Eur. Phys. J. C (2018) 78:212
https://doi.org/10.1140/epjc/s10052-018-5656-9

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

keV-Scale sterile neutrino sensitivity estimation with time-of-flight spectroscopy in KATRIN using self-consistent approximate Monte Carlo

Nicholas M. N. Steinbrink[1,a], **Jan D. Behrens**[1,2], **Susanne Mertens**[3], **Philipp C.-O. Ranitzsch**[1], **Christian Weinheimer**[1]

[1] Institut für Kernphysik, WWU Münster, Wilhelm Klemm-Str. 9, 48149 Münster, Germany
[2] Institute of Experimental Particle Physics (ETP), Karlsruhe Institute of Technology (KIT), Wolfgang Gaede-Str. 1, 76131 Karlsruhe, Germany
[3] Physics Department, TU München, James-Franck-Str. 1, 85748 Garching, Germany

Received: 25 October 2017 / Accepted: 19 February 2018 / Published online: 13 March 2018

Abstract We investigate the sensitivity of the Karlsruhe Tritium Neutrino Experiment (KATRIN) to keV-scale sterile neutrinos, which are promising dark matter candidates. Since the active-sterile mixing would lead to a second component in the tritium β-spectrum with a weak relative intensity of order $\sin^2\theta \lesssim 10^{-6}$, additional experimental strategies are required to extract this small signature and to eliminate systematics. A possible strategy is to run the experiment in an alternative time-of-flight (TOF) mode, yielding differential TOF spectra in contrast to the integrating standard mode. In order to estimate the sensitivity from a reduced sample size, a new analysis method, called self-consistent approximate Monte Carlo (SCAMC), has been developed. The simulations show that an ideal TOF mode would be able to achieve a statistical sensitivity of $\sin^2\theta \sim 5\times10^{-9}$ at one σ, improving the standard mode by approximately a factor two. This relative benefit grows significantly if additional exemplary systematics are considered. A possible implementation of the TOF mode with existing hardware, called gated filtering, is investigated, which, however, comes at the price of a reduced average signal rate.

1 Introduction

In recent years the interest has grown for sterile neutrinos with a mass scale of a few keV [1]. They are proposed as dark matter particle candidates in cold dark matter (CDM) and especially warm dark matter (WDM) scenarios [2–5]. WDM has the potential to avoid issues regarding structure formation on small scales which are not yet solved for WIMP (weakly interacting massive particle) CDM [6–12]. However, the shortcomings of WIMP CDM can possibly be mitigated via Baryonic feedback [13] while any sterile neutrino dark matter production mechanism needs to be fine-tuned to yield the correct DM density. Mass-dependent bounds on the sterile neutrino mixing with active neutrinos have been established by searches for sterile neutrino decay via X-ray satellites [14,15] and on basis of theoretical considerations in order to avoid dark matter overproduction [16], which never exceed $\sin^2\theta \lesssim 10^{-7}$. The mass range has been constrained by the DM phase-space distribution in dwarf spheroidal galaxies [17] and gamma-ray line emission from the Galactic center region [18] to 1 keV $< m_h <$ 50 keV. In order to produce the existing amount of dark matter, mass and mixing angle are linked by the production mechanism, which can be non-resonant [16,19,20] or resonant [21–24]. Moreover, possible evidence of relic sterile neutrinos with mass $m_h = 7$ keV has been reported in XMM-Newton data [25–27].

In principle, it can also be searched for keV-scale sterile neutrinos in ground-based experiments, such as in tritium β-decay [28,29]. A promising example is the Karlsruhe Tritium Neutrino Experiment (KATRIN) [30], which is the most sensitive neutrino mass experiment currently under construction. Sterile neutrinos would be visible by a discontinuity in the β-decay spectrum if they have a sufficiently large mixing angle with electron neutrinos. In order to adapt KATRIN, which is optimized for light neutrinos of $m_l \lesssim \mathcal{O}((\text{eV}))$, for keV sterile neutrinos, different approaches are discussed with the goal of enhancing statistics and managing systematics. A suitable idea is to develop a dedicated detector measuring in differential mode [31–33]. As an alternative idea, it is worthwhile to study the performance of an alternative time-of-flight (TOF) mode, which has already shown to be promising in theory for active neutrino mass measurements [34].

[a] e-mail: n.steinbrink@uni-muenster.de

In this publication the sensitivity of a keV-scale sterile neutrino search based on TOF spectroscopy with the KATRIN experiment is discussed both for an ideal measurement method as for a possible implementation with minimal hardware modifications.

2 Sterile neutrino search with TOF spectroscopy

2.1 Sterile neutrinos in tritium β-decay and KATRIN

There has been some previous work on sterile neutrinos in general in tritium β-decay. Most publications focus on eV-scale sterile neutrinos [35–39], which are proposed to address certain anomalies in oscillation experiments [40–45]. However, in recent time also dedicated studies, dealing with keV-scale neutrinos have been published, such as [29,31,32], as well as studies involving more exotic models, such as [46–48]. We will quickly summarize the main effect of a keV-scale sterile neutrino on the tritium β-spectrum, while we refer especially to [31] for deeper insights into systematics and theoretical corrections.

The tritium β-decay spectrum with a single neutrino with mass eigenstate m_i is given as

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}E} = N\frac{G_\mathrm{F}^2}{2\pi^3\hbar^7c^5}\cos^2(\theta_\mathrm{C})|M|^2\,F(E,Z')\cdot p\cdot(E+m_\mathrm{e}c^2) \cdot\sum_j P_j\cdot(E_0-V_j-E) \cdot\sqrt{(E_0-V_j-E)^2-m_i^2c^4}, \quad (1)$$

[28,49,50], where E is the kinetic electron energy, θ_C the Cabbibo angle, N the number of tritium atoms, G_F the Fermi constant, M the nuclear matrix element, $F(E, Z')$ the Fermi function with the charge of the daughter ion Z', p the electron momentum, P_j the probability to decay to an excited electronic and rotational–vibrational state with excitation energy V_j [51–53] and E_0 the beta endpoint, i.e. the maximum kinetic energy in case of $m_i = 0$.

The electron neutrino is a superposition of multiple mass eigenstates. Since the flavor eigenstate is the one which defines the interaction, but the mass eigenstate the one which describes the dynamics of the decay, the β-spectrum for the electron neutrino is an incoherent superposition of the contributions for each mass eigenstate,

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}E}(m_{\nu_\mathrm{e}}) = \sum_{i=1}^{3}|U_{ei}|^2\frac{\mathrm{d}\Gamma}{\mathrm{d}E}(m_i). \quad (2)$$

In case of an additional keV-scale sterile neutrino, a fourth mass state m_4 is introduced with a significantly lower mixing with the electron neutrino, $|U_{e4}|^2 \ll |U_{ei}|^2$ $(i \in 1, 2, 3)$. In the following we define the *heavy* or sterile neutrino mass as $m_h \equiv m_4$ and the *active-sterile mixing angle* as $\sin^2\theta \equiv |U_{e4}|^2 < 10^{-7}$ [15]. Since the light mass eigenstates 1, 2, 3 are not distinguishable by KATRIN [50], a *light neutrino mass* is defined as $m_l^2 \equiv \sum_{i=1}^{3}|U_{ei}|^2 m_i^2$. The combined β-spectrum with sterile and active neutrino can then be expressed as

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}E}(m_{\nu_\mathrm{e}}) = \sin^2\theta\frac{\mathrm{d}\Gamma}{\mathrm{d}E}(m_h) + \cos^2\theta\frac{\mathrm{d}\Gamma}{\mathrm{d}E}(m_l). \quad (3)$$

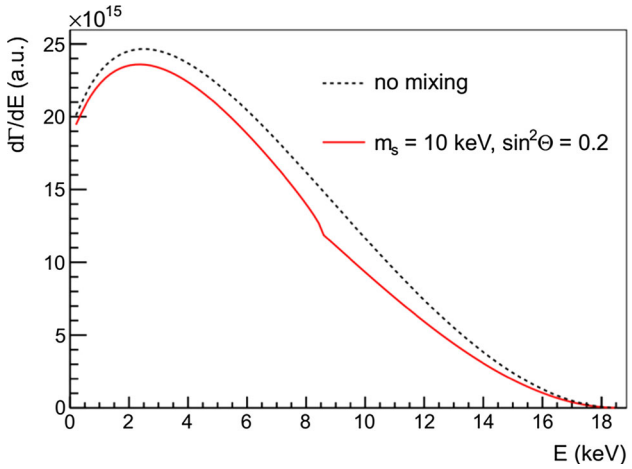

Fig. 1 Tritium β-decay spectrum without sterile neutrino contribution (dashed) and with exemplary case of exaggerated mixing with $\sin^2\theta = 0.2$ and $m_h = 10$ keV (red solid). Figure reproduced from Ref. [31]

An example with exaggerated mixing is shown in Fig. 1. In probing the absolute neutrino mass scale, the KATRIN experiment is designed to measure the light neutrino mass m_l with a sensitivity of < 0.2 eV at 90% confidence level (CL) [30]. Therefor it uses a windowless gaseous molecular tritium source (WGTS) [54] with an activity of $\sim 10^{11}$ Bq. The electrons from the β-decay are filtered in the main spectrometer based on the *magnetic adiabatic collimation with electrostatic filter (MAC-E-Filter)* principle [55]. The magnetic field in the center of the main spectrometer, the *analyzing plane*, is held small at $B_A = 3$ mT and otherwise high at $B_S = 3.6$ T in the source and at $B_\mathrm{max} = 6$ T at the exit of the main spectrometer just before the counting detector. Due to adiabatic conservation of the relativistic magnetic moment, electron momenta are aligned with the field in the analyzing plane. By additionally applying an electrostatic retarding potential qU in the analyzing plane, the MAC-E-Filter acts as a high-pass filter with a sharp energy resolution of $\Delta E/E = B_A/B_\mathrm{max} \approx 0.9\ \mathrm{eV}/E_0$. In the focal plane detector (FPD) the count rate is then measured. That way, KATRIN measures the *integral β-spectrum* as a function of qU.

