Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR MATHEMATIK, INFORMATIK
UND NATURWISSENSCHAFTEN

Bachelorarbeit

Studying black body radiation backgrounds for the ALPS II TES Detector

Malte Thoms

malte.thoms@studium.uni-hamburg.de
Studiengang Physik (B.Sc.)
Matr.-Nr. 7279241
Fachsemester: 8

Erstgutachter: Dr. Manuel Meyer
Zweitgutachter: Dr. Gulden Othman

Abgabe: 17.04.2023

Abstract

The Any Light Particle Search II (ALPS II) experiment is located at DESY in Hamburg and focuses on axion detection. These theoretical particles are weakly interacting sub-eV particles and a possible candidate for cold dark matter. The goal of the ALPS II experiment is to converted laser photons into axions inside the presence of a strong magnetic field. The converted axions pass through a wall and beyond the wall the axions are reconverted to photons. These single photons are then measured by the TES. A major challenge is the low coupling constant of axions and photons and the resulting low signal to background ratio. Therefore, a precise understanding and analysis of the background is necessary, to obtain a sufficient statistical significance for the axion detection.
This thesis focuses on the analysis and distinction of background events and light events, consisting of 1064nm photons. All events are measured by the Transition-Edge-Sensor (TES), which is the key component of the ALPS II experiment.
The analysis is split into two parts. First, the background events were fitted and analysed using a Gaussian Mixture Model. This unsupervised machine learning method assigns the parameter of the fit and the associated background events to different clusters. Comparing the background clusters with the light events indicates that the source of some of the background signal are photons. These photons are most likely produced by the black body radiation of parts of the setup.
For the second part of the analysis a deep learning algorithm was trained. This algorithm is a fully convolutional neural network used to classify the background and light events. After optimizing the algorithm the neural network achieved an accuracy of 99,62 % and a true positive rate of 99,79 %. Only light like events, most likely from black body radiation and very unusual light events, are classified incorrectly.
The results demonstrate the possibility of a sufficient background discrimination with deep learning. With this a sufficient statistical significance can be achieved. This makes axion detection with the Transition-Edge-Sensor feasible.

Zusammenfassung

Das ALPS 2 Experiment befindet sich bei DESY in Hamburg und konzentriert sich auf den Nachweis von Axionen. Diese theoretischen Teilchen sind schwach wechselwirkende Teilchen im Sub-eV-Bereich und ein möglicher Kandidat für kalte dunkle Materie. Das Ziel des ALPS II-Experiments ist es, Laserphotonen in einem starken Magnetfeld in Axionen umzuwandeln. Die umgewandelten Axionen durchqueren eine Wand und werden jenseits dieser Wand wieder in Photonen umgewandelt. Diese einzelnen Photonen werden dann mit dem TES gemessen. Eine große Herausforderung ist die niedrige Kopplungskonstante von Axionen und Photonen und das daraus resultierende geringe Signal-Hintergrund-Verhältnis. Daher ist ein genaues Verständnis und eine Analyse des Hintergrunds erforderlich, um eine ausreichende statistische Signifikanz für den Axionennachweis zu erhalten.
Diese Arbeit konzentriert sich auf die Analyse und Unterscheidung von Hintergrundereignissen und Lichtereignissen, die aus 1064nm-Photonen bestehen. Alle Ereignisse werden mit dem TES gemessen, der die Schlüsselkomponente des Experiments ALPS II darstellt.
Die Analyse ist in zwei Teile gegliedert. Zunächst wurden die Hintergrundereignisse mithilfe eines Gaußschen Mischungsmodells angepasst und analysiert. Diese unüberwachte maschinelle Lernmethode ordnet die Parameter der Anpassung und die zugehörigen Hintergrundereignisse verschiedenen Clustern zu. Der Vergleich der Hintergrund Clusters mit den Lichtereignisse deutet darauf hin, dass die Quelle eines Teils der Hintergrundsignale Photonen sind. Diese Photonen werden höchstwahrscheinlich durch die Schwarzkörperstrahlung von Teilen der Anlage erzeugt.
Für den zweiten Teil der Analyse wurde ein Deep-Learning-Algorithmus trainiert. Bei diesem Algorithmus handelt es sich um ein vollständig gefaltetes neuronales Netzwerk, das zur Klassifizierung von Hintergrund- und Lichtereignissen verwendet wird. Nach der Optimierung des Algorithmus erreichte das neuronale Netzwerk eine Genauigkeit von 99,62 % und eine True-Positive-Rate von 99,79 %. Nur lichtähnliche Ereignisse, die höchstwahrscheinlich von Schwarzer-Körper-Strahlung stammen, und sehr ungewöhnliche Lichtereignisse werden falsch klassifiziert.
Die Ergebnisse zeigen die Möglichkeit einer ausreichenden Hintergrundunterscheidung mit Deep Learning. Damit kann eine ausreichende statistische Signifikanz erreicht werden. Dies macht die Axion-Detektion mit dem Transition-Edge-Sensor machbar.

Contents

Chapter 1

Introduction

The main goal of particle physics is to understand the smallest components of the universe and thereby the history of the universe. The Standard Model of particle physics is the best description of the elementary particles so far, but it only describes matter which is about 4% of the estimate energy of the universe [1]. Hence, the study beyond the standard model is a great part of particle physics, especially the field of dark matter research. Dark matter contributes about 25 % of the estimated energy of the universe and would explain astrophysical occurrences like the angular spectrum of the cosmological microwave background [1].

One possible dark matter candidate are the axions. These theoretical particles belong to the group of weakly interacting sub-eV particles (WISPs) and would be a solution to the Strong CP-problem [2]. The main property used for axion detection is the axion-photon conversion via the Primakoff effect.[3] The ALPS II experiment at DESY uses the Primakoff effect for a light shining through wall system. A photon from a laser is converted into a axion in a strong magnetic field and behind a wall reconverted into a photon. Due to the extremely low coupling constant of axion and photons only around 1 photon is reconverted per day. Thus, a detector capable of detecting these single photons is required. Two different detectors are foreseen in the ALPS 2 experiment. The heterodyne detector and the TES where only the latter will be subject of this thesis [4].

Based on the low photon count, the signal to background ratio is low too. Therefore, this thesis focuses on the analysis of the extrinsic background of the TES and background discrimination with machine learning. Since the sources of the background are unclear the analysis uses unsupervised machine learning to create clusters of background events. With these clusters the different sources of the background can be analyzed. In the second part of the thesis a supervised deep learning algorithm learns to classify background and light events. Thereby, enabling to discriminate the background effectively and showing the possibility to measure the low light event count with a high statistical significance.

Chapter 2

Theoretical background

This chapter focuses on the relevant physical background for this thesis. In the first section 2.1 the necessity of a extension of the Standard Model (SM) is discussed. Afterwards in section 2.2 the reason for dark matter as one possible extension of the SM is explained. In section 2.3 the axion and axion like particles are introduced as one candidate for such a extension of the Standard Model. Section 2.4 explains the ALPS 2 experiment and the associated experimental setup. Lastly, in section 2.5 the TES is introduced and the operating principle explained.

2.1 Beyond the Standard Model

The Standard Model (see fig. 1) of particle physics is the best description of the elementary particles and most of their interactions with each other so far [1]. It is split into 2 groups of particles. The fermions consists of quarks and leptons and their associated anti-particles. The other group are the bosons. Bosons are the force carriers that mediate the weak, strong, and electromagnetic interactions [5].

However the fourth fundamental force, the gravitational force, is not explained by the

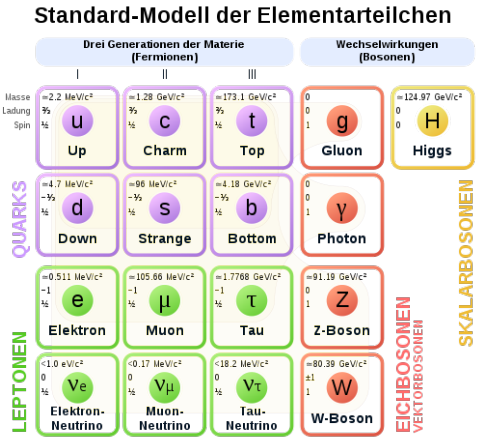


Figure 1: Standard Model of particle physics [6].

SM. Furthermore, dark matter is also not described in the SM, although it accounts for 25% of the energy of the universe [1]. Hence, an expansion of the SM, for example new particles, is necessary to accurately depict the universe. Therefore, the search for these new particles, especially dark matter candidates, is one of the main research field in particle physics.

2.2 Dark Matter

2.2.1 Evidence for dark matter

Even though no particle which can explain dark matter has been found until today plenty of cosmological phenomena indicate the existence of dark matter. One of the most significant is the cosmological microwave background (CMB) (see fig. 2)which is a remnant of the early days of the universe and consists of photons that froze in just before decoupling from the baryonic matter [1]. The fluctuation of the temperature as depicted in figure 2

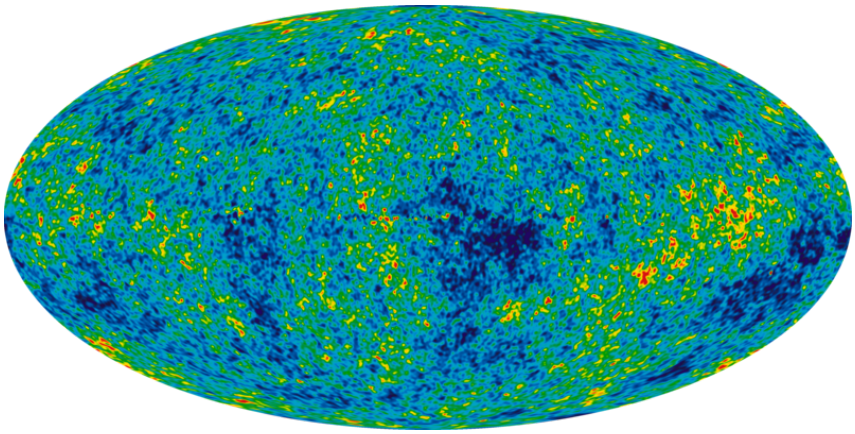


Figure 2: Temperature fluctuation of the CMB created from Wilkinson Microwave Anisotropy Probe data [7].

is only possible if on the one hand the universe is flat and on the other hand baryonic matter is about 5%, dark matter 25% and dark energy 70% of the energy in the universe [1].

2.2.2 Candidates for dark matter

To form the observed cosmological large-scale structure, dark matter must be mostly cold. In this context cold means that dark matter moves slowly compared to the speed of light (around 400 km/s) [8]. At the beginning of dark matter research it seemed reasonable that dark matter consist of stellar objects which were to faint to be discovered. These belong to the group called massive compact halo objects (MACHOs). Due to the data from the CMB, which demonstrated that a very high amount of MACHOs is needed to explain the missing mass, nonbaryonic candidates seem much more likely [1]. The two most popular

group of nonbaryonic candidates are the weakly interacting massive particles (WIMPs) and the weakly interacting sub-eV particles(WISP). The mass of WIMPs is postulated to be 10–1000 GeV and because it could easily satisfy all constraints on dark matter, imposed by cosmological and astrophysical experiments, a lot of research focused on the detection of WIMPs [9]. However, even after three decades of WIMP searches no significant evidence for them was found [10]. Therefore, many research groups focus on WISPs and especially axions [11].

2.3 Axion and axion like particles

Theories of the axion can roughly be divided into the quantum chromodynamics (QCD) axion and the axion like particles.
The QCD axion, also known as simply the axion, is postulated to explain the strong CP problem in quantum chromodynamics. The CP problem corresponds to a CP violating term in the Lagrangian. This term can be written as: [12]

$$\mathcal{L} = \theta \frac{1}{16\pi^2} F^a_{\mu\nu} \widetilde{F}^{\mu\nu a} \tag{1}$$

where

$$\widetilde{F}^a_{\mu\nu\sigma\rho} = \frac{1}{2} \epsilon_{\mu\nu\sigma\rho} F^{\sigma\rho a}. \tag{2}$$

With a bound of:

$$\theta \ll 10^{-9} \tag{3}$$

One explanation for this very low bound of θ is a new spin zero field which was introduced by Peccei and Quinn(1977) [13]. This field is denoted as the axion field and the corresponding pseudo-Goldstone boson is called axion. The field is endowed with a quasi shift symmetry $a \to a + \kappa f_a$, where f_a is the axion decay constant. This transformation parameter κ is arbitrary and can be chosen to remove the θ term. Thereby, the strong CP problem is solved [14].
The mass of a QCD axion is inversely proportional to the axion decay constant f_a. The mass is computed with

$$m_a \simeq \frac{f_\pi^2 m_\pi^2 m_u m_d}{f_a^2 (m_u m_d)^2} \tag{4}$$

where m_π is the pion mass, f_π the decay constant, and m_u the up and m_d the down quark masses. This yields to a result of $m_a = 5.691(51)\mu\text{eV}(10^{12}\text{GeV})/f_a$ [2]. Due to the high decay constant ($f_a \geq 10^{12}$ GeV), the mass of axions is low ($m_a < 1\text{eV}$).

