ eV}^2$

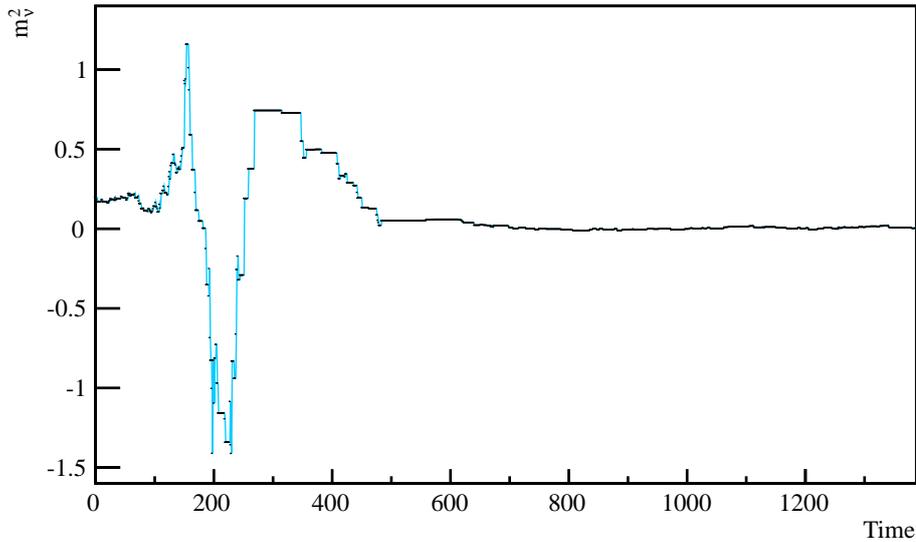


Figure 6.9: **Gaussian transition kernel, example 2:** Convergence towards the minimum of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and a Gaussian transition kernel. After about 750 steps it reaches the minimum and sampling begins.

- Signal amplitude: $R_{\text{sig}} = 100001.0(-17.7|18.4) \cdot 10^{-5}$
- Background rate: $R_{\text{back}} = 999.82(-1.89|2.09) \cdot 10^{-5}$ Hz
- Parameter convergence time: 950, 150, 750, 2150
- ESS: 5.01
- $\frac{\text{ESS}}{t}$: $0.0155 \frac{1}{\text{s}}$

Again the results can be visualized with some figures. The burn-in phase can be seen nicely in figures 6.9 and 6.10. In the latter it also becomes apparent that the sampler moves in the multi-dimensional parameter space. The figures 6.11, 6.12 and 6.13 are very similar to the respective illustrations in subsection 6.4.2.

Third example: Sampling with smearing

In this subsection an even more realistic approach is demonstrated. Not only are the start values of the parameters away from the minimum but also smearing is activated. At the beginning the sampler is almost at the same point as in subsection 6.4.2. Only the signal amplitude has a slightly lower value: $R_{\text{sig}} = 1.1$. Smearing, which was introduced in subsection 6.4.1, entails a certain blurring to the simulated counts. In real life this uncertainty can come through to unknown effects or simply because the values are Poisson distributed. Because of this uncertainty the clean results we have seen so far are not to be expected again, even though seed, length and acceptance rate remain unchanged to the previous runs. After the simulation the following median values for the parameters are obtained:

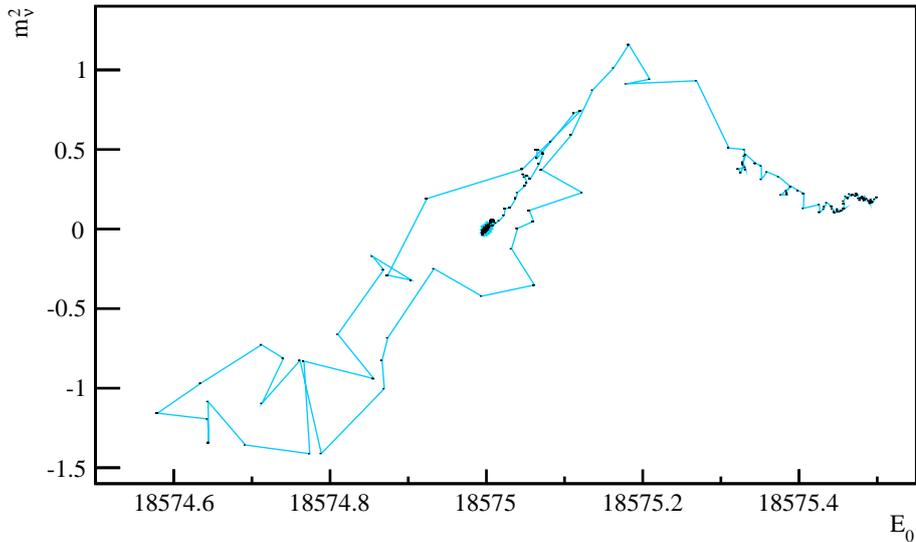


Figure 6.10: **Gaussian transition kernel, example 2:** Convergence towards the minimum of the sampler in two dimensions: the neutrino mass squared m_ν^2 and the endpoint energy E_0 . It starts at the right side, walks all the way down to the bottom left until it finally reaches the minimum in the middle of the figure. If one discards those first steps (that belong to the burn-in phase), we get the correlation plot depicted in figure 6.12.

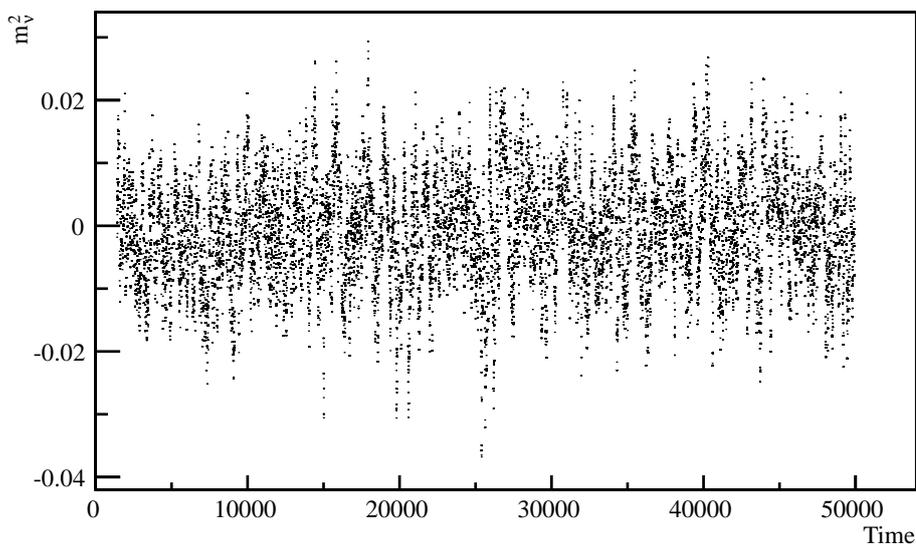


Figure 6.11: **Gaussian transition kernel, example 2:** After the burn-in phase is finished and the sampler has reached the minimum, the sampling of m_ν^2 begins. The Metropolis Hastings algorithm with the Gaussian transition kernel walks around the minimum and the result is very similar to that of figure 6.6.

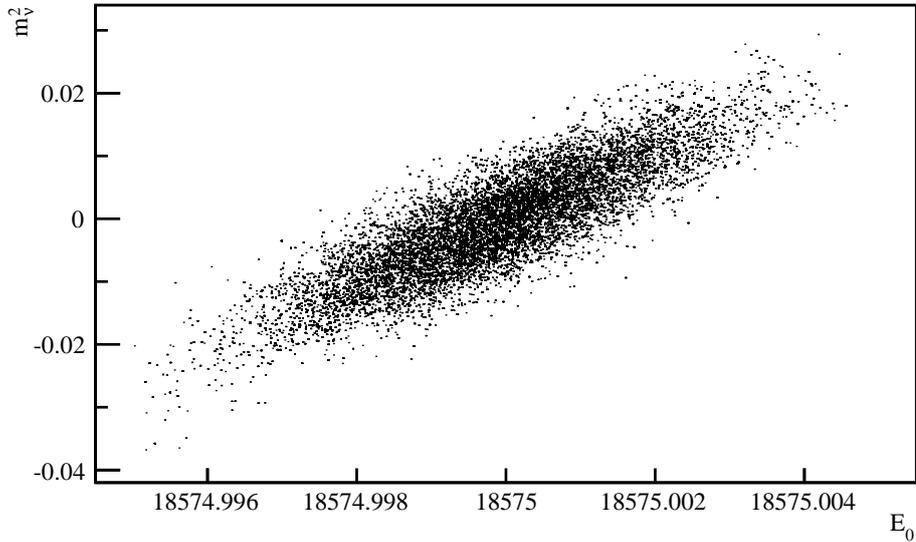


Figure 6.12: **Gaussian transition kernel, example 2:** Correlation of the neutrino mass m_ν^2 with the endpoint energy E_0 as obtained by sampling with the Metropolis Hastings algorithm and a Gaussian transition kernel. The values before the minimum is reached are discarded. One observes a correlation of 77.3 %

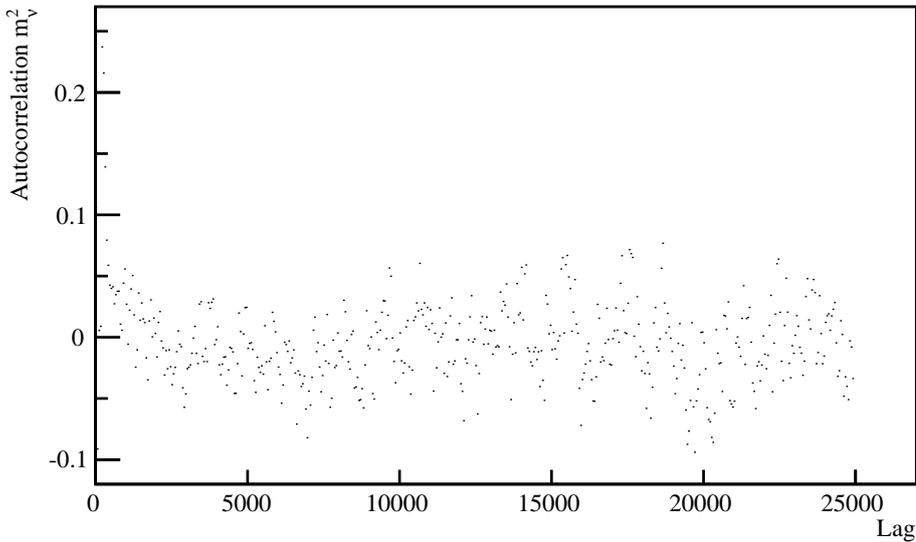


Figure 6.13: **Gaussian transition kernel, example 2:** Autocorrelation function for the neutrino mass m_ν^2 as obtained by sampling with the Metropolis Hastings algorithm and a Gaussian transition kernel.

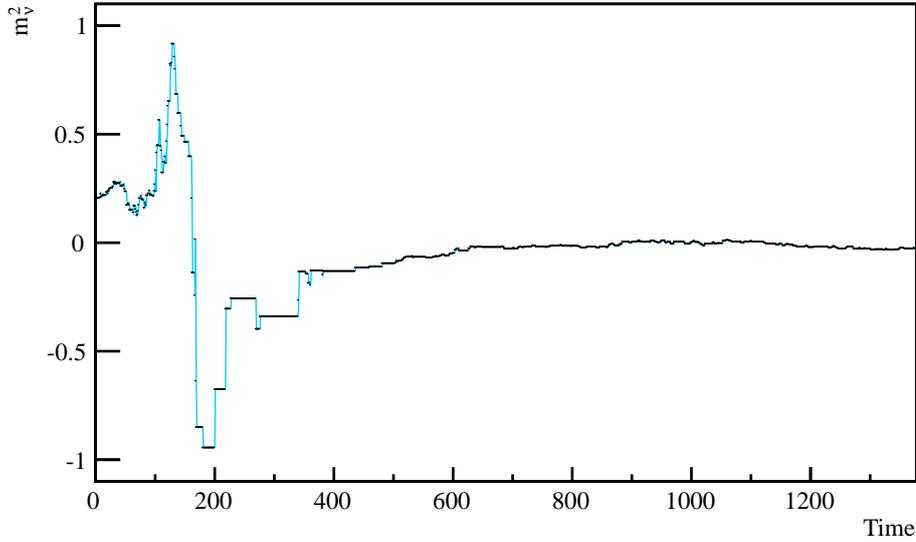


Figure 6.14: **Gaussian transition kernel, example 3:** Convergence towards the minimum of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and a Gaussian transition kernel. After about 850 steps it reaches the minimum and sampling begins. Smearing has almost no influence in this period.

- Endpoint energy: $E_0 = 18575.0(-0.00148|0.00153)$ eV
- Electron neutrino mass squared: $m_\nu^2 = -18.59(-9.04|9.10) \cdot 10^{-3}$ eV²
- Signal amplitude: $R_{\text{sig}} = 10002.30(-1.92|1.84) \cdot 10^{-4}$
- Background rate: $R_{\text{back}} = 996.70(-1.91|1.92) \cdot 10^{-5}$ Hz
- Parameter convergence time: 800, 850, 1200, 600
- ESS: 30.3
- $\frac{\text{ESS}}{t}$: $0.176 \frac{1}{s}$

They still describe the true values very well but the confidence intervals of m_ν^2 and of R_{sig} no longer yield the true values. This can be observed in figure 6.15: The sampler now does not walk along the value zero, but around -0.02 . Smearing, however, does not have a visible effect on the convergence. This can be seen in figure 6.14 which looks very similar to figure 6.9. The correlation plot of parameter m_ν^2 and E_0 (figure 6.16) does not change significantly, provided the burn-in phase is cut of. The same is true for the autocorrelation (figure 6.17).

6.4.3 Posterior Distribution

In the last three examples we have seen the behavior of the Metropolis Hastings algorithm with a Gaussian transition kernel. If it does not start at the minimum, the

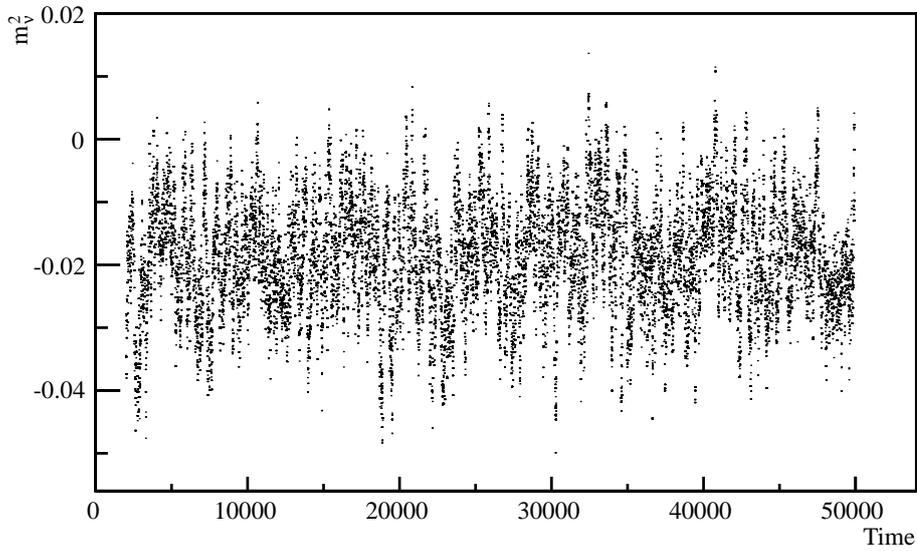


Figure 6.15: **Gaussian transition kernel, example 3:** After the burn-in phase is finished and the sampler has reached the final value, the sampling of m_ν^2 begins. The Metropolis Hastings algorithm with the Gaussian transition kernel walks around the minimum. Because of the activated smearing a slightly too low value is obtained.

