Karlsruhe Institute of Technology

Application of machine learning algorithms for improved background characterization at the KATRIN experiment

at the Department of Physics
Institute of Experimental Particle Physics

Master's Thesis

of

Johannes Heizmann

Reviewer:	Prof. Dr. Guido Drexlin
Second Reviewer:	Prof. Dr. Ulrich Husemann
Advisor:	Dr. Florian Fränkle

November, 2017 – October, 2018

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, October 31, 2018

. . . . . .

(Johannes Heizmann)

Zusammenfassung

Das Neutrino wurde erstmals von W. Pauli im Jahre 1930 postuliert [PKW64]. Er sagte damals schon voraus, dass das Neutrino sehr schwer nachzuweisen sei, weil es nur über die schwache Wechselwirkung mit anderen Teilchen interagiert. Als die Existenz des Neutrinos jedoch ca. ein Vierteljahrhundert später nachgewiesen wurde [RC53], war der Weg frei für die Neutrinophysik. Während das Neutrino im Standardmodell der Teilchenphysik keine Masse besitzt, haben Experimente um die Jahrtausendwende Neutrino Flavor-Oszillationen nachgewiesen (Kapitel 2). Für diese bahnbrechende Entdeckung wurde 2015 der Nobelpreis in Physik verliehen [Nob18]. Dieses Phänomen zeigt, dass Neutrinos eine nicht verschwindende Masse besitzen müssen, jedoch kann in diesen Oszillationsexperimenten die absolute Masse des Neutrinos nicht bestimmt werden. Der herausfordernde Nachweis dieser kleinen Neutrinomasse würde Einblicke in fundamentale Fragen der Astroteilchen Physik geben: Warum besteht das Universum zum Großteil aus Materie und nicht aus Antimaterie, wie hat das Neutrino als häufigstes Teilchen den strukturbildenden Prozess des Universums beeinflusst?

Die derzeitig stringentesten oberen Limits auf die Elektron-Antineutrino-Masse aus direkten Messungen kommen von den Experimenten in Mainz und Troitsk, welche mit Hilfe von hochauflösender Tritium-β-Spektroskopie die Form des Energiespektrums der Elektronen nahe des Endpunktes abtasten. Das derzeitige Limit von 2 eV stammt aus einer kombinierten Massenanalyse dieser beiden Experimente [TH18]. Das KArlsruhe TRItium Neutrino experiment (KATRIN) (Kapitel 3) macht sich das gleiche Messprinzip zum Vorteil und wurde gebaut, um die Elektron-Antineutrino-Masse mit einer bisher unerreichten Sensitivität von 0,2 eV bei 90 % C.L. zu messen [KAT05]. Die Basis um diese noch nie dagewesene Sensitivität zu erreichen bildet ein verbessertes und hochauflösendes MAC-E-Filter-Modul, gepaart mit einer hoch aktiven Quelle von Tritium-β-Zerfallselektronen um diese Sensitivität in einer angemessenen Zeit zu erreichen. Da nur ein winziger Bruchteil der Signalelektronen nahe des Endpunktes der Energieverteilung wichtige Informationen über die Neutrinomasse trägt, ist ein niedriger Untergrund in den Spektrometern unabdingbar. Hierfür ist ein detailliertes Verständnis der Untergrund erzeugenden Prozesse notwendig.

Seit Inbetriebnahme des Vorspektrometers gab es erfolgreiche Studien über MAC-E-Filter-abhängige Untergrundprozesse, wie Penning-Fallen oder myon- und radoninduzierte Ereignisse (Kapitel 4). Eine mögliche Untergrundquelle ist Tritium, das von der Quelle entweder in molekularer Form (HT, T_2) in den Spektrometerbereich diffundiert oder als

Ionen von den magnetischen Feldern geführt wird. Der Zerfall eines Tritiumatoms im magnetischen Flussschlauch der Spektrometer führt zu einem Elektron, welches bis zu 18,6 keV kinetische Energie besitzt und deshalb mit hoher Wahrscheinlichkeit aufgrund der MAC-E-Filter-Eigenschaften in einem der Spektrometer gespeichert ist. Durch Streuungen an Restgasmolekülen kann dieses dort bis zu mehrere hundert Sekundärelektronen erzeugen. Diese hinterlassen auf dem Detektor eine charakteristische Clustersignatur. Die gleiche Signatur wurde schon bei Elektronen mit Energien von mehreren keV beobachtet, die begleitend zu Radonzerfällen entstehen. Um den Beitrag des radoninduzierten Untergrund zu erkennen und abzuschätzen, wurde ein Clusteralgorithmus zusammen mit einer Messung bei künstlich erhöhtem Druck durch Argon oder Helium verwendet.

In dieser Arbeit wird ein neuer Ansatz zur Clustererkennung mit Hilfe von statistischen Lernmethoden verfolgt (Kapitel 5). Indem ein sogenannter Trainingsdatensatz unter Einbeziehung von Messungen und Simulationen konstruiert wird, können die Parameter der Algorithmen angepasst werden um verbesserte Klassifizierungsergebnisse zu erzielen.

Charakteristische Eigenschaften von Clusterereignissen sind Spitzen in der Rate und eine Erhöhung der Häufigkeitsverteilung der Zeiten zwischen zwei Ereignissen. Eine weitere Eigenschaft in der Häufigkeitsverteilung der Radiusdifferenzen zwischen zwei Ereignissen konnte im Zuge dieser Arbeit aufgezeigt werden (Kapitel 6). Während für die Enthüllung der bereits bekannten Eigenschaften ein erhöhter Druck in den Spektrometern nötig ist, ist diese neue Eigenschaft auch bei niedrigem Druck präsent. Sie kann deshalb nicht nur dazu benutzt werden um die Anwesenheit von Clusterereignissen robuster zu indizieren, sondern dient auch als Effektivitätsmaß für die Detektion durch Clusteralgorithmen. Diese Algorithmen werden verwendet um den Anteil an Radon, das aus den Getterpumpen des Vorspektrometers in das Hauptspektrometervolumen strömt, während der SDS3-Chrismas-Messungen abzuschätzen (Kapitel 6). Mit Hilfe der neuen Algorithmen kann dieser Beitrag genauer abgeschätzt und die Entscheidung erleichtert werden, ob das Gettermaterial im Vorspektrometer entfernt werden soll um den Untergrund des Hauptspektrometers während der Langzeit-Neutrinomassen-Messung zu senken. Abschließend werden die Ergebnisse von Untergrundreferenzmessungen im Vorspektrometer dazu verwendet, den Einfluss und die Menge von tritiuminduziertem Untergrund abzuschätzen. Die Arbeit schließt mit einer kurzen Zusammenfassung und einem Ausblick über mögliche Verbesserungen (Kapitel 7).

Contents

1 Introduction

The neutrino as an elementary particle was first postulated in 1930 by W. Pauli [PKW64]. He claimed that the existence of this particle would be difficult to observe, as it only takes part in the weak interaction. When it was found a quarter century later by [RC53], the journey of experimental neutrino physics started. While it was first assumed that the neutrino is massless, experiments around the turn of the millennium observed neutrino flavor oscillations. This phenomenon shows that neutrinos have non-vanishing masses (chapter 2). For this ground-breaking revelation, the Nobel Prize in physics was awarded in the year 2015 [Nob18]. However, these experiments do only prove that neutrinos have mass, but not the absolute value of it. The determination of the absolute mass would shed light into fundamental questions of astroparticle physics: Why is it that we live in a matter dominated universe today? How is the structure formation process in the universe influenced by its most abundant particle, the neutrino?

The current most stringent upper limits from direct observations on the electron-antineutrino mass come from experiments in Mainz and Troitsk, which precisely measure the electron's energy spectrum from tritium β-decay close to its endpoint. In a combined analysis they state an upper mass limit of 2 eV. The KATRIN experiment (chapter 3) utilizes the same measurement principle as its predecessors in Mainz and Troitsk and was built to measure the electron antineutrino mass with an unmatched sensitivity of 200 meV at 90 % C.L. [KAT05]. The key point to achieve this novel sensitivity is the high energy resolution of the Magnetic Adiabatic Collimation with Electrostatic filter (MAC-E filter) module, paired with a high luminosity source of tritium β-decay electrons. As only a tiny fraction of the signal electrons close to the energy spectrum's endpoint carries information about the neutrino mass, an ultra low background within the spectrometer section is required. Therefore detailed knowledge of the background generating processes is vital.

There have been successful studies about MAC-E filter related background processes such as Penning discharge, muon- and radon induced events, from the commissioning of the Pre-Spectrometer (PS). A possible background source in future neutrino mass measurements is tritium, which can diffuse from the source into the spectrometer section as molecules (HT, T_2), or can be magnetically guided as ions. A tritium decay in the spectrometers flux tube results in a <18.6 keV electron, which may likely be stored in the flux tube due to the MAC-E filter properties. There it can create up to several hundreds of secondary electrons by scattering off residual gas molecules. These secondary electrons leave a characteristic cluster signature on the Focal-Plane Detector (FPD) (chapter 4).

These clusters have also been observed from radon decay accompanied emissions of electrons with keV energies. In order to estimate this background contribution, a cluster algorithm paired with a measurement of the background at elevated pressure was developed. At elevated pressure, the mean time between scattering processes is shorter and therefore cluster events can be distinguished from other background events by their different time scale.

In this work, a novel approach to identify cluster events utilizing statistical learning algorithms is presented in chapter 5. In order to train and evaluate the algorithms, a hybrid training data set was constructed of both measurement data and simulations. This unique training data set allows the tuning of the algorithm's input parameters to provide unmatched classification results.

The properties of data hinting the presence of cluster events are spikes in the rate trend of events, and a surplus of the distribution of the time difference of consecutive events at short times. In the course of this work, another feature indicating cluster events is discovered in the distribution of the radii difference of consecutive events (chapter 6). While for the former properties an elevated pressure was necessary to reveal them, this feature is also present at low pressures. It can therefore be used to not only indicate the presence of cluster events in a robust way, but also as a measure for the effectiveness of a cluster detection algorithm. These algorithms are then used to estimate the influence of radon emanating from the Non-Evaporable Getter (NEG) pumping material of the PS into the Main Spectrometer (MS) (chapter 6). This provides an answer to the question whether the NEG pumps could be de-installed in the PS to decrease the MS background rate for the long-term neutrino mass measurements. Furthermore, dedicated reference background measurements in the PS are used, to estimate the impact of the background contribution due to tritium decays, as well as the number of corresponding tritium atoms. The thesis closes with a summary and an outlook on possible future improvements (chapter 7).

2 Neutrino physics

The objective of this chapter is to put the direct neutrino mass experiment KATRIN into the context of several decades of neutrino physics. It took about three decades to experimentally prove the existence of the neutrino, after its postulation by W. Pauli in 1930, with the famous Poltergeist experiment (section 2.1.1). There, in the year of 1956, the exciting journey of experimental neutrino physics started, revealing two more neutrino flavors up till now. In the Standard Model of Particle Physics (SM) (section 2.1.2), the neutrino is massless. However, experiments around the turn of the millennium discovered neutrino oscillations and therefore established, that neutrinos are not massless. This amplified the efforts of neutrino mass measurements (section 2.3), which up to today only constrain the neutrino mass.

2.1 History of the neutrino

The neutrino was introduced the first time in 1930 by Wolfgang Pauli. He added a third particle to the products of the β-decay to explain the continuous energy spectrum and therefore save energy and momentum conservation

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z+1}Y + e^{-} + \bar{\nu}_e. \tag{2.1} $$

The electron can now share the constant decay energy with this particle and therefore has a continuous energy spectrum. To fulfill charge and spin conservation, the particle needs to have charge 0 and spin 1/2. Therefore Pauli named the particle neutron.
However, two years later the neutron was discovered by Chadwick [Cha32]. It was E. Fermi, who picked up the idea of Pauli and came up with a first theoretical description of the β-decay, naming the neutral particle "neutrino".

2.1.1 Discovery

With the invention of nuclear fission reactors, the first strong (anti-) neutrino source was developed. Cowan and Reines utilized this occasion and built a detector close to a fission reactor in Los Alamos, after reconsidering their first idea to use a nuclear fission bomb as a source [RC97]. To detect the incoming antineutrinos, they used alternating layers of water as proton donators and liquid scintillator material. The protons of hydrogen nuclei in H_2O can react with the antineutrinos in an inverse β-decay into a positron and neutron

$$ \bar{\nu}_e + p^{+} \rightarrow n + e^{+}. \tag{2.2} $$

The positron can then annihilate with any available electron in water quasi instantly into a pair of photons, each with an energy of the electron's rest mass (511 keV, prompt

signal), the neutron gets moderated in the water. Additionally, $CdCl_2$ is solved in the water. This cadmium then captures the neutron and cools down to its ground state via photon emission (delayed signal). The produced photons can lose energy due to Compton scattering and ultimately being absorbed by the scintillator material at UV energies. The excited scintillator molecule then returns to its ground state under the emission of photons in the visible region. These photons can then be detected by the mounted Photo Multiplier Tubes (PMTs). The characteristic time difference between the prompt and delayed signal is then evidence for the inverse β-decay (equation 2.2). In 1956, Cowan and Reines published their results and claimed the detection of antineutrinos from inverse β-decays [RC53].
Several years later, in 1962, the muon neutrino was found as a second type by Schwartz, Lederman and Steinberger at Brookhaven National Laboratory (BNL). They used pions from a nearby accelerator, which decay in flight, due to their small lifetime, into a pair of muon and muon neutrino

$$\pi^{\pm} \longrightarrow \mu^{\pm} + \overset{(-)}{\nu}_{\mu}. \tag{2.3}$$

These muons, as well as other remaining particles, are then absorbed by a thick iron shielding. The only particles passing through this layer are neutrinos. In the detector these neutrinos created exclusively muons and no electrons, which means that the muon neutrino is different from the electron neutrino [DG62].
Finally, in the year 2001, the last neutrino flavor ν_τ in the Standard Model was detected by the DONUT collaboration. Unstable D_S mesons, produced by an 800 GeV proton beam dumping on a tungsten target, decay purely leptonical into a tauon tau-antineutrino pair.

$$D_S \longrightarrow \tau + \bar{\nu}_\tau \tag{2.4}$$

The ν_τ then can pass through a shield for all other particles and hit the detector. There they can produce tauons, which are detected via their characteristic kinematic decay signal, a kink [Kod01].

2.1.2 Standard Model of Particle Physics

The SM describes all known modules of matter and their interactions (besides gravity). The gauge group the SM is based on, is given by an $SU(3)_C \times SU(2)_L \times U(1)_Y$. Hereby, every symmetry comes along with a conservation law. In Quantum Chromo Dynamics (QCD), described by the $SU(3)_C$, the conserved observable is the color charge. Its force carriers are eight massless gauge bosons, the gluons. They can couple exclusively to the quarks and among themselves, because only they carry color charge. The remaining four gauge bosons (γ,Z,$W^{\pm}$) are the force carriers of the electro-weak theory, with the gauge symmetry $SU(2)_L \times U(1)_Y$. This theory was introduced independently by Salam [Sal68] and Weinberg [Wei67], to combine the weak interaction $SU(2)$ with the electro-magnetic interaction $U(1)$. The force carriers of the weak interaction are the gauge bosons Z, and $W^{\pm}$, which couple to the weak isospin. The force carrier of the electro-magnetic interaction is the γ, which couples to the electric charge. The neutrino is standing out in the SM among the other particles: it is the only particle which carries neither color nor electric charge

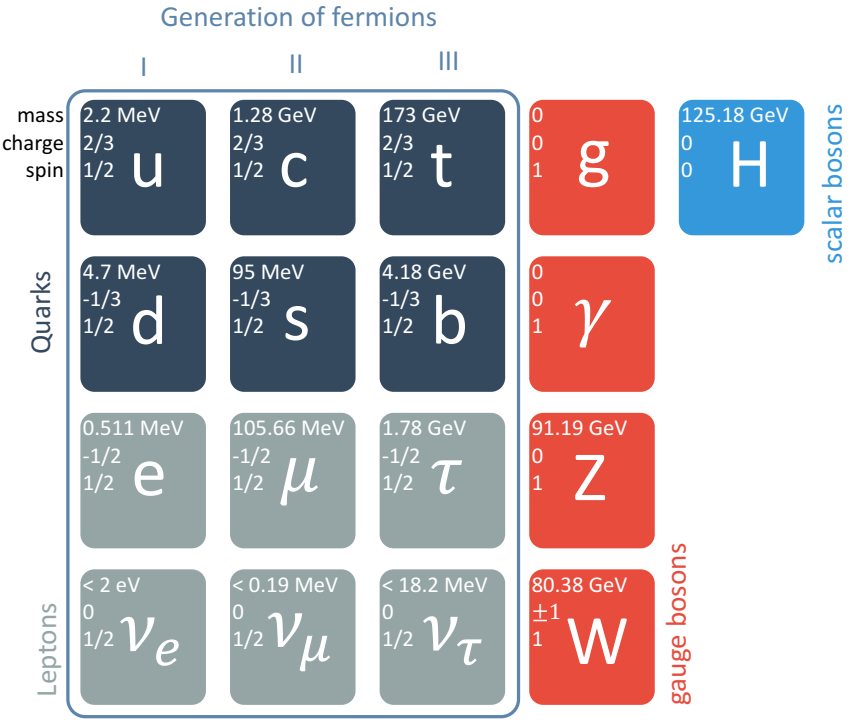


Figure 2.1.: **The Standard Model of Particle Physics (SM)** containing the three generations of matter particles, the gauge bosons, and the Higgs boson. Their masses are given, or in case of the neutrinos, their upper limit [TH18].

and therefore interacts exclusively via the weak interaction. This makes their detection a challenging effort.

2.2 Neutrino oscillations

Although the SM has performed quite well, it does have some open issues such as the impossibility of describing gravity. But also, the neutrino is massless in the SM. However, recent experiments measuring the solar neutrino flux [Aha13] and atmospheric neutrino flux [Wen10] have proven that neutrinos have a fascinating property, namely oscillations. A similar feature is shown by the quarks, where the three flavor generations mix via the interaction with a W boson. This leads to a mixing matrix called Cabibbo-Kobayashi-Maskawa (CKM) matrix. Analogously, in neutrino physics there is the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, which describes the flavor mixing. For the experimental evidence of neutrino oscillations, Arthur B. McDonald and Takaaki Kajita received the Nobel Price in physics in 2015.

2.2.1 Theory

The possibility of neutrino flavor mixing was first described by Pontecorvo [Pon58], Maki, Nakagaya and Sakata [MNS62]. They came up with a theory, in which the neutrinos are created or destroyed in their flavor eigenstates, but propagate through space and time in

their mass eigenstates. Therefore, the flavor eigenstates can be described as a superposition of mass eigenstates ν

$$\nu_j = \sum_i C_{ij}\nu_i, \qquad j = e, \mu, \tau \tag{2.5}$$

where $C_{i,j}$ is the PMNS matrix, which can be factorized into

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_D} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta_D} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\delta_{M1}} & 0 \\ 0 & 0 & e^{i\delta_{M2}} \end{pmatrix}, \tag{2.6}$$

where $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$ [TH18]. This matrix features three mixing angles θ_{ij}, one Dirac phase δ_D and two Majorana phases δ_M, which can cause CP violation and are important for studying double-beta decay experiments.
While propagating, the neutrino flavor eigenstates become a superposition, due to the different masses of the mass eigenstates, which results in different oscillation frequencies. For reasons of simplicity, the phenomenon will be described in a 2×2 space, and not in 3×3. When considering the oscillation between ν_μ and ν_τ, with one mixing angle θ, and two mass eigenstates ν_2, ν_3 the equation becomes

$$\begin{pmatrix} \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \nu_2 \\ \nu_3 \end{pmatrix}. \tag{2.7}$$

By using the ansatz for the mass eigenstates $|\nu_i\rangle = |\nu_i(0)\rangle \exp(-i\omega_i t)$ in natural units $\hbar = c = 1 \rightarrow \omega = E$, and using the constraint, that the masses are much smaller than the energy

$$E_i = p + \frac{m_i^2}{2p}, \tag{2.8}$$

the probability of ν_μ staying ν_μ becomes

$$P(\nu_\mu \rightarrow \nu_\mu) = 1 - \sin^2 2\theta \cdot \sin^2\left(1.27\frac{\Delta m^2 L}{E}\right). \tag{2.9}$$

Here, L is the oscillation length in km, $\Delta m^2 = m_3^2 - m_2^2$ the difference of the squared masses in $(\mathrm{eV/c^2})^2$ and E the energy in GeV. The numerical factor 1.27 arises from the transformation back to SI units, $1.27 = \frac{1}{4\hbar c}$. In figure 2.2, equation 2.9 is visualized at the top and the oscillation of the mass eigenstates at the bottom.

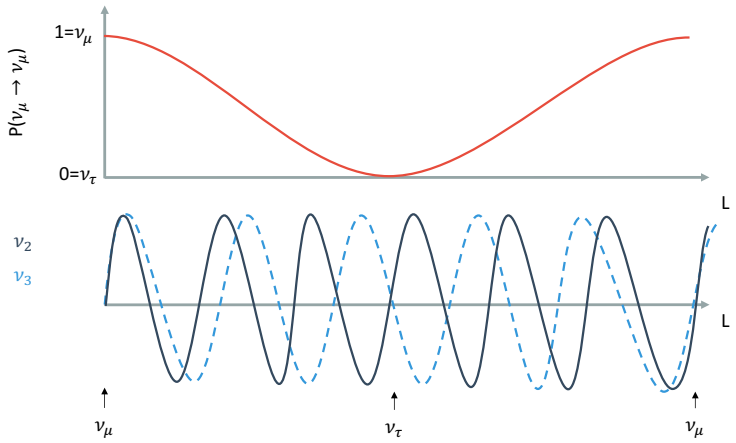


Figure 2.2.: **Neutrino oscillation with two flavors.** Here the neutrino mixing angle is assumed to be maximal, $\theta = 45°$. In the top, the probability in the disappearance channel $\nu_\mu \to \nu_\mu$ is shown, on the bottom the oscillation of the different mass eigenstates $\nu_{2,3}$.

2.2.2 Discovery

To experimentally access the three different mixing angles θ_{ij}, and mass differences Δm^2_{ij}, different neutrino sources have to be used:

- θ_{12}: Solar neutrinos
- θ_{13}: Reactor/accelerator neutrinos
- θ_{23}: Atmospheric/accelerator neutrinos

In general, the experiments set up to study these different sources either make use of a variable source-target length (atmospheric reactor experiments) or scan the energy dependent disappearance of (muon) neutrinos (accelerator). In table 2.1, recent results of neutrino oscillation parameter fits can be found.

Solar neutrino experiments

The results from radio-chemical experiments like the Homestake experiment [CD98] and GALLEX [AH93] provided first hints for solar neutrino oscillations. They measured a deficit in the solar neutrino flux, when comparing measurement results to the predicted flux by Bahcall's Standard Solar Model (SSM) [Bah64]. However, the evidence that this neutrino deficit comes from flavor oscillations was provided by the Sudbury Neutrino Observatory (SNO) experiment. They used 1000 tons of heavy water (D_2O) to detect interactions of solar neutrinos with D_2O via Cherenkov radiation.

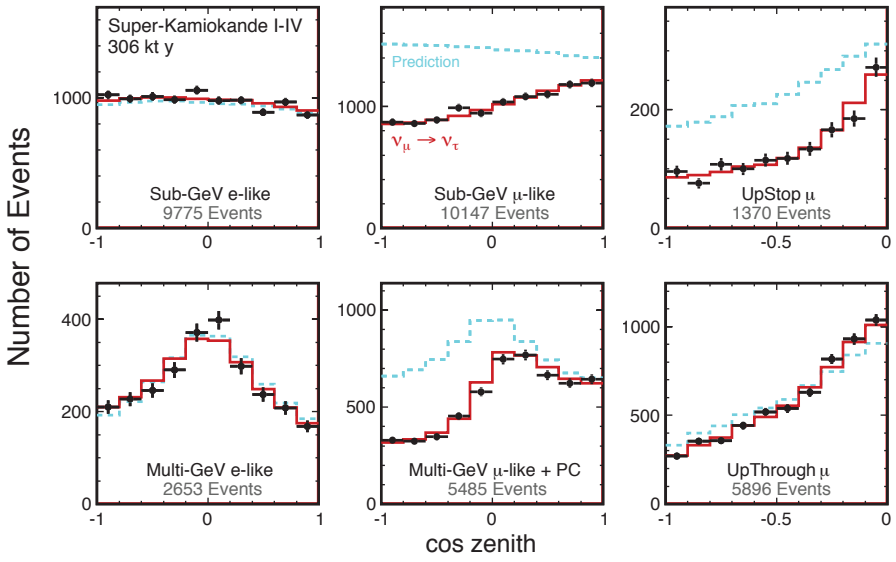


Figure 2.3.: **Distribution of e and μ-like events in SuperK in dependence of the zenith angle** with visible energy <1.33 GeV (Sub-GeV) and >1.33 GeV (Multi-GeV). The dotted lines represents predictions from MC simulations for the no-oscillation case, the black dots the measured data and the red line the best fit expectation for $\nu_\mu \leftrightarrow \nu_\tau$ oscillations. Figure taken from [TH18].

The main difference to radio-chemical experiments like the Homestake experiment, which are sensitive to Charged-Current (CC) interactions exclusively, is the additional sensitivity to Neutral-Current (NC) reactions and elastic scattering (ES)

$$\nu_e + d \rightarrow p + p + e^- \quad \text{(CC)}, \tag{2.10}$$
$$\nu_x + d \rightarrow p + n + \nu_x \quad \text{(NC)}, \tag{2.11}$$
$$\nu_x + e^- \rightarrow \nu_x + e^- \quad \text{(ES)}. \tag{2.12}$$

With these combined reactions, the SNO collaboration could solve the solar neutrino deficit and provide evidence that ν_e as a product of the fusion reactions in the sun do oscillate into ν_μ and ν_τ inside the sun (MSW effect [MS85]).

Atmospheric neutrino experiments

The first compelling evidence of atmospheric neutrino oscillations was given by the Super-Kamiokande experiment in 1998 [FH98]. They used a detector filled with 50-ktons of water to detect electron and muon neutrinos. These neutrinos can scatter with the water-molecules and produce high energy electrons or muons, whose Cherenkov radiation is detected by PMTs. They reported a significant deficit of μ-like events compared to the no-oscillation expectation, in dependency of the zenith angle, see figure 2.3. These results could later be confirmed by other atmospheric neutrino experiments like MACRO [The04].

Reactor/Accelerator neutrino experiments

Finally, man-made sources of neutrinos can be used to study their oscillation behavior. In accelerators, proton-proton collisions produce a large number of charged pions, which decay into muons (equation 2.3). A detector close to the beam dump allows for calibration, whereas the main detector is usually several kilometers away from the target. Because the baseline length is usually fixed in accelerator experiments, the energy dependent disappearance of (muon) neutrinos is utilized to study the oscillation parameters θ_{ij} and Δm^2_{ij}.
In comparison to accelerators, nuclear fission reactors do not only provide a strong source of neutrinos, but also exist more frequently. An example to mention here is the KamLAND experiment in Japan, which by the time of data-taking was surrounded by 53 commercial nuclear fission reactors within different distances to the detector. However, as the energy of neutrinos produced in nuclear fission reactions is in the few MeV range, these detectors exclusively look for oscillations in the $\bar{\nu}_e$ disappearance channel.
Instead of using different reactors, more recent experiments, such as Daya Bay, make use of an combination of near and far detectors, in order to deliver more stringent limits on the mixing angle θ_{13}. This mixing angle is much smaller (table 2.1) than the other two mixing angles and plays a crucial role in determining whether there is CP violation in the lepton sector and the resolution of the mass hierarchy [CL16].

Neutrino mass ordering

While decades of neutrino oscillation experiments shed light into the squared mass differences, the absolute mass scale, and the sign of Δm^2_{13} is still unknown. This allows for two different neutrino spectra. The spectrum with Normal Ordering (NO), where $m_1 < m_2 < m_3$ and the spectrum with Inverted Ordering (IO), where $m_3 < m_1 < m_2$, which are illustrated in figure 2.4.

To investigate the absolute scale of the neutrino masses, a different approach to neutrino oscillation measurements is needed. These experiments, called neutrino mass experiments, will be introduced in the following section.

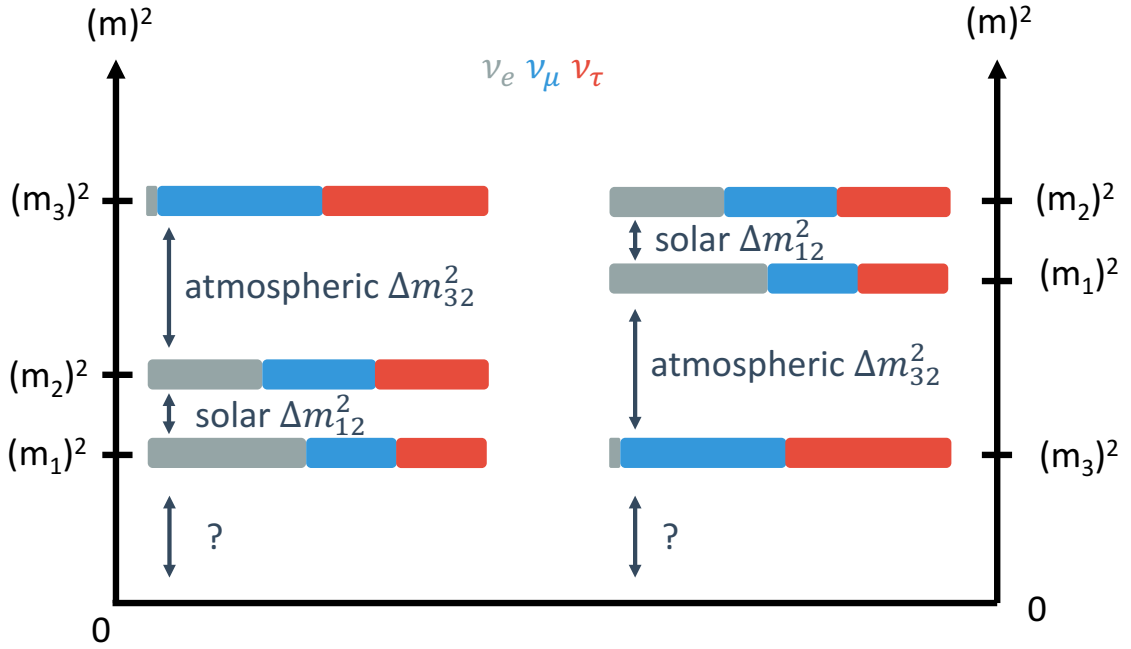


Figure 2.4.: **Two different neutrino mass ordering scenarios.** The left part shows the NO, where sign(Δm^2_{13})=−1 and the right part the IO, where sign(Δm^2_{13})=+1

Table 2.1.: **Recent results of global neutrino oscillation parameter fits** from [TH18]. The data listed here assumes normal ordered mass hierarchy. For inverted hierarchy parameters as well as the CP-violating phase, see [TH18].

Parameter	best fit	3σ
Δm^2_{12} [10^{-5} eV^2]	7.37	6.93 - 7.96
$\Delta m^2_{31(23)}$ [10^{-3} eV^2]	2.56 (2.54)	2.45 - 2.69 (2.42 - 2.66)
$\sin^2\theta_{12}$	0.297	0.250 - 0.354
$\sin^2\theta_{23}$	0.425	0.381 - 0.615
$\sin^2\theta_{13}$	0.0215	0.0190 - 0.0240

2.3 Neutrino mass experiments and constraints

The experiments described in this section focus on solving two of the most urgent questions in neutrino physics: Is the neutrino its own antiparticle and what is the absolute scale of the mass ordering? The most prominent theory of neutrino mass generation will be described briefly in the following.
The see-saw mechanism (type I) provides a natural theory to explain the smallness of neutrino masses, and therefore a smaller coupling to the Higgs field, compared to lepton and quark masses. A Right-Handed (RH) neutrino singlet is introduced, which has a Majorana mass term and couples to the SM lepton and Higgs doublets via Yukawa type coupling. After spontaneous symmetry breaking of the electro-weak scale, the mass term can be written as see e.g. [Kla18]

$$\mathcal{L} = -\frac{1}{2}(\bar{\nu}_L \bar{\nu}_R^c) M (\nu_L^c \nu_R)^T + \text{h.c.}, \tag{2.13}$$

where superscripts R, L denote RH and Left-Handed (LH) chirality, and c the charge conjugate.

$$M = \begin{pmatrix} m_M^L & m_D \\ m_D & m_M^R \end{pmatrix} \tag{2.14}$$

is the mass matrix, where D, M denotes the Dirac and Majorana mass respectively. Assuming the mass hierarchy needed for the see-saw mechanism $M_M = m_M^R >> m_D > m_M^L \approx 0$, the mass matrix reduces to [SV80]

$$M = \begin{pmatrix} 0 & m_D \\ m_D & M_M \end{pmatrix}. \tag{2.15}$$

The corresponding eigenvalues analysis to get from equation 2.14 to 2.15 will give the mass eigenstates with masses

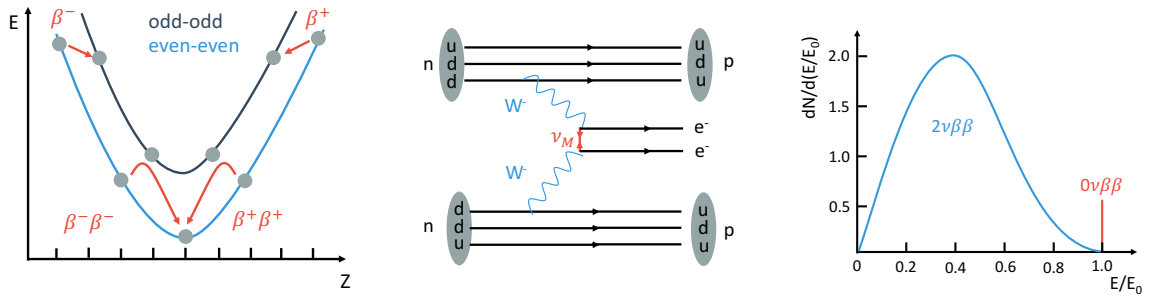
$$m_1 = \frac{m_D^2}{M_M} \tag{2.16}$$

$$m_2 = M_M, \tag{2.17}$$

where m_1 is a very light eigenstate and m_2 a heavy one.
The naming convention "see-saw" becomes visible for the eigenvalue $\lambda = 0$ by

$$m_M^L m_M^R = m_D^2, \tag{2.18}$$

which means for a fixed value of m_D, a rise in m_M^R corresponds to a lowering of m_M^L.
In order to get a better understanding of the neutrino mass generation process, it is necessary to experimentally confirm the Majorana character of the neutrino.



(a) (b) (c)

Figure 2.5.: **Double beta-decay properties.** a): Mass excess E as a function of proton number Z. This sketch shows that in some rare cases single β-decay is energetically forbidden. b): Feynman graph of $0\nu\beta\beta$-decay. c): Sketch of energy spectra of $2\nu\beta\beta$-decay (blue) and $0\nu\beta\beta$-decay in magnified size (red).

2.3.1 Cosmology

Since large amounts of neutrinos originate from their freeze out[1] from thermal equilibrium shortly (10^{-9} s) after the Big Bang, they had a major impact on structure formation on large scales. While travelling the universe, the initial relic neutrinos with a temperature of ≈ 1 MeV can smear out small perturbations of matter density by carrying matter with them. The current model dependent upper limits on the sum of neutrino masses from experiments measuring the matter distribution in the universe, is [TH18]

$$\sum_i m_{\nu,i} < 0.12 - 0.73\,\text{eV}. \tag{2.19}$$

For more detailed information, the reader is referred to [Wei08].

2.3.2 Neutrinoless double β-decay experiments

The information presented in this section was gathered with the help of [Per09]. If the neutrino is a Majorana particle, the lepton number L is violated, since the neutrino ($L = 1$) and its antiparticle, the antineutrino ($L = -1$), are identical. So in order to prove the Majorana nature of the neutrino, lepton number violating processes need to be studied. Such a process is the so-called neutrinoless double beta decay ($0\nu\beta\beta$). From single β-decay, it is known that the electron is accompanied by an antineutrino (equation 2.1). If two of these processes occur simultaneously ($2\nu\beta\beta$), the decaying nucleus will transform according to

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- + 2\bar{\nu}_e. \tag{2.20}$$

The probability of these second order weak processes scales with G_F^4, is therefore very rare and occurs only for even-even nuclei numbers (figure 2.5a). In addition, two more conditions need to be fulfilled: The single β-decay must be energetically forbidden, which

[1] Expansion of the universe Γ_H is faster than the collision rate Γ. Detailed information in [GG91].

is usually the case due to the odd-even effect of the mass. Furthermore, energy conservation requires $M(Z,A) > M(Z+2,A) + 2m_e$.
A $0\nu\beta\beta$ (figure 2.5b) can be thought of as a two staged process: First, an antineutrino is emitted due to β-decay. Second, the RH antineutrino switches to a LH neutrino and initiates a second decay with the daughter nucleus $(Z+1,A) + \nu_e \rightarrow (Z+2,A) + e^-$

$$(Z,A) \rightarrow (Z+2,A) + 2e^-. \tag{2.21}$$

For neutrinos having mass, the probability of an antineutrino being emitted as a neutrino is given by $\approx \frac{(m_\nu c^2)^2}{2E^2}$. Therefore the $0\nu\beta\beta$-decay is even unlikelier than the $2\nu\beta\beta$-decay and theoretically shows up as a discrete line in the summed energy spectrum of the electrons. However, detectors with high fiducial mass/volume are commonly used, whose energy resolution smear out the discrete line (figure 2.5c). Up to now, no experiment was able to observe $0\nu\beta\beta$-decay. The keys to success of experiments searching for $0\nu\beta\beta$-decay are a large mass M, long exposure time t and a high energy resolution combined with a low background. The GERmanium Detector Array (GERDA), located at Gran Sasso underground laboratory (LNGS), Italy, searches for $0\nu\beta\beta$-decay of ^{76}Ge using germanium detectors enriched with a fraction of ^{76}Ge, operated in a liquid argon bath for background suppression. The germanium array serve as both source and detector. Another experiment, also utilizing ^{76}Ge, is the MAJORANA experiment. Together these two experiments plan to collaborate for a future tonne scale ^{76}Ge $0\nu\beta\beta$ search [AB18a]. The current upper limit of the lifetime is set to $t_{1/2} > 10^{25}$ years, which corresponds to an effective weighted sum of Majorana neutrino masses of

$$\langle m_\nu \rangle = \sum_i \left| U_{ei}^2 m_{\nu i} \right| < 0.11 - 0.52\,\text{eV}. \tag{2.22}$$

These limits are taken from [TH18].

2.3.3 Single β-decay experiments

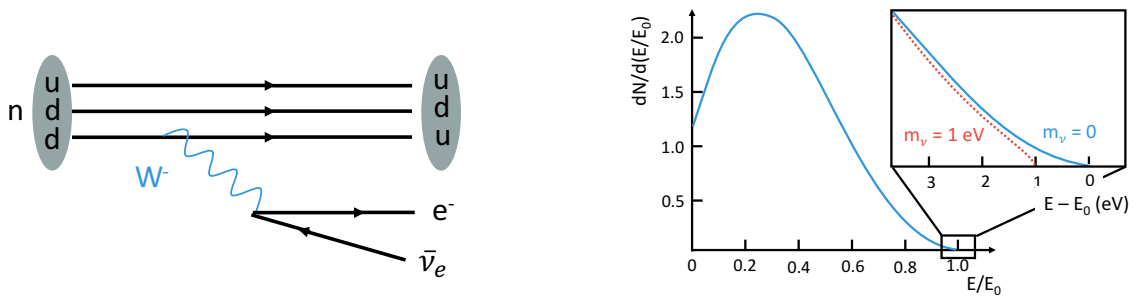
A direct way to measure the neutrino mass is by precisely measuring the kinematics of the single β-decay (figure 2.6a)

$$(Z,A) \rightarrow (Z+1,A) + e^- + \bar{\nu}_e. \tag{2.23}$$

A non vanishing neutrino mass affects the shape of the electron's energy spectrum, especially leading to a fine splitting near the endpoint of the spectrum for different neutrino masses (figure 2.6b). The differential spectrum close to the endpoint is given by

$$\frac{\mathrm{d}N}{\mathrm{d}E} = C \cdot F(E,Z) \cdot p_e \cdot (E_e + m_e c^2) \cdot (E_0 - E_e) \cdot \sqrt{(E_0 - E)^2 - \sum_i |U_{ei}|^2 m_{\nu_i}^2}, \tag{2.24}$$

where p_e, E_e, m_e denote the momentum, kinetic energy, and mass of the emitted electron. $F(E,Z)$ is the Fermi function, which takes into account the Coulomb interaction of the outgoing electron with the daughter nucleus, and E_0 is the endpoint energy for a vanishing neutrino mass.



(a) (b)

Figure 2.6.: **Single beta decay properties.** a): Feynman graph of single beta decay. b): Sketch of an energy spectrum of electrons, emitted by single beta-decay. The zoomed in figure shows the difference between a zero neutrino mass spectrum (blue) and a non-vanishing neutrino mass spectrum (red).