2.2 Time-of-flight spectroscopy

The idea of using time-of-flight (TOF) spectroscopy for a measurement of the light neutrino mass is explained in detail in Ref. [34]. In the following, we will recapitulate the approach briefly and explain the motivations for investigating this technique for a keV-scale sterile neutrino search as well.

In contrast to the standard mode of operation, as described in the last section, TOF spectroscopy allows to measure not only the count-rate, but a full TOF spectrum at a given retarding potential qU. The TOF as a function of the energy is given by integrating the reciprocal velocity over the center of motion, which we will assume for simplicity to be on the z-axis,

$$\mathscr{T}(E,\vartheta) = \int \mathrm{d}z\, \frac{1}{v_\parallel} = \int_{z_\mathrm{start}}^{z_\mathrm{stop}} \mathrm{d}z\, \frac{E + m_e c^2 - q\Delta U(z)}{p_\parallel(z)\cdot c^2}, \quad (4)$$

where E and ϑ are the initial kinetic energy and polar angle of the electron, respectively. z_start and z_stop are the positions on the beam axis between which TOF is measured, $\Delta U(z)$ is the potential difference as a function of position z and $p_\parallel(z)$ the parallel momentum. By assuming adiabatic conservation of the magnetic moment, $p_\parallel(z)$ can be expressed analytically as a function of the potential $\Delta U(z)$ and magnetic field $B(z)$ (derivation see Ref. [34]). If these are known, the integral in Eq. (4) can be solved numerically.

Since the TOF is a function of the energy, the β-spectrum can be transformed into a TOF spectrum $\mathrm{d}N/\mathrm{d}\tau$, given the initial angular distribution of the β-decay electrons. A feature in the β-spectrum such as a sterile neutrino contribution would then also have a corresponding effect on the TOF spectrum if the retarding energy qU is sufficiently low. Like the β-spectrum (2), the TOF spectrum can as well be expressed as a superposition of a component with a heavy neutrino mass m_h and a light neutrino mass m_l:

$$\frac{\mathrm{d}N}{\mathrm{d}\tau}(m_{\nu_\mathrm{e}}) = \sin^2\theta \frac{\mathrm{d}N}{\mathrm{d}\tau}(m_h) + \cos^2\theta \frac{\mathrm{d}N}{\mathrm{d}\tau}(m_l). \quad (5)$$

For each of these two components, the TOF spectrum can then formally be obtained from the β-spectrum with neutrino mass m_l and m_h, respectively, using the transformation theorem for densities [56]:

$$\frac{\mathrm{d}N}{\mathrm{d}\tau} = \int_0^{\vartheta_\mathrm{max}} \int_{qU}^{E_0} \mathrm{d}\vartheta\, \mathrm{d}E\, g(\vartheta) \frac{\mathrm{d}N}{\mathrm{d}E}(E,\vartheta)\, \delta\left(\tau - \mathscr{T}(E,\vartheta)\right), \quad (6)$$

where $g(\vartheta)$ denotes the angular distribution and $\mathrm{d}N/\mathrm{d}E(E,\vartheta)$ the response corrected energy spectrum, which itself is a function of the β-spectrum (1) for a given neutrino mass. If angular changes from inelastic scattering processes in the tritium source are neglected, the angular distribution is approximately independent from the energy spectrum and given by isotropic emission

$$g(\vartheta) = \frac{1}{2}\sin(\vartheta) \quad (7)$$

within the angular acceptance interval given by the default KATRIN field settings with $\vartheta_\mathrm{max} = \sqrt{B_S/B_\mathrm{max}} = 50.77°$. The response corrected energy spectrum $\mathrm{d}N/\mathrm{d}E(E,\vartheta)$ in Eq. (6) is given in good approximation by the β-spectrum (3), convolved with the inelastic energy loss function in the tritium source,

$$\frac{\mathrm{d}N}{\mathrm{d}E}(E|\vartheta) = \frac{\mathrm{d}\Gamma}{\mathrm{d}E} \otimes f_\mathrm{loss}(E,\vartheta) = p_0(\vartheta)\cdot\frac{\mathrm{d}\Gamma}{\mathrm{d}E} + \sum_{n=1}^{\infty} p_n(\vartheta)\cdot\frac{\mathrm{d}\Gamma}{\mathrm{d}E}\otimes f_n(E) \quad (8)$$

where the f_n is the energy loss spectrum of scattering order n which can be approximately defined via recursive convolution through the single scattering energy loss spectrum f_1. This can be written as

$$f_n = f_{n-1}\otimes f_1 \qquad (n>1). \quad (9)$$

The probability p_n that an electron is scattered n times depends on the emission angle ϑ and is given by a Poisson law

$$p_n(\theta) = \frac{\lambda^n(\vartheta)}{n!} e^{-\lambda(\vartheta)}. \quad (10)$$

The average number of scattering processes λ is given in terms of the column density ρd of the tritium source, the mean free column density ρd_free and the scattering cross section σ_scat as

$$\lambda(\vartheta) = \int_0^1 \mathrm{d}x\, \frac{\rho d\cdot x}{\rho d_\mathrm{free}\cdot\cos\vartheta} = \int_0^1 \mathrm{d}x\, \frac{\rho d\cdot x\cdot\sigma_\mathrm{scat}}{\cos\vartheta}. \quad (11)$$

Since the probability of n-fold scattering is a function of the emission angle (10), the response corrected energy spectrum (8) itself becomes dependent on the angle. Note that the scattering model is simplified, since angular changes in collisions are neglected and the scattering probabilities are averaged over a hypothetical uniform density profile in the source. We would like to clarify that in our actual implementation the n-fold energy loss spectra are not generated via convolution but via Monte Carlo, which yields, however, equivalent results. Furthermore, using Eq. (4), the radial starting position is always assumed to be $r = 0$, which is not the case in KATRIN, but we do not expect significant changes in the spectral shape for outer radii. For analysis of real experimental data a fully realistic treatment would be necessary, yet for a principle sensitivity study these approximations are reasonable.

The benefits of a TOF measurement can be understood from Fig. 2, where the TOF (4) as a function of E for different angles is shown. It can be seen that energy differences

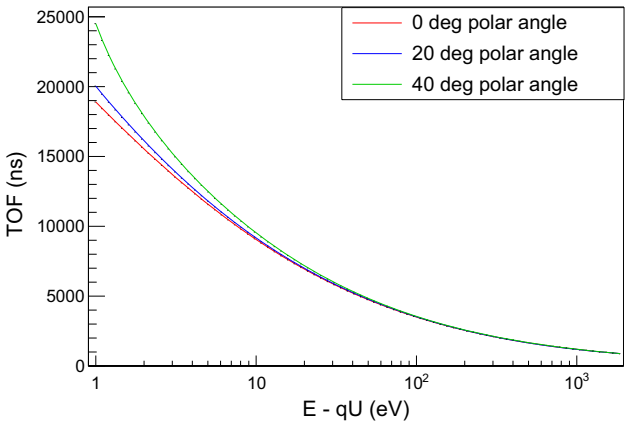

Fig. 2 TOF as a function of surplus energy $E - qU$. Significant energy differences are detectable up to a few 100 eV above the filter threshold. By combining multiple TOF spectra with different retarding energies, the TOF method will give a differential map of the energy spectrum within the measuring interval

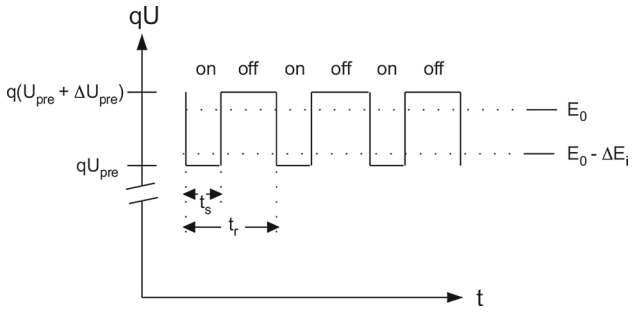

Fig. 3 Pre-spectrometer potential pulsing scheme for gated filter. X-axis: time. Y-axis: pre-spectrometer retarding potential. At the lower filter setting all electrons of the interesting region of width ΔE_i below the endpoint E_0 are transmitted while at the higher setting all electrons are blocked

up to some $\sim$ 100 eV above the retarding potential translate into significant TOF differences. Within these regions, TOF spectroscopy is thus a sensitive *differential* measurement of the energy spectrum. Combining multiple TOF spectra measured at different retarding energies thus allows to measure a differential equivalent of the β-spectrum throughout the whole region of interest. As already outlined in Ref. [31], a differential measurement has important benefits for a sterile neutrino search. On the one hand it enhances the statistical sensitivity since the sterile neutrino signature can be measured directly without any intrinsic background from higher energies as in the classic high pass mode. On the other hand, it reduces the systematic uncertainty since it improves the distinction between systematic effects and a real sterile neutrino signature in the spectrum.

2.3 TOF measurement

As the approach is rather novel, most existing ideas for TOF measurement are still in an early development phase and have not been tested. There are ongoing efforts to develop hardware which is intended to detect passing electrons with minimal interference with their energy (*electron tagger*) [34]. Approaches are amongst others to measure tiny excitations induced in an RF cavity or to detect the weak synchrotron emission of the electrons in the magnetic field via long antennas (cf. Refs. [57,58]). While promising, there has unfortunately not been any break-through in the technical realization for such an electron tagger, yet. Additionally, it seems unlikely that such an approach is also useful for keV sterile neutrino searches. For a sufficient sensitivity on $\sin^2\theta$ the count-rate needs to be as high as possible. However, count-rates much above 10 kcps would lead to ambiguities in the combination of a start signal in the electron tagger and the stop signal in the detector given the overall TOF of order $\sim$ μs (see Fig. 2).

