Dissertation

Towards next-generation cryogenic dark matter searches with superconducting thermometers

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

im Rahmen des Studiums

Doktoratsstudium der Technischen Wissenschaften

eingereicht von

Dipl.-Ing. Dipl.-Ing. Felix Wagner, B.Sc.

Matrikelnummer 01426449

ausgeführt am Atominstitut
der Fakultät für Physik der Technischen Universität Wien
in Zusammenarbeit mit dem Institut für Hochenergiephysik
der Österreichischen Akademie der Wissenschaften

Betreuer: Univ.-Prof. Dipl.-Phys. Dr. Jochen Schieck
Mitwirkung: Priv.-Doz. Dr. Wolfgang Waltenberger
Mitwirkung: Assist.-Prof. Dipl.-Phys. Dr. Florian Reindl

Wien, 24.10.2023

"So many things are possible just as long as you don't know they're impossible."

- from "The Phantom Tollbooth", by Norton Juster

Abstract

Today we observe overwhelming gravitational evidence for the existence of dark matter in the universe and its non-null abundance along the solar circle. Experiments using cryogenic calorimeters spearhead the effort to measure the scattering of sub-GeV/c^2 dark matter particles with nuclei of a detector target. Among them, the CRESST experiment achieves the strongest sensitivity for spin-independent, elastic scattering scenarios. Its spin-off, the COSINUS experiment, is set up to validate the long-standing dark matter discovery claim of the DAMA experiment, exploiting low energy thresholds and particle identification. These searches rely on careful detector design, measurement setup, and data analysis to provide insights into the nature of dark matter. We review the motivations for the searches, the operation principles, and study detector design choices in detail by constructing a dedicated detector response simulation. The analysis techniques are summarized and implemented in a modern software toolbox, allowing the execution of all established methods and the incorporation of machine learning classifiers. Building upon these methods, we summarize the characterization of two detector modules from the latest CRESST-III measurement campaign, which lead to the currently most stringent limits on spin-dependent sub-GeV/c^2 dark matter-nucleus elastic scattering. Furthermore, we study detector prototypes for COSINUS and estimate achievable energy thresholds of detector designs for the first physics runs. We describe the technological challenges for a large-scale measurement setup and analysis process and study methods based on deep and reinforcement learning with CRESST-III data to automate the required manual interventions. These methods equip the next generation of cryogenic dark matter searches with improved sensitivities and higher collected exposure.

Kurzfassung

Heute beobachten wir überwältigende gravitative Beweise für die Existenz dunkler Materie im Universum und an der Position der Erde. Experimente mit kryogenen Kalorimetern stehen an vorderster Front, die Streuung von dunklen Materie Teilchen im Sub-GeV/c²-Massenbereich mit den Atomkernen eines Detektortargets zu messen. Unter ihnen erreicht das CRESST-Experiment die höchste Empfindlichkeit für spinunabhängige, elastische Streuungsszenarien. Sein Ableger, das COSINUS-Experiment, soll den seit langem bestehenden Anspruch des DAMA-Experiments auf die Entdeckung dunkler Materie prüfen, indem es niedrige Energieschwellen und die Identifizierung von Teilchen nutzt. Diese Suchen benötigen sorgfältiges Detektordesign, Messaufbau und Datenanalyse. Wir gehen auf die Beweggründe für die Suche ein und untersuchen die Wahl des Detektordesigns im Detail, indem wir eine spezielle Simulation der Detektoren konstruieren. Die Analysetechniken werden zusammengefasst und in einer modernen Software-Toolbox implementiert, die die Ausführung aller gängigen Methoden und die Einbeziehung von Klassifikatoren des maschinellen Lernens ermöglicht. Aufbauend auf diesen Methoden fassen wir die Charakterisierung von zwei Detektormodulen aus der letzten CRESST-III Messkampagne zusammen, die zu den derzeit strengsten Grenzwerten für spinabhängige elastische Streuung zwischen dunkler Materie und Kernen im Sub-GeV/c²-Bereich führen. Darüber hinaus untersuchen wir Detektorprototypen für COSINUS und schätzen erreichbare Energieschwellen von Detektordesigns für die ersten Physik Runs ab. Wir beschreiben die technologischen Herausforderungen für einen groß angelegten Messaufbau und Analyseprozess und untersuchen Methoden auf der Grundlage von Deep Learning und Reinforcement Learning mit CRESST-III-Daten, um die erforderlichen manuellen Eingriffe zu automatisieren. Diese Methoden rüsten die nächste Generation der kryogenen Dunkle-Materie-Suche mit verbesserten Empfindlichkeiten und höherer Sammelquote aus.

Contents

Preamble

Before moving on to the main body of this thesis, I would like to address some organisational issues that readers may find relevant.

First, there are different styles of writing in scientific literature when presenting the work of authors. Personally, I find the use of a purely passive voice unengaging and boring for readers ("...has been/was done ..."). However, using the active voice ("...we did/I did ...") raises the question of how the author should address himself or herself. For a thesis, the singular form ("...I did ...") might be correct, but I like to say that the work done was done by "us" ("...we did ..."). This is because, let's face it, almost all work done in modern experimental science is the product of a team in some way, including single-author publications. Writing in a completely singular form can, in my opinion, sound a bit egocentric, even if it may be correct. Also, this style is explicitly discouraged by some journals/conferences in order to provide a review process as anonymous as possible. Therefore, in this thesis, as always, I use the plural ("we") to talk about work that I have done.

Second, I have been fortunate enough to have published much of my work during my Ph.D. studies. Since many of these papers have more than one author, I use the convention of referring to figures and data that were published with them, but are my work as part of this thesis, as "also used in Ref. XY". At the same time, work not done by me and used only as a basis for information is declared as "taken from Ref. XY" or simply marked with the reference in square brackets. The papers that overlap significantly with this thesis are listed below, including details of my contributions to them and their reference identifier.

I wish you an enjoyable read and hope that you can take something away from it!

—

- White Paper: Complications to comparing results from low-threshold direct dark matter searches. Paper not yet finalized, to be published in 2024. I wrote the section about challenges in the collective analysis of multiple experiments. Ref. identifier [1].

- Optimal operation of cryogenic calorimeters through deep reinforcement learning. Paper currently under internal review in the CRESST collaboration (corresponding author: F. Wagner). Based on my second master thesis, [2]. I planned and executed the project, wrote the code and the paper, and managed the publication process. Ref. identifier [3].

- Towards an automated data cleaning with deep learning in CRESST. The CRESST collaboration (corresponding author: F. Wagner), Eur. Phys. J. Plus 138, 100 (2023). I planned and executed the project, wrote the code and the paper, and managed the publication process. Ref. identifier [4].

- Cait: analysis toolkit for cryogenic particle detectors in Python. F. Wagner et al. (2022), Comput Softw Big Sci 6, 19 (2022). I planned and executed the project, wrote the code and the paper, and managed the publication process. Ref. identifier [5].

- Testing spin-dependent dark matter interactions with lithium targets in CRESST-III. The CRESST collaboration (2022, corresponding authors: A. Bertolini, S. Gupta, F. Wagner), Phys. Rev. D 106, 092008 (2022). I analyzed one of two detector modules, wrote half of the text, created the majority of figures, and managed the publication process. Ref. identifier [6].

- EXCESS workshop: Descriptions of rising low energy spectra. A. Fuss, M. Kaznacheeva, F. Reindl, F. Wagner (editors, 2022), SciPost Phys. Proc. 9, 001(2022). I was one of the main organizers of the workshop, reviewed the section editor contributions, wrote half of the common sections, and managed the publication process. Ref. identifier [7].

- Nonlinear pile-up separation with LSTM neural networks. F. Wagner, 2021. arXiv:2112.06792, contribution to the Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021). I planned and executed the project and wrote the code and the paper. Ref. identifier [8].

- A parametric fit model for Gaussian noise maxima in the presence of pollutions. F. Wagner (2020), CRESST-internal note. I developed the method based on previous work by M. Mancuso et al., and wrote the note. Ref. identifier [9].

2

Introduction

For decades, physicists have been observing the gravitational pull of an overabundance of non-luminous matter on the scales of galaxies [10], clusters of galaxies [11], and even the universe as a whole, i.e. on cosmological scales [12]. This missing source of gravity is called dark matter (DM) and can neither be explained satisfyingly with particles of the standard model (SM) of particle physics nor with other known physics. Many theories propose possible explanations: weakly interacting massive particles (WIMPs) would explain all observed phenomena, leading to the so-called "WIMP miracle", a convincing and elegant explanation that fits well to known physics [13]. A WIMP candidate is naturally included in supersymmetry (SUSY) theories in the form of the lightest supersymmetric particle (LSP). Alternatively, axions would solve both the DM problem and the problem of strong charge-parity symmetry violation. A thorough review of these traditional DM theories can be found in Ref. [14]. In much modern literature it is, however, assumed that not one DM particle, but a zoo of dark sector particles exist that interact via so-called "portals" with SM particles [15]. In addition, many more particle candidates exist, and several theories explain DM without the need for new particles, among them primordial black holes and theories of modified gravity [16, 17]. However, none of these theories has been experimentally verified so far. DM is therefore still one of the biggest open questions in modern physics and resolving its mystery will bring about a new era in our understanding of the universe.

Many experimental efforts focus on the detection of DM or its potential decay products. Experiments at particle colliders and astroparticle observatories test the production [18] and decay channels of DM [19], respectively. The latter depend on models of DM decay hotspots, regions with particularly high DM density, within or close to our own galaxy. However, a large DM abundance from the galactic halo is observed in our galaxy and along the solar circle [10], such that the particle density is expected to be large enough to measure it directly. Therefore direct detection (DD) experiments test the scattering of DM with nuclei or electrons of a target directly. Latter experiments have excluded large regions of the favored parameter

space of the original WIMP miracle, e.g. the XENON and LZ experiments using nuclear scattering in liquid targets [20, 21, 21]. As DD experiments with large liquid targets continue to increase their sensitivity down to the neutrino fog, a region in the parameter space of nuclear scattering that is expected to be dominated by a non-reducible neutrino background, another region raises interest in the community. While the original WIMP model does not allow for DM particles lighter than 2 GeV/c^2 due to a violation of the Lee-Weinberg bound [22], a growing set of theories predict such sub-GeV/c^2 (light) DM, e.g. freeze-out scenarios [23–26], asymmetric DM [27], freeze-in scenarios [28] and portal models [29]. These models are testable via the same observables as original WIMPs, nuclear or electron scattering. The favored parameter space is accessible with DD experiments using solid targets due to the achievable low energy thresholds. The leading experiments in the field are the cryogenic rare event search with superconducting thermometers (CRESST) and the cold DM search (CDMS). The CRESST-III experiment, located in the Laboratori Nazionali del Gran Sasso (LNGS) achieves the currently strongest sensitivity for spin-independent sub-GeV/c^2 DM-nucleus scattering under standard assumptions by employing scintillating calcium tungstate crystals as athermal phonon detectors [30]. Reaching the required energy thresholds for testing low-energy nuclear recoils requires the operation of a small target mass, effectively leading to a decrease in collected exposure compared to liquid targets. However, future upgrades of the experiments plan to operate a large number of detectors simultaneously and expect to acquire sensitivity down to the neutrino fog in the long term. Currently, the CRESST-III experiment simultaneously operates detector modules with a mix of target materials: calcium tungstate, silicon, sapphire, and lithium aluminate. The sensitivity of low threshold experiments using solid-state targets is limited by low energy excesses (LEEs) observed by many experiments [7], specifically in CRESST-III detectors below 200 eV [31]. Those are unmodelled and not yet fully understood instrumental backgrounds.

While almost all these experiments achieve increasingly strong null results, one DD experiment claims to measure a convincing DM signal. A certain aspect of that signal, namely its seasonal dependency, was previously considered a "smoking gun" evidence for its DM origin. In our standard assumptions, we expect the distribution of DM in our galaxy to be stationary compared to the rotation of our solar system around the galactic center and the rotation of Earth around the sun. The former rotation happens on the time scale of 225 million years and would be observed by DD experiments as a constant particle flux. The latter rotation happens on the time scale of a year and makes for a 5% seasonal change of the total expected DM flux.

A DD experiment would primarily observe the tail of the DM distribution with high velocity and therefore a modulating rate of typically 5 % or more of the total observed DM rate is expected [32–35]. These statements also hold for scenarios with both co- and counter-rotating DM halos [36]. The DAMA experiment observes such sinusoidal modulations with the expected phase shift, originally published in Ref. [37]. With an exposure of more than 2.86 tonne-years and an observation time of more than two decades, the statistical significance of the modulation has grown to an astonishing value of 13.7 σ [38]. This signal claim is however in vast disagreement with all other experiments that reach sensitivity in the same parameter space of tested models. Ref. [39] summarizes that no theory is compatible with a DM signal observed by the DAMA experiment while remaining consistent with all other published results. However, no convincing non-DM explanations for the DAMA signal exist either, even though several were proposed [40, 41]. The DAMA signal remains a topic of controversy. Responding to the lack of a suitable explanation, efforts were started to reproduce the results in an as model-independent way as possible, reusing the same target material, sodium iodide. The dependency of a DM signal from DM-nucleus scattering on the nucleus taking part in the interaction is typically modeled by considering the mass, spin, and nuclear form factors. However, it cannot be excluded that more complex systematics could play a role in the DM signal. The ANAIS, SABRE, and COSINE experiments use detection technologies identical to those used by DAMA, measuring scintillation light at room temperature [42–44]. A new experiment that is currently in construction at LNGS, the cryogenic observatory for signals seen in next-generation underground searches (COSINUS) experiment, uses the CRESST technology and measures both a phonon signal and scintillation light at cryogenic temperatures [45]. This two-channel readout allows for particle discrimination (PD) and resolves dependencies on the quenching factor (QF) of the scintillator. Furthermore, the achievable nuclear recoil energy thresholds for phonon readouts are much lower than for pure scintillator experiments, where the photon production from nuclear recoils is suppressed by a factor 3 or more [46]. The COSINUS experiment employs an adaption to the CRESST detector design, called remoTES, where the phonon sensor is not directly attached to the target but connected to it via a gold wire [47]. This design can provide a high athermal phonon collection efficiency, a crucial factor for the sensitivity of the detector, while no potentially deteriorating treatment of the crystal surface is required. These features of the COSINUS experiment allow for a high significance validation of the DAMA signal after a short run time of only one year under standard assumptions and two years in a model-independent

way [48]. Promising R&D results exist with moderate crystal sizes [49, 50]. The final detector design and the required nuclear recoil energy threshold with the required target size and number of detectors are yet to be experimentally validated.

The objective of both experiments includes overcoming three major challenges: first, detector designs need to be further optimized to achieve an as low as possible threshold with target masses of several tens of grams. Additionally, target materials need to be optimized for individual physics cases. Second, a large number of detectors need to be operated simultaneously. For this, as many processes as possible in manufacturing, operating, and analyzing the detectors must be automated to stay within reasonable bounds of human workforce. Finally, all types of backgrounds need to be identified and mitigated as far as possible. This includes all types of radiation and electromagnetically interacting astroparticles as well as their secondary products, but especially also the LEE and other instrumental backgrounds. These three tasks are not only relevant for the CRESST and COSINUS experiments, but they are in fact the major challenges for the whole community using cryogenic detectors for low threshold rare event searches.

This work contributes to several of the above-mentioned challenges. We report on the development of cryogenic artificial intelligence tools (Cait), a modern analysis framework that leverages the Python open-source scientific computing infrastructure and enables the usage of machine learning-based data selection methods while maintaining the run time requirements of high-performance computing tasks. Cait was first presented in Ref. [5]. Using this framework, we analyze the recorded raw data (RD) from two detector modules with lithium aluminate targets operated in CRESST-III and remoTES prototype detectors for COSINUS. We characterize the detectors with response and noise models and interpret the detector parameters extracted. The CRESST-III data set is used to test spin-dependent DM-nucleus elastic scattering, for which lithium-6 is a promising target material: its low nuclear mass provides relatively high recoil energy from interactions with light DM particles, and the odd spin number equips it with a high cross-section for spin-dependent interactions. The resulting limits were first presented in Ref. [6]. Using a dedicated detector response simulation, we project the achievable energy thresholds in several proposed final detector designs for COSINUS and show that they meet the performance goals of the experiment. Furthermore, we investigate the automation of certain tasks in the analysis and detector setup procedure. On a data set of historic CRESST data we train a deep learning model that can discriminate detector artifacts and pile-up from valid particle events without fine-tuning to individual detectors. We review the process of finding optimal

operation parameters for detectors and propose a method that leverages reinforcement learning (RL) to perform the task without the need for manual interventions. The latter method is tested in a virtual environment as well as in live operation on the CRESST experiment. The results of the tests with both methods were first presented in Refs. [3, 4]. Finally, we review the LEE and unexplained phenomena in cryogenic detectors using transition edge sensors (TESs) and review the first iteration of a dedicated workshop series on the topic of instrumental low-energy backgrounds. This report was originally published in Ref. [7].

The thesis is structured as follows:

In Chap. 1 we provide a brief review of our current knowledge of DM in our universe. We describe the observations made on galactic and cosmological scales and which distribution of DM in our galaxy we can infer from them. Furthermore, we summarize selected models of particle DM. Finally, the current status of the experimental search for DM via DD is reviewed, with a focus on sub-GeV/c² DM.

We derive the macroscopic detector response and noise models for cryogenic detectors with superconducting thermometers in Chap. 2 and connect them with the physics of solids. The LEE and other phenomena of poorly understood origin are discussed in detail.

The overall procedure of the data analysis and its implementation in the Cait package is described in Chap. 3. The statistical methods used in a signal search and in limit setting, and possible complications in the combination of data sets are discussed in detail.

The detector modules with lithium aluminate targets operated in CRESST-III and their spin-dependent DM results are described in Chap. 5.

RemoTES prototypes for COSINUS are analyzed in Chap. 4, and performance projections of proposed detector designs are simulated.

Finally, in Chap. 6 we treat the challenges of setup and analysis automation with deep and RL. We especially discuss two methods for automating the data cleaning and the optimization of operation parameter, respectively.

Chapter 1

Direct detection of sub-GeV/c^2 dark matter

Human understanding of our universe has progressed tremendously within the past century. Interactions and matter can be described with stunning precision through a zoo of elementary particles. The knowledge of the properties of these particles lets us not only describe nature at the smallest scales but also at its largest scales. The evolution of our universe bears witness to processes that are induced by the mere existence of types of particles. Today one could tell a clear story of the early universe that was still a hot soup, its cooling process towards the structures we observe today, and also about the scenarios for its progression in the future. But would this be already the whole story?

On the very contrary, it seems that the more we learned to describe nature with physics, the more questions accumulated that remain unanswered until today: the nature of dark energy, which leads to an observable acceleration of the expansion of our universe in the recent Mrds. years [12]; particle physics at highest energies and in the presence of strong gravity [51]; and finally the matter content of the universe, which consists to $(83.9 \pm 1.5)\%$ [12] of DM, a yet not directly measured quantity. Overall, only $\approx 5\%$ of the energy content of our universe is relatively well described through our SM of particle physics [52], while the remainder is yet to be understood.

While not all of these questions can be addressed in this work, we shall focus on the question of the nature of DM. The interaction of DM with ordinary matter is clearly observed through gravity, in rotational curves of galaxies, the cosmic microwave background (CMB), and colliding galaxy clusters. Many theories exist about the nature of DM, and the majority model DM as a new particle not contained in the SM. These theories include WIMPs, motivated by SUSY, and their generalizations [14]; axions, motivated by the strong CP violation problem, and their generalization [14];

dark photons that could be mediators to a hidden dark sector, containing a second zoo of particles [15]. There is no way to predict which of these theories is realized in our universe, and therefore a broad experimental search program is needed to test the parameter space of DM models. A shared parameter between all particle theories is the mass of the DM particle. The above statement can therefore be equivalently expressed by saying that the mass scale of the DM particle is not fixed. For a complete picture, it should also be mentioned that non-particle theories exist, including the idea that DM could consist of primordial black holes [16] and modifications of the theory of gravity itself [17].

The experimental search for DM has been ongoing for decades, and many dedicated experiments have been constructed. Searches are conducted with collider experiments, where DM production mechanisms at high energies are tested [18]. DM annihilation products are searched with telescopes and other astronomical instruments [19]. Finally, a significant non-null abundance of DM was measured along the solar circle. Specifically, a local DM density of 0.4 GeV/c^2/cm^3 [10] provides motivation to aim for directly measuring DM particle scattering with a detector target.