Subsequently, $(E_0 - E)$ is the total energy of the neutrino and $\sum_i |U_{ei}|^2 m_{\nu_i}^2 = \langle m_\beta \rangle^2$ gives the weighted average of all neutrino flavors and is called the effective electron antineutrino mass. This theoretically would lead to tiny distortions near the endpoint of the energy spectrum, however due to the small mass splittings $O\left(10^{-3}\,\mathrm{eV}^2\right)$, state of the art energy resolutions of single β-decay experiments are not able to resolve these distortions. The constant C is given by

$$C = \frac{G_F^2}{2\pi^3} \cos^2\theta_C |M|^2, \tag{2.25}$$

where G_F is the Fermi constant, θ_C the Cabbibo angle describing the strength of the transition from a down to an up quark (neutron to proton), and $|M|$ the nuclear matrix element.
In order to achieve a high sensitivity on the neutrino mass, single β-decay experiments need to gather high statistics in the endpoint region, which usually requires nuclei with a short life-time, as well as a low Q value[1]. Furthermore, the energy resolution needs to be sensitive to small distortions near the endpoint of the spectrum and the experiments need an overall high signal-to-background ratio, meaning well studied background sources as well as methods to suppress them.
Most stringent upper limits on the neutrino mass come from experiments utilizing tritium β-decay. Tritium has not only the third lowest endpoint energy (after holmium and rhenium) with $E_0 \approx 18.6\,\mathrm{eV}$, but also a rather short half-life of $t_{1/2} \approx 12.33 \pm 0.03$ years [LU00]. This is the key feature for a high-statistics measurement near the endpoint. The combined analysis of KATRIN's predecessor experiments in Mainz and Troitsk provide the most sensitive upper limit by now on the effective mass of the electron antineutrino [TH18]

$$m_{\bar{\nu}_e} < 2\,\mathrm{eV}\ (95\,\%\,\mathrm{C.L.}). \tag{2.26}$$

[1] amount of energy released in the decay.

The KATRIN experiment has the ambitious goal to improve this limit by one order of magnitude. To achieve this sensitivity, KATRIN uses the experience and technology developed by these experiments such as the successfully implementation of a Windowless Gaseous Tritium Source (WGTS) and the MAC-E filter principle. However, KATRIN still faces a lot of challenges, especially concerning systematics. Current issues are the the Final States Distribution (FSD) of the daughter nuclei, plasma effects of the electrons in the WGTS, and the magnetic field in the analyzing plane.

A next generation direct neutrino mass experiment named Project 8, tries to overcome the issues with the FSD by using atomic tritium combined with the cyclotron radiation emission spectroscopy technique (CRES) to push the limits of neutrino mass observation down to $m_{\bar{\nu}_e} \lesssim 40\,\mathrm{meV}$ [Ash17].

An isotope which is also regarded as a promising candidate for neutrino mass determination is ^{163}Ho. Experiments such as ECHo measure the energy spectrum emitted by electron capture of holmium, to access the electron neutrino mass in the sub-eV range. The working principle of a small scale detector has already been demonstrated and the experiment currently aims for a sub 10 eV (ECHo1k) range. In the future the detector will be scaled up to reach sub-eV level (ECHo1M) [GB17].

3 The KATRIN experiment - towards a high precision ν mass measurement

The KATRIN collaboration was founded in 2002 with the goal, to be a next generation neutrino mass experiment and increase the sensitivity by one order of magnitude compared to its predecessors in Mainz and Troitsk. To achieve this goal, a high luminosity source is required to get high statistics in the endpoint region, which is why the location to build this experiment was chosen to be Karlsruhe. Here the Tritium Laboratory Karlsruhe (TLK) has proven experience in tritium handling, as well as license for the required amount of this radioactive hydrogen isotope. However, not only the amount of source material has to be increased, but also a more sensitive MAC-E filter system needs to be deployed. This chapter will give a short overview over the different components of the KATRIN beamline and explain the systems which are of greater importance in this thesis more thoroughly.

3.1 Overview and measurement principle

The MAC-E filter is the key feature of KATRIN. With its high energy resolution and permission of large angular acceptance angles, it is prepared to precisely measure the energy of the β-electrons close to the spectrum's endpoint. The principle was first described in [BPT80] and later adapted by the experiments in Mainz and Troitsk [KB05, Ase12].

Strong solenoids produce magnetic fields in a spectrometer, to adiabatically guide the β-electrons from their origin in the source to the detector. In the source, the electrons are emitted isotropically under a polar angle θ to the magnetic field lines. The electron's momentum can be split into a transverse component $p_\perp = p \cdot \sin\theta$, which is responsible for the cyclotron motion around the magnetic field lines due to the Lorentz force, and a longitudinal component $p_\parallel = p \cdot \cos\theta$, which is responsible for the guidance of the electrons along the magnetic field lines in the beamline. In the same way the kinetic energy can be split into a component parallel to the magnetic fields $E_\parallel$ and a perpendicular component $E_\perp$.

As illustrated in figure 3.1b, the electron's momentum gets transformed to its longitudinal component towards the center of the spectrometer, due to the adiabatic change in the magnetic field. By providing electric fields parallel to the longitudinal momentum of the electrons, they get decelerated towards the center of the spectrometer and after passing the center, accelerated again towards the exit (bottom of figure 3.1b). The electrostatic barrier generated by the electrodes therefore allows high precision filtering of the electron's

energy, as electrons with the kinetic energy

$$E_{\parallel} < \|qU_0\| \tag{3.1}$$

get rejected by the potential.

To first order approximation, the change in the magnetic field can be considered as an adiabatic process with a constant magnetic moment μ in the non-relativistic approximation ($\gamma = 1.04$ for 18.6 keV electrons):

$$\mu = \frac{E_{\perp}}{B} = \text{const.} \tag{3.2}$$

To guarantee an adiabatic change, the ratio of the maximum magnetic field strength B_{max} to its minimum B_{min} must be sufficiently large and therefore limits the minimum length of the spectrometer.
The minimum diameter is limited by the conservation of the magnetic flux

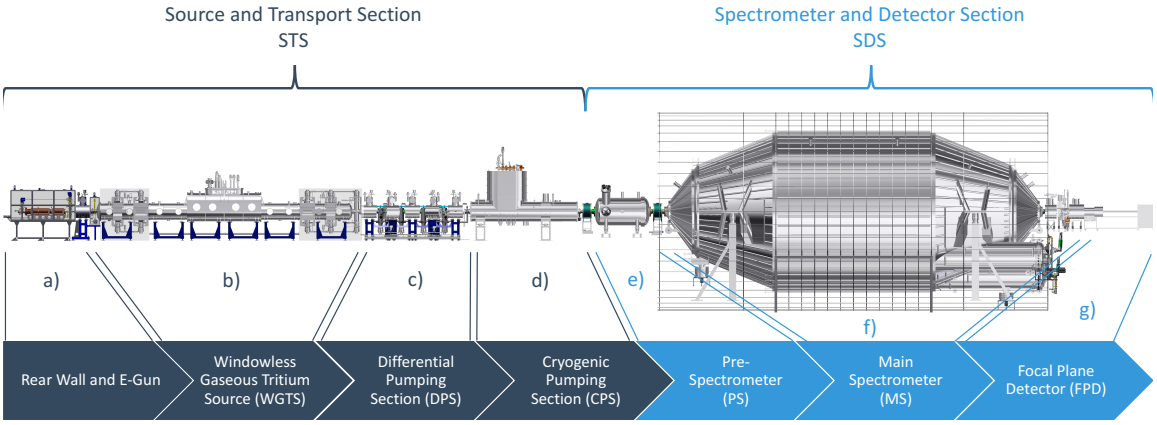
$$\Phi = \iint_A \vec{B} \mathrm{d}\vec{A} = \text{const.} \tag{3.3}$$

A decrease in the magnetic field will therefore cause an increase in area and consequently an increase in radius. The magnetic field has its minimum in the center (analyzing plane) of the spectrometer, with a diameter of

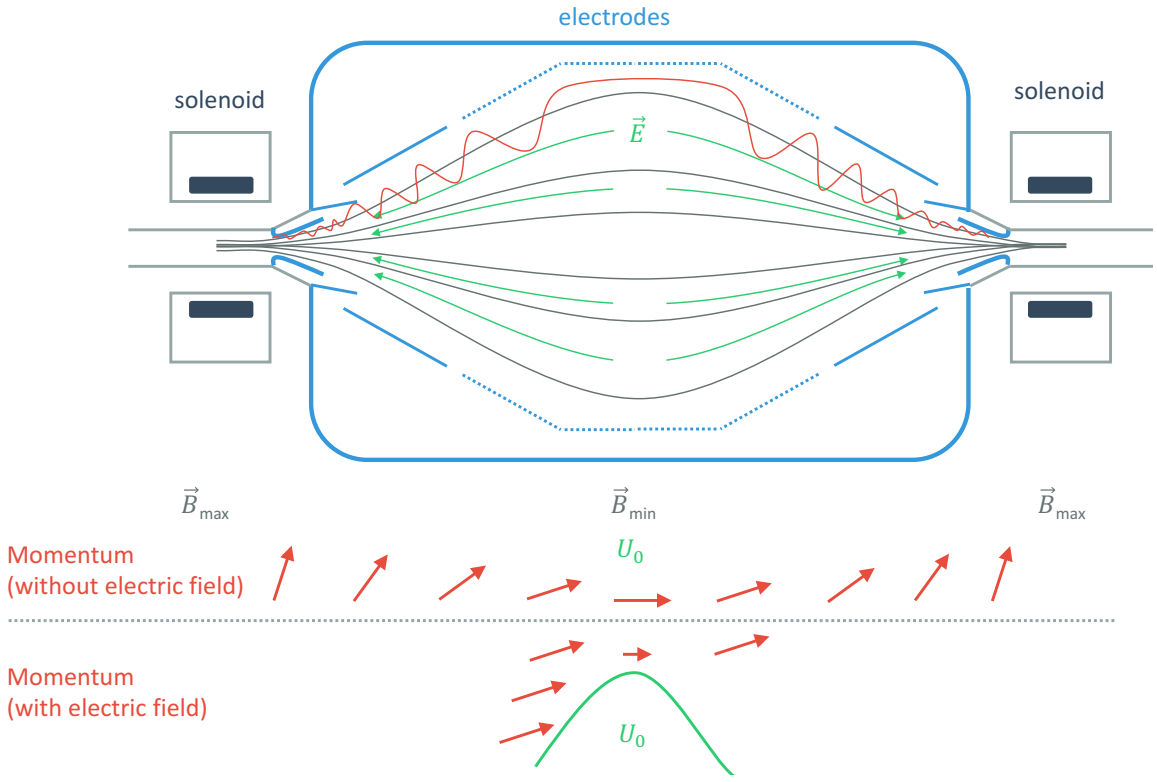
$$d_{\text{AP}} = d_{\text{S}} \sqrt{\frac{B_{\text{S}}}{B_{\text{min}}}}, \tag{3.4}$$

where d_{S} is the diameter and B_{S} the magnetic field in the source.

In figure 3.1a an overview of the whole KATRIN beamline can be found. In the following, the single components will be explained in order following the trajectory of the β-electrons from the source to the detector.



(a)



(b)

Figure 3.1.: **KATRIN overview a) and MAC-E filter principle b).** The adiabatic change in the $\vec{B}$-field (grey) from maximum to minimum causes a transformation from transverse to longitudinal momentum of the guided electrons (red). In the middle of the analyzing plane, nearly all of the electrons energy is in $E_{\parallel}$. In the presence of electric fields $\vec{E}$ (green), which are provided by the electrodes (blue), the electron therefore gets decelerated towards the center, allowing precise energy filtering of the electrons. Modified from [Wan13].

3.2 Source and Transport Section (STS)

The Source and Transport Section (STS) is responsible for providing a large β-electron flux through a high density tritium profile in the WGTS. As signal electrons are guided magnetically from the source section to the detector, the beamline is open and can only be closed at several points by valves. To prevent tritium from entering the spectrometer section, two pumping sections (Differential Pumping Section (DPS), Cryogenic Pumping Section (CPS)) are used to reduce the gas flow by 14 orders of magnitude. The pumped out tritium is then collected and re-injected into the source and thus provides a closed gas circuit.

3.2.1 Windowless Gaseous Tritium Source (WGTS)

In the center of the beam tube of the WGTS, molecular tritium with a pressure of about 10^{-3} mbar gets injected via a capillary. The molecules then travel along the beam tube and partially get pumped out at both ends of the WGTS. The isotopes get separated and the tritium is fed back in the inner loop system. The composition of hydrogen isotopes is monitored continuously by a LAser RAman spectroscopy setup (LARA) [Sch13]. The temperature is kept at 30 K to minimize distortions in the signal electrons energy due to thermal Doppler-broadening and to prohibit the accumulations of tritium clusters. Measurements during the First Tritium Campaign (FT) have shown that the relative temperature stability per run of $(6.4 \pm 0.5)\,10^{-5}$ exceeds the requirement of $2 \cdot 10^{-3}$ [KAT05] by nearly two orders of magnitude [Hei18].
To reduce possible plasma effects and potential misalignment of the beamline, only the inner part of $d_S = 8.2$ cm of the WGTS's beam tube diameter ($d = 9$ cm)is used for neutrino mass analysis. Furthermore, the solenoids need to be operated on lower currents to prevent them from quenching [AB18c]. This results into a lower magnetic field strength in the source of $B_S = 2.52$ T and a magnetic flux of

$$\Phi = 133.7\,\mathrm{Tcm}^2 \tag{3.5}$$

which has to be guided through the beamline to the detector. The reader may find more information in [Höt12, Kuc16, HS17, AB18b].

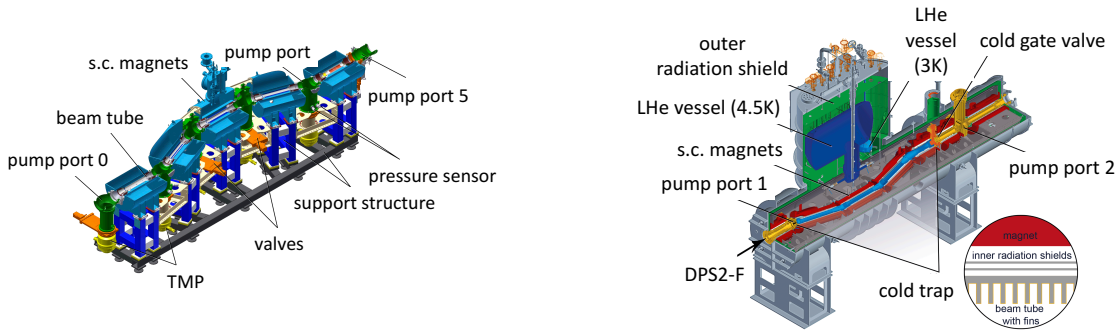
Rear wall and E-gun

At the upstream end of the beamline, the rear section monitors

- the source activity by β-induced X-ray spectroscopy (BIXS) [Röl15],
- the source density via inelastic scattering of electrons in the source with the mounted Electron Gun (E-Gun) [Beh16].

This E-Gun is also used to measure the inelastic cross section of β-electrons scattering off hydrogen isotopes and the corresponding energy loss function[1]. They both cause systematic uncertainties and have to be precisely known for an optimized neutrino mass measurement [Hei15, Gro15, Tro18].

[1] commissioning in STS-IIIa measurement campaign in September/October 2018.



(a) (b)

Figure 3.2.: **CAD drawings of the transport section's components.** a): DPS beamline. b): CPS beamline. Figures taken from [FR18].

3.2.2 The Transport Section

The transport section (figure 3.2) has the main requirements to guide the signal electrons magnetically through the beamline and reduce the gas flow. This is done by a Differential Pumping Section (DPS figure 3.1a c)) which uses four Turbo Molecular Pumps (TMPs) and the Cryogenic Pumping Section (CPS figure 3.1a d)), which uses an argon frost layer maintained at 3 K to adsorb the tritium molecules. Furthermore the beamline is tilted in both of these sections in order to increase the probability of neutral tritium hitting the beamline walls, which results in a higher pumping probability.

3.2.2.1 Differential Pumping Section (DPS)

A total of four TMPs, installed along the beamline of the DPS together with TMPs installed at the entry of the DPS, reduce the tritium gas flow along the DPS beamline by up to five orders of magnitude. The tilted segments of the DPS beamline furthermore reduce the velocity of the gas molecules and therefore actively help the TMPs to achieve a pumping speed of 2400 l/s each [Jan15]. Superconducting solenoids installed along the beamline guarantee a zero-loss magnetic guidance of the signal electrons through the tilted segments. However, as a large number of electrons are present in the WGTS, ionization of gas molecules causes tritium ions which are guided along the chicane as well. To suppress these ions reaching the spectrometer section, where they could potentially produce a background signal, several devices are installed into the DPS.

- Three dipole electrodes (half-pipe shaped stainless steel electrodes), which causes the ions to be deflected onto the walls of the beamline due to $\vec{E} \times \vec{B}$ drift [Win11, Hac15].

- An electrode with a ring shape is operated on positive voltage to block the ions.

- The Fourier Transform-Ion Cyclotron Resonance (FT-ICR) diagnostic unit aims to identify the ion species by creating a Penning trap and investigate the cyclotron radiation signal [Ubi11].

For more information about the ion production and suppression, the reader is referred to [Kle18].

3.2.2.2 Cryogenic Pumping Section (CPS)

The CPS has similar goals as the DPS: The reduction of tritium gas flow while guiding the signal electrons to the spectrometer section. However, the manner in which the reduction factor of more than 10^7 is achieved, is different to the DPS. Instead of using TMPs, a 3 K frost layer of argon with an area of about 2 m^2 is prepared on the inner surface of the beamline. The beam tube is divided into seven elements, each one being surrounded by superconducting solenoids, which provide the guiding magnetic field lines for the signal electrons. The second and fourth element are tilted by 15 ° compared to the longitudinal spectrometer axis (z) to ensure the tritium molecules hit the wall and are adsorbed there. The clean gold plated inner surface of the stainless steel beam tube allows optimal crystallization conditions for the argon frost, which will be regenerated every 60 days. For further information, the reader is referred to [FR18].

3.3 Spectrometer and Detector Section (SDS)

When leaving the transport section, the signal electrons enter the spectrometer and detector section. This section with its three main devices, the pre-spectrometer (PS), the main spectrometer (MS) and detector, has several goals: First, filter the signal electrons with energies close to the endpoint and reject all electrons below the endpoint, by utilizing MAC-E filter systems. Second, measure the energy of each of the transmitted electrons through these filters. Third, provide excellent Ultra-High Vacuum (UHV) conditions, to prevent the electrons from scattering with residual gas molecules and therefore maintain a low background.

3.3.1 Pre-Spectrometer

In the KATRIN standard neutrino mass measurement mode, the PS has the main goal to reduce the number of electrons entering the MS by rejecting most of the incoming signal electrons. This is done by setting the electrodes on a negative potential of -18.3 kV and therefore rejecting the electrons electrically. However, the PS is much more than a pre-filter. After its commissioning, it was used extensively as a test device to study several techniques, such as the MAC-E filter principle and the UHV system. These studies allowed an optimized design of the main spectrometer. Furthermore, the PS was also used to identify background sources in presence of a MAC-E filter design [Frä10].
The PS's vessel is made of 10 mm thick stainless-steel, has a length of 3.38 m, an inner diameter of 1.68 m, and is electrically isolated from its surrounding devices by ceramic insulators. The vessel can also be set on high voltage, which is a major difference to the spectrometers used in Mainz and Troitsk [Frä10].
To maintain a pressure of 10^{-11} mbar in normal neutrino mass measuring mode, two TMPs are mounted to the pump ports. Furthermore, NEG material is installed into the pump ports with a pumping speed of 27000 l/s. After heating the spectrometer in a so-called bake-out process, which removes most of the thin water-films on the surface, this NEG is

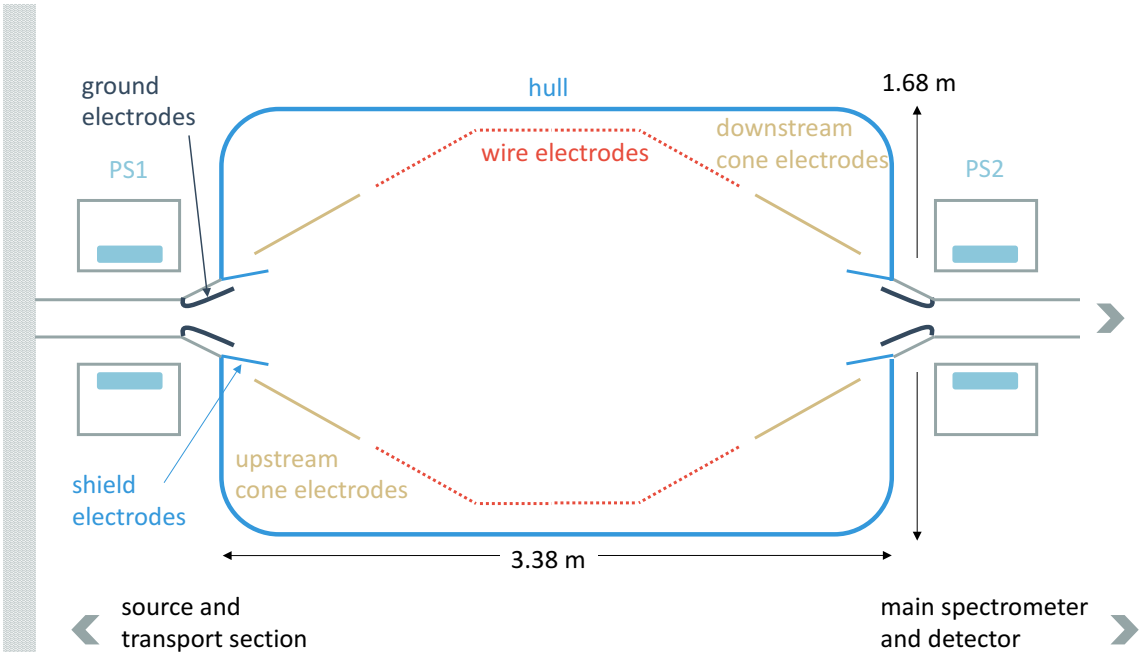


Figure 3.3.: **Schematic drawing of the PS.**

activated. It consists of porous material, which adsorbs the residual gas molecules.
An overview of the PS's inner electrode system (figure 3.3) shows the major components. All electrodes, besides the ground electrode which is fixed to ground potential, can be set on different potentials in order to

- suppress background from the inner surface of the vessel by setting the inner wire electrodes on more negative potential,
- optimize transmission properties by fine-tuning the electrode settings, and
- remove trapped particles via $\vec{E} \times \vec{B}$ drift, caused by the dipole mode of the inner electrodes.

The upstream- and downstream-cone electrodes are made of full metal sheets, whereas the wire electrodes are made of 120 wires each with 0.5 mm diameter. Two superconducting solenoids (Pre-Spectrometer Solenoid 1 (PS1) and Pre-Spectrometer Solenoid 2 (PS2)), with a maximum magnetic field of 4.5 T each, are framing the PS, each with a distance of 2.15 m between the center of the magnets and the PS's analyzing plane. They provide the magnetic guiding field for the signal electrons.

3.3.2 Main Spectrometer

The MS is a bigger, and slightly modified version of the PS. With a diameter of 10 m and a length of 23 m, it is the largest MAC-E filter spectrometer. Framed by two superconducting solenoids (PS2 and PinCH solenoid (PCH)), which provide the guiding magnetic field, the MS reaches an energy resolution of $\Delta E = 0.93$ eV at 18.6 keV with a minimum magnetic field of 0.3 mT in the analyzing plane [KAT05]. For fine-tuning the magnetic field in the analyzing plane, the MS vessel is furthermore surrounded by a Low Field Coil System (LFCS) and an Earth Magnetic field Compensation System (EMCS) [GD13]. For more information about the magnetic field systems deployed in and around the MS, take a look at [Erh16]. To fine-tune the electrostatic potential, two layers of wire electrodes are installed into the MS, which furthermore serve the purpose of electrically rejecting noise electrons originating from the surface of the vessel [Val09, Val11].

3.3.3 Detector and DAQ

The Detector Solenoid (DET) is located 1.6 m away from the PCH towards the downstream side of the beamline. Its magnetic field lines have the purpose to guide the electrons leaving the MS on their final path to the FPD wafer. A final electrode, called Post Acceleration Electrode (PAE) can accelerate the electrons and therefore move their kinetic energies in regions, where the FPD intrinsic background due to γ-radiation is lower [AB15]. Furthermore, it reduces backscattering from signal electrons off the FPD by decreasing their incident angle [Ren11]. During all measurements carried out in this thesis, the PAE was operated on 10 kV.
To count the electrons and read out their information, a segmented PIN diode array is used. While the sensitive front side of the wafer, with a diameter of 90 mm, is unsegmented, the backside is divided into 148 pixels with equal area of 44.1 mm^2. The pixels are arranged in twelve outer rings consisting of twelve pixels each, and an inner ring called the bulls eye, consisting of four pixels. After a transition of the signal from electrical to optical and back to electrical, the analogue signal from up to 24 pixels is digitized via serial Analog-to-Digital Converters (ADCs) and the event's energy and time is read out by Field Programmable Gate Array (FPGA). Both the ADC as well as FPGA are contained in First Level Trigger (FLT) cards, which are connected via a Second Level Trigger (SLT) to the Data Acquisition (DAQ) computer. There the object-oriented and real-time analysis software ORCA can access the FLT and SLT cards. Via the ORCA interface, the user can manually start and stop measurements, set the time for measurements and live-view the measurement. Another powerful feature of ORCA is the possibility to write measurement scripts, in which the user can set e.g. electrode voltages, measurement times etc. It is therefore possible to automate measurements.
Measurements are divided into runs, which itself can be divided into sub-runs After the completion of a run, the data is transferred to the KATRIN Database Management System (DBMS). For further reading, the reader is referred to [Har15] and [AB15].

4 Background contributions

4.1 Overview

Since the statistics of signal electrons in the Region Of Interest (ROI) is very poor, a low background level is key for reaching the sensitivity of 0.2 eV. The background as a free parameter in the neutrino mass fit function has a direct impact on the sensitivity and systematic budget of the neutrino mass. The current background rate of ≈400 mcps in the MS limits the neutrino mass sensitivity of KATRIN to ≈ 0.24 eV. In order to reach the maximum sensitivity of 0.2 eV, a suppression of the current background level by a factor of ≈50 is necessary. Therefore a detailed understanding of the different processes as well as active suppression methods are required. This chapter is intended to give an overview of the major background processes in the Spectrometer and Detector Section (SDS) (figure 4.1).

4.2 Energy loss of stored electrons

In general, the motion of an electron in a MAC-E filter is made up of three components (figure 4.2):

- A fast gyration around its center of motion, due to the cyclotron force.
- A longitudinal motion of the center of motion along the guiding magnetic field lines. For stored electrons, this results in a longitudinal oscillation.
- A slow oscillation around the middle axis in z-direction of the spectrometer (magnetron drift).

For non-axially symmetric fields, the magnetron drift can also cause a slight radial change. These motion plays an important role if an electron is stored. Because the electrons get reflected close to the exits of the spectrometers, they can be stored over a long period of time, which can lead to a full magnetron circle. The primary electrons can cool down by scattering off residual gas molecules and eventually escape the trap due to either energy loss or angular change. The secondary electrons resulting from scattering can only be produced along the primary's path of motion. As these electrons have much lower energies than the stored ones, they can escape the magnetic trap and reach the detector. The spatial distribution of these secondary electrons at the detector can therefore visualize the primary's path of motion. The primary's initial energy can also be indicated by the

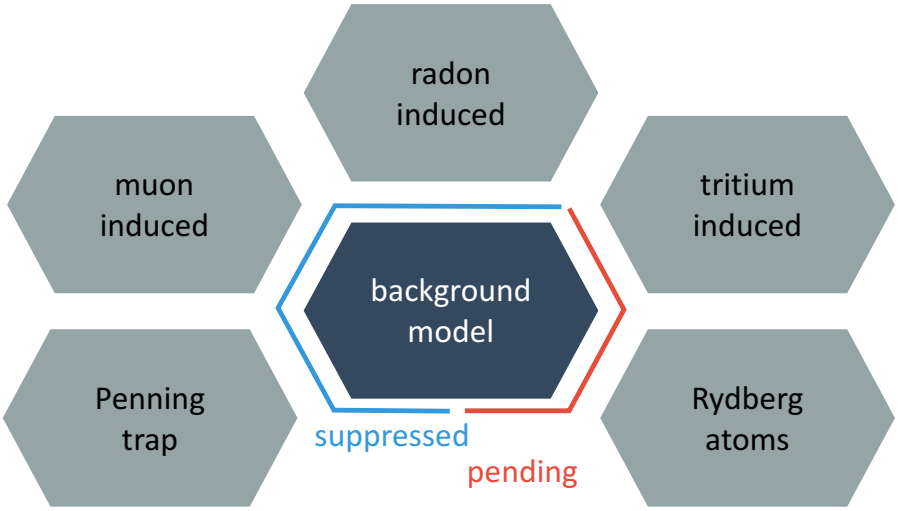


Figure 4.1.: **Different background processes**, which have been able to be identified and suppressed, meaning they do not have a significant overall contribution to the current background (blue). The major background contribution up by now is of unknown source, but may arise due to highly excited atoms, called Rydberg atoms (red). Although various suppression systems have been deployed, tritium may still be a possible background source, when KATRIN enters its long-term neutrino mass measurement (red).

number of secondary electrons produced, which depends on:

- For the PS, electrons below about 100 eV are not stored. This means that the primary electron will not deposit all of its initial energy in the production of secondary electrons. Furthermore, these secondary electrons usually have low energies (figure 5.5) and can escape the trap quite fast, almost producing no tertiary electrons.
- As the electrons perform a circular motion in the presence of a magnetic field, they can also suffer from radiative energy loss due to synchrotron radiation.
- Electrons with high kinetic energies have large cyclotron radii, which can lead to hitting a wall of the spectrometer and therefore a premature termination of secondary electron production.

For an electron at the starting point $\vec{x}_s$, the polar starting angle θ must be smaller than

$$\theta_{\min} = \arcsin\left(\sqrt{\frac{qU(\vec{x}_s)}{E_{\text{kin}}(\vec{x}_s)}\frac{B(\vec{x}_s)}{B_{\max}}}\right) \tag{4.1}$$

to escape the trap. This escape mechanism is independent of the azimuthal angle $\varphi \in 0, 360°$ and therefore has the geometric shape of a cone. Here $U(\vec{x}_s)$ is the electric potential, and

$B(\vec{x}_s)$ the magnetic field at the starting point, and E_{kin} and B_{max} the initial kinetic energy of the electron and the maximum magnetic field in the spectrometer. This angle gets smaller for increasing energies, therefore high energy electrons have a high storage probability.

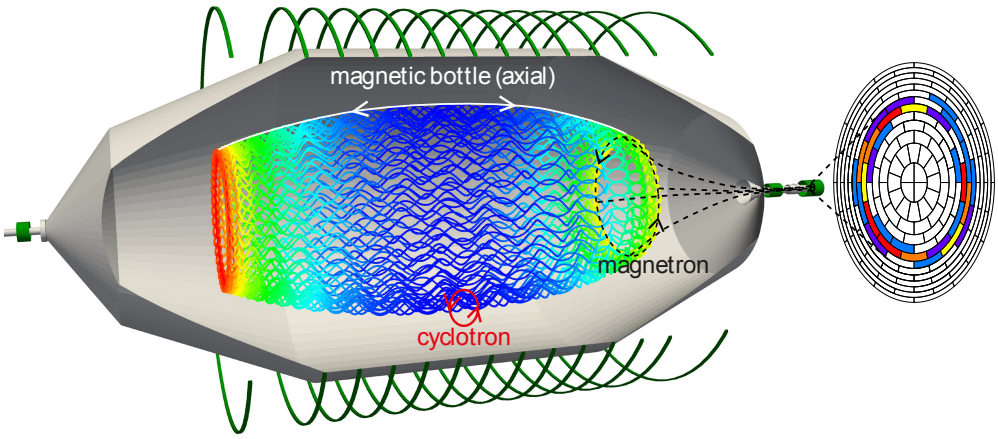


Figure 4.2.: **Cluster signature of stored electrons.** While moving back and forward through the spectrometer, the electron performs a fast cyclotron motion around its guiding path. This is accompanied by a slower and larger motion around the symmetry axis of the spectrometer, the magnetron motion. Therefore secondary electrons being produced from the primary electron will leave a ring-like pattern on the detector. Figure taken from [Har15].

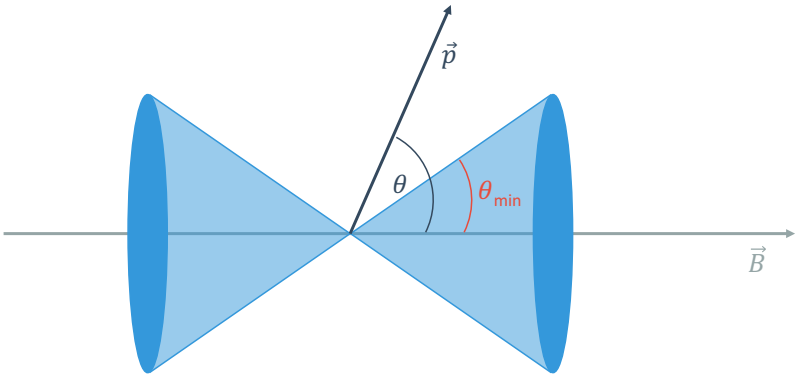


Figure 4.3.: **Escape cone of stored electrons.** If $\theta < \theta_{\text{min}}$ the electron can escape the magnetic trap. Figure inspired by [Frä10].

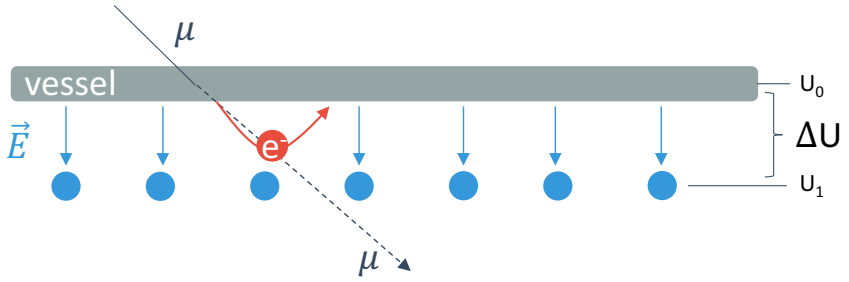


Figure 4.4.: **Muon background shielding.** Muons can scatter with atoms and molecules in the vessel of the spectrometer and cause secondary electron emission within the spectrometer's volume. In order to shield these electrons, wire electrodes are deployed with a negative potential difference to the hull. Figure inspired by [Wan13].

4.3 Muon induced events

Since the bremsstrahlung for muons is lower than for electrons due to their higher mass, a large muon flux coming from cosmic rays of about 10^5 particles per second is expected to hit the spectrometers at Sea level. When passing through, they may cause secondary electron emission because they can deposit energy in the vessel walls of the spectrometers, which are made out of stainless steel and therefore have high stopping power. However, these muon-induced electrons only have a small probability to penetrate the fiducial volume of the spectrometers due to:

- Electrostatic shielding: The vessel and inner wire electrodes are set on different potentials, causing electrons coming from the walls to be reflected (figure 4.4). The MS features a double layer of thin wire electrodes, which yields an improved suppression factor of 10^2 compared to a single layer.

- Magnetic shielding: due to the Lorentz force, the electrons perform a circular motion around the guiding magnetic field lines. Therefore electrons coming from the surface will be rejected, when they reach the magnetic field lines. Only electrons with small polar angles could penetrate the flux tube, but their path of motion would nevertheless be outside of the fiducial volume.

For electrons that do reach the fiducial volume, and therefore have the possibility to reach the detector, a muon veto system was deployed. This detector system checks for coincidence of events in both the muon detector and the FPD.
Since the background due to muons is rather small (about 7 mcps), the muon system is only active when doing measurements when the difference of 7 mcps is significant (mainly for measurements with the PS). Due to the high normal background rate of the MS of 400 mcps, the muon system is turned off, otherwise it would cause a large data overhead when taking long term measurements.

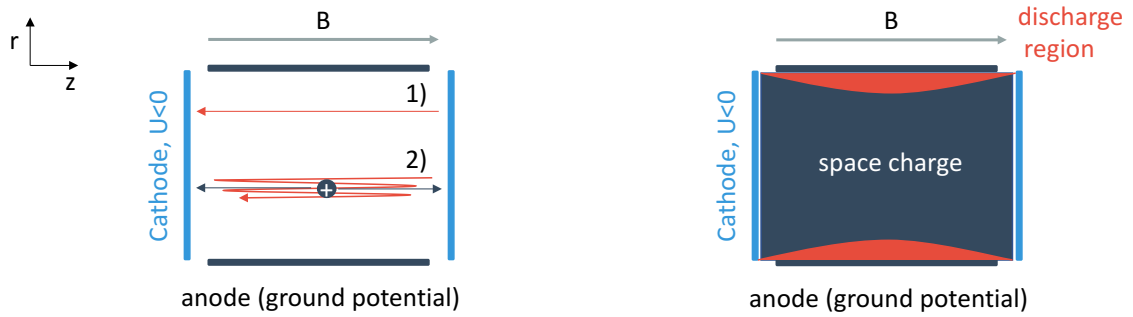


Figure 4.5.: **Penning trap mechanism (left side) and Penning discharge (right side)**, see section 4.4 for details. Figure inspired by [Frä10].

4.4 Penning traps

The Penning traps have been studied in [Frä10]. Particles can not only be magnetically trapped (section 4.2), but also electrically (figure 4.5, left side). A primary electron can be created at the surface of a cathode by for example natural radioactive processes. This electron then gets accelerated towards the middle of the anode and de-accelerated afterwards before hitting the other cathode (1). However, if this electron loses energy due to cyclotron radiation or scattering, it cannot reach the other cathode but gets reflected there, resulting into a trapped electron. This electron may now produce secondary electrons and ions by scattering off residual gas (2). The secondary electrons have low energies and can also be trapped causing further ionization. The maximum number of secondary electrons created by a primary electron is given by

$$N_{\max} = 2^{\frac{e \cdot U_{\text{trap}}}{E_{\text{ion}}}} - 1, \; U_{\text{trap}} = U_{\max}(V_p) - U_{\min}(V_p), \tag{4.2}$$

where U_{trap} is the potential difference of the maximum and minimum potential in the Penning region and E_{ion} is the required ionization energy of the residual gas molecule. The ions move towards the cathode surface where they create further electrodes, which can also be stored with probability $p(e_{\text{ion}})$. If $p(e_{\text{ion}}) \cdot N_{\max} > 1$, a self-sustained Penning discharge starts. The accumulation of electrons leads to a negative space charge, which then decreases U_{trap}. Therefore only electrons in a region close to the anode can gain sufficient energy for ionization [Kna62]. An equilibrium discharge state can be reached, causing a stable Penning discharge.

Due to thorough measurements, simulations and counter-measurements such as a modified electrode system and so called Penning wipers, the background contribution due to Penning discharge has been effectively decreased by several orders of magnitude. Therefore it does not play an important role in standard operational mode anymore.

4.5 Rydberg hypothesis

The remaining background component, after developing systems and models to suppress the known background sources, exceeds the design value of 10 mcps [KAT05] with a rate of ≈400 mcps [Har15]. Dedicated measurements to study this remaining component revealed the following properties [Har15]:

- nearly pressure independent,
- dependence on inner electrode settings, and
- uniformly distributed in the spectrometer's volume.

Possible candidates, which could cause this remaining background contribution, are highly excited atoms with high quantum numbers n, so-called Rydberg atoms [Tro18, Hin18]. In order to show the implications of a high quantum number, the Bohr model is useful. The radius of an atom's orbital electron r_n is described by

$$r_n \propto \frac{n^2}{Z}, \tag{4.3}$$

where n is the quantum number, and Z the nucleus charge number. The quadratic dependency of the radius on n implies that an excited hydrogen atom $Z = 1$ with quantum number $n = 10$, has a 100-times higher radius compared to its ground state. This has also impact on the binding energy of hydrogen and hydrogen-like ions

$$E_n = -\frac{Z^2}{n^2} \cdot 13.6\,\text{eV}, \tag{4.4}$$

which decreases quadratic with n. Furthermore, these highly excited states usually have a high lifetime $\tau \propto n^{4.5}$ [BS57]. Currently it is assumed, that the Rydberg atoms originate from the α-decay of implanted ^{210}Po, a daughter nucleus of ^{222}Rn. The resulting ion ^{206}Pb can scatter with atoms of the stainless steel, causing some of the scattered atoms to leave their lattice and penetrate the spectrometer's volume, or leave the vessel itself, enter the volume and cause a release of Rydberg atoms by hitting the opposite site of the vessel. Since the Rydberg atoms are electrically neutral, they are not affected by the inner electrodes potential and may likely be in highly excited states [Kel82, FMO82, WB94].
To emulate the observed constant volume background density, ionization processes have to take place constantly. However, these ionizations must not come from scattering off residual gas molecules, as no sufficiently large correlation of rate and vessel pressure was observed. Furthermore, the excited states must be low enough, such that the electric field between the vessel and inner electrodes does not cause an instant ionization of the Rydberg atoms. Recent and ongoing studies show that metastable auto-ionization processes are valid candidates to explain the Rydberg atoms, but do not have a sufficiently long lifetime [Hin18].

4.6 Radon induced events

Radon as an indirect source for increased background has been studied in [Frä10, Wan13, Har15]. In these works, a detailed model of the processes leading to electron emission has been developed. Radon can emanate from the stainless steel vessel into the spectrometer volume and decay there via alpha decay to polonium. This may lead to an excited state of Po, causing the emission of electrons in the range of $(1, 10^5)$ eV. The isotopes ^{219}Rn and ^{220}Rn in particular play a major role, due to their short half-life of 3.96 s and 55.6 s. The isotope ^{222}Rn is not of importance, since the half-life of 3.82 d is much larger than the average gas pump out time of the MS of 300 s [Wan13]. As the alpha particles are much heavier than electrons, they will hit somewhere on the inner surface of the spectrometer, because they are not guided by the magnetic field lines. The FPD is only sensitive to electrons which can accompany the radon decays by the following processes (figure 4.6):

- Shake off: The emitted alpha particle can interact on its way with an inner shell electron, transferring some of its energy to the electron, so it can be lifted into a higher shell or ultimately leave the Coulomb field. The necessary energy for a shake off corresponds therefore to the binding energy of the electrons in a specific shell, spanning a continuous spectrum up to 10 keV.