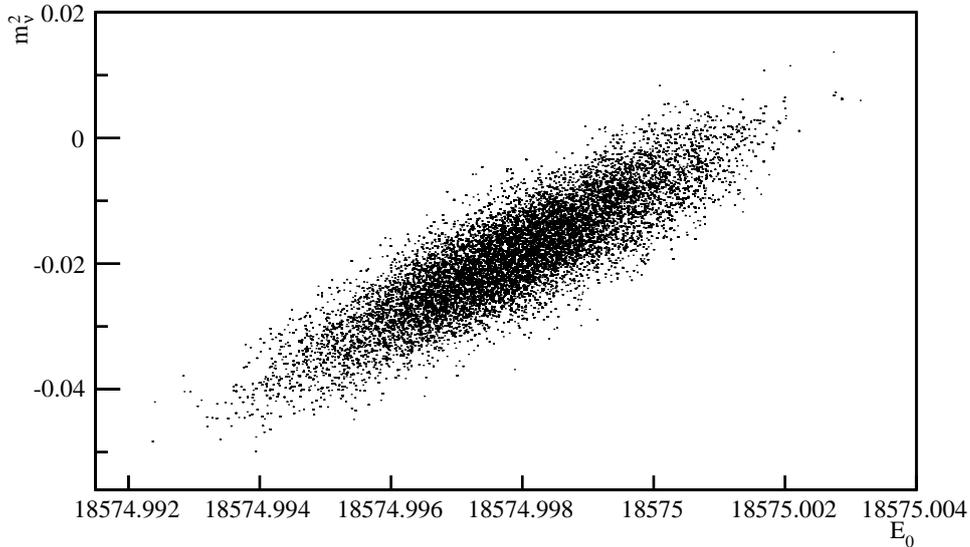


Figure 6.16: **Gaussian transition kernel, example 3:** Correlation of the neutrino mass m_ν^2 with the endpoint energy E_0 as obtained by sampling with the Metropolis Hastings algorithm and a Gaussian transition kernel. The values before the minimum is reached are discarded. One observes a correlation of 69.6%

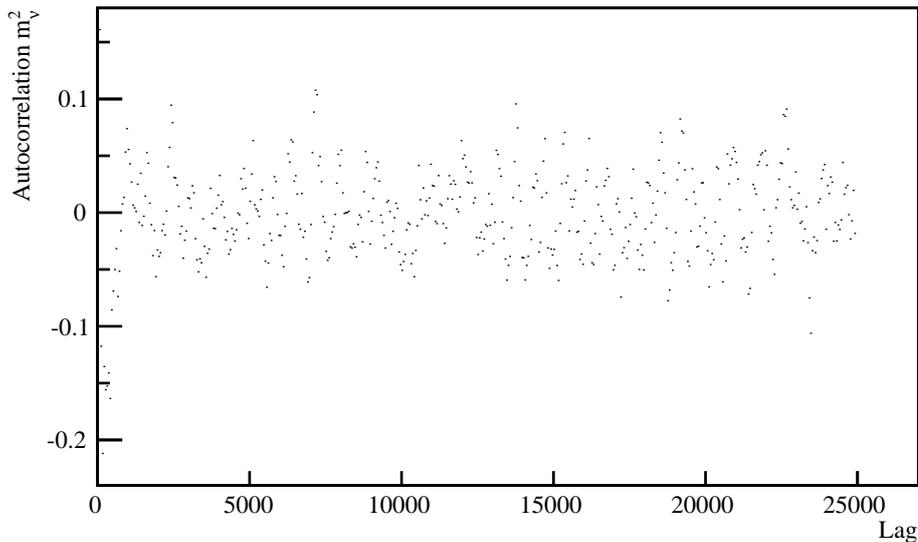


Figure 6.17: **Gaussian transition kernel, example 3:** Autocorrelation function for the neutrino mass m_ν^2 as obtained by sampling with the Metropolis Hastings algorithm and a Gaussian transition kernel and enabled smearing.

sampler will find it in the multi-dimensional parameter space (thanks to the evaluation with the Metropolis ratio) and then starts sampling. This sampling results in a Markov Chain that was visualized for the neutrino mass m_ν^2 for all three examples. The median of the values in the respective chains is the value that was given for the parameters. Beyond that it is possible to depict even more: For example the correlation of different parameters, as was shown for m_ν^2 and E_0 .

But how do we get the posterior distribution that was praised as being the real advantage of these Monte Carlo methods? The values of the Markov Chain (after the burn-in period) are simply filled into a histogram. After normalization we get a histogram with the posterior probabilities for this parameter. In figure 6.18 this was done for the Markov Chain of the neutrino mass m_ν^2 of our first example (the corresponding Markov Chain is depicted in figure 6.6). One can observe the probabilities for m_ν^2 being in one bin or accumulate the probabilities to see how likely it is that the neutrino mass lies within a certain interval. Besides the median that was already given in subsection 6.4.2 we can also calculate the mean, or read off the value with the highest probability:

$$\begin{aligned}
 m_{\nu, \text{median}}^2 &= -1.33 \cdot 10^{-5}, \\
 m_{\nu, \text{mean}}^2 &= +9.29 \cdot 10^{-6}, \\
 m_{\nu, \text{max. prob.}}^2 &= -9.27 \cdot 10^{-4}.
 \end{aligned}
 \tag{6.14}$$

6.5 Hamiltonian Monte Carlo

In the last chapter, we have seen that the Metropolis Hastings algorithm with a Gaussian transition kernel works fine for the sampling of the parameters. However,

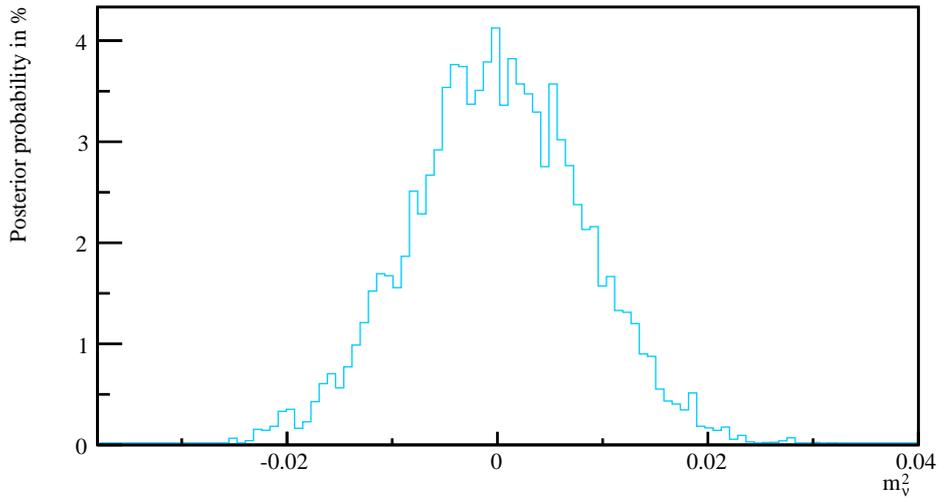


Figure 6.18: Normalized posterior distribution for the Markov Chain from the first example. The y-axis gives the posterior probability for m_ν^2 having the value on the x-axis, given the input data.

figures 6.9 and 6.14 show one shortcoming of this method: It takes a lot of steps to converge towards the minimum. Before it is reached the algorithm occasionally moves away in all directions. This random behavior does not come as a big surprise - the proposal function is symmetrical and therefore about half of the suggested states do not lie in the direction of the minimum - for each dimension. The evaluation with the Metropolis ratio still leads the sampler to the right direction, but one could wonder if there is not a more efficient way to do this. One could, for example, use the gradient information at the current position and let the sampler “walk” in this direction. Exactly this is carried out in the procedure presented in this section. Because of similarities to the classical Hamiltonian mechanics, this approach is called *Hamiltonian Monte Carlo* (HMC) and was first introduced by S. Duane, A. Kennedy, B. Pendleton and D. Roweth in 1987 [Neal12].

This more sophisticated method does not change the way the proposed states are evaluated. The Metropolis Hastings algorithm, which was described in 6.3.2, is still the underlying principle. That means we can still rely on the detailed balance equation (eq. (6.7)) to be fulfilled and therefore the proof of convergence (see section 6.3.3) to be valid. Only the way changes, how new states in the multi-dimensional likelihood space are proposed.

6.5.1 Concept of Hamiltonian Monte Carlo

We want to use the gradient information of the likelihood space in order to improve the efficiency of convergence. To do this we assign a mass to the sampler and interpret the likelihood distribution as some kind of gravitational potential. The consequence can be visualized with figure 6.19, where a ball is used to represent the sampler. This ball will roll towards the minimum. Its potential energy will become smaller and it will build up momentum. We define:

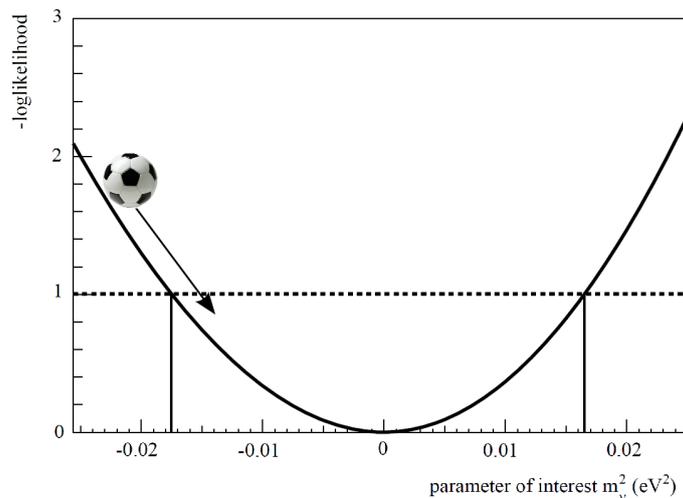


Figure 6.19: The sampler - here visualized with a soccer ball - is assigned a mass (the weight of the ball). If the likelihood distribution is understood as a gravitational potential, the ball will roll towards the minimum. If it experiences a rising slope then it can continue, thanks to its built-up momentum. After all the kinetic energy is transformed to potential energy, the ball will turn around and roll back down. This enables the sampling around the minimum but also - to some extent - the overcoming of local minima.

- A position variable \vec{q} , that describes a point in the parameter space. This parameter is saved in the Markov Chain.
- A momentum variable \vec{p} for the sampler. This goes hand in hand with the introduction of the mass of the moving sampler.
- An Hamiltonian energy function that can be seen in eq. (6.15). The kinetic energy $K(\vec{p})$ is equal to $\frac{|\vec{p}|^2}{2m}$ and the potential energy $U(\vec{q})$ is minus the logarithm of the probability density for the variables at the current position [Neal12],

$$H(\vec{p}, \vec{q}) = K(\vec{p}) + U(\vec{q}) = \frac{|\vec{p}|^2}{2m} - \log(L(\vec{q})). \quad (6.15)$$

This is completely analogous to classical mechanics. Therefore we can use the Hamilton equations of motion to describe the movement of the sampler in the parameter space:

$$\begin{aligned} \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i}, \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i}. \end{aligned} \quad (6.16)$$

The momentum variables define the direction of movement and make large step-sizes possible, without fear of a high rejection rate. The HMC algorithm obtains

a Markov Chain by alternating between updates of the position and momentum variables. The position values are saved and form the Markov Chain, from which - similar to the previous section - the posterior distribution can be obtained.

6.5.2 Implementation

The implementation that is described in this subsection was carried out as an extension of the code for MCMC with a Gaussian transition kernel. That means the programs that were written for the Metropolis-Hastings algorithm are still used. These basic functions were described in section 6.4.1 and in appendix B. The other files that are to be presented in this section can be found in the *Kasper/KaFit*-folder. The HMC program is started in the same way, only some adjustments in the configuration file are needed⁸.

In order to implement HMC on a computer a discretization of the time for the Hamilton equations (eq. (6.15)) is necessary. A standard approximation for a system of differential equations is Euler's method. The values for $p_i(t)$ and $q_i(t)$ are evaluated at discrete values of time. Sadly, it can be shown that \vec{p} diverges to infinity for simple examples [Neal12]. That is why we had to implement a more complicated approximation: the *leapfrog method*. The equations for $p_i(t)$ and $q_i(t)$ are again broken down to discrete values by evaluating them only at certain time steps. As advancement to Euler's method there is a third equation that represents a half-step. The whole method can be described with the following three equations [Neal12]:

$$p_i \left(t + \frac{\epsilon}{2} \right) = p_i(t) - \frac{\epsilon}{2} \frac{\partial U}{\partial q_i} (q(t)), \quad (6.17)$$

$$q_i(t + \epsilon) = q_i(t) - \epsilon \frac{p_i(t + \frac{\epsilon}{2})}{m_i}, \quad (6.18)$$

$$p_i(t + \epsilon) = p_i \left(t + \frac{\epsilon}{2} \right) - \frac{\epsilon}{2} \frac{\partial U}{\partial q_i} (q(t + \epsilon)). \quad (6.19)$$

Here m_i is the value for the assigned mass and ϵ is the length of a "frog-leap". In eq. (6.17) a half step for the momentum variable is performed. The new value of $p_i(t + \frac{\epsilon}{2})$ is used in eq. (6.18) for the calculation of a full step of the position. Finally, another half step for the position variables is executed (eq. (6.19)), again with the values obtained in the previous equation. To describe the whole time evolution these equations can be executed successively. Then it makes sense to combine eq. (6.17) and eq. (6.19) to form a whole step of the momentum. Only in the beginning and in the end of the algorithm a half step is performed. After a certain number of such leapfrog steps the new position is evaluated using the Metropolis-Hastings algorithm. If the new value is accepted it is saved in the Markov Chain. If, however, the evaluation fails because the Metropolis ratio worsened, the latest leapfrog-steps

⁸The sampling method has to be changed from "Metropolis" to "Hamiltonian". A number of frog-leaps (variable: *NumberOfFrogLeaps*) has to be defined. And the acceptance rate should be set to 65.1% for an efficient sampling process [Besk⁺10]. See also appendix C for an exemplary configuration file.

are discarded and the sampler moves back to the last accepted state. This can be visualized with the metaphor of a frog, which gave this algorithm its name: At the beginning the frog sits at the starting position. It has a fixed mass and a random initial momentum is diced. Then it starts jumping and the gravitational potential pulls it towards the minimum. After a pre-fixed number of leaps the new position is evaluated and the frog is either left at his current place or picked up and set back. In both cases the algorithm starts again and is repeated until the maximum length of the Markov Chain is reached. The formulas of this algorithm preserve the volume, are time-reversible and energy-preserving and therefore leave the posterior distribution invariant [ChQL01].

Even though the formulas eq. (6.17) - eq. (6.19), that form the core of the HMC algorithm, are quite short, the implementation is much longer. It took place in the *Walk-Routine* of the file *KFMCHamiltonian.cxx*. With 300 lines of code this program is about twice as long as the simpler Gaussian transition kernel. The extra lines are due to programming procedures and extra implementations like the treatment of constraints. This code covers the case that the sampler moves out of the allowed region. A simple way of handling such boundaries - for example that the neutrino mass squared cannot be negative - would be to simply evaluate the position after each run. The downside of this unpretentious algorithm is that if the sampler is indeed in a forbidden region it will simply be moved back. If the minimum lies near such a region this will happen a lot and the computation time will be increased massively. A better approach is the one of an infinite potential rise at the boundary. The sampler will move up the potential barrier and on the way lose all its kinetic energy. Then it stops, turns around and rolls back into an allowed region - it is literally reflected at the potential wall. Therefore the sampler never really enters forbidden regions without a significant increase in computation time [Neal12].

6.5.3 Results

Now that we have seen the concept of HMC, how it is implemented and how it can be used for the KATRIN analysis, the results of some sample-runs are to be presented. Following the approach of section 6.4.2 examples with different starting conditions are presented. The basis for the analysis is again a simulated KATRIN likelihood with a simplified final state distribution.

First example: Sampling around the minimum

This first example deals with the easiest starting conditions and will demonstrate HMC's ability to perform the sampling process. Therefore the sampler is set to the optimal values of the parameters, so that no minimum has to be found:

- Endpoint energy: $E_0 = 18575.0 \text{ eV}$
- Electron neutrino mass squared: $m_\nu^2 = 0.0 \text{ eV}^2$
- Signal amplitude: $R_{\text{sig}} = 1.0$
- Background rate: $R_{\text{back}} = 0.010 \text{ Hz}$

Besides that the following settings were used:

- Seed: 12345
- Length: 50000
- Acceptance Rate: 65.1 %
- Number of frog leaps: 10

Seed and length are the same as in section 6.4.2. The acceptance rate is tweaked to the optimal value for HMC and the new parameter *NumberOfFrogLeaps* is introduced [Besk⁺10]. It is the number of times that the algorithm “jumps”, before the position is evaluated with the Metropolis ratio. The values for the parameters that are determined by this algorithm are very well in correspondence with the true values:

- Endpoint energy: $E_0 = 18575.0(-0.00147|0.00147)$ eV
- Electron neutrino mass squared: $m_\nu^2 = -0.25(-8.70|8.93) \cdot 10^{-3}$ eV²
- Signal amplitude: $R_{\text{sig}} = 10000.1(-1.86|1.84) \cdot 10^{-4}$
- Background rate: $R_{\text{back}} = 1000.02(-2.02|2.01) \cdot 10^{-5}$ Hz
- Parameter convergence time: 200, 400, 200, 250
- ESS: 632
- $\frac{\text{ESS}}{t}$: $0.0321 \frac{1}{\text{s}}$

The graphs, too, show that the algorithm works very well. Figure 6.20 depicts the sampling process. The sampler moves symmetrically around the true value zero and the states are more randomly distributed than with the Gaussian transition kernel (see figure 6.6) - which is exactly what we want. The correlation is visualized in figure 6.21. Finally, figure 6.22 shows that there is almost no autocorrelation.

Second example: Sampling away from the minimum

Now that we have seen that the algorithm does a great job at sampling we can have a look at the convergence behavior. Therefore the starting values are set to those of section 6.4.2, which is relatively far away from the minimum. All the other parameters are the same as in the previous run. The results bear no surprises; the algorithm converges quite well.