Since one of the well-known properties of DM is that it does not, or maximally in a suppressed way, interact with the photon field, the search for massive WIMP-like DM particles is usually conducted by assuming nuclear scattering as the primary interaction channel. For the assumption that this scattering, since it happens at relatively low energies, would follow the principles of elastic scattering, we can calculate an expected scattering rate with a target. This rate further depends on assumptions about the astrophysics properties and has the interaction cross section and the DM mass as free parameters of the theory. For such searches, no convincing positive signals were seen in the past decades, but more and more of the potential parameter space keeps getting excluded through experimental results. Recently, an interest of the community in lighter DM candidates has risen. Among them is ample motivation for sub-GeV/c^2 DM particles, e.g. freeze-out scenarios [23–26], freeze-in scenarios [28], and asymmetric DM [27].

A highly relevant factor for DD experiments is the detector technology. Technologies that are able to measure sub-GeV/c^2 DM scattering must be sensitive towards $\approx$ eV energy in their target. The currently strongest limits in this region of the DM parameter space are superconducting thermometers operated at $\approx$ mK temperatures. Besides low recoil energy thresholds, low backgrounds, and a suitable target mass are necessary for such searches. They are usually conducted in underground laboratories,

and recently low threshold experiments are starting to scale up their operated target mass.

In this chapter, we provide an introduction to the field of experimental searches for sub-GeV/c^2 DM. We start with discussing the gravitational evidence for DM in Sec. 1.1 on different length scales: on a cosmological scale, the scale of galaxies, and specifically its distribution in the Milky Way. We briefly discuss some widely known DM models in Sec. 1.2. The expected scattering rates from elastic DM-nucleus scattering are introduced in Sec. 1.3. We conclude with currently applied experimental techniques and their latest results in Sec. 1.4.

1.1 Observations of gravitational interaction

The gravitational evidence for DM presented in this section is structured by the length scale of observation, starting with a cosmological scale, moving over the galactic scale, and ending with the Milky Way. A historical order would look different and would start with the galactic scale. We will also discuss some of the important breakthroughs in history in the corresponding section. An extensive review of the historical development around the DM phenomenon can be found in Ref. [53].

1.1.1 Cosmological scale

Today the model used to describe the evolution of our universe is the Lambda cold DM (ΛCDM)model. It is based on the Friedmann equations [54] that govern the expansion of space in homogenous and isotropic models of the universe. They are a product of Einstein's equations of general relativity [55], and include the redshift, Hubble parameter, gravitational and cosmological constants, and parametrizations of the energy content and curvature of the universe. The cosmological constant can be understood as a non-zero vacuum energy that pushes the universe apart, called dark energy. Einstein's cosmological constant was historically debated since its motivation is purely empirical, and its scale seems unphysically small. However, it crucially determines the expansion phases in the evolution of the universe and is responsible for the acceleration of the expansion of our universe that we observe today. The ΛCDM model also naturally includes the existence of DM, with its mass scale and interaction strength not fixed. Furthermore, the model includes baryonic matter and radiation. When a universe expands, its energy content shifts. While the matter-energy content per volume element scales proportional to the expansion, the radiation energy content additionally decreases linearly since the radiation wavelength is also stretched by

the expansion. This also motivates the commonly used units to measure the age of a universe and quantify distances: the redshift. It quantifies the relative decrease in energy of traveling photons and measures by the time that they already existed while their wave length kept expanding. The redshift can be used to determine the distance of stars by measuring the shift in the light from spectral lines of known energy. The dark energy content per volume stays constant. While the universe was radiation-dominated in its very early moments of existence, the matter-dominated era took over rapidly. Our universe is in an intermediate era, where the matter content is still relevant, even though dark energy already represents the majority of the energy content in the universe.

The SM of cosmology foresees that our universe came into existence with very small spacial expansion and high energy density. In this state, all particles were in a hot equilibrium of production and annihilation. With the expansion of the universe and the resulting lowered energy density, the production mechanisms of high-mass particles started freezing out, followed by lower-mass particles. When the energy density dropped below the binding energy of hydrogen atoms, the average photon energy in the plasma was not sufficient anymore to cause ionization of electrons in bound state[1]. Photons started traveling freely without a relevant amount of scattering partners left, and the universe became transparent. These photons travel still today and produce an omnipresent and almost perfectly isotropic background, the CMB. The CMB was accidentally discovered in 1965 by the radio astronomers Arno Penzias and Robert Wilson. Since its characteristics are similar to signals that would be caused by instrumental backgrounds, the signal was first believed to be the result of "white dielectric material" on the measuring antenna that was produced by a close-by nesting bird family [57]. The CMB has a thermal black body spectrum at a temperature of 2.72548 ± 0.00057 K and was emitted about 380000 years after the big bang when the temperature of our universe was about 3000 K [56]. A heatmap of the CMB is depicted in Fig. 1.1.

The CMB radiation does not reach Earth perfectly isotropic but has slight anisotropies. These anisotropies are due to the non-homogeneous structure of the universe at the time of the last scattering. Gravitational wells would require the photons to have more energy to escape, causing a higher photon wavelength. The measured anisotropies therefore contain information about the size of gravitationally bound structures at the time of photon decoupling. This information can be extracted

---

[1]To be precise, this so-called time of last scattering is not at the energy density of 13.6 eV, but some time later, at an average energy density of 0.26 eV [56].
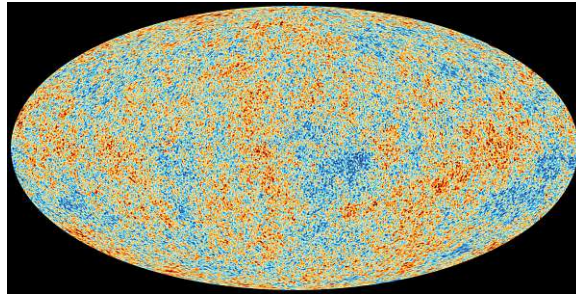
Figure 1.1: The fluctuations observed in the CMB, as observed by ESA's Planck mission. The figure covers the area visible to an observer at the position of the earth, with the azimuth on the horizontal axis and the altitude on the vertical axis. The fluctuations (red) correspond to deviations of $\approx \mu K$ from the average value of the CMB. Fig. from Ref. [58].

by fitting a theoretical model of the expected CMB cluster size to the measurement in Fourier space. The fit crucially determines the DM content of the universe: without DM, the radiation counter pressure that is felt by baryonic matter when coupled to a gravitational well would have caused an immediate wash-out of the fluctuation. Only a significant amount of matter that does not feel this radiation counter pressure enabled the buildup of stable and large structures. The latest measurements of the CMB by the European Space Agency(ESA)'s Plack mission today provide the most precise measurement of the universe's energy content [12].

Amazingly, not only the structure formation before and until the production of the CMB shows evidence for the existence of DM, but also the very first structure formation that took place in our universe shortly after the period of inflation. This period was marked by a rapid expansion of the universe, fast enough that random quantum fluctuations were prohibited from equilibrating. These random vacuum effects induced the very first density fluctuations, which later grew into galaxies, stars, and planets. Without DM, they would have washed out immediately after the inflation era ended. DM contributed already in the earliest moments of our universe to shape it into the form we observe today.

There are more observations that strongly suggest the existence of DM. One of them is the upper limit on the amount of baryonic DM in the universe from Big Bang nucleosynthesis [59], which is far below the total amount of matter in the universe. We omit a discussion of this evidence in detail in this work. A comprehensive summary can be found e.g. in Ref. [14].

1.1.2 Galactic scale

The historically first observations of DM were made by observing the movement, velocity, and brightness of stars and galaxy clusters. The obvious stars to observe are within our galactic local group, consisting of the Milky Way and the Andromeda galaxy M31 which is a distance of 765 kpc away from Earth. The Andromeda has a satellite galaxy, the Triangulum M33 galaxy, which is the third and third largest galaxy completing our local group [53].

The first to make the observation of missing mass in galaxies was William Thomson Kelvin, who compared the light of galaxies with their mass, measured with the velocity of their stars [53]. The first quantitative estimate of the DM abundance in the universe was calculated by Fritz Zwicky [60]. He estimated the mass of the Coma cluster from the number of observed galaxies and the average mass of a galaxy. With an estimate of the system's size, he further estimated its potential energy. The virial theorem predicts a relation between the virial of a dynamic system of gravitationally interacting masses and its total kinetic energy T in equilibrium:

$$T = -\frac{1}{2}\sum_{i=1}^{N} \overrightarrow{F_i} \cdot \overrightarrow{r_i} \tag{1.1}$$

Zwicky approximated the virial by using the system's potential energy and could predict the velocity dispersion. However, the observed velocity dispersion was over 10 times larger than the predicted one from visible matter. The velocity can be extracted from the Doppler shift of spectral lines in the rotating stars. Therefore he concluded that the Coma cluster should contain much more mass than is visible. He was then the first to use the phrase DM to describe the absent matter in his calculation.

These first observations were made on galaxy clusters. Later, Vera Rubin [61] studied the rotational curves of individual stars in galaxies in detail. The velocity of stars far away from the center of a galaxy is expected to fall with $1/\sqrt{r}$, with r the distance from the center of the galaxy. This statement is based on equating the centripetal force and the attracting gravitational force in the simple calculation of the gravity of a point mass:

$$\frac{mv^2}{r} = G\frac{mM}{r^2}, \tag{1.2}$$

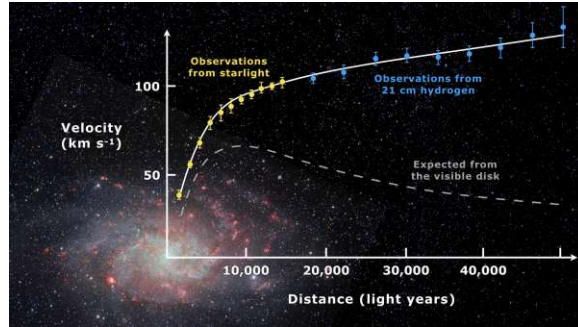$$v = \sqrt{G\frac{M(r)}{r}}. \tag{1.3}$$

Figure 1.2: The velocity distribution of visible stars depending on the distance from the galactic center of the spiral galaxy M33. The expectation from the gravitational pull of the visible matter (grey dashed) deviates strongly from the observed velocities (white continuous) that are fitted to the observations from starlight (yellow markers) and hydrogen (blue markers). Fig. from Ref. [62].

Here G is the gravitational constant, m the mass of the point mass, M the mass of the total mass of the visible orbit, which we assume to be closer to the center of the system than the point mass is, and v the velocity of the point mass. However, the experimental observation showed a velocity curve falling with a much weaker dependency on the distance from the galaxy center or staying almost flat. Fig. 1.2 depicts the situation for another galaxy. This behavior of the rotation curve could be explained by the existence of a larger halo of invisible matter that extends beyond the visible and rotating stars. M would then effectively be a quantity that grows with r, and the point mass would still be inside the growing orbit. For the assumption $M = r\rho$, we would obtain a constant velocity:

$$v = \sqrt{G\rho}. \tag{1.4}$$

Another piece of important evidence for the existence of DM was the observation of galaxy cluster 1E 0657-56, dubbed the bullet cluster [11]. An image of the observation is shown in Fig. 1.3. This galaxy cluster consists of two (sub-)clusters that collided 150 million years ago, observed from Earth. In the collision, the luminous matter interacted much stronger than the DM, leading to a shift in the centers of gravity, visible light, and radiation produced by the interstellar gas. The center of gravity is nowadays measured through gravitational lensing, where the fact that light bends towards centers of gravity is used. This leads to distortions in the image of stars and galaxies in the background, which are otherwise homogeneous distributed. This observation puts an upper limit on the self-interaction strength of DM and is, until

14
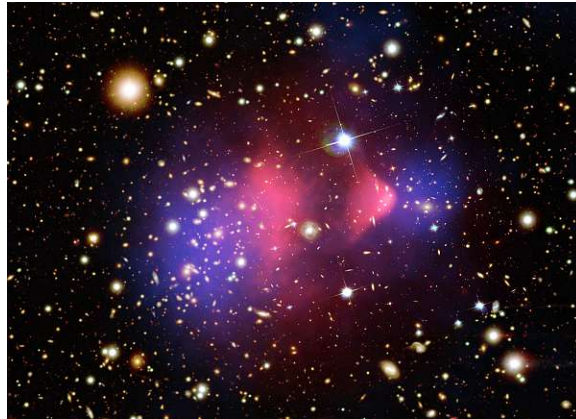
Figure 1.3: Color image of the merging cluster 1E0657-558, reconstructed from the observations of the Chandra X-ray observatory. The distribution of stars and the interstellar gas that is supposed to make up the largest share of the gravitational pull (red) is severely displaced from the mass distribution reconstructed from gravitational lensing (blue). Fig. from Ref. [63].

today, the major argument excluding modified theories of gravity from substituting a particle theory of DM.

Further evidence is contained in gravitational lensing measurements that allow us to map the location of the DM content of distant galaxies. Many more historical contributions discussed the gravitational evidence for DM, and this text solely focuses on the most prominent ones. A very relevant distribution of DM for experimental efforts is that along the solar circle, and we discuss the most recent results on this DM abundance in the following.

1.1.3 Dark matter in the Milky Way

The presence of DM along the solar circle was historically debated since galactic rotation curves were of poor data quality, leading to large uncertainties in the local DM density estimations. However, Ref. [10] provided clear evidence for the existence of DM in the inner Milky Way in a model-independent analysis. They compare the Milky Way's recently observed high-quality rotation curves with the expectation from visible matter alone. They used a data set of 2780 measurements from gas kinematics, star kinematics, and masers. Furthermore, they benefit from a lowered uncertainty in the baryonic matter distribution. By bracketing over a large number of models of baryonic matter distribution, they obtain their estimates on the uncertainties in the baryonic matter content. For all considered baryonic distribution models, they obtain the result that the observed gravity is not explicable solely with the observed baryonic matter,
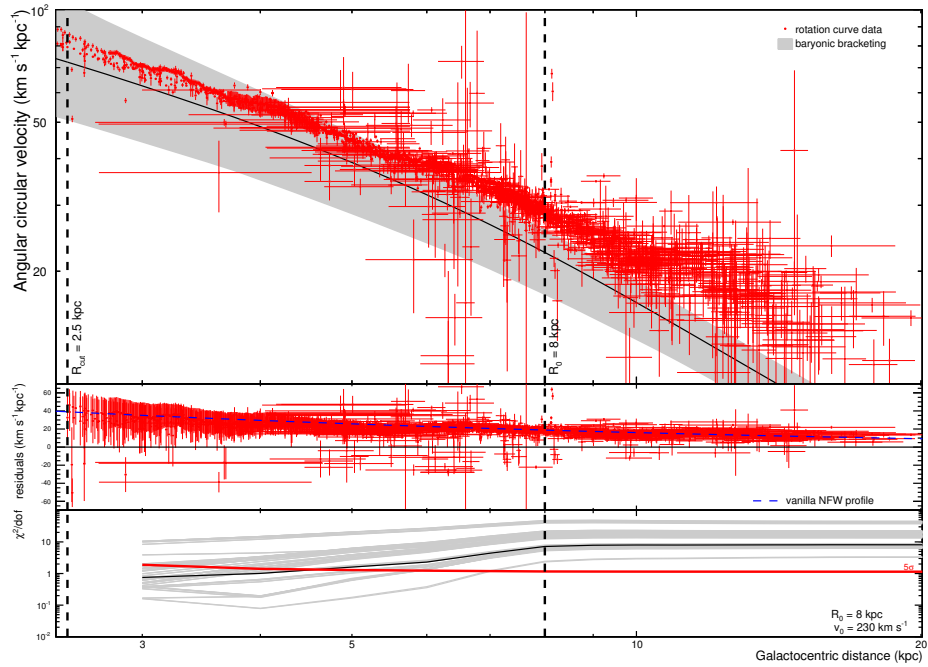
Figure 1.4: (top) Angular velocity measurements and their uncertainties (red) together with the bracketing of the contribution of all baryonic models (grey) as a function of Galactocentric radius, and contribution of the fiducial baryonic model (black). (middle) Residuals between observed and predicted angular velocities for the fiducial baryonic model are shown in the central panel. The blue dashed line shows the contribution of a NFW profile (see text for values). (bottom) The cumulative reduced χ^2 for each baryonic model as a function of Galactocentric radius. The black line shows the case of the fiducial model, and the thick red line represents the reduced χ^2 corresponding to 5 σ significance. Fig. and caption taken from Ref. [10]

with a significance level above 5 σ in a χ^2 test, for distances between 5 and 8 kpc from the galactic center, which is the assumed galactic orbit of the sun. Furthermore, they obtain an excellent match of the resulting residuals with the residuals obtained with the widely used Navarro–Frenk–White (NFW) [64] profile of the DM halo, with a local DM density of 0.4 GeV/c^2/cm^3 and a local circular velocity of $v_0 = 230$ km/s.

The above-summarized results are essential for earth-based experiments. However, such experiments have been performed already before, with less precise parameters and models. To retain compatibility with older results, the conventions that are still used today differ slightly from those most precise values, and we explain them in the following.

DM is assumed to be in an equilibrium state in the galaxy, spherically Maxwellian distributed and without significant movement in the galactic rest frame. The sun moves through the galaxy and passes through the DM distribution. The speed at

16

which an earth-bound observer would see DM passing by modulates since the Earth moves on an annual cycle around the sun. There are also daily modulations expected since an observer on the surface of the earth would rotate around the earths core, which are usually neglected since they are too small to detect experimentally today. [2] This leads to an overall expectation of the velocity distribution of DM observed from Earth, of

$$f(v)dv = \frac{1}{\mathcal{N}} \left(\frac{3}{2\pi v_{\text{rms}}^2} \right)^{\frac{3}{2}} \exp \left(-\frac{3v^2}{2v_{\text{rms}}^2} \right) \theta \left(v - v_{\text{esc}} \right) dv \tag{1.5}$$

with normalization

$$\mathcal{N} = \text{erf}(z) - \frac{2}{\sqrt{\pi}} z \exp \left(-z^2 \right) \tag{1.6}$$

and

$$z := \frac{v_{\text{esc}}}{v_{\odot}} \tag{1.7}$$

and an annual modulation of

$$v(t) = v_{\odot} + v_{\text{earth}} \, \cos(\gamma) \cos \omega \left(t - t_0 \right) \tag{1.8}$$

We followed the notation used in Ref. [66]. Conventionally the following values are used, based on Refs. [32–35]. The angle between the DM flux direction and the rotation plane of the Earth around the Sun is $\gamma = 60°$. The average solar velocity is $v_{\odot} = 220$ km/s, which is an outdated value, but still the most widely used one. The rotation velocity of earth around sun is $v_{earth} = 30$ km/s, and the galactic escape velocity $v_{\text{esc}} = 544$ km/s. The local DM density is taken to be $\rho_{\text{DM}} = 0.3 \text{GeV/cm}^3$ which is also an outdated but widely used value. The phase of the annual modulation, determined by t_0, is such that peak is at June 1. The variation of the WIMP velocity is smaller than 10 %, but one would expect to see the peak of the distribution mostly. Therefore the modulation in an experimentally observed event rate should be larger. Finally, we define $v_{\text{rms}} = \sqrt{3/2} v_{\odot}$.

We assumed in these numbers, that the DM halo does not rotate itself in the galactic rest frame. Effects of maximally co- and counter rotating DM are studied in Ref. [36] and can induce an uncertainty of a factor ≈ 2 in the DM velocity.

---

[2] However, daily modulations were used to conduct DM searches, e.g. in Ref. [65], not by the daily modulating strength of the DM velocity, but by the daily modulating effect of earth shielding.

This section does not require DM to be of particle nature for the validity of the statements. They hold of any alternative theory of DM as well. In the following section, we discuss some of the most widely probed DM models in more detail.

1.2 Selected dark matter candidate models

Many theories exist that seek to explain the DM phenomenon. Most assume DM to be a massive particle since only massive particles have the energy to inflict a significant gravitational pull in currently known physics. This being said, there are also theories that suggest DM to be made of black holes and general modifications of the laws of gravity. While some of these more famous theories have been disproven, e.g. the modified Newtonian dynamics theory by observing the Bullet Cluster and non-primordial black hole theories by observing the CMB, others persist. Especially noteworthy, the theory of primordial black holes seems to provide explanations of all observed phenomena, and experimentalists expect that it can be tested with the next generation of gravitational wave observatories. Also, theories of modified gravity remain but require that the centers of gravity decouple from the centers of mass.