Axion like particles (ALPS) share most properties with the QCD axion [2] and also acts as cold dark matter in nearly all cosmological relevant scales.[15] However ALPS do not solve the strong CP problem and as opposed to Eq.(4) the mass m_a and the decay constant f_a

are independent parameters[2]. For the rest of the thesis the QCD axion and ALPs will not be differentiated. The main property used in axion detection is the Primakoff effect and the associated photon axion coupling $g_{a\gamma\gamma}$. This coupling can be calculated by:[2]

$$g_{a\gamma\gamma} = \frac{a}{2\pi f_a}\left(\frac{E}{N} - 1,92(4)\right) \tag{5}$$

Where, E and N are the electromagnetic and QCD anomaly coefficients and a is the fine structure constant.

2.4 The ALPS II Experiment

The ALPS II experiment is stationed in the HERA tunnel at DESY in Hamburg. As most axion detection experiments, the ALPS II Experiment focuses on the photon axion coupling $g_{a\gamma\gamma}$ to detect axions. The fundamental layout of the experiment is a light shining through a wall setup as depicted in figure 3. The photon source is a 1064 nm

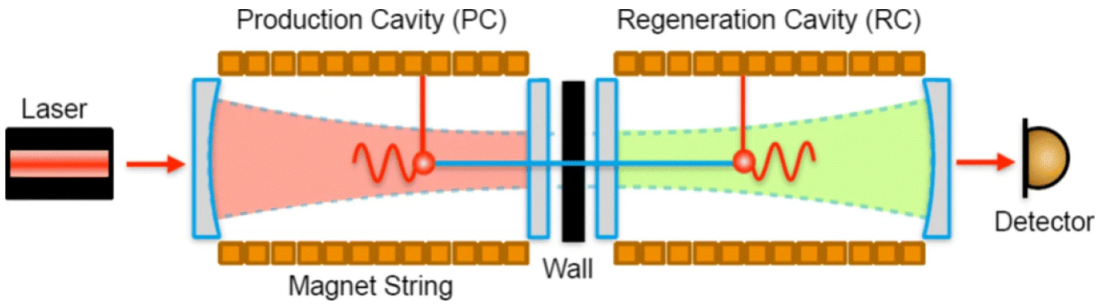


Figure 3: Sketch of the experimental setup for the ALPS 2 experiment. The light of the laser is enhanced in the production cavity(PC). The wall blocks the photons and behind the wall the axion photon reconversion is increased by the regeneration cavity (RC). The photons are detected by the detector [16].

laser with a optical power of up to 70 W. The emitted photons are locked in the 106m long production cavity(PC). On both sides of the PC and the regeneration cavity(RC) are superconducting cables which produce a static magnetic field of 5 Tesla [4]. Here, the photons get converted into axions in the presence of the magnetic field via the Primakoff effect [3]. Unlike photons, axions are weakly interacting and therefore they can pass through the wall and reconvert into photons in the static magnetic field of the RC. Since ALPS II targets a photon axion coupling $g_{a\sigma\sigma}$ of 2×10^{11} GeV^{-1} the numbers of expected reconverted photons is around 1 photon per day [4]. These single photons are measured by the ALPS II detector.

2.5 ALPS II Detector

To detect these single low energy photons, the ALPS II detector needs to be a very sensitive detector with a high efficiency and very low dark count rate with stability over

a couple of days. Two different detectors are foreseen in the ALPS II experiment. The heterodyne detector [17] and the Transition-Edge-Sensor(TES) where the latter will be the focus of this thesis [4].

2.5.1 Superconductivity

The measurement of single photons by the TES is only possible trough superconductivity which reduces the resistance of a material to zero. Most elements below a certain critical temperature T_c become superconducting [18]. A superconductor is on the one hand perfectly conducting that means it has zero resistance, and on the other hand perfectly diamagnetic, which means external magnetic fields remain on the surface of the superconductor. This effect was first explained with the BSC-theory by Bardeen, Cooper and Schrieffer in 1957. The Bardeen–Cooper–Schrieffer-theory is based on electron pairs, the so called Copper pairs. Together they move trough a grid in such a way that the first electron polarize the grid and thereby enabling the second electron to more easily follow it. The strength of the interaction of the electron pair determines the critical temperature T_c [19].

2.5.2 TES

The core part of the TES are two tungsten chips, which are kept near their critical temperature T_c of about 140 mK and are connected to a thermal bath. Therefore, if the TES absorbs energy from, for example a photon, the temperature of the tungsten chip is increased above T_c and the tungsten becomes normal conducting as can be seen in figure 4 [4].

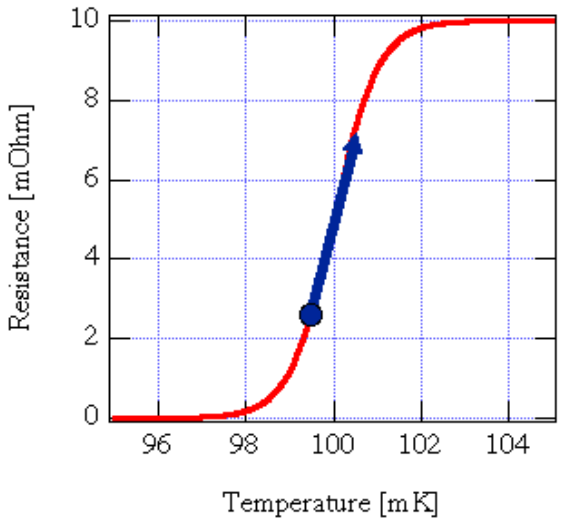


Figure 4: Resistance change of a TES with a critical temperature T_c of about 100 mK from the absorption of energy [20].

To detect the emerging change of resistance the TES is connected to a electric circuit with

a constant bias voltage. A change of resistance leads to a change in current which leads to an inductance of a B field. This is measured by a superconducting interference device (SQUID). Afterwards the chip cools down to his working temperature via a thermal bath. The resulting data appears as a pulse in a timeline [4].

2.5.3 Black body radiation

One of the expected background sources for the extrinsic data is black body radiation [4]which is a thermal radiation consisting of photons emitted from objects with a non zero temperature. This radiation has a characteristic spectrum for each temperature, which can be calculated with Planck's law [21]. Using natural Planck units it can be written as:

$$B(\nu, T) = 2\nu^3 \frac{1}{e^{\frac{\nu}{T}} - 1} \tag{6}$$

With the frequency ν and the temperature T. The peak black body radiation of each spectrum and therefore the most likely wavelength λ_{max} of a photon can be derived from Wien's approximation [21]:

$$\lambda_{max} = \frac{2.9 \times 10^{-3}\text{mK}}{T}, \tag{7}$$

where λ_{max} is the peak of the black body radiation spectrum(see fig. 5).

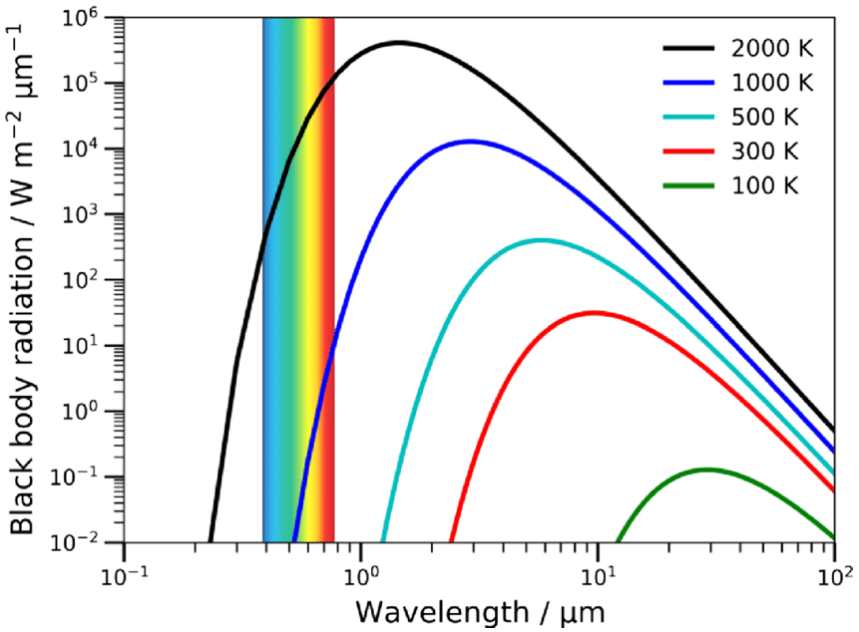


Figure 5: Black body radiation for various temperatures in the range 100-2000 K [22].

The corresponding energy of the photon can be derived from special relativity and can be calculate with:

$$E = \frac{hc}{\lambda} \tag{8}$$

With the Plank constant h and the speed of light c [23].

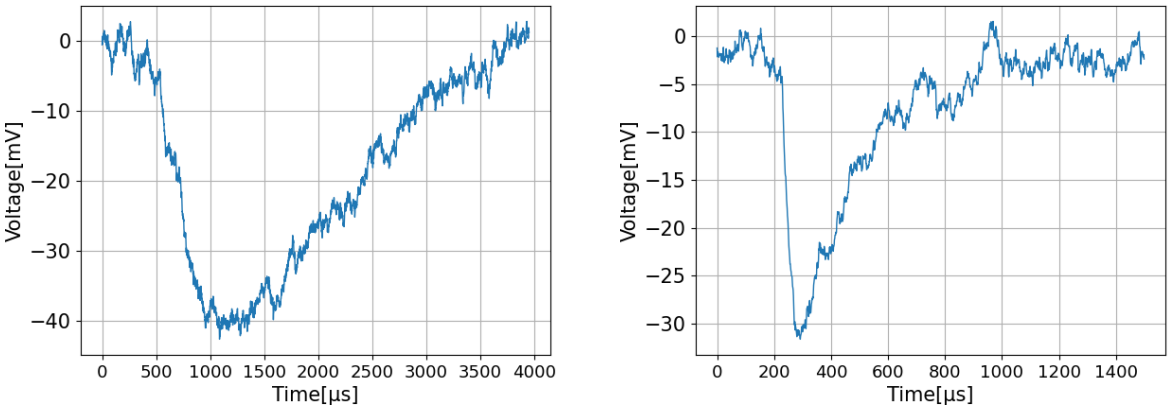
Chapter 3

Identification of background sources

The following chapter is about the analysis of the extrinsic background data from the TES. In section 3.1 the fitting procedure of the extrinsic data is covered. The next section 3.2 covers first the theoretical background of the machine learning algorithm used for the data analysis and afterwards the results of the Gaussian Mixture models are presented.

3.1 Fitting

As mentioned in the previous section, the data from the TES is a pulse in a timeline. The timeline consist of the measured voltage and the associated time. This thesis differentiate between two data set. The light data which was taken with a 1064 nm laser connected to the TES with a optical fiber and the extrinsic background where the laser is not connected with the optical fiber but the laser is turned off. In contrast, for the intrinsic background the optical fiber is not connected to the TES. This thesis focus on the extrinsic background and therefore in the rest of the thesis the extrinsic background is called background. Figure 6 shows exemplary background and light events.

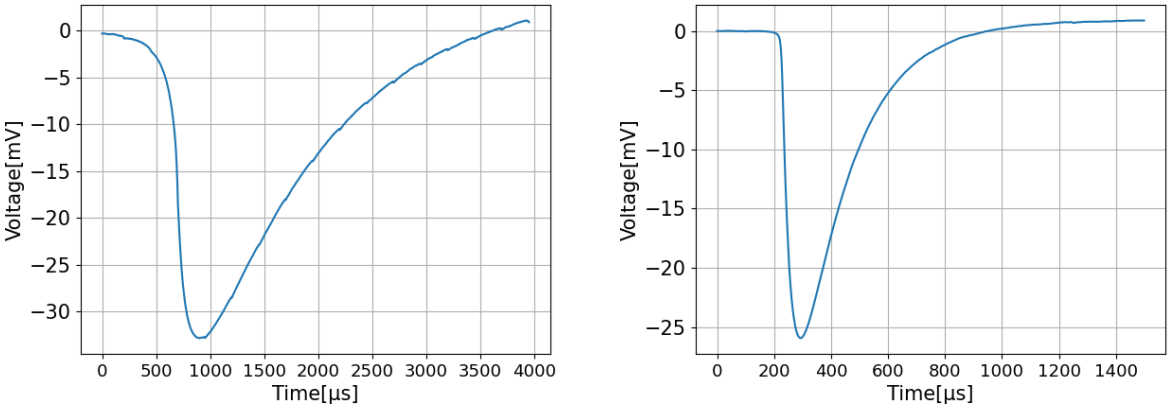


(a) Extrinsic background pulse

(b) Light pulse

Figure 6: Timelines from the TES.