- Endpoint energy: $E_0 = 18575.0(-0.00148|0.00146)$ eV
- Electron neutrino mass squared: $m_\nu^2 = -0.24(-8.74|8.85) \cdot 10^{-3}$ eV²
- Signal amplitude: $R_{\text{sig}} = 10000.10(-1.84|1.84)$

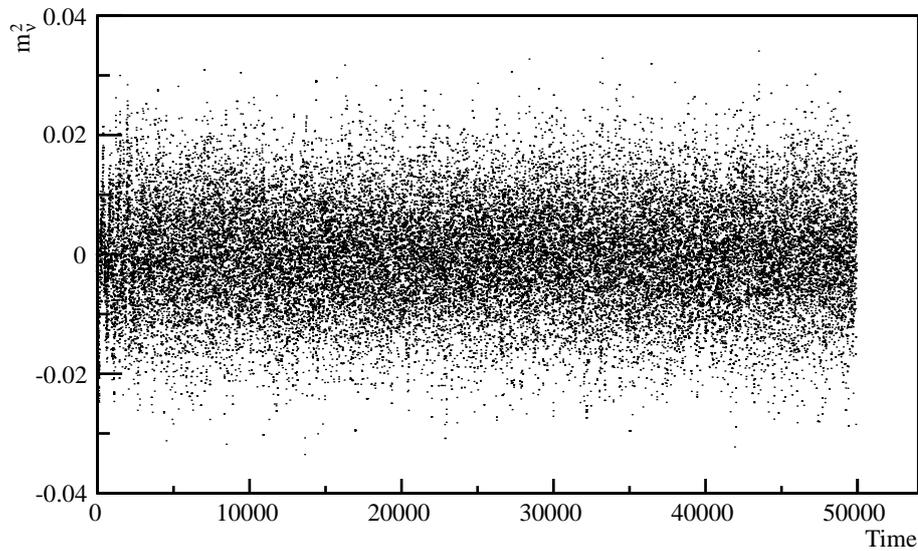


Figure 6.20: **HMC, example 1:** Sampling of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and the Hamiltonian Monte Carlo method around the minimum. The sampler walks nicely around the expected value zero.

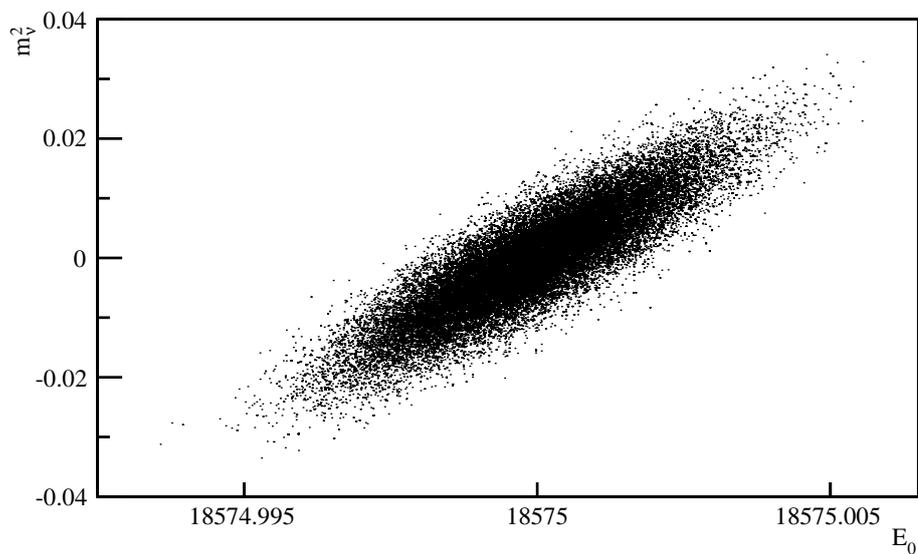


Figure 6.21: **HMC, example 1:** Correlation of the neutrino mass m_ν^2 with the endpoint energy E_0 as obtained by sampling with the Metropolis Hastings algorithm and the HMC method. One observes a high correlation of 87.3%.

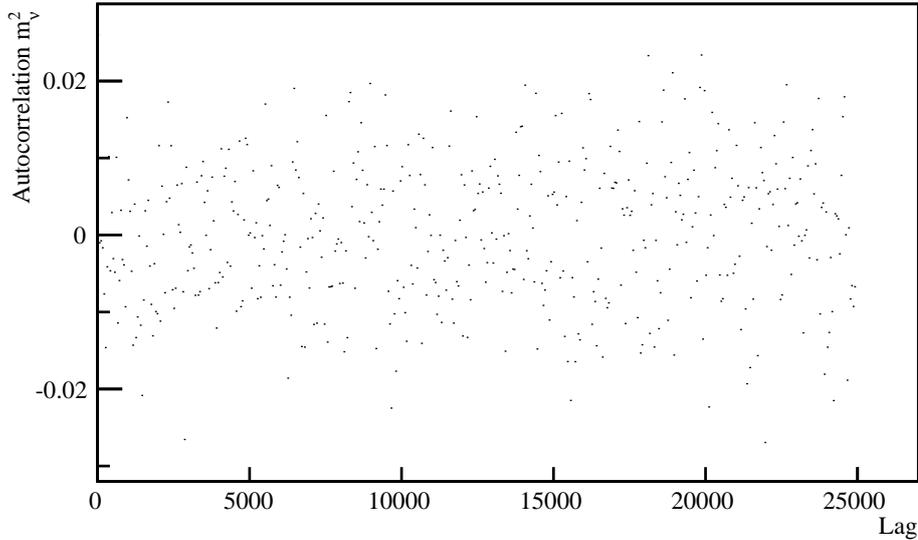


Figure 6.22: **HMC, example 1:** Autocorrelation function for the neutrino mass m_ν^2 as obtained by sampling with the Metropolis Hastings algorithm and HMC. The very symmetric distribution shows that there is almost no autocorrelation.

- Background rate: $R_{\text{back}} = 1000.03(-2.01|1.99) \cdot 10^{-3}$ Hz
- Parameter convergence time: 150, 150, 250, 350
- ESS: 464
- $\frac{\text{ESS}}{t}$: $0.0408 \frac{1}{\text{s}}$

Figures 6.23 and 6.24 make the differences to the simple Gaussian transition kernel even more apparent than the previous example. HMC converges very fast and after less than 20 steps the minimum is reached and sampling (figure 6.25) begins. The correlation plot (figure 6.26) looks very similar to the one before (figure 6.21) and the correlation has the same value of 87.3%. Finally, figure 6.27 gives the autocorrelation plot.

Third example: Sampling with smearing

This third test is carried out analogously to section 6.4.2. The start values are away from the minimum and an uncertainty - called smearing - is applied to the simulated rates. This makes the simulation realistic. The other settings are the same as in the last two example runs. Hamiltonian Monte Carlo shows no difficulties dealing with this new setup which can be seen in the following four graphs. Figure 6.28 shows that the gradient-led sampler still converges very fast. Once it is converged, the sampling begins (figure 6.29). Due to the smearing it is slightly shifted towards negative values. The correlation plot in figure 6.30 shows the same shift. The autocorrelation (figure 6.31) does not seem to be affected. The results underline what we have seen so far:

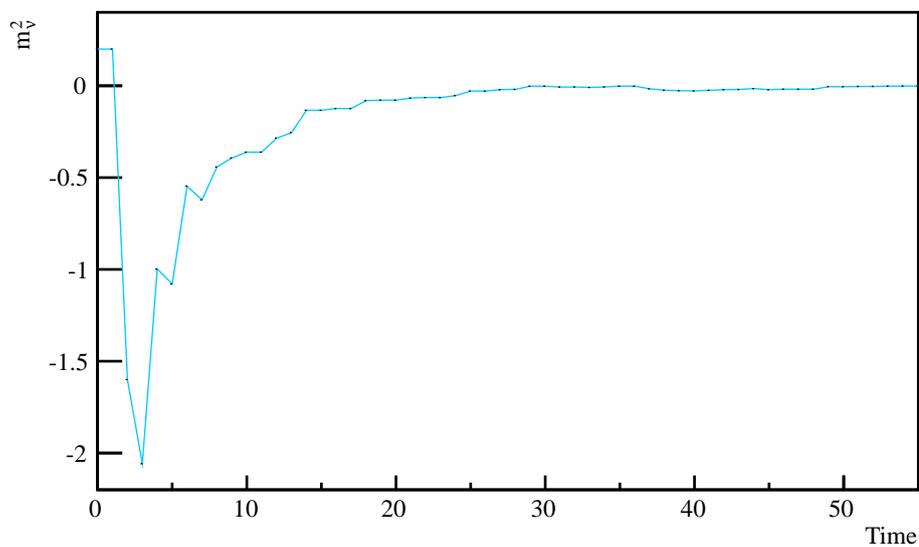


Figure 6.23: **HMC, example 2:** Convergence towards the minimum of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and the Hamiltonian Monte Carlo method. After less than 20 steps it reaches the minimum and sampling begins.

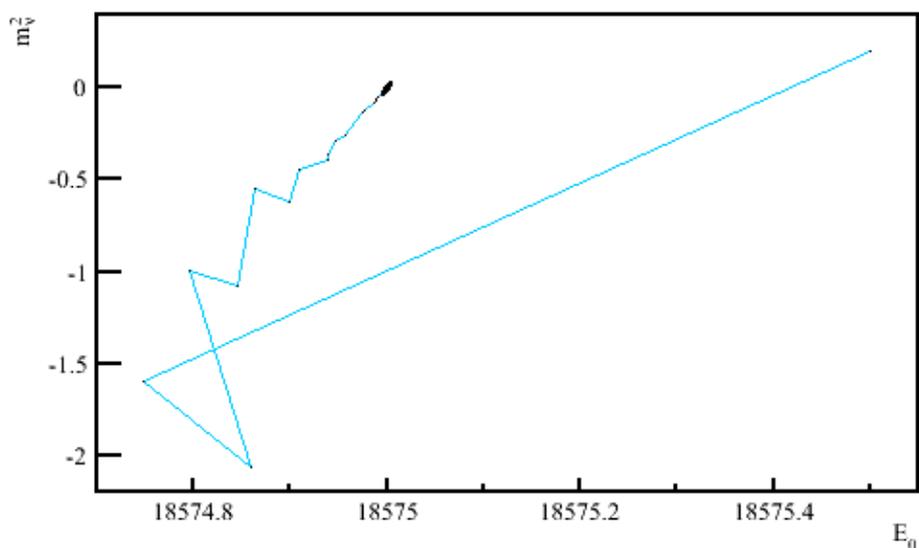


Figure 6.24: **HMC, example 2:** Convergence towards the minimum of the sampler in two dimensions: the neutrino mass squared m_ν^2 and the endpoint energy E_0 . It starts at the top right corner from where it moves quickly towards the minimum. This fast convergence is thanks to the use of gradient information. If one discards the first 20 steps the correlation plot depicted in figure 6.21 is obtained.

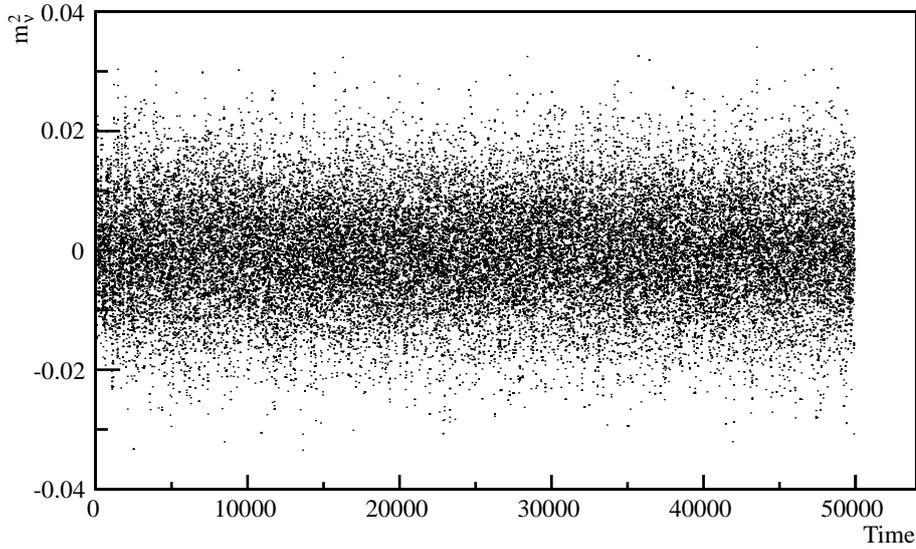


Figure 6.25: **HMC, example 2:** Sampling of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and the Hamiltonian Monte Carlo method around the minimum. The sampler walks nicely around the expected value zero.

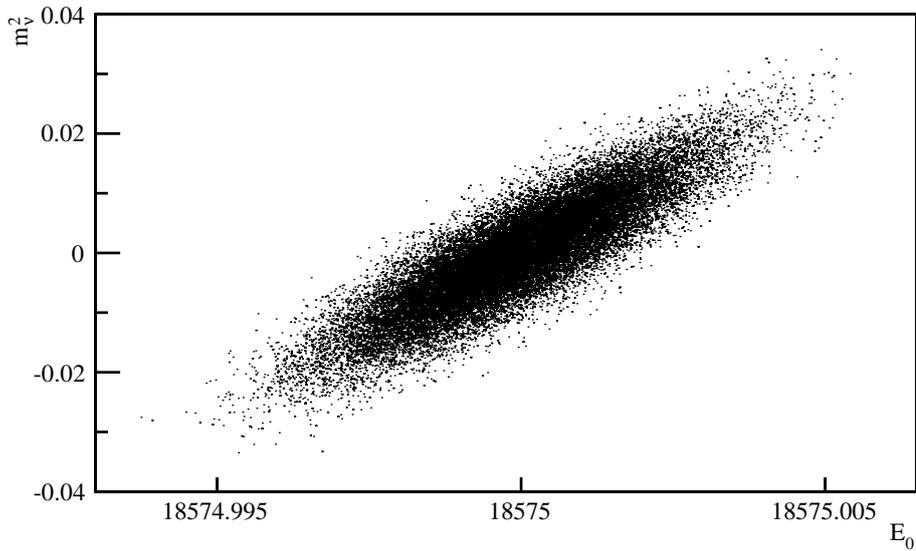


Figure 6.26: **HMC, example 2:** Correlation of the neutrino mass m_ν^2 with the endpoint energy E_0 as obtained by sampling with the Metropolis Hastings algorithm and the Hamiltonian Monte Carlo method. One observes a high correlation of 87.3 %.

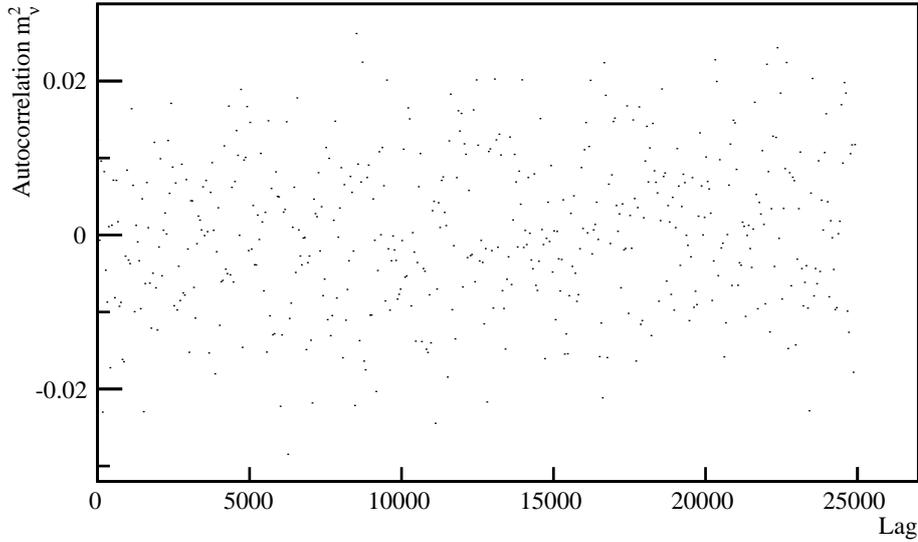


Figure 6.27: **HMC, example 2:** Autocorrelation function for the neutrino mass m_ν^2 as obtained by sampling with the Metropolis Hastings algorithm and the Hamilton Monte Carlo method. The very symmetric distribution shows that there is almost no autocorrelation.

- Endpoint energy: $E_0 = 18575.0(-0.00149|0.00146)$ eV
- Electron neutrino mass squared: $m_\nu^2 = -18.96(-8.77|8.83) \cdot 10^{-3}$ eV²
- Signal amplitude: $R_{\text{sig}} = 10002.30(-1.84|1.86) \cdot 10^{-4}$
- Background rate: $R_{\text{back}} = 996.65(-2.00|2.00) \cdot 10^{-5}$ Hz
- Parameter convergence time: 100, 100, 100, 350
- ESS: 303.061
- $\frac{\text{ESS}}{t}$: $0.0187185 \frac{1}{\text{s}}$

Besides the endpoint energy all the values are slightly too little. This is in good accordance with what one expects for enabled smearing.

6.5.4 Influence of the Parameters

In this subsection, the results of dozens of tests are shown on how the different parameters and routines influence the results.

- *Start values:* A variation of the start values is the most obvious modification. An insight into these tests was also given in the previous examples. Longer series of tests do not yield other results. When the start values are equal

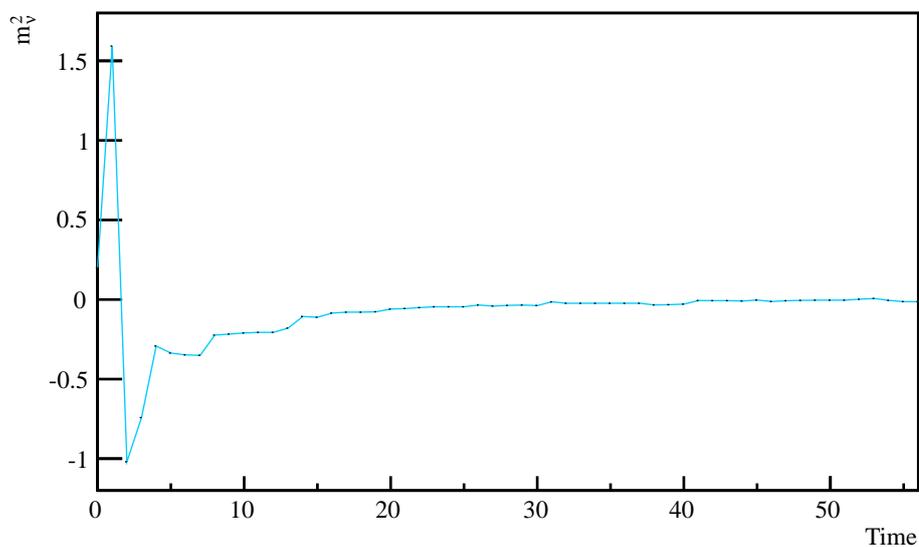


Figure 6.28: **HMC, example 3:** Convergence towards the minimum of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and HMC with activated smearing. After less than 20 steps it reaches the minimum and sampling begins.