Particle DM theories provide explanations of all observed phenomena by introducing a new particle that would make up the DM energy density. Since there is no observation that gives hints about the mass of a DM particle, the mass scale is not fixed, and different theories provide different candidates for DM particles. The requirements for a DM particle are that it has to be cold at the time of decoupling, which also excludes the SM neutrinos, only weakly interacting, stable on the time scale of the universe, and massive. There are two major types of particle classes that require different types of experiments: WIMP-like and axion-like particles. The WIMP-like particles are the relevant class for the work at hand, we shall therefore start by discussing them.

The WIMP model is motivated by the idea that the weak nuclear force could be the primary interaction mechanism between dark and ordinary matter. There is a particle predicted by SUSY, the LSP, that cannot decay further and would therefore be stable. There would be a production mechanism for this particle in the early universe, above a certain energy density, and also annihilation mechanisms. When the energy density drops below this energy, both mechanisms freeze out. This scenario is called a thermally produced DM particle. Different masses and cross sections with other particles would lead to different residual DM abundances in our universe today. This model gained significant attention when it was discussed that a DM particle mass on the scale of 100 GeV/c^2 and interaction via the weak nuclear force would lead

to the DM abundance observed today. The elegance of this model leads to the name "WIMP miracle". The picture for the classical WIMP has changed since then. Collider experiments have not measured SUSY particles in the previous runs of the large hadron collider (LHC), which does not mean that any SUSY theory has been excluded, but the initially favored parameter space of the most favored models is. Furthermore, the Lee-Weinberg bound [22] restricts a WIMP interacting with the weak nuclear force to masses above 3 GeV/c². The parameter space that is experimentally accessible there was extensively explored by DD experiments, as LZ and XENON [20, 21, 21]. It is not forbidden that a WIMP could hide at weaker cross sections, covered by backgrounds from neutrinos (see details in Sec. 1.3), but additional mechanisms that weaken the interaction via the weak nuclear force would be required. However, there is generally no reason to favor the weak nuclear force as an interaction mechanism. It is possible that another force mediator exists and carries the interaction between DM and SM particles or that an existing force carrier experiences a form of shielding of its interaction, e.g. by a fractional charge of the DM particles [67]. Therefore many former WIMP searches are today conducted as searches for generalized WIMPs, or thermally produced DM. Such searches extend down to energies in the sub-GeV/c² range. The parameter that is tested is usually the cross-section σ_{DM} with nucleons, where searches can be focused on a spin-dependent cross-section with protons or neutrons or independently of the spin. This will be discussed in more detail in Sec. 1.3.

For thermally produced particles with very weak interaction strengths, another effect occurs. The lower the interaction probability, the rarer the initial production and annihilation interactions of DM in the early universe would have happened, and the longer a time their thermalization would require after the freeze out of those channels. Below a certain interaction probability, this thermalization process would still be going on today. We call such a scenario a freeze-in scenario, and theories exist that predict such DM candidates [28].

The common expectation for particles in the mass regime of MeV and above is that they would inflict recoils since their De-Broglie wavelength is much smaller than their distance to each other. This is different from another DM search regime that was motivated by the problem of strong CP violation [68]. The Peccei-Quinn mechanism is a proposed solution to the question why quantum chromodynamics (QCD) does preserve charge-parity (CP) symmetry. The mechanism predicts a pseudo-Nambu-Goldstone boson at low energies, called the axion [69]. Its mass scale is also not fixed but could possibly extend from $\approx$ eV/c² downwards, without a definite lower limit. Since the axion would be produced through a spontaneous symmetry breaking it is

not a thermal DM candidate. A theoretically predicted coupling with the photon field would lead to distortions in electromagnetic (EM) waves, called the Primakov effect, which can be used to conduct searches for this axion interaction. Most axions searches focus on the meV/c^2 to μeV/c^2 mass range, testing the coupling parameter $g_{a\gamma}$. Generally, particles with such a low mass are called wave-like DM since their de Broglie wavelength would by far exceed their typical distance to each other, which can possible lead to long-distance interactions. From the literature, wave phenomena are expected to appear below DM masses of 30 eV [70].

More wave-like DM candidates exist. The dark photon is a predicted boson that could be a massive mediator between dark sector particles and SM particles. It could couple weakly to the EM field through a coupling parameter ϵ, and therefore be detectable. However, this coupling would be introduced per hand in the Lagrangian and not predicted from an underlying mechanism.

Another DM candidate exists in an intermediate-mass range of about keV/c^2. While the SM neutrino is excluded as a DM particle due to its too-small mass that would make it move at relativistic speeds at the time of decoupling and later, dubbed hot DM, its hypothetical right-handed partner, a sterile neutrino, could be more massive due to the Seesaw mechanism and a viable DM candidate [71]. Since the interaction cross section with the sterile would be very low due to its ignorance of the mediators of the weak nuclear force, it could likely only indirectly be detected as missing mass from a neutrino spectrum.

In principle, it is not forbidden that DM could only gravitationally interact with SM particles. In this case, the experimental search would be much harder, and none of the currently conducted experiments would reach the required sensitivity for a reasonable signal expectation. There is no lower limit on the non-gravitational interaction strength with baryonic matter and the self-interaction strength.

Furthermore, it is not excluded that DM has a non-trivial internal structure or could exist of more than one particle. This motivates DM models as the two mediator model [29], or strongly interacting DM [72, 73]. Also, if DM is not a Majorana particle, it should have an antiparticle, and possibly the same matter-antimatter annihilation rules would apply to DM as to SM particles, leading to the same asymmetry in relic particles observed today. Asymmetric DM [28] is a model that predicts a DM mass of ≈ 5 GeV/c^2 range, motivated by the idea that a similar number of relic particles would exist for baryonic and DM that survived the annihilation after the freeze out in the early universe, and to match the factor 5 difference in energy density the typical energy should be roughly 5 times the proton energy.

Overall we can summarize that many possible explanations for the DM phenomenon exist, but for none of them, any convincing experimentally evidence could be collected yet. The nature of DM remains a mystery and requires further intensive experimental efforts. In the following section, we will explore the experimental efforts to measure the scattering of WIMP-like DM particles via nuclear recoils in a target.

1.3 Direct detection of dark matter-nucleus interactions

Having introduced a model for the DM velocity distribution at the position of the Earth, we can derive an expected scattering rate of DM particles in a detector. We will do so under the assumption of a WIMP-like particle that would inflict only nuclear recoils, no electronic scattering, in a target. The scattering rate scales with the measurement time and target mass for a weak interaction. The product of these quantities is known as the exposure. This is different from the expected scattering phenomena of EM interactions, where a penetration depth of the target would need to be considered. For a low interaction probability, as for DM, the scattering is approximately uniformly distributed across the target. In the simplest scenario, the scattering does not depend on the spin of the target. However, the scattering could also occur spin-dependent, and we shall examine both cases in the following.

The differential rate formula for elastic spin-independent scattering is in Ref. [74] calculated as:

$$\frac{d\sigma_{SI}}{dE_R} = \frac{\sigma_0}{E_r^{\max}(v)} F^2(E_R), \tag{1.9}$$

with the definition of the point-like DM-nucleus cross-section

$$\sigma_0 = \frac{4A^2 f^2 \mu_N^2}{\pi}. \tag{1.10}$$

It is useful to define the normalized cross section to one nucleus

$$\sigma_{\text{DMN}} = \left(\frac{1 + m_\chi/m_N}{1 + m_\chi/m_p}\right)^2 \cdot \frac{\sigma_0}{A^2} \tag{1.11}$$

The maximal possible energy transfer between the incident particle of velocity v and that at rest is given by

$$E_r^{\max}(v) = \frac{2v^2 \mu_N^2}{m_N}, \tag{1.12}$$

and the reduced mass by

$$\mu_N = \frac{m_\chi m_N}{m_\chi + m_N}.$$ (1.13)

We followed again the notation used in Ref. [66]. Here M_{Target} is the target mass, m_N the mass of the nuclear scattering partner, ρ_χ the local DM density, m_χ the DM mass, v the incident DM velocity and $\frac{d\sigma(v)}{dE_R}$ the DM-nucleus cross-section that itself depends on the velocity. f the coupling strength. The minimal required velocity to inflict a recoil is given by

$$v_{\text{min}} = \sqrt{(E_R m_N)/(2\mu_N^2)}.$$ (1.14)

The spin-dependent interaction cross section can be written as [6]

$$\frac{dR}{dE_R} = \frac{2\rho_0}{m_\chi}\sigma_{p/n}^{SD}\sum_{i,T} f_{i,T}\left(\frac{J_{i,T}+1}{3J_{i,T}}\right)\left(\frac{\left\langle S_{p/n,i,T}\right\rangle^2}{\mu_{p/n}^2}\right)\eta\left(v_{\text{min}}\right).$$ (1.15)

The parameter $f_{i,T}$ is the fraction of each nucleus in the target scaled by its mass:

$$f_{i,T} = \frac{n_T \zeta^i m_T^i}{\sum_{i,T'} n_{T'} \zeta^i m_{T'}^i}.$$ (1.16)

Here we follow the notation used in Ref. [6]. Since we are mostly interested in low energy scattering, we put the form factor to unity, it would be required for higher energy scattering to include it as proposed e.g. in Ref. [75]. n_T is the multiplicity of the nucleus T, ξ^i is the natural abundance of the isotope i, and m_T^i is its mass. $J_{i,T}$ is the nuclear ground state angular momentum of the isotope i of nucleus T; $\left\langle S_{p/n,i,T}\right\rangle$ is the expected value of the proton/neutron spins in the target isotope i of nucleus T and $\mu_{p/n}^2$ the nucleon-DM reduced mass, and $\eta(v_{min})$ is the mean inverse velocity in the SM halo where v_{min} is the minimum velocity required to produce a nuclear recoil of energy ER.

The spectra for different target materials and DM masses are calculated and exemplary plotted in Fig. 1.5, normalized to the reference cross-section for spin-independent scattering. What is clearly visible is the inverse proportionality of the scattering rate with the DM mass and the proportionally required energy threshold. The same holds for the nuclear mass: higher mass leads to a higher rate but also to a lower required energy threshold. Therefore, lighter target nuclei will generally be beneficial for light DM searches since the required energy threshold for detecting DM particles of a fixed mass can be higher. However, the overall expected rate will also
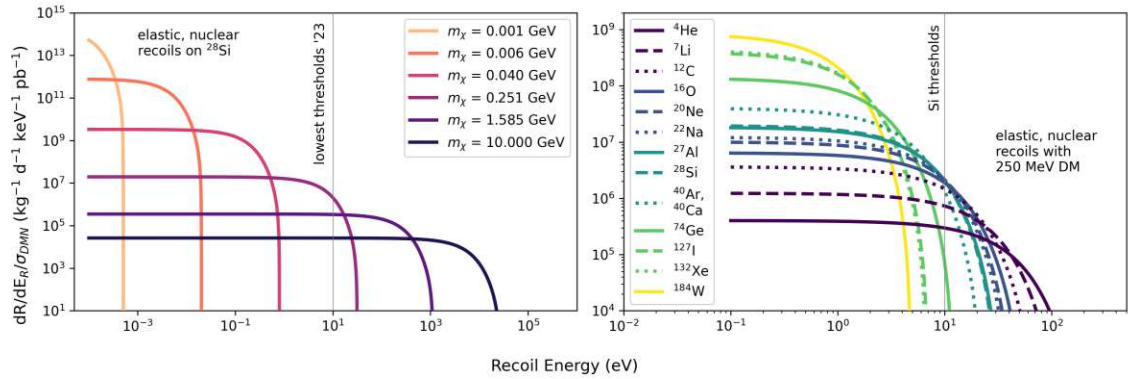
Figure 1.5: Expected DM scattering rates under standard elastic DM nucleus scattering assumptions. (left) Different scenarios for DM masses (purple-red colored lines), (right) scenarios for different target materials (blue-yellow colored lines).

be lower, which is why heavier materials can more efficiently test heavier DM masses at a fixed amount of collected exposure.

The scattering rate does furthermore depend on seasonal effects due to the movement of the Earth around the sun, as discussed in Sec. 1.1.3. This effect is fully represented in the velocity of the DM particles.

This model includes only the very basic assumptions about DM scattering. It was recently shown that the expected spectra can be significantly changed by additionally considering inelastic effects such as Bremsstrahlung or the Migdal effect [76,77]. These do not intrinsically change the assumptions about DM, only about the scattering process. In principle, comparing results with different assumptions about the scattering process is valid. However, it is common in the community to avoid this comparison since it could lead to misinterpretations about the potential performance reach of the individual experiments. Very light DM is only accessible by assuming an additional DM electron scattering channel [78]. This assumption does change the DM model since the classical WIMP-like DM would not have a relevant cross-section with electrons. Such results can not be compared directly.

There are different experimental techniques to measure particle scattering in a detector. We classify them by the detected signal quanta: photons, phonons, and electron-hole pairs. The channels are visualized in Fig. 1.6. Ionization detectors are probably the simplest to realize. A particle recoil inside a target will cause a certain amount of ionized atoms. Simply putting the target under an electrical potential difference leads the electrons and holes to drift into opposite directions, inducing a measurable current spike. The amount of produced electron-hole pairs depends on the material's technological realization and band gap. A lower band gap generally leads
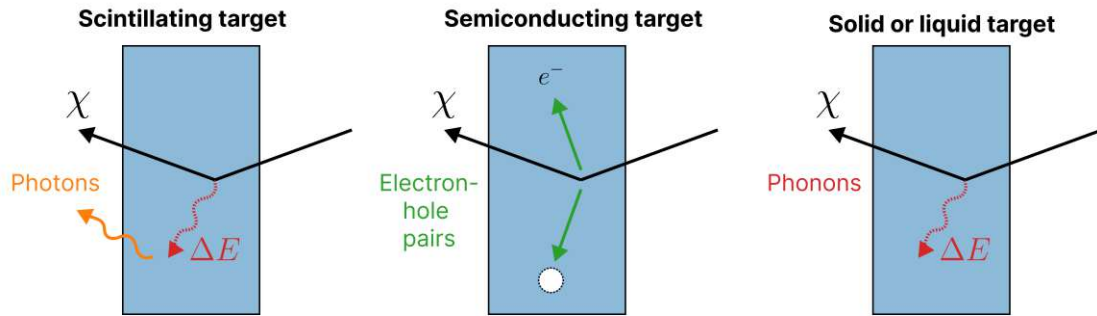
Figure 1.6: Schematic visualization of the three experimental detection channels of particle scattering. (left) A scintillating target produces photons in a particle interaction that can be measured with a nearby detector. (center) Electron-hole pairs are created in a particle interaction in a target. Especially sensitive are targets with low band gaps as semiconductors. (right) Particle interactions in a target produce excitations in the target structure, a lattice for crystalline solids. These excitations can be measured with dedicated phonon detectors.

to a higher number of signal quanta and a finer level of signal quantization, which is beneficial for low-threshold detectors. Charge coupled devices (CCDs), e.g. used in Ref. [79], can be used to collect charge over a long exposure time and read it out at a later point, leading to good spatial resolution of the interaction points. The spatial resolution can be used for event discrimination since different particles leave different traces in the detector. However, such devices usually have no useful temporal resolution. A significant advancement was the Skipper CCD [80] technology that has the advantage that the charge can be read out multiple times, lowering the readout noise significantly. These devices achieve energy resolutions significantly below their single electron resolution, they therefore reached their natural quantization limit. Modern low-threshold ionization detectors use a more sophisticated detection regime by combining it with phonon detectors [81, 82].

Phonon detectors are detectors that measure the amount of produced thermal and athermal phonons in a particle interaction. They are often thermistors, e.g. TES and neutron transmutation doped (NTD) sensors, measuring the temperature increase in a small-scale sensor, but other types of readout exist that detect magnetization (see also Sec. 2.2.4) or changes in resonator frequencies, e.g. kinetic inductance detectors (KIDs), see Ref. [83]. While thermal phonons follow the expected energy distribution of the observed temperature increase in the material from the deposited energy in the particle recoils, athermal phonons are out-of-equilibrium short-lived excitations that can have energies far above the thermal distribution. While the early phonons detectors were

mostly optimized for thermal signals, real thermometers, modern phonons detectors are optimized for the efficient collection of athermal phonons and can reach with that much better sensitivity. For this technology, the deposited energy needs only to increase the temperature of the sensor, not that of the whole target, which typically has a larger mass and heat capacity. Differently from the ionization channel, the energy measured in the phonon channel is in good approximation independent of the type of particle interaction and provides a total energy scale.

It is possible to manufacture a two-channel readout, combining ionization and phonon signal. This can be done by applying a moderate or no voltage via the ionization readout channel and simultaneously reading out the phonon channel. The significant advantage of two-channel readouts is that they enable PD between charged and neutral particles and leptons flavors, baryons, and photons since they typically induce different electron-hole pairs for the same total recoil energy. We discuss the analysis techniques for PD in more detail in Sec. 3.2. A pure ionization detector needs to be calibrated individually for different recoil types.

A modern approach to improve sensitivity is to apply a strong voltage to the ionization channel and exploit the Neganov-Trofimov Luke (NTL) effect. A particle interaction induces electron-hole pairs, which get accelerated by the strong voltage, producing more phonons than in the initial recoil. This can effectively lead to an energy resolution below the single electron limit [84], while retaining temporal resolution of $\approx$ milliseconds and lower. However, in this scenario, the channels cannot be used independently for event discrimination.

Finally, scintillation detectors measure the produced scintillation light in a target from particle scattering. The light is collected in an external, spatially separated detector with a direct line of sight to the target, which is itself an ionization or phonon detector. While this technology sounds as if it was the most complex one, it is likely the easiest to realize, since strong scintillators and sensitive light detectors, as e.g. photomultiplier tubes (PMTs), are also industrially produced and can be bought with sufficient sensitivity. Also, scintillation channels can be combined with the phonon and ionization channels for event discrimination. For the combination of scintillation and phonons readouts, scintillating crystals are used, where a phonon sensor sits directly on the crystal, and a light detector collects the scintillation light close by [30].

Nuclear recoils typically inflict a much weaker signal than electron recoils for ionization and scintillation channels. This effect can additionally depend on the recoil

energy. The function quantifying this effect is called the QF. Dedicated measurements of target QFs are routinely conducted in the preparation of experiments.

An innovative approach for light and phonon read-out is based on superfluid helium-4. Quasiparticles that scatter with the liquid surface cause evaporation of helium atoms, which can then be collected with sensors separated through a vacuum gap [85]. The scintillation light can be collected with remote sensors as well and potentially with submerged sensors. This technology has great potential since helium as a light target would cause high expected DM recoil energies, and the adsorption of helium atoms on suitable sensor surfaces can induce an additional boost in the measured energy.

The combination of scintillation and ionization channels is often used in time projection chambers (TPCs) with liquid targets [86]. An incident particle causes a trace of signal quanta that travels to sensors on the edges of the containing vessel. The prompt scintillation photons and the electron-hole pairs have different drift times, which leads to different time delays and amounts of measured signal quanta in the channels. This effect can be used for PD and for localization of the recoil sight.

The presented signal readout channels are used by various experiments to test DM models. The following section summarizes some of their most recent results.