- Conversion: If the decay of ^{219}Rn ends in an excited state of ^{215}Po*, the cool down to its ground state ^{215}Po can either happen via radiation or by emission of an internal conversion electron with energies in the range (40, 500) keV.

- Atomic relaxation: If conversion or shake-off electrons have been emitted, a vacancy is left in the inner shell, which can be occupied by an electron from higher shells. These transfer their binding energy to a second shell electron, which can then be emitted. This process doubles the number of vacancies and can therefore lead to a cascade of relaxation processes being propagated to the outermost shell accompanied by the emission of several electrons with energies up to 10 keV. The vacancy can also be filled via radiation, transferring the initial vacancy to a higher shell.

Another source of natural radon isotopes is the NEG material, which is used together with the TMPs to maintain the UHV conditions in the spectrometers. To prevent the radon from reaching the spectrometer volume, a nitrogen cooled baffle system has been deployed in front of the NEG pumps to adsorb the radon [Har15].

The experimental approach to study the background contribution of radon is usually done at an artificially elevated pressure. At an elevated pressure, the number of residual gas molecules is increased, making the ionization of gas molecules due to stored electrons more likely and thus leading to a decreased cool down time and faster energy loss of the stored electron. This decreased mean average time between two ionization processes is especially useful for the algorithm used to cluster these events and classify them as secondary electrons due to a radon decay (section 5.3).

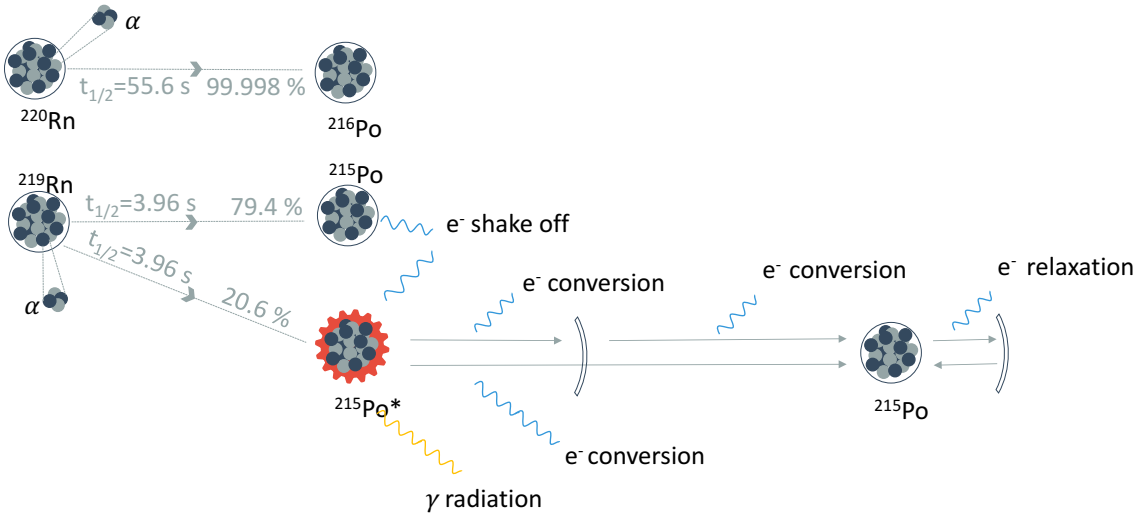


Figure 4.6.: **Sketch of radon accompanied electron emission processes.** Explanations in section 4.6. Sketch based on [Wan13, Sch14].

4.7 Tritium induced events

Tritium can enter the spectrometer section in either molecular or ionized form. Although simulations of the CPS [FR18] have revealed a higher retention factor of the transport section than designed [KAT05], small number of tritium molecules can still diffuse from the source section to the PS and afterwards the MS.
Despite the suppression methods of the DPS (section 3.2.2.1), there is still a small possibility that ions enter the PS. There the electric field works not as a filter but as an accelerator, causing the ions to hit the downstream side of the PS hull or the full metal electrodes. The ions then get implanted and neutralized there. If these implanted tritium atoms decay in the vessel steel or on its surface, this should not result in an increased and therefore measurable background rate, because the MAC-E filter works as a magnetic and electric shield. However, if they get emitted back into the volume and decay there, this would lead to secondary electron production through ionization, analogue to keV electrons emitted by a radon decay accompanied process.

A significant amount of tritium in the spectrometer section is therefore unfavorable, because an increased background rate due to tritium has a direct implication on the neutrino mass sensitivity, as the background is a free fit parameter in the neutrino mass fit. This is especially important if the tritium is able to reach the MS, as the PS in standard neutrino mass measurement mode is operated on a lower potential than the MS. Therefore the electrons due to tritium decay will get rejected in the analysis plane of the MS.

It is therefore of high importance, to develop and have high precision methods to detect tritium in the PS, before it can reach the MS. In the course of this work, statistical methods have been applied to highly increase the sensitivity to tritium induced electrons.

5 Machine learning algorithms for active background characterization

This chapter describes both unsupervised and supervised machine learning approaches as an superior alternative to the previous clustering algorithm.

Machine learning can be divided into two domains (figure 5.1). In the supervised approach, each observation of the predictor measurement $x_i, i = 1...n$ is associated with a response measurement y_i. The goal is to fit a model that relates the response observations to the predictor observations in such a way that either the learned model can predict future response observations y_k with x_k as input (prediction), or learn more about the relationship of the data (inference). A popular example is a linear regression, where a function $f(x) = a \cdot x + b$ is fit to $(x, y)_i$ in order to approximate y.
In contrast, unsupervised learning tries to learn features about the observation vector $x_i, i = 1...n$, without knowing the response y_i. Here, a popular example is the k-means clustering method, where the observation space V is divided into k-distinct, non-overlapping subspaces $V_j, j = 1..k$ (Voronoi regions) and each observation within these regions belongs to the same cluster.
Although the first learning algorithms, like the perceptron algorithm by F. Rosenblatt [Ros57], have been proposed in the last century, the popularity has increased in recent times. This has a lot of reasons, but the ones with the most impact include:

- The number of data taking devices (sensors) has increased rapidly due to digitalization and the need to store information as data.
- Increased and improved methods and infrastructure to store and process this data rapidly (Big Data).
- The success stories of tech-companies such as Google and Facebook, who put a lot of their financial power into research about machine learning algorithms, as their business is based on these techniques.

Nowadays, machine learning algorithms have a broad application spectrum. Most often, we do not even recognize their deployment in everyday situations such as text auto-completion [Sho13], and traffic prediction [LD15]. In economics, machine learning methods are used e.g. in sales prediction [SC08], customer relationship management [NXC09], and personalized advertisement [Dis18]. However, in science these methods have already been

established for decades. Here, these algorithms are used, for example, in molecular biology for genome classification [SR02], in medical imaging for assisted cancer detection [MG11], and also in high-energy physics for particle identification. A popular example is the multi-variate analysis of the Higgs boson conducted by the ATLAS collaboration. When they had the first results of their analysis, they provided a data challenge, opening simulation data to interested data scientists to explore the data and develop their own analysis work-flows [OLu14, Kag18].

In comparison to the Compact Muon Solenoid (CMS) detector, which has several different detectors and therefore can do a full multi-variate analysis, KATRIN features only one detector. However, the information gathered by the detector can also be split into different channels, making a multi-variate analysis possible, at least for background characterization of cluster events.

To apply and study both supervised and unsupervised methods, a combination of particle storage simulations (section 5.1) and measurements is used to construct a training data set (section 5.2). Then this training data set is used to optimize the performance of the cluster algorithm, which was used up until now (section 5.3). Afterwards, examples of unsupervised (section 5.4) and supervised methods (section 5.5) are described in detail and applied to simulation data to check the performance. The chapter closes with a short comparison of the algorithms and a recommendation in section (5.6).

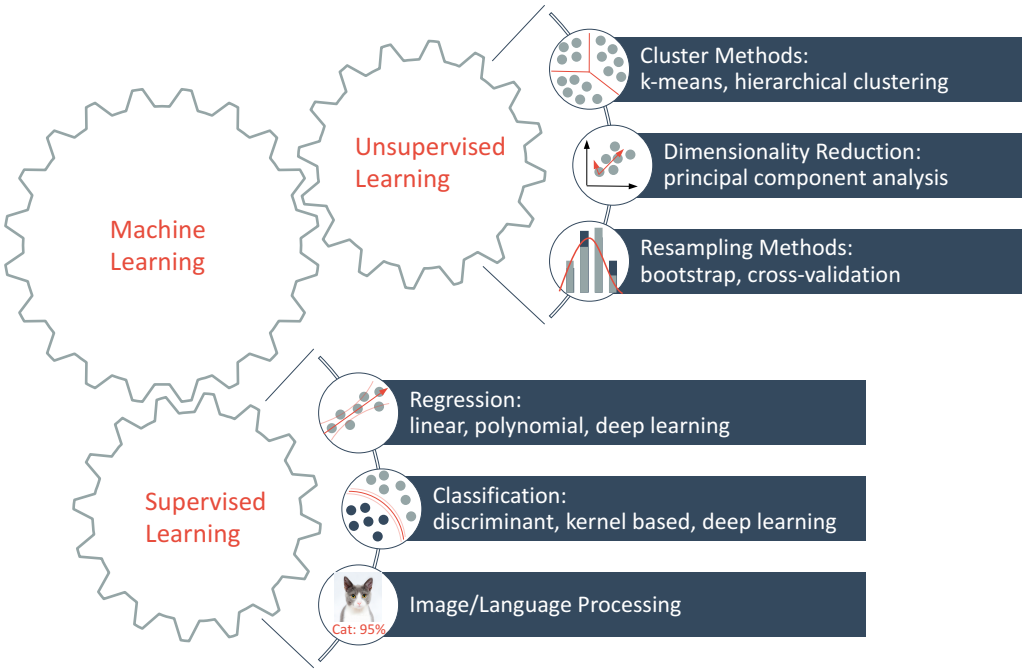
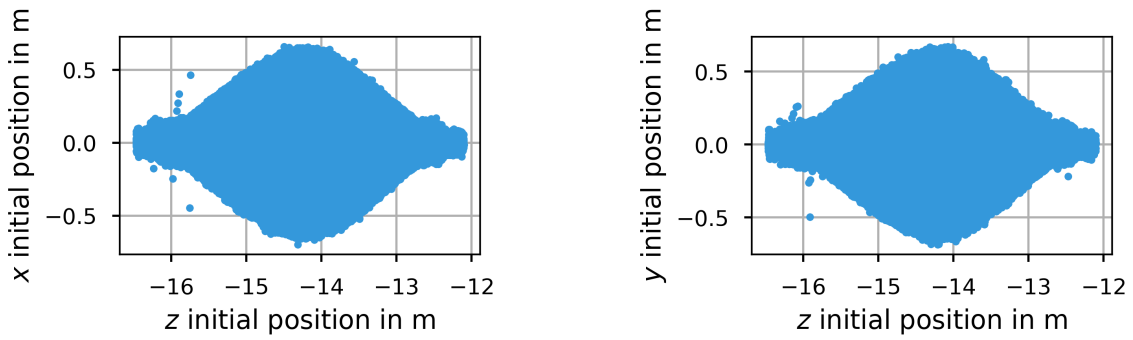


Figure 5.1.: **Application of machine learning algorithms** are summarized in two areas: Unsupervised learning, which does not require prior knowledge in form of a training data set, and supervised learning, which first needs to learn features from a training data set.



(a) (b)

Figure 5.2.: **Initial position of the primary and secondary electrons** within the PS. The figure shows that electrons are generated homogeneously both in (x, z) as well as in (y, z) direction. The tiny number of outliers is not important, as these electrons are not able to reach the PS2 anyway. z-axis normed to FPD coordinates and x, y-axis to center of beamline.

5.1 Storage simulations

In order to understand the behavior of stored electrons due to tritium β-decays within a MAC-E filter based spectrometer, simulations are performed. The simulation configuration is given in subsection 5.1.1 and the results in subsection 5.1.2.

5.1.1 Simulation configuration

Electrons are generated from a tritium β-energy spectrum homogeneously in the flux tube of the PS (figure 5.2). The KASSIOPEIA generator used to do so, calculates the initial position of the particle with a flux tube value Φ, and the magnetic field strength, which depends on the z-position [Ahm16]. However, as the magnets will be operated at 70 % of their strength, the conservation of the flux tube (equation 3.3) would cause a rise in the radius of the flux tube, which could exceed the dimensions of the PS. To emulate realistic conditions, the nominal flux tube of 191 Tcm^2 needs to be modified to:

$$\Phi_{\text{nominal}} = B \cdot A = B \cdot \left(\pi \cdot r^2\right) \tag{5.1}$$

$$\Phi_{\text{final}} = 0.7 \cdot \Phi_{\text{nominal}} = 133.7\ \text{Tcm}^2. \tag{5.2}$$

However, electrons created outside of the flux tube can still penetrate it frequently, due to their rather large cyclotron radius

$$r_{\text{cyclotron}} = \frac{m \cdot c\sqrt{\gamma^2 - 1}}{|q| \cdot B}, \tag{5.3}$$

which can lead to an adiabatic guidance through the PS's magnetic field. The new flux tube is therefore simply scaled by a factor of 0.8 instead of 0.7, to include the generation of these electrons

$$\Phi_{\text{sim}} = 0.8 \cdot B \cdot \left(\pi \cdot r^2\right) = 152.8\ \text{Tcm}^2. \tag{5.4}$$

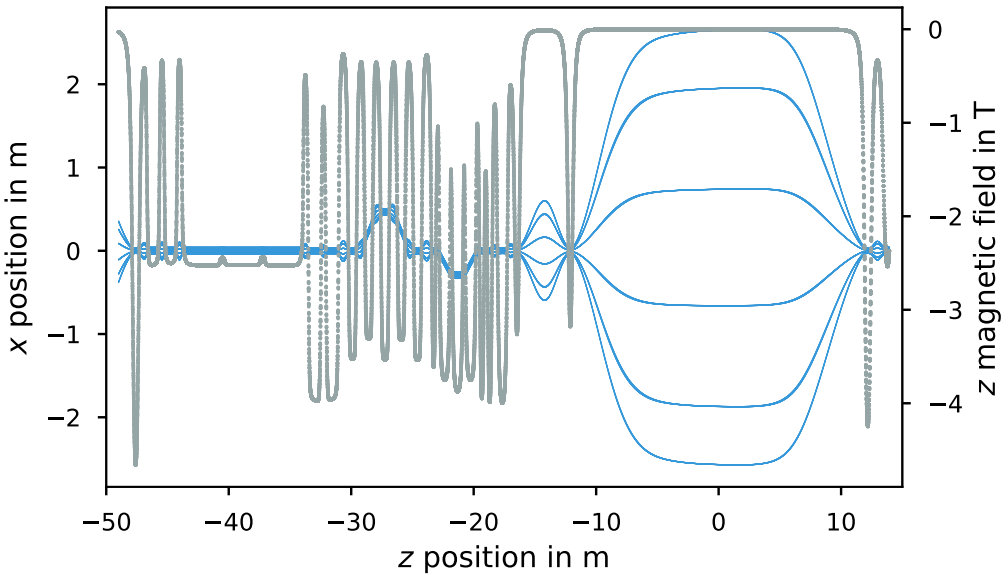


Figure 5.3.: **Magnetic field strength at different z positions** of the whole beamline is displayed in grey, the corresponding trajectory of electrons starting on the outer ring of the FPD projected to the x-axis is displayed in blue. A 0.6 mT setting of the LFCS was used.

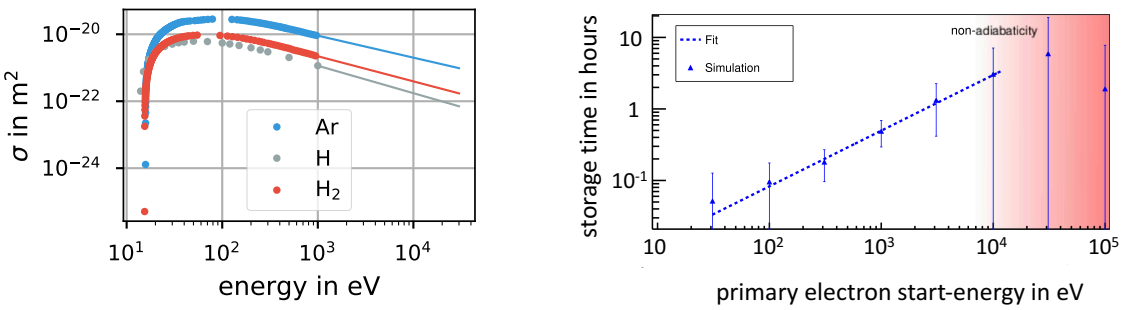
Magnetic fields are simulated in the whole beamline, because they have a non-negligible influence on the magnetic field in the PS created by the solenoids PS1 and PS2. In particular, the magnetic field of the CPS has a major impact on the shape and strength of the magnetic field in the PS. In figure 5.3 the magnetic field strength in dependence of the z-position (grey), as well as the corresponding trajectory of an electron starting from the pixels of the outer FPD ring projected on the x-axis in dependence of the z-position are shown. The hull of the PS is operated at −18600 V and the full metal electrodes as well as the inner wire electrodes with an offset to the hull of −400 V. The MS electrodes are grounded. Tracking of the electrons is performed using a symplectic integrator[1] with a step size of 1 mm. This means every mm the equations of motion are numerically solved and the particle shifted according to the solution. Interactions with residual gas are also included, which in this case is hydrogen at a pressure of $3{\times}10^{-9}$ mbar. For further information about the software and simulation framework KASSIOPEIA, have a look at [Fur15, FG17].

Termination of tracking

To minimize computational effort, the tracking is terminated if:

- `term_min_z`: The particle passes the z position of the center of the PS1 magnet.
- `term_max_z`: The particle passes the z position of the center of the PS2 magnet.

[1] implementation in KASSIOPEIA is based on [Chi08]



(a) (b)

Figure 5.4.: **a): Ionization cross sections of Ar, H and H_2** up to 1 keV derived from the LxCat database [LxC13, LxC12, LxC10]. To extrapolate the energies up to the endpoint of the tritium spectrum, a power law fit is performed. At the tritium peak of ≈2-5 keV, the mean ionization cross section is $\approx 1.38 \cdot 10^{-21}\,\mathrm{m}^2$. **b): Storage time of electrons**. Electrons with energies over 1 keV can have a total storage time of up to several hours. Figure taken from [MD13a].

- `term_max_r`: The particle exceeds the maximum radius and is considered to be no longer in the flux-volume.
- `term_surface`: The particle passes a surface of the PS.
- `term_electrode`: The particle passes an electrode.
- `term_max_stepsize`: The particle tracking passes the threshold number of steps (10^9).
- `term_max_t`: If real calculation time between two steps exceeds the limit. This can be the case for electrons coming close to the electrodes. Here, the approximation of electric fields by the so-called zonal-harmonic field solver is valid no longer and the calculation method switches to an exact one, which is much slower. This can cost computation time, and particles close to the electrodes are not of interest in this storage simulation in any case.

Dynamical enhancement

Since the small scattering cross section of ≈ 20 keV electrons and residual gas molecules can imply a storage time of up to thousands of seconds (figure 5.4b [MD13a, Wan13]), storage simulations require large computational power. This arises mainly due to the exponential decrease in the scattering cross section of high energy electrons and residual gas molecules (figure 5.4a). To compensate for this energy dependent cross section, a dynamical enhancement factor is used. This factor is multiplied with the scattering cross section to make interactions, as well as the energy loss due to synchrotron radiation at high energies, more likely. It is dynamical such that its impact is higher for higher energies and decreases when the electron is losing energy. This factor can therefore speed up the

storage simulations and save a lot of computational effort [Ahm16]. In the simulations performed in this thesis, the dynamical enhancement factor was set to 10^6. Furthermore, H_2 was used as a residual gas to provide a scattering medium for the electrons. While this is a realistic representation for emulating nominal pressure measurements, it is not entirely correct for elevated pressure measurements. These measurements use argon and make use of its rather small ionization energy and higher cross section, compared to hydrogen atoms and molecules (figure 5.4a), which results in a higher number of secondary electrons.

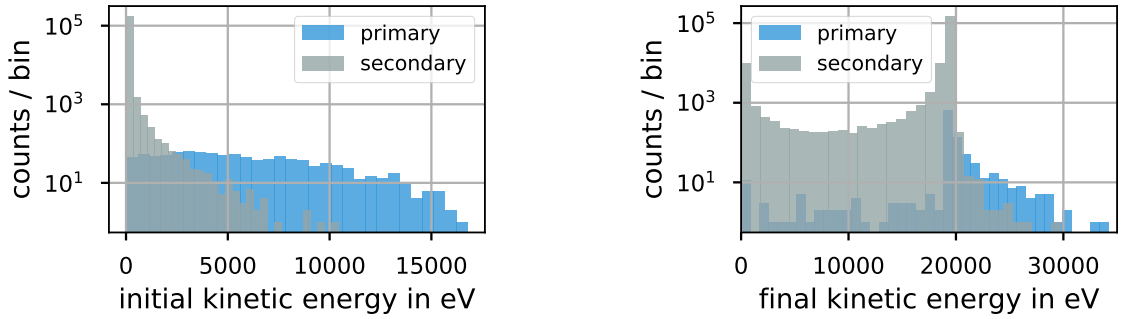
5.1.2 Simulation results

The analysis of the simulated data is performed in the following way:
First all electrons which reach the PS2 are located (`term_max_z`), because only these have the possibility to reach the detector. Here it is assumed that all electrons reaching the PS2 would be guided through the magnetic field of the MS and reach the detector. Furthermore, a ROI-cut is performed in the energy range of $[-22, -16]$ keV corresponding to all electrons with energies 19±3 keV. The total number of electrons generated in the simulation is 176480, of which 78182 were able to reach the PS2 and 73485 lie within the ROI. This is equal to a fraction of $\frac{73485}{176480} \approx 0.42$. With a symmetric cone electrode configuration and homogeneously generated electrons within the flux tube, a theoretical fraction of 0.5 would have been expected to reach the PS2. However, the additional terminators and a reasonable `step_size` may lead to an early abort of the simulation and therefore the fraction of electrons reaching the PS2 relative to the total number of electrons in the simulation is smaller.
Second, the position of the electrons terminated at PS2 needs to be mapped onto the detector pixels. To derive a mapping procedure, a second simulation was conducted. In this simulation only the trajectory of the electrons from the detector throughout the whole beamline to the source section was tracked, with magnetic fields on and electric fields off. The most recent beamline configuration (Very-First Tritium campaign (VFT), FT) was used. To derive the maximum radius electrons can have in the middle of the PS2 magnet to reach the detector, the radius for every pixel in the outer ring is determined in dependence of the z-position. Next, an interpolation between these value pairs allows one to approximate the radius to an exact z-value. Then the mean of all radii of electrons starting at the center of the pixels in the outer FPD ring is determined as a measure of the maximum radius, which is 0.041 m. The final position at termination by `term_max_z` is then scaled by a factor of

$$f = \frac{r_{\text{detector}}}{r_{\text{PS2}}} = \frac{0.045}{0.041} \tag{5.5}$$

to map the electrons in the PS2 to the detector. Since all events (primary and secondary) are cluster events, these clusters need to be distinguished from each other. This is simply done by their `simulation_id`, because each simulation only simulates one primary electron. As storage simulations can consume a lot of time, the simulations are performed in parallel on the TESLA cluster. In total, two storage simulations consisting of ≈1000 primary electrons each were performed at two different pressures: $3{\times}10^{-9}$ mbar (elevated pressure) and $4{\times}10^{-11}$ mbar (nominal pressure).

The energy spectra of both primary and resulting secondary electrons at creation and



(a) (b)

Figure 5.5.: **Initial and final energy distribution of simulated electrons.** Initial energy distribution of primary electrons (blue) and secondary electrons (grey) is shown in a). While the primary electron energy distribution shows the one of tritium β-decay, the secondary electron distribution shows a typical ionization distribution. Energy distributions at the termination of tracking are shown in b). Most primary and secondary electrons have an energy close to 19 keV. The peak around 0 eV arises from electrons being created close to electrodes and terminated there and electrons created close to the PS1 or PS2, where the acceleration due to the electric field is not strong any more.

termination can be found in figure 5.5. Here it can be assured that the initial primary electrons follow a tritium β-decay energy distribution (figure 5.5a blue) and the secondary electrons an energy distribution due to ionization (grey) by visual confirmation. Furthermore, it can be double-checked that the electric potentials are set correctly at 19 kV by having a look at the energy distributions of the electrons at their termination (figure 5.5b), which both peak at an energy of 19 keV.
The energy distribution of all electrons which can reach the PS2 and be contained in the ROI can be found in figure 5.7a. A projection of a cluster event, terminated at PS2, to the detector can be found in figure 5.6c as well as the figure of a simple ring fit in 5.6a. This can be done for all simulation clusters. The histogram of their mean radii can be used to get information on how the cluster events appear within the PS's volume (figure 5.6b).
Furthermore, the total number of secondary electrons, produced by scattering of the primary electron off residual gas molecules (H_2), can be investigated. Figure 5.7b shows an energy dependent number of secondary electrons and figure 5.7c confirms this nearly linearly dependent behavior by an exponential fit. For a condensed view of the input and output parameters of the simulations, table A.1 may be useful.

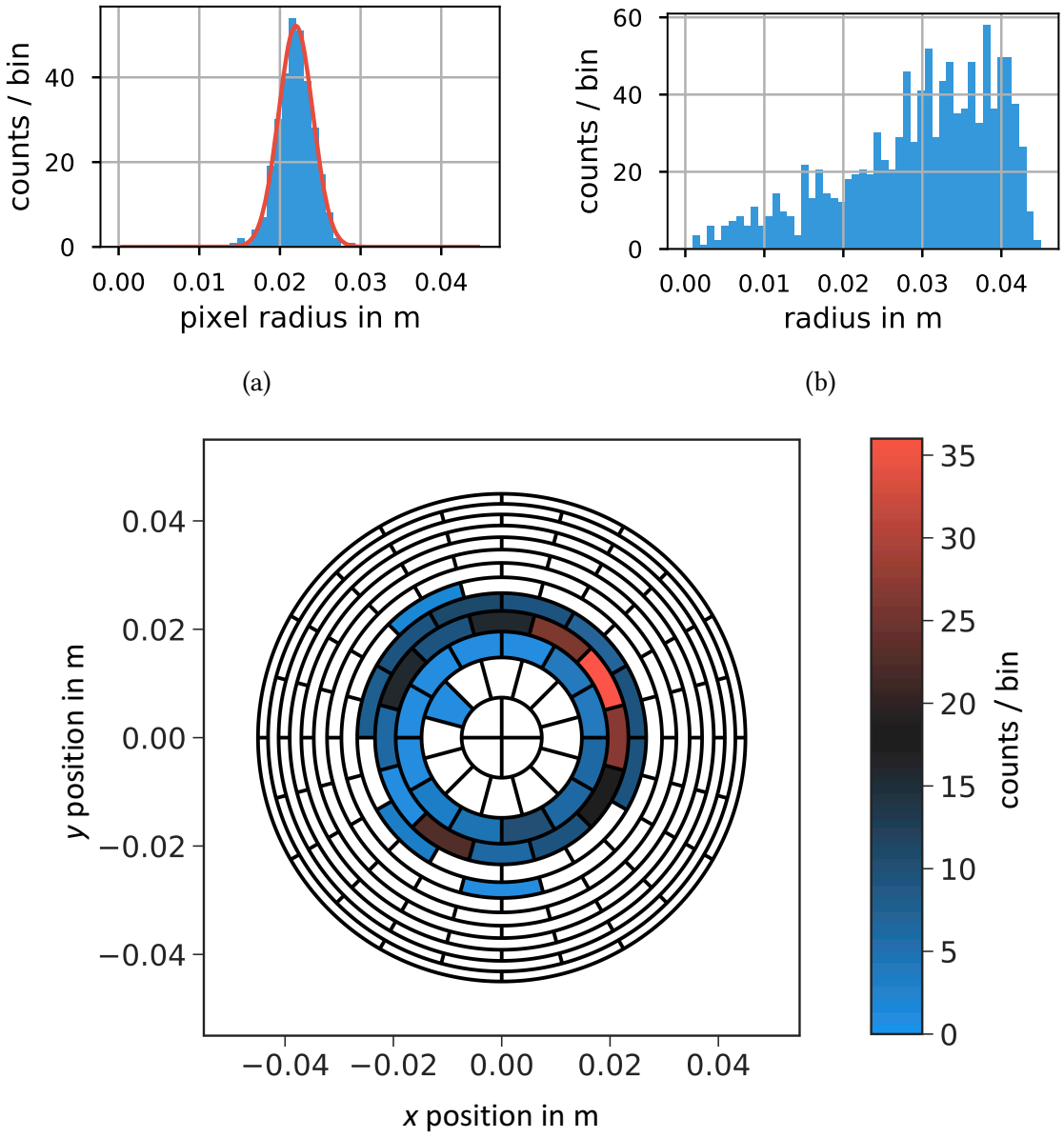


Figure 5.6.: **Simulated cluster properties.** a): Simple ring fit by counting the frequency of the radii of the events within one cluster and fitting with a Gaussian distribution. b): Histogram of all fit cluster radii. This visualizes the probability of a cluster appearing at a certain radius. For electrons being generated homogeneously in the flux tube, a larger number of events is expected with increasing radius. c): Events within one cluster projected on the FPD.

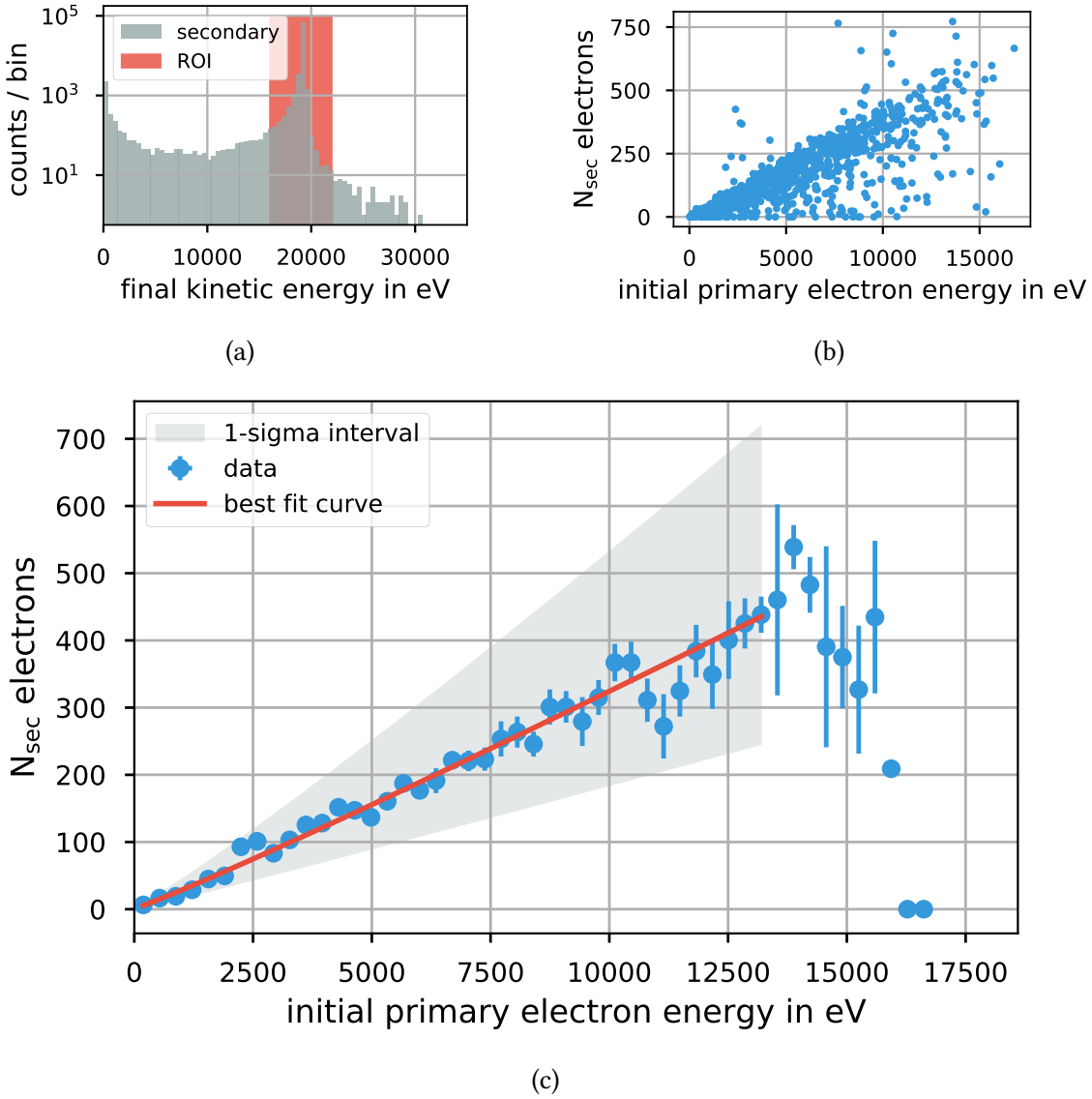
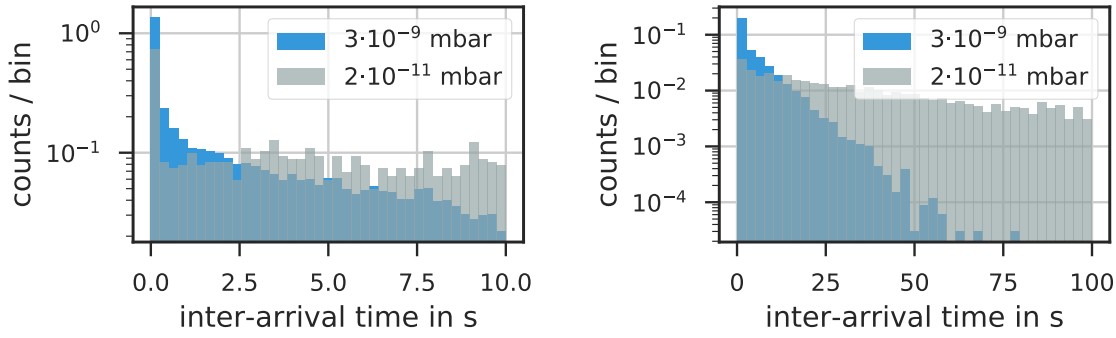


Figure 5.7.: **Storage simulation ROI and number of secondary electrons per primary.** a): Energy spectrum of electrons being terminated at PS2 (grey). As in a measurement, a ROI cut is performed. b): For each simulation and corresponding initial primary electron energy, the produced number of secondary electrons is visualized. c): The number of secondary electrons is binned and represented by the mean of the bin as a data point. The standard deviation of the data within one bin is the error. Then a fit is performed, leaving out the last ten data points, because the statistics are too low in this region. The exponential fit $a \cdot \exp(b \cdot E)$ confirms the linear behavior with the parameters $a = (1.98 \pm 0.41)10^{-2}$, $b = 1.05 \pm 0.02\,\mathrm{eV}^{-1}$, which is also in good agreement with simulations within the MS [MD13b].



(a) (b)

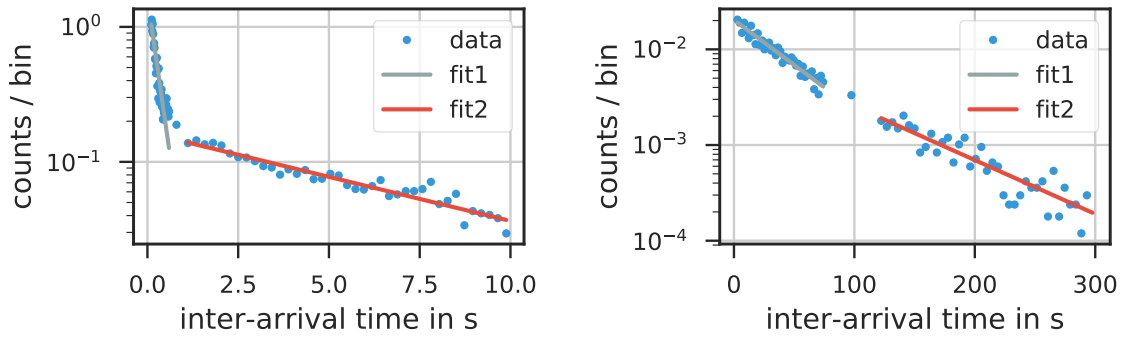
Figure 5.8.: **Inter-arrival times at different pressures** with density normalized to 1. a): Time scale up to 10 s. b): Time scale up to 100 s.

5.2 Construction of training set

In order to evaluate the performance of the different algorithms two training data sets are constructed one at elevated pressure ($3{\times}10^{-9}$ mbar) and one at nominal pressure ($4{\times}10^{-11}$ mbar). This is done by combining the data from the storage simulations (section 5.1) with sampled data according to measurements at both elevated (subsection 5.2.2) and nominal pressure (subsection 5.2.3).

5.2.1 Known characteristics of cluster events

The cluster events are distinguished from normal background events by looking at the distribution of inter-arrival times. Under the assumption, that normal background events are independent, uncorrelated, and follow a Poisson process, their inter-arrival distribution should follow an exponential distribution. The proof, based on [Kos14], can be found in appendix A. In a logarithmic representation, the distribution therefore should fall off linearly. However, the correlated cluster events tend to have lower inter-arrival times, leading to a second and different exponential part in the distribution. This effect can be enhanced by artificially increasing the pressure in the spectrometers, which causes scatterings of the stored electrons off residual gas molecules to be more likely. Therefore the mean-free path length is decreased and ultimately the mean time between two ionization processes is also decreased. In the case of low pressure, the distribution of the inter-arrival times transforms and becomes more flat, making it nearly impossible to distinguish cluster events from normal events (see figure 5.8).
In a cluster analysis, the frequency of inter-arrival times of the events are therefore counted and plotted in a logarithmic representation and then the time-threshold for the cluster algorithm can be read out of this plot (see figure 5.10 for an example).



(a) (b)

Figure 5.9.: **Exponential fits to inter-arrival times at different pressures.** a): Elevated pressure, parameters in table 5.1. b): Nominal pressure, parameters in table 5.2.

Table 5.1.: **Fit $A \cdot \exp(-t/\tau)$ to the counted inter-arrival times (density normalized to 1) at elevated pressure.**

	Name	Value	Error
Fit1	A	1.66	1.15
Fit1	τ in s	0.23	0.16
Fit2	A	0.17	0.56
Fit2	τ in s	6.61	34.39

Table 5.2.: **Fit $A \cdot \exp(-t/\tau)$ to the counted inter-arrival times (density normalized to 1) at nominal pressure.**

	Name	Value	Error
Fit1	A	0.02	0.69
Fit1	τ in s	46.18	36.40
Fit2	A	0.01	0.97
Fit2	τ in s	77.29	50.01

Sampling of data

In order for the training data to have similar properties as the inter-arrival distribution of the measurement, the distribution of measurement data is fit. Then the inter-arrival times of the simulation data are sampled from this fit distribution. Hereby Fit1 corresponds to the fit of smaller inter-arrival times under the assumption, that these events are purely cluster events. Analogously, Fit2 corresponds to purely uncorrelated background events at larger inter-arrival times (figure 5.9).

5.2.2 Elevated pressure

The sampled simulation data is derived from the measurement with the FPD runs 35160 - 35164. In a first step, measurement data with an elevated pressure is used, because the times between two ionization processes are shorter than at lower pressure. This makes electrons due to ionization distinguishable from the normal background events by counting the frequencies of the time difference of two consecutive events. Then this distribution is fit with two exponential distributions (figure 5.9a). The first one describes the probability distribution of ionization inter-arrival times and the second one the distribution of inter-arrival times of normal background events. The fit parameters returned by iMinuit[1] can be found in table 5.1. This measurement was done at a pressure of $3{\times}10^{-9}$ mbar. To derive an analytical calculation of the mean time between two ionization events, it is assumed that the pressure in the spectrometers in equilibrium can be estimated via the ideal gas law

$$p \cdot V = N \cdot k_b \cdot T, \tag{5.6}$$

where p is the pressure, V the volume, N the number of particles in the volume, k_b the Boltzmann constant and T the temperature. Expressing the number of particles as a density

$$n_v = \frac{N}{V} \tag{5.7}$$

allows one to rewrite this equation with the mean free path length of an electron $\langle x \rangle$, which describes the average length an electron can propagate before scattering with a residual gas molecule, with a scattering cross section σ

$$\langle x \rangle = \frac{1}{n_v \cdot \sigma}, \tag{5.8}$$

$$p = \frac{k_b T}{\langle x \rangle \, \sigma}, \tag{5.9}$$

$$\langle x \rangle = \frac{k_b T}{p \sigma}. \tag{5.10}$$

The β-electrons of tritium most likely have a kinetic energy of about $E_{\text{kin}} = 2$ to 5 keV. To include higher energetic electrons, an average of 10 keV is assumed in all calculations,

[1] Python interface [IMi18] for the minimizer Minuit [JR75].

which results in a velocity of

$$\gamma = \frac{E_{\text{kin}}}{\text{mc}^2} + 1, \tag{5.11}$$

$$\beta = \sqrt{1 - 1/\gamma^2}, \tag{5.12}$$

$$v = \beta \cdot c = 0.195 \cdot c. \tag{5.13}$$

$$\tag{5.14}$$

Using $v = \frac{\langle x \rangle}{\langle t \rangle}$, equation 5.10 can be expressed as

$$\langle t \rangle = \frac{k_b T}{v \cdot p\sigma}, \tag{5.15}$$

where $\langle t \rangle$ is the mean time between two ionization processes and has an impact on the automatic derivation of the distance measure parameter of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm introduced in section 5.4 [EK96]. In all analytical calculations for the mean time between two ionization processes, an average kinetic energy of 10 keV, an average cross section of 10^{-21} m^2 for both argon and hydrogen (5.4a) and a room temperature of 20 °C is assumed. For more detailed calculations, also considering the different ionizations cross sections, the reader is referred to the appendix A.