A method which has already been tested in the predating Mainz experiment [59] is a periodic blocking of the electron flux, called *gated filtering* (GF). If electrons are only transmitted during a short fraction of the time, the arrival time spectrum would approximate the TOF spectrum. In KATRIN, this could for instance be achieved by pulsing the pre-spectrometer potential between one setting with full transmission and one setting with zero transmission (Fig. 3). The main downside of the method is that it sacrifices statistics in order to get time information. However, it would require minimal hardware modifications since only the capability to pulse the pre-spectrometer potential by some keV would have to be added. Since the focal plane detector of KATRIN is optimized for low rates near the endpoint, the method could also in principle be utilized for an early keV sterile neutrino search by using a small duty cycle with sharp pulses and thereby reducing the count-rate. However, in this scenario with small hardware modifications, it is unlikely that the pre-spectrometer potential can be pulsed by more than some keV. Due to the capacity of the pre-spectrometer, there is possibly a non-vanishing ramping time involved, depending on the ramping interval. If electrons arrive within the ramping time, they become either accelerated or retarded, giving rise to non-isochronous background. The problem can be mitigated partly by using a voltage supply with higher power. Alternatively, a mechanical high-frequency beam shutter could be use. However, this would come at the cost of larger modifications of the set-up and a lower flexibility regarding fine-tuning of the timing parameters. We will not discuss this problem further and just assume an ideally efficient method of periodically blocking the beam. However, we will restrict the sensitivity study of the sterile neutrino search with the GF method to a measurement region spanning only a few keV below the endpoint.

3 Monte Carlo sensitivity estimation

The TOF spectrum (6) can not be calculated analytically, since the magnetic field $B(z)$ and electron potential $q\Delta U(z)$ are only known numerically. There are two remaining possibilities of simulating TOF spectra. The first approach is to evaluate the δ function in the TOF spectrum (6) via numerical integration. This method has been used in Ref. [34] since it delivers generally precise results and is well scalable. The bottleneck of this method is, however, the convolution of the β-spectrum with the n-fold energy loss spectra (8). The convolution routine is rather performance-intensive especially for a large spectral surplus $E_0 - qU$ (as present in case of keV scale sterile neutrino search) and requires complicated optimizations to work successfully. Furthermore, if the addition of further effects such as angular-changing collisions might be requested for future studies, the implementation will become more difficult.

Therefore, we chose to apply the second approach which is to generate the TOF spectra (6) via Monte Carlo (MC) simulation. This especially avoids the convolution of the β-spectrum with the energy loss function (8), since the energy loss can be randomly generated individually without additional expensive convolutions. While a MC approach is generally very flexible when it comes to the addition of more detailed effects and systematics, it is generally not as scalable in terms of the expected number of events as a purely numerical approach. KATRIN is designed for measurements near the β-endpoint with low rates on the order of several cps. The measurements for the keV-scale sterile neutrino detection, however, have to be performed over a significantly broader region of the β-spectrum and thus count rates up to $\sim 10^{10}$ cps can be expected. For a data taking period of three years, one would thus expect up to $\sim 10^{18}$ cps. If a realistic model for a sensitivity analysis shall be simulated event-by-event, it is obvious that the sample size needs to be significantly larger than the expected number of events. In our case, the calculation of flight times of more than 10^{18} events is simply not possible within a reasonable computing time.

However, we will show that, if the signal is sufficiently small compared to the total expected rate, the dominating "background part" of the model (corresponding to the $\cos^2\theta$-term in Eq. (5)) can be approximated. This works due to the fact that for a pure sensitivity study, as opposed to an analysis of real data, only the fidelity of the signal is relevant.

3.1 Self-consistent approximate Monte Carlo

In this section, we argue that a modified Monte Carlo strategy, from here on called *self-consistent approximate Monte Carlo* (SCAMC), will be able to reduce the necessary total sample size in a sufficient amount to address the problems mentioned above. This works if two requirements are met. These are

1. that the model can be separated into a background part and a signal part, with the latter sufficiently smaller than the first, and
2. that model and toy data are self-consistent, i.e. the toy data are sampled directly from the model.

We will first discuss this approach for a generic case. Assume, the model distribution Φ can be expressed by a linear combination

$$\Phi = c_\mathrm{S}\Phi_\mathrm{S} + c_\mathrm{B}\Phi_\mathrm{B}, \tag{12}$$

consisting of a *signal* contribution $c_\mathrm{S}\Phi_\mathrm{S}$, sampled with maximum precision, and an approximated *background* contribution, $c_\mathrm{B}\Phi_\mathrm{B}$. The distribution of interest is then replaced by a modified distribution

$$\Phi' = c_\mathrm{S}\Phi_\mathrm{S} + c_\mathrm{B}\Phi'_\mathrm{B}, \tag{13}$$

with $\Phi'_\mathrm{B} \sim \Phi_\mathrm{B}$, where the background component is either approximated by an analytic expression or simulated by MC with a reduced sample size. We demand that Φ_B is independent of any parameter of interest, μ (and of any parameter which is strongly correlated with a parameter of interest):

$$\frac{\mathrm{d}\Phi_\mathrm{B}}{\mathrm{d}\mu} = 0. \tag{14}$$

The approximate model (13) can then be used as replacement for the real model. The sensitivity estimation can then be continued in the standard frequentist way: toy data are sampled from Φ' for given parameter choices and the confidence region for the parameter of interest μ can then be determined via χ^2 fits.

The benefit of this strategy can be understood in the following way. Since the data have been sampled from the model, any error in the model will also be passed over to the data. However, while the total approximated distribution Φ' itself is inaccurate, it still contains all essential information about the sensitivity, since $\Phi' - c_\mathrm{B}\Phi'_B = c_\mathrm{S}\Phi_\mathrm{S}$ holds exactly (Fig. 4). Since only the fidelity of the signal is relevant for the sensitivity analysis (which we assured with condition (14)), both the error in the model and in the data approximately cancel each other in the fit. It can be shown in this case that the width of the χ^2 minimum stays the same as long as the background component is at least approximately correct. A simplified proof can be found in Appendix A.

We shall discuss the method now on the initial case of the keV scale sterile neutrino search with TOF spectroscopy. As derived above, the electron TOF spectrum (6) with added sterile neutrinos can be expressed as a superposition of two TOF spectra with a light or heavy neutrino mass, m_l and m_h, respectively. We identify the signal with the sterile neutrino

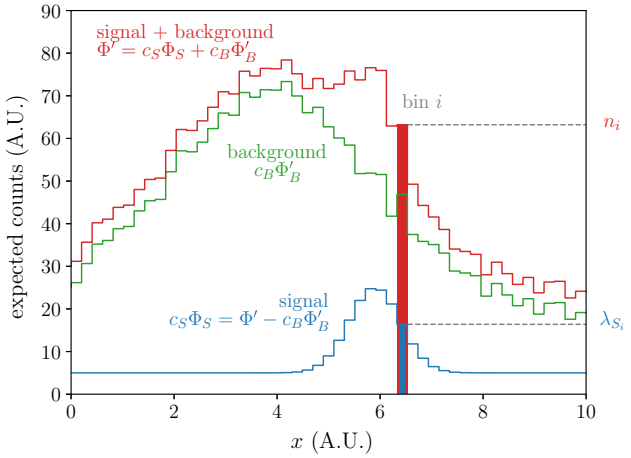

Fig. 4 Illustration on the SCAMC approximated model for the example of a sum of two Gaussians. The signal term $c_S\Phi_S$ is an analytic Gaussian, while the background term $c_B\Phi'_B$ has been sampled with low statistics MC. The total approximated distribution Φ' is then inaccurate but contains all essential information about the signal, because $\Phi' - c_B\Phi'_B = c_S\Phi_S$ holds exactly

component of the TOF spectrum (5) and the background with the active neutrino contribution,

$$\Phi_S = \frac{dN}{d\tau}(m_h) \quad \Phi_B = \frac{dN}{d\tau}(m_l). \tag{15}$$

The coefficients are then given by the active-sterile mixing,

$$c_S = \sin^2\Theta \quad c_B = \cos^2\Theta. \tag{16}$$

It is obvious that for a small signal fraction of, e.g., $\sin^2\theta \lesssim 10^{-6}$, only a small fraction of the total expected events needs to be simulated now. However, since signal and background are always measured together and not separately, the required sample size is reduced even more. For demonstration purposes, let us define the signal expectation value in bin i as

$$\lambda_{S_i} = n \cdot c_S \cdot \Phi_S(x = X_i), \tag{17}$$

with n as total number of expected events (see Fig. 4). We will denote the number of expected events in bin i as n_i. To approximate the necessary sample size, we require that the numerical uncertainty of λ_{S_i} needs to be smaller than the expected measurement uncertainty of the number events in the corresponding bin, σ_i:

$$\Delta\lambda_{S_i} \ll \sigma_i, \tag{18}$$

Assuming a Poissonian measurement uncertainty, $\sigma_i = \sqrt{n_i}$ and using $\Delta\lambda_{S_i}/\lambda_{S_i} = 1/\sqrt{N_{S_i}}$, where N_{S_i} denotes the signal sample size in bin i, Eq. (18) gives

$$\frac{1}{\sqrt{N_{S_i}}} \ll \frac{\sqrt{n_i}}{\lambda_{S_i}} \iff N_{S_i} \gg \frac{\lambda_{S_i}^2}{n_i}, \tag{19}$$

We now define the total signal sample size as $N_S = \sum_i N_{S_i}$. If we assume the signal-background ratio to be roughly within a constant order of magnitude, we get the required minimum signal sample size:

$$N_S \gg \sum_i \frac{\lambda_{S_i}^2}{n_i} \approx \frac{n_i^2 \cdot c_S^2}{n_i} = n \cdot c_S^2. \tag{20}$$

Naively, one would suppose that the signal part still needs to be sampled with full statistics, i.e. $N_S \gg n \cdot c_S$. However, due to the fact that the signal part is always measured with background, we have shown that an additional suppression factor of c_S applies. Assuming $\sin^2\theta \sim 10^{-6}$ and a total event size of $n \sim 10^{18}$, we thus get

$$n \cdot c_S^2 = n \cdot \sin^4\theta \sim 10^{18} \cdot 10^{-12} = 10^6. \tag{21}$$

Note that $\sin^2\theta \sim 10^{-6}$ represents roughly the upper bound from astrophysical observations. Likewise, $n \sim 10^{18}$ is approximately the maximum number of counts which will decrease with higher retarding potentials. Thus, for lower values of either one, the necessary sample size is reduced even more according to condition (20).