1.4 Latest experimental results

Experiments are using a variety of technologies to test DM. Results are quantified for elastic DM nuclear scattering in the parameter space of the reference cross-section and DM mass. This is depicted in Fig. 1.7 for several relevant recent results in the classical WIMP mass region and its lighter generalizations. The strongest results in the GeV/c^2 mass region are obtained by the XENONnT and LZ experiments [86, 93], and the DarkSide experiment [92] by using large TPCs filled with xenon or argon. Those have the significant advantage that a liquid-containing vessel can be easily scaled up to high masses, to the order of tons of target mass, and therefore high exposure can be collected. Furthermore, the heavy targets increase the expected DM scattering rate for high DM masses. Their results already reach the background from elastic nuclear scattering of solar and other neutrinos. Testing this parameter space is significantly harder since the omnipresent background can hardly be discriminated from DM recoils. Directional approaches could, in principle, discriminate, but with current technology, those are not realizable for a high mass and high exposure regime [97]. The HeRALD experiment, which is a part of the larger TESSERACT collaboration,
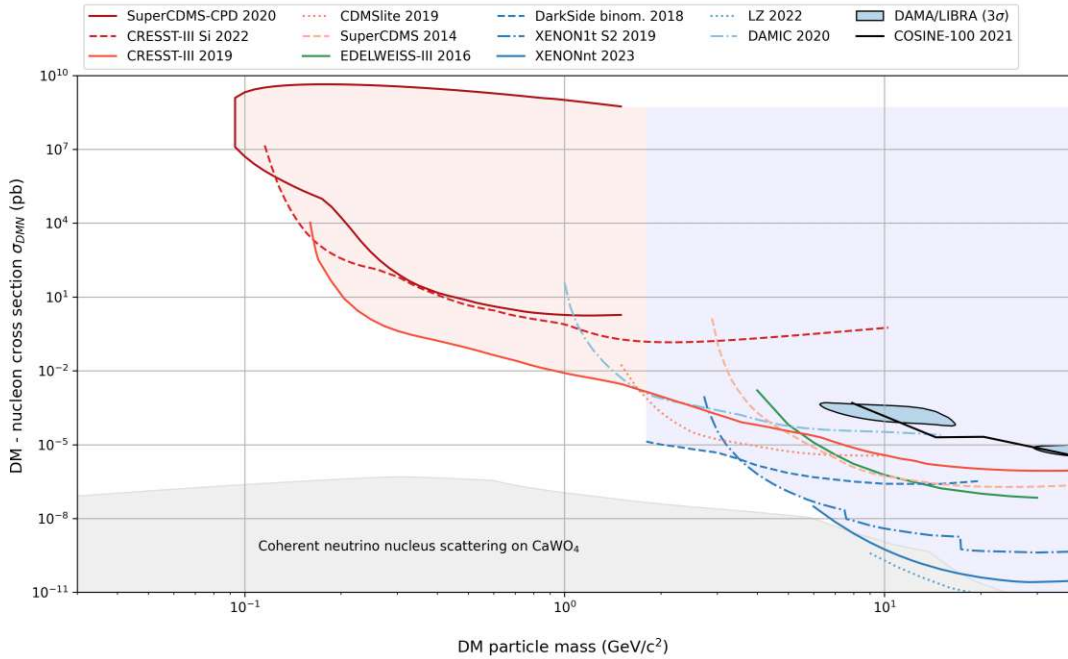
Figure 1.7: Upper limits on the cross-section of elastic DM-nucleus scattering by experiments using cryogenic solid state detectors (red) [30, 87–90] and liquid noble gas TPCs (blue) [86, 91–93]. The DAMIC experiment (light blue dash-dotted) [79] uses CCD sensors, the EDELWEISS experiment germanium ionization detectors [94]. The two islands (light blue, black edges) are from discovery claims of the DAMA experiment [37, 95]. A limit after background subtraction from the COSINE experiment (black) constrains the parameters space of the DAMA claim under the standard assumptions discussed in the text [43]. The lower part of the parameter space is harder to access experimentally due to the presence of the so-called neutrino fog (grey) [96].

and the DeLIGHT experiment planned using superfluid helium-4 targets and could reach a new optimum between target mass and low detection threshold [85, 98, 99].

Another technology that was formerly competing with TPCs is germanium ionization detectors used by the EDELWEISS experiment [94]. Those have better scaling properties than phonon readouts, but do not reach competitive energy thresholds, and not the same exposure as TPCs.

Solid state detectors are today the most sensitive devices for detection of low mass nuclear scattering, i.e. testing sub-GeV/c^2 DM masses. They use phonon readouts through superconducting thermometers to reach recoil energy threshold down to energies of eV [87]. These low thresholds are additionally enabled by collecting athermal high-energy phonons from targets with high Debye temperature and low mass. Using relatively light target nuclei enables them to detect the low energy scattering of light DM particles that would remain below the detection threshold of larger

27

targets. The strongest results in the low mass DM region are currently obtained by the CRESST and SuperCDMS [30, 87–90] experiments, where CRESST traditionally uses as scintillation light channel for PD and SuperCDMS an ionization channel. Promising preliminary results were also presented by the SPICE experiment, which is also a part of the TESSERACT collaboration. Solid state detectors are currently challenged by a limiting LEE of events whose origin is not yet fully understood. An external particle origin has been excluded, and the most likely hypothesis is stress effects in the target crystal or its interface with the sensor. The technology of superconducting thermometers and their challenges are discussed in great detail in Sec. 2. The CRESST experiment and its latest results are discussed in more detail in Sec. 5.

Scintillating targets without PD were used in the early days of DM searches. It is difficult to reach low energy thresholds with such detectors, and the impossibility of discriminating EM backgrounds requires pure target crystals and a good knowledge of the expected EM rates prior to the measurement. Often sodium iodide is used as a target due to its strong scintillation efficiency. Such searches are relatively easy to scale to larger target masses. The most prominent experiment using solely scintillating targets is the DAMA/LIBRA experiment. This experiment has gained popularity since it is the only one that claims to measure a signal compatible with a DM hypothesis. With a large measurement time of over 22 years and an exposure of 2.86 tonne years, they measure a modulation in their rate with 13.7 σ confidence that is within the expectation of the annual modulation of a DM signal [38]. Measuring the DM modulation was considered a smoking gun evidence of a discovery. However, the parameter space in which DAMA measures the signal is excluded by numerous other experiments under standard assumptions. Currently, no model of astrophysics and DM known would lead to a positive signal in the DAMA measurements but to null results in all other experiments. The DAMA experiment could so far exclude all reasonable alternative explanations of their signal modulation. The signal claim remains, therefore, a mystery and potentially one of the hottest leads to a discovery of new physics, be it DM or another yet unknown effect.

Multiple other experiments are on the quest to validate the DAMA signal claim, using the same target material to exclude a dependency on wrongly modeled interaction mechanisms with different nuclei. Out of those experiments, the ANAIS experiment has already obtained good results, excluding the DAMA islands with 3 σ confidence [42]. However, a discussion about the compatibility of the light outputs of the different crystals remains unresolved in the community, which could lead to an additional uncertainty of the confidence of the exclusion. The SABRE experiment is currently

28

built up with the special feature that they have two identical experimental sights on the northern and southern hemispheres to investigate potential environmental and seasonal effects in case they would reproduce the DAMA result [44]. The COSINE experiment is currently built up as well. They have published preliminary results from above-ground R&D measurements with which they exclude a part of the DAMA discovery space under standard assumptions [43]. However, in the used data set, they subtracted the expected background, the validity of this method is debated in the community (see Chap. 3 for analysis methods). An upcoming experiment, the COSINUS experiment, uses a two-channel readout in a cryogenic detector setting with a phonon and light channel on a sodium iodide target [45]. This technology enables the COSINUS experiment to validate the DAMA claim with much less exposure through its much lower recoil energy threshold for nuclear recoils. Furthermore, the two-channel read-out would immediately enable the identification of the recoil type in the scenario of a measured signal and resolve dependencies on the QF that can be measured in situ. The COSINUS experiment and its status are discussed in more detail in Chap. 4.

A number of experiments are specialized to measure not nuclear recoils but electron recoil DM. While this requires additional assumptions on the DM model, it also does enable searches for by far lower masses and should therefore be mentioned in the context of light particle-like DM searches. The DAMIC and SENSEI [100] experiments and their successors DAMIC-M [101] and OSCURA [102] operate large numbers of silicon CCDs to reach low thresholds and large exposure, especially for electron recoil DM. Also, the SPICE experiment plans to operate low bandgap semiconductors such as gallium arsenite, which could potentially provide them with very good energy resolutions for these DM models.

This section provided a brief overview of DM models and the experimental efforts in searching for WIMP-like DM. We will use the developed models in the following chapters when we introduce the CRESST and COSINUS experiments to test DM models and argue detector design choices for optimized sensitivity, and argue scaling techniques with machine learning for the large-scale operation of solid-state detectors in Chap. 6. Before that, we will discuss the technology of superconducting phonons sensors in more detail in the following Chap. 2.

Chapter 2

Cryogenic particle detectors with superconducting thermometers

In current experiments seeking to measure low-mass DM interactions with nuclei of a target material, the lowest recoil energy thresholds and the strongest sensitivity are obtained with cryogenic solid state detectors with TES readout. In Ref. [87], the currently lowest nuclear recoil energy threshold of only 10 eV$_{nr}$ was reported. TES are a quantum sensor technology extensively used for physics searches: they are combined with macroscopic absorber crystals for DM searches and coherent elastic neutrino nucleus scattering (CEvNS) and used as pixel arrays of X-ray sensors in astrophysics observatories. Today's technology used for DM searches is mostly based on the devices introduced in Refs. [103] and [104].

The TES is a superconducting film with low transition temperature T_c that is thermally coupled to a monocrystalline absorber. The system is cooled to mK temperatures in a dilution refrigerator, and the temperature of the TES is fine-tuned to its transition from superconducting to the normal conducting state. Within this transition, it features a strong relative change of its resistance R_f for small temperature fluctuations. The TES is operated in a readout circuit, in parallel connection with a shunt resistance R_s and a pickup coil of inductance L. This current is turned into a magnetic field by the pickup coil, then translated into a voltage signal by a superconducting quantum interference device (SQUID) amplifier and subsequently digitized. This situation is depicted in Fig. 2.1. For completeness, it should be mentioned that there are also setups where the SQUID is placed in the TES branch.

A particle recoil in the absorber creates an athermal phonon population, which thermalizes in the crystal and the sensor, producing heat and a subsequent increase in temperature. The energy threshold of TES-based detectors scales proportionally to the heat capacity of the crystal or sensor, depending on the detailed design and operation
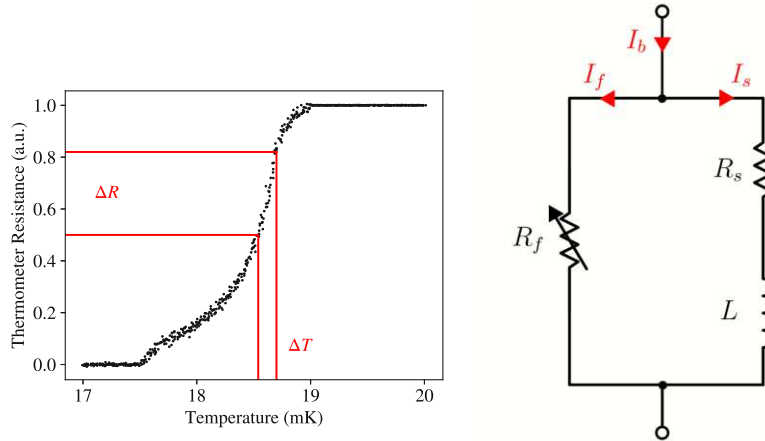
Figure 2.1: (left) Schematic view of the strong temperature dependency of a TES within the transition from its superconducting to its normal conducting state. A small temperature increase ΔT causes a strong increase of its resistance ΔR, relative to its normal conducting resistance. (right) A TES readout circuit. The sensor with temperature-dependent resistance R_f is in an electrically parallel connection with a shunt resistance R_s and a pickup coil with inductance L. The bias current I_b distributes in the two branches accordingly.

mode. Therefore, a low transition temperature of the sensor and low operation temperature, and low heat capacities are crucial to achieve suitable sensitivity.

In the first section of this chapter Sec. 2.1 we introduce a mathematical model to describe the response and noise of TES-based cryogenic detectors. Following that, we discuss experimental realizations of such detectors and elements of the involved microphysics, such that the built mathematical model can be used to obtain quantitative descriptions. A series of yet unexplained phenomena are seen across TES devices, and we introduce them in Sec. 2.2.3. Recently, other quantum sensor technologies that can be used as sensors for particle scattering have emerged. Several of them are briefly summarized in Sec. 2.2.4.

2.1 Electrothermal response model

A thermal model describing the temperature evolution in the absorber crystal and the thermometer was originally introduced in Ref. [103]. In Ref. [105], an electrothermal description of temperature response and noise for an isolated TES was studied in detail. We adapt the calculation done in these references into a combined description. For building a tractable model, we make the following assumptions:

- We include only temperature-dependencies of physical quantities that have a
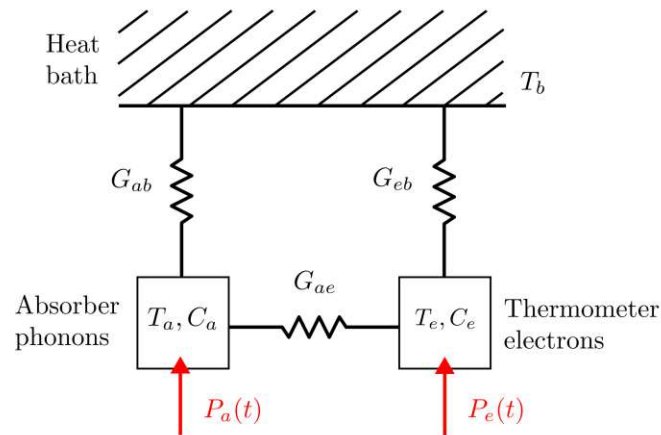
Figure 2.2: The thermal equivalent circuit of the cryogenic detector system. The phononic system of the absorber and the electronic system of a thermometer, described by their temperatures and heat capacities, are thermally coupled to each other and coupled to a joint heat bath individually. Power inputs (red) impact the temperature evolution.

strong impact in close proximity to the critical temperature of the thermometer. This includes the temperature-dependent resistance of the TES, and the heat capacity of the TES, which is expected to increase by a factor 2.43 when transitioning from the normal conducting to the superconducting state [106]. For comparability with the literature, especially Ref. [105], it is to note that the linearized temperature dependency of the TES resistance is sometimes absorbed into a parameter α, and the current dependency in a parameter β. This work considers situations where the bias current does not deviate too strongly from a reference point, which is equivalent to assuming $\beta = 0$.

- We neglect the spatial resolution of all involved components.

- We assume a monochromatic distribution of the initially produced athermal phonon population after particle recoils that thermalizes through surface scattering.

- We assume the heat capacity of the non-conducting absorber crystal to be dominated by its lattice phonons and that of the metallic sensor by its electrons.

- We assume that all thermal components are coupled to an ideal heat bath with infinite heat capacity.

These assumptions are sufficiently met in typical devices. However, they restrict the reliability of the model to a temperature range close to the critical temperature.

Furthermore, predictions derived from the model by changing physical quantities have to be controlled for consistency with measurement, as several of the effects neglected through our assumptions are absorbed in the parameters of the model. The evolution of temperatures of the absorber T_a and the thermometer T_e are in this model fully described by the heat capacities of the absorber's phononic system C_a and the thermometer's electronic system C_e, their thermal coupling to each other G_{ae} and to a joint heat bath of temperature T_b, the time-dependent power inputs in the photonic system $P_a(t)$ and electronic system $P_e(t)$ and the initial temperatures. This situation can be described as a thermal-equivalent circuit and is depicted in Fig. 2.2.

The thermal and electronic circuits can be expressed individually by systems of ordinary differential equations (ODEs). They accumulate into a joint ODE system through the coupling between them caused by the temperature-dependency of the TES's resistance:

$$C_e \frac{dT_e}{dt} + (T_e - T_a)\,G_{ea} + (T_e - T_b)\,G_{eb} = P_e(t), \tag{2.1}$$

$$C_a \frac{dT_a}{dt} + (T_a - T_e)\,G_{ea} + (T_a - T_b)\,G_{ab} = P_a(t), \tag{2.2}$$

$$L \frac{dI_f}{dt} + R_s I_b - (R_f(T_e) + R_s) I_f = 0. \tag{2.3}$$

Here, Eqs. (2.1) and (2.2) describe the temperature evolution in the thermometer and absorber components with a heat equation of zero spatial dimensions, respectively. Eq. (2.3) is a mesh analysis of the currents in the readout loop. The power inputs in the components have several contributions:

- The thermalization of an athermal phonon population induced by particle recoils in the crystal releases heat in the crystal and thermometer on an exponential time scale τ_n. A share ϵ of the phonon population thermalizes in the thermometer. The total released heat over time equates to the deposited energy ΔE in the particle recoil.

- The electrical resistance of the TES causes Joule heating on the TES. This effect is also called self-heating, bias heating or electrothermal feedback (ETF).

- A constant heating is induced by a heating resistor R_H deposited on the crystal. For controlling the magnitude of the heating, we adopt the conventions of the electronics of the CRESST experiment, where a digital-analog converted (DAC) value between zero and ten volt is the output, interpolating linearly between

the maximal heating current I_H and no heating. We introduce an additional parameter δ_H that absorbs any deviation of the power released to the TES from the average temperature increase in the heater. This absorbs any spacial dependency of the temperature distribution in the crystal and can also be used to model non-standard heater placement, e.g. in direct thermal connectivity with the TES.

- Additionally to the heating, heater pulses can be injected through the heating resistor. Such heater pulses mimic the released heat from phonon thermalization on an exponential time scale τ_{TP}, and their magnitude and distribution of released heat are analogous to the constant heating controlled by parameters TPA and δ. The test pulse amplitude (TPA) is the parameter controlling the height of injected heater pulses, and we will use is extensively in Chap. 3 for energy calibration and other analysis. The parameter δ can deviate from the parameter δ_H, which additionally absorbs potential athermal phonon populations produced by the heater and other non-equilibrium effects.

The linear and uncorrelated control scales of heating and heater pulses can, in practice, be realized by square-rooting the sum of the heater and pulse outputs. In summary, the power inputs for particle recoils, or heater pulse injection at $t = 0$, for times $t > 0$, read:

$$
\begin{aligned}
P_e(t) &= \epsilon \frac{\Delta E}{\tau_n} \exp(\frac{t}{\tau_n}) + R_f I_f^2 \\
&\quad + \delta \frac{TPA}{10} \exp(\frac{t}{\tau_{TP}}) \beta R_H I_H^2 + \delta_H \frac{DAC}{10} R_H I_H^2, \\
P_a(t) &= (1-\epsilon) \frac{\Delta E}{\tau_n} \exp(\frac{t}{\tau_n}) \\
&\quad + (1-\delta) \frac{TPA}{10} \exp(\frac{t}{\tau_{TP}}) \beta R_H I_H^2 + (1-\delta_H) \frac{DAC}{10} R_H I_H^2.
\end{aligned}
$$

(2.4)

(2.5)

The ODE system (2.1-2.3) is in general not analytically solvable. However, it can be simplified in certain practically useful scenarios to derive analytic solutions of the response to particle recoils and noise conditions in equilibrium. Furthermore, a numerical solution can always be calculated with standard ODE solvers. For this, the equations can be generalized in matrix-vector notation:

$$\dot{\underline{T}}(t) = \operatorname{diag}(\underline{C})^{-1}\bigg(\underline{P}\Big(t, \underline{T}(t), \underline{I_f}(t)\Big)$$

$$+ \operatorname{diag}\big(\underline{G_b}\big)\big(\underline{T_b} - \underline{T}(t)\big) + (\underline{\underline{G}} - \operatorname{diag}(\underline{\underline{G}}\underline{1}))\underline{T}(t)\bigg), \tag{2.6}$$

$$\dot{\underline{I}_f}(t) = \operatorname{diag}(\underline{L})^{-1}\bigg(\operatorname{diag}\big(\underline{R_s}\big)\,\underline{I_b} - \operatorname{diag}\Big(\underline{I_f}(t)\Big)\Big(\underline{R_f}(\underline{T}(t)) + \underline{R_s}\Big)\bigg).$$

Here, underlined (double underlined) quantities are vectors (matrices). This system has the advantage that it can describe an arbitrary number of thermal components and readout loops, which is useful for detector designs with multiple TES or additional separate thermally coupled crystals. $\underline{\underline{G}}$ is the symmetric matrix containing the thermal couplings between all components. All other quantities are analogous to their scalar versions generalized to vectors.

To study analytical solutions, we separate the system into three temperature scales that interact with each other only negligibly:

$$T = T_b + \Delta T_{\text{heating}} + \Delta T_{\text{pulse}} + \Delta T_{\text{noise}},$$

where

$$T_b + \Delta T_{\text{heating}} \gg \Delta T_{\text{pulse}} \gg \Delta T_{\text{noise}}.$$

In the following, we discuss solutions of the system on the temperature scale of low energy particle recoils in the absorber ΔT_{pulse}, noise fluctuations ΔT_{noise} and the equilibrium state $T_b + \Delta T_{heating}$.