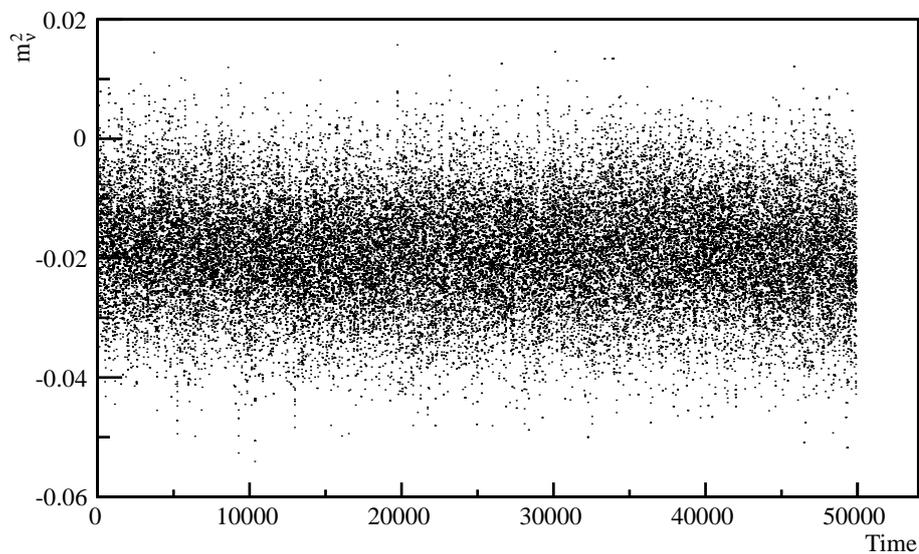


Figure 6.29: **HMC, example 3:** Sampling of the neutrino mass m_ν^2 with the Metropolis Hastings algorithm and HMC with activated smearing. The sampler walks around the determined minimum. Because of smearing this value is slightly smaller than the input value zero.

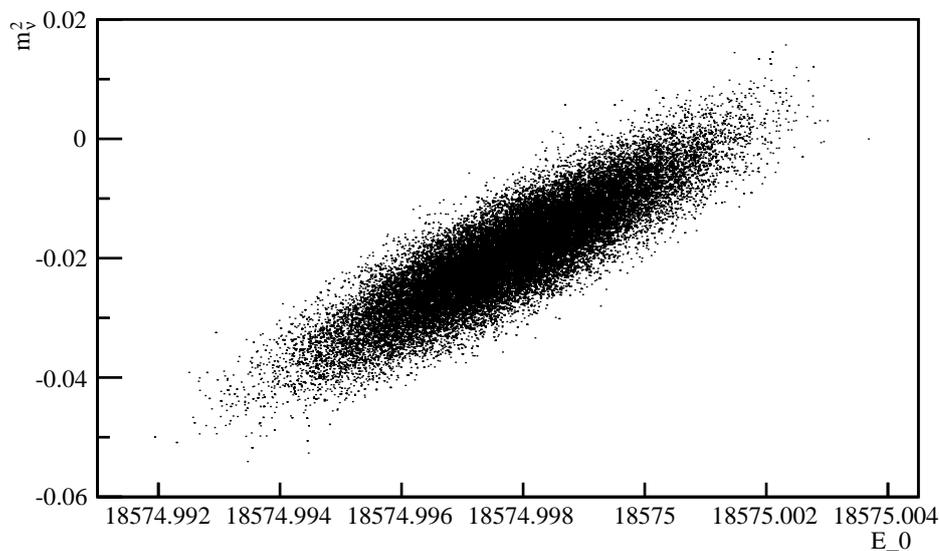


Figure 6.30: **HMC, example 3:** Correlation of the neutrino mass m_ν^2 with the endpoint energy E_0 as obtained by sampling with the Metropolis Hastings algorithm and HMC with activated smearing. One observes a high correlation of 87.4 % which is almost the same as without smearing.

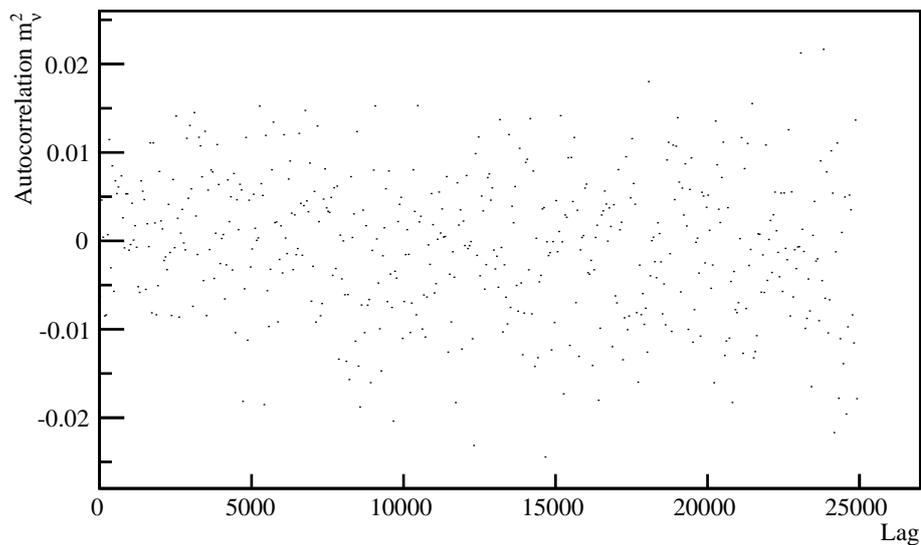


Figure 6.31: **HMC, example 3:** Autocorrelation function for the neutrino mass m_ν^2 as obtained by sampling with the Metropolis Hastings algorithm and the Hamilton Monte Carlo method with activated smearing. The very symmetric distribution shows that there is almost no autocorrelation - similar to the figure without smearing.

to those of the location of the minimum, sampling starts right away. If the sampler begins at a position away from the minimum it converges towards the minimum. This process may take a while - especially if one uses the Gaussian transition kernel instead of HMC. In extreme cases, e.g. if the starting values are nowhere near the minimum, it may happen that the algorithm does not converge at all. In this case the program will give an error message.

- *Optimal number of frog-leaps L* : This is the number of frog-leaps that are conducted before the new position is evaluated again with the Metropolis ratio. Small numbers lead to very little movement, but if the number is too big, a lot of additional calculations are required. This raises the runtime unnecessarily. Throughout literature there is no consensus on what the optimal number is. In my tests, however, it became apparent that stepsizes between five and 15 yield very good results for the KATRIN likelihood. Therefore I would recommend a value of ten, which leads to faster convergence than $L = 5$, but takes less calculation time than $L = 15$.
- *Variation of the number of frog-leaps*: The sampling process can be even more randomized by varying the number of frog-leaps. This was accomplished by adding a randomly chosen value of a Gaussian distribution to L . The normal distribution has a width of one. If the input value for L is ten the program will also sample with values between seven and 13 (and occasionally even smaller or bigger values).
- *Coupling of ϵ and L* : ϵ is the length of a frog-leap. That means the product of ϵ and L is the stepsize of the HMC algorithm. It makes sense to keep this stepsize at a fixed value that is reasonable for the parameter - for example the size of the error of that parameter. Running tests without this coupling show that it is still possible, but it does not achieve any improvement.
- *Stepsize and error control*: Tests have shown that a stepsize in the area of the errors of the parameters works very well. That is about 0.0015 for E_0 , 0.009 for m_ν^2 , 0.0002 for the signal amplitude and $2 \cdot 10^{-5}$ for the background rate. But if the values are too small or too big it does not matter - as long as *error control* is activated. This feature will compare the current acceptance rate with the target value and adjust the stepsize accordingly. After several cycles error control deactivates itself. The samples obtained while error control was still active should be discarded.
- *Influence of smearing*: As demonstrated in the previous examples, smearing has an influence on the results. The convergence towards the minimum is not significantly slower, but the position of the minimum may be changed slightly.⁹
- *Influence of boundaries*: As long as the minimum lies in the allowed region (and not outside the pre-set boundaries) the algorithm will still converge. If the boundary is close to the minimum, sharp cuts can be seen in the visualization.

⁹For HMC it is advisable to set the value for error scaling to 0.1, if smearing is activated.

- *Mass of the sampler:* The mass of the sampler was fixed to the value one. In the Riemannian Manifold Hamiltonian Monte Carlo algorithm the mass will play a more important role. This is discussed in the next chapter.

6.6 Riemannian Manifold Hamiltonian MC

In the last two sections, two powerful algorithms were presented: Markov Chain Monte Carlo with either a Gaussian transition kernel or with the Hamiltonian Monte Carlo method. Both programs are well implemented in the KATRIN-analysis framework and fulfill all of our expectations. The examples that have been presented prove that these Bayesian methods provide a good alternative to MINUIT. Thanks to their abilities in dealing with many dimensions and very complex likelihood distributions, it is also possible to make fits with even more parameters. The tritium purity or the high voltage fluctuations can easily be implemented as well - only minor changes in the configuration file are needed. Additionally, these algorithms, especially the HMC method, make it feasible to consider the effects of a sterile neutrino. This will result in a second minimum of the likelihood function, which is difficult to handle for ordinary algorithms. Thanks to parallel tempering the MCMC algorithms are able to deal with this complexity. This has been shown in [Haag13].

Even though these two algorithms seem to address all the issues one can imagine with the KATRIN data analysis, another, even more sophisticated approach was implemented: *Riemannian Manifold Hamiltonian Monte Carlo* (RMHMC). This method is presented in this section.

6.6.1 Concept of RMHMC

For very high dimensional distributions and very strong correlations and therefore really complex likelihood functions, the HMC method might not be sufficient. This can be exemplified with a 100-dimensional funnel distribution (see figure 6.32), which has been done in [Beta12]. The HMC stepsize does not take the curvature of the likelihood space into account. Therefore an efficient stepsize is too large for the small entry of the funnel and the HMC sampler is not able to explore it properly. The general idea behind RMHMC is to replace the mass that was introduced for HMC with a complex metric that takes the curvature of the space into account. This raises the computation time, but it may come in handy, if future modifications of the KATRIN experiment require the handling of highly curved likelihood distributions.

But let us start at the beginning. For the Hamiltonian Monte Carlo algorithm the kinetic energy has a simple form (see eq. (6.15)). With RMHMC a stepsize is wanted that can also explore areas with very high curvature. Therefore the kinetic energy is replaced with

$$T(\vec{p}, \vec{q}) = \frac{1}{2} \vec{p}^T \cdot \vec{\Sigma}^{-1}(\vec{q}) \cdot \vec{p} + \frac{1}{2} \log |\vec{\Sigma}(\vec{q})|, \quad (6.20)$$

where $\vec{\Sigma}(\vec{q})$ is a metric whose proper choice can “dynamically decorrelate and rescale the target distribution to avoid inefficiencies in the numerical integration, while also yielding a dynamic determinant whose variations can compensate for much larger variations in the potential” [Beta12]. The trajectories that one obtains with

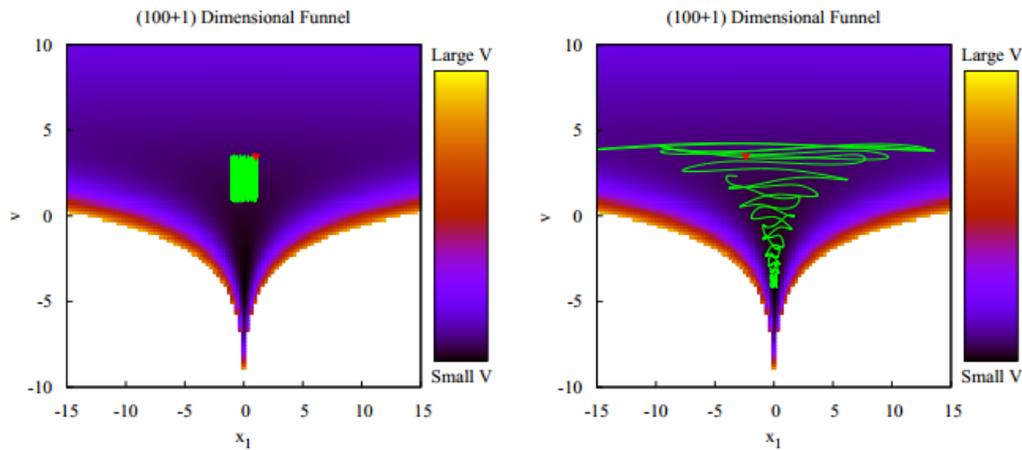


Figure 6.32: This picture from [Beta12] shows one of the advantages of RMHMC over HMC. A high-dimensional funnel with a high curvature is analyzed by both algorithms. HMC only explores a small area of the funnel, whereas the RMHMC (here with a diagonal SoftAbs metric) is able to explore the funnel more efficiently.

this algorithm, after applying the Hamiltonian equations of motion, are similar to geodesics on a Riemannian manifold, hence the name *Riemannian Manifold Hamiltonian Monte Carlo*. But what exactly should one choose for the metric $\vec{\Sigma}(\vec{q})$? One of the motivations for this algorithm was that it should be sensitive to the curvature. Therefore the metric should contain the second derivatives of the likelihood space. At first sight a reasonable choice seems to be the Hessian matrix,

$$H_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j}, \quad (6.21)$$

because in a convex neighborhood the target distribution can be approximated with the use of it [Beta12]. However this is only true if the target distribution is globally convex - something we cannot take for granted for all possible likelihood distributions. In 2012 Michael Betancourt presented a solution to this problem. His so-called *SoftAbs metric* represents a way to create a well-behaved metric from the Hessian matrix. This metric is based on exponential maps and requires a lot of complicated calculations.

6.6.2 Implementation

Like HMC, RMHMC uses many of the standard routines that were presented in section 6.4.1. The programs that are specific to RMHMC can be found in the files *KFMC Riemannian.cxx* and *KFMC Riemannian.h*. The main parts of the program can be found in the first file. It starts with the computation of the needed matrices (program-section *KFMC Riemannian::calculateMatrices*). In the following 600 lines of code all the other needed calculations are carried out, before the *Walk*-algorithm is started. This method is the heart of the program because it proposes the next step. In the end the new step is evaluated with the Metropolis-algorithm and either accepted or discarded. This implementation was carried out according to [Beta12]. Important parts of the algorithm are described in the following.

- At first an eigendecomposition of the Hessian \mathbf{H} is conducted:

$$\mathbf{H} = \mathbf{Q} \cdot \boldsymbol{\lambda} \cdot \mathbf{Q}^T. \quad (6.22)$$

\mathbf{Q} is the matrix of eigenvectors and $\boldsymbol{\lambda}$ is a matrix whose diagonal elements are the eigenvalues λ_i of \mathbf{H} . According to the SoftAbs algorithm, the metric for \mathbf{H} can be written as

$$\wr \mathbf{H} \wr = \mathbf{Q} \cdot \wr \boldsymbol{\lambda} \wr \cdot \mathbf{Q}^T, \quad (6.23)$$

where $\wr \boldsymbol{\lambda} \wr = \text{Diag}(\lambda_i \coth(\alpha \lambda_i))$.

- For the Hamiltonian evolution the program needs the gradient of the quadratic form, $\mathbf{p}^T \cdot \wr \mathbf{H} \wr^{-1} \cdot \mathbf{p}$,

$$\begin{aligned} \partial(\mathbf{p}^T \cdot \wr \mathbf{H} \wr^{-1} \cdot \mathbf{p}) &= \mathbf{p}^T \cdot \partial \wr \mathbf{H} \wr^{-1} \cdot \mathbf{p} \\ &= -\mathbf{p}^T \cdot \wr \mathbf{H} \wr^{-1} \cdot \partial \wr \mathbf{H} \wr \cdot \wr \mathbf{H} \wr^{-1} \cdot \mathbf{p} \\ &= -(\mathbf{Q}^T \cdot \mathbf{p})^T [\mathbf{J} \circ \mathbf{Q}^T \cdot \partial \mathbf{H} \cdot \mathbf{Q}] (\mathbf{Q}^T \cdot \mathbf{p}). \end{aligned} \quad (6.24)$$

For the last step the use of the Hadamard product \circ is needed¹⁰. \mathbf{J} is a matrix represented by

$$J_{ij} \equiv \frac{\lambda_i \coth(\alpha \lambda_i) - \lambda_j \coth(\alpha \lambda_j)}{\lambda_i - \lambda_j}. \quad (6.25)$$

- For the solution of the Hamiltonian equations the gradient of the log determinant, $\log |\wr \mathbf{H} \wr|$, is required:

$$\partial \log |\wr \mathbf{H} \wr| = \text{Tr}[\mathbf{Q}(\mathbf{R} \circ \mathbf{J})\mathbf{Q}^T \cdot \partial \mathbf{H}]. \quad (6.26)$$

The matrix \mathbf{R} is given by

$$\mathbf{R} = \text{Diag} \left(\frac{1}{\lambda_i \coth(\alpha \lambda_i)} \right). \quad (6.27)$$

6.6.3 Results

The algorithm was implemented according to [Beta12] and is fully functioning for simple likelihood distributions. However, it runs into difficulties for our KATRIN likelihood. This is due to the values that show up during the calculations. Some can be very big (in the area of $10^{20} - 10^{50}$ and even larger), while others almost tend towards zero. This results in problems because the numerical precision of the computer is not good enough to deal with a mixture of such extreme values. This is a topic of current “work in progress” and the matter will hopefully be resolved soon. Until this has happened the workings of this method will be demonstrated on a basic function:

$$L(x, y) = x^2 + y^2. \quad (6.28)$$

¹⁰The Hadamard product, as it is defined in [Mill07]: \mathbf{A} and \mathbf{B} are $m \times n$ matrices. The Hadamard product of \mathbf{A} and \mathbf{B} is given by $[\mathbf{A} \circ \mathbf{B}]_{ij} = [\mathbf{A}]_{ij}[\mathbf{B}]_{ij}$ for all $1 \leq i \leq m, 1 \leq j \leq n$.