4 Simulation

4.1 Probabilistic model: TOF spectra

Using a Monte Carlo algorithm, the TOF spectra given by the transformation (6) can be determined in a straightforward way. For each MC sample, at first an initial energy and starting angle is generated. The angular distribution is given by Eq. (7). For the initial energy, the electronic excited state is generated from the final state distribution in Eq. (1) and then the energy is generated from the respective β-spectrum component in Eq. (2). Given the initial energy and the starting angle, the number of inelastic scattering process in the source is generated from Eq. (10) and for each process the energy loss is generated from Eq. (9) and subtracted from the energy. In order to further optimize the Monte Carlo method for a parametrizable heavy neutrino mass, the TOF spectra have additionally been decomposed into elements corresponding to different sterile neutrino mass phase space segments, which is explained in detail in Appendix B. The advantage of such a scheme is that already simulated Monte Carlo events can be reused for different sterile neutrino masses.

We found that a sample size of 10^8 for each sterile subcomponent is feasible in finite calculation time and sufficient for an accurate simulation. The active neutrino component, which contains $\sim 1/\sin^2\theta$ more counts than the total sterile component, was approximated with a sample size of 10^9, according to the SCAMC approach. The active neutrino mass was set to $m_l = 0$ and the endpoint held constant at

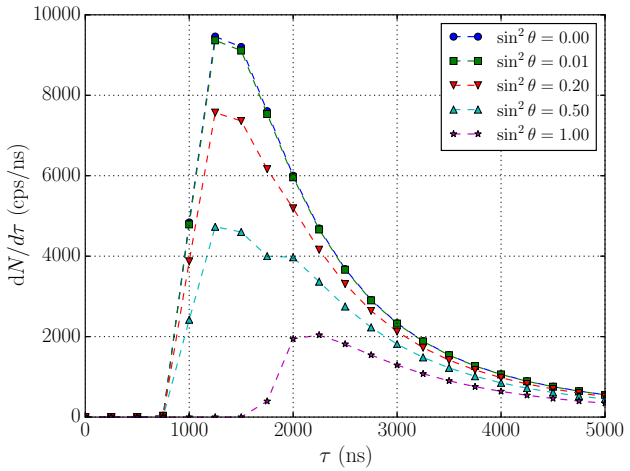

Fig. 5 Electron TOF spectra for a keV-scale sterile neutrino of $m_h = 1.1$ keV and different mixing angles at a fixed retarding potential of 17 keV. The mixing angles have been exaggerated to enhance the signature and comprise additionally the case of no mixing ($\sin^2\theta = 0$) as well of pure sterile contribution ($\sin^2\theta = 1$). Similar to the tritium β-decay energy spectrum, the signature of a sterile neutrino is a kink-like discontinuity at a certain point in the TOF spectrum. Figure first published in [1]

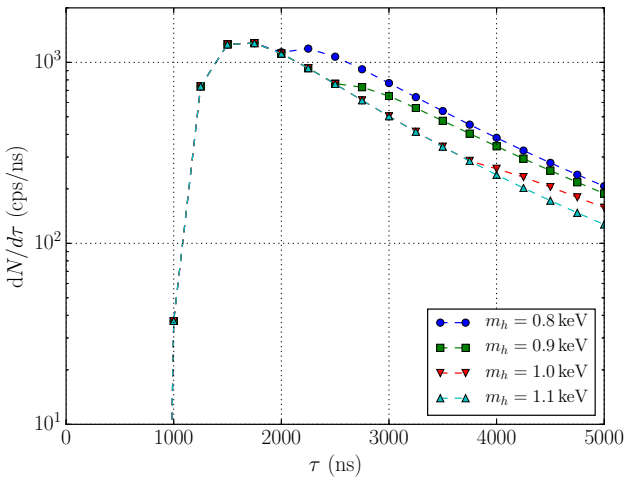

Fig. 6 Electron TOF spectra for different sterile neutrino masses at a fixed retarding potential of 17 keV. The mixing has been set to $\sin^2\theta = 0.5$ to enhance the signature. The heavy neutrino mass determines the position of the kink on the TOF-axis. The on-set TOF for a certain sterile neutrino mass can be estimated from Fig. 2

$E_0 = 18.575$ keV, since there is no correlation to expect with the sterile neutrino. The bin width was chosen to be 250 ns (compared to the FPD time resolution of about 50 ns) for reasons of performance and robustness. However, it is unlikely to expect for any measurement method to achieve a higher resolution. To all spectra a Gaussian time uncertainty of $\Delta\tau = 50$ ns was added to account for the detector time resolution and a isochronous background of $b = 10$ mcps.

Figures 5 and 6 show exemplary simulated TOF spectra for different active-sterile mixings and heavy neutrino masses, respectively. It can be seen that the spectra show a dominating peak within the first 2 μs which consists of the fast electrons more than some 100 eV above the retarding potential. They are, however, followed by a long tail where the electron velocity becomes slower and the TOF difference per given energy difference (see Fig. 2) becomes more significant. In this region the TOF spectrum is to a good extent a differential representation of the β spectrum, while the fast peak region consists only of some bins, thus contributing to the sensitivity more by its integral. If the sterile neutrino mass is some 100 eV smaller than the difference of endpoint and retarding potential, the sterile neutrino signal becomes similar to that one in the tritium beta spectrum. The sterile neutrino contribution appears as a discontinuity in shape of a "kink" at a certain position in the spectrum. Since the relationship between energy and TOF is non-linear, the position of the kink allows no direct analytical conclusion about the sterile neutrino mass. However, given the retarding potential, the relation in Fig. 2 can be used for an estimation.

4.2 Ideal TOF mode sensitivity

The model described in the last section was utilized to estimate the sensitivity according to the procedure described in Sect. 3. The fits have generally been performed by a χ^2 minimizations using MINUIT [60] . For statistical sensitivity estimation, the mixing $\sin^2\theta$ and overall amplitude S are free fit parameters, using a range of fixed values for m_h. In those simulations, where the uncertainty on m_h is of interest, also the squared heavy neutrino mass m_h^2 has been included as fit parameter. Since each fit incorporates a set of multiple measurements at different retarding potentials, the χ^2 functions of each measurement are added and fitted with global fit parameters. Instead of a pure ensemble approach, the parameter uncertainties have been calculated using the module MINOS from MINUIT [60], averaged over multiple simulations, which gives in case of an approximately quadratic χ^2 near the minimum an identical result.

Exemplary systematics

In addition to the statistical sensitivity, an exemplary systematic effect has been studied, which is the inelastic scattering cross section due to fluctuation in the column density as described in Eq. (9). This is one of two main systematics when it comes to keV sterile neutrino search, the other being the final state distribution [51–53]. To incorporate the systematics, the χ^2 function has been modified by an additional term:

$$\chi^2 = \chi_0^2 + \frac{(\rho d - \langle \rho d \rangle)^2}{(\Delta \rho d)^2}, \tag{22}$$

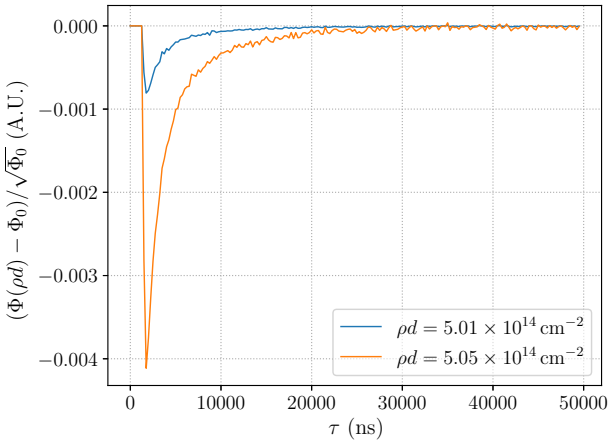

Fig. 7 Difference between TOF spectra with shifted ρd, $\Phi(\rho d)$ and default value $\langle \rho d \rangle = 5 \times 10^{14}\,\mathrm{cm}^{-2}$, Φ_0, weighted proportionally with the expected Poissonian uncertainty of the data $\propto \sqrt{\Phi_0}$. The imprint of a shifted column density is present foremost at lower flight times, due to missing events near the endpoint because of the energy loss. Fluctuations at higher flight times near the retarding potential are suppressed by a lower differential count rate. The spectra consist only of the active neutrino component, $\sin^2\theta = 0$, and the retarding potential is $qU = 18$ kV