Using equation 5.15 and a pressure of $3{\times}10^{-9}$ mbar yields a mean time between two ionization processes of 0.23 s. This can be compared to the fit parameter τ of Fit1 (figure 5.9, table 5.1), because this parameter is also the expectation value of the exponential distribution ($\text{E}[X] = \tau$) and therefore represents the mean time between two consecutive events. The value obtained by the fit is $\tau = 0.23 \pm 0.16$ s and is therefore in good agreement with the calculated value of the mean time between two ionization processes.

In a next step, the number of cluster and non-cluster events in FPD runs 35160 - 35164 is derived, by using the DBSCAN algorithm with a fixed set of hyper parameters. The total rate of both cluster and non-cluster events over the total time of the measurement is calculated, as well as the number of clusters and number of cluster events.
The two fit distributions are then used to sample simulation data, having the same signature of inter-arrival times as the measurement data by using rejection sampling. First, the number of total events N_{sim} is specified. Then the total time (of the sample measurement data) T_{sim} is calculated by scaling with the observed rate R_{meas} from the measurement

$$T_{\text{sim}} = \frac{N_{\text{sim}}}{R_{\text{meas}}}. \tag{5.16}$$

For every non-cluster event, an inter-arrival time $\tau_{\text{non-cluster}}$ is sampled from Fit2 (figure 5.9a, table 5.1). Then these times are added up (cumulative sum) to obtain the timestamp $t_{\text{non-cluster},j}$ of each non-cluster event.

$$t_{\text{non-cluster},j} = \sum_{i=0}^{i=j} \tau_{\text{non-cluster},i} \tag{5.17}$$

The cluster events are treated differently: A start timestamp is drawn uniformly between 0 and the total time $t_{\text{cluster},j} \in U(0, T_{\text{sim}})$, under the assumption that cluster events are distributed uniformly in time. Next, a random cluster from the simulation data is selected. The number of events k in this cluster is then subtracted from the cluster event budget, which was determined by using the cluster algorithm on measurement data. The timestamps of the events within the cluster are derived by sampling the inter-arrival times from Fit1 (figure 5.9a, table 5.1) and adding them to the start timestamp

$$t_{\text{cluster},j+k} = \sum_{i=0}^{i=k} \tau_{\text{cluster},j+i}. \tag{5.18}$$

This process is repeated until the budget of cluster events is less than 0, meaning a cluster rate close to the one obtained in the measurement is reached.

5.2.3 Nominal pressure

In the same way as in subsection 5.2.2, a training set simulating a low pressure measurement is constructed. Here, the runs 35173 - 35189 are used as reference data for the inter-arrival times and the cluster rate. As discrimination between cluster and non-cluster events is not as easy as in the elevated pressure regime, the calculated mean time between two ionization processes corresponding to a pressure of $1.5{\times}10^{-11}$ mbar is used as a starting parameter for the first fit. The fit value of 46.18 ± 36.40 s (figure 5.9b, table 5.2) agrees quite well with the calculated one 47.52 s, however the errors of the fit parameters are quite large (table 5.2).

5.3 Previous algorithm

The idea to distinguish cluster events from normal background events is to use the characteristic of the inter-arrival time distribution at elevated pressure (section 5.2.1). This clustering algorithm has been used extensively (together with an elevated pressure in the spectrometers) to identify cluster events [Frä10, Sch14, Har15, Ahm16]. However, the algorithm used in this thesis does not make use of the accidental detection parameter ΔT [Har15]. This parameter does only have a minor impact on the overall performance of the algorithm. The algorithm is based on two parameters:

1. `delta_t`: The maximum time between events to be considered as within a cluster.
2. `min_samples`: The minimum number of events within a cluster to be considered as a cluster.

A major drawback of this algorithm is the `delta_t`. The value is usually derived by looking at the inter-arrival times of background events and set to a value where the exponential distribution of the inter-arrival times of normal background events does apply no longer (figure 5.10). At low pressure, the manual choice of `delta_t` becomes difficult (figure 5.9b). However, as shown later, the hyper parameters of this algorithm can still be optimized for the low pressure regime with the aid of the analytically calculated time between two ionization events.

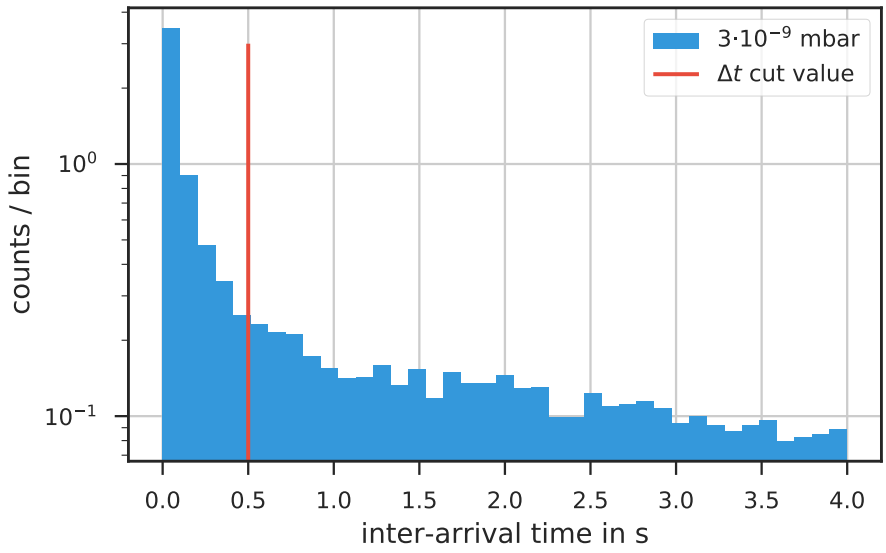


Figure 5.10.: **Inter-arrival times and cut value.** Inter-arrival times of background events at an elevated pressure. At around a time of 0.3 s, the exponential distributions can be separated into cluster events below the cut (red) and normal background events above the cut.

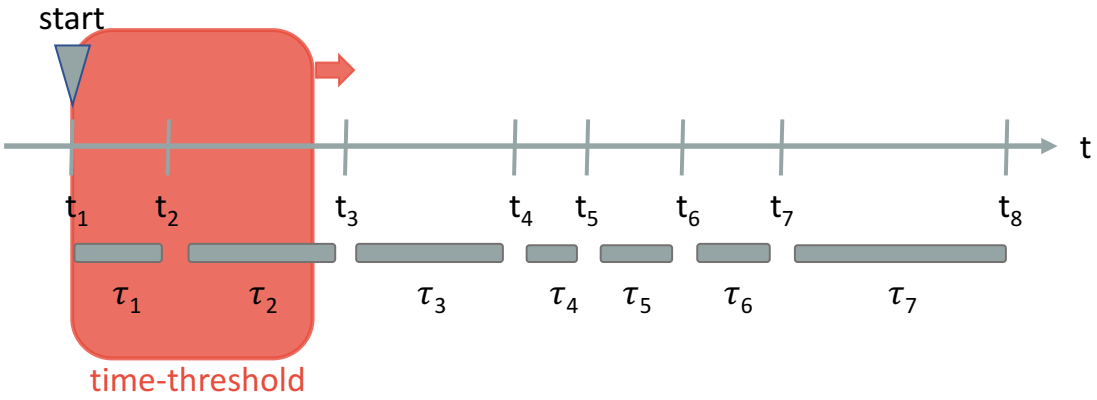


Figure 5.11.: **Schematic sketch of the cluster algorithm's working principle.**

Cluster detection

The cluster detection is straightforward (illustrated in figure 5.11):

1. Start at event 1 with time t_1. Check if the inter-arrival time $\tau_1 = t_2 - t_1$ is smaller than the reference `delta_t`. If true, tag events as clusters and continue with event 2. If false, continue and start again with event 2.

2. Check for all following events if the inter-arrival time is smaller than `delta_t`. The algorithm terminates if the condition is no longer met. The number of events tagged as a cluster are then counted. If this value is larger than the minimum cluster-size-threshold, these events are tagged as a cluster,otherwise they are not tagged.

In the sketch (figure 5.11) if `min_samples` is, for example, 5 all events besides event 8 have inter-arrival times below the given threshold `delta_t` and are tagged as a cluster.

The advantages of this algorithm are that it works well at an elevated pressure, it is easy to understand, and it is reasonable fast. However, if pressure decreases and the mean time between two ionization events becomes larger, determining the time-threshold manually becomes difficult and can lead to a higher misclassification rate. Another major drawback is that this univariate approach only considers correlations in time. However, the storage simulations in section 5.1 have shown that the secondary electrons, produced by scattering, will leave a ring pattern on the detector due to the magnetron movement of the primary electron and therefore have a correlation in radius. This is where the new approach tackles the problem to give a superior classification of the background events.

Hyper parameter estimation

In general, the performance of an algorithm is derived by comparing the event prediction of the algorithm to its true label (cluster or non-cluster). This allows four different combinations[2] (table 5.3):

- True Positive (TP), events that were classified correctly as cluster events.
- True Negative (TN), events that were classified correctly as non-cluster events.
- False Negatives (FN), events that were classified incorrectly as non-cluster events.
- False Positive (FP), events that were classified incorrectly as cluster events.

Next, a measure of success is defined. The goal of this measure is to quantify how well a classifier can predict cluster events and normal events compared to the true labels. Therefore a low number of FP and FN and a high number of TP and TN is favorable. The measure is defined by the sum of TP and TN and further modified by taking the negative

[2] In the case of binary classification (0,1).

of the sum and adding the sum of the false values.

$$M = \mathrm{TN} + \mathrm{TP}, \tag{5.19}$$
$$M^{-} = -M, \tag{5.20}$$
$$M_{\mathrm{final}} = M^{-} + (\mathrm{FP} + \mathrm{FN}), \tag{5.21}$$
$$= -(\mathrm{TP} + \mathrm{TN}) + (\mathrm{FP} + \mathrm{FN}). \tag{5.22}$$

The best parameter combination will then be a minimum value of this sum (the negative was taken, because optimizers favor determining a minimum).
In the following, a training set (section 5.2) is used to find the optimum parameter combination of `min_samples` and `delta_t`. For this, parameter grids of $[2, 10] \otimes [0.1, 0.4]$ s and $[2, 10] \otimes [0.4, 0.8]$ s are evaluated. In figure 5.12, the sum (equation 5.21) is shown for each corresponding parameter combination. The lowest misclassification rate is achieved by the combination `min_samples` = 7 and `delta_t` =0.8 s.

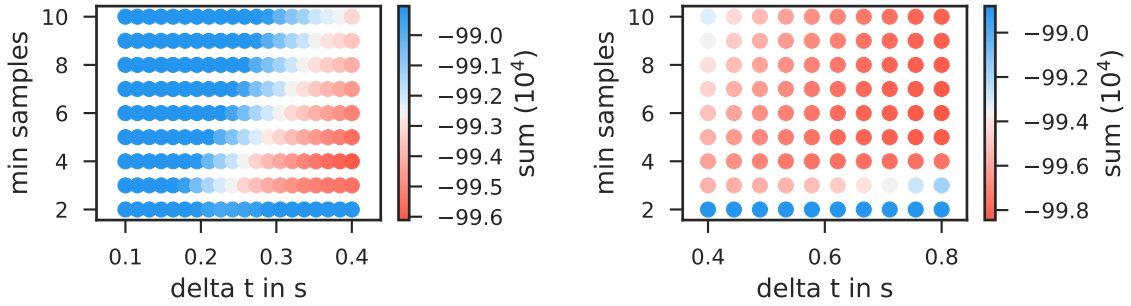


Figure 5.12.: **Parameter estimation previous algorithm**. Parameter grids and the measure of success (sum) are displayed. Smaller values of the sum indicate better performance (red areas) and bigger values worse performance (blue area). a): Smaller `delta_t` values. b): Higher `delta_t` values.

Table 5.3.: **Typical binary (0,1) classification outcome.** The predicted event labels by the algorithm are compared to the true simulation labels. The diagonal entries represent true classifications and the off-diagonal false ones.

	predicted 0	predicted 1
simulation 0	TN	FP
simulation 1	FN	TP

5.4 Unsupervised approach

Considering additionally the spatial correlation of cluster events produced by stored electrons makes the cluster algorithm more robust against misclassification. The process from the raw data to the tagged cluster events after application of the algorithm is split into two parts. The first one (subsection 5.4.1) describes the algorithm's working principle. The second one (subsections 5.4.2, 5.4.3) describes the data preparation, and parameter optimization of the algorithm.

5.4.1 Cluster algorithm (DBSCAN)

There are two types of clustering algorithms. Partitioning algorithms divide the space into k non-overlapping subspaces. All events within one subspace then belong to the same cluster. A prominent application is the k-means algorithm, where the space is divided into k Voronoi regions with k centers of gravity. Each event is then assigned to its closest center of gravity. Therefore the shape of these clusters is heavily restricted because the cluster space is convex. A major drawback of this type of algorithms is that they require the number of clusters as an input parameter. Therefore the user should have some prior knowledge of the dataset, which is often not the case.
On the other hand hierarchical clustering algorithms decompose the data space into a dendrogram, which is a tree representation where the space is split iteratively into smaller subsets until a termination condition is met. A native approach would be to calculate the distance for every point, resulting in a distance matrix

$$D_{ij} = d(q_i, q_j), \qquad 0 < i, j < n \tag{5.23}$$

and then grouping events with a distance lower than the threshold-distance together. However, the computational effort of this approach would scale quadratically with the number of events $O(n^2)$.
An alternative approach is density based and divides a k-dimensional space S into clusters and noise. A precise explanation of DBSCAN can be found in [EK96], here a more visual presentation will be given (figure 5.14a and 5.14b). At first, a random point p is selected. Then every point, which is reachable wrt. ϵ (the maximum distance between two points), and N_{min} (the minimum number of core points) to p is calculated. If p is a core point, points which are reachable within ϵ are a cluster, and if p is a border point the points are not a cluster. The algorithm then continues with a non-visited point.
In this thesis, the *scikit-learn* implementation of the DBSCAN algorithm is used [PV11, Dev18].

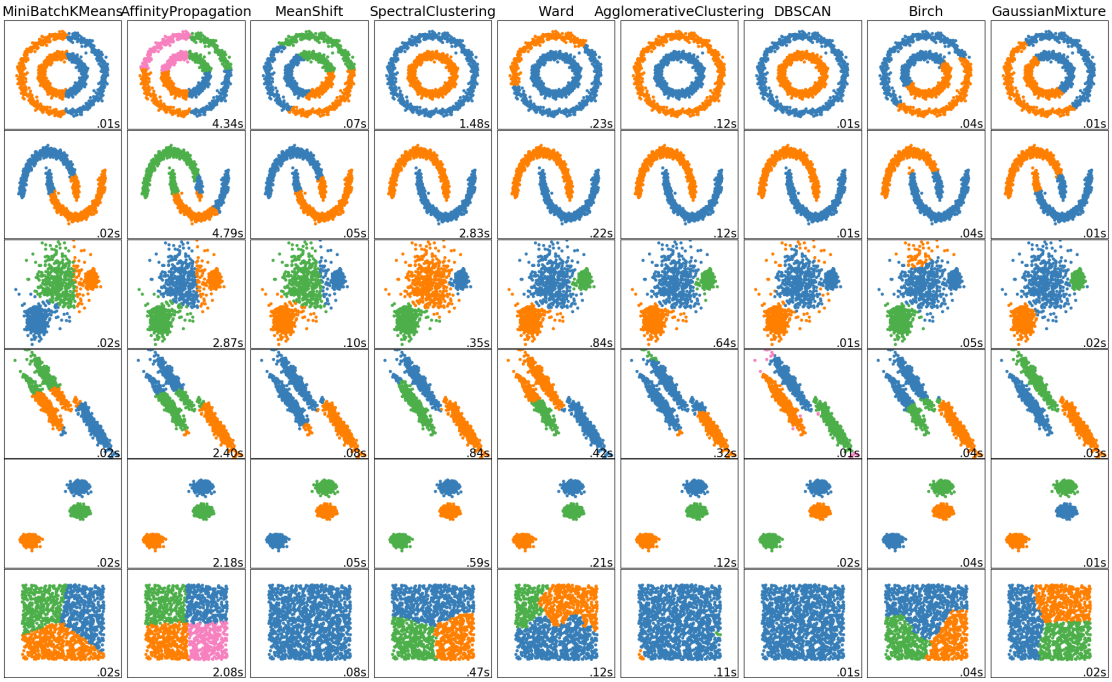
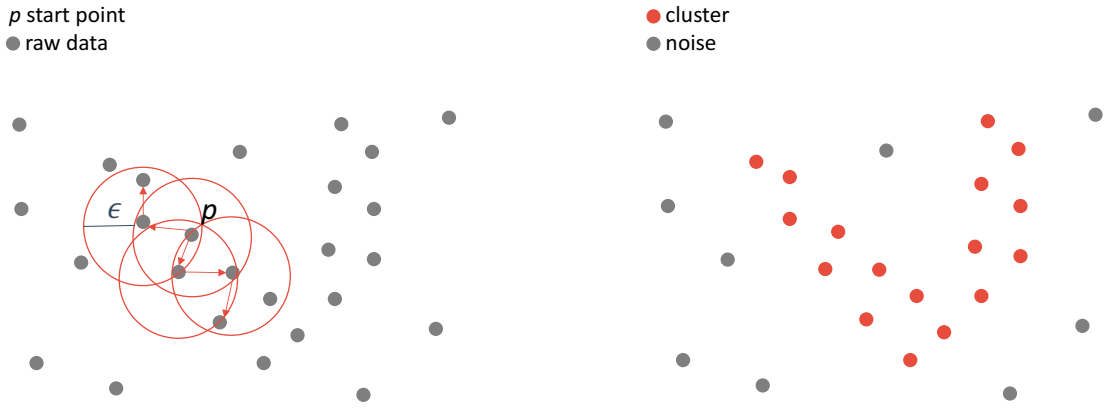


Figure 5.13.: **Overview of different clustering algorithms on toy datasets of the scikit-learn library** ([LNB18, PV11]). Although several algorithms may seem to be suitable for the specific use-case, a big advantage of DBSCAN is to separate clusters from noise.



(a) (b)

Figure 5.14.: **Working principle of DBSCAN algorithm.** a): Starting from a random point p, the space is checked for other points within distance ϵ (circle with radius ϵ). b): Result of the algorithm, the high density regions (red) are events within one cluster, other points are noise (grey).

5.4.2 Data preparation

The cluster algorithm uses two parameters as input variables (predictors):

- timestamp of the event, and
- radius of the pixel where the electrons hit the FPD.

While the former already comes as an output of BEANS[1], the radius has to be derived from the pixel which was hit. This is done by assigning the center of each pixel to (x, y) coordinates and calculating the radius via $r^2 = x^2 + y^2$.
However, these two variables have different scales (time of the measurement can be $\mathcal{O}\left(10^3\right)$ s versus radius $\mathcal{O}\left(10^{-2}\right)$ m) and most unsupervised cluster algorithms need the data to be scaled to have mean 0 and variance normalized to 1, otherwise they will have difficulties interpreting the input data correctly.
An overview of unsupervised clustering algorithms is given in figure 5.13. Most cluster algorithms have the disadvantage that the number of clusters have to be defined beforehand and serve as an input parameter to the algorithm (e.g. k-means clustering). This is not the case for the DBSCAN algorithm, which is similar to the previous algorithm, however extended towards multi-dimensionality. It also has a `min_samples` threshold and additionally a distance measure ϵ as input parameters. In particular the distance measure allows for further improvement compared to the previous algorithm. By deriving this parameter analytically from the pressure, it is possible to use this algorithm at lower pressures.
For a suitable distance measure, the user has several choices such as the Mahalanobis distance, Euclidean distance, correlation etc. [Dev18]. The Euclidean distance was chosen for simplicity. Therefore the maximum distance between two points considered to be in the same cluster is given by

$$d(t_{1,2}, r_{1,2}) = \sqrt{((t_2 - t_1)^2 + f \cdot (r_2 - r_1)^2)}, \tag{5.24}$$

$$\tag{5.25}$$

Here, f is a scaling factor transforming the unit of r to match the unit of t. Since the radii of the FPD pixels stretch from 0.00375 m to 0.045 m, differences in radius have a negligible impact on the distance measure (equation 5.24). As mentioned earlier, cluster events due to stored electrons also have a correlation in radius, therefore it does not make much sense to consider two consecutive events, where the first one hits the outer most ring and the second one the most inner ring, to be within the same cluster. For this reason, the radii have to be scaled to have bigger impact on the distance measure. Since the primary electrons have different cyclotron radii in the different spectrometers, this scaling can be set accordingly.

[1]Building Analysis Sequence (BEANS) is a data collection and analysis framework for KATRIN.

To make this algorithm dependent only on the pressure, this scaling factor is also derived from the mean time between two ionization events and set to

- MS: f=0.2·$l \cdot \frac{\langle t \rangle}{\Delta r_{\text{mean}}}$
- PS: f=0.4·$l \cdot \frac{\langle t \rangle}{\Delta r_{\text{mean}}}$

where Δr_{mean} is the mean average distance of the detector rings, which are not equally spaced. On the downside this puts a higher penalty on the inner rings, since they are broader, and a lower penalty on the outer rings. On the upside the algorithm is then only dependent on the pressure. Due to the higher magnetic field in the analyzing plane of the PS, electrons have smaller cyclotron radii (equation 5.3) compared to the main spectrometer and therefore a higher penalty term was chosen. To provide an input value for ϵ,

$$\epsilon = l \cdot \langle t \rangle , \qquad l = 7. \tag{5.26}$$

was chosen, because the distribution of times between two ionization processes may most likely be asymmetric, particularly having a tail towards large times for high primary electron energies. The derivation of the optimal value will be described in subsection 5.4.3. In figure 5.15 the scaling is visualized. Without scaling (blue) the algorithm would cluster every event into the cluster, even if the event is not related to a cluster. The scaling leads to a broadening of the data points (grey), forming more and less dense regions which can be better distinguished.

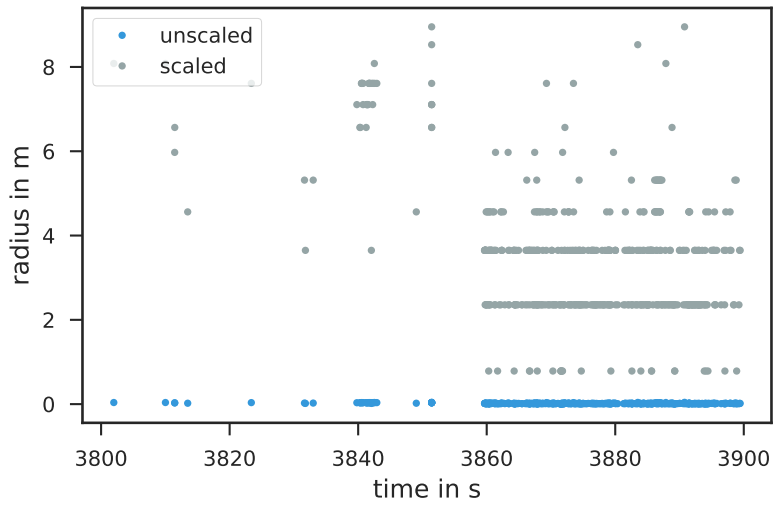
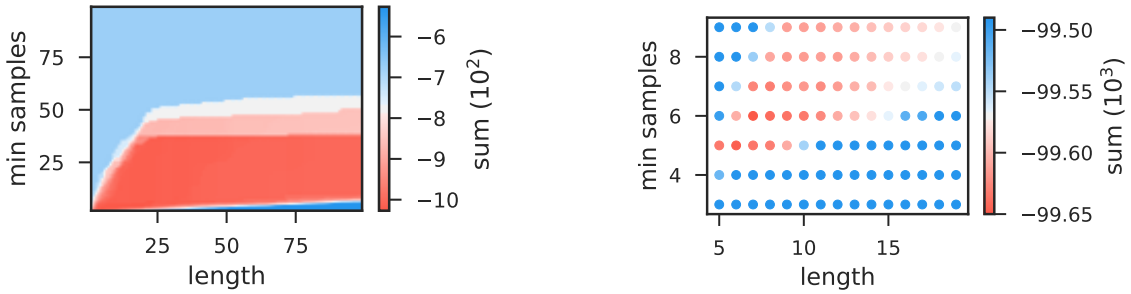


Figure 5.15.: **Representation of a cluster event in the event space of radius and time.** The blue points have unscaled radii. By scaling the radii (grey), the order of radii difference and time difference becomes similar, which makes cluster events more distinguishable from other background events. Unit of r for the grey events is not meters here, but seconds.



(a) (b)

Figure 5.16.: **Parameter estimation DBSCAN algorithm.** a): Large parameter grid of a small training data set. b): Small parameter grid of a larger training data set.

5.4.3 Derivation of optimal parameters

The two optimal parameters of the DBSCAN algorithm (initially ϵ and `min_samples`, later modified to `length` and `min_samples`) are derived. This is done in a brute-force approach, looping through every combination of `length` and `min_samples` ($[1, 100] \otimes [2, 100]$). The clustered data is then evaluated according to equation 5.21. A plot for the parameter combinations and the corresponding value of the sum can be found in figure 5.16a as well as a zoom in figure 5.16b. The parameter grids are interpolated for visualizing the trend of the sum, in reality the combinations are always of integer type. The minimum obtained is a `length` of about 7 and `min_samples` between 5-6, which corresponds to misclassifying only 691 events false, corresponding to a percentage of 691/999417=False/True=0.07 %.

5.5 Supervised approach

Distinguishing cluster events from normal background events is a classification problem. Either the event belongs to a cluster or not. There is a large number of classification algorithms, ranging from simple and easy interpretable methods such as Linear Discriminant Analysis (LDA), over the popular out-of-the-box methods such as Support Vector Machines (SVMs), and tree methods, to deep-learning methods such as neural networks. In this approach a Boosted Decision Tree (BDT) is used, as in first attempts the results looked more promising than the output from a convolutional neural network (CNN).
In the following, a quick introduction to tree-based methods will be given in subsection 5.5.1, followed by the application to simulation data and a performance evaluation in subsection 5.5.3. The performance evaluation is based on the simulation output described in section 5.1, which is used as a training sample.

5.5.1 Theory

The information summarized in this section is based on [JW17, HTF17].
The basic idea of tree based supervised learning methods is, to divide the predictor Space V into a number of simple regions by segmenting it according to splitting rules. Decision trees are suitable for both classification and regression because the principle is similar. Trees are grown upside down using recursive binary splitting, which means that by applying a split criterion the predictor space V is divided into two smaller, non overlapping subsets of V. In the next step these subsets are again divided into two smaller subsets. This process is repeated until a termination criterion is reached, such as the maximum number of observations within a node contains fewer than a threshold number of observations or the maximum depth of the tree is reached. These nodes are also called terminal nodes or leaves of the tree. The cut point s on each cut is chosen such that a measure of optimization is minimized. In the case of a regression tree, this is for example the residual sum of squares (RSS)

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \,, \tag{5.27}$$

where y_i is the i-th value of the variable to predict, and $\hat{y}_i = f(x_i)$ is the predicted value from the fit model f. In the case of classification the Gini index

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \tag{5.28}$$

or the cross entropy

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log(\hat{p}_{mk}), \tag{5.29}$$

is used, where $\hat{p}_{mk}$ is the proportion of training observations from the k-th class within the m-th region of the data. Figure 5.18 visualizes that both measures take on small values,

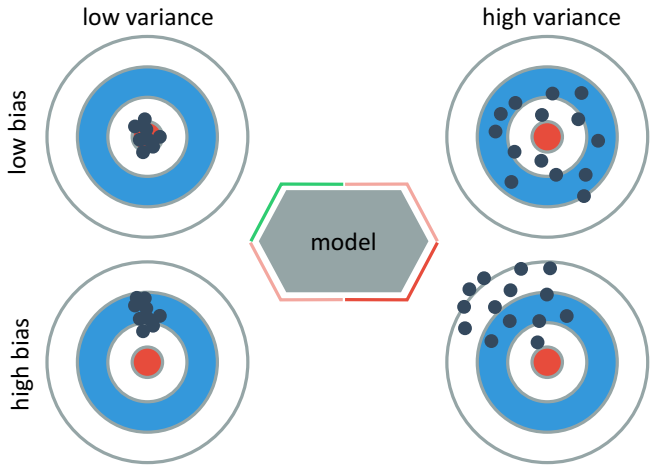


Figure 5.17.: **Bias-variance tradeoff**. Figure inspired by [For18].

if $\hat{p}_{mk}$ is either close to 0 or 1, meaning that this node has a high purity and contains predominantly observations with the same class label. Therefore the split is set at the position s, where these indices are minimized. The measure of the classifier used in this work is set to the Gini index G.

Bias-variance tradeoff

Models fit to data can suffer from error, which can be decomposed into

$$Err(x) = \left(\mathrm{E}\left[\hat{f}(x)\right] - f(x)\right)^2 + \mathrm{E}\left[\left(\hat{f}(x) - \mathrm{E}\left[\hat{f}(x)\right]\right)^2\right] + \sigma_e^2 \tag{5.30}$$

$$= \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{irreducible\ error} \tag{5.31}$$

where $\hat{f}(x)$ is an estimate of the model $f(x)$, which describes the true relationship $Y = f(X) + \epsilon$ [HTF17]. Bias error describes the offset of predicted values by the model to the correct values and the variance error the spread of the predicted values. In figure 5.17 a visual representation is given. If the true model is known and we have infinite data to repeatedly adapt the model, it is possible to reduce both errors to 0. However, in reality this is not possible and there is a tradeoff between bias and variance. If we optimize our model to represent the fit data too much, it is likely that this model's prediction accuracy will suffer if new data is spread differently than the fit one (overfitting, variance error), i.e. our model is too complex for the data. If we choose a model that is too simple and can therefore not extract all important features of the data, an offset of predicted values to the correct values is present (underfitting, bias error).

Since trees obtained by recursive binary splitting are likely to overfit the data, a regularization method similar to the Lasso regression [HTF17] is used. In a first step, a large tree T_0 is grown and then made smaller by cutting off branches and leaves, which is also called pruning. A subtree $T \subset T_0$ corresponds to each value of a non-negative value α,

such that

$$\sum_{m=1} G + \alpha \cdot N_T \tag{5.32}$$

$$= \sum_{m=1} \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) + \alpha \cdot N_T \tag{5.33}$$

is minimized. N_T is the number of terminal nodes in tree T. For $\alpha = 0$, this will result in only the training error. However, with increasing α, the measure favors smaller trees, as trees with larger N_T have higher measures. Therefore α represents a trade-off between training error and tree size. The parameter α can be obtained by using k-fold Cross-Validation (CV). k-fold CV is an approach to avoid overfitting the training data. The observations are divided into k non-overlapping regions (illustrated in figure 5.19). Then the algorithm is trained on all except the i-th sample, which is used to test the trained model. This process is repeated for all k regions. The total error is then estimated by the mean of all single errors

$$CV_k = \frac{1}{k} \sum_{i=1}^{k} \text{error}_i. \tag{5.34}$$

Since autocorrelation is present in the case of cluster data (event i is correlated with event $i - 1$), the k-fold CV splitting is not done randomly, but ordered in time.

Another method to avoid overfitting is boosting, which is similar to bagging. In bagging the training data is re-sampled (bootstrapped) and a different tree is fit for each of these independent samples. These trees are then combined to a single predictive model. The general idea of boosting is to combine the output of several weak classifiers (output only slightly better than random guessing) to powerful ones. In order to do so, a weak classifier is sequentially applied to repeatedly modified versions of the data. The prediction of these sequence of classifiers is then weighted, where the output from good classifiers is weighted more and vice versa [HTF17]. Boosting in the context of trees means that trees are grown sequentially, i.e. each tree is grown using information of the prior tree. The BDT algorithm has three hyper parameters:

- Number of trees B.
- Shrinkage parameter λ, which controls how fast the algorithm learns.
- The interaction depth d, which controls the interaction order of variables, because with $d = k$, k splits are performed and therefore k variables can interact at most.

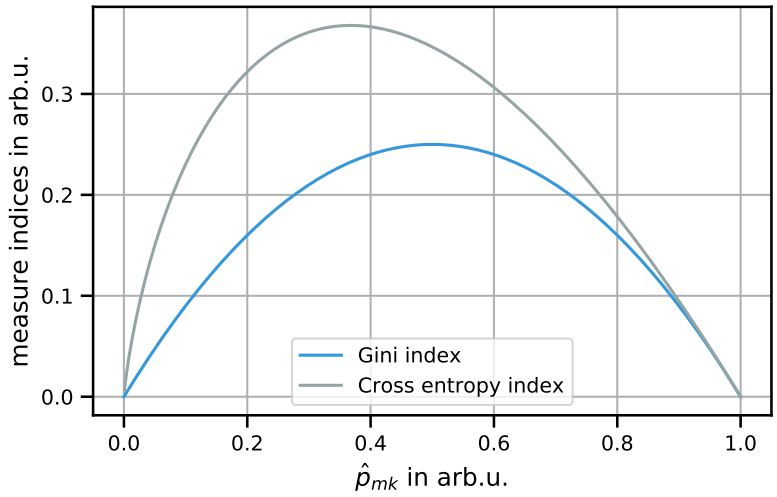


Figure 5.18.: **Gini and cross entropy index** as a function of node purity $\hat{p}_{mk}$. The indices take on small values for $\hat{p}_{mk}$ being either close to 0 or 1, which corresponds both to a high node purity.

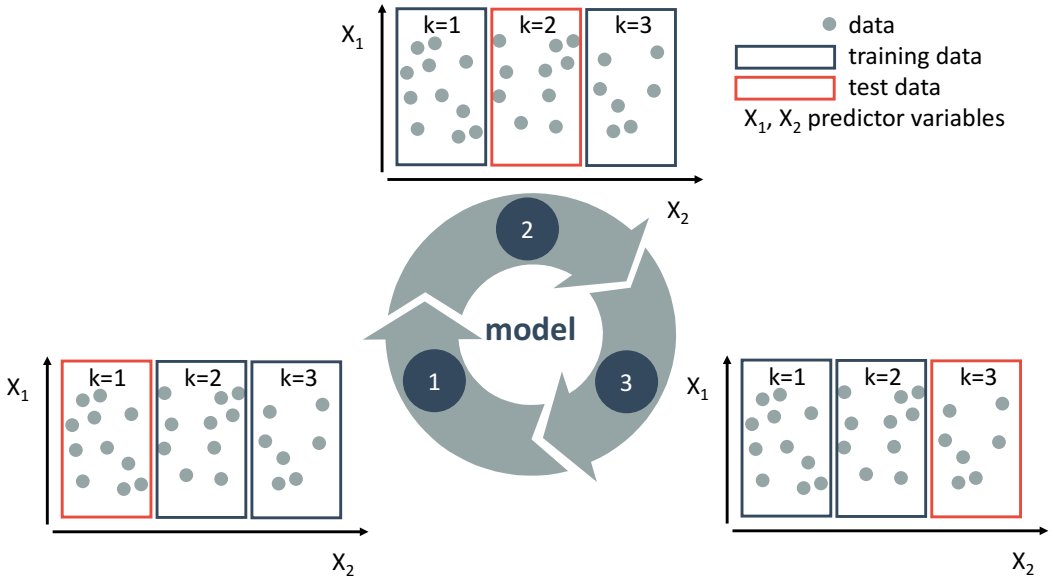


Figure 5.19.: **Schematic figure of a k(=3)-fold CV process.** The initial data is split into k non-overlapping regions of about the same size. Then the algorithm is trained on the data, leaving the first of the k-parts out for testing this trained algorithm and calculating a measure of performance. This procedure is then repeated for every k-parts. The final performance measure is then the mean of the k-measures. Most supervised learning algorithms suffer from overfitting the data, which makes the algorithm vulnerable to outliers. CV is a procedure to maintain a solid bias-variance tradeoff [HTF17].

5.5.2 Data preparation

In order to train the BDT model on data, the data needs to include at least one predictor. First, the data from the detector readout is cut to include only the timestamps (`time`) and channel (`pixel_number`). With the help of a channel → radius table, the channel is then transformed to radius (`r`). Furthermore, the time between two events is taken as a third predictor variable (`Interarrival`). To improve node purity and therefore prediction accuracy, the predictor space is enhanced by three more variables: The difference in radius of event k and $k+1$ (`r_diff1`), as well as difference in time and radius of event k and $k+2$ (`Interarrival2`, `r_diff2`). Therefore the BDT will map a six dimensional space onto a one dimensional space $V^6 \rightarrow V^1$, where in the training data set the binary output $\{0, 1\} \in V^1$ is also known to the algorithm. After the algorithm is trained, this output variable is then predicted using the input from the six predictor variables. The output values are probabilities of the event being a non-cluster (0) or cluster (1) event, $p \in (0, 1)$.

5.5.3 Application to simulation data at elevated pressure

The hyper parameters of the BDT are estimated by a grid search 3-fold CV. A first parameter grid is defined by `n_trees` $\otimes$ `learning_rate` $[1, 10, 100, 500, 1000] \otimes [1, 0.1, 0.05, 0.01]$ and for each combination a 3-fold CV on the simulation data is performed. The mean score is calculated for each parameter combination and the highest score was yielded by 500 trees and a learning rate of 0.01 (figure 5.20a). This grid search is then repeated with values closer to this first optimal parameter combination, which yielded 1000 trees and a learning rate of 0.00464 as the best combination (figure 5.20b). The final combination is used to perform a 3-fold CV on the simulation data and checked for the TP classification compared to the FP classification, which is shown in a so called Receiver Operating Characteristic (ROC) curve in figure 5.20c. This curve ideally should have a strong increase in the True Positive Rate (TPR) when the False Positive Rate (FPR) is low, which means the classification algorithm achieves a high classification rate of TPs while only allowing a small number of misclassifications with FPs. Furthermore, the performance of the algorithm can be evaluated by visualizing the decision score and the cut value of the classifier in figure 5.20. Another possibility to evaluate the performance is to look at the FPs and FN in dependency of various cut values (figure 5.20d). Usually it is desired to want them both as low as possible, however there can be cases when it is desirable to allow an increased misclassification to obtain an increased TPR. In the appendix (section A.2.3), the same results for the training data at nominal pressure can be found.

Although it has been shown that BDT classifiers are valid candidates for detecting cluster events, a minor drawback is the cluster reconstruction. As the output of the BDT is only 0 or 1, the affiliation of the single cluster events to specific clusters is not given. To reconstruct clusters, some assumptions about the time scale of clusters have to be made. By applying something similar as the Δt cut mentioned earlier to the labeled data allows an assignment of single cluster events to a cluster. While this might be easier for smaller storage times and therefore elevated pressure regimes, in nominal pressure regimes often multiple cluster events can happen at the same time due to the larger storage times [Sch14]. Another and probably more fruitful approach would be to first apply the

BDT for classification to the raw data and then feed only the labeled cluster data (1) into the DBSCAN algorithm, which has an automated implementation of cluster separation.

5.6 Comparison

To compare the three algorithms described in section 5.3, section 5.4 and section 5.5, a common evaluation dataset is used. Therefore, one dataset taken at high pressure and another one taken at low pressure have been created (section 5.2).

Elevated pressure

This dataset consists of a total of 1000108 events, where 142965 (≈15 %) are cluster events and 857143 (≈85 %) are normal background events. The misclassification rate without any usage of a cluster algorithm would therefore be 15 %. The performance of the different algorithms is summarized in figure 5.21. In terms of overall classification, all algorithms achieve a good performance (figure 5.21b). However, when looking at the total number of falsely classified events, small differences become visible. The lowest misclassification rate is achieved by the DBSCAN algorithm, followed by the established algorithm and lastly the BDT.
Although the established algorithm still achieves a good result, it has to be kept in mind that this dataset represents the background at elevated pressure and the hyper parameters were optimized using a training data set. This makes the distinction between cluster events, which have small time differences between two ionization events at high pressures, and normal background events, which have higher inter-arrival times, possible. The new algorithms do not face this pressure problem and are therefore far more flexible. In case of the DBSCAN algorithm, the scaling of the radii is indirectly dependent on the pressure and in case of the BDT, the pressure can simply be used as a further predictor parameter for training. Finally, the DBSCAN algorithm also outperforms the previous algorithm. In the following analyses (chapter 6), cluster events at an elevated pressure are tagged using the DBSCAN algorithm with optimized hyper parameters.

Nominal pressure

In the following analyses (chapter 6), cluster events at nominal pressure are tagged using the DBSCAN algorithm with optimized hyper parameters. Although the BDT outperforms the DBSCAN at nominal pressure, the DBSCAN is chosen due to the ease of automatic cluster separation.

Summary

In this chapter, it has been shown that supervised and unsupervised statistical methods are valid tools for identifying cluster events. In fact, due to the unique composition of the training data set, the hyper parameters of the previous algorithm could be optimized to achieve better performance in cluster detection as well. It has been shown that a larger time window together with a higher cluster size threshold achieve a better performance than just setting the time window cut where the inter-arrival times distribution is distinguishable from the normal background distribution. However, both at elevated as well at nominal pressure, this conventional algorithm is outperformed by the new algorithms described in section 5.4 and section 5.5. These new algorithms therefore represent a robust and sensitive method to detect cluster events due to tritium β-decay within the spectrometers. While the DBSCAN algorithm is by now the most convenient, the BDT has the most potential. To unlock the full potential of this supervised learning algorithm, a well suited predictor variable choice, as well as a clever cluster reconstruction process are necessary.