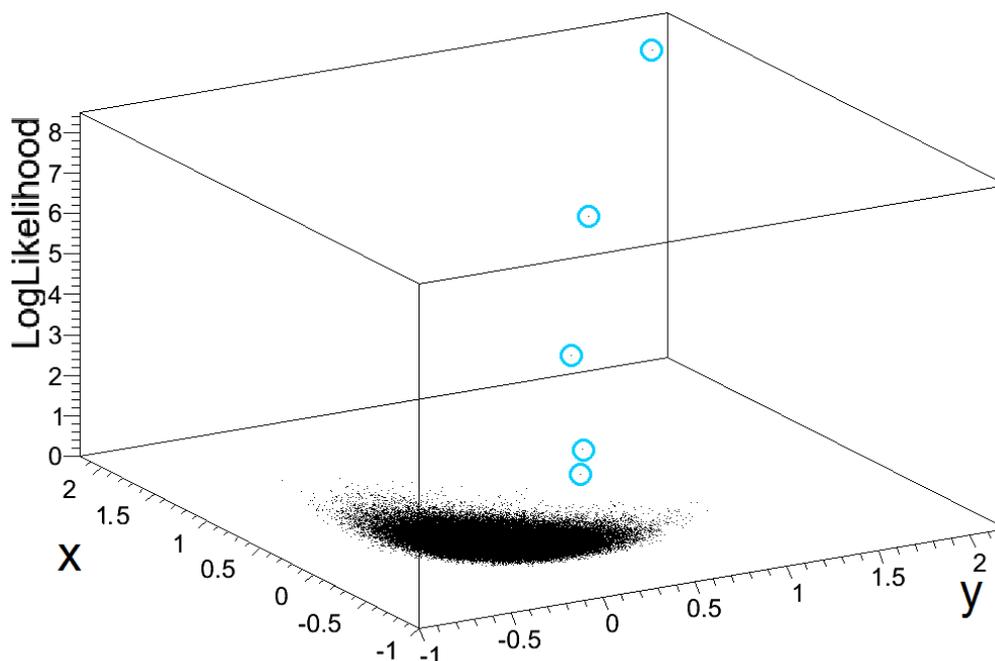


Figure 6.33: Convergence of the Riemannian Manifold Hamiltonian Monte Carlo method for a two-dimensional parabolic likelihood. The algorithm converges within a few steps towards the minimum at $x = y = 0$, where it begins sampling (see figure 6.34).

Calculations with modest input functions like this do not cause any problems, even when they are in more than two dimensions. The start values were $x = 2$ and $y = 2$. In figure 6.33 the convergence can be observed. Similar to HMC, this algorithm converges directly towards the minimum (which is at $x = y = 0$). During the sampling process the area near this minimum is explored and thereby an image of the likelihood function is created. This can be seen in figure 6.34, which clearly depicts the two-dimensional parabola. If the sampling is depicted in one dimension (see figure 6.35), the similarities to the other Monte Carlo methods, HMC and Gaussian transition kernel, can be seen clearly. The sampler walks around near the minimum and the values are added into the Markov Chain. From this file one can easily obtain the posterior distribution. Smearing and other variations are also possible for this algorithm.

When the RMHMC method is compared to other sampling approaches, it can be seen that the value of the time-normalized effective sample size $\frac{ESS}{t}$ is, for some likelihood functions, even higher than that of other methods (see e.g. [GiCa11]).

6.7 Comparison of the Algorithms

In this section the pros and cons of the different methods are discussed with regard to their runtime, the convergence time and the sampling process. This is done with particular emphasis on the comparison between the Gaussian transition kernel and

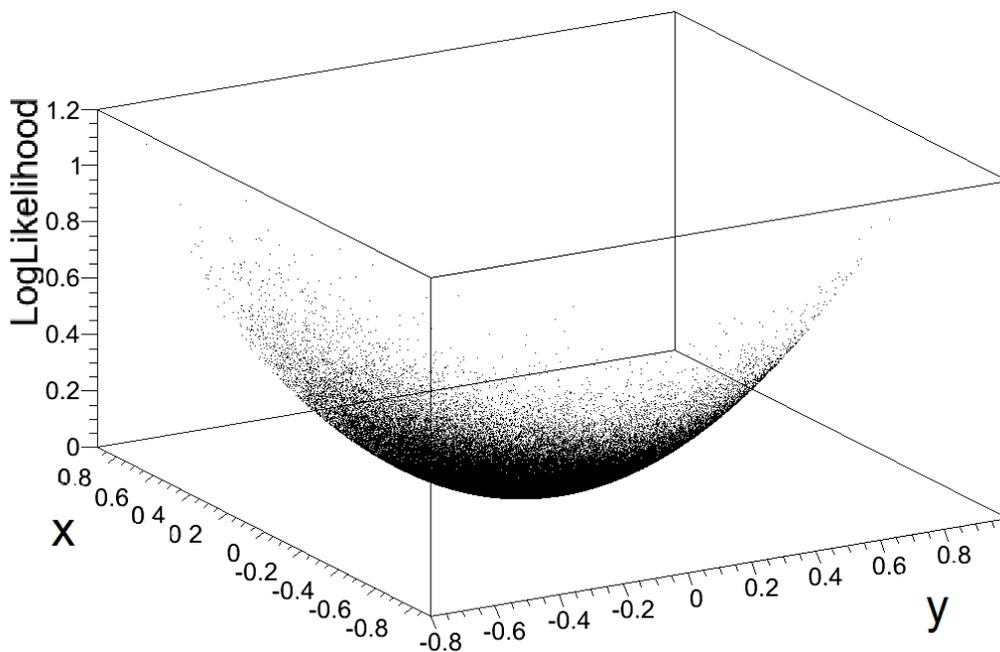


Figure 6.34: The sampling process of the RMHMC method clearly yields the input function, a two-dimensional parabola.

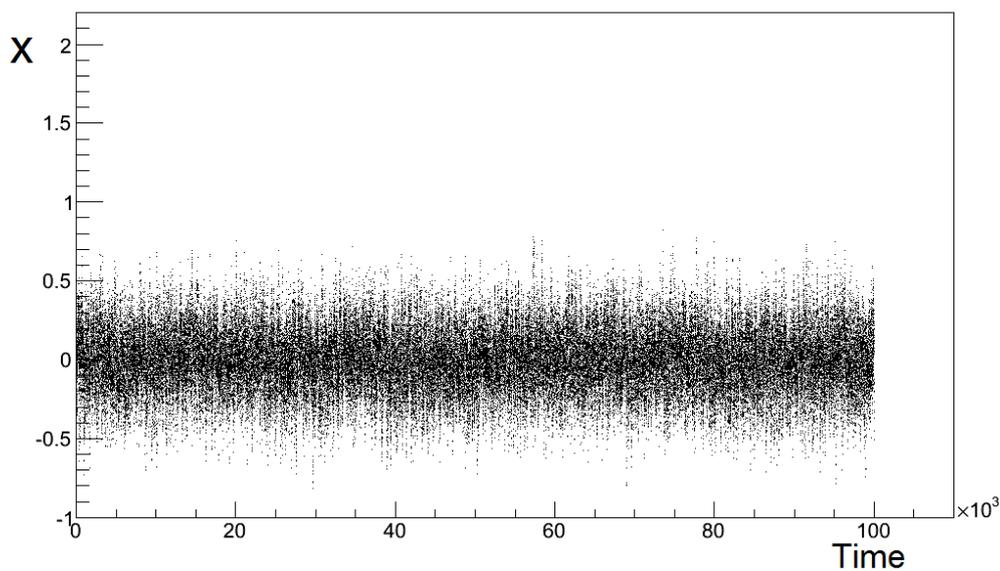


Figure 6.35: Sampling in one dimension results in a Markov Chain that is similar to those obtained by the other Monte Carlo methods, HMC and Gaussian transition kernel.

the Hamiltonian Monte Carlo method. These two have the highest significance for present-day KATRIN analysis.

6.7.1 Runtime and Effective Sample Size

The runtime was not always mentioned in the above calculations. This is due to the fact, that the KATRIN computation cluster, on which the calculations were performed, allocates its calculations capacity according to the demand. This means sometimes calculations are performed faster, just because a lot of cores and memory are not in use. However, the average computation times for the examples with the Metropolis algorithm with a Gaussian transition kernel (section 6.4.2) were in the area of five minutes. The Hamiltonian Monte Carlo examples (section 6.5.3) took between three hours and two days. This runtime seems to be extremely high compared to frequentist approaches like MINUIT. That is true and due to the fact that the calculations are more complicated. However, the runtime can be reduced significantly by disabling the computation of ESS. This is only needed for the classification of the algorithms and especially for their comparison, but does not influence the results in any way. Without ESS the runtime of the Metropolis algorithm with Gaussian transition kernel is reduced to less than one minute and for the Hamiltonian Monte Carlo algorithm it is reduced to about 45 minutes¹¹. This is in about the same area of the standard frequentist methods MINUIT and MINOS that need several minutes to accomplish a similar task - finding the minimum of the KATRIN likelihood and defining confidence limits.

This runtime difference is one of the most apparent differences between the two methods. At first sight the Gaussian transition kernel seems to be a clear winner. However, the values for the time-normalized effective sample size $\frac{\text{ESS}}{t}$ indicate that this is not always the case. This value shows how many usable samples (i.e. samples without autocorrelation) are created per time unit. Because the HMC method samples more effectively, it creates more usable samples and the ESS value is always up to ten times higher.

In the first and third example, the time-normalized effective sample size of HMC is about ten times smaller, but for the second example (the one where the start-values were the farthest away from the minimum) its value is about three times higher. This shows, again, the advantage of HMC, when confronted with complicated starting conditions.

The runtime for the RMHMC algorithm will be even higher than the one of HMC. The calculation of the gradient informations slows HMC down, but the calculations for RMHMC do not stop there. At some points even the third derivatives are required. However RMHMC can be thought of as an additional tool and never as a standard approach, so this high runtime should not matter. For some likelihood distributions the time-normalized effective sample size of RMHMC is even expected to improve [GiCa11].

¹¹Even though the second calculations were not performed on the computation cluster, but on a regular computer (E8400 @ 2 times 3.00 GHz).

6.7.2 Convergence Rate

Another obvious difference between the algorithms is the convergence time. Figures 6.9 and 6.28 show the steps it takes for the sampler to converge towards the minimum. The gradient information that is used in the HMC method leads the sampler straight towards the minimum, so that sampling can begin after less than 20 steps. The Gaussian transition kernel, however, does not have this information. This fact can be seen clearly in figure 6.9: The sampler walks around randomly and even moves away from the minimal value. It takes about 750 steps (and for the signal amplitude even more than 2000) to converge.

That means, if the approximate area of the minimal value is not known, the HMC method should be the instrument of choice. It will use gradient information to converge faster and therefore has a huge advantage over the basic Gaussian transition kernel. Only in cases where the likelihood function is so complicated that an evaluation with this method fails, the RMHMC approach should be used. Provided of course its implementation is complete - but the preliminary findings (see figure 6.33) are already very promising.

6.7.3 Sampling Process

Once the minimum is found the sampling process is the heart and soul of the algorithms. It has been demonstrated that all methods yield good results in this area (see figures 6.6, 6.11, 6.15 for the Gaussian transition kernel, figures 6.20, 6.25, 6.29 for HMC and figure 6.35 for RMHMC). However some differences become apparent. For the Gaussian transition kernel random-walk behavior can be observed (the vertical stripes in the figures). According to [Neal12] this avoidance of random-walk behavior is one of the major benefits of Hamiltonian Monte Carlo. The Markov Chains obtained with this method are very randomly distributed. This leads to the higher numbers for the effective sample size - more of the samples, obtained with HMC, can be used. The sample created with RMHMC looks similar to the one obtained using HMC. Here almost no random-walk behavior is to be observed, too.

6.7.4 Autocorrelation

The fact that HMC's effective sample size is so much higher is due to the instance that the autocorrelation of the samples is significantly lower. This can be seen when the figures 6.8, 6.13 and 6.17 (Gaussian transition kernel) are compared with 6.22, 6.27 and 6.31 (Hamiltonian Monte Carlo). The limits, between which the autocorrelation values lie, are about five times smaller for HMC. Also, some kind of random-walk behavior for the Gaussian transition kernel can be observed, which reflects the random walks we have seen in subsection 6.7.3.

6.7.5 Conclusion

In this chapter three Markov Chain Monte Carlo algorithms have been presented and their implementations into the KASPER-software, which is the KATRIN analysis framework, have been discussed. Two of these algorithms, MCMC with a Gaussian transition kernel and Hamiltonian Monte Carlo, work extremely well and cover everything that is required for the KATRIN analysis - now and in the near future. Also,

ground work has been done for the rare case that even more complicated likelihood functions need to be dealt with. For this case Riemannian Manifold Hamiltonian Monte Carlo was implemented into the analysis framework. Even though currently it cannot prove its full potential, it will soon be able to provide another good alternative to the well-established frequentist methods.

All of these Bayesian Markov Chain Monte Carlo methods have the advantage that they provide a direct way to obtain the posterior distribution (see figure 6.18), by sampling of the likelihood function. Thanks to their flexibility and their power to deal with a high number of parameters, they are also able to deal with likelihood distributions that implement a sterile neutrino, or even multiple sterile neutrinos. For this task the HMC method seems to be the most appropriate because of its better convergence behavior and the higher $\frac{ESS}{t}$ -value.

Despite the benefits over the frequentist methods that have been used for the simulated data analysis of KATRIN for a long time, the Bayesian methods are not only an alternative, but also an addition. They provide the chance to benchmark KATRIN results internally and compare different methods with each other. This will ensure a high quality analysis.

7. Summary and Outlook

During the course of this Diploma thesis we took a journey through the realms of neutrino- and astrophysics. We started at the very beginning, in 1930, with the postulation of a tiny particle, called neutrino, in order to save the laws of energy- and momentum conservation. After a basic introduction to the topic, we had an excessive look at the data analysis and simulation of the KATRIN experiment that has the ability to unravel one of the last secrets of neutrinos: their absolute mass.

The neutrino was detected less than 60 years ago and still not all of its properties are completely understood. The measurement of one of its features, the ν -mass, is a focus of modern research because of its huge implications for astrophysics and cosmology. There are several different approaches to measure this mass, e.g. using CMBR and LSS data, measuring supernova neutrinos or β -decay experiments. The KATRIN experiment is a direct, model-independent measurement, which is more precise than any of its predecessors.

It can be roughly divided into four main parts. The WGTS is the tritium source that ensures a stable flow of β -electrons that originate from the decay of tritium. These electrons are then guided through the transport section towards the spectrometers. The DPS and CPS that make up the transport section ensure that no tritium enters the ultra-high vacuum areas. The main spectrometer is operated as a MAC-E filter and allows only electrons above a certain energy threshold to pass. These electrons are measured inside the focal plane detector. With the use of extensive data analysis algorithms the mass of the electron antineutrino can be extracted from the resulting integrated energy spectrum.

After these introductory chapters, this work was dedicated to a fascinating part of astrophysics: the dark matter phenomenon. Even though it makes up the majority of matter in the universe it is still unknown today what dark matter consists of. A possible candidate for a dark matter particle is a sterile neutrino in the keV range. This hypothetical particle only interacts gravitationally and is exceptionally difficult to determine. However, sterile neutrinos do mix with regular neutrinos. Provided the mixing angle is not too small, their existence could be accountable for a kink

in the spectra of β -decay experiments. A kink that a future version of KATRIN might be able to discover. To do so, KATRIN needs to be modified to measure not only the endpoint region, but the entire range of the decay-spectrum. Also, the data analysis and simulation need to be augmented to meet the new requirements. It was the task of this thesis to implement some modifications for this purpose.

The most important corrections for the tritium β -spectrum were shown in chapter 5. The complete Fermi function, for one thing, takes the electro-magnetic interactions between the β -electron and the daughter nuclei HeT^+ into account. Additionally, radiative corrections that can be used over the whole range were implemented into the software framework. These take care of the energetic influences of emitted real and virtual photons. The excitations of the daughter nuclei are accountable for energy “lost” in the decay and need to be considered as well. For the standard KATRIN experiment the excitation energy is described with a final state distribution. For this thesis a more complete function was implemented.