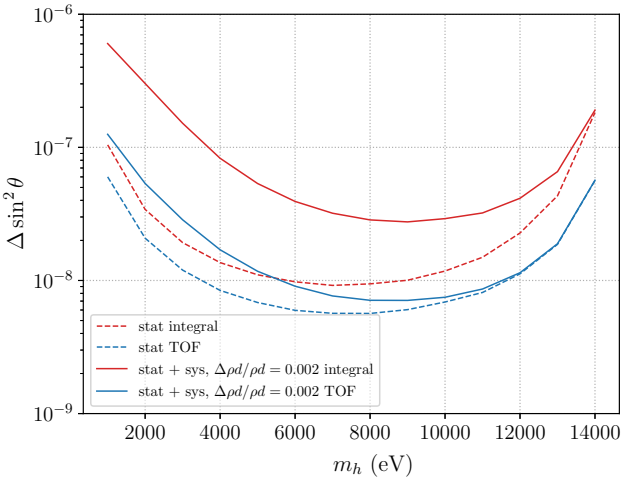

Fig. 8 Sensitivity ($1\,\sigma$) of ideal TOF mode (blue) compared with integral mode (red). Both statistical uncertainty (dashed lines) and combined uncertainty with exemplary systematics (full lines) in form of column density uncertainty $\Delta\rho d/\rho d = 0.002$ affecting the inelastic scattering cross section in the WGTS. It can clearly be seen that the sensitivity gain by a TOF mode is especially significant if the uncertainty of the column density is accounted for. The results are based on 3 years measurement time, distributed uniformly on the retarding potential within an interval of [4; 18.5] keV with steps of 0.5 keV

where χ_0^2 is the default binned χ^2 function, ρd the fitted column density, $\langle \rho d \rangle$ its expectation value and $\Delta\rho d$ the systematic uncertainty. In order to be able to have ρd as free fit parameter, the complete model has additionally been separated by number of inelastic scattering processes and weighted with the l-fold energy loss probability $p_l(\rho d)$ as given by Eq. (10), instead of randomly generating the number of inelastic scattering events,

$$\frac{\mathrm{d}N}{\mathrm{d}\tau} = \sum_l p_l(\rho d) \cdot \left(\frac{\mathrm{d}N}{\mathrm{d}\tau}\right)_l . \tag{23}$$

To determine the influence of the uncertainties $\Delta\rho d$ on the sensitivity, the column density has been shifted by for the data generation by $\rho d = \langle \rho d \rangle + \Delta\rho d$ while still using the unshifted expectation value $\langle \rho d \rangle$ in Eq. (22). By this approach the MINOS error will increase plus a possibly slight bias in average which is then quadratically added to the average error bars.

To illustrate the imprint of the systematic uncertainty of ρd in the TOF spectrum, Fig. 7 shows the difference between a TOF spectrum with shifted column density, $\Phi(\rho d) = \mathrm{d}N/\mathrm{d}\tau(\rho d + \Delta\rho d)$ and a TOF spectrum with mean column density, $\Phi_0 = \mathrm{d}N/\mathrm{d}\tau(\langle \rho d \rangle)$, weighted by $\sqrt{\Phi_0}$ which is proportional to the expected Poissonian uncertainty of the data. By doing so, the signature becomes visible proportionally to its impact in the χ^2 function. It can be seen that the imprint of a shifted column density is present foremost at lower flight times, which is since the energy loss causes the count-rate near the endpoint to drop. There are fluctuations at higher flight times near the retarding potential arising from the energy loss spectrum (9). However, these are weighted minimally since the differential rate in the TOF spectrum drops with higher flight times (see Fig. 5).

Results

Figure 8 shows the sensitivity for an ideal TOF mode. The results are based on three years measurement time which was distributed uniformly on the retarding potential within an interval of [4; 18.5] keV with steps of 0.5 keV. The setting was chosen in that way that a 7 keV neutrino signal [25] would roughly lie in the center of the potential distribution. For the exemplary inelastic scattering systematics an initial uncertainty of $\Delta\rho d/\rho d = 0.002$ has been assumed in accordance with Ref. [30]. The statistical sensitivity of the integral mode in this simulation is in good agreement with Ref. [31]. The statistical sensitivity of the ideal TOF mode is close to that one of an ideal differential detector in the aforementioned publication. However, if the uncertainties of the column density are incorporated, the benefit by the TOF mode grows even further, since a shifted column density has a unique imprint on the TOF spectrum (see Fig. 7), which is not the case in the integral mode.

It should be noted, however, that for low retarding potentials as used in Fig. 8, adiabaticity of the electron transport is limited. Yet, that can be maintained by increasing the magnetic field in the main spectrometer. This lowers the energy resolution and thus the transformation of transverse into longitudinal momentum, which would manifest in a stronger

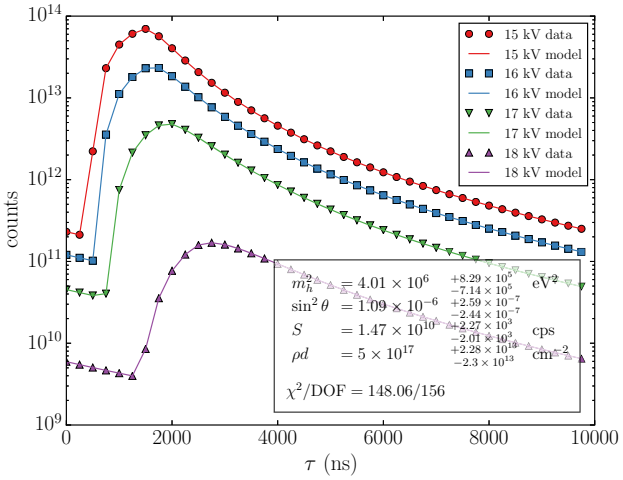

Fig. 9 Exemplary fit of a sterile neutrino with mass $m_h = 2$ keV and low mixing $\sin^2\theta = 10^{-6}$ (not visible by eye), assuming an ideal TOF measurement and using four exemplary retarding potentials of 15, 16, 17 and 18 keV. The fit includes the systematic uncertainty of the column density ρd, as well as the sterile neutrino mass as free fit parameters. The overall count rate increases with decreasing retarding potential

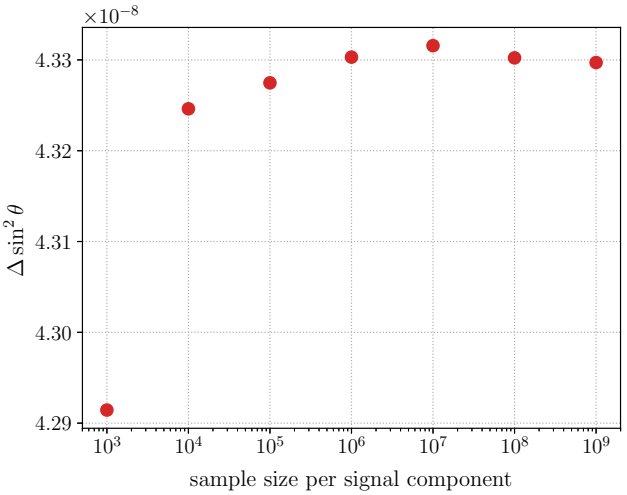

Fig. 10 Estimated statistical sensitivity with ideal TOF mode for a 2 keV neutrino as a function of the MC sample size per signal component (see Appendix B, Eq. (B.6)), using a measurement interval of [4; 18.5] keV with steps of 0.5 keV. The total signal sample size per TOF spectrum is given by Eq. (24). The background has been simulated using 10 times the signal component sample size

angular-dependence of the energy-TOF relation in Fig. 2. Though, this should have no significant influence on the sensitivity since the measurement takes place on a keV-scale where the requirements for magnetic adiabatic collimation are more relaxed.

An exemplary fit is shown in Fig. 9 for a sterile neutrino with mass $m_h = 2$ keV and a mixing of $\sin^2\theta = 10^{-6}$ assuming an ideal TOF measurement and using four exemplary retarding potentials of 15, 16, 17 and 18 keV. In this case the sterile neutrino mass has not been fixed but used as a free fit parameter to test the ability to fit the sterile neutrino mass, given a sufficiently high active-sterile mixing. While it is in principle sufficient to use only one retarding potential closely below the sterile neutrino kink, in practice a multitude of retarding potentials is necessary. The reasons are that, on one hand, the mass of the sterile neutrino is unknown and, on the other hand, that it is also necessary in order to determine the other parameters. In contrast to the pure sterile active mixing sensitivity estimation (Fig. 8) the heavy neutrino mass has been used as free fit parameter. It shows that the method is capable of a sensitive mass determination as well, in case the mixing angle is large enough. However, since most parts of the sensitive regions of the TOF method are disfavored by X ray satellite measurements [15], it seems unlikely that a mass fit will be possible.

4.3 Optimization and integrity

SCAMC variance

In order to show that the SCAMC method is really working as expected, it has been tested using different Monte Carlo sample sizes. A necessary condition is convergence of the result towards a constant value with growing sample size. As described in Appendix B, the signal itself is split into signal components defined by slices of the signal TOF spectrum in m_h-space (Eq. (B.6)). Figure 10 shows the ideal TOF mode statistical sensitivity for a 2 keV neutrino as a function of the sample size used for each such component of the signal. The components have been sampled with steps of 0.1 keV in terms of m_h. The total signal sample size per TOF spectrum is thus given by

$$N_S = N_C \cdot \frac{(E_0 - m_h - qU)}{0.1\ \text{keV}}, \tag{24}$$

where N_C denotes the sample size per component. For minimum qU and m_h it amounts to $\sim 150 \cdot N_C$. The background has been simulated with a sample size of $10 \cdot N_C$.

It can be seen that convergence is met and that with subsample sizes such as 10^4 per component the expected result is approximated with less than 1% uncertainty.