The background data was taken in May 2021 and consist of 5179 timelines which were taken in around 60 hours . The 4646 timelines of the light data were taken in September 2021. For both data sets each timeline consist of 10000 data points. The mean timelines of each data set is visualized in figure 7.



(a) Average extrinsic background pulse

(b) Average light pulse

Figure 7: Average timelines from the TES.

On average the background pulses have a higher voltage and the light pulses seem to be sharper. Since the Gaussian Mixture model needs parameters for the cluster formation, the data is fitted and the fitting parameters used for the GMM. For the fitting a model function is used. This function can be written as:

$$V(t) = c - 2A\Big(exp(-\frac{t-t_0}{\tau_{rise}}) + (exp\frac{t-t_0}{\tau_{decay}})\Big)^{-1} \tag{9}$$

The variables in the equation, except the time t, are the the fitting parameters. The constant C represents the shift of the fit produced by noise. The amplitude A is the voltage value of the peak from the fitting function. t_0 is the trigger time and is a time value where the voltage drops below the trigger value of 20 mV. The rise time τ_{rise} describes the time the fit takes to reach the peak. Corresponding the decay time τ_{decay} is the time the peak needs to get to a normal level. From the resulting fitted timeline (see fig. 8) the integral is determined.

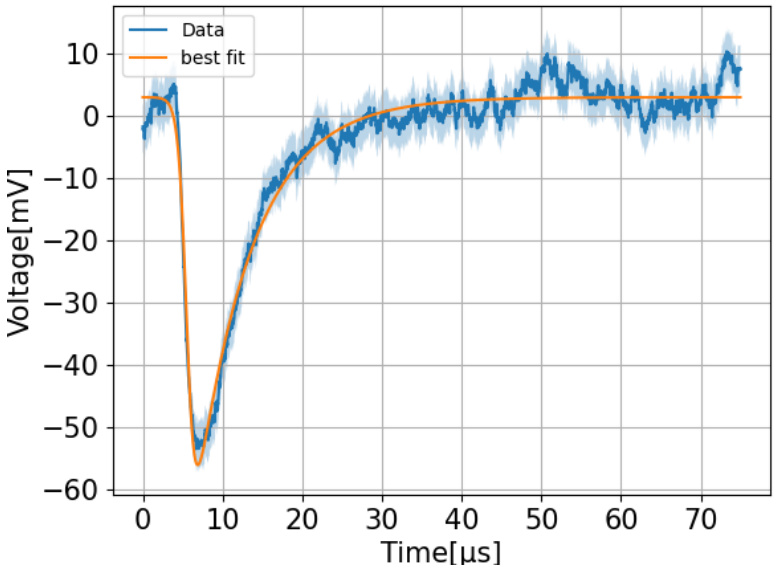


Figure 8: Fitted timeline.

To evaluate the fitting procedure, the χ^2 value of every fit is calculated with the following formula:

$$\chi^2 = \sum_{n=1}^{N} \left(\frac{y_n - f(x_n; \vec{\theta})}{\sigma_n} \right)^2. \tag{10}$$

Where f is the fitting model with $\vec{\theta}$ parameters to N data points y_n and σ_n the Gaussian error at position x_n. The Gaussian error can be calculated with:

$$\sigma_n = \frac{2}{\sqrt{\pi}} \int_0^{x_n} e^{-t^2} dt. \tag{11}$$

Due to the fact that this value depends on the number of data points, the reduced χ^2 is calculated. For this χ^2 is divided by the number of degrees of freedom.

$$\chi^2_{red} = \frac{\chi^2}{K} \tag{12}$$

For a ideal fit this value would be $\chi^2_{red} = 1$[24]. The results are depicted in figure 9.

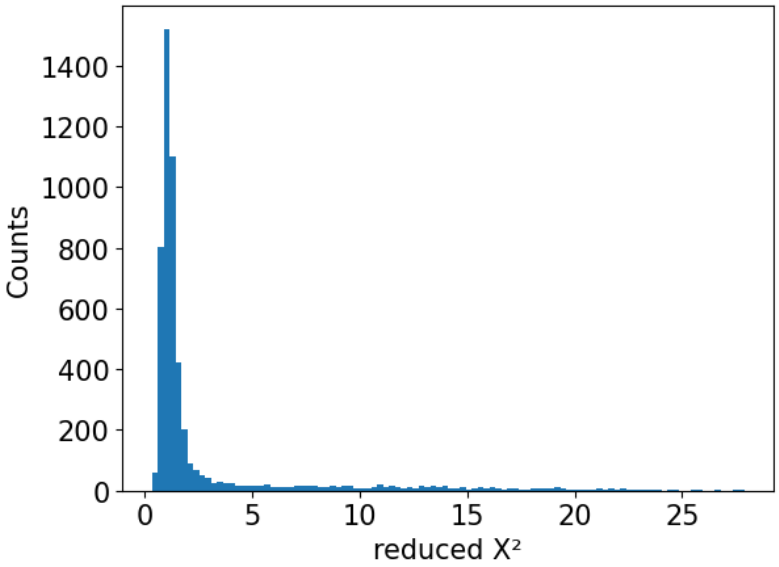


Figure 9: Reduced χ^2 of all fits.

Since the reduced χ^2 was high for some of the data, only fits with a reduced $\chi^2 \geq 4$ are used for the Gaussian Mixture in section 3.3. Thus, assuring only fits which represent their according timeline accurate are used. The modified distribution of χ^2 is shown in figure 9. Afterwards the distribution of every parameter is visualized in figure 11.

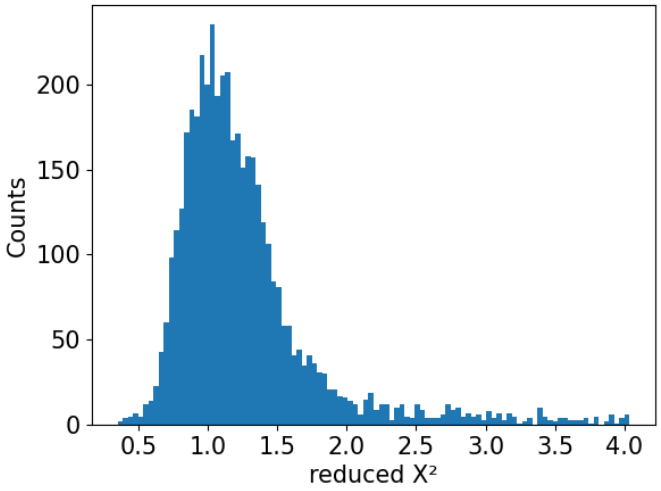
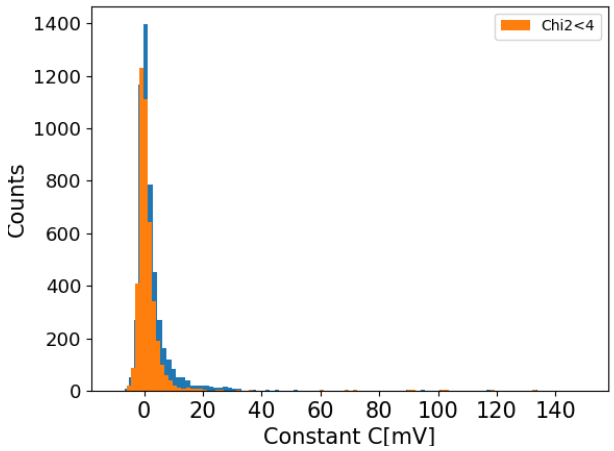
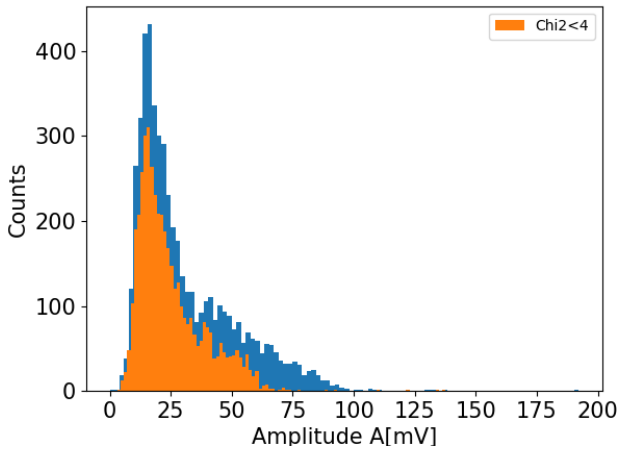


Figure 10: Cut reduced χ^2.

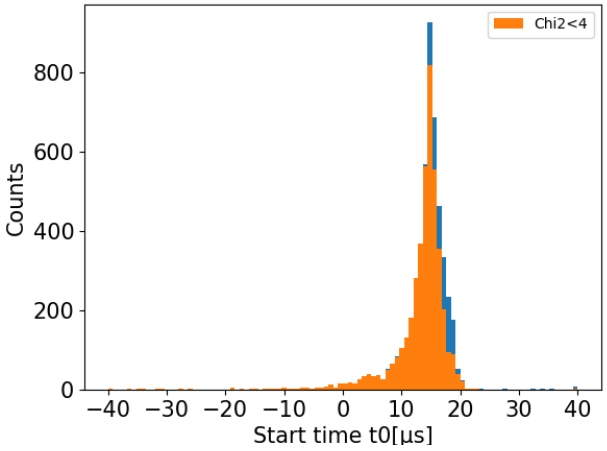
The histograms (b) and (f) of figure 11 indicate that high amplitude and high pulse integral fits often result into high reduced χ^2. This is because the fit was optimized for light events and the higher energy background events are too different from the light events. The distribution of the parameters is used in the next section to determine clusters of background events and thus different potential background sources.



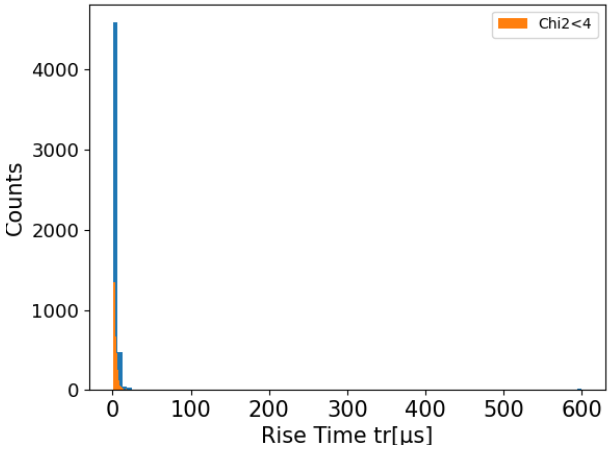
(a) Histogram of the C



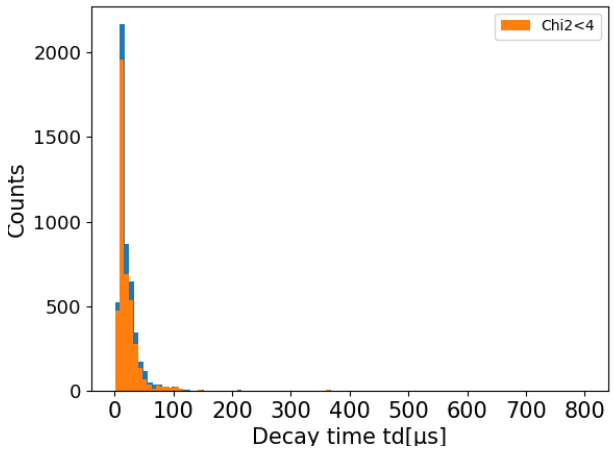
(b) Histogram of the A



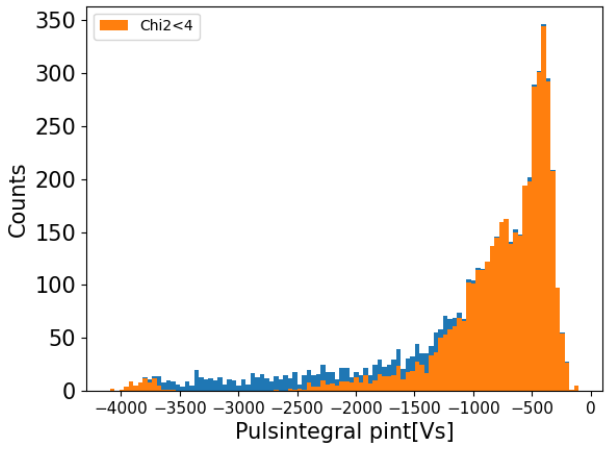
(c) Histogram of the t_0



(d) Histogram of the τ_{rise}



(e) Histogram of the τ_{decay}



(f) Histogram of the pulse integral

Figure 11: Histogram of the fitting parameters.