2.1.1 Small signal approximation

Consider a particle recoiling in the absorber crystal. We introduced as part of Eqs. (2.4 - 2.5) the resulting power inputs in the thermal system, which we call $P_{e,\ particle}(t)$ and $P_{a,\ particle}(t)$. We assume a low energy particle recoil, causing a small enough temperature rise that we can linearly approximate the resistance change in the TES, and treat the heat capacity of the TES as constant. Furthermore, we ignore the feedback of the pickup coil in the readout circuit, $L = 0$. In this scenario, we only need to consider the thermal equations of our ODE system Eqs. (2.4) and (2.5). To isolate the effect of the particle recoil from the equilibrium state, we separate constant and time-dependent parts of the quantities by substitution:

$$T_e(t) = T_{e0} + \Delta T_e(t),$$
$$I_f(t) = I_{f0} + \Delta I_f(t),$$
$$P_e(t) = P_{e0} + \Delta P_e(t),$$

and equivalent for the absorber equation. The explicit time-dependency is in the further considerations omitted for better readability. Before we insert these substitutions in the system of equations, we treat the self-heating term in Eq. (2.4). It is the only nonlinear term, and we use its subdominant effect on the total heating to approximate it linearly. In the following, we depict with $\approx$ a Taylor series in first order:

$$R_f I_f^2 \approx R_{f0} I_{f0}^2 + \frac{d}{dT_e} \left(R_f I_f^2 \right) (T_{e0} + \Delta T_e),$$

where the total derivative can be found as

$$
\begin{aligned}
\frac{d}{dT_e} \left(R_f I_f^2 \right) &= \frac{d}{dT_e} \left(\frac{R_f R_s^2 I_0^2}{(R_f + R_s)^2} \right) \\
&= \left(\frac{-2 R_{f0} R_s^2 I_0^2}{(R_f + R_s)^3} \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} + \frac{R_s^2 I_0^2}{(R_f + R_s)^2} \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \right) \\
&= \left(\frac{-2 R_{f0} R_s^2 I_0^2}{(R_f + R_s)^3} + \frac{R_s^2 I_0^2 (R_f + R_s)}{(R_f + R_s)^3} \right) \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \\
&= -I_f^2 \frac{(R_{f0} - R_s)}{(R_{f0} + R_s)} \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \\
&\stackrel{\text{def}}{=} -G_{ETF}.
\end{aligned}
$$

Here, we used the current divider rule for the currents in the two branches of the readout circuit. This description of the self-heating effect as an additional coupling to the heat bath allows us to absorb it in the coupling constant G_{eb} and express them as a joint and effective thermal coupling:

$$G_{eff} = G_{eb} + G_{ETF}$$

The equations can then be summarized into the standard matrix-vector form of a first-order ODE system:

$$\frac{d}{dt}\begin{pmatrix}\Delta T_e \\ \Delta T_a\end{pmatrix} = -\begin{pmatrix}\frac{G_{ea}+G_{eff}}{C_e} & -\frac{G_{ea}}{C_e} \\ -\frac{G_{ea}}{C_a} & \frac{G_{ea}+G_{ab}}{C_a}\end{pmatrix}\begin{pmatrix}\Delta T_e \\ \Delta T_a\end{pmatrix} + \begin{pmatrix}\frac{\Delta P_e}{C_e} \\ \frac{\Delta P_a}{C_a}\end{pmatrix}. \tag{2.7}$$

This procedure is also legitimate if the absorber and thermometer have different base temperatures in their equilibrium state. An analytical solution to this system can be derived by using the fact that the expression

$$x = X(t)\int_0^t X(s)^{-1}b(s)ds \tag{2.8}$$

is a solution to a linear ODE system

$$x' = A(t)x + b(t), \tag{2.9}$$

where $X(t)$ is the fundamental matrix whose columns are $e^{\lambda_1 t}\vec{v}_1, e^{\lambda_2 t}\vec{v}_2$ with (λ_i, v_i) the eigenvalues and -vectors of the coefficient matrix of Eq. (2.7):

$$X = \begin{pmatrix}e^{\lambda_1 t}v_1, & e^{\lambda_2 t}v_2\end{pmatrix}.$$

This statement also holds for coefficient matrices with non-constant coefficients, where the integral can always be solved numerically. In our case, the coefficient matrix of Eq. (2.7) consists of constants, and the eigenvalues read

$$\lambda_{1,2} = \frac{-a \mp \sqrt{a^2 - 4b}}{2} \stackrel{\text{def}}{=} -\frac{1}{\tau_{1,2}} \qquad \left(\stackrel{\text{def}}{=} -\frac{1}{\tau_{in,t}} \stackrel{\text{def}}{=} -s_{in,t}\right),$$

with

$$a \stackrel{\text{def}}{=} \frac{(G_{ea} + G_{eff})C_a + (G_{ea} + G_{ab})C_e}{C_e C_a}, \qquad b \stackrel{\text{def}}{=} \frac{(G_{ea} + G_{eff})(G_{ea} + G_{ab}) - G_{ea}^2}{C_e C_a}.$$

The eigenvalue λ_1 (λ_2) induces a smaller (larger) time constant $\tau_1 = \tau_{in}$ ($\tau_2 = \tau_t$) that can be interpreted as the thermal relaxation time of the thermometer (absorber). We fixate an orthonormal set of eigenvectors by choosing a specific parametrization that will be useful later in this section:

$$v_{1,2} = \begin{pmatrix}\alpha_{1,2} \\ 1\end{pmatrix}\left(1 + \alpha_{1,2}^2\right)^{-1/2}, \qquad \alpha_{1,2} \stackrel{\text{def}}{=} 1 + \frac{G_{ab}}{G_{ea}} + \lambda_{1,2}\frac{C_a}{G_{ea}}. \tag{2.10}$$

The matrix $X(t)$ and its inverse can now be derived:

$$X(t) = \begin{pmatrix} (1+\alpha_1^2)^{-1/2} e^{\lambda_1 t}\alpha_1 & (1+\alpha_2^2)^{-1/2} e^{\lambda_2 t}\alpha_2 \\ (1+\alpha_1^2)^{-1/2} e^{\lambda_1 t} & (1+\alpha_2^2)^{-1/2} e^{\lambda_2 t} \end{pmatrix}, \tag{2.11}$$

$$X(t)^{-1} = \frac{1}{(\alpha_1 - \alpha_2)} \begin{pmatrix} \sqrt{(1+\alpha_1^2)}e^{-\lambda_1 t} & -\sqrt{(1+\alpha_1^2)}e^{-\lambda_1 t}\alpha_2 \\ -\sqrt{(1+\alpha_2^2)}\,e^{-\lambda_2 t} & \sqrt{(1+\alpha_2^2)}\,e^{-\lambda_2 t}\alpha_1 \end{pmatrix}. \tag{2.12}$$

With that, we can calculate solutions for arbitrary inhomogeneities $\Delta P_e(t)$ and $\Delta P_a(t)$. By using the power inputs expected from scattering particles, we introduce a third time constant τ_n in the system that corresponds to the effective thermalization time of phonons in the combined system of crystal and thermometer:

$$\Delta P_e(s) = \epsilon \frac{\Delta E}{\tau_n} \exp\left(\frac{-t}{\tau_n}\right) \qquad \Delta P_a(s) = (1-\epsilon)\frac{\Delta E}{\tau_n} \exp\left(\frac{-t}{\tau_n}\right).$$

The integrant of Eq. (2.8) can then be written as

$$I(s) = \frac{1}{(\alpha_1 - \alpha_2)} \begin{pmatrix} \sqrt{(1+\alpha_1^2)}e^{-\lambda_1 s} & -\sqrt{(1+\alpha_1^2)}e^{-\lambda_1 s}\alpha_2 \\ -\sqrt{(1+\alpha_2^2)}\,e^{-\lambda_2 s} & \sqrt{(1+\alpha_2^2)}\,e^{-\lambda_2 s}\alpha_1 \end{pmatrix} \dots$$
$$\dots \begin{pmatrix} \frac{\epsilon \frac{\Delta E}{\tau_n} \exp\left(\frac{-s}{\tau_n}\right)}{C_e} \\ \frac{(1-\epsilon)\frac{\Delta E}{\tau_n} \exp\left(\frac{-s}{\tau_n}\right)}{C_a} \end{pmatrix}$$
$$= \frac{\frac{\Delta E}{\tau_n}}{(\alpha_1 - \alpha_2)} \begin{pmatrix} \sqrt{(1+\alpha_1^2)}\left(\frac{\epsilon}{C_e} - \alpha_2\frac{(1-\epsilon)}{C_a}\right)\exp\left(-\frac{s}{\tau_n} + \frac{s}{\tau_1}\right) \\ -\sqrt{(1+\alpha_2^2)}\left(\frac{\epsilon}{C_e} - \alpha_1\frac{(1-\epsilon)}{C_a}\right)\exp\left(-\frac{s}{\tau_n} + \frac{s}{\tau_2}\right) \end{pmatrix},$$

and the integral in Eq. (2.8) as

$$\int_0^t I(s)ds = \frac{\frac{\Delta E}{\tau_n}}{(\alpha_1 - \alpha_2)} \begin{pmatrix} \sqrt{(1+\alpha_1^2)}\left(\frac{\epsilon}{C_e} - \alpha_2\frac{(1-\epsilon)}{C_a}\right)\left(-\frac{1}{\tau_n} + \frac{1}{\tau_1}\right)^{-1}\dots \\ \dots\left(\exp\left(-\frac{t}{\tau_n} + \frac{t}{\tau_1}\right) - 1\right) \\ -\sqrt{(1+\alpha_2^2)}\left(\frac{\epsilon}{C_e} - \alpha_1\frac{(1-\epsilon)}{C_a}\right)\left(-\frac{1}{\tau_n} + \frac{1}{\tau_2}\right)^{-1}\dots \\ \dots\left(\exp\left(-\frac{t}{\tau_n} + \frac{t}{\tau_2}\right) - 1\right) \end{pmatrix}.$$

By multiplying again with $X(t)$, we reach the solution of the system, which we state separately for the temperature rise in the thermometer and absorber induced by the particle recoil:

$$\Delta T_e(t) = \frac{\alpha_1 \frac{\Delta E}{\tau_n}}{(\alpha_1 - \alpha_2)} \left(\alpha_2 \frac{(1-\epsilon)}{C_a} - \frac{\epsilon}{C_e} \right) \left(\frac{1}{\tau_n} - \frac{1}{\tau_1} \right)^{-1} \left(e^{-\frac{t}{\tau_n}} - e^{-\frac{t}{\tau_1}} \right) \dots$$

$$\dots + \frac{\alpha_2 \frac{\Delta E}{\tau_n}}{(\alpha_1 - \alpha_2)} \left(\alpha_1 \frac{(1-\epsilon)}{C_a} - \frac{\epsilon}{C_e} \right) \left(\frac{1}{\tau_n} - \frac{1}{\tau_2} \right)^{-1} \left(e^{-\frac{t}{\tau_2}} - e^{-\frac{t}{\tau_n}} \right)$$

$$\Delta T_e(t) \stackrel{\text{def}}{=} A_n \left(e^{-\frac{t}{\tau_n}} - e^{-\frac{t}{\tau_1}} \right) + A_t \left(e^{-\frac{t}{\tau_2}} - e^{-\frac{t}{\tau_n}} \right), \tag{2.13}$$

and

$$\Delta T_a(t) = \frac{\frac{\Delta E}{\tau_n}}{(\alpha_1 - \alpha_2)} \left(\alpha_2 \frac{(1-\epsilon)}{C_a} - \frac{\epsilon}{C_e} \right) \left(\frac{1}{\tau_n} - \frac{1}{\tau_1} \right)^{-1} \left(e^{-\frac{t}{\tau_n}} - e^{-\frac{t}{\tau_1}} \right) \dots$$

$$\dots + \frac{\frac{\Delta E}{\tau_n}}{(\alpha_1 - \alpha_2)} \left(\alpha_1 \frac{(1-\epsilon)}{C_a} - \frac{\epsilon}{C_e} \right) \left(\frac{1}{\tau_n} - \frac{1}{\tau_2} \right)^{-1} \left(e^{-\frac{t}{\tau_2}} - e^{-\frac{t}{\tau_n}} \right)$$

$$\Delta T_a(t) = \frac{1}{\alpha_1} A_n \left(e^{-\frac{t}{\tau_n}} - e^{-\frac{t}{\tau_1}} \right) + \frac{1}{\alpha_2} A_t \left(e^{-\frac{t}{\tau_2}} - e^{-\frac{t}{\tau_n}} \right). \tag{2.14}$$

Note that these equations describe two pulses with individual amplitudes and three overall time constants. The faster pulse, which incorporates τ_1, can be interpreted as the thermalization of athermal phonons in the thermometer and the subsequent relaxation of the thermometer. The slower pulse that incorporates τ_2 can be interpreted as a similar process in the absorber. The values α_1 (α_2) are the values by which the faster (slower) pulse shape in the thermometer temperature is amplified or suppressed relative to the absorber temperature. We considered our equation only for times $t > 0$. To generalize it to any real-values time, we need to write an additional Heavyside function in the power inputs and solution to preserve the physical description of a sudden power input induced by scattering:

$$\Delta T_e(t) = \Theta(t) \left[A_n \left(e^{-t/\tau_n} - e^{-t/\tau_{in}} \right) + A_t \left(e^{-t/\tau_t} - e^{-t/\tau_n} \right) \right], \tag{2.15}$$

$$\Delta P_e(s) = \epsilon \frac{\Delta E}{\tau_n} \exp \left(\frac{-t}{\tau_n} \right), \tag{2.16}$$

$$\Delta P_a(s) = (1 - \epsilon) \frac{\Delta E}{\tau_n} \exp \left(\frac{-t}{\tau_n} \right). \tag{2.17}$$

The amplitudes of the pulses in the thermometer can be written in compact forms

$$A_1 \stackrel{\text{def}}{=} A_n \stackrel{\text{def}}{=} \frac{\frac{\Delta E}{\tau_n}}{\left(1 - \frac{\alpha_2}{\alpha_1} \right)} \left(\alpha_2 \frac{(1-\epsilon)}{C_a} - \frac{\epsilon}{C_e} \right) \left(\frac{1}{\tau_n} - \frac{1}{\tau_1} \right)^{-1} \tag{2.18}$$

$$= - \frac{P_0(s_{in} - (G_{ab}/C_a))}{\varepsilon(s_{in} - s_t)(s_{in} - s_n)} \left(\frac{s_t - (G_{ab}/C_a)}{G_{eff} - (C_e/C_a)G_{ab}} - \frac{\varepsilon}{C_e} \right), \tag{2.19}$$

C_e (pJ/mK)	C_a (pJ/mK)	G_{eff} (pW/mK)	G_{ab} (pW/mK)	G_{ea} (pW/mK)	ϵ	τ_n (ms)
$1 \cdot 10^{-3}$	$3.8 \cdot 10^{-2}$	3	0.7	$4.5 \cdot 10^{-2}$	0.08	2

Table 2.1: Physical quantities of an exemplary cryogenic detector, adapted from Ref. [103], at the operation temperature of 20 mK. The calculated relaxation times of the thermometer and absorber are $\tau_{in} = 0.373$ ms and $\tau_{in} = 51.1$ ms. The amplitudes in the thermometer are $A_n = 154$ μK and $A_t = 3.61$ μK, and the amplitude of the thermal temperature rise in the absorber is 243 μK.

and

$$A_2 \stackrel{\text{def}}{=} A_t \stackrel{\text{def}}{=} -\frac{\frac{\Delta E}{\tau_n}}{\left(1 - \frac{\alpha_1}{\alpha_2}\right)} \left(\alpha_1 \frac{(1-\epsilon)}{C_a} - \frac{\epsilon}{C_e}\right) \left(\frac{1}{\tau_n} - \frac{1}{\tau_2}\right)^{-1} \tag{2.20}$$

$$= \frac{P_0(s_t - (G_{ab}/C_a))}{\varepsilon(s_t - s_{in})(s_t - s_n)} \left(\frac{s_{in} - (G_{ab}/C_a)}{G_{eff} - (C_e/C_a)G_{ab}} - \frac{\varepsilon}{C_e}\right), \tag{2.21}$$

where the second lines match the formulas Ref. [103], except for a global sign (Eq. (2.19), marked in red), which was accidentally omitted in Ref. [103], according to private communications with the authors. The temperature rises in the absorber can be derived by dividing A_n (A_t) by α_1 (α_2).

We confirm our formula by comparing it with a numerically calculated solution of Eqs. (2.7). We use values for all physical quantities from Ref. [103]. They are summarized in Tab. 2.1, jointly with the calculated values for the relaxation times and amplitudes. The resulting pulse shape from a particle recoil is shown in Fig. 2.3. The rise time of the temperature pulse in the absorber corresponds to τ_n, while that in the thermometer corresponds for the chosen detector to $\tau_{in} < \tau_n$. This scenario is called the bolometric mode because the thermometer relaxes faster than it heats up, and the temperature rise effectively scales with the power input. The scenario $\tau_{in} > \tau_n$ would be called calorimetric mode because the thermometer accumulates the power, and the temperature rise scales with the total amount of deposited energy. In this scenario, the rise time of the pulses in the absorber and thermometer are both τ_n. These two modes, and their impact on the sensitivity of the device, are discussed in more detail in Sec. 2.2.2. We can observe a discrepancy between the calculated thermal pulse amplitudes and the observed PHs of a factor $\approx 1.8(1.2)$ for the thermometer (absorber), observable by comparing Tab. 2.1 with the pulses in Fig. 2.3. This discrepancy arises because the values of τ_{in} and τ_n are relatively close
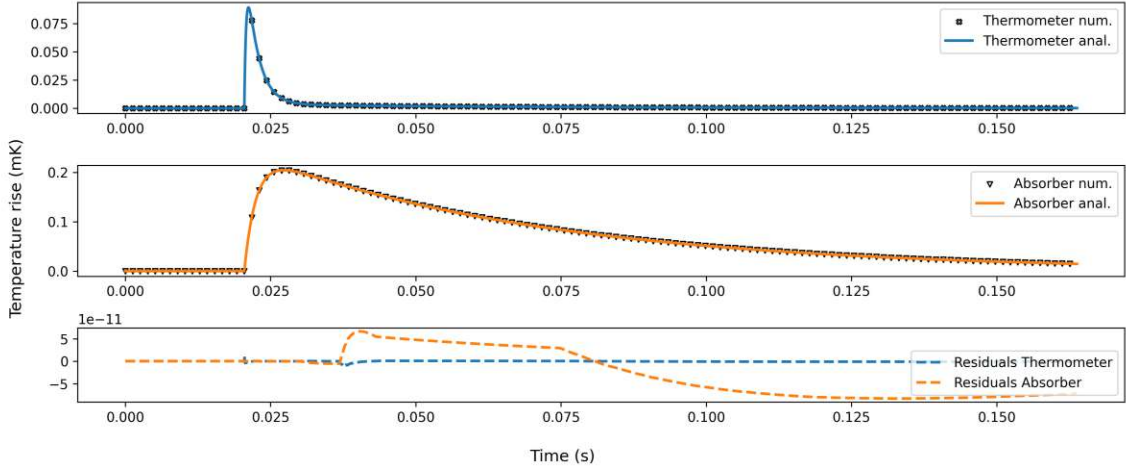
Figure 2.3: Resulting pulse-shaped temperature rise from a 60 keV absorber recoil in the thermometer (blue, top panel) and absorber (orange, central panel) of a cryogenic detector with the physical quantities summarized in Tab. 2.1. Shown are the analytic solutions calculated with Eqs. (2.13) and (2.14) (continuous lines), numerical solutions calculated from Eqs. (2.7), and their residuals (bottom panel). The pulse in the thermometer (absorber) has a height of 86.4 (204) μK, consistent with the pulse heights (PHs) reported in Ref. [103].

together: the athermal pulse component starts to fall again, before it has risen fully. The value A_n is therefore only in a fully calorimetric or fully bolometric mode a useful description of the height of the non-thermal pulse component.

To derive the observable voltage pulse shape, we need to calculate the proportional current change in the SQUID branch of the readout circuit caused by ΔT_e. This is given by rearranging Eq. 2.3 in a small signal approximation while neglecting the impact of the pickup coil $L \approx 0$:

$$\Delta I_s(\Delta T_e) \approx \frac{I_b R_s}{(R_s + R_f(T_{e0}))^2} \left.\frac{dR_f}{dT_e}\right|_{T_{e0}} \Delta T_e \tag{2.22}$$

We will use this formula in the next Sec. 2.1.2, to compare the magnitude of voltage pulses with the omnipresent sensor noise. Neglecting the coil is justified, since it typically has values of $\approx \mu$H and causes therefore only a pole in frequencies far above the interesting spectrum.