Measurement interval

Figure 11 shows the sensitivity to $\sin^2\theta$ in a similar way as Fig. 8, but for a measurement interval of [15; 18.5] keV, roughly centered around a 2 keV neutrino, as favored in Ref. [3]. It can be seen in comparison that there is no benefit of restricting the measurement interval to a narrow region in search for a sterile neutrino with a given energy. This seems counter-intuitive at first, but is has to be kept in mind that the sterile neutrino signal is not localized at the kink, but instead contributes to the whole spectrum below. In contrast

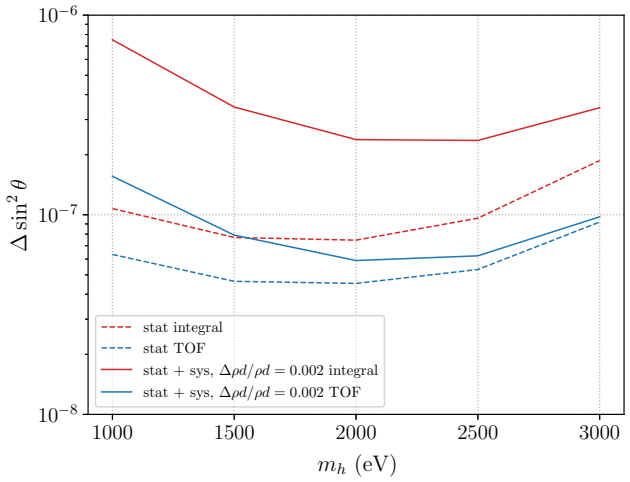

Fig. 11 Same settings as in Fig. 8 but with a measurement interval of [15; 18.5] keV. The narrowing of the measurement interval shows no benefit even if the sterile neutrino kink is within the interval

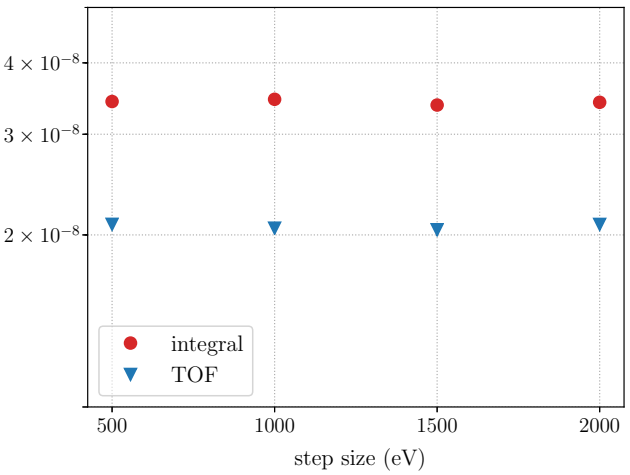

Fig. 12 Statistical sensitivity as a function of the step size between measurement points of the retarding potential qU for a sterile neutrino with $m_h = 2$ keV, with constant total sample size. The measurement interval is [4; 18.5] keV for a total measurement time of 3 years

to dedicated 'kink-search' methods [32], all spectral parts contribute to the sensitivity in a χ^2 fit. While the relative difference made by a sterile neutrino signal might be smaller at lower retarding potentials, this drawback is however balanced by a larger count-rate at lower potentials.

Measurement step size

Figure 12 shows the statistical sensitivity as a function of the spacing between different measurement points of the retarding potential qU. The total sample size has been kept constant. The simulations show no preference towards any particular value. That appears unintuitive, since one would expect a narrower spacing to have beneficial effects on a distinct kink search. Yet, as mentioned in the last paragraph, the sterile neutrino signal is not localized, but manifests itself in relative count rate differences between the measurement points with a spectral feature as broad as the mass of the sterile neutrino m_h. Therefore, a larger step size does not weaken the sensitivity in principle because the measurement time is distributed over less points. Anyway, it is in general recommended to use a step size lower then the smallest possible heavy neutrino mass, since otherwise it is possible that there are not enough vital measurement points above the kink.

The benefit of a TOF measurement can be explained in this context as follows: TOF spectra carry extra information about the differential energy distribution closely above each measurement point. That equates to knowledge about the slope of the integral spectrum at these measurement points.

4.4 Gated filter sensitivity

Figure 13 shows exemplary TOF spectra using Gated Filtering (GF, see Fig. 3). It illustrates how GF works: without the gate (cyan points), the arrival time spectrum is isochronous. However, with activated gate, a certain portion is cut away from the isochronous spectrum. For a given repetition time t_r and duty cycle ξ, the duration in which the gate is open is given by $t_r \cdot \xi$. The GF arrival time filter thus is smeared with a step function when compared to the raw TOF spectrum. Reducing the duty cycle ξ makes the arrival time spectrum approximate the TOF spectrum of Fig. 9, however with a loss of overall rate. Electrons with a TOF greater than the repetition time t_r lead to the wrongful attribution of the corresponding events to a later period, which can be seen in the first few bins. However, since TOF spectra at several keV below the endpoint are rather sharp, this effect is small for repetition times of ~ 10 µs.

Figure 14 shows the sensitivity for two exemplary gated filter scenarios with a constant duty cycle of 0.1. The scenario is based on the assumption that the existing focal plane detector (FPD) of KATRIN is used, which is optimized for a measurement near the endpoint of the β spectrum and thus can not maintain much higher count-rates. The bottleneck is particularly the per-pixel rate which should not exceed $\sim 10^3$ cps within a window of some µs. This limitation holds for the current data acquisition and might be improved in the future. In this simulation, an exemplary overall reduction of the signal rate by a factor 10^5 has been chosen which will ensure a flux compatible with the current hardware. Since the gated filter periodically blocks the flux of electrons, the rate can be increased again with respect to the integral mode. The actual allowed rate with the gated filter depends on the readout electronics and will effectively be between two extremal values. In an optimistic case, short-time excess of the rate is tolerable, while the average rate has to be at the same level as with the integral mode. In a conservative case, also short-time excess leads to pile-up, which means that instead

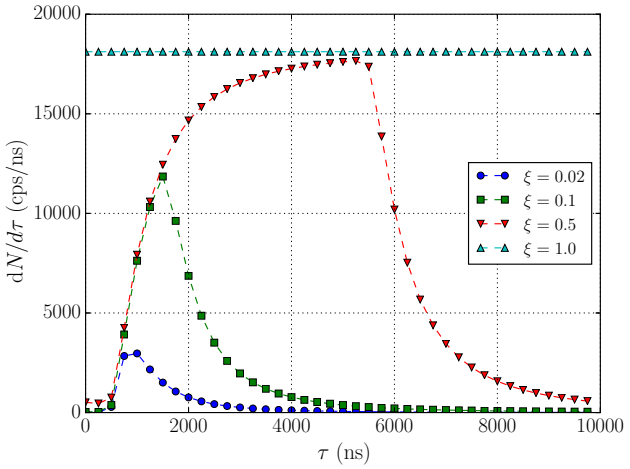

Fig. 13 Exemplary Gated Filter arrival time spectra for different duty cycles. Retarding energy is $qU = 17$ keV and repetition time $t_r = 10\,\mu$s. The active-sterile mixing has been set to $\sin^2\theta = 0$. Activating the gate and decreasing the duty cycle cuts away portions of the arrival time spectrum, which is isochronous without gate

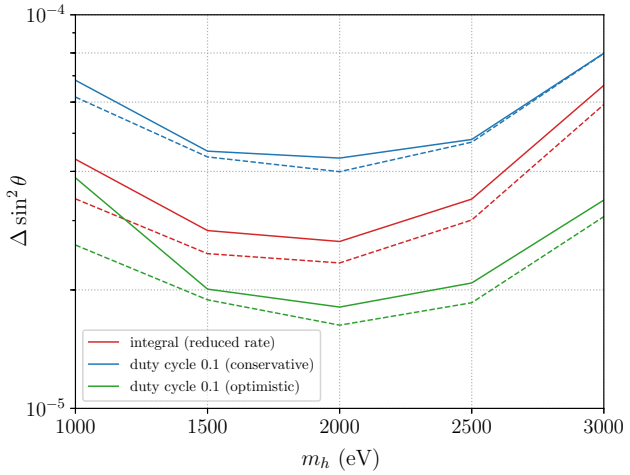

Fig. 14 Sensitivity ($1\,\sigma$) of integral mode with the rate reduced by a factor 1×10^5 (red), compared with a conservative gated filter TOF scenario (blue) with same peak rate as the integral mode and an optimistic gated filter TOF scenario (green) with the same total rate as the integral mode. Both statistical uncertainty (dashed lines) and combined statistical plus systematic uncertainty including a column density uncertainty $\Delta\rho d/\rho d = 0.002$ (full lines, see Sect. 4.2) are plotted. The duty cycle is 0.1 for both gated filter scenarios. Measurement interval has been [15; 18.5] keV for three years data taking. The repetition time is $t_r = 10\,\mu$s for all retarding potentials

the peak rate may not exceed the constant rate of the integral mode (see Fig. 13). The repetition rate has been fixed at 10 μs, which will ensure coverage the vast part of the TOF spectrum. The measurement interval has been limited to [15; 18.5] keV since it is not believed to be viable to pulse the pre-spectrometer more than several keV.

It can be seen that in the gated filter beats the integral mode in the optimistic case, but not in the conservative case. This means that the loss of rate in the conservative case is too high to be compensated by the additional TOF information. In case the detector readout electronics sufficiently tolerates short-time excesses, the loss of statistics by the gated filter can, nevertheless, be compensated and additional TOF information is gained. Note, however, that in the scenario of an upgraded future detector which tolerates the full rate from the tritium source, the integral mode outperforms the gated filter mode, since there is now way in this case to increase the rate further.