3.2 Gaussian Mixture Model

After fitting the extrinsic background data from the TES, the background sources are analyzed. The expected sources from the background are black body radiation from the optical fiber and electronic noises[4]. To check this hypotheses the parameters from the fitting procedure are analyzed with the Gaussian Mixture Model (GMM) to determine different potential background sources. As mention in section 3.1 only fits with a reduced $\chi^2<4$ are considered. Furthermore, since the constant C should not depend on the background source, C will not be used for the GMM. Since the sources of the background is unknown, each timeline has no known class and the number of sources is unknown as well. Therefore the algorithm learns to distinguish between different clusters of timelines without a labeled data set to learn from. This is called a unsupervised machine learning algorithm and is ideal for creating cluster of timelines for unknown background sources. Furthermore the number of cluster is a hyper-parameter which can be determined with the Akaike information criterion.

3.2.1 Theory of the Gaussian Mixture Model

A Gaussian Mixture Model is a multidimensional probability density function consisting of a weighted sum of Gaussian component densities. This sum can be written as:

$$p(x \mid \lambda) = \sum_{i=1}^{M} w_i g(x \mid \mu_i \Sigma_i) \tag{13}$$

where x is a data vector of the different measurements from the fit, w_i are the weights, Σ_i is the covariance matrix, μ_i is the mean vector and $g(x \mid \mu_i, \Sigma_i)$ are the component Gaussian densities. The covariance matrix is a matrix, where the j, k entry is calculated with [25]:

$$\Sigma_{j,k} = E(x_j \bar{x_k}) - E(x_k)E(\bar{x_k}) \tag{14}$$

where E is a operator which denotes the mean value of its argument. A D-variate Gaussian function of the form,

$$g(x \mid \mu_i \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} \mid \Sigma_i \mid^{\frac{1}{2}}} exp\Big(-\frac{1}{2}(x-\mu_i)\Sigma_i^{-1}(x-\mu_i)\Big) \tag{15}$$

describes each component Gaussian density. The sum over the weights equals 1. To determine the best estimate parameters for the GMM the Maximum Likelihood (ML) method is used. If each data vector x_t is independent, the GMM likelihood for T vectors with $X = \{x_1, .., x_T\}$ can be written as: [26]

$$p(X \mid \lambda) = \prod_{t=1}^{T} p(x_t \mid \lambda) \tag{16}$$

Due to the non-linearity of this function direct maximization is not possible. Therefore the expectation-maximization (EM) algorithm is utilized. The idea behind this algorithm is to start with a initial model λ and to determine a $\bar{\lambda}$ with $p(x \mid \bar{\lambda}) \geq p(x \mid \lambda)$. Afterwards $\bar{\lambda}$ is the new initial model and the process is repeated until $\bar{\lambda}$ and λ converge. To ensure a monotonic increase in the model's likelihood value in each iteration the following re-estimation formulas are used:[26]

Mixture Weights

$$\bar{w_i} = \frac{1}{T} \sum_{i=1}^{T} Pr(i \mid \boldsymbol{x}_t, \lambda) \tag{17}$$

Means

$$\bar{\boldsymbol{\mu}_i} = \frac{\sum_{i=1}^{T} Pr(i \mid \boldsymbol{x}_t, \lambda) \boldsymbol{x}_t}{\sum_{i=1}^{T} Pr(i \mid \boldsymbol{x}_t, \lambda)} \tag{18}$$

Variance(diagonal covariance)

$$\bar{\sigma_i}^2 = \frac{\sum_{i=1}^{T} Pr(i \mid \boldsymbol{x}_t, \lambda) x_t^2}{\sum_{i=1}^{T} Pr(i \mid \boldsymbol{x}_t, \lambda)} - \bar{\mu_i}^2 \tag{19}$$

with

$$Pr(i \mid \boldsymbol{x}_t, \lambda) = \frac{w_i g(\boldsymbol{x}_t \mid \boldsymbol{\mu}_i, \sum_i)}{\sum_{k=1}^{M} w_k g(\boldsymbol{x}_t \mid \boldsymbol{\mu}_k, \sum_k)} \tag{20}$$

where σ_i^2, x_t and μ_i are arbitrary elements of the vectors $\boldsymbol{\sigma}_i^2$, $\boldsymbol{x}_t$ and $\boldsymbol{\mu}_i$.

Depending on the constraints of covariance matrices, number of components and if parameters are shared between the Gaussian components the GMM can be optimized for the task (see fig. 12) [26].

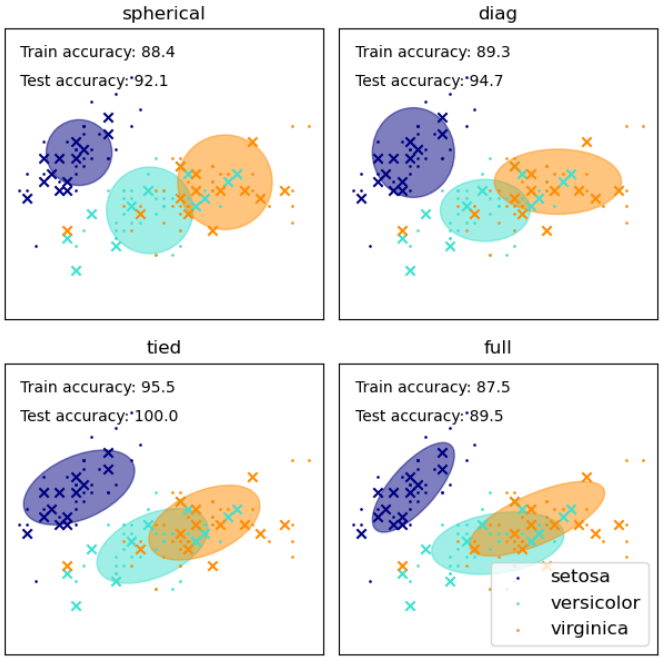


Figure 12: Gaussian Mixture model for different covariance matrices[27].

Akaike information criterion

A method to determine the optimal number of components, in our case the number of clusters, is the Akaike information criterion(AIC). For this the AIC value for different number of components is calculated via this formula[28]:

$$\mathrm{AIC} = -2 \log p(X \mid \lambda) + 2k \tag{21}$$

Where k is the number of parameters and $p(X \mid \lambda)$ the maximized likelihood. By plotting this function for different k a so called "elbow" plot is created. The model with the lowest AIC value is the "best". However, a model with a comparable low AIC value and lower number of components can be better suited for the task. Such a model is depicted as a kink in the "elbow" plot and this point is called the inflection point(see fig. 13).

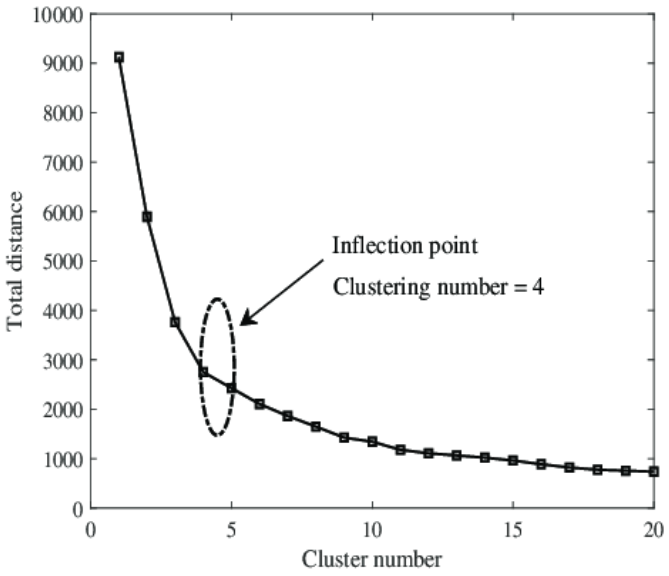


Figure 13: "Elbow" plot for a data set with a inflection point at 4 clusters[29].

3.2.2 Results of the Gaussian Mixture Model

First the appropriate amount of clusters is identified with the Akaike information criterion. As described in section 3.2.1 the AIC value for different cluster is calculated and then plotted in a "elbow" plot to decide the inflection point. This "elbow" plot is depicted in figure 14. Since the inflection is at 6 clusters, the number of components for the Gaussian Mixture Model is chosen to be 6. The resulting GMM for the background data is depicted in figure 15. Each timeline is assigned to one of the six clusters. To get a better understanding of these clusters the mean timeline from each cluster is calculated and plotted in figure 16.

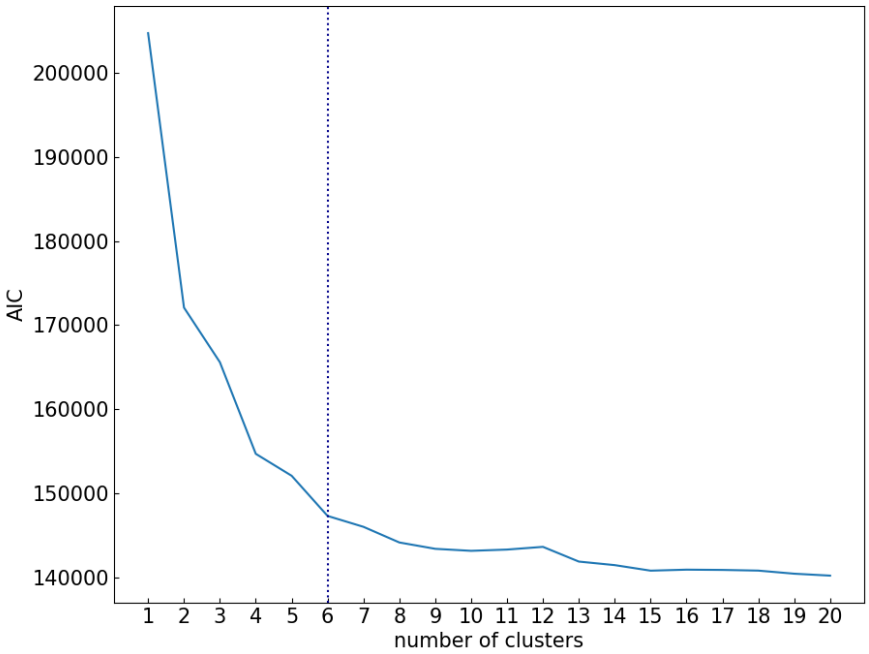


Figure 14: "Elbow" plot for the extrinsic background of the TES.

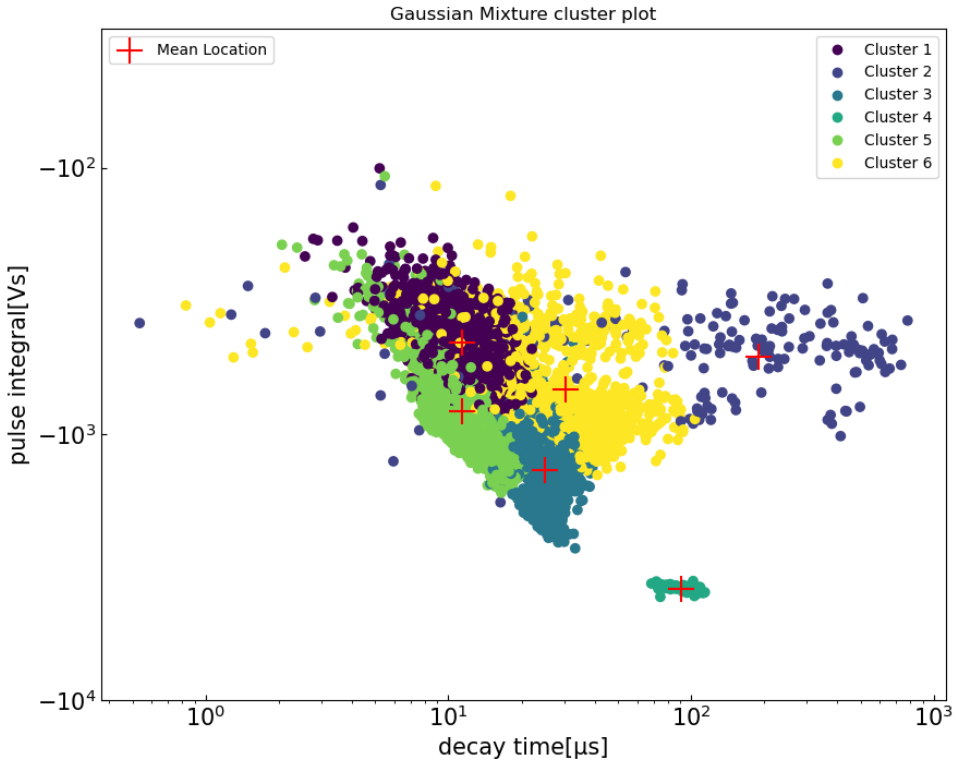
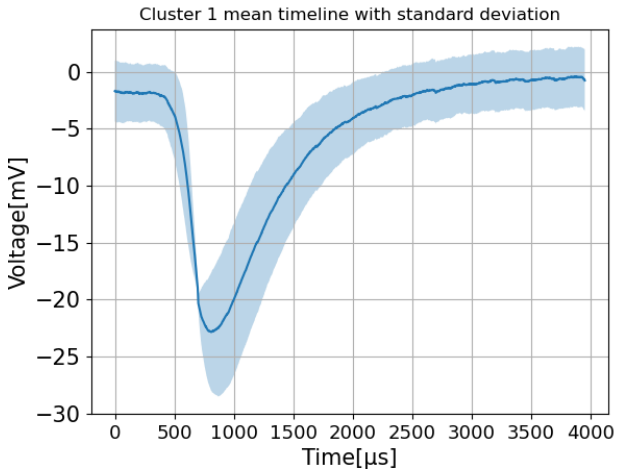
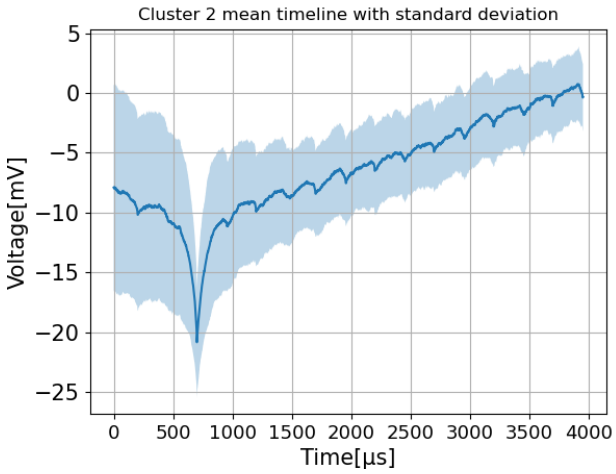