2.1.2 Noise fluctuations

To quantify the sensitivity of a cryogenic detector, we are interested in its thermal response to particle recoils and the omnipresent noise in the sensor, above which

the recoil signature has to rise to induce a detectable event. A noise model can be derived by adapting the procedure described in the previous section, as was done in Ref. [105]. To simplify the model to its most relevant contributions, we consider only the thermal equation of the TES Eq. (2.4) and the electrical equation of its readout circuit Eq. (2.3). This effectively means that we ignore noise contributions that arise in the absorber by thermally decoupling the TES from the absorber, i.e. $G_{ea} = 0$. While this simplification seems justified, since energy deposition or temperature fluctuations in the TES should have a much larger effect on the observables, the exact contribution of noise from the absorber is unclear and is briefly discussed in Sec. 2.2.3. The system then reads:

$$C_e \frac{dT_e}{dt} + (T_e - T_b)G_{eb} = \Delta P_{e,noise} + P_J, \qquad (2.23)$$

$$-L\frac{dI_f}{dt} + R_S I_b - (R_f(T_f) + R_S)I_f = \Delta U_{noise}, \qquad (2.24)$$

where we consider power and voltage inputs in the system from noise fluctuations and Joule heating. The equivalent substitutions as in the previous section can be applied, i.e. $T_e \approx T_{e0} + \Delta T_e$, $I_f \approx I_{f0} + \Delta I_f$ and $P_e \approx P_{e0} + \Delta P_e$. A small signal approximation is used by applying a first-order Taylor approximation on nonlinear terms. For this, we linearize the self-heating differently from the previous section, namely with partial derivatives in the state variables. The expression of the Joule heating that induces fluctuations above the equilibrium state differs for different types of noise contributions that originate internally in the TES and externally. For internally produced noise, additional work done by the bias current on an internal source has to be taken into account:

$$P_{J,ext} = R_f I_f^2,$$
$$P_{J,int} = I_f \left(R_f I_f + \Delta U_{noise} \right).$$

Here, we introduce color coding to highlight that calculations are applied equivalently to the two versions of the equations. Treating the external term first, we can linearize:

$$P_{J,ext} = R_f I_f^2 \approx R_{f0} I_{f0}^2 + \frac{\partial}{\partial I_f} \left(R_f I_f^2 \right) (I_{f0} + \Delta I_f) + \frac{\partial}{\partial T_e} \left(R_f I_f^2 \right) (T_{e0} + \Delta T_e),$$

$$\frac{\partial}{\partial I_f} \left(R_f I_f^2 \right) = 2R_{f0} I_{f0} \Delta I_f,$$

$$\frac{\partial}{\partial T_e} \left(R_f I_f^2 \right) = \frac{\partial}{\partial R_f} \left(R_f I_f^2 \right) \frac{\partial R_f}{\partial T_e} = I_{f0}^2 \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \Delta T_e.$$

We have used $\Delta R_f \approx \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \Delta T_e$. Inserting the substitutions and linearized terms in Eqs. (2.23) and (2.23) leads to the system:

$$C_e \frac{d\Delta T_e}{dt} + \Delta T_e G_{eb} = \Delta P_e + 2R_{f0} I_{f0} \Delta I_f + I_{f0}^2 \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \Delta T_e,$$

$$L \frac{d\Delta I_f}{dt} + (R_{f0} + R_S) \Delta I_f + I_{f0} \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \Delta T_e = \Delta U ,$$

where we omitted the extended subscripts of the power and voltage inputs for better readability. This system can be treated conveniently in Fourier space, where time derivatives can be expressed as complex coefficients. Exploiting this, we rearrange the equations into a system of linear equations:

$$\left(G_{eb} - I_{f0}^2 \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} + 2\pi w i C_e \right) \Delta T_e - 2R_{f0} I_{f0} \Delta I_f = \Delta P_e,$$

$$I_{f0} \left. \frac{dR_f}{dT_e} \right|_{T_{e0}} \Delta T_e + (R_{f0} + R_S + 2\pi w i L) \Delta I_f = \Delta U.$$

In Ref. [105], several time constants were introduced to express the underlying physics and improve the symmetries of the system. These are:

$$\tau \overset{\text{def}}{=} \frac{C_e}{G_{eb}}, \qquad\qquad \tau_{el} \overset{\text{def}}{=} \frac{L}{R_{f0} + R_S},$$

$$\tau_I \overset{\text{def}}{=} \frac{\tau}{(1 - \mathcal{L}_I)}, \qquad\qquad \mathcal{L}_I \overset{\text{def}}{=} \frac{I_{f0}^2}{G_{eb}} \left. \frac{dR_f}{dT_e} \right|_{T_{e0}},$$

where τ is the intrinsic relaxation time of the isolated thermometer, without considering the effect of self-heating, τ_{el} is the characteristic time constant of the RL circuit that is the readout loop, τ_I is the relaxation time of the thermometer, including the effect of the self-heating, and $\mathcal{L}_I$, called the low-frequency loop gain under constant current,

43

characterizes the sensitivity of the thermometer to temperature fluctuations. The value $\mathcal{L}_I$ is a useful control variable for the stability of the TES: a scenario with $\mathcal{L}_I > 1$ is called thermal runaway, and can cause unstable operation conditions. The system can be written in a compact matrix-vector form:

$$\begin{pmatrix} \left(\frac{1}{(\tau_I)_{ext}} + 2\pi wi\right) C_e & (-2R_{f0}I_{f0})_{ext} \\ \frac{G_{eb}\mathcal{L}_I}{I_{f0}} & \left(\frac{1}{\tau_{el}} + 2\pi wi\right) L \end{pmatrix} \begin{pmatrix} \Delta T_e \\ \Delta I_f \end{pmatrix} = \begin{pmatrix} \Delta P_e \\ \Delta U \end{pmatrix}. \tag{2.25}$$

When comparing with Ref. [105], be aware that the element (2,1) of this matrix contains the equilibrium current through the TES I_{f0} for us, and the bias current I_b for them. This is because they are using a voltage-biased circuit, where the current through the TES equals the total current in the loop, while we treat a current-biased parallel loop, where currents distribute in the branches. The equivalent calculation can be done for noise fluctuations internal to the TES, where the green marked $(\cdot)_{ext}$ terms need to be replaced with their internal equivalents. We derive these elements by linearizing the internal expression for Joule heating:

$$\begin{aligned} P_J &= I_f \left(R_f I_f + \Delta U \right) \\ &= I_f \left(R_s I_s + L\frac{dI_s}{dt} \right) \\ &= I_f \left(R_s(-I_f + I_b) - L\frac{dI_f}{dt} \right) \\ &= -I_f^2 R_s + I_f I_b R_s - I_f L\frac{dI_f}{dt} \\ &\approx -I_{f0}^2 R_s - 2R_s I_{f0}\Delta I_f + I_{f0} I_b R_s + \Delta I_f I_b R_s - (I_{f0} + \Delta I_f)L\frac{d(I_{f0} + \Delta I_f)}{dt} \\ &= I_{f0}(-I_{f0} + I_b)R_s + (-2I_{f0} + I_b) R_s \Delta I_f - I_{f0}L\frac{d\Delta I_f}{dt}, \end{aligned}$$

and expressing the result in Fourier space as done for the external form, where we also use the formula $I_b R_s = I_{f0}(R_s + R_{f0} + 2i\pi wL)$ from the initial mesh analysis of the readout loop:

$$\begin{aligned} P_J(w) &= -I_{f0}^2 R_s + I_b R_s + (-2I_{f0}R_s + I_b R_s - 2i\pi w I_{f0}L)\,\Delta I_f \\ \frac{-\Delta P_J(w)}{\Delta I_f} &= (2I_{f0} - I_b)\, R_s + 2i\pi w I_{f0}L \\ &= I_{f0}(R_s - R_{f0})_{int}. \end{aligned}$$

Additionally, we do not observe in this derivation the self-heating term depending on ΔT, and the green marked element in the (1,1) element of Eq. 2.25 should be replaced by its internal equivalent. The $(\cdot)_{ext}$ terms in the (1,1) and (1,2) elements are, therefore, to be replaced with:

$$(\tau)_{int}, \ (I_{f0}R_s - I_{f0}R_{f0})_{int}. \tag{2.26}$$

Eq. (2.25) can be solved by matrix inversion, which provides a transfer matrix for the resulting temperature fluctuations in the TES, and current fluctuations through the TES, from initial power fluctuations in the TES and voltage fluctuations in the readout circuit:

$$\begin{pmatrix} \Delta T_e \\ \Delta I_f \end{pmatrix} \overset{\text{def}}{=} \begin{pmatrix} s_{11}(w) & s_{12}(w) \\ s_{21}(w) & s_{22}(w) \end{pmatrix} \begin{pmatrix} \Delta P_e \\ \Delta U \end{pmatrix}.$$

The calculation, exemplary for the external form, reads:

$$\begin{pmatrix} s_{11}(w) & s_{12}(w) \\ s_{21}(w) & s_{22}(w) \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{(\tau_I)_{ext}} + 2\pi wi\right) C_e & (-2R_{f0}I_{f0})_{ext} \\ \frac{G_{eb}\mathcal{L}_I}{I_{f0}} & \left(\frac{1}{\tau_{el}} + 2\pi wi\right) L \end{pmatrix}^{-1}$$

$$= \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \text{ ... Inversion formula 2x2 matrix}$$

$$= \frac{1}{\left(\frac{1}{(\tau_I)_{ext}} + 2\pi wi\right) C_e \left(\frac{1}{\tau_{el}} + 2\pi wi\right) L - \frac{G_{eb}\mathcal{L}_I}{I_{f0}}(-2R_{f0}I_{f0})_{ext}} ...$$

$$... \begin{pmatrix} \left(\frac{1}{\tau_{el}} + 2\pi wi\right) L & (2R_{f0}I_{f0})_{ext} \\ -\frac{G_{eb}\mathcal{L}_I}{I_{f0}} & \left(\frac{1}{(\tau_I)_{ext}} + 2\pi wi\right) C_e \end{pmatrix},$$

resulting in the solutions for the relevant matrix elements of the external transfer matrix:

$$s_{21}^{ext}(w) = \frac{-\frac{G_{eb}\mathcal{L}_I}{I_{f0}}}{\left(\frac{1}{(\tau_I)_{ext}} + 2\pi wi\right)C_e\left(\frac{1}{\tau_{el}} + 2\pi wi\right)L - \frac{G_{eb}\mathcal{L}_I}{I_{f0}}(-2R_{f0}I_{f0})_{ext}}$$

$$= -\frac{1}{I_{f0}}\left[\frac{L\tau}{(\tau_I)_{ext}\tau_{el}\mathcal{L}_I} - \frac{(-2R_{f0}I_{f0})_{ext}}{I_{f0}}\right.$$

$$\left. + 2\pi wi\frac{L\tau}{\mathcal{L}_I}\left(\frac{1}{(\tau_I)_{ext}} + \frac{1}{\tau_{el}}\right) - \frac{4\pi^2 w^2\tau L}{\mathcal{L}_I}\right]^{-1},$$

$$s_{22}^{ext}(w) = \frac{\left(\frac{1}{(\tau_I)_{ext}} + 2\pi wi\right)C_e}{\left(\frac{1}{(\tau_I)_{ext}} + 2\pi wi\right)C_e\left(\frac{1}{\tau_{el}} + 2\pi wi\right)L - \frac{G_{eb}\mathcal{L}_I}{I_{f0}}(-2R_{f0}I_{f0})_{ext}}$$

$$= s_{21}^{ext}(w)I_{f0}\frac{\mathcal{L}_I - 1}{\mathcal{L}_I}(1 + 2\pi wi\tau_I).$$

The equivalent calculation for the internal form results in the solutions:

$$s_{21}^{int}(w) = \frac{-\frac{G_{eb}\mathcal{L}_I}{I_{f0}}}{\left(\frac{1}{(\tau)_{int}} + 2\pi wi\right)C_e\left(\frac{1}{\tau_{el}} + 2\pi wi\right)L - \frac{G_{eb}\mathcal{L}_I}{I_{f0}}(I_{f0}R_s - I_{f0}R_{f0})_{int}}$$

$$= -\frac{1}{I_{f0}}\left[\frac{L\tau}{(\tau)_{int}\tau_{el}\mathcal{L}_I} - \frac{(I_{f0}R_s - I_{f0}R_{f0})_{int}}{I_{f0}}\right.$$

$$\left. + 2\pi wi\frac{L\tau}{\mathcal{L}_I}\left(\frac{1}{(\tau)_{int}} + \frac{1}{\tau_{el}}\right) - \frac{4\pi^2 w^2\tau L}{\mathcal{L}_I}\right]^{-1},$$

$$s_{22}^{int}(w) = \frac{\left(\frac{1}{(\tau)_{int}} + 2\pi wi\right)C_e}{\left(\frac{1}{(\tau)_{int}} + 2\pi wi\right)C_e\left(\frac{1}{\tau_{el}} + 2\pi wi\right)L - \frac{G_{eb}\mathcal{L}_I}{I_{f0}}(I_{f0}R_s - I_{f0}R_{f0})_{int}}$$

$$= -s_{21}^{int}(w)I_{f0}\frac{1}{\mathcal{L}_I}(1 + 2\pi wi\tau).$$

If the superconductor is in its fully normal conducting or superconducting phase, the *int* and *ext* matrices degenerate to the relevant elements:

$$s_{21}^{nc,sc}(w) = 0,$$

$$s_{22}^{nc,sc}(w) = \frac{1}{R_{f0} + R_s + 2\pi wiL}.$$

These results are consistent with Ref. [105], and provide us with the tools to calculate the observable current noise in the readout circuit from fluctuations in the system

46

caused by underlying physical processes. There are six of these processes that jointly make up the largest, explainable, share of the observed noise:

- The thermal fluctuation noise from the coupling of the absorber to the heat bath, called thermal noise or phonon noise, induces a power input in the TES:

$$|\Delta P_e|^2_{ph} = 4k_B T_e^2 G_{eb} \frac{2}{5} \frac{1 - \left(\frac{T_b}{T_{e0}}\right)^5}{1 - \left(\frac{T_b}{T_{e0}}\right)^2},$$

$$|\Delta I_f|^2_{ph} = \left|s_{21}^{int}(w)\right|^2 |\Delta P_e|^2_{ph}.$$

- The electrical Johnson noise in the TES is caused by thermal fluctuations of the electrons in the resistor:

$$|\Delta U|^2_{Jf} = 4k_B T_e R_{f0},$$

$$|\Delta I_f|^2_{Jf} = \left|s_{22}^{int}(w)\right|^2 |\Delta U|^2_{Jf}.$$

- We scale the Johnson noise in the shunt resistors with a factor E_J to account for excess electrical noise. This phenomenon is further discussed in Sec. 2.2.3.

$$|\Delta U|^2_{J_s} = 4k_B T_s R_s E_J,$$

$$|\Delta I_f|^2_{J_s} = \left|s_{22}^{ext}(w)\right|^2 |\Delta U|^2_{J_s}$$

- The noise in the SQUID amplifier is frequency-independent and decoupled from the feedback effects in the readout loop. It is therefore added directly to the output current, with magnitude i_{sq}.

- Additional low-frequency excess noise, called 1/f noise or flicker noise, is observed in TES devices and has a spectral dependency of $1/f^\alpha$. This phenomenon is discussed in Sec. 2.2.3 as well. Its magnitude is described following Ref. [107] with a device- and operation-dependent constant $\left(\frac{\Delta R_{f,flicker}}{R_{f0}}\right)$, to be found empirically.

$$|\Delta P_e|^2_{flicker} = \frac{\left(\frac{\Delta R_{f,flicker}}{R_{f0}}\right)^2 R_{f0}^2}{w^\alpha} I_{f0}^2,$$

$$|\Delta I_f|^2_{flicker} = \left|s_{21}^{int}(w)\right|^2 |\Delta P_e|^2_{flicker}.$$

- Electromagnetic inferences cause additional poles in the observed frequency spectrum. The most prominent contribution is usually the pole from the power supply and its harmonics at 50 Hz, 150 Hz, and 250 Hz, in the continental European standard. We include them in noise power spectra and scale them empirically with constants $p_0(w)$, $p_1(w)$, and $p_2(w)$, respectively.

T_b (mK)	I_b (μA)	T_c (mK)	k (1/mK)	R_{t0} (Ω)	R_s (Ω)	p_0	$\frac{\Delta R_{f,flicker}}{R_{f0}}$	α	i_{sq} (pA/$\sqrt{\text{Hz}}$)
15	3	40	0.05	0.4	0.05	$5 \cdot 10^{-6}$	$5 \cdot 10^{-3}$	1.5	1.2

Table 2.2: Additional physical quantities for the noise simulation, shown in Fig. 2.4. Additional values not included in the table are $p_1 = p_2 = \delta = \delta_H = 0$, $E_J = 1$ and $\eta = 57.8$ (V/μA), where η is the default conversion factor of the CRESST SQUID system, from current flow in the SQUID branch of the readout loop to an observable voltage. The TES curve is modeled as a truncated logistics function with derivative $4kR_{f0}$ at its steepest point. A lowpass filter is applied at 10 kHz. The transition temperature of the iridium-gold bilayer TES is higher than typical for modern tungsten TES. T_c marks the steepest point of the transition and the fully normal conducting state is only reached above 100 mK (see Fig. 2.4).

Because the noise contributions are independent, they can be summed quadratically. The absolute value of the observable current fluctuation in the SQUID branch of the readout circuit is equal to the calculated current fluctuations in the TES branch, as the total current in the loop has to remain constant. We illustrate this noise model by calculating the noise contributions of the detector used in Ref. [103] and the previous section. For this, we need several more physical quantities, which are chosen to be consistent with the operation at 20 mK in Ref. [103], and summarized in Tab. 2.2. Note that transition temperatures are lower in modern devices, and transitions are steeper due to different materials and manufacturing choices. Furthermore, operation points (OPs) are usually chosen higher in the transition. The resulting transition curve, observable voltage pulse and noise, and NPS with its individual contributions are shown in Fig. 2.4. For producing a noise trace from the NPS, we used a method described in Ref. [108], where uniformly sampled phases are added in Fourier space, inducing the random noise structure after transformation back into a time-dependent array.

In the previous calculation, we implicitly assumed all noise to be Gaussian distributed around the equilibrium state. Such a process can be described as an Ornstein-Uhlenbeck (OU) process [109], with characteristic time and damping constants. This theory is again discussed in Sec. 2.2.3. Each OU process has two phases of frequency dependency in its spectral shape: an inverse frequency dependence for high frequencies, and a flat frequency dependency for low frequencies. The cutoff frequency that marks the transition between the phases is given by $1/(2\pi\tau)$, where τ is the intrinsic time constant of the process. We can apply this theory of OU processes to a record window
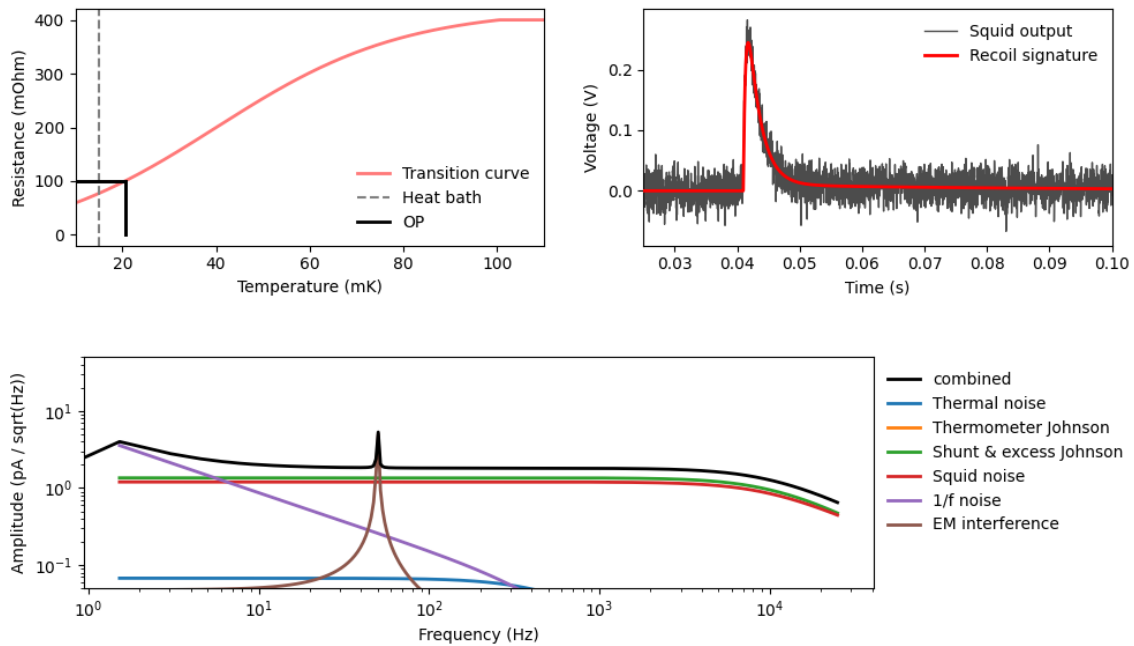
Figure 2.4: Simulated 60 keV X-ray event in a cryogenic detector modeled after Ref. [103]. (top left) The transition curve (red) and the OP within it (black), 5 mK elevated above the heat bath (grey, dashed). (top right) The observable voltage pulse and noise (black) and the part that is attributed to the energy deposition from a particle recoil (red). (bottom) The noise power spectrum (NPS, black) and its individual contributions. The dominating noise contribution in the relevant bandwidth is the Johnson noise internally produced in the TES.

containing a pulse induced from a particle recoil to obtain characteristic cutoff frequencies corresponding to the pulse shape's characteristic time constants. The frequency $f_{c,in} = 1/(2\pi\tau_{in})$ ($f_{c,t} = 1/(2\pi\tau_t)$) is the upper limit of the bandwidth relevant for the athermal (thermal) pulse component. For the modeled detector, $f_{c,in} = 426$ Hz and $f_{c,t} = 3.12$ Hz. The value $f_{c,in}$ is also the cutoff frequency of the thermal noise. The cutoff frequencies for the electronic (flicker) noise are above (below) the observed bandwidth.