5 Summary and discussion

It has been shown that TOF spectroscopy in a KATRIN context is in principle able to boost the sensitivity of the sterile neutrino search significantly. Figure 11 suggests an improvement of up to a factor two in terms of pure statistical uncertainty down to at maximum $\sin^2\theta \sim 5 \times 10^{-9}$ for a sterile neutrino of $m_h = 7$ keV at one σ. If the exemplary systematic uncertainty of the inelastic scattering cross section is considered, the sensitivity is only mildly weakened in contrast to the integral mode, which is in that case outperformed by the TOF mode by up to a factor five. However, the practical realization of a sensitive TOF measuring method is still work in progress. Given the current hardware, which requires a reduction of the signal rate, the gated filter method might be able to realize a TOF mode with a slight sensitivity increase compared to the integral mode under the condition that the detector tolerates short-time excesses of the rate and that it is possible to ramp the pre-spectrometer potential some keV within $\sim 0.1\,\mu$s. From a long-term point of view, the concept of an upgraded differential detector [31] which is capable of extreme rates up to 10^{10} cps is very promising. If there is sufficient progress in developing a sensitive TOF measurement method, a beneficial strategy could be a combined measurement to eliminate systematics and perform cross-checks.

Acknowledgements We would like to thank S. Enomoto for discussions. This work is partly funded by BMBF under Contract no. 05A11PM2 and DFG GRK 2149.

Funded by SCOAP[3].

Appendix A: Unchanged χ^2 properties with SCAMC

In the following it is shown that the properties of the χ^2 function defining the sensitivity, which are position and width

of the minimum with respect to any parameter of interest, are independent of the choice of the background model Φ'_{B}. This works as well for a Poissonian log-likelihood, but for brevity we show it on a χ^2 example. First we define the expectation value for the i-th bin,

$$\lambda'_i = \lambda_{\mathrm{S}_i} + \lambda'_{B_i} = n\,(c_{\mathrm{S}}\Phi_{\mathrm{S}_i} + c_{\mathrm{B}}\Phi'_{B_i}), \tag{A.1}$$

using the definition of the approximated model (13), and assume that the background prediction λ'_{B_i} is independent of the parameter of interest μ,

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\lambda'_{B_i} = 0. \tag{A.2}$$

For the proof we differentiate χ^2 with respect to μ and demand that the result is approximately independent of the choice of the background model Φ'_{B}:

$$\begin{aligned}
\chi^2(\mu) &= \sum_i \frac{(n_i - \lambda'_i(\mu))^2}{\lambda'_i(\mu)} \qquad (A.3)\\
\frac{\mathrm{d}}{\mathrm{d}\mu}\chi^2 &= \sum_i \frac{\lambda'_i \frac{\mathrm{d}}{\mathrm{d}\mu}(n_i - \lambda'_i)^2 - (n_i - \lambda'_i)^2 \frac{\mathrm{d}}{\mathrm{d}\mu}\lambda'_i}{\lambda'^2_i}\\
&= \sum_i \frac{-2\lambda'_i(n - \lambda'_i)\frac{\mathrm{d}}{\mathrm{d}\mu}\lambda_{\mathrm{S}_i} - (n - \lambda'_i)^2 \frac{\mathrm{d}}{\mathrm{d}\mu}\lambda_{\mathrm{S}_i}}{\lambda'^2_i}\\
&= -\sum_i \frac{(n_i^2 - \lambda'^2_i)\frac{\mathrm{d}}{\mathrm{d}\mu}\lambda_{\mathrm{S}_i}}{\lambda'^2_i}\\
&= \sum_i \left(1 - \frac{n_i^2}{\lambda'^2_i}\right)\frac{\mathrm{d}}{\mathrm{d}\mu}\lambda_{\mathrm{S}_i}\\
&= \sum_i \left(1 - \frac{n_i^2}{(\lambda'_{B_i} + \lambda_{\mathrm{S}_i})^2}\right)\frac{\mathrm{d}}{\mathrm{d}\mu}\lambda_{\mathrm{S}_i}\\
&= \sum_i \left(1 - \left(\frac{\lambda'_{B_i}}{n_i} + \frac{\lambda_{\mathrm{S}_i}}{n_i}\right)^{-2}\right)\frac{\mathrm{d}}{\mathrm{d}\mu}\lambda_{\mathrm{S}_i} \qquad (A.4)
\end{aligned}$$

The variable n_i is Poisson distributed with mean $\lambda'_i(\mu_0) = \lambda_{\mathrm{S}_i}(\mu_0) + \lambda'_{B_i}$, where μ_0 is the null-hypothesis for μ. Due to self-consistency, $\frac{\lambda'_{B_i}}{n_i}$ is approximately independent from the choice of Φ'_{B}, as long as the order of magnitude is in agreement $\Phi'_B \sim \Phi_{\mathrm{B}}$. The latter condition ensures that the Poissonian uncertainty of n_i, which is given by $\sqrt{\lambda'_i(\mu_0)}$, is approximately correct. □

Note that the proof is only correct in the simplified case of one parameter of interest and no correlation with nuisance parameters. However, the simulation results in this paper show that there is valid reason to expect the method to work also for more complex problems as long as there is no heavy parameter correlation.

Appendix B: Sterile neutrino mass decomposition of TOF spectra

The simulation of the TOF spectra has further been optimized with the aim of being able to use the sterile neutrino mass m_l as a free parameter with a minimum of computational overhead. The idea is to decompose the sterile neutrino components of the TOF spectra, Φ_{S}, into sub-spectra Φ_{S_k} which can be added subsequently to obtain the signal for a given sterile neutrino mass m_h. That works as follows: at first a number J of grid points with heavy neutrino masses m_j are chosen. For each grid-point j, the signal spectrum is given as the sum of all sub-signals from j up to J,

$$\Phi_{\mathrm{S}}(m_j) = \sum_{k=j}^{J} \Phi_{S_k}. \tag{B.5}$$

The sub-signals Φ_{S_k} constitute the difference of two TOF spectra with adjacent sterile neutrino masses. The total TOF spectrum for the sterile component can then be written as

$$\frac{\mathrm{d}N}{\mathrm{d}\tau}(m_j) = \frac{\mathrm{d}N}{\mathrm{d}\tau}(m_J) + \sum_{k=j}^{J-1}\left(\frac{\mathrm{d}N}{\mathrm{d}\tau}(m_k) - \frac{\mathrm{d}N}{\mathrm{d}\tau}(m_{k+1})\right). \tag{B.6}$$

Each sub-component in the sum will be sampled separately. The difference between two TOF spectra can be sampled just like any TOF spectrum, as outlined, by replacing the β-spectrum in (6) also with the difference of two β spectra corresponding to the neutrino masses m_k and m_{k+1}. Via

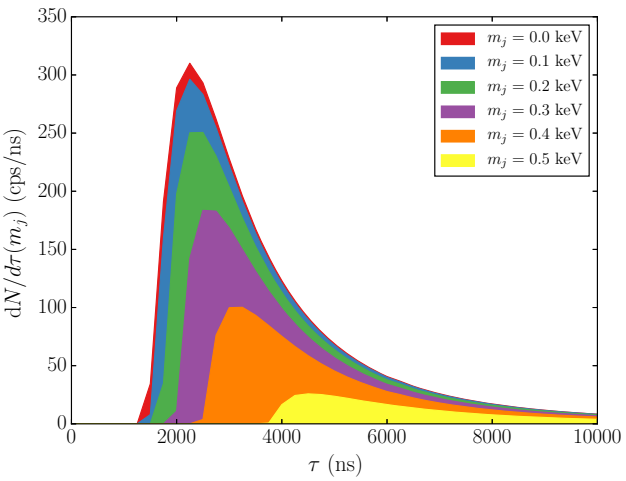

Fig. 15 Illustration of the calculation of sterile component of the electron TOF spectrum via subsequent addition of sub-components according to (B.6). The figure shows the sterile components of the TOF spectrum (5) for different sterile neutrino masses m_j on a grid for a retarding potential of $qU = 18$ kV. Each colored area corresponds to a sub-component between two adjacent mass values. The component for any sterile neutrino mass m_j is then given by the sum of all areas below the envelope

(B.6), that gives then the sterile contribution of the TOF spectrum for each mass value m_j on the grid. For sterile neutrino masses between the grid points, the resulting spectrum is then calculated by cubic spline interpolation. The strategy is illustrated in Fig. 15.

In addition to the reuse of already simulated Monte Carlo events, this strategy has the possible advantage of a smoother interpolation in bins with small statistics, which are possible for high flight times $\gtrsim 40$ μs. By the de-composition and subsequent addition of the components, monotony between the interpolation grid points is guaranteed. However, if a sufficient overall sample size is chosen, this effect should not matter significantly.