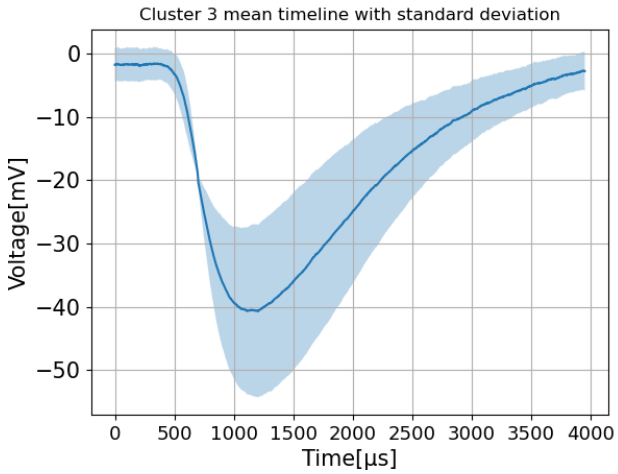
Figure 15: 2 dimensional depiction of the GMM of the extrinsic background of the TES.



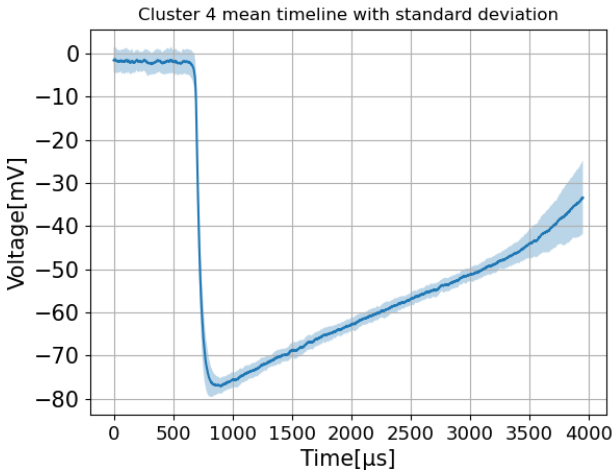
(a) Mean timeline of Cluster 1(1365 timelines)



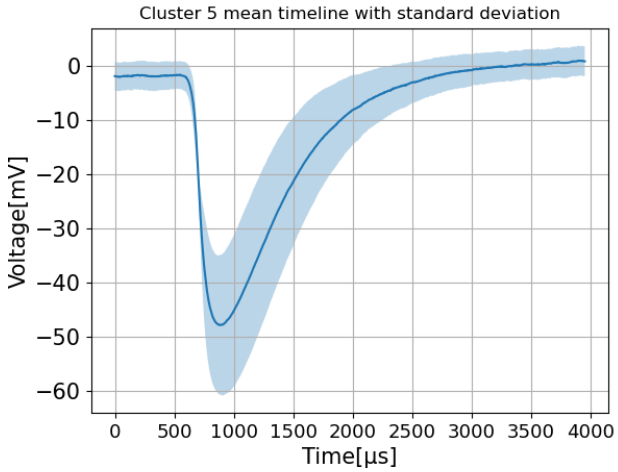
(b) Mean timeline of Cluster 2(194 timelines)



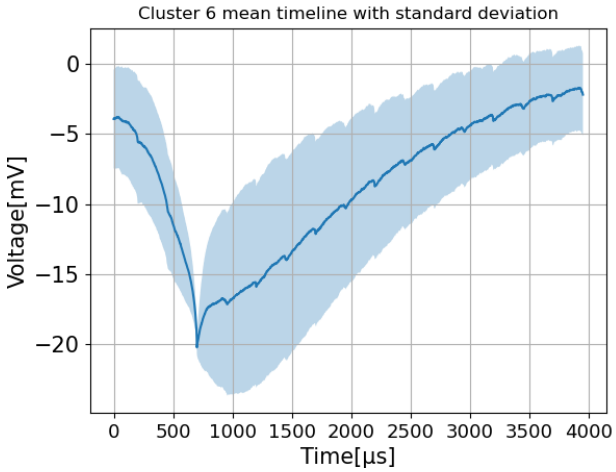
(c) Mean timeline of Cluster 3(656 timelines)



(d) Mean timeline of Cluster 4(70 timelines)



(e) Mean timeline of Cluster 5(998 timelines)



(f) Mean timeline of Cluster 6(1126 timelines)

Figure 16: Mean Timeline of each cluster.

Due to the high deviation within cluster 2 and cluster 6 and the relative low peak, they are probably generated by noise from the electronic parts of the TES. In comparison to the other clusters, cluster 3 has a very high amplitude and pulse integral. Therefore, the source must be high energetic and along with the linear decay, the source is most likely not black body radiation. Contrary, cluster 1, 3, and 5 are photon like events and hence the source is most likely black body radiation from the optical fiber. Especially compared to the average light pulse (see fig. 7b) cluster 2 consist of photons with a similar wavelength to the 1064 nm laser. Due to the twice as high peak of cluster 5 in comparison to cluster 1 and the similar timeline shape, cluster 5 consist of photons with around half the wavelength of the laser or two 1064 nm photons which hit the TES simultaneous. If two photons hit the TES not simultaneously but slightly delayed a timeline like the cluster 5 could be achieved. In total over 3000 of the around 5000 timeline have a similar shape to the light and events and could be from black body radiation. This would correspond to a rate of 50.5 black body photons per hour.

Chapter 4

Background discrimination with machine learning

In addition to a good understanding of the background sources, the possibility to discriminate the background from the light events is necessary for the success of the ALPS II experiment. Therefore, this chapter focuses on background discrimination with deep learning. In section 4.1 the theory of deep learning is introduced and thereafter in section 4.2 the classifier used for the background discrimination is explained.

4.1 Theory of convolutional neural networks

One of the most common forms of data analysis today is machine learning, and recently neural networks in particular have become popular [30]. Machine learning can be divided into supervised and unsupervised algorithms. If the data used to train the model is labeled, as in the case of background discrimination, it is supervised machine learning. Neural networks in general are modeled after the human brain and consist of different layers which are connected via general matrices. Although, the input and the output of the neural network is accessible, only the architecture of the neural network can be chosen (see fig. 17) [31].

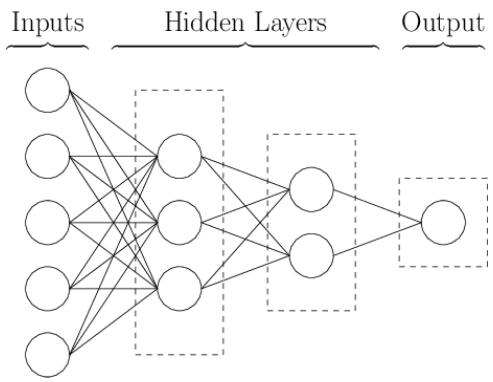


Figure 17: Sketch of a Neural network [32].

A specialized kind of neural networks are the convolutional neural networks (CNNs). CNNs are optimized for processing time-series data and image recognition. As the name suggests CNNs are neural networks that use convolutions[1] instead of general matrices [30]. Every CNN consist of a specific number of different layers, which are optimized for the different tasks, but they share some common layers, which are shown in figure 18 and explained below [31].

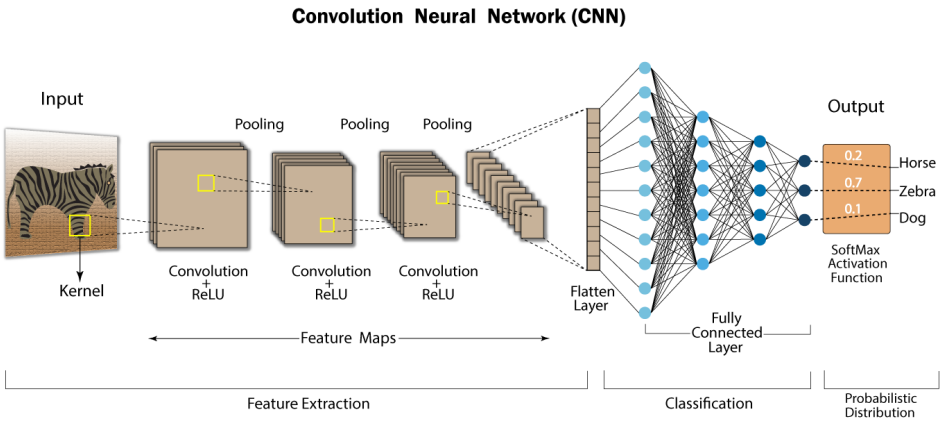


Figure 18: Layers of a CNN used for Image recognition [33].

4.1.1 Layers of CNN

The feature extraction of a CNN consist of convolutional layers, ReLU layers, pooling layers and a flattening layer. Afterwards a fully connected layer classifies the data. Each type of layer is explained below.

Convolutional Layer:

The convolutional layers are the key component of a CNN architecture. Convolutional layers convolve a part of the input data with a filter or kernel and transform it into one part of the output data as visualized in figure 19.

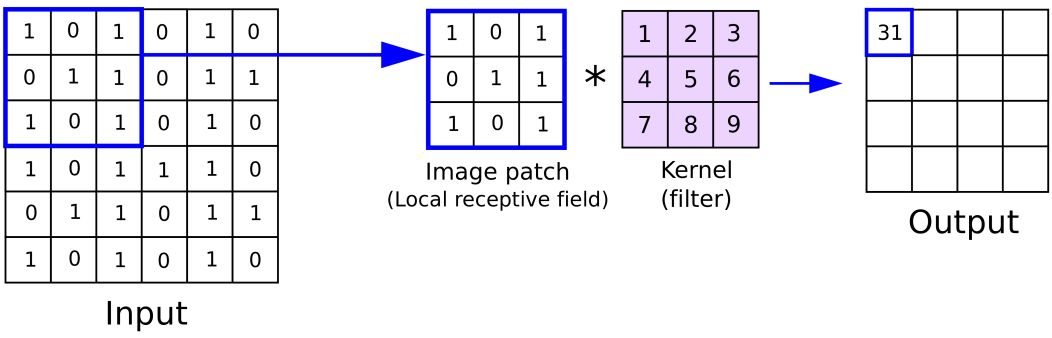


Figure 19: Convolution of an image patch [34].

[1]The convolution of two functions is defined as $(f * g)(t) = \int f(t)g(t-\tau)d\tau$

Thereafter, the kernel is shifted over the input data until the whole data is convolved. The whole output is called an activation map. Each value of the output reflects the learning of a feature [31]. Since each value of the kernel is used at every position of the input, except at the edge of the image, the network has shared parameters. Thus, rather than learning a separate set of parameters for every location, the algorithm learns only one set of parameters. Therefore CNNs are position invariant [30].

ReLU Layer:

To reduce training time and training error the network uses the Rectified Linear Unit (ReLU) activation function. The function is defined as:

$$f(x) = \max(0, x) \tag{22}$$

and is applied element wise as the output for the input x. Another advantage is that ReLU does not need normalized inputs and each input with a positive values displays that learning happened in the corresponding neuron [31].

Pooling Layer:

Pooling is used after a convolutional layer to reduce the spatial size of the activation maps. Thus, the use of pooling layers avoid overfitting. The different versions of pooling summarize the area they cover in different ways. For example, a 2×2 max pooling filter takes the largest value out of the four corresponding values [31]. This is depicted in figure 20.