As an outlook, a background data analysis workflow (figure 5.23) could implement these new cluster detection methods in an automatic way. First, the pressure from either one of the spectrometers is read out as slow-control data for the runs to be analyzed, then the average of these pressure values is taken and fed into the algorithm. Together with the FPD data the algorithm then can detect all cluster events in the dataset automatically.

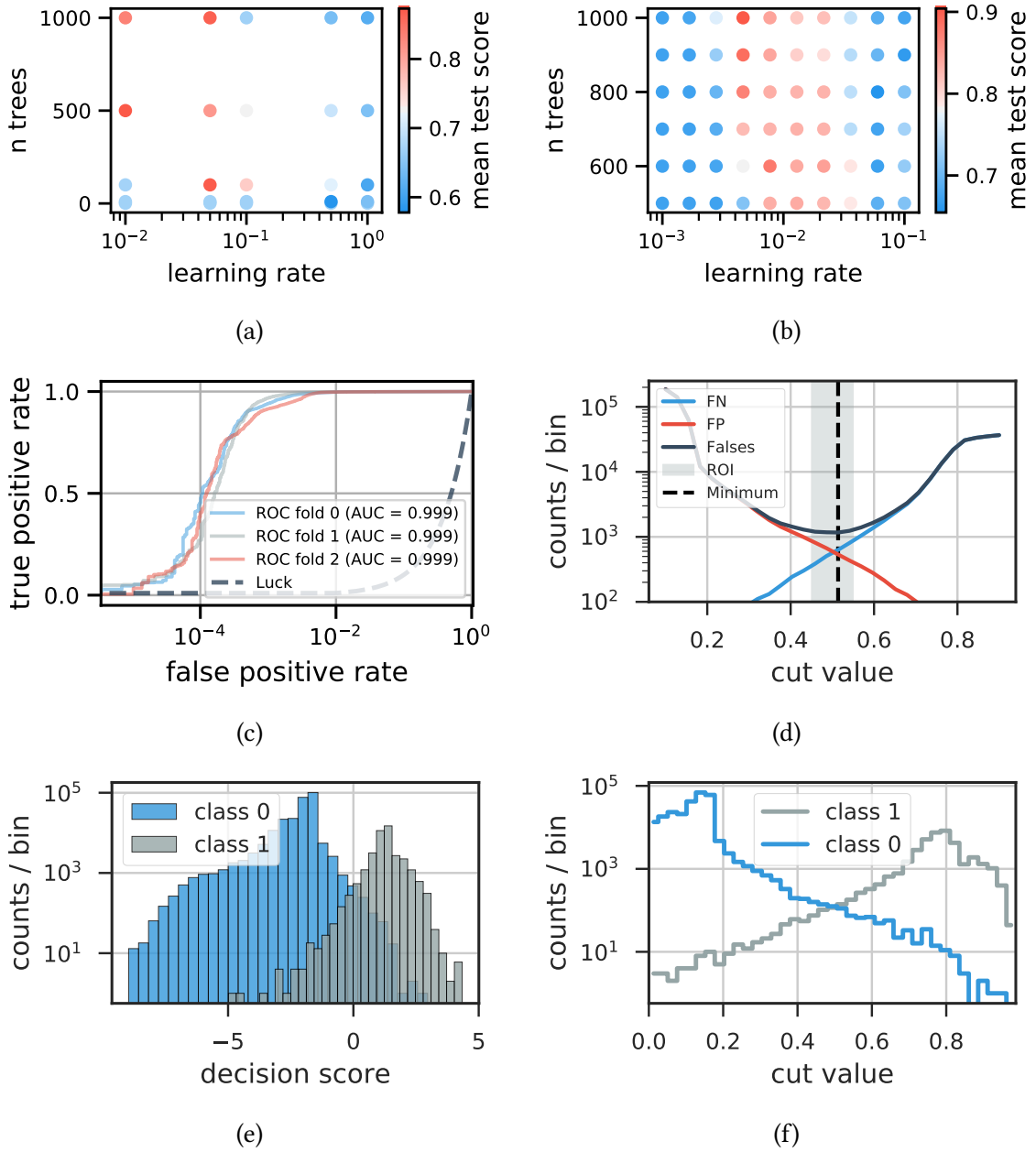
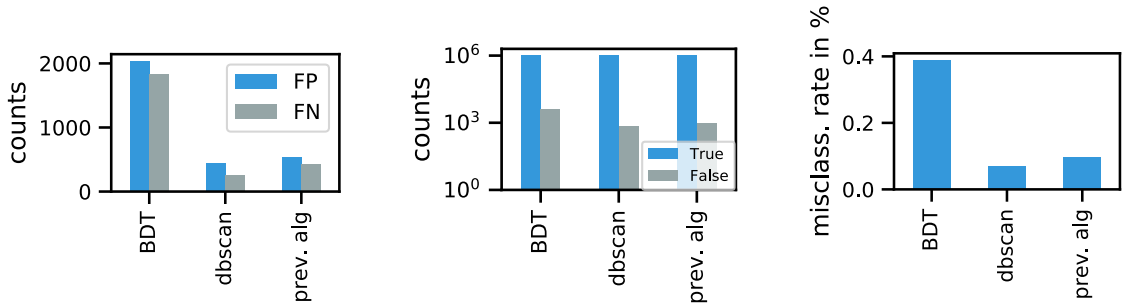
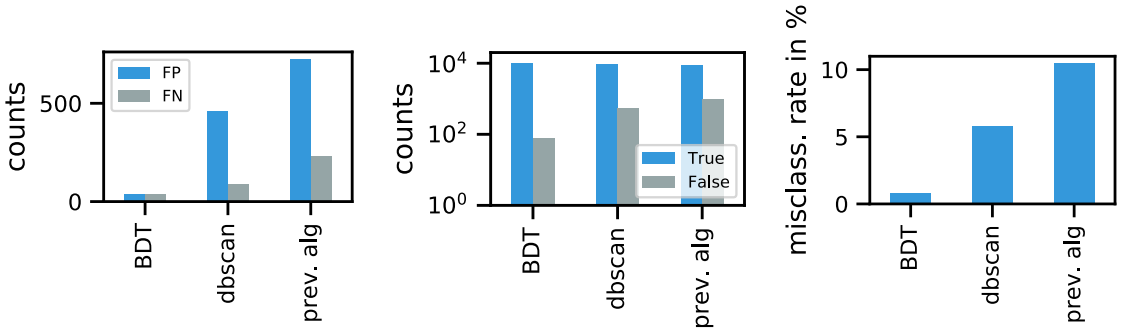


Figure 5.20.: **Performance plots of the BDT classifier.** a), b): Heatmap of different parameter grids, to visualize classification precision for different parameter configurations. c), d): ROC curve for 3-fold CV in c) and false classifications in dependence on the cut value in d). The ROI indicates region of favored cut. e), f): e) shows the decision score of the classifier and the probability value cut figure f) shows if the probability, as an output of the classifier for each event being a normal background event (close to 0) or a cluster event (close to 1), would be cut at a specific value, how many of these observations would be true positives.



(a) (b) (c)

Figure 5.21.: **Algorithm comparison at elevated pressure.** a): Performance of the algorithms in terms of FPs and FNs. The DBSCAN algorithm has the overall lowest numbers of false events, followed by the previous algorithm and the BDT. b): Performance of the different algorithms in terms of true and false events. All algorithms have a good performance, because the number of true events is several orders higher than the number of false events. Differences are only slightly visible here. c): Performance of the algorithms in terms of misclassification $\left(\frac{\text{False}}{\text{True}}\right)$. All algorithms have a low misclassification rate, but the DBSCAN algorithm again performs best, followed by the previous algorithm and the BDT.



(a) (b) (c)

Figure 5.22.: **Algorithm comparison at nominal pressure.** a): Performance of the different algorithms in terms of FPs and FNs. The BDT has the overall lowest numbers of false events, followed by the DBSCAN and the previous algorithm. b): Performance of the algorithms in terms of true and false events. c) Performance of the different algorithms in terms of misclassification $\left(\frac{\text{False}}{\text{True}}\right)$. The misclassification rates vary far more than in the case of an elevated pressure. Here the BDT achieves the best performance, outreaching the previous algorithm by one order of magnitude. The DBSCAN algorithm lays in between these two, with a misclassification rate of about 6 %.

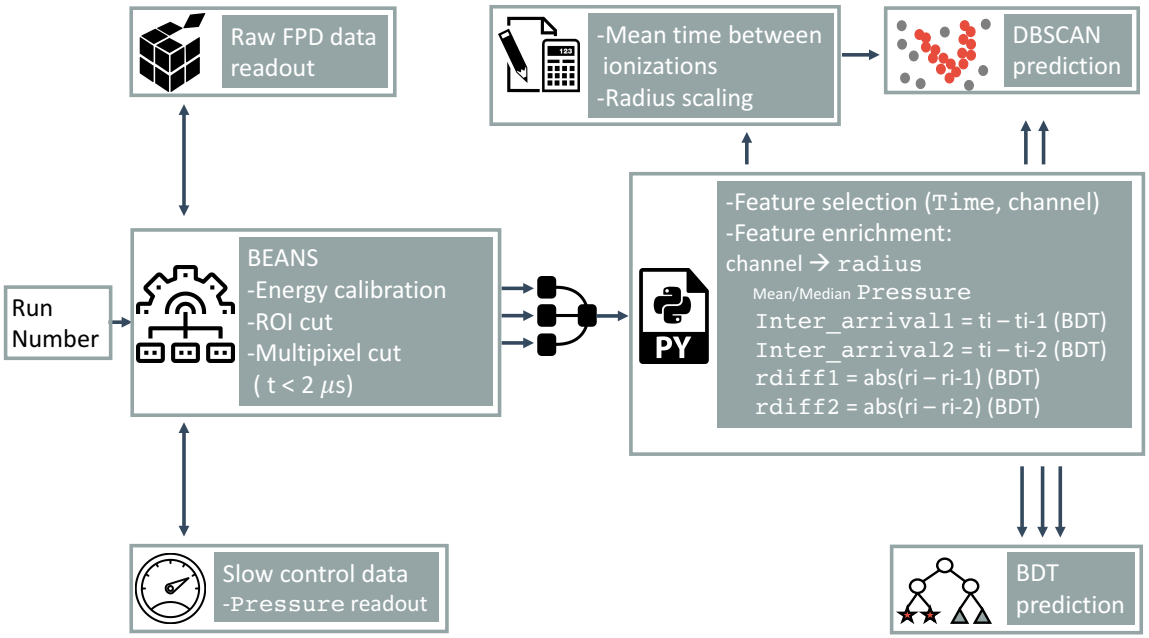


Figure 5.23.: **Cluster analysis workflow.** Data is accessed via the RunNumber. BEANS collects and processes the data. Then the data is read in Python, selecting the required predictors (`time` and `pixel/channel`). FPD Channel is converted to `radius` and the mean/median `pressure` calculated. Depending on the algorithm, the data is either enriched with four additional predictors (BDT) or the mean time between ionization events is calculated from the `pressure` and the `radius` scaled accordingly in the case of the DBSCAN.

6 Cluster identification

6.1 Characteristics of cluster events

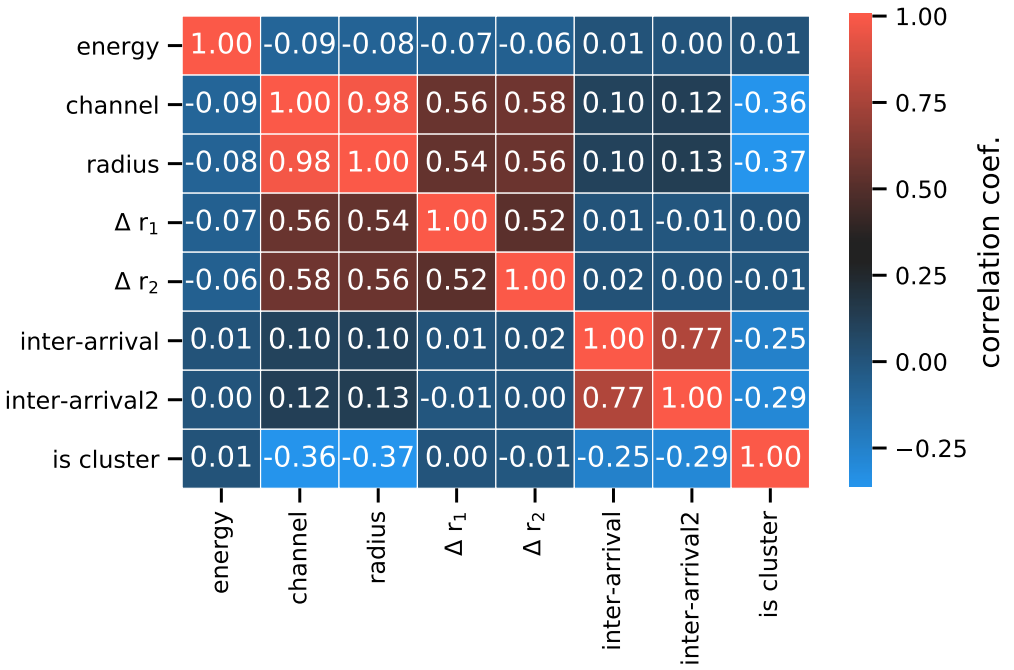
In order to explain the predictor variable choice of the BDT in section 5.5, both the BDT and DBSCAN algorithms are applied to measurement data with the run numbers 35160 - 35171, which is a measurement of the PS background at elevated pressure p=3×10^{-9} mbar. The variables `time`, `radius`, $\Delta r_1 = r_i - r_{i-1}$, Δr_2, `Interarrival`, and `Interarrival2` were used to predict cluster events. Figure 6.1 shows the pairwise correlation coefficients. The high correlation coefficient of `channel` and `radius` is natural, when considering how the `radius` is derived from the pixel map. However, by using both `channel` and `radius`, the classifier does not gain any new information. Therefore the `channel` was dropped as a variable. Additionally, the `energy` was dropped as it has no correlation with either the output variable (`is_cluster`) or any of the other predictor variables and is consequently not helpful for the classifier. The remaining variables either have a small correlation with the output variable, or correlation among the other predictors.

For further investigations, the pairwise joint-distribution plots of the predictor variables were investigated (figure 6.2). Here it can be seen that cluster events tend to have high peaks in the predictor variables Δr_1, Δr_2 around 0. This feature will be investigated in the following more precisely. Recalling the change in the exponential distribution of cluster-events as the current cluster-event identification characteristic, it will be shown that this sharp peak in the radius difference of consecutive events is also a strong indicator.

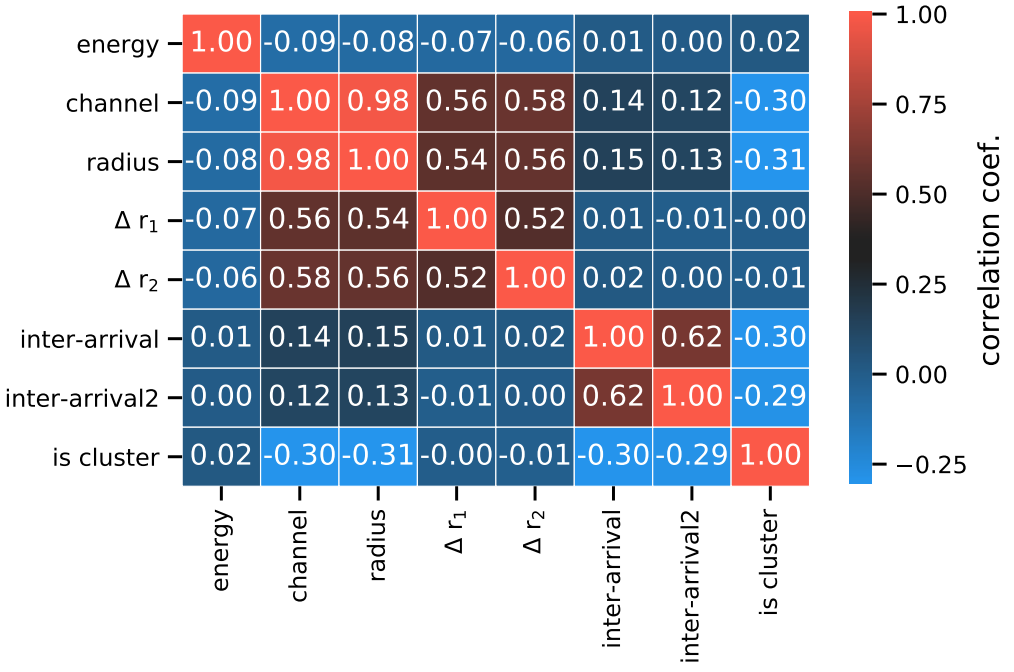
In order to motivate the distribution of Δr_i, it is assumed that the set of background events $\{ x \}_{i=1}^{i=n}$ can be regarded as i.i.d. (independent, identically distributed) random variables. Taking a look at the distribution of radii (figure 6.3a) it can be considered as either an exponential distribution ($r \sim \text{Exp}(\lambda)$) or as a Gaussian distribution ($r \sim N(\mu, \sigma)$). Here an exponential distribution was chosen. If r is distributed exponentially (Gaussian), the difference of two independent, identical exponential distributions is given by the Laplace distribution La(λ) (normal distribution $N(\mu_2 - \mu_1, \sigma_2^2 + \sigma_1^2)$). The proofs can be found in the appendix A. Under the assumption that normal background events are uncorrelated and independent, their distribution should follow a La(λ) function. However, cluster events are correlated and the characteristic ring structure causes a narrow peak around zero in the distribution of radius differences. The sum of these two distributions can therefore be fit with a Laplace function plus a narrow Gaussian (figure 6.3b). After application of both

supervised or unsupervised cluster-cut, the narrow peak around zero is much smaller and therefore indicates an effective reduction of background events due to clusters (figure 6.3c, 6.3d).

To show that the new algorithms also work at nominal pressure, the distribution of radii after the cluster cut of en elevated pressure measurement and nominal pressure measurement (run number 35362 - 35378) was investigated and is shown in figure 6.3e and figure 6.3f. It is clearly visible that the excess in the peak value is reduced to fit the distribution more naturally, both for the measurement at elevated pressure and nominal pressure. This is a strong indicator that the new algorithms also work at nominal pressure. For pairwise plots of the nominal pressure measurement, the reader may have a look at figure 6.4.



(a)



(b)

Figure 6.1.: **Predictor variable correlation coefficients**. a): after the cluster cut is applied through the DBSCAN algorithm. b): after the cluster cut applied is through the BDT algorithm. Due to the different approaches, the correlation matrix differs slightly. Furthermore, the inter-arrival time after the DBSCAN cluster cut is reprocessed to be the time between two consecutive events with the same label. Therefore, its correlation also differs slightly from the one after the BDT cluster cut.

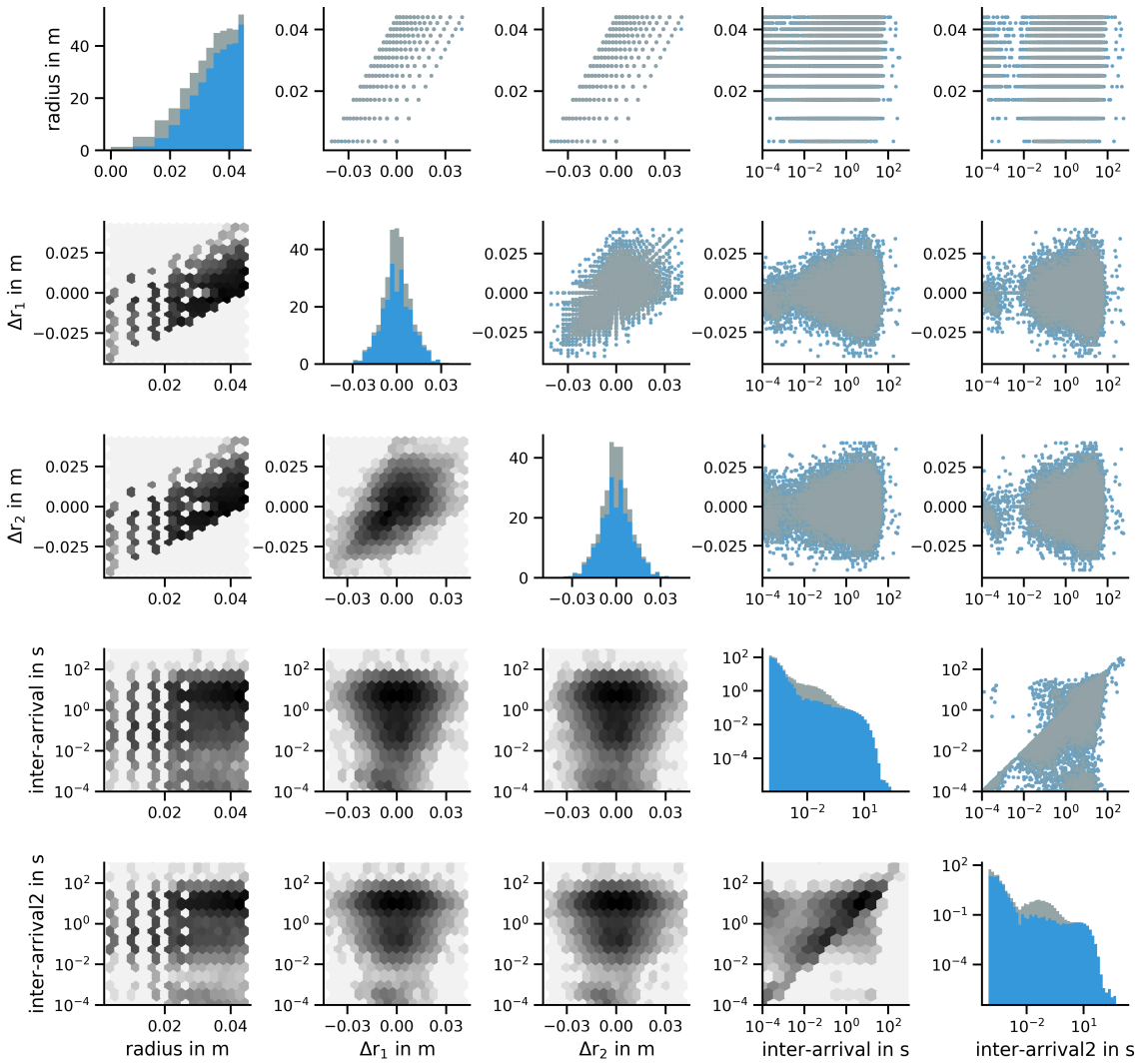


Figure 6.2.: **Pairwise plots of the predictor variables** at elevated pressure. The diagonal elements are histograms (density normalized to 1), the upper elements of the diagonal are scatter plots, where blue labels indicate non-cluster events and grey labels cluster events. On the lower part of the diagonal, two dimensional histograms are shown to illustrate dense data regions, with black color indicating a high number of counts in this bin and white color a low number of counts.

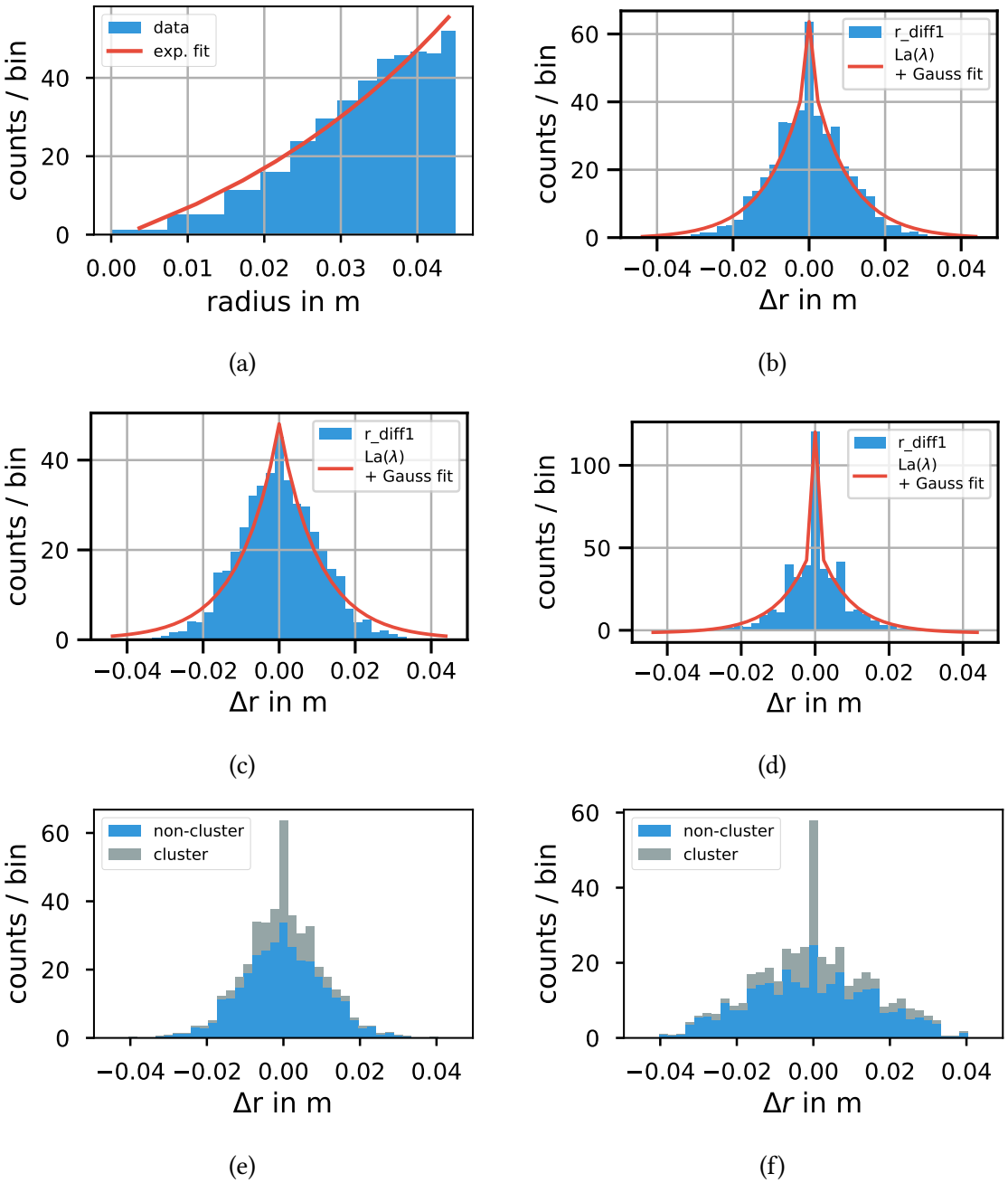


Figure 6.3.: **Distribution of radii and difference of radii**. All histograms are normalized to 1. a): As the widths of the detector rings are not equally spaced, the normalization has a penalty on the broader inner rings, which causes the distribution of the radii to follow an exponential (or half-Gaussian) distribution. b): Distribution of radius differences of all events. A clear excess at 0 is visible. c) and d): Distribution of radius differences after DBSCAN cluster cut. Non-cluster events on the left and cluster events on the right. e), f): Stacked histograms of non-cluster and cluster events at elevated pressure e) and nominal pressure f).

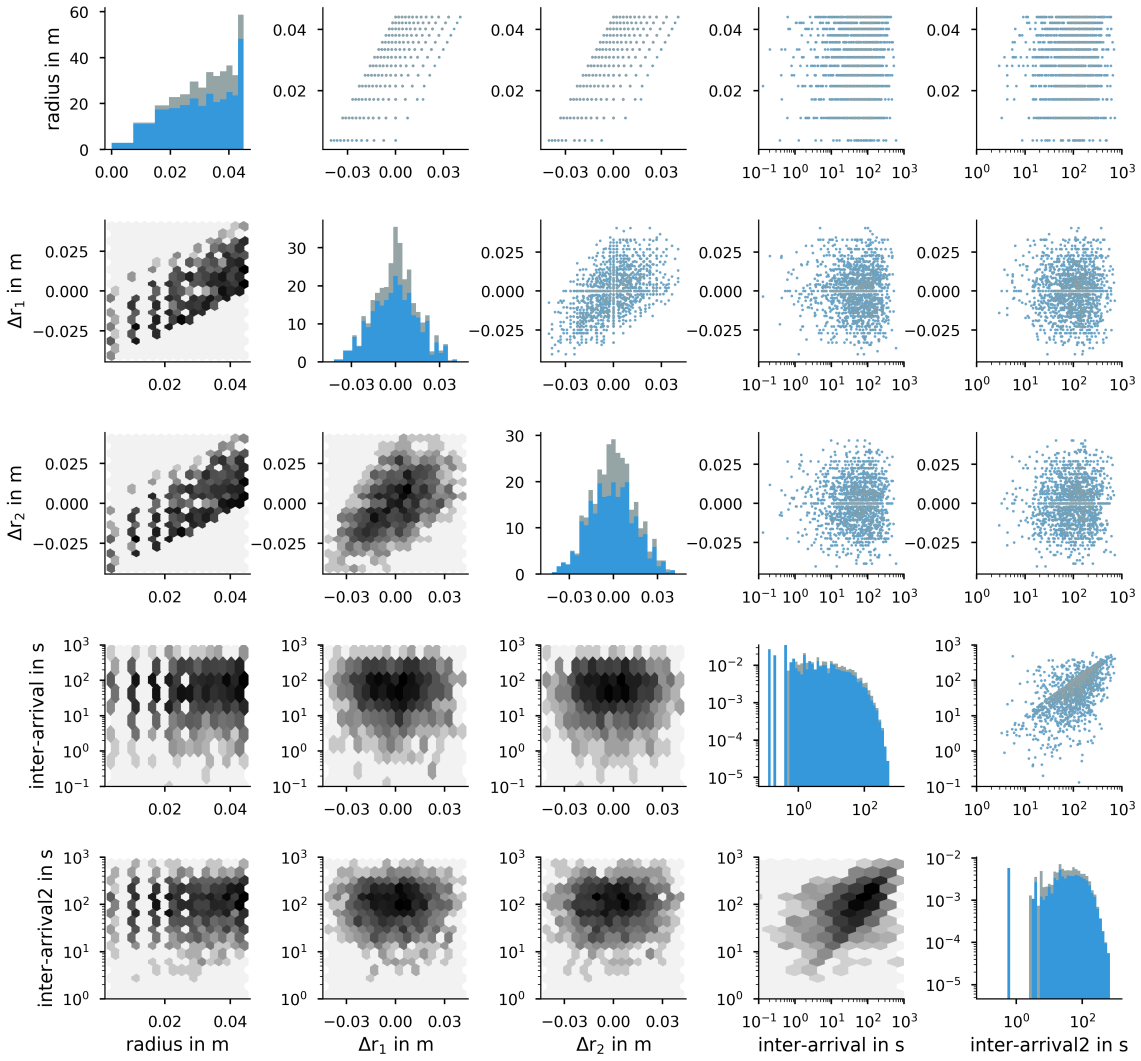


Figure 6.4.: **Pairwise plots of the predictor variables** at nominal pressure. The diagonal elements are histograms (density normalized to 1), the upper elements of the diagonal are scatter plots, where blue labels indicate non-cluster events and grey labels cluster events. On the lower part of the diagonal, two dimensional histograms are shown to illustrate dense data regions, with black color indicating a high number of counts in this bin and white color a low number of counts.

6.2 Influence of NEG material on radon cluster rate in the main spectrometer

The background rate in the MS with ≈400 milli-counts per second (mcps) currently exceeds the limit for the designed KATRIN sensitivity of 0.2 eV. While the remaining background component is still under investigation (section 4.5), radon emanating from the PS's NEG pumps into the MS's volume may have a non-negligible influence on the background rate. In order to estimate this impact, a reference background measurement with opened valve between PS and MS and another one with closed valve were conducted, both at elevated pressure p=3.1×10^{-9} mbar.
With the the help of the tuned cluster algorithms, the cluster rate of both measurements is determined and compared. This comparison will give insight into the overall radon contribution of the PS's NEG material to the MS background.

Open valve

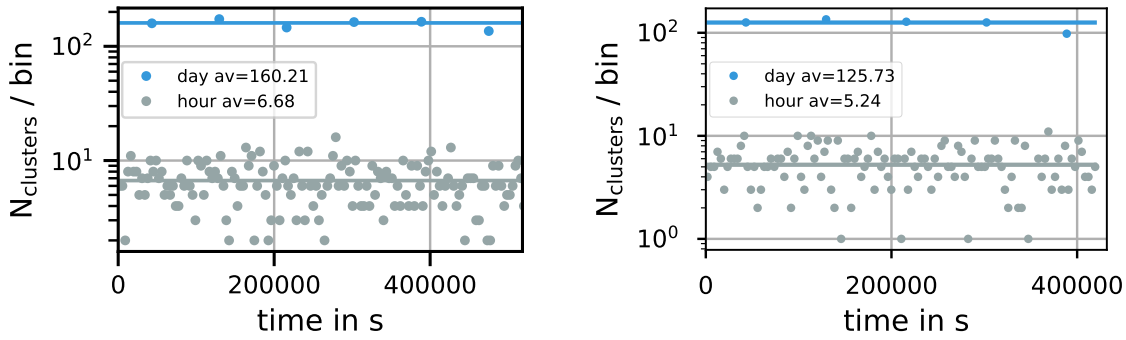
When the valve is opened, an overall rate of 432.29±0.85 mcps is observed. Applying the DBSCAN algorithm with a scaling factor of 0.2 (MS scaling) instead of 0.4 (PS scaling) at elevated pressure yields a cluster rate of 17.85±0.17 mcps, which corresponds to a fractional contribution of 4.13 % from radon to the non-radon background. The average number of clusters per day is 160.21±11.65 (figure 6.5). In figure 6.6a the rate with different binnings, the projection of the rate and a fit, the residuals of the fit and the FPD pixel distribution over all runs is shown. After application of the DBSCAN cluster cut the rate distribution is more Poissonian (figure 6.6b). For detailed information about cluster size, cluster duration, and FPD maps of clusters have a look at appendix figure A.5.

Closed valve

With the valve closed, an overall rate of 421.51±1.00 mcps is observed. Applying again the DBSCAN algorithm for cluster cuts reveals a cluster rate of 12.38±0.17 mcps, which corresponds to a fractional contribution of 2.94 % from radon to the non-radon background. The average number of clusters per day is 125.73±12.74 (figure 6.5). In figure 6.7a the rate with different binnings, the projection of the rate and a fit, the residuals of the fit and the FPD pixel distribution over all runs is shown. After application of the DBSCAN cluster cut the rate distribution is more Poissonian (figure 6.7b). For detailed information about cluster size, cluster duration, and FPD maps of clusters have a look at appendix figure A.6.

Table 6.1.: **Influence of NEG material on radon cluster rate in the MS.**

Valve	Rate in mcps	Cluster rate in mcps	clusters per day
open	432.29±0.85	17.85±0.17	160.21±11.65
closed	421.51±1.00	12.38±0.17	125.73±12.74

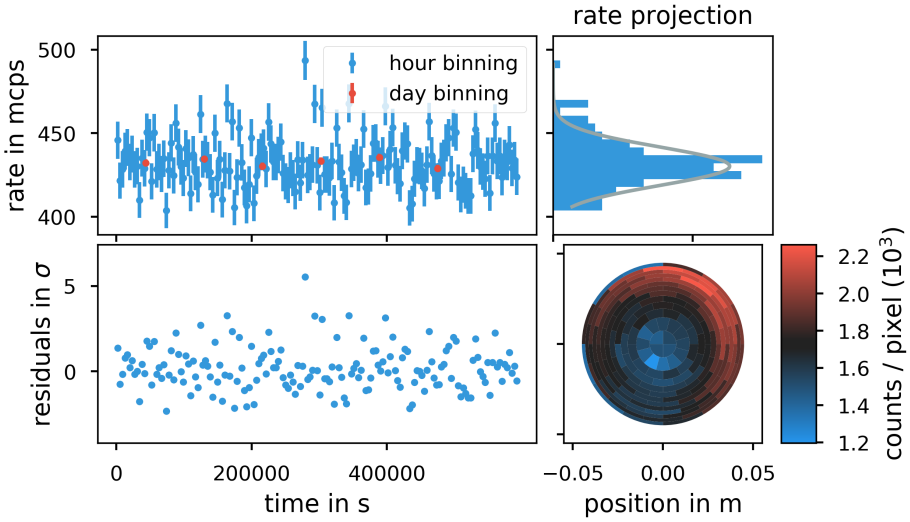


(a) (b)

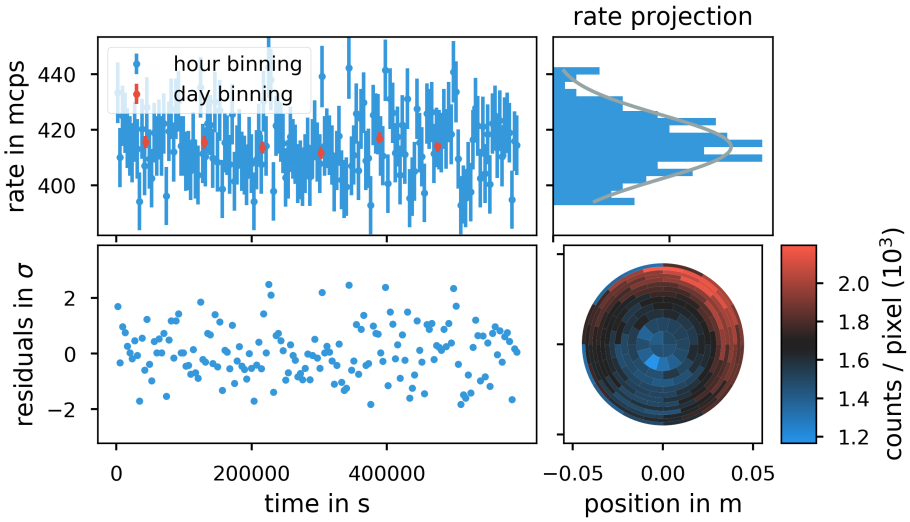
Figure 6.5.: **Number of clusters per day (hour) in the MS.** a): Valve is opened. b): Valve closed.

Summary

The information for opened and closed valve is summarized in table 6.1. Removal of the NEG material in the PS would lead to a decrease of the background rate in the MS by about $17.85 - 12.38 = 5.47 \pm 0.34$ mcps, which is about 1.5 % of the total background rate in the MS. However, removing the NEG material from the PS would potentially lead to an overall increased background rate in the PS, as the effective pumping speed is two orders of magnitude smaller than before. This may have a negative impact on the UHV conditions and therefore increases the background rate. At the same time the contribution due to radon decay induced cluster events would be smaller. This can further increase the sensitivity on tritium decays in the PS, which also would leave a cluster signature on the detector.

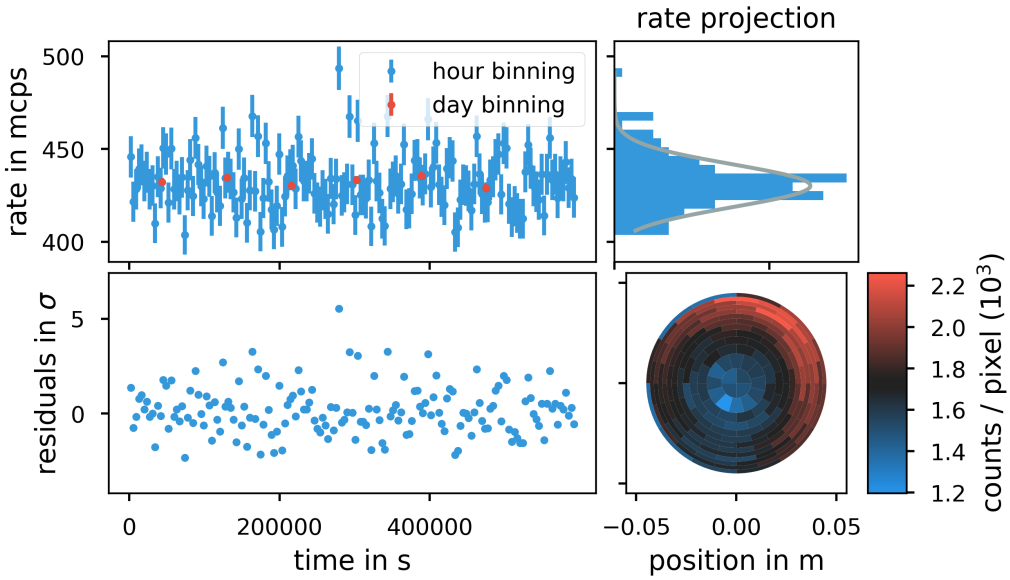


(a)

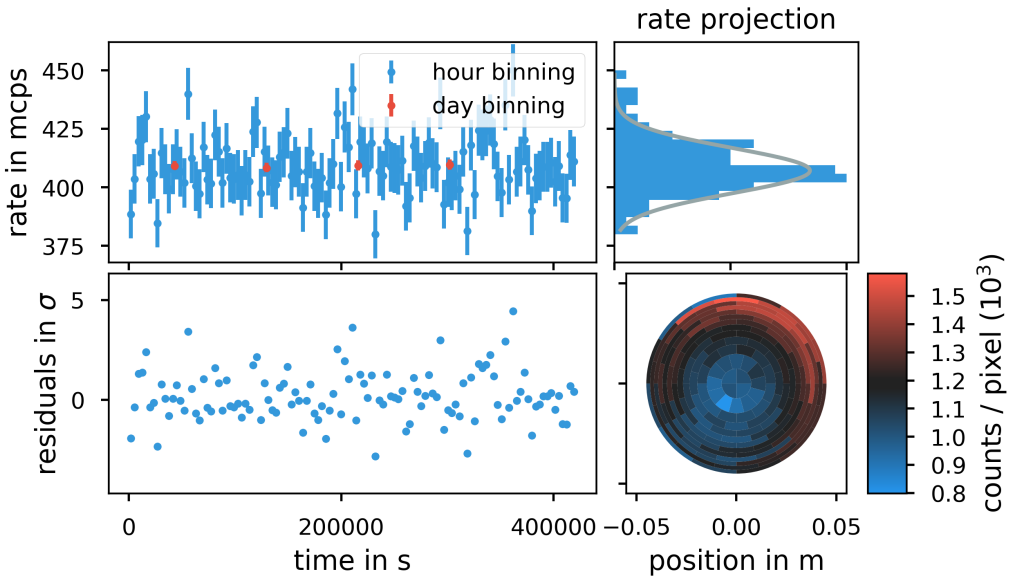


(b)

Figure 6.6.: **Summary plot of runs 35435 - 35476**, with bin sizes of one hour (blue) and one day (red). The projection of the hour binned rate is shown in the top right corner with a Gaussian fit in grey. Bottom left shows the residuals between fit and data points and bottom right the pixel distribution for all runs. a): Before cluster cut and b) after cluster cut.