Having obtained descriptions of the detector's response to particle recoils and its omnipresent noise, we can quantify the sensitivity of the detector in terms of its signal-to-noise ratio (SNR). This we define as the ratio of the observed pulse (see Eq. 2.22) and the NPS, in Fourier space, integrated over the observable bandwidth:

$$SNR_{opt} = \sqrt{\int_0^\infty \left| \frac{\Delta I_{f,pulse}(w)}{\Delta E \cdot NPS(w)} \right|^2 dw}, \qquad (2.27)$$

$$\Delta I_{f,pulse}(w) = \int_0^\infty \Delta I_{f,pulse}(t) e^{-2\pi wt} dt. \qquad (2.28)$$

This is an optimal estimate in the sense that only the bandwidth where the characteristic frequencies of the pulse shape dominate over the noise is accumulated, as is done in optimum filtering (see Sec. 3). We calculate SNR_{opt} numerically[1], this provides us with a SNR of 0.991 per keV recoil energy for the given detector. Its inverse matches the energy resolution of 1 keV reported in Ref. [103]. This description assumes a constant energy resolution equal to the baseline resolution, which is only valid in a small signal approximation. The relevance and harmfulness of the individual noise contributions are further discussed in Sec. 2.2.2.

2.1.3 Operating conditions in equilibrium

The developed detector response model provides us with reliable estimates of the underlying physical quantities in a neighborhood of the critical temperature. However, strongly extrapolating from this point will violate the temperature-independent heat capacities and coupling strengths assumption. Nevertheless, for using the model at hand as a tool to simulate detector response, we need to make assumptions for several nuisance quantities that are not directly observable in the experiment and for which we are not interested in their exact physical values. This especially concerns the absorber crystal's base temperature and the heating resistor's resistance. We can derive effective values for these quantities, which have only meaning within our model, by considering Eqs. 2.4 and 2.5 in an equilibrium state, and the power inputs from the constant and self-heating:

$$(T_e - T_a)G_{ea} + (T_e - T_b)G_{eb} = R_f I_f^2 + \delta_H R_H I_H^2,$$

$$(T_a - T_e)G_{ea} + (T_a - T_b)G_{ab} = (1 - \delta_H) R_H I_H^2.$$

---

[1]The normalization ζ of the discrete Fourier transform of $\Delta I_{f,n}$:

$$\Delta I_{f,k} = \zeta \sum_{n=0}^{N-1} \Delta I_{f,n} \cdot e^{-\frac{i2\pi}{N}kn},$$

has to be chosen to preserve energy $\zeta = \Delta t$, instead of power $\zeta = \sqrt{\frac{\Delta t}{N}}$, where Δt is the inverse sampling frequency, N is the number of samples in the record window, n is the time index, and k is the frequency index.

This system can be solved to derive the base temperatures of the thermometer and absorber when we assume the heater resistance as known:

$$T_e = \frac{(G_{ea} + G_{eb})\left(G_{eb}T_b + R_f I_f^2 + \delta_H R_H I_H^2\right) + G_{ea}\left(G_{ab}T_b + (1-\delta_H)R_H I_H^2\right)}{(G_{ea} + G_{eb})(G_{ea} + G_{ab}) - G_{ea}^2},$$

$$T_a = \frac{G_{ea}\left(G_{eb}T_b + R_f I_f^2 + \delta_H R_H I_H^2\right) + (G_{ea} + G_{eb})\left(G_{ab}T_b + (1-\delta_H)R_H I_H^2\right)}{(G_{ea} + G_{eb})(G_{ea} + G_{ab}) - G_{ea}^2},$$

or for the absorber temperature and heater resistance, if we assume the thermometer temperature as known:

$$R_H I_H^2 = \frac{(G_{ea} + G_{ab})\left(G_{ea}T_e + G_{eb}(T_e - T_b) - R_f I_f^2\right) - G_{ea}(G_{ea}T_e + G_{ab}T_b)}{\delta_H(G_{ea} + G_{ab}) + (1-\delta_H)G_{ea}},$$

$$T_a = \frac{(1-\delta_H)\left(G_{ea}T_e + G_{eb}(T_e - T_b) - R_f I_f^2\right) + \delta_H(G_{ea}T_e + G_{ab}T_b)}{\delta_H(G_{ea} + G_{ab}) + (1-\delta_H)G_{ea}}.$$

As an artifact of the approximation, the mismatch between phonon and electron temperature in the TES may be overestimated. In the scope of our model, this is not harmful, as Eq. 2.5 is purely linear, and the base temperature of the absorber has no observable effect. However, the constant δ_H can partially absorb the magnitude of this effect.

This section has established a simplified mathematical description of cryogenic detectors with superconducting thermometers. So far, we have avoided any specifications about used materials and their properties, which we will discuss in the following.

2.2 Experimental realization

Many low-threshold experiments use superconducting thermometers, and considerable progress has been made in their manufacturing in the past decades. The operation at the lowest temperatures has advantages and disadvantages: many theoretical models from solid-state physics are more reliable at low temperatures due to the freeze-out of many thermally induced effects. On the other hand, most measurements of material properties are done at higher temperatures. Their extrapolated values to lower temperatures are sometimes subject to large uncertainties. Furthermore, a series of yet unexplained phenomena appear when devices are pushed towards low recoil energy thresholds, which impose limits on their sensitivity in practice. In Sec 2.2.1, we explain which quantities of the cryogenic detector model established in Sec. 2.1

can be calculated from theory and which need to be extracted empirically. We review the current state of knowledge of excess phenomena in Sec. 2.2.3. The implications for designing sensitive devices are studied in Sec. 2.2.2, where we again use our exemplary detector, modeled after Ref. [103]. Finally, we discuss the similarities and differences with other quantum sensors and devices and hypothesize which phenomena are or can become shared technological challenges.

2.2.1 Determination of physical quantities

The heat capacities of metals at low temperatures are dominated by their electronic heat capacity. This can be well approximated with the Sommerfeld free electron gas model [110]:

$$C_e = \frac{\pi^2}{2} n_e k_B \frac{T}{T_F} = \gamma T,$$

(2.29)

where γ is the Sommerfeld constant of the material, n_e is the number of conduction electrons, k_B is the Boltzmann constant and T_F is the Fermi temperature. The detector from Ref. [103] uses a TES made of an iridium-gold bilayer. Modern devices usually use superconductors with lower transition temperatures, e.g. tungsten. The heat capacity of superconductors additionally increases by $\Delta c = 1.43\gamma T_c$, following the Bardeen-Cooper-Schrieffer (BCS) theory, below its transition curve [106]:

$$C_{e,op} = C_e \left(2.43 - 1.43 \frac{R_{op}}{R_n} \right),$$

(2.30)

where R_{op} is the resistance of the superconductor in its OP, and R_n is the normal conducting resistance. The physical constants for several metals often used for components of cryogenic detectors are summarized in Tab. 2.3.

The contribution of its phonons dominates the heat capacity of the non-metallic absorber crystal. At low temperatures, it can be well described with the Debye model [110]:

$$C_{ph} = \frac{12\pi^4}{5} n_a k_B \left(\frac{T}{\Theta_D} \right)^3,$$

(2.31)

where n_a is the number of atoms in the lattice and Θ_D the Debye temperature. Several properties of materials often used as absorber crystals are summarized in Tab. 2.4.

52

Quantity	W	Au	Al	Ir	Cu
T_F (kK)	27 [111]	63.9 [112]	135 [112]	-	8.12 [112]
γ (nJ/mole/K^2)	1.01 [111]	0.729 [112]	1.36 [112]	3.1 [112]	0.7 [112]
G_{ep} (fW/mK^6/mm^3)	21 [113]	$\approx$210 [114]	-	-	-
T_c (K)	0.015 [111]	-	1.14 [112]	0.14 [112]	-

Table 2.3: Fermi temperature T_F and Sommerfeld constant γ for several metals. For tungsten and gold, values of the strength of electron-phonon coupling G_{ep} are shown. References are stated individually. The values vary in the literature depending on calculation or measurement methods. The electron phonon coupling for gold was in Ref. [114] stated as 10 times higher than that for tungsten.

Quantity	CaWO$_4$	Si	LiAlO$_2$	NaI	Al$_2$O$_3$
Density (mg/mm^3)	6.06 [115]	2.33 [115]	2.62 [115]	3.67 [115]	9.22 [115]
Molar mass (g/mole)	287 [116]	28.1 [116]	65.9 [116]	150 [116]	102 [116]
Θ_D (K)	228 [117]	648 [112]	429 [115]	165 [115]	1041 [117]
$\nu_D = k_B\Theta_D/h$ (THz)	4.7	13.5	8.94	3.4	21.7

Table 2.4: Density, molar mass, Deybe frequency ν_D and Debye temperature Θ_D of several materials often used as targets for cryogenic particle detectors. References are stated individually. The values vary in the literature depending on calculation or measurement methods.

The downconversion process of athermal phonons in the absorber crystal was studied in detail in Ref. [103] and is visualized in Fig. 2.5. The initial particle interaction in the crystal created a population of high-energy optical phonons, which downconvert into acoustic phonons on the time scale of $\approx$ ns. Optical phonons do, in current devices, not reach the sensor directly [118]. The resulting acoustic phonons have energies of about half the Debye frequency and continue to decay with a strongly frequency-dependent decay rate. The system reaches an average phonon frequency much lower than the Debye frequency, e.g. a factor 27 lower in silicon, after $\approx 100\mu$s, which stays approximately constant for several ms. The athermal phonon distribution fills the absorber crystal uniformly in less than a μs, depending on its size and sound velocity, which effectively eliminates any dependency of the observed signal on the position of the scattering sight. The athermal phonon distribution thermalizes via surface scattering in the crystal and interaction with the electrons in the thermometer. The effective thermalization time of the phonons, which is observed as either the rise or fast decay time of pulses, is, therefore, a parallelization of these competing processes:
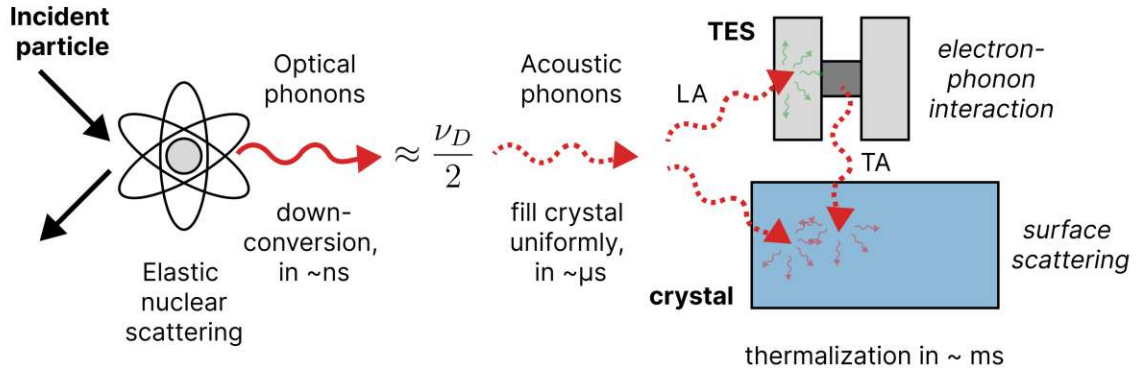
53

Figure 2.5: Schematic visualization of the phonon downconversion process. An incident particle scatters elastically with a nucleus in the crystal. This induces high-energy optical phonons that rapidly break down into acoustic phonons of about half the Debye frequency. A share of the phonons in the crystal can travel to the thermometer film, and longitudenal (LA) phonons interact with the film electrons, while transveral (TA) phonons don't. Therefore a share of the athermal phonons thermalize in the thermometer, and the remainder in the crystal. Fig. based on the description of the process in Ref. [103].

$$\tau_n = \left(\frac{1}{\tau_{\text{film}}} + \frac{1}{\tau_{\text{crystal}}} \right)^{-1}, \tag{2.32}$$

where τ_{film} ($\tau_{crystal}$) is the thermalization time in the thermometer (crystal). The thermalization generally depends on the frequency of phonons. However, we shall follow Ref. [103] and treat them as frequency-independent, which is equivalent to the assumption of a monochromatic phonon distribution. The thermalization time in the crystal increases with its volume and decreases with its Debye temperature (and speed of sound) and with its surface:

$$\tau_{crystal} \propto \frac{V_a}{A_a \Theta_D}. \tag{2.33}$$

Currently, no theoretical values for the absolute scale of $\tau_{crystal}$ exist in the literature. This quantity has to be extracted from measured data. The thermalization time in the thermometer has more complex dependencies:

$$\tau_{\text{film}} = \frac{\tau_0}{\bar{\eta}}, \qquad \qquad \tau_0 = \frac{2V_a}{A \langle v_\perp \alpha \rangle}. \tag{2.34}$$

54

Here τ_0 is the thermalization time in the thermometer in a scenario where all phonons that interact with the superconducting film thermalize immediately. $\bar{\eta}$ is the film's absorption probability of athermal phonons. Ref. [103] measured an approximately linear temperature dependency between $\bar{\eta}$ and the operation temperature:

$$\bar{\eta} \propto T. \tag{2.35}$$

We will use this heuristic dependency in Chap. 4 and Sec. 2.2.2, because it provides a conservative estimate of the change in the athermal phonon collection efficiency ϵ when the operation temperature is lowered. Furthermore, the Hamiltonian for electron-phonon interaction in Ref. [110] also features a linear temperature dependency. However, measurements on individual devices are required for precise statements about the temperature dependency of the athermal phonon absorption probability. It is possible that the measured temperature dependency of $\bar{\eta}$ in Ref. [103] was due to the state of superconductor, especially the amount of the film that was in a normal conducting state, and not due to the operation temperature. The quantity $\langle v_\perp \alpha \rangle$ is the transition probability of phonons from the crystal to the superconducting film, averaged over all phonon modes, wave velocities, and angles. This quantity is proportional to the speed of sound in the crystal and, therefore, the Debye temperature of the crystal material:

$$\langle v_\perp \alpha \rangle \propto \Theta_D. \tag{2.36}$$

The collection efficiency of athermal phonons in the superconducting film can be written as a competition between the thermalization times in crystal and film:

$$\epsilon = \frac{\tau_{\text{crystal}}}{\tau_{\text{crystal}} + \tau_{\text{film}}}. \tag{2.37}$$

The collection efficiency is a crucial quantity for the sensitivity of optimized devices, and scales approximately with the crystal to film surface area ratio. We will analyze this quantity in more detail in Sec. 2.2.2.

The thermal couplings between the phononic and electronic systems in crystal, film, and the heat bath can be described as follows. The thermal coupling G_{ea} between crystal and film must be described as a two-stage process: the thermal coupling between the phononic systems in the components G_K, also called the Kapitza

resistance and the electron-phonon coupling in the metallic film. Since the heat capacity of the film phonons is negligibly small, the total coupling can be described as an in-series connection:

$$G_{ea} = \left(\frac{1}{G_{ep}} + \frac{1}{G_K} \right)^{-1}.$$

(2.38)

This expression is dominated by the smaller of the two quantities. Which of them is smaller depends on the details of individual devices. However, in most standard device designs, the electron-phonon coupling is much weaker than the Kapitza resistance. The ratio can change strongly with the operation temperature, to which both quantities have a higher order dependence:

$$G_{ep} \propto T^5 V_f, \qquad\qquad G_K \propto T^3.$$

(2.39)

The electron-phonon coupling additionally scales with the volume of the superconducting film since the interaction of electrons and phonons is a property of the bulk of the material. Only LA phonons take part in the electron-phonon interaction, which additionally weakens this type of thermal coupling. A reliable theoretical model of the Kapitza resistance between solids exists, e.g. the acoustic and diffusive mismatch model [119]. The thermal coupling from the crystal directly to the heat bath is determined by the technology used to mount the crystal and typically has the temperature dependency of a phononic system. The thermal link between the film and the heat bath has contributions from the applied bonding wire and is, in some modern devices, additionally suppressed through a thin gold film. Its thermal connectivity is subject to the Wiedemann-Franz law [110], which predicts a linear scaling with the temperature and the electrical resistance R_{eb} of the metal:

$$G_{ab} \propto T^3, \qquad\qquad G_{eb} \propto T R_{eb}.$$

(2.40)

The coupling between crystal and bath G_{ab} is hard to model theoretically and is usually purely extracted from measured data.

Further electrical properties that have an impact on the observed pulse are the amplification setting of the SQUID system, which can, for our purposes, be summarized as a conversion factor η between the current in the SQUID branch of the TES readout loop and the observes voltage on the SQUID output. The resistance of TES and shunt

resistor are usually measured at higher temperatures than the operation temperature during the measurement and can therefore be subject to errors. An expert operator sets the currents applied to the heater and TES readout circuit for each device individually.

The parameters p_0, α, i_{sq} and $\frac{\Delta R_{f,flicker}}{R_{f0}}$ of the noise contributions need to be extracted from data of individual devices. No reliable theory for these noise sources exists today. These are not the only phenomena that are still currently under research and not perfectly understood about particle detectors with superconducting thermometers yet. In Sec. 2.2.3, we will look at such unexplained phenomena in more detail. In the following section, we will explore the detector design choices and operation parameters that mainly drive the sensitivity of the device by using the established equations determining the dynamics of the detector response to simulate devices and their energy resolutions.

2.2.2 Optimizing designs for sensitivity

This subsection is dedicated to developing an intuition for design choices of particle detectors with superconducting thermometers and how those relate to the sensitivity of the detector. We start with discussing the two thermal detector modi and follow up with discussing the sizes of components, the strengths of thermal couplings, and the operation temperature. All statements are underlined with results from a dedicated sensitivity simulation based on the theory introduced in the previous sections. We call our lumped-element, electrothermal detector response simulation software "CryoEnv". Finally, we discuss a technology to increase the collection efficiency of athermal phonons by the deposition of additional aluminum films on the crystal.