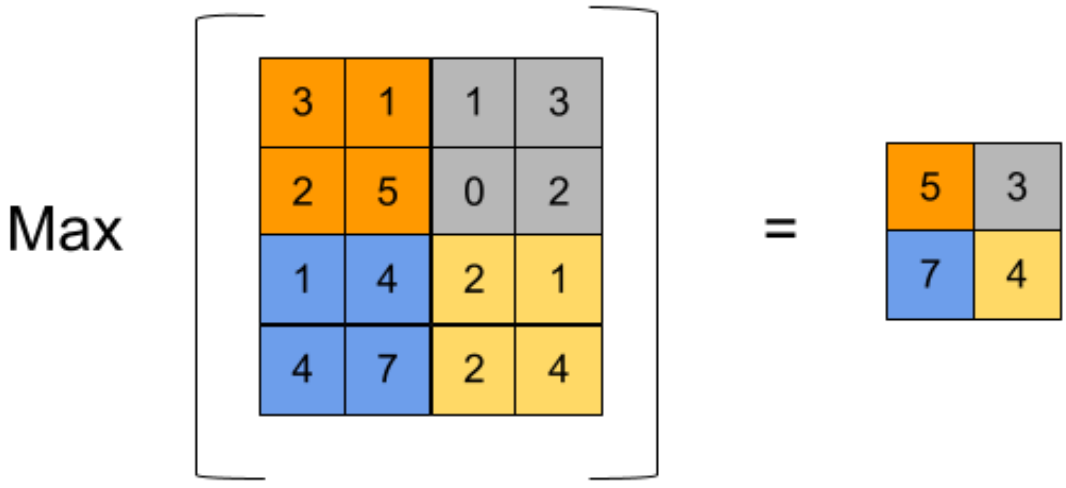


Figure 20: Max pooling of a 4*4 input [35].

Flattening Layer:

The flattening layer reduces any tensor shape into a 1 dimensional tensor, while keeping all values of the tensor. For example a tensor (10, 10, 32) will be flattened to (10 * 10 * 32). Another flattening method is the global average pooling. It applies average pooling on each spatial dimensions until every spatial dimension is one, and leaves the other dimensions unchanged. For example, a tensor (10, 10, 32) would be transferred to

(1, 1, 32). In comparison to the first method, global average pooling reduces overfitting and is more robust to spatial translations of the input [36].

Fully Connected Layer:

The fully connected layer or dense layer calculates the probabilities of the input to belong to an output class . The output for this layer is a vector which consist of N numbers, each representing the the probability of one of the N output classes [31]. For a classifier a threshold is set. If the probability is over the threshold the input belongs to the according class.

Batch Normalization:

Another layer used in the classifier is the batch normalization layer. The aim of a batch normalization layer is to reduce the change in the internal covariate shift. The internal covariate shift is a change in the distribution of network activations due to the change in network parameters during training. The batch normalization layer fix the distribution of the layer inputs as the training progresses. This reduces the internal covariate shift and thereby improving the training speed. An in-depth discussion of batch normalization is found in [37].

4.2 Results of the CNN classifier

For the classifier, the background data from the previous chapters is used as well as the light data from section 3.1. Since the light and background data consist of time series a CNN is used for the background discrimination. To optimize the accuracy of the classifier the data is prepared as proposed in [38]. First, the data is z-normalized. This can be done with the following formula:

$$f(x) = (x - \mu)/\sigma \tag{23}$$

where x is the original data point, μ the mean of the data and σ the standard deviation of the data. A z normalized time series sample has a mean equal to zero and a standard deviation equal to one. Thereby, each time series has arbitrary units for the voltage as visualized in figure 21.

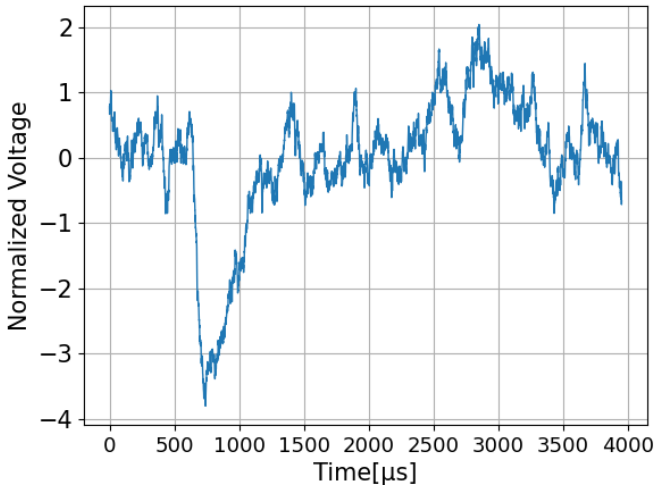


Figure 21: A z normalized time series.

Furthermore, the data is transformed into a multivariate timeseries via a reshaping. In this way, the algorithm is applicable to multivariate time series as well. Another method used to reduce overfitting of the algorithm is the stratified K-Fold cross validation. The K-Fold method splits the training data in k different splits and trains the algorithm for each fold. After every split the performance is evaluated with the according test fold and a new test fold is selected. Thereby, reducing overfitting while ensuring every part of the data is used for testing and training. A stratified K-Fold ensures that of each class, in this case background and light events, a representative part is used for the training.

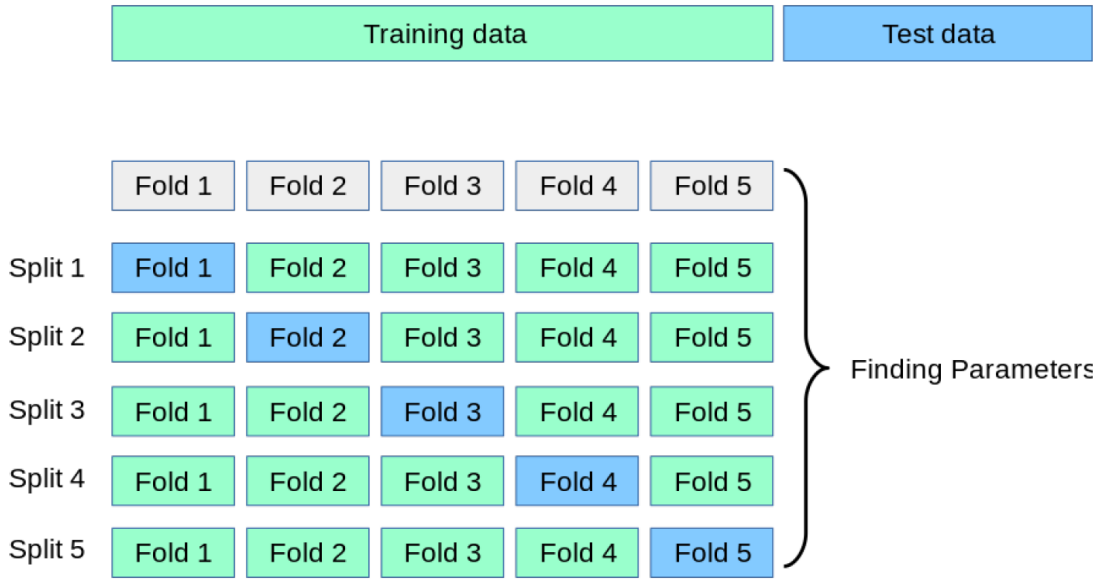


Figure 22: 5 fold cross validation.[39]

The layers of the CNN for a time series classification is proposed in [40]. The different layers are depicted in table 1. Since the pulses are shorter then the whole timeline, only 3950 data points around the amplitude are used for the algorithm.

Layer (type)	Output Shape
Input layer	(None, 3950, 1)
Convolution layer	(None, 3950, 8)
Batch normalization	(None, 3950, 8)
ReLu layer	(None, 3950, 8)
Convolution layer	(None, 3950, 8)
Batch normalization	(None, 3950, 8)
ReLu layer	(None, 3950, 8)
Global average pooling	(None, 8)
Dense layer	(None, 2)

Table 1: Architecture of the final CNN.

The changeable hyper-parameters are the kernel size, the number of filters, the batch size and the number of epochs. The batch size is the size of the training set before the internal model parameters are updated and the number of epochs is how often the algorithm works through the whole training data set. The algorithm was first trained with a kernel size of 11, 16 filters, a batch size of 50 and 100 epochs per split.
After achieving a accuracy of $p = 100\%$ the parameters were changed to reduce the number variables of the algorithm. For the first algorithm the network used 4,274 variables. The batch size was raised to 100 and the number of filters was first changed to 10, then to 8 and at the end to 4. The accuracy maintained at $p = 100\%$ for 10 and 8 filters. With 4 filters the accuracy dropped to $p = 0.9998\%$ and the number of variables was reduced to 350. To visualize the training the class activation maps (CAMs) of different time series were plotted (see fig. 23) with the following formula [41]:

$$L_{\text{CAM}}(A) = \text{ReLU}(\sum_{k=1}^{N_l} \alpha_k A_k), \tag{24}$$

with $A = f^{[l]}(x)$, where $f^{[l]}()$ is the output of the l-th layer. A_k is a k-th activation map of A and α_k is the coefficient (i.e., the importance) of A_k.

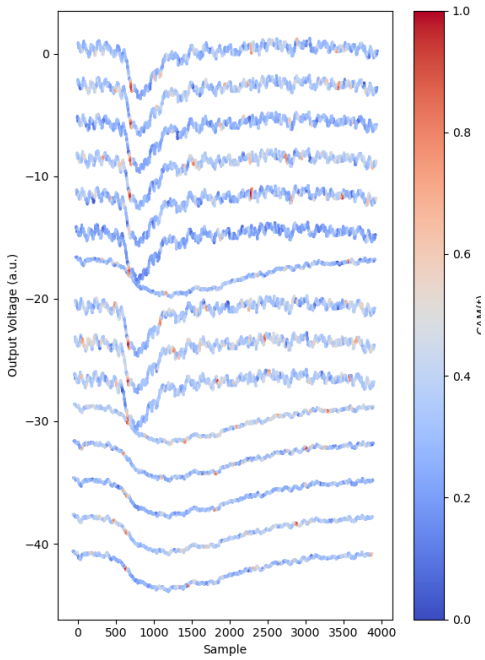


Figure 23: Class activation map of different time series. The colour visualize how important different parts of the time series are for the classification. The unit of the y axis should be microseconds.

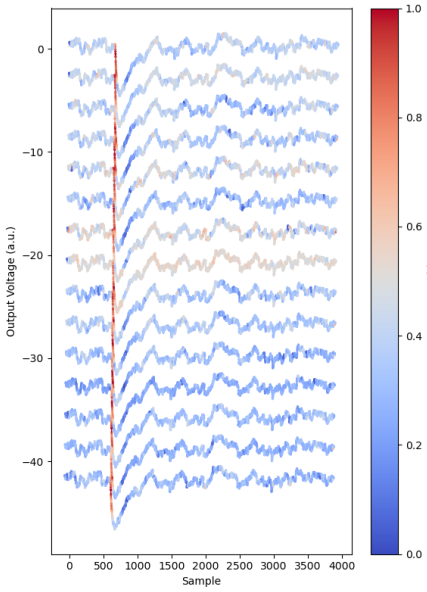
Since the peak of the time series are not the focus of the algorithm, but some arbitrary points, the algorithm decides between light and background events with the help of the noise. This indicates that the setup and therefore the time series between the two data set changed too much. Since the resulting network is not usable for new data the algorithm was retrained with new light data set which was taken in June 2021. Since now both data sets are from the same cool down the difference between the two data should be reduced. The new light data consist of 9608 time series. The settings for the new algorithm was a kernel size of 11, a batch size of 100 and 100 epochs per split. The first algorithm used 16 filters, the second 8 and the last 4. The results of the different algorithms are depicted in table 2-3 and in figure 24.

Number of filters	4	8	16
Acc. split 1	98.85%	99.22%	99,12%
Acc. split 2	99.25%	99.39%	99.36%
Acc. split 3	99.56%	99.96%	99.59%
Acc. split 4	99.35%	99.66%	99.49%
Acc. split 5	99.45%	99.63%	99.63%
Mean Acc.	99.30%	99.49%	99.45%

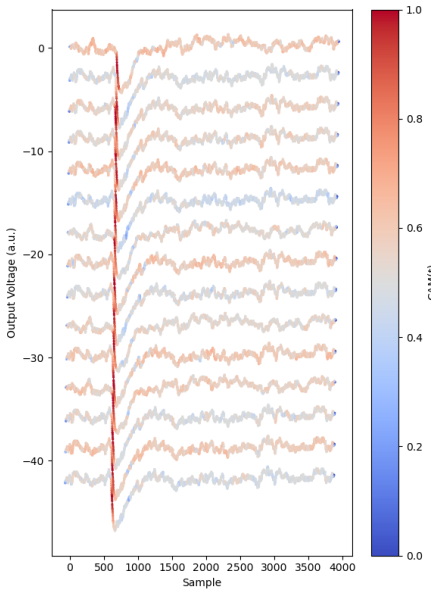
Table 2: Accuracy of the different splits.