(a)



(b)

Figure 6.7.: **Summary plot of runs 35477 - 35506**, with bin sizes of one hour (blue) and one day (red). The projection of the hour binned rate is shown in the top right corner with a Gaussian fit in grey. Bottom left shows the residuals between fit and data points and bottom right the pixel distribution for all runs. a): Before cluster cut and b) after cluster cut.

6.3 Sensitivity study with the pre-spectrometer on tritium decays

A major goal of this thesis is to find sensitive methods for tritium detection within the spectrometer volumes, in order to enable safe long term tritium operation. A general approach would be to just take regular background measurements before and at some point during tritium operation and then compare the measured rates of both. If they differ significantly, it may be due to tritium. However, in the previous chapter, the characteristic ring pattern on the detector of stored electrons due to tritium have been utilized to derive a more sensitive method.
In this chapter, the result of different background measurement settings will be presented which have a different significance on the tritium sensitivity. An overview of the various measurements, which have been conducted can be found in the appendix (figure A.14).
Tritium can reach the PS either as ions or molecules (most likely HT, minimal fraction of T_2). The ions are expected to be accelerated towards the downstream cone electrode and be implanted there or hit the hull. Although the tritium ions are implanted there, they are neutralized and can evaporate back into the volume.

6.3.1 Estimation of sensitivity

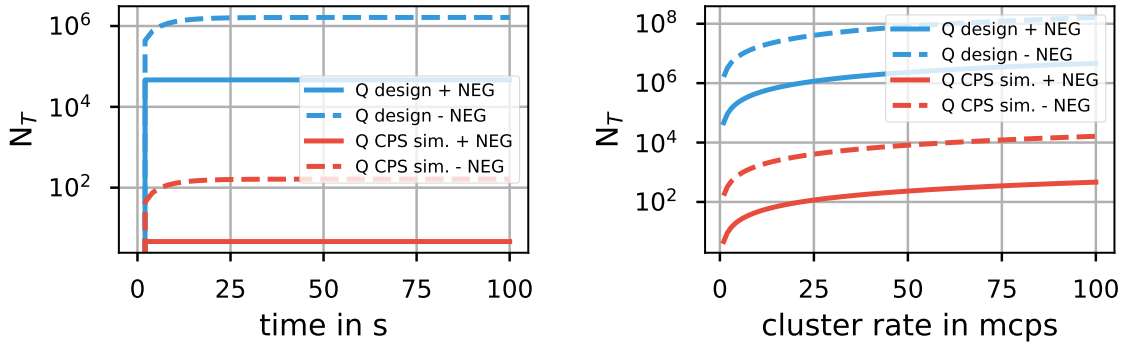
In the following measurements cluster rates will be determined. These rates are then used to estimate a sensitivity to tritium decays within the flux volume.
The number of tritium atoms within the PS volume depends on the gas flow from the CPS into the PS, for which the design report states a limit of

$$Q^{\mathrm{CPS}\to\mathrm{PS}} = 2.5 \cdot 10^5 \frac{1}{\mathrm{s}}, \tag{6.1}$$

under the assumption of a reduction factor of 10^{14} between WGTS and PS1 magnet. However, recent CPS simulations have yielded a better performance of the CPS in terms of tritium adsorption [FR18]. Instead of the reduction factor of 10^7 stated in [KAT05], a reduction factor of 10^{11} is expected. This leads to a combined reduction factor of 10^{18} between WGTS and the PS1 magnet. Therefore, the gas flow is expected to be four orders smaller than the one in the design report, $Q^{\mathrm{CPS}\to\mathrm{PS}}_{\mathrm{sim}} = 25\frac{1}{\mathrm{s}}$
The main molecule considered here will be HT, which gets pumped out by TMPs or the NEG material, but it can also decay. The rate change of tritium atoms within the PS is therefore given by the differential equation

$$\frac{dN}{dt} = -\underbrace{\lambda_T N(t)}_{\text{decay}} - \underbrace{\frac{S^{\mathrm{PS}}}{V^{\mathrm{PS}}} N(t)}_{\text{pumped out}} + \underbrace{Q^{\mathrm{CPS}\to\mathrm{PS}}}_{\text{const. gas flow into PS}} \tag{6.2}$$

$$= -(\lambda_T + k)N(t) + Q^{\mathrm{CPS}\to\mathrm{PS}}, \text{ with}, k = \frac{S^{\mathrm{PS}}}{V^{\mathrm{PS}}} \tag{6.3}$$



(a) (b)

Figure 6.8.: **Number of tritium atoms.** a): Number of tritium atoms over time. b): Number of tritium atoms for a given cluster rate. Equation 6.9 was used and scaled with the cluster rate in mcps.

which is solved by

$$N(t) = \frac{Q^{\mathrm{CPS}\to\mathrm{PS}}}{\lambda_T + k}\left(1 - \exp\left(-(\lambda_T + k)t\right)\right). \tag{6.4}$$

S^{PS} here denotes the effective pumping speed of combined NEG and TMP pumps (table 6.2) and $V^{\mathrm{PS}} = 8.5\,\mathrm{m}^3$ the PS volume. The saturation of this function is the maximum number of tritium atoms in the PS

$$N_{\max} = \frac{Q^{\mathrm{CPS}\to\mathrm{PS}}}{\lambda_T + k} = 46423. \tag{6.5}$$

The maximum number is reached nearly instantly, which can be seen in figure 6.8a. The background rate to which this amount of tritium corresponds to, is given by

$$R_T(t) = N(t) \cdot \lambda_T \cdot \mathrm{E}\left[n_e(E)\right] \cdot R_{\mathrm{V(flux)}}, \tag{6.6}$$

because the radioactive decay of tritium follows an exponential distribution and λ_T is the expectation value of this distribution. Furthermore, $\mathrm{E}\left[n_e(E)\right]$ denotes the mean number of secondary electrons produced by a stored primary β-electron (figure 5.7c). The average value here is 179, but with a symmetric cone electrode configuration, only half of these electrons reach the detector, the other half travels back to the source section. Therefore, an observed rate of R =1 mcps due to tritium corresponds to

$$N_T = \frac{R}{\lambda_T \cdot \mathrm{E}\left[n_e(E)\right]} = \frac{1\,s}{1000\,s \cdot 1.8 \cdot 10^{-9} \cdot 90} = 6.193 \cdot 10^3 \tag{6.7}$$

atoms in the sensitive volume. To derive the total number of atoms in the total PS volume, the number has to be multiplied with the ratio of the total PS volume to the flux volume.

$$R_{\mathrm{V(flux)}} = \frac{V_{\mathrm{PS}}}{V_{\mathrm{flux,PS}}} = \frac{8.5\,\mathrm{m}^3}{4.06\,\mathrm{m}^3} = 2.09 \tag{6.8}$$

$$N_T(\mathrm{tot}) = N_T \cdot R_{\mathrm{V(flux)}} = 46.42 \cdot 10^3 \tag{6.9}$$

Table 6.2.: **Expected number of tritium atoms in the total PS volume**, with pumping speed, decay and constant flux into the PS being in equilibrium. Pumping section reduction factor 10^{14} from [KAT05] and 10^{18} based on [FR18].

	Reduction factor (10^{14})	Reduction factor (10^{18})
TMPs and NEG ($S^{\mathrm{PS}} \approx 45771 \frac{\mathrm{l}}{\mathrm{s}}$)	46423	5
TMPs only ($S^{\mathrm{PS}} \approx 1305 \frac{\mathrm{l}}{\mathrm{s}}$)	1628352	163

However, it has been considered to take the getter material out of the PS, to reduce the radon background (Rn emanates into the volume from this material). This would reduce not only the effective pumping speed of the PS from 45771 $\frac{\mathrm{l}}{\mathrm{s}}$ to 1305 $\frac{\mathrm{l}}{\mathrm{s}}$ [Sch18], but also the cluster rate due to radon. Together with the two reduction factors (design report and simulation), the number of tritium atoms in the total PS volume is given in table 6.2. As long as the saturation of tritium atoms in the total PS volume is smaller than $46.42 \cdot 10^3$ corresponding to a rate of 1 mcps, it is nearly impossible that an elevated background due to neutral tritium can be measured, which would be the case for the results from the CPS simulation.
If the cluster rate of future measurements is twice as high as the one obtained in the reference background measurements in subsection 6.3.2 or subsection 6.3.3, it is likely the case that this additional rate is due to stored tritium.

6.3.2 Surface method

In the first method the PS1 magnet is switched off, which causes the magnetic field lines to point towards the upstream side surface of the PS. If tritium decays there, the β-electrons are guided by the magnetic field lines to the detector and are counted there. However, only about 50 % of the PS's surface (upstream side) can be covered by this method although more surface activity is expected on the downstream side, as tritium ions are accelerated towards the electrodes. Furthermore it has to be taken into account, that the magnetic field lines do not reach through the full metal upstream cone electrodes, but only to the wire electrodes. A schematic drawing of this setup, as well as the relevant magnets and electrodes can be found in figure 6.9.
An overview plot of this measurement (figure 6.10a) shows an overall high rate, but also a rise of the latter within the measurement. When the PS1 is turned back on the magnetic field lines start to bend down towards the PS1 and then no longer point towards the surface any more, but the volume. Thus the rate also starts to fall, because much fewer electrons are created in the volume. A time cut is performed to evaluate the rate before the rise in rate at about t=4000 s (figure 6.11a). The events are then binned in minute and hour intervals, projected, and fit with a Gaussian distribution. The residuals are plotted on the lower part and on the lower right the corresponding pixel distribution can be found, which shows a hotspot consisting of three pixels. This hotspot contributes to a big part of the overall rate, but is not understood in a physical way so far.

Nevertheless, with a rate of ≈ 1000 counts per second (cps) it is impossible to be sensitive to tritium decays, which is why a second measurement with the same configuration was done to check for reproducibility of the results (figure 6.9). The rate overview (figure 6.10b) shows a drop in rate, which is due to the failure of the air coil system. When the system comes back up, the rate is slightly higher compared to before. Here, a time cut is also performed to remove the data during the air coil failure. This time, the rate (≈30 cps) is one order of magnitude smaller than before, but the hotspot still exists (figure 6.11b). Due to the lack of reproducibility of the results, this measurement setup cannot be used as a sensitive method to measure the background contribution due to tritium decays.

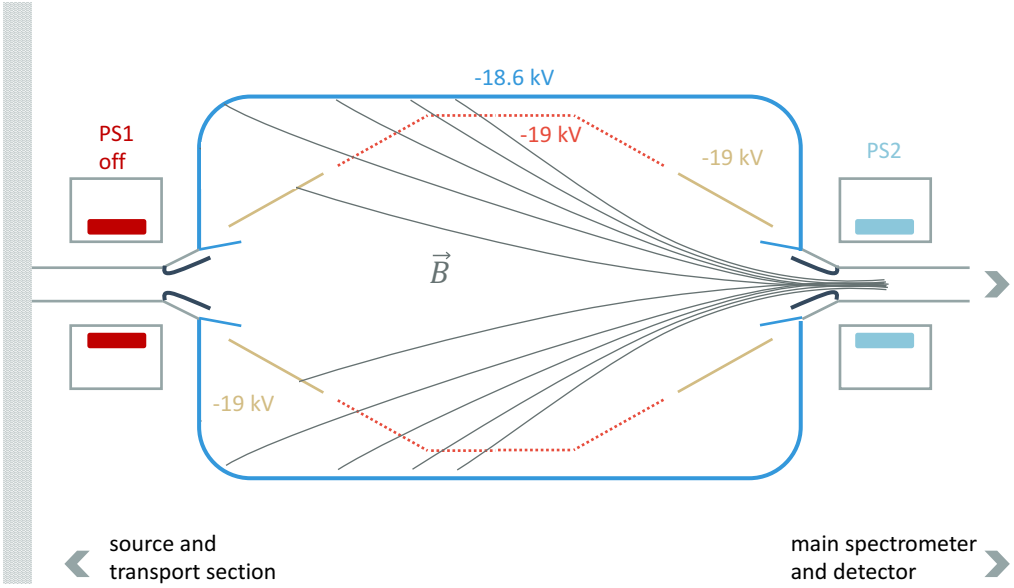
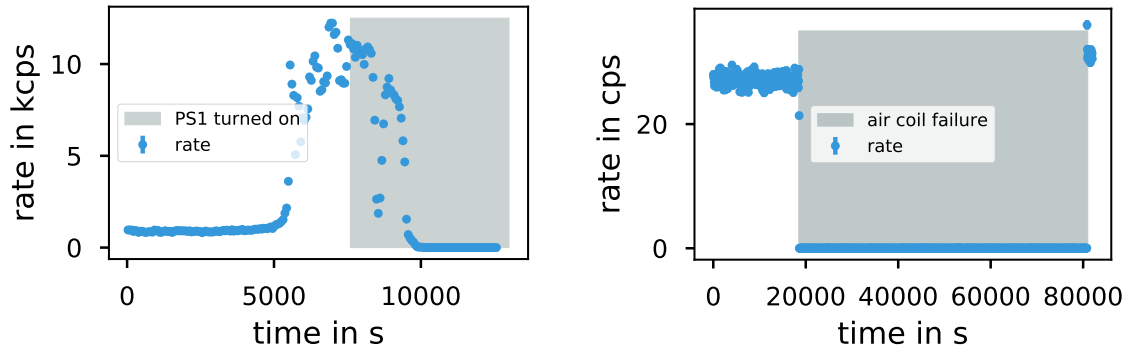
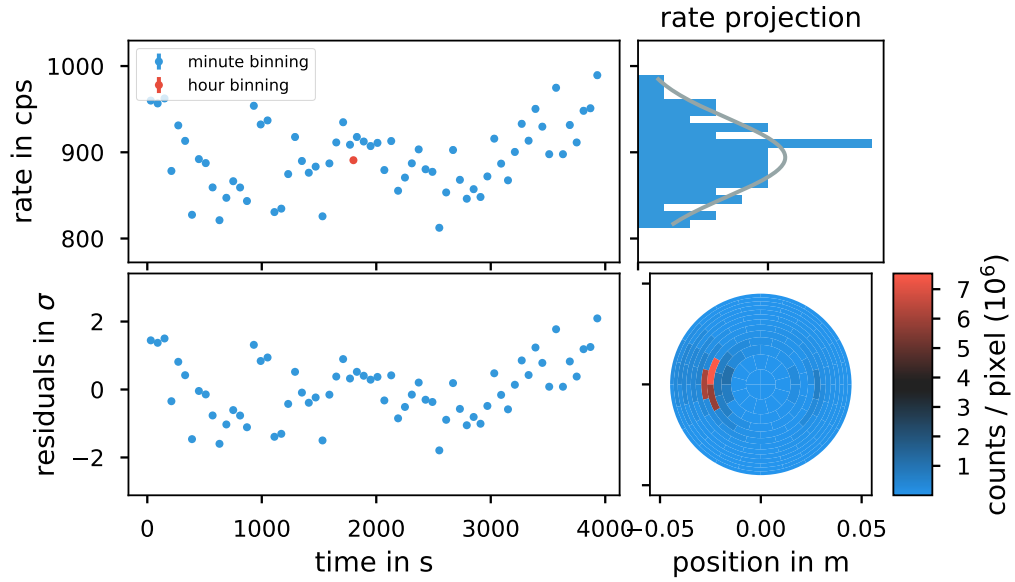


Figure 6.9.: **Schematic drawing of the PS and its relevant magnetic and electric components for the surface configuration.**

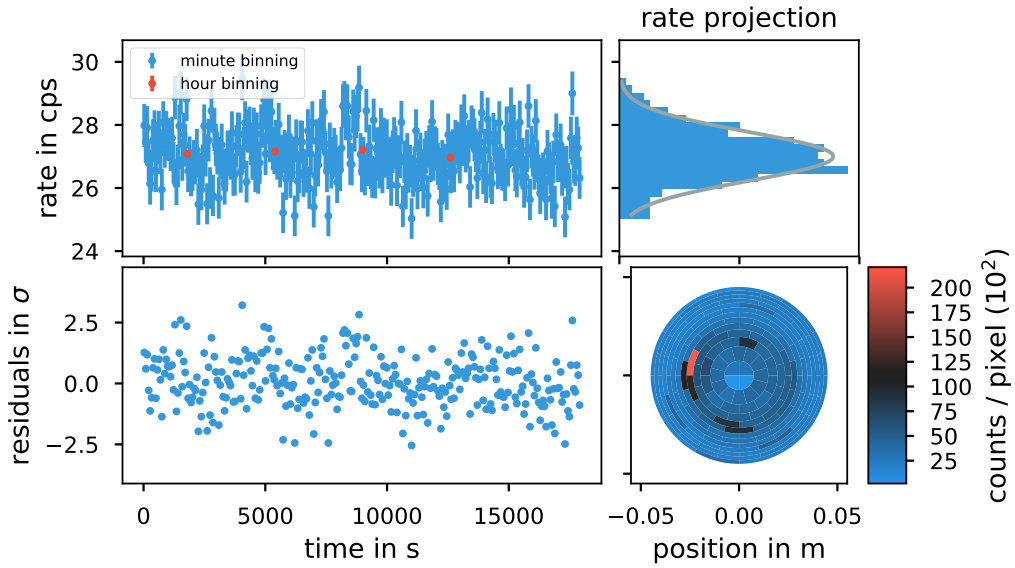


(a) (b)

Figure 6.10.: **Overview plots surface method measurements.** a): Rate overview plot of run 35153 - 35159. Besides an overall high rate, the measurement also shows a strange behavior because the rate suddenly starts to rise higher. Rate decreases due to the ramp-up of PS1 magnet. b): Rate overview plot of run 35385 - 35390. Rate is stable at first, but drops when the air coil system fails. After system is operational again near the end of the measurement, the rate is slightly higher compared to before.



(a)



(b)

Figure 6.11.: **Summary plots for surface method measurements.** a): Summary plot of runs 35153 - 35159, with bin intervals of one minute (blue) and one hour (red) (upper left). The projection of the minutely binned rate is shown in the top right corner with a Gaussian distribution fit in grey. Bottom left shows the residuals between fit and data points and bottom right the pixel distribution for the whole run. b): Summary plot of runs 35385 - 35390, with bin intervals of one minute (blue) and one hour (red) (upper left). The projection of the minute binned rate is shown in the top right corner with a Gaussian distribution fit in grey. Bottom left shows the residuals between fit and data points and bottom right the pixel distribution for the whole run.

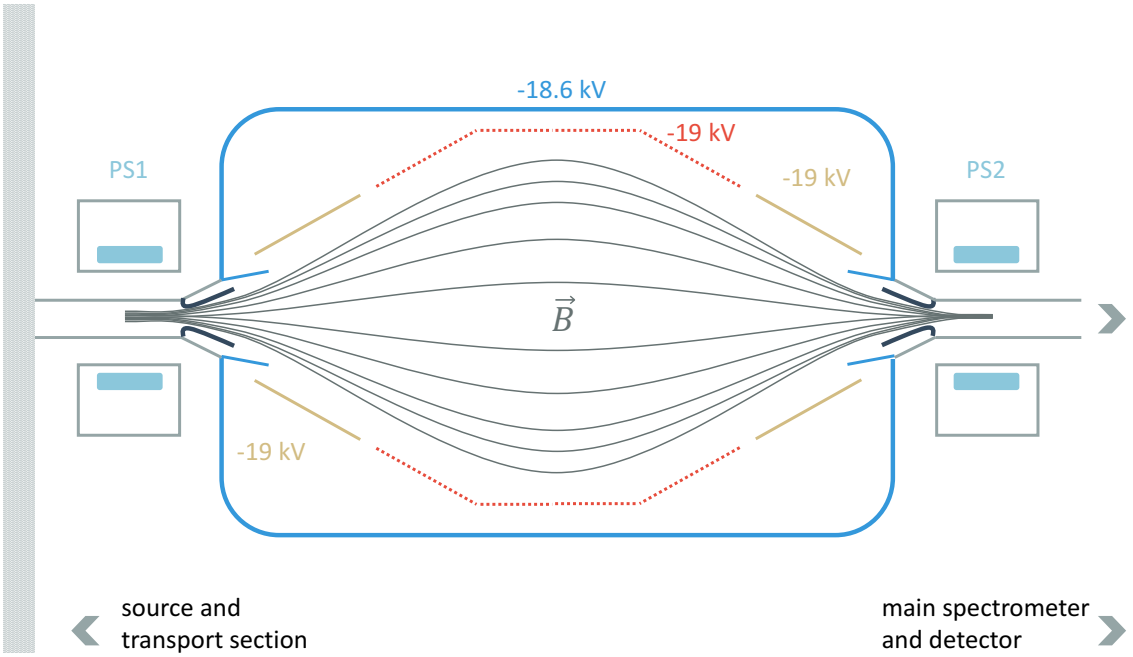


Figure 6.12.: **Schematic drawing of the PS and its relevant magnetic and electric components for the volume configuration.**

6.3.3 Volume method

The measurement setup checking for background due to tritium decays in the volume (6.12) may not be able to be sensitive to the area on the surface, where most tritium is expected, but this configuration has the advantage of an overall low background. In the following it will be shown that this method is the more sensitive one.
Measurements with this configuration have been carried out before, during, and after the FT campaign. At first, all measurements at an elevated pressure are presented, followed by measurements at nominal operational pressure $\approx 10^{-11}$ mbar.

Elevated pressure p=3.1×10^{-9} mbar

In the first measurement configuration, the upstream cone electrode is set on a more negative potential (-19 kV) than the wire- and downstream cone electrode (-18.9 kV) in order to reflect the electrons moving back to the source towards the detector. Therefore an increased rate compared to a symmetric electrode configuration is expected. The rate summary plot in figure 6.14a clearly shows spikes in the rate, which is a first indicator of cluster events (radon), which results in a non Poissonian distribution of the background. Furthermore, a pixel hotspot in the lower right part of the detector is visible which has been observed in several other measurements as well. During the FT campaign, a background measurement while ramping up the CPS magnetic field has revealed a high dependency of the pixel hotspot on the magnetic field strength. In figure 6.13a, the background rate and CPS current is shown for the time the CPS magnets are ramped up. In figure 6.13b, the value pairs are binned with a bin size of 60 s and a moving average with a window

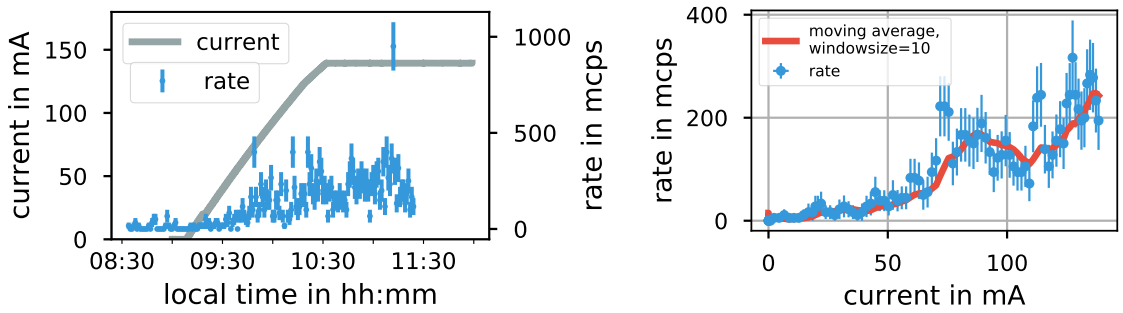
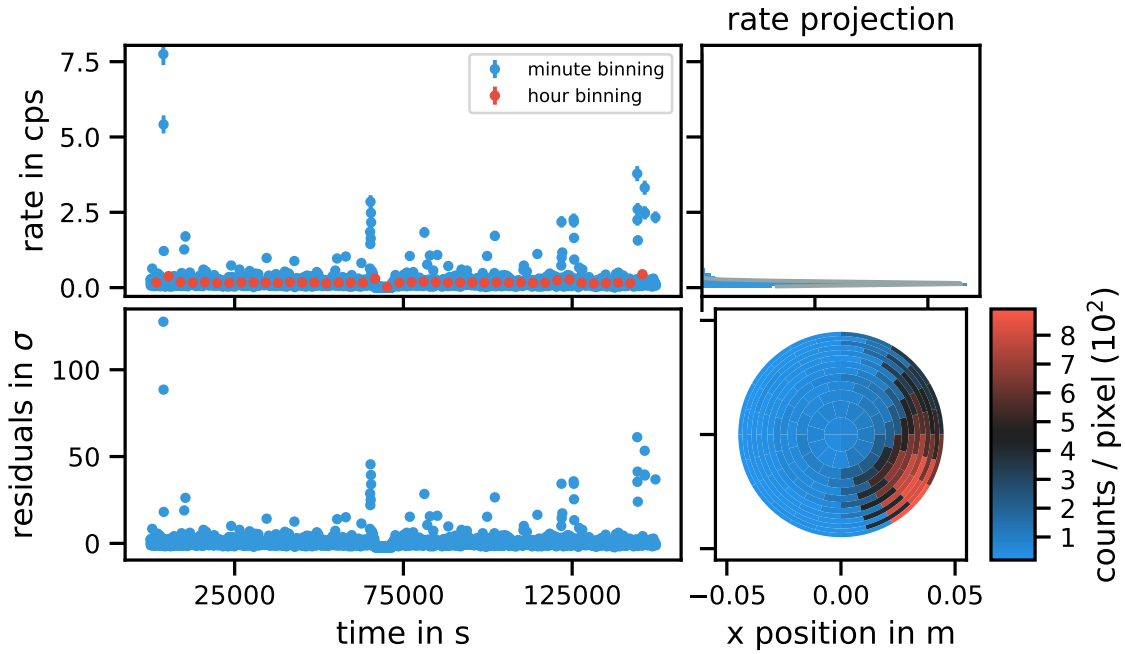


Figure 6.13.: **CPS current and rate while ramping up the CPS magnetic field** in figure a). b): CPS current and rate with a bin width of 60 s. Poisson uncertainties on the rate and Gaussian uncertainties on the current are assumed. In red, a moving average for the time-series with a window size of 10 bins is shown.

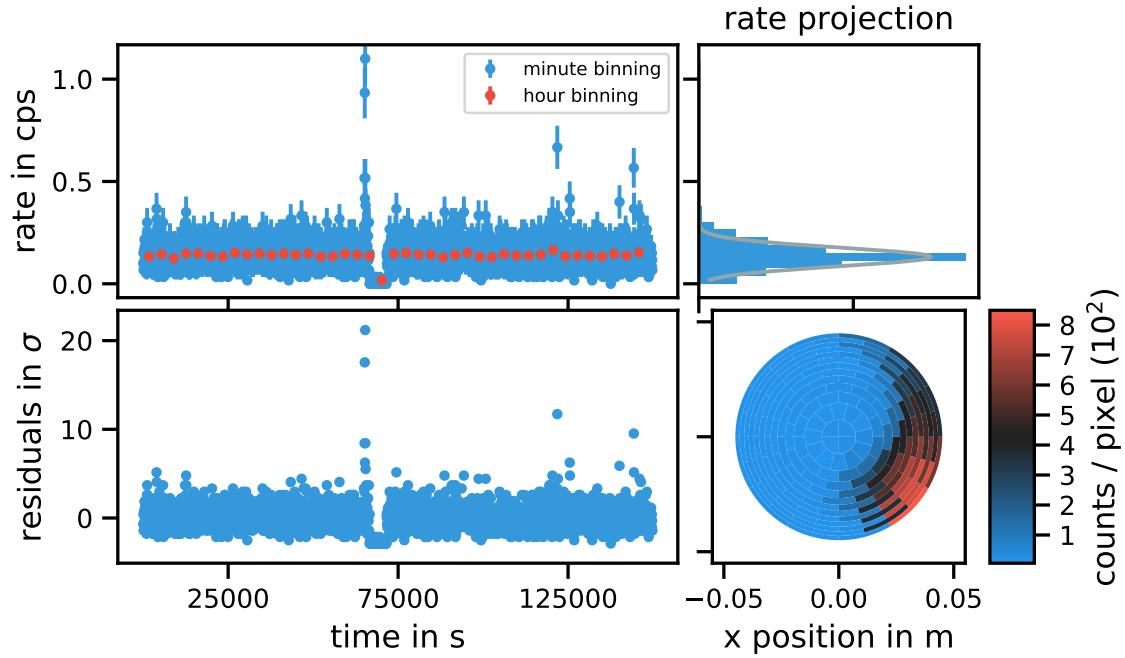
size of ten bins visualizes the trend that the rate increases with the current. This is also confirmed by a high correlation factor of $r \approx 0.88$ between the rate and the current.
Next, the cluster cut with the DBSCAN algorithm is applied, which leads to a significant decrease of the spike structures in the rate and therefore to the expected Poisson distributed background (figure 6.14b). The total background rate for this setup is 190.61 ± 1.69 mcps and the cluster rate is 46.71 ± 0.84 mcps, leading to a rate of 143.90 ± 1.47 mcps after the cluster cut. In the appendix A.3.2, additional information about the clusters can be found. The number of tritium atoms to which this cluster rate equals can be read out from figure 6.8b. Four different scenarios are possible:

- i): With NEG material and gas flow from the design report [KAT05] (solid blue).
- ii): Without NEG material and gas flow from the design report [KAT05] (dotted blue).
- iii): With NEG material and gas flow from simulation [FR18] (solid red).
- iv): Without NEG material and gas flow from simulation [FR18] (dotted red).

According to the scenarios i)-iv) the number of tritium atoms in the volume this cluster rate would equal can be read out from figure 6.8b.



(a)



(b)

Figure 6.14.: **Summary plots of runs 35160 - 35171** before a) and after cluster cut b) at elevated pressure in the PS.

Nominal pressure p=1.5×10^{-11} mbar

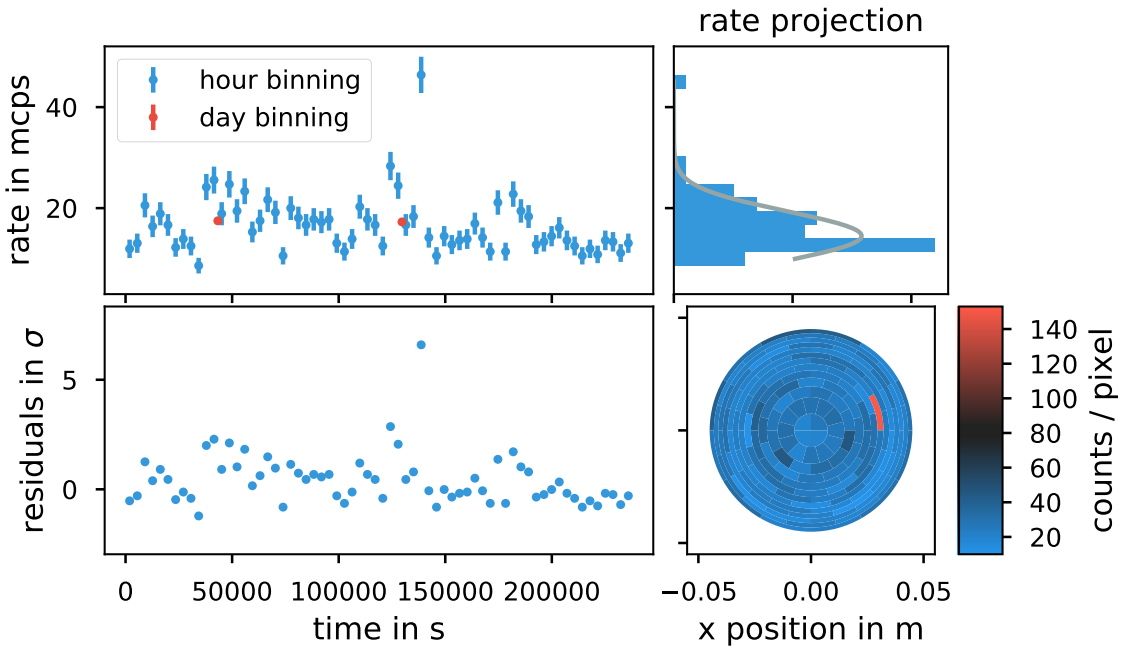
At low pressure, cluster identification with the previous algorithm is nearly impossible. However, as shown in chapter 5, the machine learning algorithms perform well. The electro-magnetic configuration of the PS is the same as in figure 6.12. In figure 6.15a and figure 6.15b the rate, its distribution, and the pixel distribution are shown for the run numbers 35171 -35189. For more detailed information about the clusters figure A.10 may be useful. The total background rate for this setup is 16.62 ± 0.26 mcps and the cluster rate is 10.06 ± 0.21 mcps, leading to a rate of 6.56 ± 0.17 mcps after the cluster cut. Similar results were obtained by a second measurement with the same configuration (figure A.11). Here, the total background rate is 15.29 ± 0.39 mcps and the cluster rate is 8.32 ± 0.29 mcps, leading to a rate of 6.97 ± 0.26 mcps after the cluster cut. This similarity of the results shows that the background and its properties are stable and can be reproduced, which is favorable since these measurements are intended to be reference measurements. Similar results are achieved at a slightly higher pressure (3×10^{-11} mbar) with a total background rate of 21.20 ± 0.30 mcps and the cluster rate is 9.03 ± 0.19 mcps, leading to a rate of 12.17 ± 0.23 mcps after the cluster cut. For detailed cluster information have a look at figure A.12.

Summary

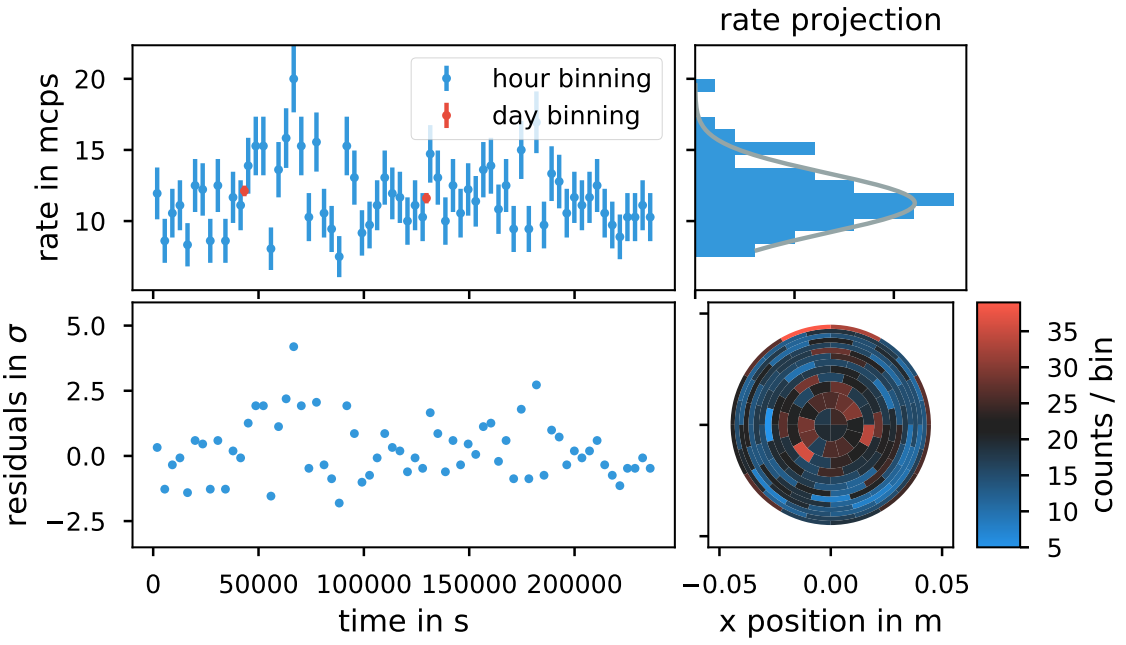
The goal of the work described in this chapter was to:

- Investigate the characteristics of cluster events for both different pressures and target devices (PS and MS).
- Find a sensitive measurement configuration to detect the presence of tritium in the spectrometer section.

Hereby, the excess around zero in the distribution of radius differences of consecutive events has been found in addition to the known characteristics of inter-arrival times distribution. This new feature allows one to approximate the cluster cut efficiency independent of the pressure, compared to the pressure dependent inter-arrival times.
In order to find a sensitive measurement setup to detect the presence of tritium in the spectrometer section, reference background measurements with two different methods have been carried out. The method where the magnetic field lines point towards the surface of the PS generally suffered from a high and unstable count rate. The other method where the field lines are confined to the volume shows not only better reproducibility, but also the rate is several orders of magnitude lower compared to the surface method. Due to the increased performance of the machine learning algorithms, the cluster rate can be determined more precisely which is a direct measure for the amount of tritium in the spectrometer section.



(a)



(b)

Figure 6.15.: **Summary plots of runs 35173 - 35189** before a) and after cluster cut b) at nominal pressure in the PS.

7 Conclusion

KATRIN is a next generation neutrino mass experiment with the ambitious goal to determine the neutrino mass with a sensitivity of $m(\nu_e) = 0.2\,\text{eV}$ at 90% C.L. by precision electron spectroscopy near the tritium β-decay spectral endpoint. As the statistics of signal electrons in this energy region are low, maintaining a low background is key to achieving this unprecedented sensitivity. Electrons accompanying radon decays in the spectrometer volumes have a high probability of being stored in the flux tube for up to thousands of seconds due to the MAC-E filter design. While moving back and forth in the flux tube, these stored electrons can produce hundred of secondary electrons via scattering off residual gas. The secondary electrons escape the trap quickly due to their low energy and appear on the detector as events with a correlation in time and radius, due to the unique magnetron motion of the primary electron. Electrons originating from tritium β-decay hereby have the same signature.
A time-cut based clustering algorithm was developed in order to discriminate these cluster events from normal background events and estimate the overall contribution of cluster events. As the precision of this cluster identification algorithm is key to sensitive background characterization, novel approaches have been developed in this thesis.

A training data set which is based on both measurement and simulation data has been used to compare and optimize both the previous algorithm, as well as the unsupervised and supervised machine learning methods. For meaningful results, the quality of the training data set is important, which is why storage simulations of electrons up to 18.6 keV within the PS, generated by tritium β-decays, were carried out. Here, the most recent beamline configuration, as well as realistic magnetic field settings were used in order to achieve realistic results. Moreover, the analytical derivation of the mean time between two ionization events has been confirmed with measurements at elevated (3×10^{-9} mbar) and nominal pressure (1.5×10^{-11} mbar). Due to this unique training set, the hyper parameters optimization of the previous algorithm has revealed that larger time window cuts than previously used, lead to an increased classification score.

While the previous algorithm needs the user to provide a reasonable choice of the time window input parameter, the parameters of the DBSCAN algorithm are derived automatically and analytically by only providing the pressure as an input parameter. The pressure for the corresponding measurement can be read out from a pressure sensor in the spectrometers and the parameter could then be represented by the mean or median of these pressure values. The results from this cluster cut have the advantage of being nearly independent

of the pressure. While the hyper-parameter input of the DBSCAN is derived from the underlying physics of ionizations processes, the BDT as a supervised learning method adjusts its splits in order to optimize classification precision of the training data. This makes the BDT a generally flexible method, but also the most dependent one on a high quality training data set.

With the optimization of these algorithms for cluster identification, the excess in the distribution of the difference of the radii of consecutive events at zero as a novel, and independent feature of cluster events has been discovered. In this work, this was used to estimate the contribution of radon emanating from the getter material of the PS into the MS volume. These measurements have revealed, that this effect has a non-negligible influence on the background rate of the MS. The NEG material causes an additional rate of 5.47 ± 0.34 mcps, which equals about 1.5 % of the total MS background rate. Furthermore, a robust and more sensitive measurement setup to detect decays of neutral tritium within the PS volume has been tested. With this method it is possible to immediately estimate the contribution of tritium to the background in the PS and take counter-measures, before most of the tritium reaches the MS and causes a non-negligible rise in the background rate.

For further improvements of the DBSCAN algorithm, the scaling of the radius can be optimized. As the pixel rings are not equally spaced, the average of the ring distances is used for scaling of the individual pixel radii. For a more precise result, every pixel ring could have its own scaling factor. In case of the BDT algorithm, a new definition of predictor variables as well as the parameter tuning of the interaction depth could lead to further cluster identification precision.