All of these corrections are integrated into the framework and can be used to simulate a spectrum. They have a significant influence on the shape of the spectrum, which is raised on average by about half of its original height. The influence of the different corrections was exemplified in figure 5.7.

The other main and even more extensive part of this work are new tools for the KATRIN data analysis. The implementation of Markov Chain Monte Carlo (MCMC) methods can be used for two things. On the one hand, it is vital for the future analysis of sterile neutrino data. The additional parameters that are needed for the proper description of at least one further neutrino render the analysis for basic frequentist methods almost impossible. MCMC methods, however, are able to handle many free parameters with ease. They can, on the other hand, also be used today. As Bayesian methods, they provide the opportunity to easily obtain credibility intervals and parameter correlations. As an alternative approach to the already implemented ways of data analysis, they represent a chance to benchmark and double check results internally and verify the credibility of results.

The discussed methods include MCMC with a Gaussian transition kernel, Hamiltonian Monte Carlo and Riemannian Manifold Hamiltonian Monte Carlo. The former is the most straight-forward algorithm. The sampler walks around randomly in the likelihood space and is evaluated with the use of the Metropolis ratio. This method is quite fast and delivers good results, however, it takes many steps to converge towards the minimum and the obtained samples display some autocorrelation.

Hamiltonian Monte Carlo uses gradient information to guide the sampler towards the minimum. This permits fast convergence and, thanks to the specific architecture of the algorithm, it displays almost no autocorrelation. It is better suited for more complicated likelihood functions, which makes it a great match for the requirements of the sterile neutrino analysis.

A way to deal with even more complicated input functions is represented by Riemannian Manifold Hamiltonian Monte Carlo. This algorithm takes the curvature of the space into account and shows its strength analyzing curved distributions like a funnel. Such functions are not expected for the sterile neutrino analysis, therefore it plays a subordinate role within the KATRIN data analysis. It can, nevertheless,

be an important tool to validate the results of the other algorithms, even though its computation time is significantly higher. It will also come in handy if new likelihood functions that display a high curvature, need to be analyzed in the future.

A detailed comparison of the algorithms can be given in section 6.7.

KATRIN will start taking data for the electron antineutrino mass measurements in 2015. The measurements will take about three years, in order to collect enough data for the sensitive analysis. Even if sterile neutrino measurements could start right afterwards, it would leave us with plenty of time to improve the simulation and analysis software even further. MCMC with Gaussian transition kernel and HMC do not require any form of modification, but the intricacies of RMHMC open some areas of further research.

In the case that the sterile neutrino mixing angle is very small, it may also be advisable to implement even more corrections for the spectrum calculation. Possible effects were presented in section 5.4. They can be added to the program-branch that was written for this thesis.

With the use of the implemented corrections and analysis algorithms, KATRIN has a great chance to pin down the mass of the electron antineutrino. The new spectrum calculations and analysis methods also extend KATRIN's abilities towards another ambitious project: the discovery of the mysterious particle that is the sterile neutrino. Both observations would be of great significance, as they complete mankind's understanding of astrophysics and cosmology and help us perceive "whatever holds the world together in its inmost folds" [Goet08].

A. Implementation of the Corrections for the Tritium β -Spectrum

It was shown in chapter 5 that in order to be able to detect sterile neutrinos with a modified version of the KATRIN experiment a new, more detailed tritium β -spectrum is required. It is shown in this appendix how these corrections are implemented into the KASPER analysis framework. The used programming language is C++ and the code is fully functional.

Comments within the code enable an easy understanding.

- All of the corrections can be found in a separate branch within the KASIOPEIA-framework. The name of the branch is *sschams_SpectrumComplete*. The following files can be found in that branch (the file extensions are .cxx and .h).
- *KFSandbox*: This is the program where all the routines are started. It also includes algorithms that enable comparisons between the different corrections.
- *KSCDifferentialSpectrumComplete*: In this file the code with the full Fermi function and the radiative corrections can be found. There are also approximations for the Fermi function, in case this is needed. The final state distribution is not dealt with in the same way. Because of the structure how the final states are implemented in KASPER, they have to be in a separate file (see items *SSCFinalStatesNew* and *FSD_Tail_Universal.txt*).
- *SSCDifferentialSpectrumFSD*: In this program the regular corrections that are only valid in the endpoint region can be found. This is important for comparisons with the regular KATRIN spectrum calculations.
- *SSCFinalStatesNew*: The code in this file deals with the final state distribution. The file “FSD_Tail_Universal.txt” is used for the full distribution.

- *FSD_Tail_Universal.txt*: This file includes a table of the calculated final state distribution for the whole decay spectrum.

Beyond that there were two calculations that might require further information:

- For the *Fermi corrections* the $\Gamma(a)$ -function with a complex argument was needed. C++ and ROOT cannot handle this intricate function. So the problem was solved by the implementation of the GNU scientific library.
- The *radiative corrections* require the Spence function $L(z) = \int_0^z \frac{\ln|1-t|}{t} dt$. This integral was approximated to a very high precision using the following identity and a for-loop:

$$L(z) = \int_0^z \frac{\ln|1-t|}{t} dt = \sum_{k=1}^{\infty} \frac{z^k}{k^2}. \quad (\text{A.1})$$

This approximation is only valid if z is not close to 1, which is the case.

B. Implementation of the MCMC Analysis Methods

In the following an overview over the implementation of the Markov Chain Monte Carlo methods into the KASPER-framework is given. The implementation will, in all likelihood, be described in further detail in [Haag14].

The presented programs can be found in the *Kasper/KaFit*-folder, mainly in the *MCMC*-sub-folder. These basic routines also provide the foundation for the Hamiltonian Monte Carlo and Riemannian Manifold Hamiltonian Monte Carlo methods (see sections 6.5 and 6.6).

Just a reminder: We want to get a sample of the posterior distribution by only using the KATRIN likelihood and prior information.

- The routine is started from the shell by typing `./KFmcmc`.
- A configuration file contains all the necessary configuration data. It can be found and edited in the *config*-folder. In this file all the parameters can be set and the method (“Metropolis”, “Hamiltonian”, or “Riemannian”) can be chosen. A typical configuration file is shown in appendix C.
- The Markov Chain starts either from a randomly chosen point, or from the value specified in the config file. This is initialized by the file *KFMarkovChainMC.cxx* where the main procedures can be found.
- In *KFMCMetropolis.cxx* we find the two steps of the Metropolis-Hastings algorithm. “Walk” suggests a new state and “Test” evaluates this proposed state according to the Metropolis ratio.
- *KFMCMCState.cxx* provides us with the current state if this is needed.
- The file *KFMCMCChain.cxx* contains the list with all the elements that are the Markov Chain. Using this program we can get different values, like the Median of the Markov Chain.

- With the use of an input function the program *KFMCMCProposal.cxx* finds the next state which is then evaluated in *KFMCMetropolis.cxx*. This is also the place where boundaries are implemented.

C. Configuration File for the MCMC Analysis

Simulations within the KATRIN analysis framework KASPER can be started from the shell of any Linux system where it is installed. Simulations with Markov Chain Monte Carlo (see chapter 6) for example are executed with the command *KFmcmc*. All the settings for the simulation do not need to be set for each run separately; they can be predefined in a configuration file. An example for such a configuration file is following.

```
<KaFit
  OutputFile="{output=HamAwayTest.root}"
  NFits="{nfits=100}"
  Fitter="myMinuit2"
  Seed="{seed=12345}"
/>

<Fitter Type="KFMinuit2"
  Name="myMinuit2"
  EstimationMethod="myKatrinLogL"
  RandomizeStartValues="True"

  Minimizer="minimize"
  Strategy="2"
  UseMinos="true">

  <Parameter Name="E0" StartValue="0.0" Error="2.0"
LowerLimit="-200.0" UpperLimit="200.0" Fixed="false" />
  <Parameter Name="mnu2" StartValue="0.0" Error="2.0"
LowerLimit="-200.0" UpperLimit="200.0" Fixed="false" />
```

```
<Parameter Name="Sig" StartValue="1.0" Error="0.05"  
LowerLimit="0.0" Fixed="false" />  
<Parameter Name="Bg" StartValue="0.01" Error="0.01"  
LowerLimit="0.0" Fixed="false" />  
<Parameter Name="epsT" StartValue="0.95" Error="0.001"  
UpperLimit="1.0" LowerLimit="0.5" Fixed="false" />  
<Parameter Name="varQU" StartValue="0.0" Error="0.01"  
LowerLimit="0.0" Fixed="true" />
```

```
</Fitter>
```

```
<Fitter Type="KFmcmc"  
Name="myMarkovChain"  
EstimationMethod="myKatrinLogL"  
RandomizeStartValues="False"
```

```
Sampler="Hamiltonian"
```

```
Epsilon="1.0"
```

```
Alpha="0.1"
```

```
NumberOfFrogLeaps="10"
```

```
Length="{length=50000}"  
Thinning="1"  
FunctionType="LogLikelihood"
```

```
ErrorScaling="{scaling=0.1}"
```

```
CycleLength="50"
```

```
BurnInModeCycles="0"  
ErrorControlMode="0"  
AcceptanceRate="0.651"  
ErrorControlTolerance1="1.5"
```

```
ErrorControlTolerance2="1.8"  
ErrorControlDampening="0.7"  
ErrorControlDeactivation="50"
```

```
ParameterProbingCycleLength="50"

ParallelTemperingFrequency="50"
Betas="1.0"
>

  <Parameter Name="E0"    StartValue="18575.0"  Error="0.0015"
LowerLimit="18525.0" UpperLimit="18600.0" />
  <Parameter Name="mnu2"  StartValue="0.0"    Error="0.0085"
LowerLimit="-10.0" UpperLimit="10.0" />
  <Parameter Name="Sig"   StartValue="1.0"    Error="0.00019"
LowerLimit="0.0" />
  <Parameter Name="Bg"    StartValue="0.010"  Error="0.00002"
LowerLimit="1E-5" />

  <Parameter Name="epsT"  StartValue="0.95"  Error="0.0001"
LowerLimit="0.5" UpperLimit="1.0" Fixed="true" />
  <Parameter Name="varQU" StartValue="0.0"    Error="0.5"
LowerLimit="0.0" Fixed="true" />

</Fitter>

<EstimationMethod Type="KFLoglikelihoodKatrin"
  Name="myKatrinLogL"
  Function="Poisson"
  RunGenerator="myRungenStd"
  SpectrumSimulator="mySpecSim"
  EnableInterpolation="False">

  <Systematic ParameterIndex="4" Name="epsT" Mean="0.95" Error="0.01" />
  <Systematic ParameterIndex="5" Name="varQU" Mean="0.0" Error="0.01" />

</EstimationMethod>

<SpectrumSimulator
  Name="mySpecSim"
  Spectrum="myintspectmeasurement"
  MixingStrategy="BeforeIntegration"
  TritiumPurity="0.95"
  SigmaQU="0.0"
  IntegrationPrecision="1E-5"
  IntegrationMinSteps="16"
/>
```

```
<RunGenerator
  Name="myRungenStd"
  Spectrum="myintspectmeasurement"
  Smearing="{smear=true}"
  E0="18575"
  SquaredNeutrinoMasses="0.0"
  SquaredMixingParameters="1.0"
  MixingStrategy="BeforeIntegration"
  TritiumPurity="0.95"
  SigmaQU="0.0"
IntegrationPrecision="1E-5"
IntegrationMinSteps="16"
  >

  <RuntimeSchedule Type="KFRuntimeSchedule"
    InputFile="time-3y-jopti30eV10mHz.dat"
    TotalMeasurementTime="94608000" />

  <Background Type="KFBackgroundPoisson" Rate="0.01" />

</RunGenerator>
```

Deutsche Zusammenfassung

Das **K**arlsruhe **TR**itium Neutrino (KATRIN) Experiment ist ein bemerkenswertes Projekt. Über 150 Wissenschaftler, Ingenieure und Studenten arbeiten in fünf verschiedenen Ländern daran, die Masse des Elektron-Antineutrinos mit noch nie dagewesener Genauigkeit zu bestimmen. Darüber hinaus könnte eine Erweiterung des Experiments dazu führen, in noch unbekanntere Gebiete der Physik vorzustoßen. Im Jahre 2011 wurde in [dVeg⁺11] vorgeschlagen, mit der Hilfe von KATRIN im Zerfallsspektrum von Tritium nach Signaturen steriler Neutrinos zu suchen.

Die Aufgabe der vorliegenden Arbeit war es, das β -Spektrum des Tritiumzerfalls zu diesem Zweck genauer zu berechnen und die Datenanalyse-Methoden so zu erweitern, dass man sie zur Suche nach sterilen Neutrinos nutzen kann.

Diese Zusammenfassung orientiert sich an der Gliederung der Diplomarbeit. Zunächst werden der astrophysikalische Hintergrund und insbesondere die zugrundeliegende Neutrinophysik erläutert. Dann wird das KATRIN Experiment vorgestellt. Im Anschluss daran wird der Rahmen dieser Arbeit auf dunkle Materie und sterile Neutrinos ausgedehnt. Das sterile Neutrino könnte das Teilchen sein, das einen Großteil der dunklen Materie ausmacht. Um seine Existenz nachzuweisen, muss das β -Spektrum des Tritiumzerfalls sehr genau bekannt sein. Aus diesem Grund werden im darauffolgenden Abschnitt Korrekturen vorgestellt, die das bisher bekannte Spektrum vervollständigen. Der letzte Teil dieser Arbeit handelt von Bayes'schen Analysemethoden, die zum einen für die Standard-KATRIN m_ν^2 -Suche, aber auch für die Analyse von sterilen Neutrino Signaturen verwendet werden können.

Neutrinos wurden im Jahre 1930 von W. Pauli postuliert, um die scheinbare Verletzung der Energie- und Impulserhaltungssätze beim β -Zerfall zu erklären. Es sind schwer nachzuweisende Spin- $\frac{1}{2}$ Fermionen, die keinerlei Ladung tragen. Bisher wurden lediglich linkshändige Neutrinos und rechtshändige Antineutrinos beobachtet. [Zube12]

Neutrinos sind die am häufigsten vorkommenden bekannten Teilchen in unserem Universum und es gibt theoretisch viele Möglichkeiten, ihre Masse zu bestimmen. So können sowohl kosmologische Daten (beispielsweise von der kosmischen Mikrowellenhintergrundstrahlung) als auch astrophysikalische Beobachtungen (Supernova-Neutrinos, ultrahochenergetische kosmische Strahlen) für die Bestimmung herangezogen werden. Auch Experimente mit neutrinolosem doppel- β -Zerfall bieten sich an, sofern Neutrinos Majorana-Teilchen sind. Direkte Messungen mit Hilfe des einfachen β -Zerfalls sind besonders geeignet, da sie ohne zusätzliche Modell-Annahmen

die Neutrinomasse bestimmen können. Das Wissen über diese Masse ist wichtig für Teilchenphysik und Kosmologie und könnte unter anderem offene Fragen bezüglich der Strukturbildung und Evolution des Universums beantworten. [KATR04]

Das KATRIN Experiment ist eine solche modellunabhängige Messung. Bei der Beobachtung des Tritiumzerfalls kann durch die Messung der entstehenden β -Elektronen auf die Masse des Neutrinos zurückgeschlossen werden. KATRIN besteht aus folgenden Komponenten. In der WGTS („Windowless Gaseous Tritium Source“, die fensterlose Tritiumquelle) zerfällt das molekulare Tritium in einen Tochterkern, ein Elektron und ein Elektron-Antineutrino. Mit der Hilfe starker Magnetfelder werden die Elektronen durch die DPS („Differential Pumping Section“) und die CPS („Cryogenic Pumping Section“) zu den Spektrometern geleitet. Eventuell ausgetretenes Tritium wird auf dem Weg eingefangen und zur Quelle zurückgeführt. Die Spektrometer werden als MAC-E-Filter („Magnetic Adiabatic Collimation with an Electrostatic Filter“) betrieben und lassen nur Elektronen passieren, die eine gewisse Mindestenergie haben. Im Detektor wird dann ein integriertes Elektronen-Spektrum aufgenommen. Da die Neutrinos eine Masse haben und folglich Energie davontragen, lässt sich aus den gemessenen Raten die Neutrinomasse extrahieren. [KATR04]

Auch sterile Neutrinos mit einer Masse von bis zu 18 keV könnten Spuren im β -Spektrum hinterlassen. Zwar unterliegen diese hypothetischen Teilchen nicht der schwachen Wechselwirkung, jedoch mischen sie mit aktiven Neutrinos und könnten folglich beim Tritiumzerfall Energie davontragen. Sterile Neutrinos sind rechtshändig und manche Wissenschaftler vermuten, dass sie den Hauptbestandteil dunkler Materie ausmachen [dVeg⁺11].

Es gibt etwa fünfmal so viel dunkle wie baryonische Materie in unserem Universum. Eine der ersten Evidenzen für ihre Existenz wurde 1933 von F. Zwicky veröffentlicht. Er hatte beobachtet, dass die hohen Geschwindigkeiten einzelner Galaxien im Coma-Galaxienhaufen nur dann in Übereinstimmung mit dem Virialsatz sind, wenn eine wesentlich höhere Massendichte vorherrscht als die beobachtete. Mit der Existenz dunkler Materie ließe sich dieses Phänomen erklären, genau wie Beobachtungen mit dem Gravitationslinseneffekt und die Materietrennung am Bullet-Galaxienhaufen. Bisher wurde kein Teilchen gefunden, das diese mysteriöse Substanz erklären kann.