Number of filters	4	8	16
True Light	1918	1916	1917
False Light	7	5	9
True Background	1028	1031	1027
False Background	4	5	4

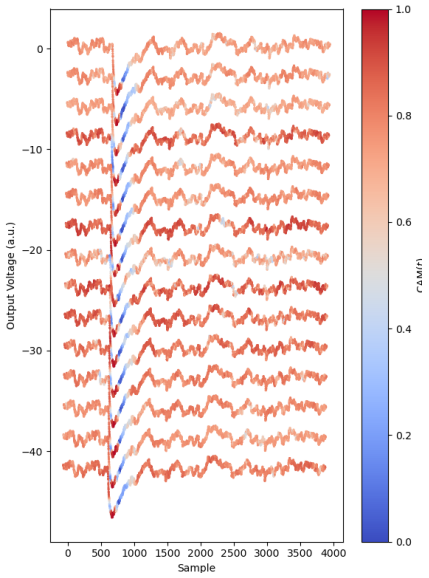
Table 3: Results of best splits of each algorithm.



(a) CAM of the 16 filter algorithm. The unit of the y axis should be microseconds.



(b) CAM of the 8 filter algorithm. The unit of the y axis should be microseconds.



(c) CAM of the 4 filter algorithm. The unit of the y axis should be microseconds.

Figure 24: CAMs of the different algorithms.

The CAMs of the new algorithms, especially the 16 and 8 filter CNN, display that the networks focus on the peaks of the time series. This demonstrate that the light and background data should be measured during the same cool down. Since the accuracy of the 8 filter network is the highest, the 8 filter network is used from now on. To further optimize the CNN, the wrongly classified time series are plotted in figure 25-27.

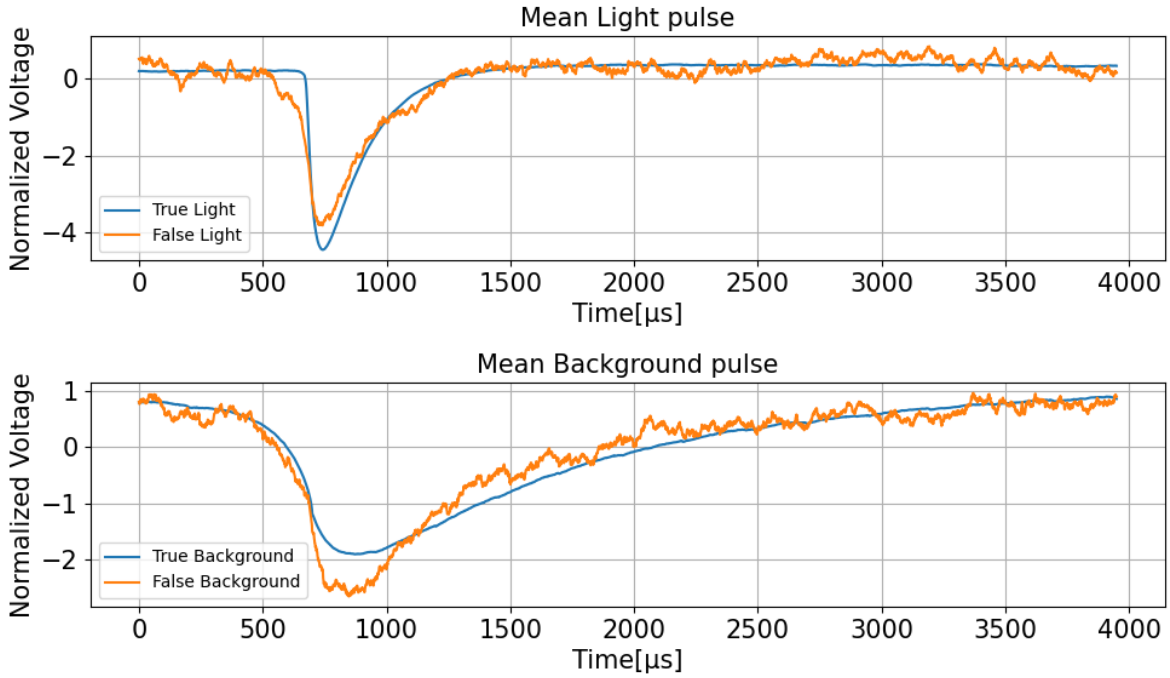
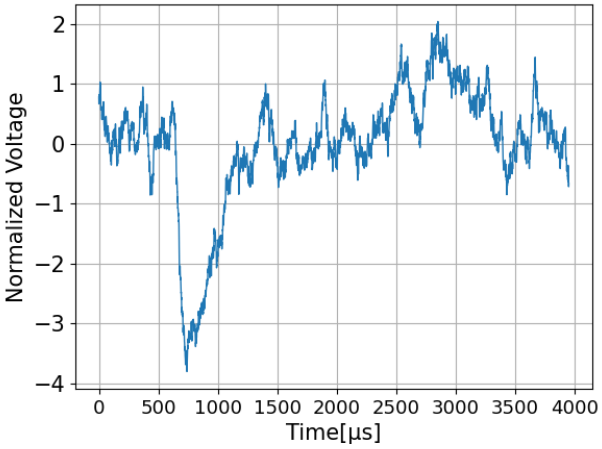
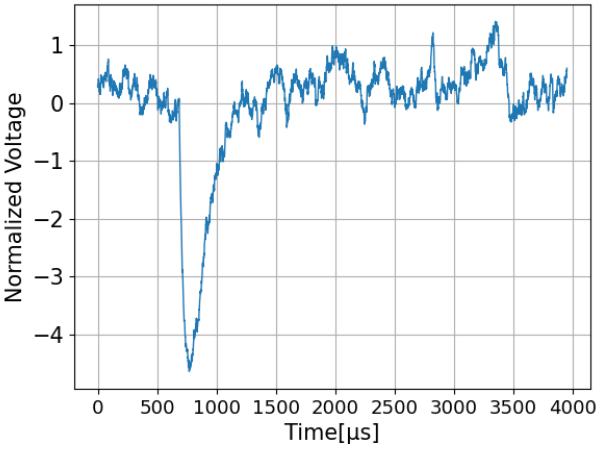


Figure 25: Mean of the misidentified timelines compared to the mean of the correctly classified data. False light is the background data which was selected as light and false background the light data which was selected as background.

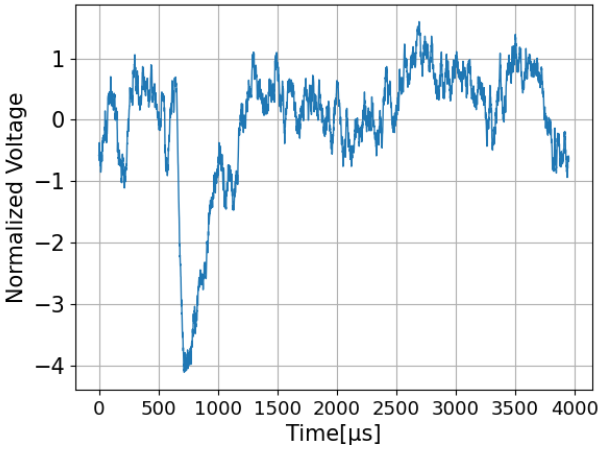
The false light looks very similar to the true light. This aligns with the Gaussian mixture results. Some timelines in the background data are most likely due to black body radiation with a similar wavelength as the light pulses and thus similar energy to the photons from the laser. The resulting timeline is very similar to the light timelines. To optimize this part of the background discrimination more data or a reduction of the black body radiation via, for example cooling of the optical fiber or laboratory, is necessary. Since the false background data is very atypical for light data it appears that either the light data contains background data or the data recording from the TES was faulty. In both cases the light data could be analyzed before the training to exclude data of this kind.



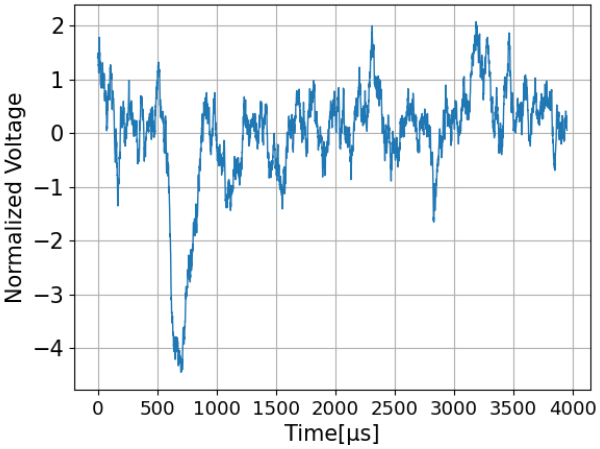
(a) False Background 1



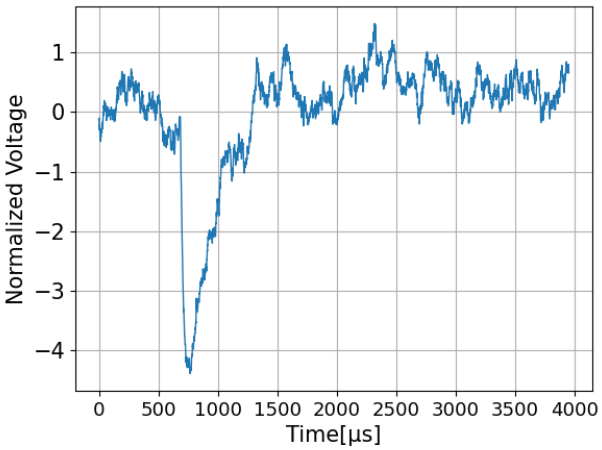
(b) False Background 2



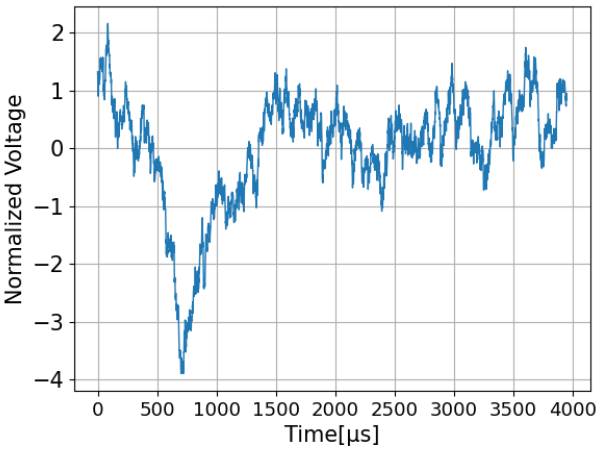
(c) False Background 3



(d) False Background 4

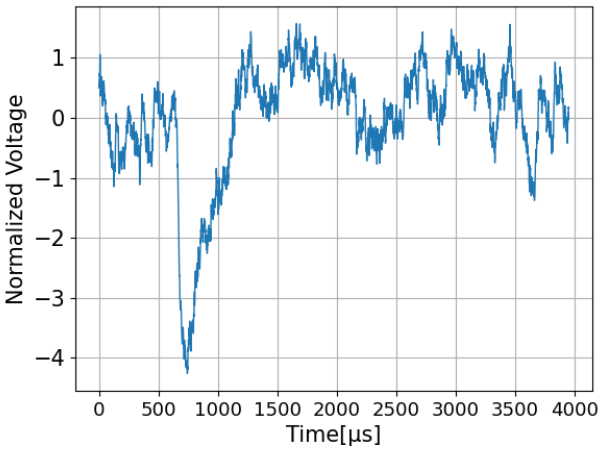


(e) False Background 5

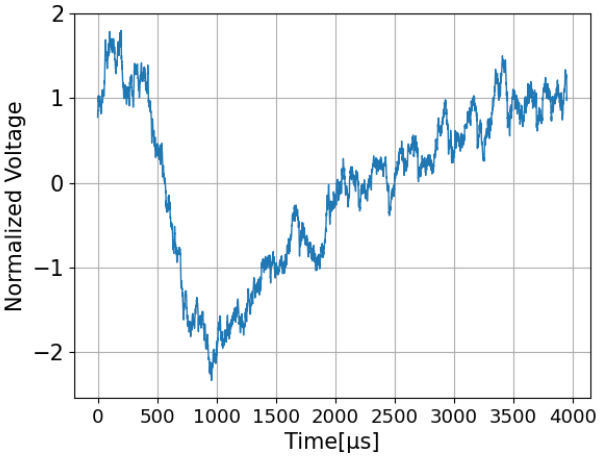


(f) False Background 6

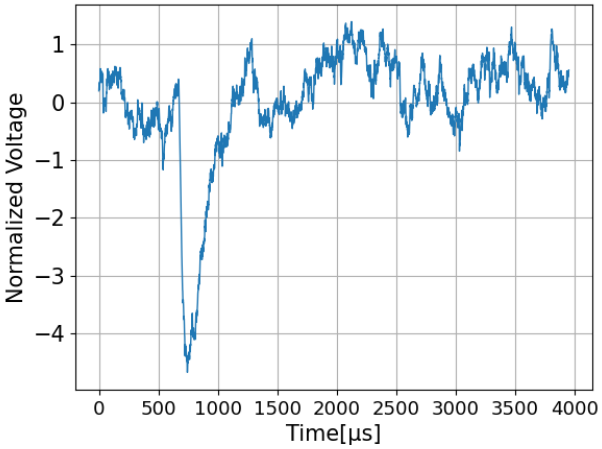
Figure 26: Wrongly classified time series of the last split of the 8 filter algorithm.