A Appendix

A.1 Proof of Theorems

Proof. Let N_t be the number of events arriving in $[t]$ and X_t the time it takes for one additional event to occur. Then $(X_t > x)$ means that no event has arrived in $[t, t+x]$, which is equivalent to $(N_t = N_{t+x})$.

$$(X_t > x) \equiv (N_t = N_{t+x}) \quad \text{or by law of complement} \tag{A.1}$$

$$P(X_t < x) = 1 - P(X_t > x) = 1 - P(N_t = N_{t+x}) = 1 - P(N_{t+x} - N_t = 0). \tag{A.2}$$

Since the increments N_{t+x}, N_t of a Poisson process are independent stochastic variables, it follows

$$P(N_{t+x} - N_t = 0) = P(N_x = 0) = \exp(-\lambda x) \cdot \frac{(\lambda x)^0}{0!} = \exp(-\lambda x) \tag{A.3}$$

$$1 - P(X_t > x) = \begin{cases} 1 - \exp(-\lambda x), & \text{for } t \geq 0 \\ 0, & \text{for } t < 0 \end{cases} \tag{A.4}$$

which is the cumulative distribution function (cdf) of the exponential distribution. □

Proof. Let $X, Y \sim \text{Exp}(\lambda)$ be independent random variables, with the probability density function (pdf) given by

$$f(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right), \qquad \lambda, x > 0 \tag{A.5}$$

and the characteristic function by

$$\varphi_X(t) = \frac{1}{1 - \lambda i t}, \qquad \lambda > 0. \tag{A.6}$$

Then the random variable's $Z = X - Y$ characteristic function is given by

$$\begin{aligned}
\varphi_Z(t) &= \mathrm{E}\left[\exp\left(-it(X-Y)\right)\right] && \text{(A.7)}\\
&= \mathrm{E}\left[\exp(-itX)\exp(-it(-Y))\right] && \text{(A.8)}\\
\text{independence} &= \mathrm{E}\left[\exp(-itX)\right]\mathrm{E}\left[\exp(-it(-Y))\right] && \text{(A.9)}\\
&= \varphi_X(t)\cdot\varphi_Y(-t) && \text{(A.10)}\\
&= \frac{1}{1-\lambda it}\frac{1}{1-\lambda i(-t)} && \text{(A.11)}\\
&= \frac{1}{1+\lambda^2 t^2}, && \text{(A.12)}
\end{aligned}$$

which is the characteristic function of the Laplace distribution $\mathrm{La}(\lambda)$ with the pdf

$$f(\lambda) = \frac{1}{2\lambda}\exp\left(-\frac{|x|}{\lambda}\right). \tag{A.13}$$

□

Proof. Let $X, Y \sim \mathrm{N}(\mu_i, \sigma_i^2)$ be independent random variables, with the characteristic function given by

$$\varphi_X(t) = \exp\left(it\mu_x - \frac{1}{2}t^2\sigma_x^2\right) \tag{A.14}$$

Then the random variable's $Z = X - Y$ characteristic function is given by

$$\begin{aligned}
\varphi_Z(t) &= \mathrm{E}\left[\exp\left(-it(X-Y)\right)\right] && \text{(A.15)}\\
&= \mathrm{E}\left[\exp(-itX)\exp(-it(-Y))\right] && \text{(A.16)}\\
\text{independence} &= \mathrm{E}\left[\exp(-itX)\right]\mathrm{E}\left[\exp(-it(-Y))\right] && \text{(A.17)}\\
&= \varphi_X(t)\cdot\varphi_Y(-t) && \text{(A.18)}\\
&= \exp\left(it\mu_x - \frac{1}{2}t^2\sigma_x^2\right)\cdot\exp\left(-it\mu_y - \frac{1}{2}t^2\sigma_y^2\right) && \text{(A.19)}\\
&= \exp\left(it(\mu_x-\mu_y) - \frac{1}{2}t^2(\sigma_x^2+\sigma_y^2)\right) && \text{(A.20)}
\end{aligned}$$

which is the characteristic function of the normal distribution $\mathrm{N}(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$ with the pdf

$$f(\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \tag{A.21}$$

□

A.2 Appendix to machine learning

A.2.1 Simulation Parameters

Table A.1.: Simulation input parameters and condensed output parameters.

parameters	Simulation 1	Simulation 2
beamline configuration	FT	FT
pressure	$3{\cdot}10^{-9}$ mbar	$4 \cdot 10^{-11}$ mbar
enhancement	10^6	10^8
integrator	symplectic	symplectic
MS electrodes	grounded	grounded
PS upstream electrode	-19 kV	-19 kV
PS downstream electrode	-19 kV	-19 kV
PS hull electrode	-18.6 kV	-18.6 kV
terminator_z_min	-16.46375-(10^{-8}) m	-16.46375-(10^{-8}) m
terminator_z_max	-12.10375+(10^{-8}) m	-12.10375+(10^{-8}) m
number of started simulations	10^3	10^3
successful simulations	986	979
number of secondary electrons	149437	175501
number of electrons terminated at PS2	70605 (47 %)	78182 (44 %)
number of electrons in the ROI	63535 (42 %)	73485 (42 %)
number of electrons scaled to the detector	62900 (42 %)	72832 (41 %)

A.2.2 Calculation of mean time between ionization processes

In the calculations of the mean time between two ionization events in equation 5.15, it is assumed that the electron has a kinetic energy of 10 keV, and the mean cross section for that energy is $\approx 1.38 \cdot 10^{-21}$ m^2 (figure 5.4a). To prove, that these assumptions are valid, a more detailed analysis is necessary. At low pressures ($\approx 10^{-11}$ mbar), mainly hydrogen atoms and molecules are the dominating isotopes contributing to the pressure. A first approximation is, that at a pressure of $1 \cdot 10^{-11}$, H and H_2 both have half of the fraction and therefore both a partial pressure of $0.5{\cdot}10^{-11}$ mbar. If the pressure now is artificially increased due to pumping in argon, the partial pressure of argon can be derived from the total pressure $p = \sum \partial p_i$

$$\partial p_{\mathrm{Ar}} = p - \sum_{i=\mathrm{H},\mathrm{H}_2} \partial p_i. \tag{A.22}$$

Therefore, the single ionization cross sections of Ar, H and H_2 can be taken into account individually. Rewriting equation 5.15 to

$$\langle t \rangle = \frac{k_B T}{\beta \mathrm{c} \cdot p\sigma} = \frac{k_B T}{\beta \mathrm{c}} \sum_{i=\mathrm{H},\mathrm{H}_2,\mathrm{Ar}} \frac{1}{\partial p_i \sigma_i}. \tag{A.23}$$

Taking into account the energy dependence of the scattering cross section, equation A.23 becomes

$$\langle t \rangle (E) = \frac{k_B T}{\beta(E)\mathrm{c}} \sum_{i=\mathrm{H,H_2,Ar}} \frac{1}{\partial p_i \sigma_i(E)}. \tag{A.24}$$

Introducing the fraction of the partial pressure to the total pressure

$$w_i = \frac{\partial p_i}{p}, \tag{A.25}$$

equation A.24 can be written as

$$\langle t \rangle (E) = \frac{k_B T}{\beta(E)\mathrm{c} \cdot p} \sum_{i=\mathrm{H,H_2,Ar}} \frac{1}{w_i \sigma_i(E)}. \tag{A.26}$$

This represents not an average mean scattering cross section, but a weighted mean scattering cross section, taking into account that at high pressures, the cross section of argon dominates. In figure A.1c and figure A.1d this impact becomes visible, since the mean times between two ionization processes are smaller in figure A.1d than in figure A.1c.

To show that a valid mean time between two ionization events to a corresponding pressure is given by the approximation in equation 5.15, figure A.1a can be useful. It shows, that at constant pressure ($3 \cdot 10^{-9}$ mbar), the mean time between two ionizations below 0.3 s cover energies between 40 eV and 10 keV. Considering the fact that the primary electrons with energies below 100 eV can escape the magnetic trap of the PS, this value as an upper limit will cover the main energy range of a tritium β-primary electron while cooling down due to scattering.

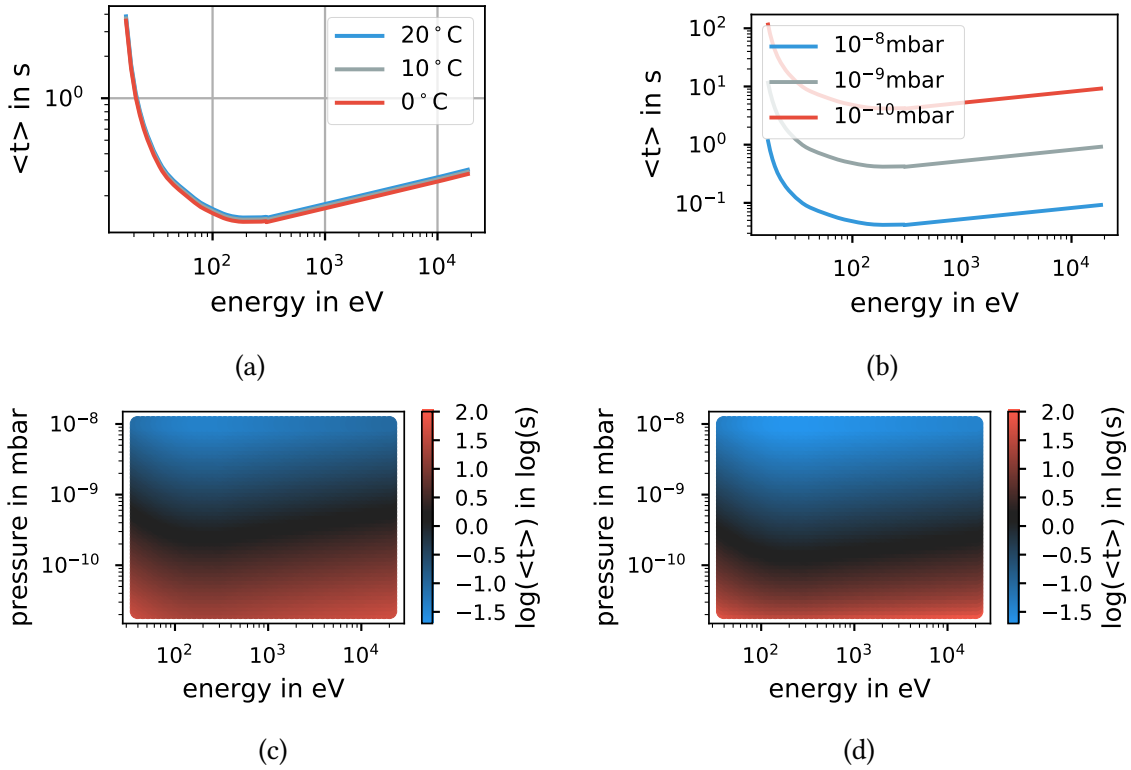


Figure A.1.: **Mean time between ionization processes in dependence of temperature, pressure and energy.** a): For ordinary environmental temperatures (0° C - 20° C), the mean ionization time does not vary much. b): The pressure has a much bigger impact on the mean time between two ionization processes. c): The mean time between two ionization processes in dependence on pressure and energy is displayed. For the calculation, the mean ionization cross section is taken on the left, and the weighted mean ionization cross section on the right (d)).

A.2.3 Hyper parameter estimation at nominal pressure

The training data set (section 5.2.3) for this parameter tuning has a size of 10075. It is therefore two orders of magnitude smaller than the one used in the elevated pressure regime (section 5.2.2), simply due to performance reasons.

A.2.3.1 Previous algorithm

In the same procedure as described in section 5.3 at elevated pressure, the hyper parameters are estimated for nominal pressure. Here the parameter grid is defined by `delta_t` $\otimes$ `min_samples` $[70, 90]\,\mathrm{s} \otimes [5, \ldots, 20]$ and for each combination, the measure of success (equation 5.21) is calculated. Best results are achieved by combinations around $[85, 90]\,\mathrm{s} \otimes [14, 16]$, which is visualized in figure A.2.

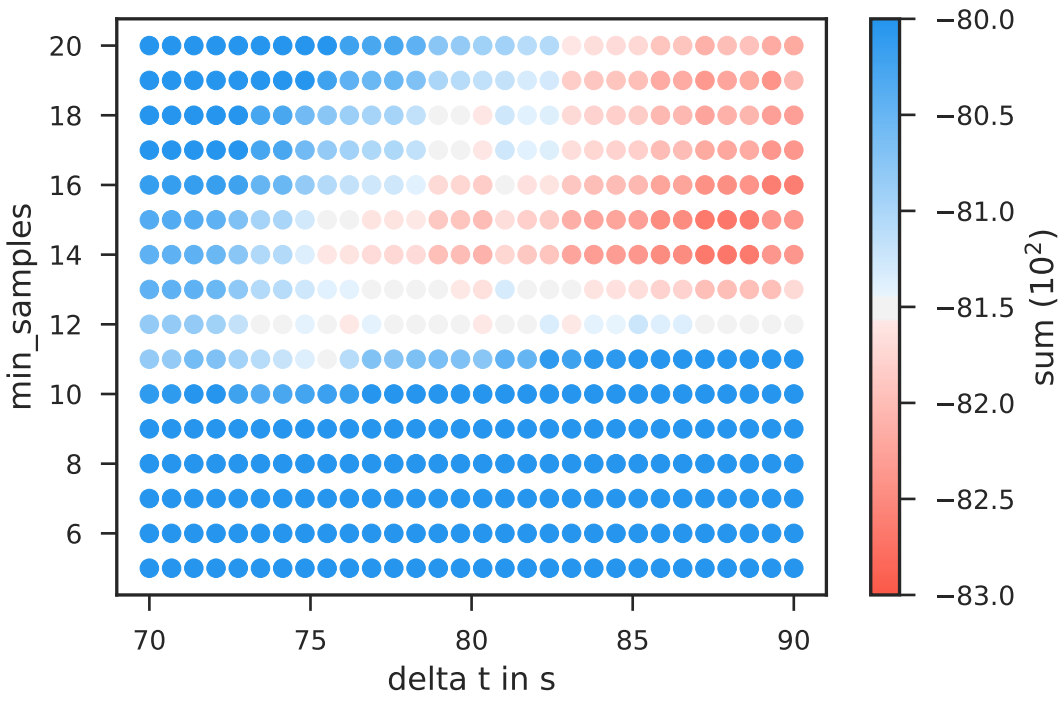


Figure A.2.: **Heatmap of the parameter grid.** Smaller values of the sum indicate better performance (red areas) and bigger values worse performance (blue area).

A.2.3.2 DBSCAN

The input parameters for the DBSCAN algorithm are also optimized. However, as first results did not look promising, the factor in equation 5.24 was also used for evaluation. This yielded best results with a factor of $f = 0.81$, in comparison to the elevated pressure with a factor of $f = 0.4$. In figure A.3, the measure of success is shown with respect to the parameters `length` and `min_samples`.

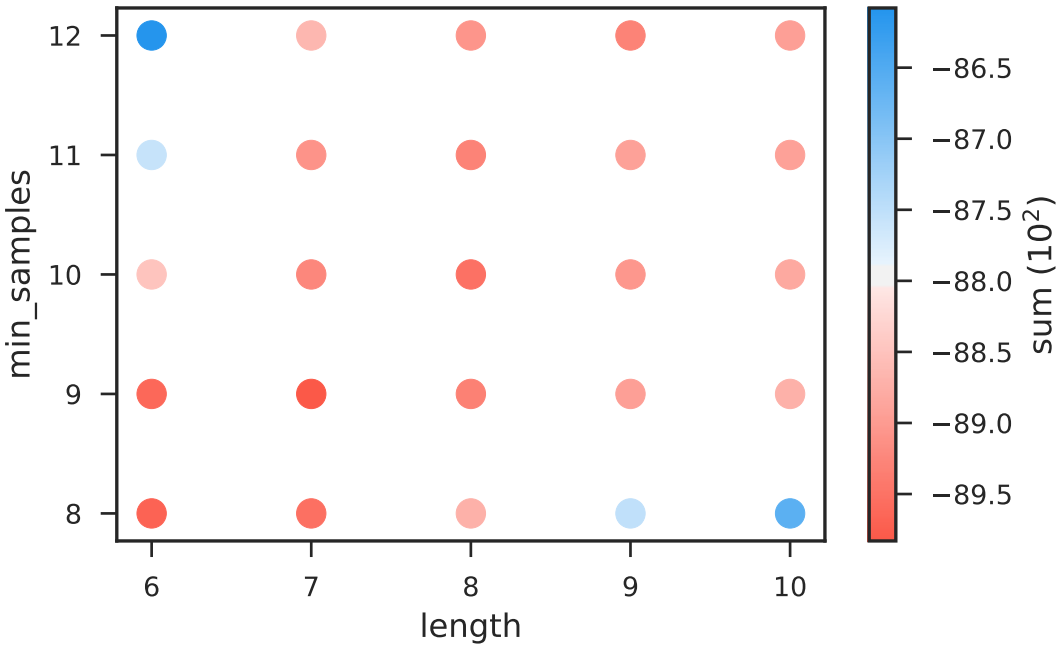
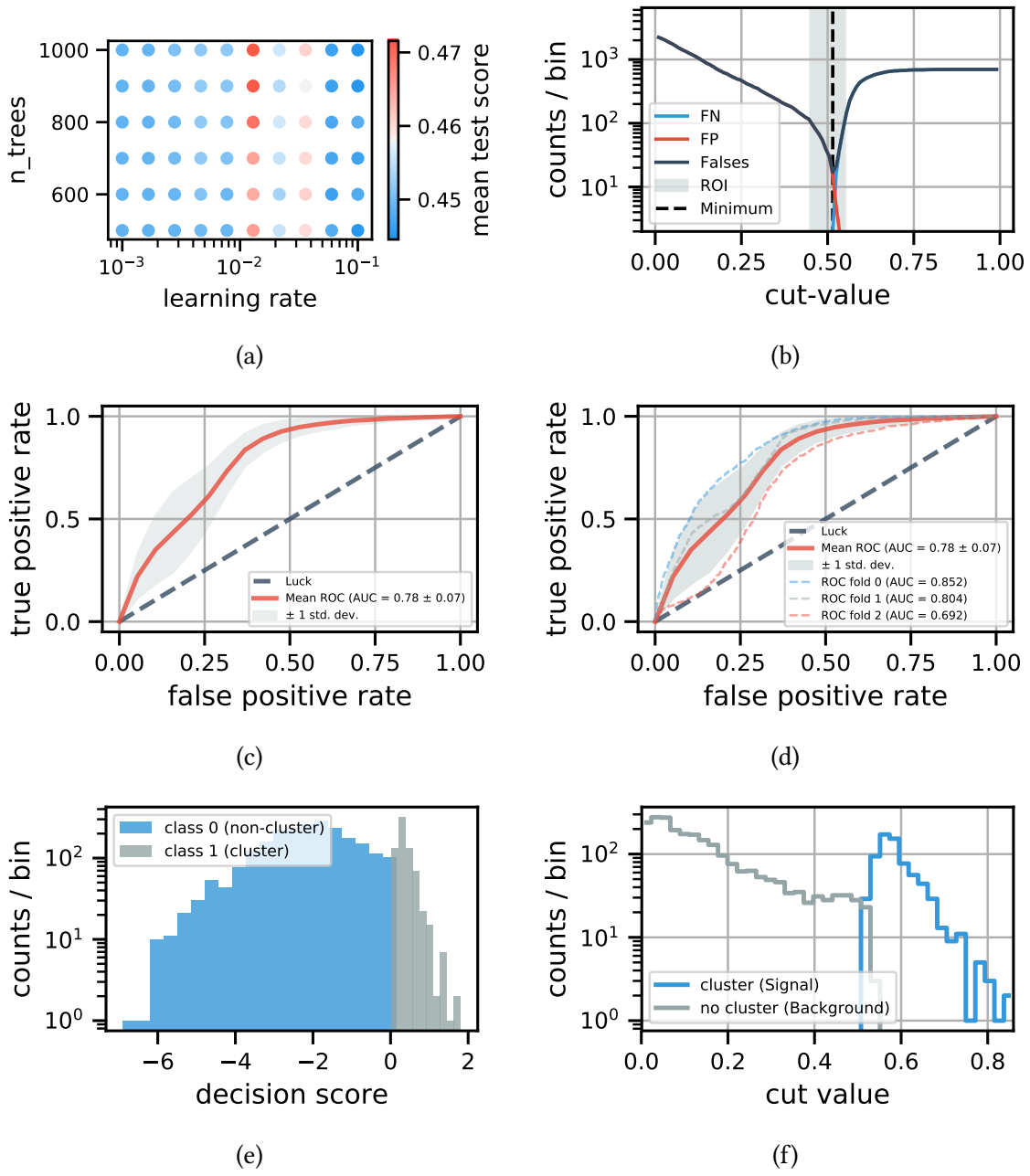


Figure A.3.: **Heatmap of the parameter grid.** Smaller values of the sum indicate better performance (red areas) and bigger values worse performance (blue area).

A.2.3.3 BDT

The hyper parameters are estimated by a grid search 3-fold CV. The parameter grid is defined by `n_trees` $\otimes$ `learning_rate` $[500, ..., 1000] \otimes [0.001, ..., 0.1]$ and for each combination a 3-fold CV on the simulation data is performed. The mean score is calculated for each parameter combination and the highest score was yield by 1000 `trees` and a `learning_rate` of 0.0129 (figure A.4a). This combination is then used to perform a 3-fold CV on the simulation data and checked for the true positive classification compared to the false positive classification, which is shown in a so called ROC curve in figure A.4c. This curve ideally should have a strong increase in the TPR when the FPR is low, which means the classification algorithm achieves a high classification rate of TPs while only allowing a small number of misclassifications with FPs. Furthermore, the performance of the algorithm can be evaluated by visualizing the decision score and the cut value of the classifier in figure A.4f. Usually it is desired to want both FP and FN as low as possible, however there can be cases when it is desirable to trade higher misclassification for an increased TPR.



(a) (b) (c) (d) (e) (f)

Figure A.4.: **Performance plots of the BDT classifier at nominal pressure.** a): Heatmap of parameter grid, to visualize classification precision for different parameter combinations. b): Number of FP and FN in dependence on the cut-value. The ROI indicates region of favored cut. c), d): Mean ROC curve for 3-fold CV on the left and ROC curve for 3-fold CV on the right. e), f): The probability value cut figure 5.20f shows if the probability, as an output of the classifier for each event being a normal background event (close to 0) or a cluster event (close to 1), would be cut at a specific value, how many of these observations would be true positives.

A.3 Appendix to cluster identification

A.3.1 Influence of NEG material on radon cluster rate

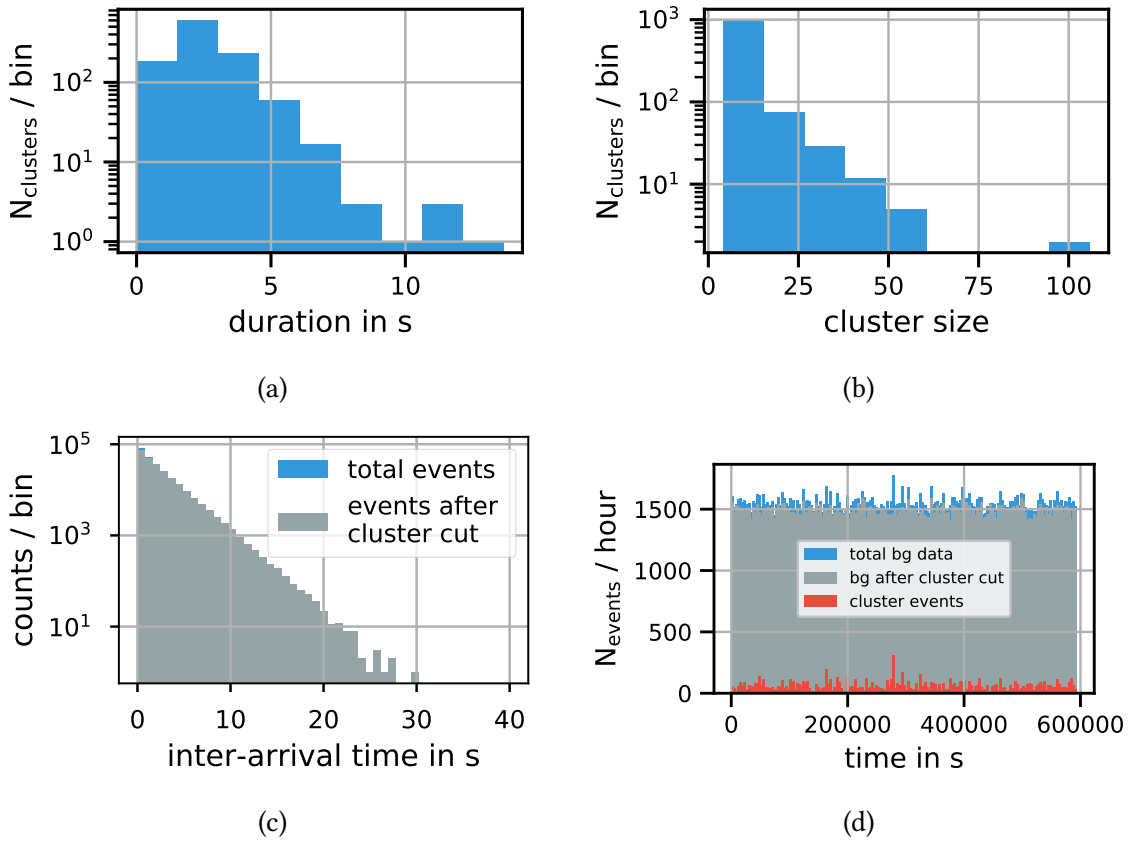


Figure A.5.: **Cluster information for runs 35435 - 35476 (elevated pressure), MS.** a), b): Histogram of cluster durations a) and histogram of cluster size b). c), d): Histogram of inter-arrival times c) and rate trend of raw data and after cluster cut with bin width of one hour is shown d).

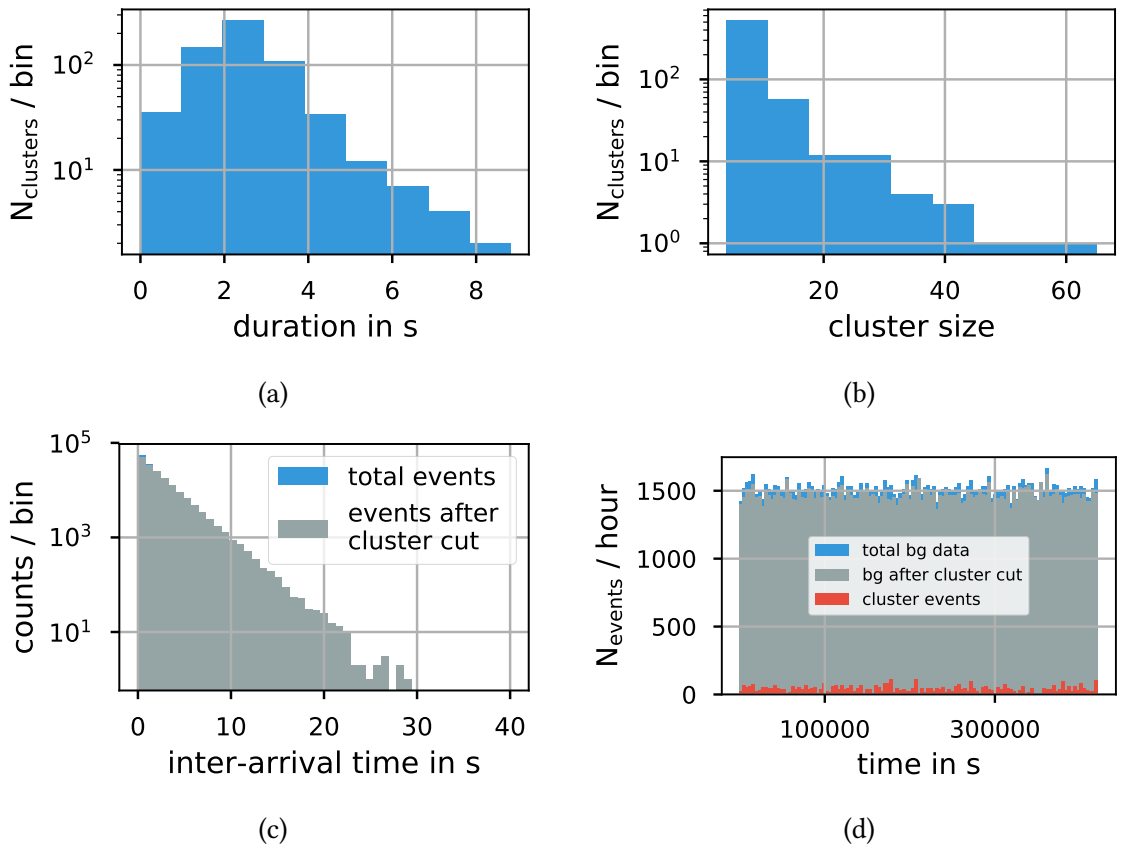


Figure A.6.: **Cluster information for runs 35477 - 35506 (elevated pressure), MS.** a), b): Histogram of cluster durations a) and histogram of cluster size b). c), d): Histogram of inter-arrival times on the left c) and rate trend of raw data and after cluster cut with bin width of one hour is shown d).

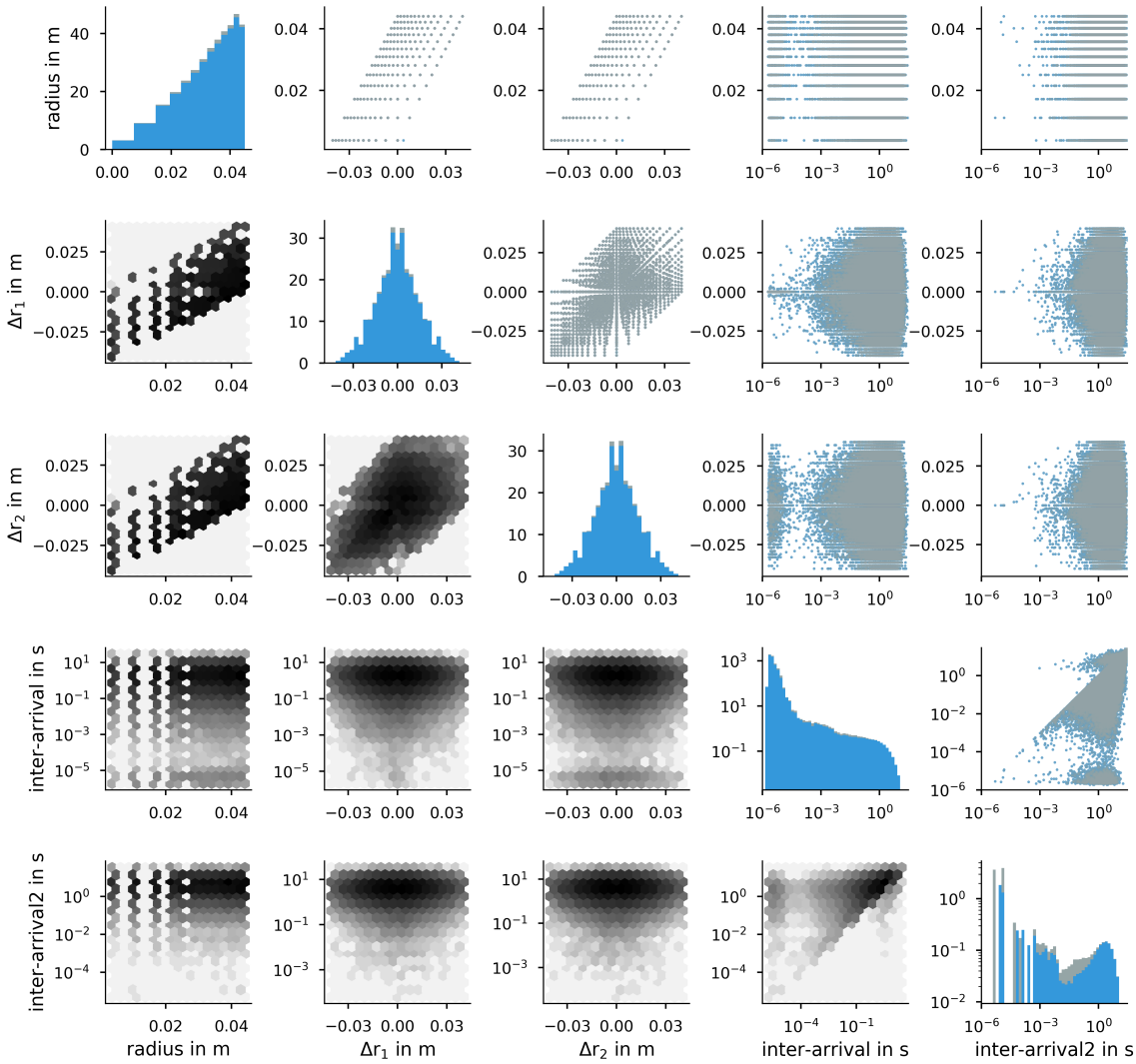


Figure A.7.: **Pairwise plots of the predictor variables** at elevated pressure in the MS. The diagonal elements are histograms, the upper elements of the diagonal are scatter plots, where blue labels indicate non-cluster events and grey labels cluster events. On the lower part of the diagonal, two dimensional histograms are shown, with black color indicating high counts in this bin and white color low counts.

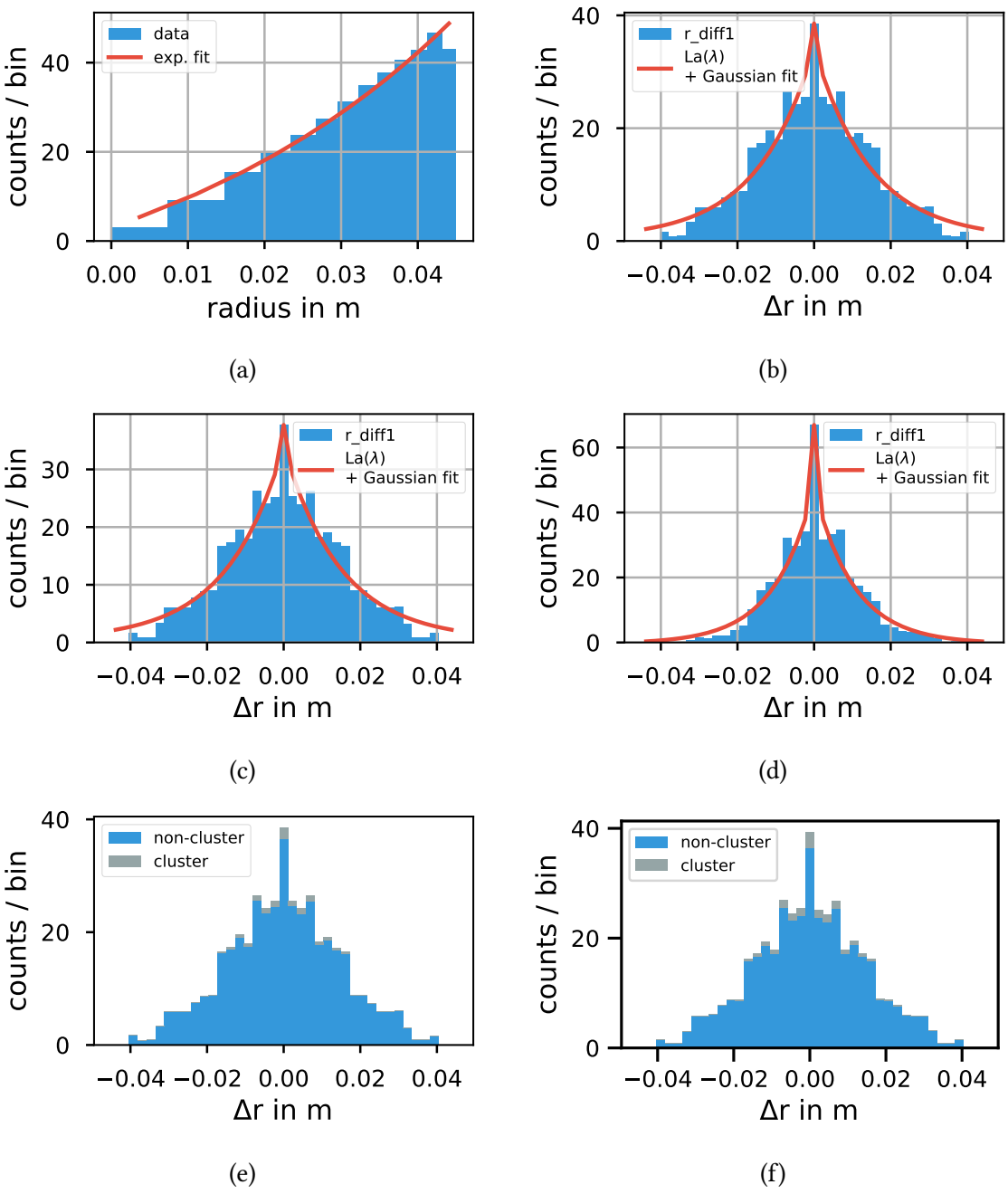


Figure A.8.: **Distribution of radii and difference of radii in the MS**. All histograms are normalized to 1. a): As the width of the detector rings are not equally spaced, the normalization has a penalty on the broader inner rings, which causes the distribution of the radii to follow an exponential (or half-Gaussian) distribution. b): Distribution of radii differences of all events. No clear excess at 0 is visible, compared to the distribution in the PS in figure 6.3. Therefore the ratio of contribution of radon induced events is smaller. c), d): Distribution of radii differences after DBSCAN cluster cut. Non-cluster events on the left c) and cluster events on the right d). e), f): Stacked histograms of non-cluster and cluster events with valve closed e)) and opened f).

A.3.2 Sensitivity study with the pre-spectrometer due to tritium decays

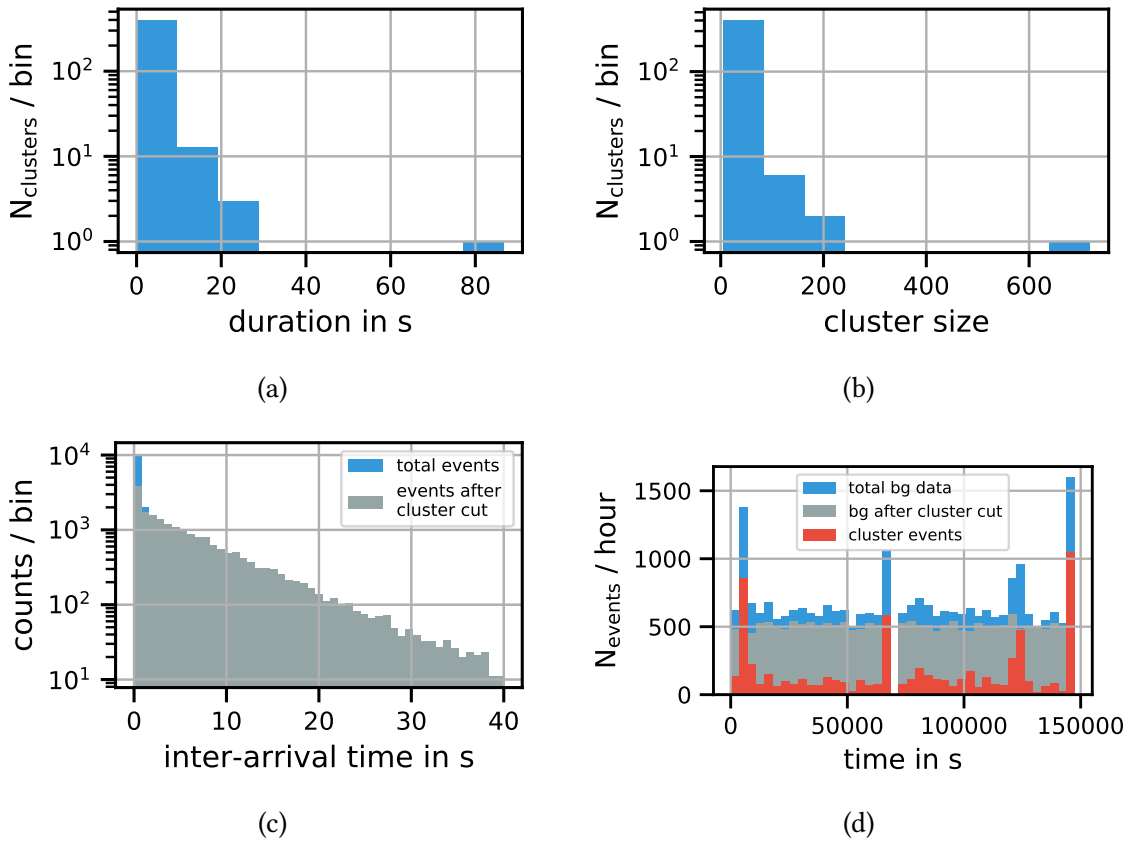


Figure A.9.: **Cluster information for runs 35160 - 35171 (elevated pressure), PS.** a), b): Histogram of cluster durations on the left a) and histogram of cluster size on the right b). c), d): Histogram of inter-arrival times on the left c) and rate trend of raw data and after cluster cut with binwidth of one hour is shown d).