Mit dem Nachweis des sterilen Neutrinos könnte sich dies ändern. Um die Sensitivität des KATRIN Experiments in diesen Bereich zu erweitern, sind genaue Berechnungen zum β -Spektrum erforderlich. Die Grundstruktur des Spektrums erhält man mit Hilfe von Fermis Goldener Regel, die den Übergang von einem quantenmechanischen Zustand zu einem anderen beschreibt. Zusätzlich zu dieser Grundform müssen verschiedene Korrekturen, die einen Einfluss auf den Verlauf des Spektrums haben, berücksichtigt werden. Die Korrekturen, die für die Messung der Elektron-Antineutrinomasse verwendet werden, sind lediglich Näherungen, die ausschließlich in der Endpunktregion des Spektrums zulässig sind. Daher mussten weiterreichende Verbesserungen implementiert werden. Im Rahmen dieser Arbeit wurden *Fermi-Korrekturen*, *Strahlungskorrekturen* und eine *vollständige Endzustandsverteilung* in das Simulations-Framework implementiert. Der Einfluss der beiden größten Korrekturen ist in Abb. C.1 dargestellt.

Die Fermi-Korrekturen beschreiben die Wechselwirkung des davonfliegenden β -Elek-

trons mit dem elektromagnetischen Feld des Tochterkerns. Diese energieabhängige Korrektur lässt sich aus der Dirac-Gleichung herleiten und macht sich deutlich im Verlauf des Spektrums bemerkbar. [Wilk82]

Die Emission von virtuellen und reellen Photonen während des Zerfalls, hat ebenfalls einen Einfluss auf die Energie des β -Elektrons. Mit der Einführung von so genannten Strahlungskorrekturen kann man diesem Energieverlust Rechnung tragen. Auch der Einfluss dieser Korrekturen ist deutlich in Abb. C.1 zu sehen. [Glü93]

Einen wesentlich geringeren Einfluss auf die Energie des Elektrons hat die Berücksichtigung der vollständigen Endzustandsverteilung des Tochtermoleküls. Die bei dem Zerfall freiwerdende Energie teilt sich folgendermaßen auf. Ein Teil geht in die Masse und die kinetische Energie des Elektrons. Ein anderer sorgt für das Neutrino und dessen Energie. Der Rest bleibt beim Tochter-Molekül zurück, zum einen als Rückstoßenergie, zum anderen als Anregungsenergie. Da es sich bei dem Tochterkern HeT^+ um ein zweiatomiges Molekül handelt, sind neben der elektronischen Anregung auch Anregungen der Vibrations- und Rotationszustände möglich. Diese statistischen Anregungen werden mit Hilfe der Endzustandsfunktion beschrieben. Auch hier ist für das gegenwärtige KATRIN Experiment lediglich die Funktion für das Ende des Spektrums implementiert. Für diese Diplomarbeit wurde die Verteilung erweitert. [SaJF00]

Der nächste große Teil dieser Arbeit befasst sich mit der Datenanalyse des KATRIN Experiments. Um aus den gemessenen Elektronenraten die Neutrinomasse zu extrahieren, wird ein Spektrum an die Messwerte gefittet. Der Abstand des theoretischen Spektrums zu den Messpunkten wird solange minimiert, bis die optimale Konfiguration gefunden ist. Diese Minimierung findet standardmäßig mit MINUIT statt. Bei Fits mit vielen freien Parametern, wie es sich bei der Suche nach sterilen Neutrinos ergibt, stößt diese frequentistische Methode an ihre Grenzen. Daher wurden im Rahmen dieser Diplomarbeit Bayes'sche *Markov Chain Monte Carlo* (MCMC) Algorithmen implementiert. Neben der Möglichkeit zur Analyse von sterilen Neutrino Signaturen, stellen diese auch für das Standard KATRIN Experiment eine hervorragende Ergänzung dar. Mit ihnen lässt sich nicht nur die Posterior-Wahrscheinlichkeitsverteilung sampeln, die angibt mit welcher Wahrscheinlichkeit der Messwert dem wahren Wert der Neutrinomasse entspricht, sondern auch Konfidenzintervalle berechnen und Korrelationen sichtbar machen. Außerdem bieten sie die Möglichkeit als Vergleich zu den anderen Algorithmen herangezogen zu werden, um intern die Analyseergebnisse zu evaluieren und Fehler auszuschließen.

Für diese Methoden benötigt man die KATRIN Likelihood-Funktion,

$$L(\vec{x}|\theta) = \prod_{i=1}^N f(x_i, \theta), \quad (\text{C.1})$$

wobei x_i die Messwerte sind und θ den Parameter beschreibt, an dessen wahren Wert man interessiert ist. Die Wahrscheinlichkeitsdichtefunktion des Messwertes x_i wird durch $f(x_i, \theta)$ beschrieben.

Alle vorgestellten Methoden sampeln die Posterior-Wahrscheinlichkeitsverteilung, indem sie das Minimum der KATRIN Likelihood-Funktion suchen und dann den Bereich darum mehr oder weniger zufällig ablaufen. Je höher die Posterior-Wahr-

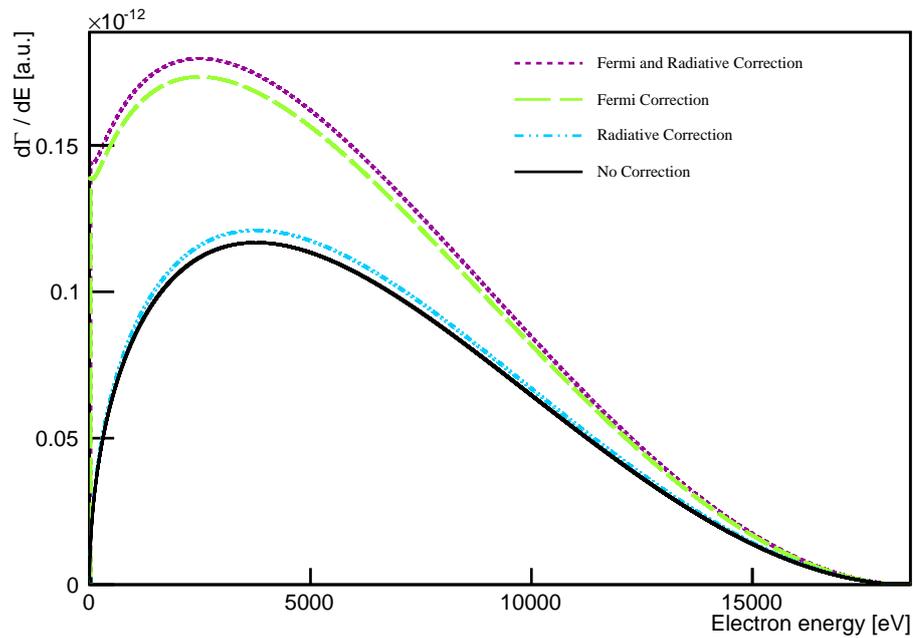


Figure C.1: Das β -Spektrum des Tritiumzerfalls ohne Korrekturen ist in schwarz dargestellt. Die Strahlungskorrekturen (blau) fñhren zu einem Faktor, der das Spektrum leicht anhebt. Eine noch stärkere Erhöhung, sowie eine Verschiebung des Maximums in Richtung niedriger Energiewerte, werden durch die Fermi-Korrekturen (grün) hervorgerufen. Die Kombination beider Korrekturen ist in lila dargestellt. Im Vergleich zu diesen Korrekturen sind die Auswirkungen der vollständigen Endzustands-Verteilung des Tochtermoleküls so gering, dass man bei diesem Maßstab keine Veränderung wahrnehmen kann.

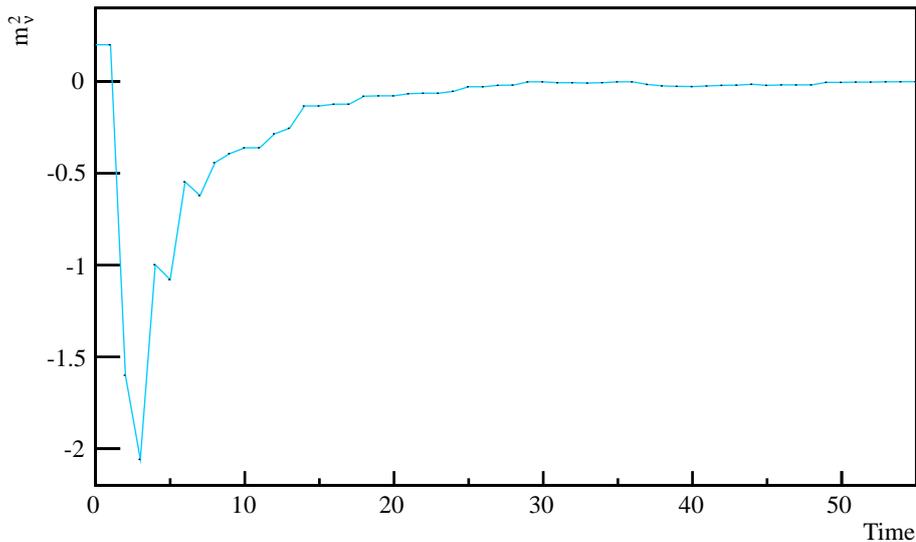


Figure C.2: Diese Abbildung zeigt die Konvergenz und das anschließende Sampling mit Hilfe der Hamilton Monte Carlo Methode. Dargestellt ist lediglich die Bewegung in der m_ν^2 -Dimension. In Wirklichkeit bewegt sich der Sampler mindestens im vierdimensionalen Raum. Nach weniger als 20 Schritten hat der Sampler das Minimum – in diesem Fall den Wert Null – erreicht und beginnt mit dem Sampeln. Durch das wiederholte Ablaufen der Gegend um das Minimum wird die Posterior-Wahrscheinlichkeitsverteilung nachgebildet.

scheinlichkeit an einem Punkt ist, desto häufiger wird der Sampler an ihm vorbeikommen. Wenn dieser Algorithmus mit Hilfe des Metropolis Ratios evaluiert wird, kann man zeigen, dass die resultierende Verteilung tatsächlich der gewünschten Posterior-Verteilung entspricht. [Greg10]

Die erste vorgestellte Methode ist *MCMC mit einem Gauß'schen Transition Kernel*. Hier ist die Vorschlagfunktion für die nächste Position des Samplers eine mehrdimensionale Gaußverteilung. Dieses vergleichsweise einfache Prinzip resultiert in einem robusten Algorithmus, der jedoch langsam konvergiert und dessen gesampelte Stichproben Autokorrelationen aufweisen. [Greg10]

Bei der *Hamilton Monte Carlo* Methode ist die Vorschlagfunktion ausgeklügelter. Die Likelihood-Funktion wird hier als Gravitationspotential aufgefasst und dem Sampler eine Masse zugewiesen. Anschaulich kann man sich nun vorstellen, dass der Sampler in Richtung des Minimums rutscht und auf dem Weg dorthin seine potentielle Energie in kinetische umwandelt. Der gewonnene Impuls trägt den Sampler über das Minimum hinaus und ermöglicht so eine genaue Abbildung der Minimumsumgebung. Als Input werden für diesen komplizierteren Algorithmus Gradienteninformationen benötigt, was die Rechenzeit signifikant erhöht. Belohnt wird man dafür mit einem schnellen Konvergenzverhalten, sehr niedrigen Autokorrelationen und einem sehr robusten Algorithmus, der auch für komplizierte Likelihood-Funktionen geeignet ist. Diese Eigenschaften machen die Hamilton Monte Carlo Methode zu einem idealen Werkzeug für die sterile Neutrino Analyse. [Neal12]

Für den Fall, dass eine zukünftige KATRIN-Likelihood noch komplizierter wird als bisher angenommen und insbesondere über starke Krümmungen verfügt, wurde noch eine weitere Methode implementiert. Diese heißt *Riemannian Manifold Hamiltonian Monte Carlo* und berücksichtigt die Raumkrümmung mit Hilfe einer komplizierten Metrik. Für die Berechnung benötigt dieser Algorithmus Ableitungen dritter Ordnung, was sich negativ auf die Rechenzeit auswirkt. Dafür liegen seine Stärken bei stark gekrümmten Verteilungen. Da für KATRIN in absehbarer Zeit keine solchen Verteilungen von Bedeutung sind, spielt diese Methode vorerst eine untergeordnete Rolle. [Beta12]

Exemplarisch für diese drei MCMC Algorithmen ist in Abbildung C.2 das Konvergenzverhalten von Hamilton Monte Carlo bei einer simulierten KATRIN Messung dargestellt.

Mit der Hilfe der in dieser Arbeit vorgestellten und implementierten Spektrumskorrekturen und Analysetools, ist das KATRIN Experiment bestens für die Zukunft gerüstet. Die Markov Chain Monte Carlo Methoden können entweder direkt für die Analyse der m_ν^2 -Daten verwendet werden, oder als Vergleich zu frequentistischen Methoden. Darüber hinaus ermöglichen sie, zusammen mit den Erweiterungen zum Tritium β -Spektrum, die Suche nach Signaturen steriler Neutrinos. Beide Entdeckungen hätten große Bedeutung für Teilchenphysik und Kosmologie und würden einen Blick darauf ermöglichen, „was die Welt im Innersten zusammenhält“ [Goet08].

Acknowledgements

When I started writing this diploma thesis, I imagined writing the acknowledgments must be a great feeling, because it implies that all the hard work is finished. Now that I have written about a hundred pages about the research that I had the honor to be involved in, I feel kind of sad. Sad because it also means that a time is ending that was interesting, informative and above all beautiful – especially because of the following people.

I would love to thank:

- Prof. Dr. G. Drexlin, who not only enthralled me about the KATRIN experiment, but also offered me, like all the other bachelor, master and PhD students, fantastic working conditions in an exciting field.
- Prof. Dr. T. Müller for being the second reviewer of this thesis.
- Prof. J. Formaggio, who looked very well after me during my time in the United States, fascinated me with his knowledge about analytical methods and made me feel extremely welcome at MIT.
- Dr. M. Steidl, who was a great supervisor and gave me the feeling that I could always contact him with all questions about KATRIN.
- Dr. F. Glück for supporting me especially in regard to the first part - the spectrum calculations - of my thesis. His knowledge in this area is overwhelming and his involvement enhanced this thesis a lot.
- Dr. Alan Poon for giving me a first insight into scientific work and for being a fantastic host during my time at LBNL.
- Dr. M. Hötzel, who deserves my greatest thanks. From the very beginning he challenged and encouraged me at the same time. He was available for many helpful discussions, which strongly quickened my work.
- M. Haag, whose fantastic work for the KATRIN analysis made this thesis possible in the first place. Like Markus Hötzel he supported me with advice and action since my first hours at KATRIN.
- S. Pawlowski, who was very dedicated to helping me with the formalities I had to deal with for my study and research visits in the United States.

- The DAAD for sponsoring my visits in Berkeley and Cambridge with two PROMOS scholarships.
- All those who helped me by proofreading this thesis: Prof. Dr. G. Drexlin, Dr. M. Steidl, Dr. F. Glück, Dr. M. Hötzel, Dr. E. Gordon, M. Haag, M. Moeferdts and F. Nickel.

Thank you very much to all of you.

Finally, I would like to highlight the great teamwork of all the students, postdocs and professors involved in the KATRIN experiment. Not only the great professional collaboration, but also the friendly daily interactions and the “KATRIN Stammtisch” will, for sure, be remembered fondly and missed greatly.

Bibliography

- [AcGL07] M. A. Acero, C. Giunti und M. Laveder. Limits on electron neutrino and anti electron neutrino disappearance from Gallium and reactor experiments. *Phys.Rev.D*, Vol. 78, 2007.
- [Arti64] E. Artin. *The gamma function*. Holt, Rinehart and Winston, New York. 1964.
- [Babu⁺11] M. Babutzka et al. The comprehensive guide to KASSIOPEIA. *Internal document*, 2011.
- [Babu⁺12] M. Babutzka et al. Monitoring of the operating parameters of the KATRIN Windowless Gaseous Tritium Source. *New Journal of Physics*, 14(10), 2012, P. 103046.
- [BaMT04] V. Barger, D. Marfatia und A. Tregre. Neutrino mass limits from SDSS, 2dFGRS and WMAP. *Physics Letters B*, 595(1–4), 2004, P. 55 – 59.
- [Barl89] R.J. Barlow. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Manchester Physics Series. Wiley. 1989.
- [Barn⁺96] R. M. Barnett et al. Review of Particle Physics. *Phys. Rev. D*, Vol. 54, Jul 1996, P. 1–708.
- [Bart12] H. Barthel. Cusped and Cored Profiles (URL: <http://www.schattenblick.de/infopool/natur/astronom/nafor366/s9-abb5.jpg>), 2012.
- [BBC13] BBC. Offbeat photos of the year 2006 (URL: <http://news.bbc.co.uk>), 2013.
- [Beac99] J. F. Beacom. Supernova Neutrinos and the Neutrino Masses. *arXiv:hep-ph/9901300*, 1999.
- [BeHL09] F. Bezrukov, H. Hettmansperger und M. Lindner. keV sterile neutrino dark matter in gauge extensions of the standard model. *Phys.Rev.D*, Vol. 81, 2009.
- [BePT80] G. Beamson, H. Q. Porter und D. W. Turner. The collimating and magnifying properties of a superconducting field photoelectron spectrometer. *Journal of Physics E: Scientific Instruments*, 13(1), 1980, P. 64.