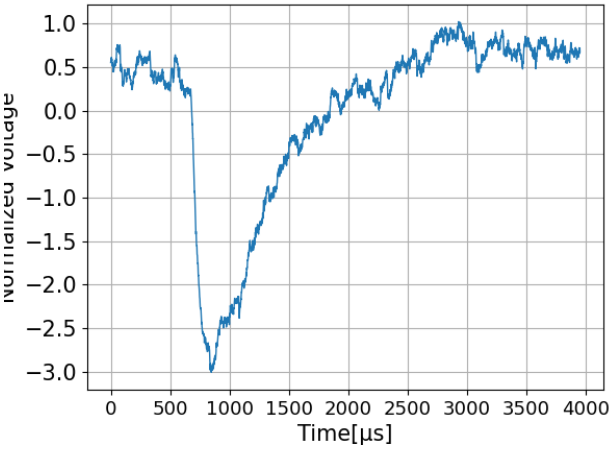
(a) False Background 7



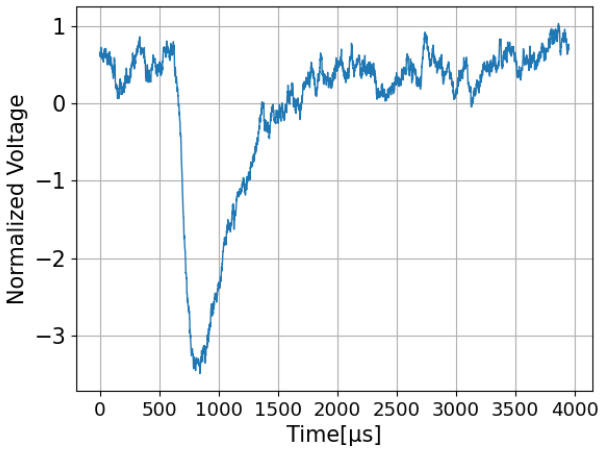
(b) False Light 1



(c) False Light 2



(d) False Light 3



(e) False Light 4

Figure 27: Wrongly classified time series of the last split of the 8 filter algorithm.

Lastly the statistical significance of the best split of the 8 filter algorithm is calculated. The statistical significance can be calculated with [42]:

$$S = 2(\sqrt{\epsilon_d \epsilon_a n_s T + n_b T} - \sqrt{n_b T}) \tag{25}$$

where ϵ_d is the detection efficiency and is assumed to be ϵ_d=0.5. The True positive rate(TPR) is ϵ_a and can be calculated from the confusion matrix.

$$\epsilon_a = \frac{\text{True Positive}}{\text{True Positive + False Negative}} = \frac{1916}{1921} = 99,74\%. \tag{26}$$

The signal rate n_s is $n_s = 2 \times 10^{-5}$.The observation time of the background data is T and $T = 215,925\text{s} = 59.979\text{h}$. Lastly n_b is the rate of false positive events. Since the test data is 20% of the whole data the number of miss-classified background events is 5/0.2=25. The rate is determined with:

$$n_b = \frac{25}{215925\text{s}} = 1,1578 \times 10^{-4}\text{Hz} \tag{27}$$

With these parameters the statistical significance is S=0.42. As the acceptance threshold for the experiment is S $\geq$ 5, the TPR and especially the rate of false positive events must be improved. To achieve this, the n_b has to be around $\sim 10^{-6}$. Another option is to change the threshold of the dense layer. A higher threshold as the default threshold of 0.5 could reduce the false positive events and thus improve the statistical significance.

Chapter 5

Summary and Outlook

In this thesis the different background sources of the TES were visualized and the analysed. The background was divided into different clusters with the Gaussian mixture model. With this, the main background source of the extrinsic background in the ALPS II experiment was identified to be likely due to black body radiation. With this improved understanding of the background sources a further reduction of the background should be possible. In particular improvements in the laboratory, for example cooling of the optical fiber, could reduce the black body radiation.

The second part of this thesis showed the potential of deep learning for data analysis and especially background discrimination. Although the accuracy of the background discrimination was high, further improvements are needed for the ALPS II experiment. Especially the false positive rate was too high to achieve the threshold of $S \geq 5$. Due to the systematic change between the different data sets, new data from the exact same setup and bigger data sets should improve the false positive rate. In addition the light data used for training seem to include some background events. By removing those, the learning could be improved further. Another option is to increase the threshold for the dense layer, which should reduce the number of false positives events further. By implementing these improvements, the statistical significance should rise and together with improvements in the laboratory the threshold of $S \geq 5$ could be achieved.

Acknowledgments

I would like to thank Dr. Manuel Meyer for the opportunity to work on a current topic like the ALPS II experiment and his help on this thesis. In addition i would like to thank Dr. Gulden Othman for her input and the laboratory visits. Lastly I would like to thank my family and friends for always supporting me.

Bibliography

[1] K. Freese, "Status of dark matter in the universe," *International Journal of Modern Physics D*, vol. 26, p. 1730012, mar 2017. https://doi.org/10.1142%2Fs0218271817300129.

[2] J. Jaeckel, G. Rybka, and L. Winslow, "Axion dark matter," 2022. https://arxiv.org/abs/2203.14923.

[3] R. Battesti, B. Beltrán, H. Davoudiasl, M. Kuster, P. Pugnat, R. Rabadán, A. Ringwald, N. Spooner, and K. Zioutas, "Axion searches in the past, at present, and in the near future," in *Lecture Notes in Physics*, pp. 199–237, Springer Berlin Heidelberg, 2008. https://doi.org/10.1007%2F978-3-540-73518-2_10.

[4] K.-S. I. and, "The any light particle search experiment at DESY," *Moscow University Physics Bulletin*, vol. 77, pp. 120–125, apr 2022. https://doi.org/10.3103%2Fs002713492202045x.

[5] T. W. B. Kibble, "The standard model of particle physics," 2014.

[6] "Standardmodell der Teilchenphysik." https://de.wikipedia.org/wiki/Standardmodell_der_Teilchenphysik. Accessed: 01.08.2022.

[7] "All-sky mollweide map of the CMB, created from Wilkinson Microwave Anisotropy Probe data." https://en.wikipedia.org/wiki/Cosmic_microwave_background. Accessed: 09.02.2023.

[8] J. Herzog-Arbeitman, M. Lisanti, and L. Necib, "The metal-poor stellar halo in RAVE-TGAS and its implications for the velocity distribution of dark matter," *Journal of Cosmology and Astroparticle Physics*, vol. 2018, pp. 052–052, apr 2018.

[9] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, "SEVEN-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: COSMOLOGICAL INTERPRETATION," *The Astrophysical Journal Supplement Series*, vol. 192, p. 18, jan 2011. https://doi.org/10.1088%2F0067-0049%2F192%2F2%2F18.

[10] L. Roszkowski, E. M. Sessolo, and S. Trojanowski, "WIMP dark matter candidates and searches—current status and future prospects," *Reports on Progress in Physics*, vol. 81, p. 066201, may 2018. https://doi.org/10.1088%2F1361-6633%2Faab913.

[11] I. G. Irastorza and J. Redondo, "New experimental approaches in the search for axion-like particles," *Progress in Particle and Nuclear Physics*, vol. 102, pp. 89–159, sep 2018.

[12] M. Dine, "TASI Lectures on The Strong CP Problem," 2000. https://arxiv.org/abs/hep-ph/0011376.

[13] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of pseudoparticles," *Phys. Rev. Lett.*, vol. 38, pp. 1440–1443, Jun 1977.

[14] L. D. Luzio, M. Giannotti, E. Nardi, and L. Visinelli, "The landscape of QCD axion models," *Physics Reports*, vol. 870, pp. 1–117, jul 2020. https://doi.org/10.1016%2Fj.physrep.2020.06.002.

[15] J. chan Hwang and H. Noh, "Axion as a cold dark matter candidate," *Physics Letters B*, vol. 680, pp. 1–3, sep 2009. https://doi.org/10.1016%2Fj.physletb.2009.08.031.

[16] "Schematic layout of the experiment. The probability for the generation of an ALP by a photon gets increased in an optical resonator. Behind the light tight wall a second optical resonator increases the probability of re-conversion to a photon in the magnetic field. (Figure from Aaron Spector)." https://epjtechniquesandinstrumentation.springeropen.com/articles/10.1140/epjti/s40485-020-00060-5. Accessed: 10.02.2023.

[17] A. Hallal, G. Messineo, M. D. Ortiz, J. Gleason, H. Hollis, D. Tanner, G. Mueller, and A. Spector, "The heterodyne sensing system for the alps ii search for sub-ev weakly interacting particles," *Physics of the Dark Universe*, vol. 35, p. 100914, 2022.

[18] "Superconductivity." https://home.cern/science/engineering/superconductivity. Accessed: 12.02.2023.

[19] N. GREENWOOD and A. EARNSHAW, "28 - copper, silver and gold," in *Chemistry of the Elements (Second Edition)*, pp. 1173–1200, Oxford: Butterworth-Heinemann, second edition ed., 1997. https://www.sciencedirect.com/science/article/pii/B9780750633659500341.

[20] "Transition Edge Sensors (TES)." http://web.mit.edu/figueroagroup/ucal/ucal_tes/. Accessed: 12.02.2023.

[21] H. Mavani and N. Singh, "A concise history of the black-body radiation problem," 2022.

[22] M. Hlavatsch and B. Mizaikoff, "Advanced mid-infrared lightsources above and beyond lasers and their analytical utility," *Analytical Sciences*, vol. 38, 07 2022.

[23] N. L. Zakamska, "Theory of special relativity," 2015.

[24] R. Andrae, T. Schulze-Hartung, and P. Melchior, "Dos and don'ts of reduced chi-squared," 2010.

[25] M. H. Gordon, M. Cerezo, L. Cincio, and P. J. Coles, "Covariance Matrix Preparation for Quantum Principal Component Analysis," *PRX Quantum*, vol. 3, p. 030334, Sept. 2022.

[26] D. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometrics*, 01 2008.

[27] "2.1. Gaussian mixture models — scikit-learn 1.2.1 documentation." `https://scikit-learn.org/stable/modules/mixture.html`. Accessed: 16.02.2023.

[28] S. Hu, "Akaike information criterion," *Center for Research in Scientific Computation*, vol. 93, p. 42, 2007.

[29] J. Zhang, L. Liu, Y. Fan, L. Zhuang, T. Zhou, and Z. Piao, "Wireless channel propagation scenarios identification: A perspective of machine learning," *IEEE Access*, vol. PP, pp. 1–1, 03 2020.

[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[31] I. Aziz, "Deep learning: an overview of convolutional neural network (cnn)," 2020.

[32] "Vessel classification through convolutional neural networks using passive sonar spectrogram images - scientific figure on researchgate.." `https://www.researchgate.net/figure/Simple-neural-network-diagram_fig1_332158639`. Accessed: 11.04.23.

[33] "Convolutional Neural Network | Deep Learning | Developers Breach." `https://developersbreach.com/convolution-neural-network-deep-learning/`. Accessed: 21.02.2023.

[34] "Anh H. Reynolds." `https://anhreynolds.com/blogs/cnn.html`. Accessed: 22.02.2023.

[35] "Explain Pooling layers: Max Pooling, Average Pooling, Global Average Pooling, and Global Max pooling. - Knowledge Transfer." `https://t1p.de/bt46w`. Accessed: 22.02.2023.

[36] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013.

[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.

[38] A. Bagnall, A. Bostrom, J. Large, and J. Lines, "The great time series classification bake off: An experimental evaluation of recently proposed algorithms. extended version," 2016.

[39] "3.1. Cross-validation: evaluating estimator performance — scikit-learn 1.2.1 documentation." `https://scikit-learn.org/stable/modules/cross_validation.html`. Accessed: 23.02.2023.

[40] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," 2016.

[41] H. Jung and Y. Oh, "Towards better explanations of class activation mapping," 2021.

[42] S. Bityukov and N. Krasnikov, "On the observability of a signal above background," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 452, no. 3, pp. 518–524, 2000.

Eidesstaatliche Erklärung

Ich versichere, dass ich die beigefügte schriftliche Bachelorarbeit selbstständig angefertigt und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem einzelnen Fall unter genauer Angabe der Quelle deutlich als Entlehnung kenntlich gemacht. Dies gilt auch für alle Informationen, die dem Internet oder anderer elektronischer Datensammlungen entnommen wurden. Ich erkläre ferner, dass die von mir angefertigte Bachelorarbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung im Rahmen meines Studiums war. Die von mir eingereichte schriftliche Fassung entspricht jener auf dem elektronischen Speichermedium.

Ich bin damit einverstanden, dass die Bachelorarbeit veröffentlicht wird.

Hamburg, 17.04.23 Malte Thoms

Ort , Datum Unterschrift

Malte Thoms