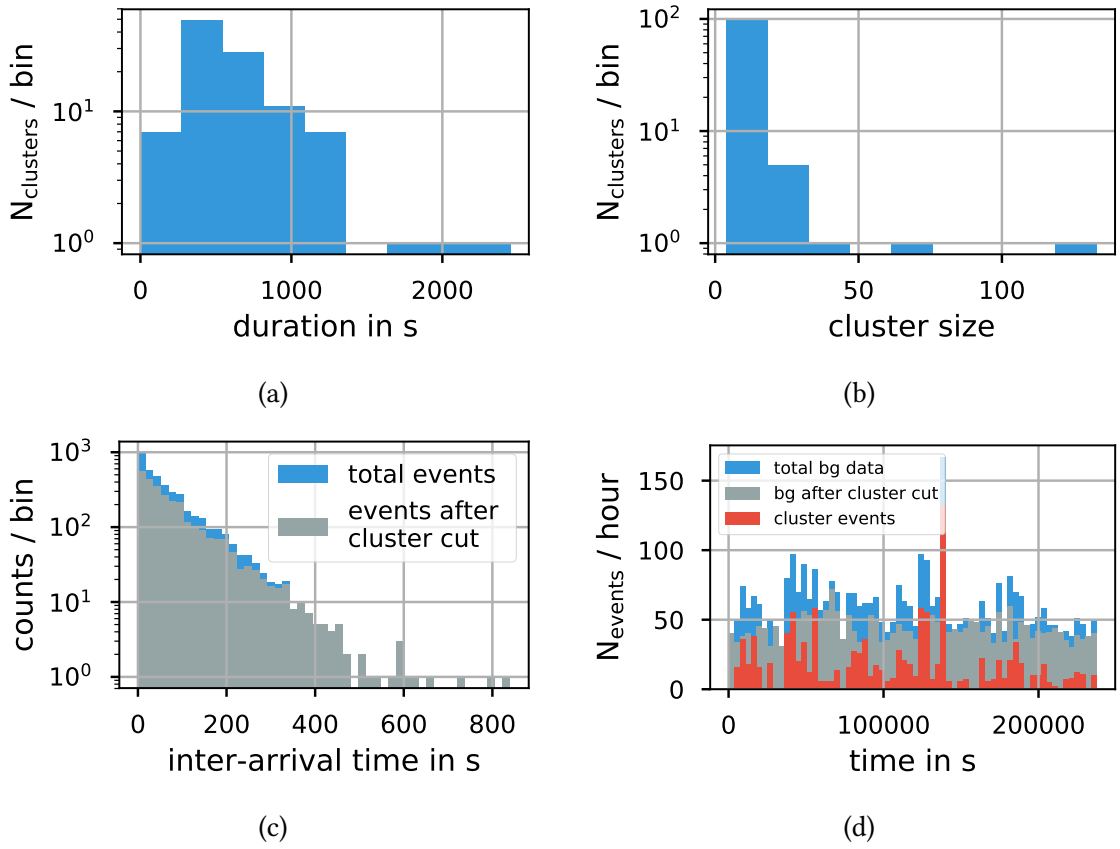


Figure A.10.: **Cluster information for runs 35173 - 35189 (nominal pressure), PS.** a), b): Histogram of cluster durations on the left a) and histogram of cluster size on the right b). c), d): Histogram of inter-arrival times on the left c) and rate trend of raw data and after cluster cut with bin width of one hour is shown d).

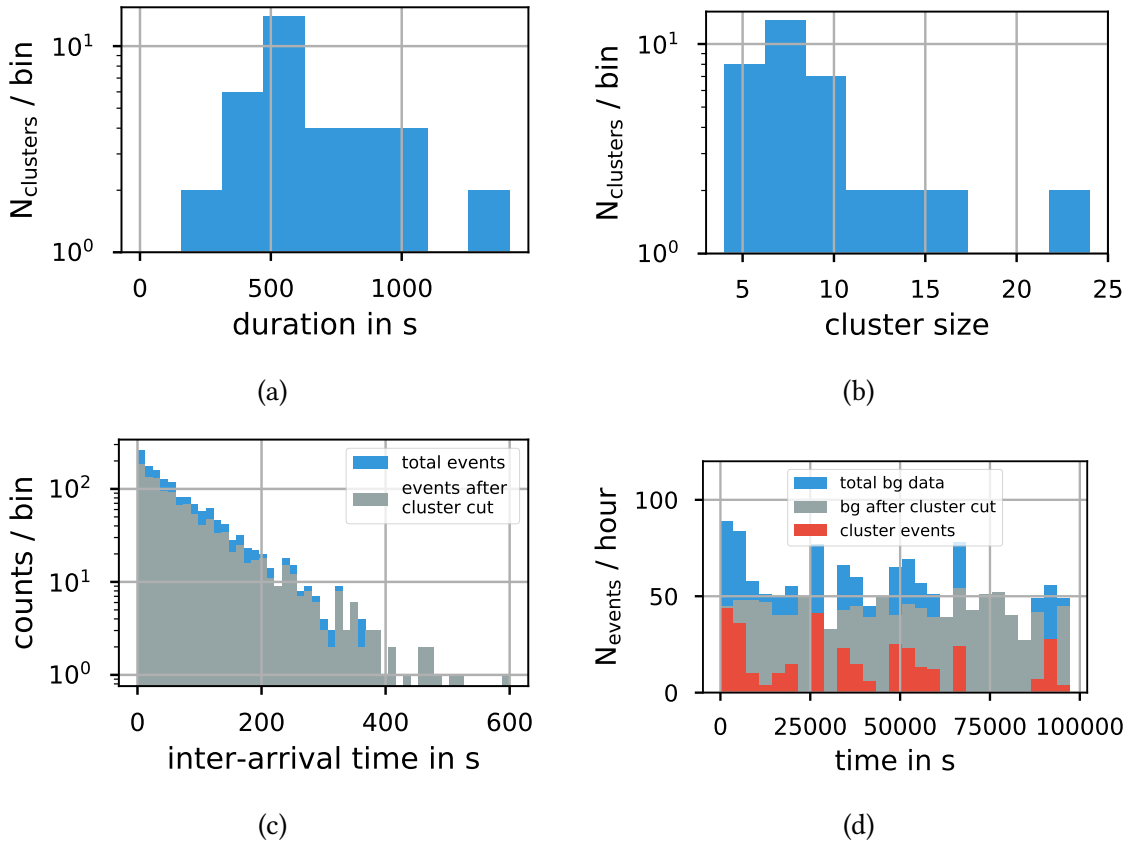


Figure A.11.: **Cluster information for runs 35362 - 35378 (nominal pressure), PS.** a), b): Histogram of cluster durations on the left a) and histogram of cluster size on the right b). c), d): Histogram of inter-arrival times on the left c) and rate trend of raw data and after cluster cut with bin width of one hour is shown d).

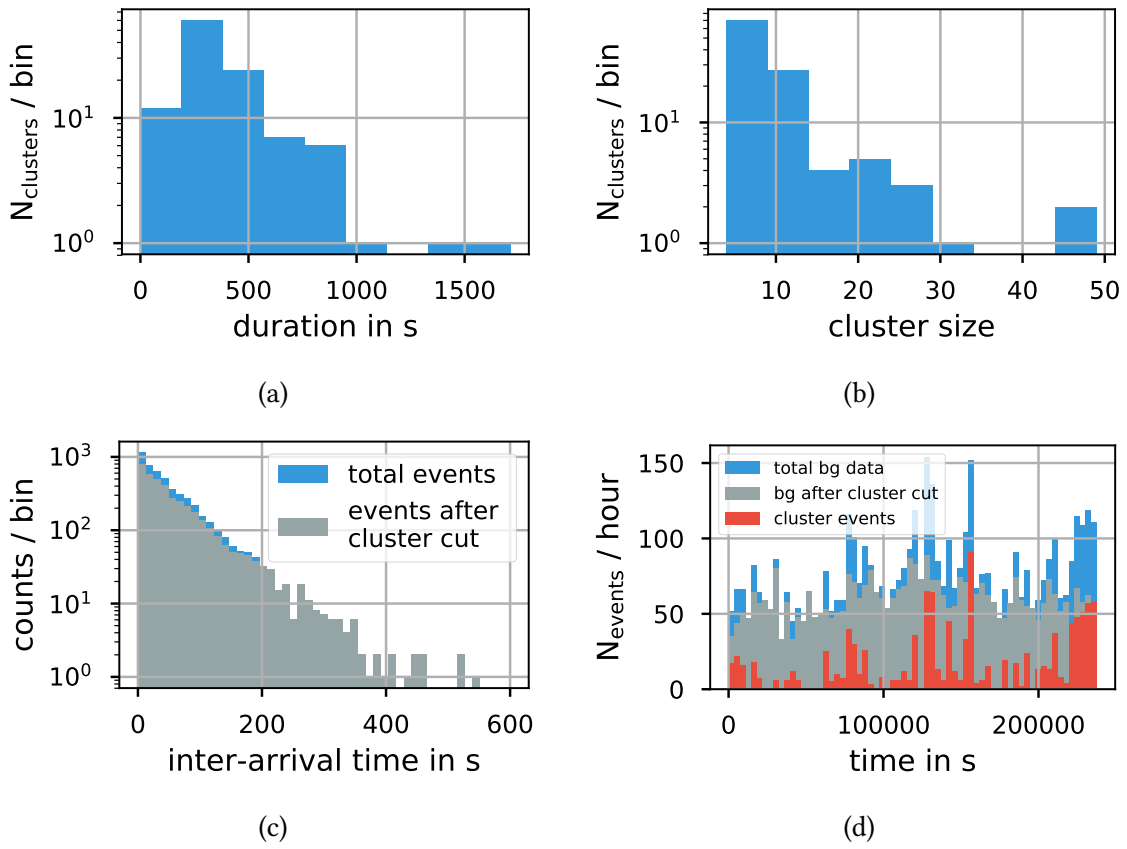


Figure A.12.: **Cluster information for runs 40270 - 40286 (nominal pressure), PS.** a), b): Histogram of cluster durations on the left (a)) and histogram of cluster size on the right (b)). c), d): Histogram of inter-arrival times on the left (c)) and rate trend of raw data and after cluster cut with binwidth of one hour is shown (d)).

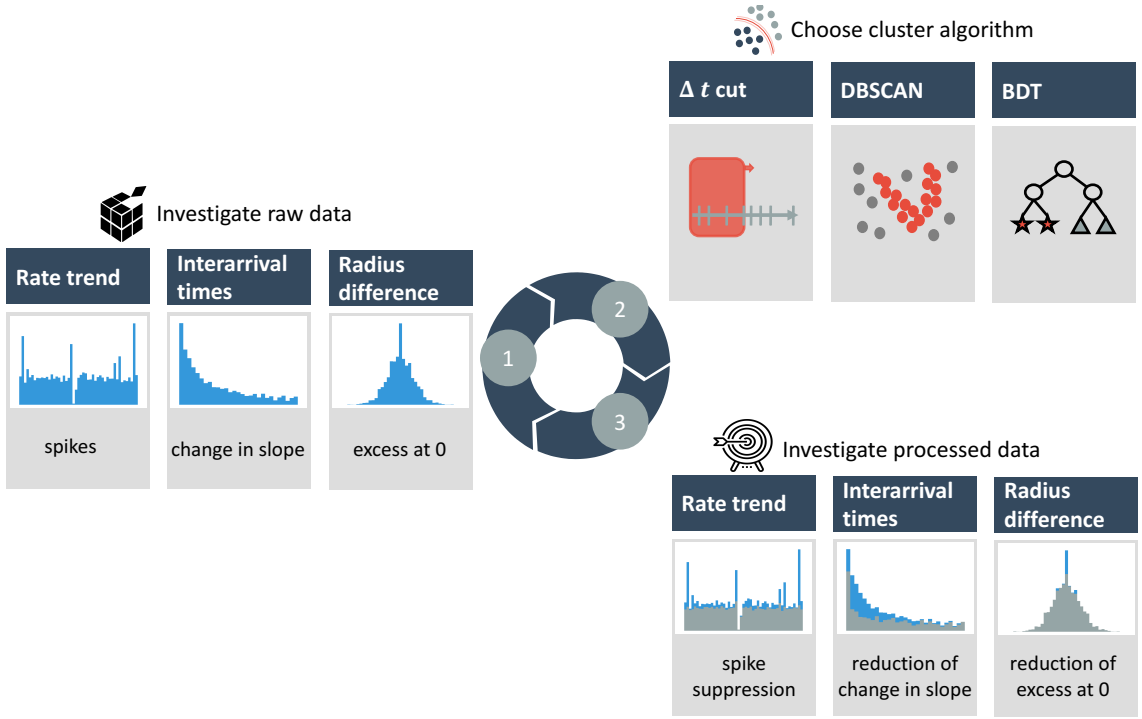


Figure A.13.: **Cluster manual.** This sketch illustrates a workflow for identifying the presence of clusters, which tools are available to perform a cluster cut, and what characteristics to look for in the data after the cluster cut.

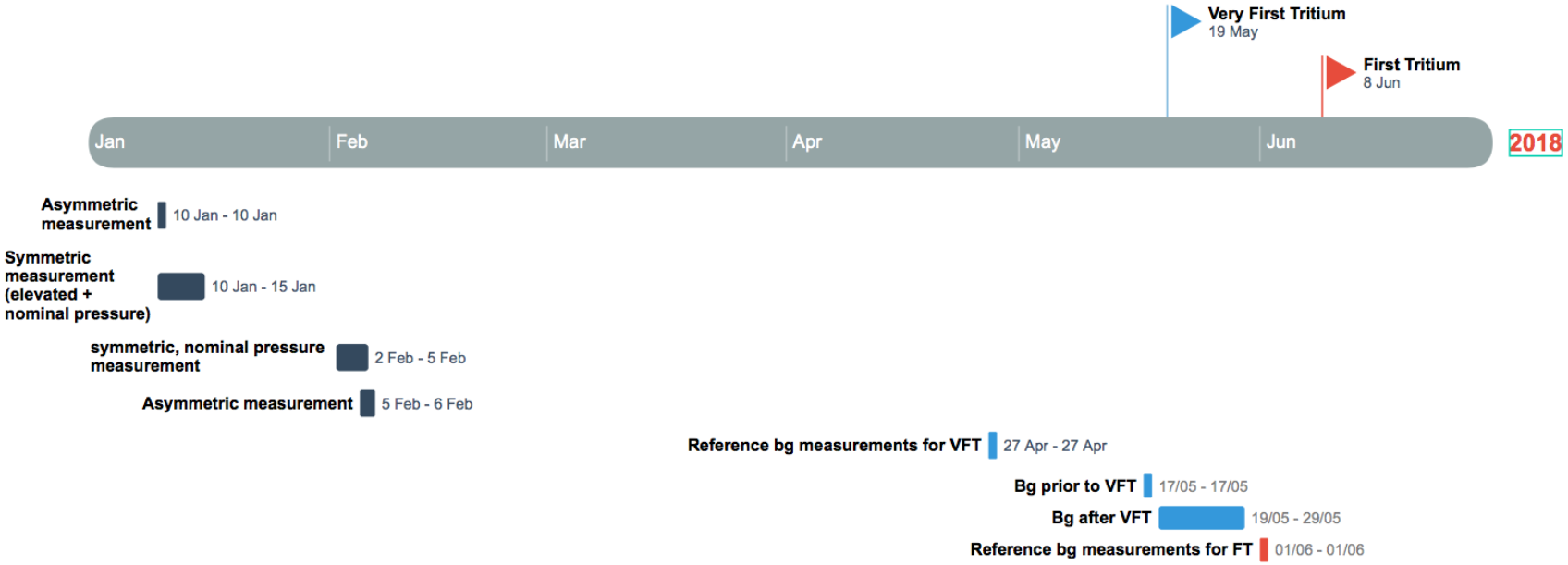


Figure A.14.: **Timeline of measurements analysed in this work**. The dark blue color indicates measurements with closed Valve, the light blue ones are within the context of VFT and the red ones correspond to FT, respectively

B List of Acronyms

ADC	Analog-to-Digital Converter 24
BDT	Boosted Decision Tree 55
BEANS	Building Analysis Sequence 52
BIXS	Beta Induced X-Ray Spectroscopy
CC	Charged-Current 8
cdf	cumulative distribution function 89
CKM	Cabibbo-Kobayashi-Maskawa 5
CMS	Compact Muon Solenoid 34
cps	counts per second 78
CPS	Cryogenic Pumping Section 20
CV	Cross-Validation 57
DAQ	Data Acquisition 24
DBMS	Database Management System 24
DBSCAN	Density-Based Spatial Clustering of Applications with Noise 45
DET	Detector Solenoid 24
DONUT	Direct Observation of the Nu Tau
DPS	Differential Pumping Section 20
E-Gun	Electron Gun 20
EMCS	Earth Magnetic field Compensation System 24
FLT	First Level Trigger 24
FN	False Negatives 48
FP	False Positive 48
FPR	False Positive Rate 59
FPD	Focal-Plane Detector 1
FPGA	Field Programmable Gate Array 24
FSD	Final States Distribution 15
FT	First Tritium Campaign 20

FT-ICR Fourier Transform-Ion Cyclotron Resonance . . . 21
GERDA The GERmanium Detector Array . . . 13
IO Inverted Ordering . . . 9
KATRIN KArlsruhe TRItium Neutrino experiment . . . i
LARA LAser RAman spectroscopy setup . . . 20
LDA Linear Discriminant Analysis . . . 55
LFCS Low Field Coil System . . . 24
LH Left-Handed . . . 11
MAC-E filter Magnetic Adiabatic Collimation with Electrostatic filter . . . 1
mcps milli-counts per second . . . 71
MS Main Spectrometer . . . 2
NC Neutral-Current . . . 8
NEG Non-Evaporable Getter . . . 2
NO Normal Ordering . . . 9
PAE Post Acceleration Electrode . . . 24
PCH PinCH solenoid . . . 24
pdf probability density function . . . 89
PMNS Pontecorvo-Maki-Nakagawa-Sakata . . . 5
PMT Photo Multiplier Tube . . . 4
PS Pre-Spectrometer . . . 1
PS1 Pre-Spectrometer Solenoid 1 . . . 23
PS2 Pre-Spectrometer Solenoid 2 . . . 23
QCD Quantum Chromo Dynamics . . . 4
RH Right-Handed . . . 11
ROC Receiver Operating Characteristic . . . 59
ROI Region Of Interest . . . 25
SDS Spectrometer and Detector Section . . . 25
SLT Second Level Trigger . . . 24
SM Standard Model of Particle Physics . . . 3
SNO Sudbury Neutrino Observatory . . . 7
SSM Standard Solar Model . . . 7
STS Source and Transport Section . . . 20
SVM Support Vector Machine . . . 55
TLK Tritium Laboratory Karlsruhe . . . 17

TMP	Turbo Molecular Pump	21
TN	True Negative	48
TP	True Positive	48
TPR	True Positive Rate	59
UHV	Ultra-High Vacuum	22
VFT	Very-First Tritium campaign	38
WGTS	Windowless Gaseous Tritium Source	15

List of Figures

List of Tables

Bibliography

[AB15] J.F. Amsbaugh et al. “Focal-plane detector system for the KATRIN experiment”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 778 (2015), pp. 40–60. DOI: 10.1016/j.nima.2014.12.116.

[AB18a] M. Agostini et al. “Improved Limit on Neutrinoless Double-β Decay of ^{76}Ge from GERDA Phase II”. In: *Phys. Rev. Lett.* 120 (2018). DOI: 10.1103/PhysRevLett.120.132503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.132503.

[AB18b] M. Arenz et al. “First transmission of electrons and ions through the KATRIN beamline”. In: *Journal of Instrumentation* 13.04 (2018), P04020. URL: http://stacks.iop.org/1748-0221/13/i=04/a=P04020.

[AB18c] M. Arenz et al. “The KATRIN superconducting magnets: overview and first performance results”. In: *Journal of Instrumentation* (2018). URL: http://stacks.iop.org/1748-0221/13/i=08/a=T08005.

[AH93] P. Anselmann et al. “GALLEX solar neutrino observations. The results from GALLEX I and early results from GALLEX II”. In: *Physics Letters B* 314.3 (1993), pp. 445–458. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(93)91264-N. URL: http://www.sciencedirect.com/science/article/pii/037026939391264N.

[Aha13] B. Aharmim et al. “Combined analysis of all three phases of solar neutrino data from the Sudbury Neutrino Observatory”. In: *Physical Review C* 88 (2 2013), p. 025501. DOI: 10.1103/PhysRevC.88.025501.

[Ahm16] Momin Ahmad. “Simulations of tritium induced background in the KATRIN pre-spectrometer”. Masters Thesis. Karlsruher Institut für Technologie (KIT), 2016.

[Ase12] V. N. Aseev et al. “Measurement of the electron antineutrino mass in tritium beta decay in the Troitsk nu-mass experiment”. In: *Physics of Atomic Nuclei* 75.4 (2012), pp. 464–478. ISSN: 1562-692X. DOI: 10.1134/S1063778812030027.

[Ash17] Ali Ashtari Esfahani et al. “Determining the neutrino mass with cyclotron radiation emission spectroscopy—Project 8”. In: *J. Phys.* G44.5 (2017), p. 054004. DOI: 10.1088/1361-6471/aa5b4f. arXiv: 1703.02037 [physics.ins-det].

[Bah64] John N. Bahcall. “Solar Neutrino Cross Sections and Nuclear Beta Decay”. In: *Phys. Rev.* 135 (1B July 1964), B137–B146. DOI: 10.1103/PhysRev.135.B137. URL: https://link.aps.org/doi/10.1103/PhysRev.135.B137.

[Beh16] Jan D. Behrens. “Design and commissioning of a mono-energetic photoelectron source and active background reduction by magnetic pulse at the KATRIN spectrometers”. PhD thesis. Westfälische Wilhelms-Universität Münster, 2016. URL: http://www.katrin.kit.edu/publikationen/phd%5C_behrens.pdf.

[BPT80] G Beamson, H Q Porter, and D W Turner. “The collimating and magnifying properties of a superconducting field photoelectron spectrometer”. In: *Journal of Physics E: Scientific Instruments* 13.1 (1980), p. 64. URL: http://stacks.iop.org/0022-3735/13/i=1/a=018.

[BS57] Hans A. Bethe and Edwin E. Salpeter. *Quantum Mechanics of One- and Two-Electron Atoms*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1957. ISBN: 978-3-662-12871-8. DOI: 10.1007/978-3-662-12869-5.

[CD98] Bruce T. Cleveland et al. “Measurement of the Solar Electron Neutrino Flux with the Homestake Chlorine Detector”. In: *The Astrophysical Journal* 496.1 (1998), p. 505. URL: http://stacks.iop.org/0004-637X/496/i=1/a=505.

[Cha32] J. Chadwick. “Possible Existence of a Neutron”. In: *Nature* 129 (1932), p. 312. DOI: 10.1038/129312a0.

[Chi08] Siu A. Chin. “Symplectic and energy-conserving algorithms for solving magnetic field trajectories”. In: *Phys. Rev. E* 77 (6 June 2008), p. 066401. DOI: 10.1103/PhysRevE.77.066401. URL: https://link.aps.org/doi/10.1103/PhysRevE.77.066401.

[CL16] Jun Cao and Kam-Biu Luk. “An overview of the Daya Bay reactor neutrino experiment”. In: *Nuclear Physics B* 908 (2016). Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015, pp. 62–73. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2016.04.034. URL: http://www.sciencedirect.com/science/article/pii/S0550321316300724.

[Dev18] Scikit-learn Developers. *DBSCAN implementation in scikit-learn*. Accessed: Oct 09, 2018. 2018. URL: http://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html.

[DG62] G. Danby et al. “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos”. In: *Phys. Rev. Lett.* 9 (1962), pp. 36–44. DOI: 10.1103/PhysRevLett.9.36.

[Dis18] Jerry Dischler. *Putting machine learning into the hands of every advertiser*. Accessed: Oct 25, 2018. 2018. URL: https://www.blog.google/technology/ads/machine-learning-hands-advertisers/.

[EK96] Martin Ester et al. “A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. KDD’96. Portland, Oregon: AAAI Press, 1996, pp. 226–231. URL: http://dl.acm.org/citation.cfm?id=3001460.3001507.

[Erh16] Moritz G. Erhard. “Influence of the magnetic field on the transmission characteristics and neutrino mass systematic of the KATRIN experiment”. PhD thesis. Karlsruher Institut für Technologie, 2016. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-650034.

[FG17] Daniel Furse et al. “Kassiopeia: a modern, extensible C++ particle tracking package”. In: *New Journal of Physics* 19.5 (2017), p. 053012. DOI: 10.1088/1367-2630/aa6950. URL: http://stacks.iop.org/1367-2630/19/i=5/a=053012.

[FH98] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[FMO82] G. Falcone, D. Matragrano, and A. Oliva. “Energy distribution of sputtered excited atoms”. In: *Lettere al Nuovo Cimento (1971-1985)* 35.16 (1982), pp. 465–468. ISSN: 1827-613X. DOI: 10.1007/BF02817286.

[For18] Scott Fortmann-Roe. *Understanding the Bias-Variance Tradeoff*. Accessed: Oct 09, 2018. 2018. URL: http://scott.fortmann-roe.com/docs/BiasVariance.html.

[FR18] Fabian Friedel et al. “Time dependent simulation of the flow reduction of D_2 and T_2 in the KATRIN experiment”. In: (July 2018). URL: https://arxiv.org/abs/1807.10126.

[Frä10] Florian M. Fränkle. “Background Investigations of the KATRIN Pre-Spectrometer”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2010. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-193929.

[Fur15] Daniel Furse. “Techniques for direct neutrino mass measurement utilizing tritium β-decay”. PhD thesis. Massachusetts Institute of Technology, 2015. URL: http://hdl.handle.net/1721.1/99313.

[GB17] L. Gastaldo et al. “The electron capture in 163Ho experiment – ECHo”. In: *The European Physical Journal Special Topics* 226.8 (June 2017), pp. 1623–1694. ISSN: 1951-6401. DOI: 10.1140/epjst/e2017-70071-y. URL: https://doi.org/10.1140/epjst/e2017-70071-y.

[GD13] F. Glück et al. “Electromagnetic design of the large-volume air coil system of the KATRIN experiment”. In: *New Journal of Physics* 15.8 (2013), p. 083025. DOI: 10.1088/1367-2630/15/8/083025.

[GG91] Paolo Gondolo and Graciela Gelmini. “Cosmic abundances of stable particles: Improved analysis”. In: *Nuclear Physics B* 360.1 (1991), pp. 145–179. DOI: https://doi.org/10.1016/0550-3213(91)90438-4. URL: http://www.sciencedirect.com/science/article/pii/0550321391904384.

[Gro15] Stefan Groh. “Modeling of the response function and measurement of transmission properties of the KATRIN experiment”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2015. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-465464.

[Hac15] Moritz Hackenjos. “Die differentielle Pumpstrecke des KATRIN-Experiments - Inbetriebnahme und Charakterisierung des supraleitenden Magnetsystems”. Master Thesis. Karlsruher Institut für Technologie (KIT), 2015. URL: http://www.katrin.kit.edu/publikationen/MaT_Hackenjos.pdf.

[Har15] Fabian Harms. “Characterization and Minimization of Background Processes in the KATRIN Main Spectrometer”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2015. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-500274.

[Hei15] Johannes Heizmann. “Modeling of inelastic electron-hydrogen scattering and of the energy loss function of 18.6 keV electrons for the KATRIN experiment”. Bachelors Thesis. Karlsruher Institut für Technologie (KIT), 2015.

[Hei18] Florian Heizmann. *PhD thesis, in preparation.* PhD thesis. Karlsruher Institut für Technologie, 2018.

[Hin18] Dominic Hinz. *Masters Thesis, in preparation.* Masters Thesis. Karlsruher Institut für Technologie (KIT), 2018.

[Höt12] Markus Hötzel. “Simulation and analysis of source-related effects for KATRIN”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2012. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-312594.

[HS17] F. Heizmann and H. Seitz-Moskaliuk. “The Windowless Gaseous Tritium Source (WGTS) of the KATRIN experiment”. In: *Journal of Physics: Conference Series* 888.1 (2017), p. 012071. DOI: 10.1088/1742-6596/888/1/012071.

[HTF17] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* 12th printing. 2017. ISBN: 978-0-387-84858-7. DOI: 10.1007/978-0-387-84858-7. URL: https://web.stanford.edu/~hastie/ElemStatLearn/.

[IMi18] IMinuit. *iminuit – A Python interface to Minuit.* https://github.com/iminuit/iminuit. Accessed: Mar 05, 2018. 2018.

[Jan15] Alexander Jansen. “The Cryogenic Pumping Section of the KATRIN Experiment - Design Studies and Experiments for the Commissioning”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2015. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-471467.

[JR75] F. James and M. Roos. “Minuit – a system for function minimization and analysis of the parameter errors and correlations”. In: *Computer Physics Communications* 10 (Dec. 1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.

[JW17] Gareth James et al. *An Introduction to Statistical Learning. with Applications in R.* 7th printing. 2017. ISBN: 978-1-4614-7138-7. DOI: 10.1007/978-1-4614-7138-7. URL: https://www-bcf.usc.edu/~gareth/ISL/.

[Kag18] Kaggle. *Higgs Boson Machine Learning Challenge*. Accessed: Sep 11, 2018. 2018. URL: https://www.kaggle.com/c/higgs-boson.

[KAT05] KATRIN collaboration. *KATRIN Design Report*. FZKA scientific report 7090. 2005. URL: http://bibliothek.fzk.de/zb/berichte/FZKA7090.pdf.

[KB05] C. Kraus et al. "Final results from phase II of the Mainz neutrino mass search in tritium β decay". In: *The European Physical Journal C - Particles and Fields* 40.4 (2005), pp. 447–468. DOI: 10.1140/epjc/s2005-02139-7.

[Kel82] Roger Kelly. "On the origin of sputtered excited atoms". In: *Nuclear Instruments and Methods in Physics Research* 194.1-3 (1982), pp. 583–588. ISSN: 01675087. DOI: 10.1016/0029-554X(82)90585-7.

[Kla18] Robert D. Klauber. *The Seesaw Mechanism*. Accessed: Apr 09, 2018. 2018. URL: http://www.quantumfieldtheory.info/TheSeesawMechanism.htm.

[Kle18] Manuel Klein. "Tritium ions in KATRIN: blocking, removal and detection". *PhD thesis, in preparation*. PhD thesis. Karlsruher Institut für Technologie, 2018.

[Kna62] W. Knauer. "Mechanism of the Penning Discharge at Low Pressures". In: *Journal of Applied Physics* 33.6 (1962), pp. 2093–2099. DOI: 10.1063/1.1728902. eprint: https://doi.org/10.1063/1.1728902. URL: https://doi.org/10.1063/1.1728902.

[Kod01] K. Kodama et al. "Observation of tau neutrino interactions". In: *Phys. Lett.* B504 (2001), pp. 218–224. DOI: 10.1016/S0370-2693(01)00307-0. arXiv: hep-ex/0012035 [hep-ex].

[Kos14] Timo Koski. *Lecture Note: Probability Theory and Random Processes at KTH. for sf2940 Probability Theory*. 2014th ed. Stockholm, Sweden, 2014.

[Kuc16] Laura Kuckert. "The Windowless Gaseous Tritium Source of the KATRIN Experiment – Characterisation of Gas Dynamical and Plasma Properties". PhD thesis. Karlsruher Institut für Technologie (KIT), 2016. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-650776.

[LD15] Yisheng Lv et al. "Traffic Flow Prediction With Big Data: A Deep Learning Approach". In: *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS* 16.2 (2015), pp. 865–873. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6894591.

[LNB18] Robert Layton, Joel Nothman, and Lars Buitinck. *Comparing different clustering algorithms on toy datasets*. Accessed: Jul 09, 2018. 2018. URL: http://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html#sphx-glr-auto-examples-cluster-plot-cluster-comparison-py.

[LU00] L. L. Lucas and M. P. Unterweger. "Comprehensive Review and Critical Evaluation of the Half-Life of Tritium". In: *Journal of Research of the National Institute of Standards and Technology* 105.4 (July 2000), pp. 541–549. DOI: 10.6028/jres.105.043. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877155/.

[LxC10] LxCat. *H cross sections extracted from PROGRAM Morgan (Kinema Research & Software), UPDATED Jul 2010*. Accessed: Sep 13, 2018. 2010. URL: http://fr.lxcat.net/.

[LxC12] LxCat. *H2 cross sections extracted from PROGRAM MAGBOLTZ, VERSION 8.97 Oct 2012*. Accessed: Sep 13, 2018. 2012. URL: http://fr.lxcat.net/.

[LxC13] LxCat. *Ar cross sections extracted from PROGRAM MAGBOLTZ, VERSION 8.97 Sept 2011*. Accessed: Sep 13, 2018. 2013. URL: http://fr.lxcat.net/.

[MD13a] S. Mertens et al. "Background due to stored electrons following nuclear decays in the KATRIN spectrometers and its impact on the neutrino mass sensitivity". In: *Astroparticle Physics* 41 (2013), pp. 52–62. DOI: 10.1016/j.astropartphys.2012.10.005.

[MD13b] S. Mertens et al. "Background due to stored electrons following nuclear decays in the KATRIN spectrometers and its impact on the neutrino mass sensitivity". In: *Astroparticle Physics* 41 (2013), pp. 52–62. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.10.005. URL: http://www.sciencedirect.com/science/article/pii/S0927650512001892.

[MG11] Vincent Michel et al. "A supervised clustering approach for fMRI-based inference of brain states". In: *CoRR* abs/1104.5304 (2011). arXiv: 1104.5304. URL: http://arxiv.org/abs/1104.5304.

[MNS62] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. "Remarks on the Unified Model of Elementary Particles". In: *Progress of Theoretical Physics* 28.5 (1962), pp. 870–880. DOI: 10.1143/PTP.28.870. eprint: /oup/backfile/content_public/journal/ptp/28/5/10.1143/ptp.28.870/2/28-5-870.pdf. URL: http://dx.doi.org/10.1143/PTP.28.870.

[MS85] S. P. Mikheyev and A. Y. Smirnov. *Resonance enhancement of oscillations in matter and solar neutrino spectroscopy*. 1985.

[Nob18] Nobelprize.org. *The Nobel Prize in Physics 2015*. Sept. 12, 2018. URL: https://www.nobelprize.org/prizes/physics/2015/summary/.

[NXC09] E.W.T. Ngai, Li Xiu, and D.C.K. Chau. "Application of data mining techniques in customer relationship management: A literature review and classification". In: *Expert Systems with Applications* 36.2, Part 2 (2009), pp. 2592–2602. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2008.02.021. URL: http://www.sciencedirect.com/science/article/pii/S0957417408001243.

[OLu14] Cian O'Luanaigh. *Higgs boson machine-learning challenge*. Accessed: Sep 11, 2018. 2014. URL: https://home.cern/about/updates/2014/05/higgs-boson-machine-learning-challenge.

[Per09] Donald H. Perkins. *Particle astrophysics*. 2. ed. Oxford master series in physics ; no. 10. Includes bibliographical references and index. - Previous ed.: 2003. Oxford: Oxford University Press, 2009. ISBN: 0-19-954546-4; 978-0-19-954546-9; 0-19-954545-6; 978-0-19-954545-2.

[PKW64] W. E. F. Pauli, R. Kronig, and V. F. Weisskopf. *Collected scientific papers*. Offener Brief an die Gruppe der Radioaktiven bei der Gauvereinstagung zu Tübingen (datiert 4. Dez. 1930). New York, NY: Interscience, 1964.

[Pon58] B. Pontecorvo. "Inverse beta processes and nonconservation of lepton charge". In: *Sov. Phys. JETP* 7 (1958). [Zh. Eksp. Teor. Fiz.34,247(1957)], pp. 172–173.

[PV11] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[RC53] F. Reines and C. L. Cowan. "Detection of the Free Neutrino". In: *Physical Review* 92 (3 1953), pp. 830–831. DOI: 10.1103/PhysRev.92.830.

[RC97] F. Reines and C. Cowan. "The Reines-Cowan experiments: Detecting the Poltergeist". In: *Los Alamos Sci.* 25 (1997), pp. 4–27. URL: https://fas.org/sgp/othergov/doe/lanl/pubs/00326606.pdf.

[Ren11] Pascal Renschler. "KESS - A new Monte Carlo simulation code for low-energy electron interactions in silicon detectors". PhD thesis. Karlsruher Institut für Technologie (KIT), 2011. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-249597.

[Röl15] Marco Röllig. "Tritium analytics by beta induced X-ray spectrometry". PhD thesis. Karlsruher Institut für Technologie (KIT), 2015. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-540500.

[Ros57] Frank Rosenblatt. "The Perceptron – a perceiving and recognizing automaton". In: *empty* (1957). URL: https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf.

[Sal68] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc.* C680519 (1968), pp. 367–377.

[SC08] Zhan-Li Sun et al. "Sales forecasting using extreme learning machine with applications in fashion retailing". In: *Decision Support Systems* 46.1 (2008), pp. 411–419. ISSN: 0167-9236. DOI: https://doi.org/10.1016/j.dss.2008.07.009. URL: http://www.sciencedirect.com/science/article/pii/S0167923608001371.

[Sch13] Magnus Schlösser. "Accurate calibration of the Raman system for the Karlsruhe Neutrino Experiment". PhD thesis. Karlsruher Institut für Technologie (KIT), 2013. URL: https://publikationen.bibliothek.kit.edu/1000034967/2598164.

[Sch14] Johannes S. Schwarz. "The Detector System of the KATRIN Experiment - Implementation and First Measurements with the Spectrometer". PhD thesis. Karlsruher Institut für Technologie (KIT), 2014. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-427724.

[Sch18] Christophe Schwachtgen. "Simulation of the vacuum properties of the KATRIN Pre-Spectrometer for different gas types". Bachelors Thesis. Karlsruher Institut für Technologie (KIT), 2018.

[Sho13] Milad Shokouhi. “Learning to Personalize Query Auto-completion”. In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’13. Dublin, Ireland: ACM, 2013, pp. 103–112. ISBN: 978-1-4503-2034-4. DOI: 10.1145/2484028.2484076. URL: http://doi.acm.org/10.1145/2484028.2484076.

[SR02] Margaret A. Shipp et al. “Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning”. In: *Nature Medicine* 8 (Jan. 2002), 68 EP -. URL: http://dx.doi.org/10.1038/nm0102-68.

[SV80] J. Schechter and J. W. F. Valle. “Neutrino masses in SU(2) $\bigotimes$ U(1) theories”. In: *Phys. Rev. D* 22 (9 Nov. 1980), pp. 2227–2235. DOI: 10.1103/PhysRevD.22.2227. URL: https://link.aps.org/doi/10.1103/PhysRevD.22.2227.

[TH18] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[The04] The MACRO Collaboration. “Measurements of atmospheric muon neutrino oscillations, global analysis of the data collected with MACRO detector”. In: *The European Physical Journal C - Particles and Fields* 36.3 (Aug. 2004), pp. 323–339. ISSN: 1434-6052. DOI: 10.1140/epjc/s2004-01951-9. URL: https://doi.org/10.1140/epjc/s2004-01951-9.

[Tro18] Nikolaus Trost. “Modeling and measurement of Rydberg-State mediated Background at the KATRIN Main Spectrometer”. PhD thesis. Karlsruher Institut für Technologie, 2018.

[Ubi11] Marta Ubieto Díaz. “Off-line commissioning of a non-destructive FT-ICR detection system for monitoring the ion concentration in the KATRIN beamline”. PhD thesis. Ruperto-Carola University of Heidelberg, 2011. URL: http://pubman.mpdl.mpg.de/pubman/item/escidoc:1255588:2/component/escidoc:1255587/2Ubieto_Thesis.pdf.

[Val09] Kathrin Valerius. “Spectrometer-related background processes and their suppression in the KATRIN experiment”. PhD thesis. Westfälische Wilhelms-Universität Münster, 2009. URL: http://nbn-resolving.de/urn:nbn:de:hbz:6-28479494638.

[Val11] K. Valerius. “The wire electrode system for the KATRIN main spectrometer”. In: *Progress in Particle and Nuclear Physics* 64.2 (2011). Neutrinos in Cosmology, in Astro, Particle and Nuclear Physics: International Workshop on Nuclear Physics, 31st course, pp. 291–293. DOI: 10.1016/j.ppnp.2009.12.032.

[Wan13] Nancy Wandkowsky. “Study of background and transmission properties of the KATRIN spectrometers”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2013. URL: http://nbn-resolving.org/urn:nbn:de:swb:90-366316.

[WB94] A. Wucher et al. “Detection of sputtered metastable atoms by autoionization”. In: *Physical Review A* 49.3 (1994), pp. 2188–2190. ISSN: 1050-2947. DOI: 10.1103/PhysRevA.49.2188.

[Wei08] Steven Weinberg. *Cosmology*. Oxford University Press, Feb. 21, 2008, p. 624. ISBN: 9780198526827.

[Wei67] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[Wen10] R. Wendell et al. "Atmospheric neutrino oscillation analysis with subleading effects in Super-Kamiokande I, II, and III". In: *Physical Review D* 81 (9 2010), p. 092004. DOI: 10.1103/PhysRevD.81.092004.

[Win11] Alexander Windberger. "Berechnungen und Simulationen zum Verhalten von Ionen in der differenziellen Pumpstrecke des KATRIN-Experiments". Diploma Thesis. Karlsruher Institut für Technologie (KIT), 2011. URL: http://www.katrin.kit.edu/publikationen/dth-windberger.pdf.

Acknowledgements

At the end, I want to thank the KATRIN collaboration for giving me the chance to gain an insight into the challenging efforts of neutrino mass measurements and the possibility to contribute to this great project. It was such an exciting moment, when we saw the first tritium spectrum during VFT and the stress relief in all these happy faces. In this moment, it became visible how hard the scientists at KATRIN fight to reach their ambitious goals. Special thanks go to

- Prof. Dr. Guido Drexlin for permitting me to do this thesis at the KATRIN experiment,
- Prof. Dr. Husemann for agreeing to act as second reviewer,
- Florian Fränkle for his supervision, proofreading, and answering questions about background related processes,
- the Richard-Winter Stiftung for financial support,
- Carsten Röttele (Stahl) for chauffeur service, and Lutz Schimpf for sharing the office,
- Leonard Köllenberger for an unforgettable analysis workshop in Munich,
- Stephanie Hickford for accurate proofreading,
- all other Masters students doing their work in parallel,
- the Approximierer for sharing and delighting the daily routine of a physics student,
- my flatmates Felix Laufer and Jonas Schneider for pure awesomeness,
- Anna Carle and my family for the support during this work.