- [Besk⁺10] A. Beskos et al. Optimal tuning of the Hybrid Monte-Carlo Algorithm. *arXiv:1001.4460*, 2010.
- [Beta12] M. Betancourt. A General Metric for Riemannian Manifold Hamiltonian Monte Carlo. *arXiv:1212.4693v1*, 2012.
- [Bezr11] F. Bezrukov. Light sterile neutrino dark matter in extensions of the Standard Model. *arXiv:1109.3187*, 2011, P. 7–8.
- [Blec10] J. Bleck-Neuhaus. *Elementare Teilchen: moderne Physik von den Atomen bis zum Standard-Modell*. Springer-Lehrbuch. Springer, Berlin. 2010.
- [BoUr01] D. Borthwick und A. Uribe. The spectral density function for the Laplacian on high tensor powers of a line bundle. *arXiv:math/0103062*, 2001.
- [Cald01] David O. Caldwell (Hrsg.). *Current aspects of neutrino physics*. Physics and astronomy online library. Springer, Berlin. 2001.
- [CERN13] CERN-Collaboration. Root documentation on the CERN website (URL: <http://root.cern.ch/drupal/content/documentation>), 2013.
- [ChQL01] L. Chen, Z. Qin und J. S. Liu. Exploring Hybrid Monte Carlo in Bayesian Computation. *Proceedings of ISBA 2000*, Vol. 2, 2001, P. 2–5.
- [Clow⁺06] D. Clowe et al. A direct empirical proof of the existence of dark matter. *Astrophysical Journal Letters*, 2006, P. 109–113.
- [CMS-12] CMS-Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, Vol. 716, 2012.
- [CoGr07] W. N. Cottingham und D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press, Cambridge. 2007.
- [CoSa99] E. Corbelli und P. Salucci. The Extended Rotation Curve and the Dark Matter Halo of M33. *arXiv:astro-ph/9909252*, 1999.
- [Danb⁺62] G. Danby et al. Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos. *Phys. Rev. Lett.*, Vol. 9, Jul 1962, P. 36–44.
- [Daya12] DayaBay-Collaboration. Observation of Electron-Antineutrino Disappearance at Daya Bay. *Phys. Rev. Lett.*, Vol. 108, Apr 2012, P. 171803.
- [Dode03] S. Dodelson. *Modern cosmology*. Academic Press, Amsterdam. 2003.
- [Domb06] C. Dombacher. Die komplexwertige Gammafunktion (<http://www.telecomm.at/documents/gamma.pdf>), 2006.
- [DoMS06] S. Dodelson, A. Melchiorri und A. Slosar. Is Cosmology Compatible with Sterile Neutrinos? *Phys. Rev. Lett.*, Vol. 97, Jul 2006, P. 041301.

- [Doss07] N. Doss. *Calculated final state probability distributions for T2 beta-decay measurements*. Dissertation, University College London, 2007.
- [Drex11] G. Drexlin. *Astroteilchenphysik I*. University Lecture, Karlsruhe Institute of Technology, 2011.
- [dVeg⁺11] H. J. de Vega et al. Search of Sterile Neutrino Warm Dark Matter in the Rhenium and Tritium Beta Decays. *Nucl. Phys. B*, 2011, P. 177–195.
- [dVSS10] H. J. de Vega, P. Salucci und N. G. Sanchez. The mass of the dark matter particle from theory and observations. *New Astronomy*, 2010, P. 653 –666.
- [Eins05] A. Einstein. Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig? *Annalen der Physik*, Vol. 323, 1905, P. 639–641.
- [EsPe12] A. Esmaili und O. L. Peres. KATRIN Sensitivity to Sterile Neutrino Mass in the Shadow of Lightest Neutrino Mass. *Phys.Rev.D*, Vol. 85, 2012.
- [EXO-12] EXO-Collaboration. Search for Neutrinoless Double-Beta Decay in ¹³⁶Xe with EXO-200. *Phys. Rev. Lett.*, Vol. 109, Jul 2012, P. 032505.
- [FeSV03] Ferruccio Feruglio, Alessandro Strumia und Francesco Vissani. Addendum to: Neutrino oscillations and signals in beta and neutrinoless double beta experiments. *Nuclear Physics B*, 659(1–2), 2003, P. 359 – 362.
- [Fisc⁺11] S. Fischer et al. Monitoring of Tritium Purity During Long-Term Circulation in the KATRIN Test Experiment LOOPINO Using Laser Raman Spectroscopy. *Fusion Science and Technology*, 60(3), 2011, P. 925–930.
- [FoBa11] J. A. Formaggio und J. Barrett. Resolving the Reactor Neutrino Anomaly with the KATRIN Neutrino Experiment. *Physics Letters B*, 2011, P. 68–71.
- [FoKR02] Z. Fodor, S. D. Katz und A. Ringwald. Determination of Absolute Neutrino Masses from Bursts of Z Bosons in Cosmic Rays. *Phys. Rev. Lett.*, Vol. 88, Apr 2002, P. 171101.
- [GeGR97] A. Gelman, W. R. Gilks und G. O. Roberts. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1), 1997, P. 110–120.
- [GiCa11] M. Girolami und B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society B*, 73(2), 2011, P. 123–214.
- [Gil⁺10] W. Gil et al. The Cryogenic Pumping Section of the KATRIN Experiment. *Applied Superconductivity*, Vol. 20, 2010.
- [GiLa10] C. Giunti und M. Laveder. Statistical Significance of the Gallium Anomaly. *Phys.Rev.C*, Vol. 83, 2010.

- [Glü93] F. Glück. Measureable distributions of unpolarized neutron decay. *Physical review D*, Vol. 47, 1993.
- [Glü13] F. Glück. Personal communication, 2013.
- [Goet08] J. W. von Goethe. *Faust. Der Tragödie erster Teil*. 1808.
- [GrBu12] C. Grupen und I. Buvat. *Handbook of Particle Detection and Imaging*. Springer, Berlin, Heidelberg. 2012.
- [Greg10] P. C. Gregory. *Bayesian logical data analysis for the physical sciences: a comparative approach with Mathematica support*. Cambridge University Press, Cambridge. Originally published: 2005, 2010.
- [Groh⁺13] S. Grohmann et al. The thermal behaviour of the tritium source in KATRIN. *Cryogenics*, 55–56(0), 2013, P. 5 – 11.
- [Grou07] Particle Data Group. Number of Neutrino Type. *J. Phys. G*, Vol. 33, 2007.
- [Haag13] M. Haag. Talk for the 24. KATRIN Collaboration Meeting, 2013.
- [Haag14] M. Haag. *In preparation*. Dissertation, Karlsruhe Institut of Technologie (KIT), 2014.
- [Hans00] K. M. Hanson. Tutorial on Markov Chain Monte Carlo (URL: <http://public.lanl.gov/kmh/talks/>), 2000.
- [Hoff⁺07] Y. Hoffman et al. Evolution of the Phase-Space Density in Dark Matter Halos. *The Astrophysical Journal*, 671(2), 2007, P. 1108.
- [Höt09] M. Hötzel. Berechnung von KATRIN Messspektren unter Einbeziehung der Eigenschaften der fensterlosen gasförmigen Tritiumquelle. Diplomarbeit, Universität Karlsruhe, Karlsruhe, 2009.
- [Höt12] M. Hötzel. *Simulation and analysis of source-related effects for KATRIN*. Dissertation, Karlsruhe Institute of Technology, 2012.
- [Jame08] F. James. *Statistical methods in experimental physics*. World Scientific, New Jersey. 2. ed., repr.. Auflage, 2008.
- [Jayn12] E. T. Jaynes. *Probability theory: the logic of science*. Cambridge Univ. Press, Cambridge. 10. print.. Auflage, 2012.
- [K.As⁺96] K. Assamagan et al. Upper limit of the muon-neutrino mass and charged-pion mass from momentum analysis of a surface muon beam. *Phys. Rev. D*, Vol. 53, Jun 1996, P. 6065–6077.
- [KATR04] KATRIN-Collaboration. KATRIN Design Report, 2004.
- [KATR06] KATRIN-Collaboration. Wie viel wiegt ein Neutrino? Das KATRIN-Experiment am KIT (URL: <http://www.youtube.com/watch?v=cnu79iC0C1M>), 2006.

- [KATR13] KATRIN-Collaboration. KATRIN website (URL: <http://www.katrin.kit.edu/>), 2013.
- [KKVi02] H. V. Klapdor-Kleingrothaus und R. D. Viollier. Dark matter in astro- and particle physics: proceedings of the international conference, DARK 2002, Cape Town, South Africa, 4-9 February 2002. Physics and astronomy online library, Berlin, 2002. Springer.
- [Klyp⁺99] A. Klypin et al. Where Are the Missing Galactic Satellites? *The Astrophysical Journal*, 522(1), 1999, P. 82.
- [Koll12] D. Koller. Lecture about the Metropolis Hastings Algorithm (URL: <http://www.youtube.com/watch?v=P20TdCy-siE>), 2012.
- [KoTu94] E. W. Kolb und M. S. Turner. *The early universe*. Frontiers in physics; 69. Westview Pr., Boulder, Col. 1994.
- [Lan⁺12] S. Lan et al. Lagrangian Dynamical Monte Carlo. *arXiv:1211.3759*, 2012.
- [Luki⁺12] S. Lukic et al. Measurement of the gas-flow reduction factor of the KATRIN DPS2-F differential pumping section. *Vacuum*, 86(8), 2012, P. 1126 – 1133.
- [Mach05] H. Machner. *Einführung in die Kern- und Elementarteilchenphysik*. Lehrbuch Physik. WILEY-VCH, Weinheim. 2005.
- [Ment⁺11] G. Mention et al. The Reactor Antineutrino Anomaly. *Phys.Rev.D*, Vol. 83, 2011.
- [Mert13] S. Mertens. Paper about KATRIN and sterile neutrinos. *In preparation*, 2013.
- [Mill07] E. Million. The Hadamard Product (URL: <http://buzzard.ups.edu/courses/2007spring/projects/million-paper.pdf>). 2007.
- [MoPa04] R. N. Mohapatra und P. B. Pal. *Massive neutrinos in physics and astrophysics*. World Scientific, Singapur. 3.. Auflage, 2004.
- [Muel⁺11] T. A. Mueller et al. Improved Predictions of Reactor Antineutrino Spectra. *Phys.Rev.C*, Vol. 83, 2011.
- [Muta10] T. Muta. *Foundations of quantum chromodynamics: an introduction to perturbative methods in gauge theories*. World scientific lecture notes in physics ; 78. World Scientific, Singapore. 3. ed.. Auflage, 2010.
- [NASA04] NASA. Kepler's Supernova Photomontage, 2004.
- [Nasa11] Nasa. What is the Universe Made Of? (URL: <http://map.gsfc.nasa.gov/universe/universe.html>), 2011.

- [NAWe06] NASA/CXC und M. Weiss. Bullet Cluster Photomontage (URL: <http://chandra.harvard.edu/photo/2006/1e0657/more.html>), 2006.
- [Neal12] R. M. Neal. MCMC using Hamiltonian dynamics. *Published as Chapter 5 of the Handbook of Markov Chain Monte Carlo*, *arXiv:1206.1901*, 2012.
- [O.El⁺02] O. Elgar et al. New Upper Limit on the Total Neutrino Mass from the 2 Degree Field Galaxy Redshift Survey. *Phys. Rev. Lett.*, Vol. 89, Jul 2002, P. 061301.
- [OPER11] OPERA-Collaboration. Measurement of the neutrino velocity with the OPERA detector in the CNGS beam. *arXiv:1109.4897*, 2011.
- [oSta54] United States National Bureau of Standards. Computation Laboratory. *Table of the gamma function for complex arguments*. U.S. Govt. Print. Off., Washington. 1954.
- [Perl⁺75] M. L. Perl et al. Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation. *Phys. Rev. Lett.*, Vol. 35, Dec 1975, P. 1489–1492.
- [Plan13] Planck-Collaboration. Planck 2013 results. I. Overview of products and scientific results. *arXiv:1303.5062*, 2013.
- [Pres07] W. H. Press (Hrsg.). *Numerical recipes: the art of scientific computing*. Cambridge University Press, Cambridge. 3. ed.. Auflage, 2007.
- [Quas11] G. Quast. Datenanalyse und Rechnernutzung (URL: <http://www-ekp.physik.uni-karlsruhe.de/quast>). University Lecture, Karlsruhe Institute of Technology, 2011.
- [ReWu83] W. W. Repko und C. Wu. Radiative corrections to the end point of the tritium beta-decay spectrum. *Phys. Rev. C*, Vol. 28, Dec 1983, P. 2433–2436.
- [RoCa10] C. P. Robert und G. Casella. *Introducing Monte Carlo Methods with R*. Springer New York, New York, NY. 2010.
- [RuFo70] V.C. Rubin und W.K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophysical Journal*, Vol. 159, Februar 1970.
- [RuIv12] O. Ruchayskiy und A. Ivashko. Experimental bounds on sterile neutrino mixing angles. *Journal of High Energy Physics*, 2012(6), 2012, P. 1–27.
- [SaFr97] A. Saenz und P. Froelich. Effect of final-state interactions in allowed beta-decays. II. Reliability of the beta-decay spectrum for Tritium. *Phys. Rev. C*, Vol. 56, Oct 1997, P. 2162–2184.
- [SaJF00] A. Saenz, S. Jonsell und P. Froelich. Improved Molecular Final-State Distribution of HeT⁺ for the Beta-Decay Process of T2. *Phys. Rev. Lett.*, Vol. 84, Jan 2000, P. 242–245.

- [Scha12] S. Schams. Feldman Cousins Unified Approach for KATRIN. Talk at the 22. KATRIN Collaboration Meeting 2012, 2012.
- [Scha13a] S. Schams. DPG talk: Monte Carlo methods for KATRIN data analysis. Talk at the 77th Annual Meeting of the DPG in Dresden, 2013.
- [Scha13b] S. Schams. Monte Carlo methods for KATRIN data analysis. Talk at the 24. KATRIN Collaboration Meeting, 2013.
- [Schl⁺11] M. Schlösser et al. Design Implications for Laser Raman Measurement Systems for Tritium Sample-Analysis, Accountancy or Process-Control Applications. *Fusion Science and Technology*, 60(3), 2011, P. 976–981.
- [Schr⁺85] K. Schreckenbach et al. Determination of the antineutrino spectrum from U235 thermal neutron fission products up to 9.5 MeV. *Physics letters*, Vol. 160B, 1985.
- [Schw12] T. Schwetz. Status of sterile neutrino oscillations (URL: <http://www.mpi-hd.mpg.de/personalhomes/schwetz/kyoto-neutrino2012.pdf>). Heidelberg, 2012.
- [SDSS09] SDSS-Collaboration. SEGUE: A Spectroscopic Survey of 240,000 Stars with $g = 14$ -20. *The Astronomical Journal*, 137(5), 2009, P. 4377.
- [Simp81] J. J. Simpson. Measurement of the beta-energy spectrum of tritium to determine the antineutrino mass. *Phys. Rev. D*, Vol. 23, Feb 1981, P. 649–662.
- [Simp85] J. J. Simpson. Evidence of Heavy-Neutrino Emission in Beta Decay. *Phys. Rev. Lett.*, Vol. 54, Apr 1985, P. 1891–1893.
- [Sirl67] A. Sirlin. General Properties of the Electromagnetic Corrections to the Beta Decay of a Physical Nucleon. *Phys. Rev.*, Vol. 164, Dec 1967, P. 1767–1775.
- [SNO-01] SNO-Collaboration. Measurement of the Rate of Interactions Produced by Solar Neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, Vol. 87, Jul 2001, P. 071301.
- [Spot03] Spergel und D.N. others. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *Astrophys.J.Suppl.*, Vol. 148, 2003, P. 175–194.
- [Step10] S. Stepanow. *Relativistische Quantentheorie Für Bachelor: Mit Einführung in die Quantentheorie der Vielteilchensysteme*. Springer-Lehrbuch. Springer Berlin Heidelberg, Berlin, Heidelberg. 2010.
- [Sutt92] C. Sutton. *Spaceship neutrino*. Cambridge University Press, Cambridge. 1992.
- [Tayl⁺98] A. N. Taylor et al. Gravitational Lens Magnification and the Mass of Abell 1689. *The Astrophysical Journal*, 1998.

-
- [Tran94] J. Tran Thanh Van. *Particle Astrophysics, Atomic Physics and Gravitation - Proceedings of the XXIXth Rencontre de Moriond, Series: Moriond Workshops : Villars Sur Ollon, Switzerland, January 22-29, 1994*. Atlantica Séguier Frontières, Paris. 1994.
- [Voge10] P. Vogel. NNPSS-TSI-Lecture 1. University Lecture, California Institute of Technology, 2010.
- [webs12] Stony Brook University website. The Standard Model Illustration, 2012.
- [Wiki13] Wikipedia. Markov Chain, 2013.
- [Wilk82] D.H. Wilkinson. Analysis of neutron beta-decay. *Nuclear Physics A*, 377(2–3), 1982, P. 474 – 504.
- [Zube12] K. Zuber. *Neutrino Physics*. CRC Press, Boca Raton, Fla. Revised.. Auflage, 2012.
- [Zwic33] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, Vol. 6, 1933, P. 110–127.