References

1. M. Drewes, T. Lasserre, A. Merle, S. Mertens et al., J. Cosmol. Astropart. Phys. **2017**(1), 025 (2017). https://doi.org/10.1088/1475-7516/2017/01/025
2. C. Destri, H.J. De Vega, N.G. Sanchez, New Astron. **22**, 39 (2013). https://doi.org/10.1016/j.newast.2012.12.003
3. C. Destri, H. de Vega, N. Sanchez, Astropart. Phys. **46**, 14 (2013). https://doi.org/10.1016/j.astropartphys.2013.04.004
4. L. Canetti, M. Drewes, M. Shaposhnikov, Phys. Rev. Lett. **110**(6), 061801 (2013). https://doi.org/10.1103/PhysRevLett.110.061801
5. H.J. De Vega, P. Salucci, N.G. Sanchez, New Astron. **17**(7), 653 (2012). https://doi.org/10.1016/j.newast.2012.04.001
6. N. Menci, F. Fiore, A. Lamastra, Mon. Not. R. Astron. Soc. **421**(3), 2384 (2012). https://doi.org/10.1111/j.1365-2966.2012.20470.x
7. M.R. Lovell, V. Eke, C.S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S.D.M. White, A. Boyarsky, O. Ruchayskiy, Mon. Not. R. Astron. Soc. **420**(3), 2318 (2012). https://doi.org/10.1111/j.1365-2966.2011.20200.x
8. N.W. Evans, J. An, M.G. Walker, Mon. Not. R. Astron. Soc. Lett. **393**(1), L50 (2009). https://doi.org/10.1111/j.1745-3933.2008.00596.x
9. A. Schneider, R.E. Smith, A.V. Maccio, B. Moore, Mon. Not. R. Astron. Soc. **424**(1), 684 (2012). https://doi.org/10.1111/j.1365-2966.2012.21252.x
10. C. Destri, H.J. De Vega, N.G. Sanchez, Phys. Rev. D **88**(8), 083512 (2013). https://doi.org/10.1103/PhysRevD.88.083512
11. E. Papastergis, R. Giovanelli, M.P. Haynes, F. Shankar, Astron. Astrophys. **113**(2011), 15 (2014). https://doi.org/10.1051/0004-6361/201424909
12. H.J. de Vega, P. Salucci, N.G. Sanchez, Mon. Not. R. Astron. Soc. **442**, 2717 (2014). https://doi.org/10.1093/mnras/stu972
13. T. Chan, D. Keres, J. Onorbe, P. Hopkins, A. Muratov, C.A. Faucher-Giguere, E. Quataert, Mon. Not. R. Astron. Soc. **454**(3), 2981 (2015). https://doi.org/10.1093/mnras/stv2165
14. A. Boyarsky, A. Neronov, O. Ruchayskiy, M. Shaposhnikov, Mon. Not. R. Astron. Soc. **370**(1), 213 (2006). https://doi.org/10.1111/j.1365-2966.2006.10458.x
15. C.R. Watson, Z. Li, N.K. Polley, J. Cosmol. Astropart. Phys. **2012**(03), 018 (2012). https://doi.org/10.1088/1475-7516/2012/03/018
16. S. Dodelson, L.M. Widrow, Phys. Rev. Lett. **72**(1), 17 (1994). https://doi.org/10.1103/PhysRevLett.72.17
17. A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, J. Cosmol. Astropart. Phys. **2009**(03), 005 (2009). https://doi.org/10.1088/1475-7516/2009/03/005
18. H. Yüksel, J.F. Beacom, C.R. Watson, Phys. Rev. Lett. **101**(12), 1 (2008). https://doi.org/10.1103/PhysRevLett.101.121301
19. B. Shakya, Mod. Phys. Lett. A **31**(06), 1630005 (2016). https://doi.org/10.1142/S0217732316300056
20. A. Merle, A. Schneider, M. Totzauer, J. Cosmol. Astropart. Phys. **2016**(04), 003 (2016). https://doi.org/10.1088/1475-7516/2016/04/003
21. X. Shi, G.M. Fuller, Phys. Rev. Lett. **82**(14), 2832 (1999). https://doi.org/10.1103/PhysRevLett.82.2832
22. M. Shaposhnikov, J. High Energy Phys. **08**(08), 008 (2008). https://doi.org/10.1088/1126-6708/2008/08/008
23. M. Laine, M. Shaposhnikov, J. Cosmol. Astropart. Phys. **2008**(06), 031 (2008). https://doi.org/10.1088/1475-7516/2008/06/031
24. A. Schneider, J. Cosmol. Astropart. Phys. **2016**, 059 (2016). https://doi.org/10.1088/1475-7516/2016/04/059
25. E. Bulbul, M. Markevitch, A. Foster, R.K. Smith, M. Loewenstein, S.W. Randall, Astrophys. J. **789**(1), 13 (2014). https://doi.org/10.1088/0004-637X/789/1/13
26. A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, J. Franse, Phys. Rev. Lett. **113**(25), 251301 (2014). https://doi.org/10.1103/PhysRevLett.113.251301
27. A. Merle, A. Schneider, Phys. Lett. B **749**, 283 (2015). https://doi.org/10.1016/j.physletb.2015.07.080
28. R.E. Shrock, Phys. Lett. B **96**(1–2), 159 (1980). https://doi.org/10.1016/0370-2693(80)90235-X
29. H.J. de Vega, O. Moreno, E. Moya de Guerra, M. Ramon Medrano, N.G. Sanchez, Nucl. Phys. B **866**(2), 177 (2013)
30. KATRIN Collaboration, KATRIN Design Report. Scientific Report 7090, FZKA (2004). https://www.katrin.kit.edu/publikationen/DesignReport2004-12Jan2005.pdf
31. S. Mertens, T. Lasserre, S. Groh, G. Drexlin, F. Glück, A. Huber, A. Poon, M. Steidl, N. Steinbrink, C. Weinheimer, J. Cosmol. Astropart. Phys. **2015**(02), 020 (2015). https://doi.org/10.1088/1475-7516/2015/02/020
32. S. Mertens, K. Dolde, M. Korzeczek, F. Glueck, S. Groh, R.D. Martin, A.W.P. Poon, M. Steidl, Phys. Rev. D **91**(4), 042005 (2015). https://doi.org/10.1103/PhysRevD.91.042005
33. K. Dolde, S. Mertens, D. Radford, T. Bode, A. Huber, M. Korzeczek, T. Lasserre, M. Slezak, Nucl. Instrum. Methods Phys. Res. A **848**, 127 (2017). https://doi.org/10.1016/j.nima.2016.12.015
34. N. Steinbrink, V. Hannen, E.L. Martin, R.G.H. Robertson, M. Zacher, C. Weinheimer, New J. Phys. **15**, 113020 (2013). https://doi.org/10.1088/1367-2630/15/11/113020
35. J. Formaggio, J. Barrett, Phys. Lett. B **706**, 5 (2011). https://doi.org/10.1016/j.physletb.2011.10.069
36. A.S. Riis, S. Hannestad, J. Cosmol. Astropart. Phys. **2011**(02), 011 (2011). https://doi.org/10.1088/1475-7516/2011/02/011
37. A. Esmaili, O.L.G. Peres, Phys. Rev. D **85**, 117301 (2012). https://doi.org/10.1103/PhysRevD.85.117301
38. C. Kraus, A. Singer, K. Valerius, C. Weinheimer, Eur. Phys. J. C **73**(2), 1 (2013). https://doi.org/10.1140/epjc/s10052-013-2323-z
39. S. Gariazzo, C. Giunti, M. Laveder, Y.F. Li, J. High Energy Phys. **17**(6), 135 (2017). https://doi.org/10.1007/jhep06(2017)135
40. LSND Collaboration, Phys. Rev. Lett. **81**(9), 1774 (1998). https://doi.org/10.1103/PhysRevLett.81.1774
41. MiniBooNE Collaboration, Phys. Rev. Lett. **98**(23), 231801 (2007). https://doi.org/10.1103/PhysRevLett.98.231801
42. SAGE Collaboration, Phys. Rev. C **80**(1) (2009). https://doi.org/10.1103/PhysRevC.80.015807
43. F. Kaether, W. Hampel, G. Heusser, J. Kiko, T. Kirsten, Phys. Lett. B **685**(1), 47 (2010). https://doi.org/10.1016/j.physletb.2010.01.030
44. G. Mention, M. Fechner, T. Lasserre, T.A. Mueller, D. Lhuillier, M. Cribier, A. Letourneau, Phys. Rev. D **83**(7), 073006 (2011). https://doi.org/10.1103/PhysRevD.83.073006

45. C. Giunti, M. Laveder, Phys. Rev. C **83**(6), 065504 (2011). https://doi.org/10.1103/physrevc.83.065504
46. V.S. Basto-Gonzalez, A. Esmaili, O.L.G. Peres, Phys. Lett. B **718**(3), 1020 (2013). https://doi.org/10.1016/j.physletb.2012.11.048
47. J. Barry, J. Heeck, W. Rodejohann, J. High Energ. Phys. **14**(7), 81 (2014). https://doi.org/10.1007/JHEP07(2014)081
48. N.M.N. Steinbrink, F. Glück, F. Heizmann, M. Kleesiek, K. Valerius, C. Weinheimer, S. Hannestad, J. Cosmol. Astropart. Phys. **2017**(06), 015 (2017). https://doi.org/10.1088/1475-7516/2017/06/015
49. E.W. Otten, C. Weinheimer, Rep. Prog. Phys. **71**(8), 086201 (2008). https://doi.org/10.1088/0034-4885/71/8/086201
50. G. Drexlin, V. Hannen, S. Mertens, C. Weinheimer, Adv. High Energy Phys. **2013**, 293986 (2013). https://doi.org/10.1155/2013/293986
51. A. Saenz, S. Jonsell, P. Froelich, Phys. Rev. Lett. **84**(2), 242 (2000). https://doi.org/10.1103/PhysRevLett.84.242
52. N. Doss, J. Tennyson, A. Saenz, S. Jonsell, Phys. Rev. C **73**(2), 025502 (2006). https://doi.org/10.1103/PhysRevC.73.025502
53. N. Doss, J. Tennyson, J. Phys. B **41**(12), 125701 (2008). https://doi.org/10.1088/0953-4075/41/12/125701
54. M. Babutzka et al., New J. Phys. **14**, 103046 (2012). https://doi.org/10.1088/1367-2630/14/10/103046
55. A. Picard et al., Nucl. Instrum. Methods Phys. Res. B **63**(3), 345 (1992). https://doi.org/10.1016/0168-583X(92)95119-C
56. D.T. Gillespie, Am. J. Phys. **51**(6), 520 (1983). https://doi.org/10.1119/1.13221
57. B. Monreal, J.A. Formaggio, Phys. Rev. D **80**(5), 1 (2009). https://doi.org/10.1103/PhysRevD.80.051301
58. Project 8 Collaboration (Asner et al.), Phys. Rev. Lett. **114**(16), 1 (2015). https://doi.org/10.1103/PhysRevLett.114.162501
59. J. Bonn, L. Bornschein, B. Degen, E.W. Otten, C. Weinheimer, Nucl. Instrum. Methods Phys. Res. A **421**(1–2), 256 (1999). https://doi.org/10.1016/S0168-9002(98)01263-7
60. F. James, M. Roos, Comput. Phys. Commun. **10**(6), 343 (1975). https://doi.org/10.1016/0010-4655(75)90039-9