Cryogenic detectors with superconducting thermometers have two modi of operation, determined by the ratio of the thermalization time of phonons τ_n and relaxation time of the superconducting film τ_{in}. In the mode $\tau_n \gg \tau_{in}$, called the bolometric mode, the film effectively measures the flux of power induced in the film through thermalization of phonons. The rise time of pulses is in this regime given by τ_{in}, and the decay time of the first (athermal) pulse component by τ_n. The PH of the first pulse component can be approximated with:

$$A_n \approx \frac{\epsilon \Delta E}{\tau_n \left(G_{ea} + G_{eb} \right)}.$$

(2.41)

The sensitivity, which can be reasonably approximated with the PH of the athermal pulse component, is therefore mainly dependent on the strength of the thermal couplings and the thermalization time of the phonons. Intuitively, the longer the

energy is kept in the film, the more energy accumulates before the film relaxes, causing a larger temperature pulse. A limit of optimal sensitivity leads to the second mode, the calorimetric mode, where $\tau_n \ll \tau_{in}$, i.e. all athermal phonons are thermalized before the film starts to relax significantly. This effectively means that a bolometer is more sensitive the more it approaches a calorimetric mode of operation. The PH of the athermal component in a calorimetric mode can then be described by

$$A_n \approx \frac{-\epsilon \Delta E}{C_e}. \tag{2.42}$$

The sensitivity is given by the total energy deposition in the film. This total energy crucially depends on the ability to directly collect the initially produced high-energy athermal phonons in the film. For Eqs. 2.41 and 2.42 the additional assumption $C_e \ll C_a$ was used. In practice, intermediate states between bolometric and calorimetric modes exist, where τ_{in} and τ_n have comparable values, such that neither is much larger or smaller. In such a regime, the simplified formulas can fail as a model for the detector sensitivity. For our further studies, it is convenient to define a quantity CAL, that we call "calorimeterness", as the ratio between the relaxation time of the thermometers and the thermalization time of phonons:

$$CAL = \frac{\tau_{in}}{\tau_n} \tag{2.43}$$

We will discuss in the following the impact of the design choices of the cryogenic detector on the expected energy resolution, the calorimeterness, and the collection efficiency. First, we will discuss the different thermal modi in more detail. For this, we refer again to the detector studied in Ref. [103], for which we already implemented a pulse shape and noise simulation shown in Fig. 2.4 and discussed in Sec. 2.1.2. We study this detector again but with variations of its original thermal parameters.

The strengths of the thermal links between components, in relation to the heat capacities, mostly determine the time constants of the system[2]. Fig. 2.6 shows the effect of variations of the thermal link between TES and heat bath G_{eb} in a set of scenarios (see also caption of the figure). The scenarios are chosen to cover all relevant types of detector designs. The sensitivity of the original detector from Ref. [103] is mostly determined by its ability to collect athermal phonons. We, therefore, call it an athermal design. Furthermore, it operates in a bolometric mode. Lowering the thermal link between the thermometer and the heat bath causes a transition to a

---

[2]The self-heating of the TES can also have an observable impact on the relaxation time of the TES.

variable: G_{eb}, G_{ae}, ε; fixed: $C_e = 0.001$ pJ/mK, $C_a = 0.038$ pJ/mK, $G_{ab} = 0.7$ pW/mK (except for blue, solid)

Legend:
- $\varepsilon = 0$, $G_{ae} = 0.045$ pW/mK
- $\varepsilon = 0$, $G_{ae} = 1$ pW/mK
- $\varepsilon = 0$, $G_{ae} = 3$ pW/mK
- $\varepsilon = 0$, $G_{ae} = 0.045$ pW/mK, $G_{ab} = 3$
- $\varepsilon = 0.08$, $G_{ae} = 0.045$ pW/mK
- $\varepsilon = 0.08$, $G_{ae} = 1$ pW/mK
- $\varepsilon = 0.08$, $G_{ae} = 3$ pW/mK
- $\tau_{in} = \tau_n$ (for solid lines)
- G_{ab} (except for blue, solid)
- athermal, bolometric
- athermal, calorimetric
- thermal, absorber calorimeter
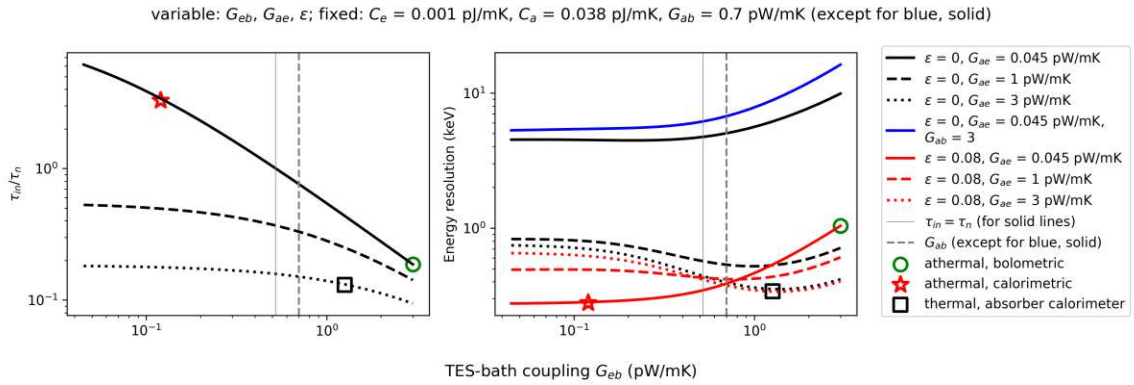
Figure 2.6: Impact of changing G_{eb} on the calorimeterness (left) and the energy resolution (right) for different scenarios: the detector from Ref. [103], in a bolometric mode and with its sensitivity governed by the collection of athermal phonons, is represented by the green circle. The red continuous line depicts the transition to a calorimetric mode (red star) with lowered G_{eb}. The solid black line shows a scenario where the detector sees only thermal phonons ($\epsilon = 0$), which generally decreases the energy resolution. Increasing the thermal coupling between crystal and heat bath G_{ab} causes a loss in energy resolution in this scenario but has less impact in the other scenarios. Increasing the strength of the thermal link between crystal and film (G_{ae}, dashed and dotted lines) has a negative impact on the athermal calorimeter but a positive impact on detectors that benefit from the thermal signal. In the extreme case, the strongly coupled crystal-thermometer system can act as a joint, large calorimeter with low energy resolution (black square). In all scenarios, a stronger G_{ae} decreases the thermometer relaxation time, moving the detector closer to a bolometric mode (left). The quantities on the left side are identical for the different scenarios and, therefore, only shown for the purely thermal case.

calorimetric mode. This continuously improves the energy resolution, underlining the statement that a high calorimeterness is for athermal designs generally preferable over bolometric designs. However, the situation can be different for thermal designs. While for athermal designs, all thermal couplings from the TES to other components should be as weak as possible, for thermal designs, the sensitivity generally increases with a stronger coupling between thermometers and crystal. Furthermore, the combined coupling from the crystal via the thermometer to the heat bath should be stronger than the coupling of the crystal directly to the heat bath, such that the preferred heat decay channel heats the TES. For small crystal dimensions, such thermal designs can reach comparable energy resolutions as athermal designs, since the combined system of crystal and TES jointly acts as a calorimeter. The PH of the dominant thermal component in the calorimeter can then be approximated by

$$A_t \approx \frac{\Delta E}{C_a} \left(\frac{1}{1 + G_{ab} \left(G_{ae}^{-1} + G_{eb}^{-1} \right)} \right) \tag{2.44}$$

We assume here $C_a \gg C_e$ and $\epsilon \approx 0$. However, these designs are more vulnerable to changes in dimensions and temperature. We have identified three distinct scenarios for detector designs, the athermal bolometer, the athermal calorimeter, and the thermal absorber calorimeter. We will discuss their scaling behavior in the following. We can already expect that the athermal designs' scaling properties will be similar since we identified the calorimetric modus as the optimized limit of the bolometric modus already earlier.

Even though this section is mainly concerned with developing a simplified model for superconducting thermometers, the experimental realization of the thermal couplings deserves a brief discussion. The coupling between the crystal and the heat bath directly is mainly determined by the mounting scheme of the crystal. Common mounting schemes are metallic clamps or sticks or insulating sticks or balls. The materials and holder forces can significantly change the strength of the thermal coupling. The coupling between the crystal and TES is, as discussed above, given by the electron-phonon coupling in the TES, and the Kapitza resistance between film and crystal. These can be controlled by adjusting the TES size and the crystal material. Since the electron-phonon coupling scales intensively with the temperature, this coupling is, in practice, mostly controlled by applying a low operation temperature, i.e. transition temperature of the TES. A similar freeze-out of the thermal coupling is not possible for the metallic coupling from the TES to the heat bath. However, a low value for this coupling is desired for optimal sensitivity[3]. The coupling is typically dominated by the gold bonds that connect the TES to the heat sink. However, the bond wire cannot be manufactured arbitrarily thin for its mechanical stability. Therefore, in modern designs, a thin gold film on the crystals builds a bridge between the TES with the bond wire, requiring the heat to pass. The coupling strength can effectively be controlled through the film's thickness and width.

The first interesting dimension to study is the area of the superconducting film. Increasing this area increases the TES' heat capacity, but also its collection efficiency. The situation is visualized in Fig. 2.7. While the TES area has only a minor impact on the energy resolution of the thermal design since it neither significantly increases the heat capacity of the combined system nor is the collection efficiency the dominating

---

[3]A practical lower limit for the coupling strength exists, below which the film does not cool properly in the cryogenic setup.

variable: $\varepsilon(A)$, $\tau_n(A)$, $C_e \propto A$, $G_{ae} \propto A$; fixed: G_{eb}, $C_a = 0.038$ pJ/mK, $G_{ab} = 0.7$ pW/mK, $C_e(A_{ref}) = 0.001$ pW/mK, $\tau_n(A_{ref}) = 2$ ms, $\varepsilon(A_{ref}) = 0.08$
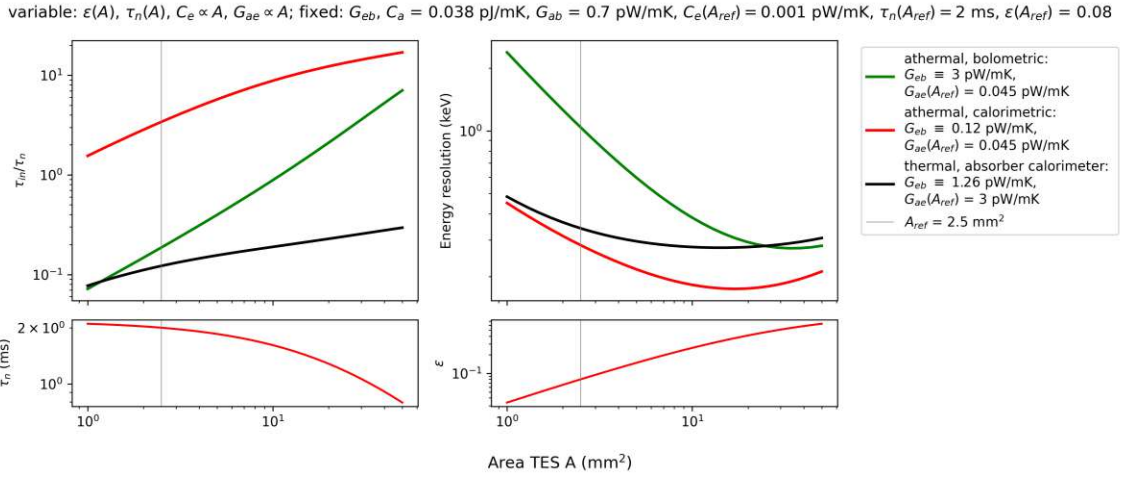
Figure 2.7: Impact of a changing area of the TES on the calorimeterness (upper left), the energy resolution (upper right), the phonon thermalization time (lower left), and the collection efficiency (lower right), for the three scenarios (established in Fig. 2.6) of an athermal bolometer (green), athermal calorimeter (red) and thermal absorber calorimeter (black). The quantities in the lower row are identical for the different scenarios and, therefore, only shown for the athermal calorimetric case. A larger TES area simultaneously results in a larger TES heat capacity. In all scenarios, a broad minimum of the energy resolution, depending on the TES area, exists.

quantity for the sensitivity. For athermal designs, the TES area can have a relevant impact for two reasons: first, the effects of increased heat capacity and collection efficiency cancel each other only approximately, and a theoretical optimum does exist. Second, the increased heat capacity of the TES increases the calorimeterness, which can have a relevant impact if the device is not in a fully calorimetric mode already. We can therefore turn our athermal bolometric design into an athermal calorimeter by increasing the TES area, which causes a decrease in energy resolution by a factor ≈ 5. For the other designs, only a much smaller impact is seen.

A quantity with an even stronger impact on the detector performance is the volume of the absorber crystal. It is especially relevant since rare event searches typically require a certain amount of exposure. A larger target volume and mass would therefore be beneficial for such searches. However, in practice, the larger volume leads to a degradation of the energy resolution. The magnitude of this phenomenon depends again on the detector design. The energy resolution scales approximately $\propto (C_e + C_a)$ for thermal detectors. Since the absorber heat capacity is typically the largest in the system, one would expect the energy resolution to scale linearly with the absorber volume. A different scaling law was first formulated and experimentally confirmed for athermal designs in Ref. [120]. With all other technological factors kept
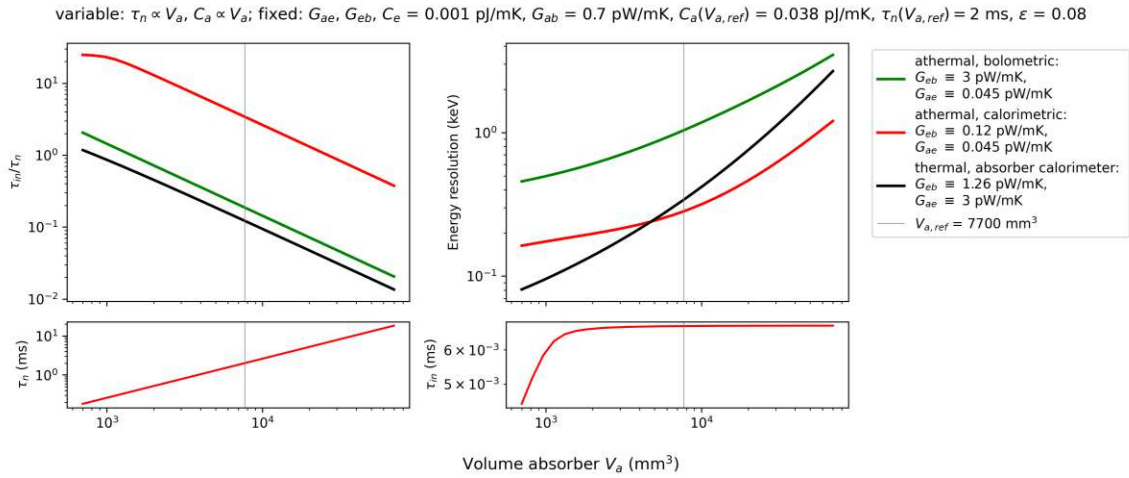
61

変数: variable: $\tau_n \propto V_a$, $C_a \propto V_a$; fixed: G_{ae}, G_{eb}, $C_e = 0.001$ pJ/mK, $G_{ab} = 0.7$ pW/mK, $C_a(V_{a,ref}) = 0.038$ pJ/mK, $\tau_n(V_{a,ref}) = 2$ ms, $\varepsilon = 0.08$

Figure 2.8: Impact of a changing absorber volume on the calorimeterness (upper left), the energy resolution (upper right), the phonon thermalization time (lower left), and the thermometer relaxation time (lower right), for the three scenarios (established in Fig. 2.6) of an athermal bolometer (green), athermal calorimeter (red) and thermal absorber calorimeter (black). The quantities in the lower row are identical for the different scenarios at high crystal volumes and, therefore, only shown for the athermal calorimetric case. A larger crystal results in a higher heat capacity of the crystal and generally degrates the energy resolution, with a more dramatic impact on the thermal design. At a volume around 100 mm^3, the crystal's relaxation time decreases below the thermometer's relaxation time. The time constant τ_{in}, defined with the lowest eigenvalue appearing in the system of ODEs discussed in Sec. 2.1 is above this transition not anymore the relaxation time of the thermometer, but that of the crystal.

constant or equivalent (e.g. the realization of phonon collectors and the thickness of the superconducting film), the collection efficiency ϵ is expected to scale with the ratio of the film and absorber crystal surface $\epsilon \propto \frac{A_f}{A_a}$. Inserting this relationship into Eq. 2.42 leads to an inverse scaling of the energy resolution with the surface of the absorber:

$$\epsilon \frac{\Delta E}{C_e} \propto \frac{A_f}{A_a} \frac{\Delta E}{C_e} \propto \frac{1}{A_c} \propto M^{2/3} \tag{2.45}$$

Here, the fact was used that the heat capacity of the film C_e grows proportionally to its volume, and that for a cubic crystal with fixed geometry, one has $A_c \propto d^2$ and $M \propto d^3$ with d the edge length. This result is intuitively clear: phonons can scatter with the crystal surface and thermalize in the absorber, effectively preventing them from thermalizing in the thermometer. We can see the expected scaling behavior of the thermal and athermal designs in Fig. 2.8.

Not all materials are similarly suitable as targets for cryogenic detectors. While

variable: $\tau_n \propto \theta_D$, $C_a \propto 1/\theta_D^3$; fixed: G_{ae}, G_{eb}, $C_e = 0.001$ pJ/mK, $G_{ab} = 0.7$ pW/mK, $C_a(\theta_{D,ref}) = 0.038$ pJ/mK, $\tau_n(\theta_{D,ref}) = 2$ ms, $\varepsilon = 0.08$
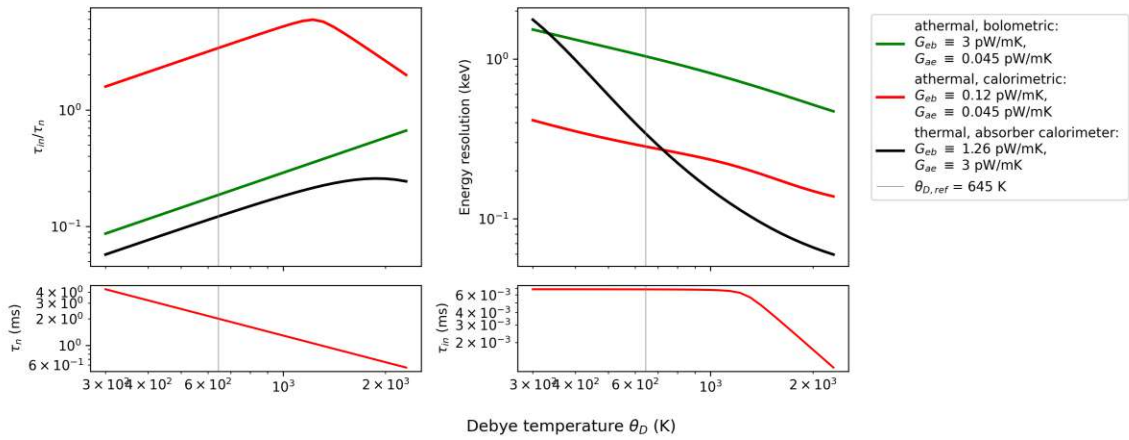
Figure 2.9: Impact of a changing Debye temperature of the crystal on the calorimeter-ness (upper left), the energy resolution (upper right), the phonon thermalization time (lower left), and the thermometer relaxation time (lower right), for the three scenarios (established in Fig. 2.6) of an athermal bolometer (green), athermal calorimeter (red) and thermal absorber calorimeter (black). The quantities in the lower row are identical for the different scenarios at low Debye temperatures and, therefore, only shown for the athermal calorimetric case. A larger Debye temperature results in a lower heat capacity of the crystal and generally improves the energy resolution, with a more dramatic impact on the thermal design. At a Debye temperature slightly above 1000 K (red) and slightly below 2000 K (black), the similar effect of τ_{in} as discussed in the caption of Fig. 2.8 appears.

materials are, in reality, characterized by various properties that affect their suitability as targets, we will only treat one of them in detail, which is also included in our simplified model: the Debye temperature. This quantity strongly changes the heat capacity of the crystal, but also the speed of sound in the crystal, and therefore the thermalization time of phonons and rise time of measured pulses. A faster thermalization time increases the calorimeterness of the TES. We can see the impact of a changing Debye temperature in Fig. 2.9. While an impact is observable for the athermal designs, it is significantly stronger for the thermal design. The cubically lowered heat capacity of the absorber with rising Debye temperature is directly proportional to the energy resolution.

There are two reasons why particle detectors with superconducting thermometers are operated nowadays mostly with tungsten TES, the superconductor with the lowest transition temperature, effectively choosing an as low as possible operation temperature. The first reason is the scaling of the TES and absorber heat capacities with the operation temperature. For a thermal design, the cubically lowered heat
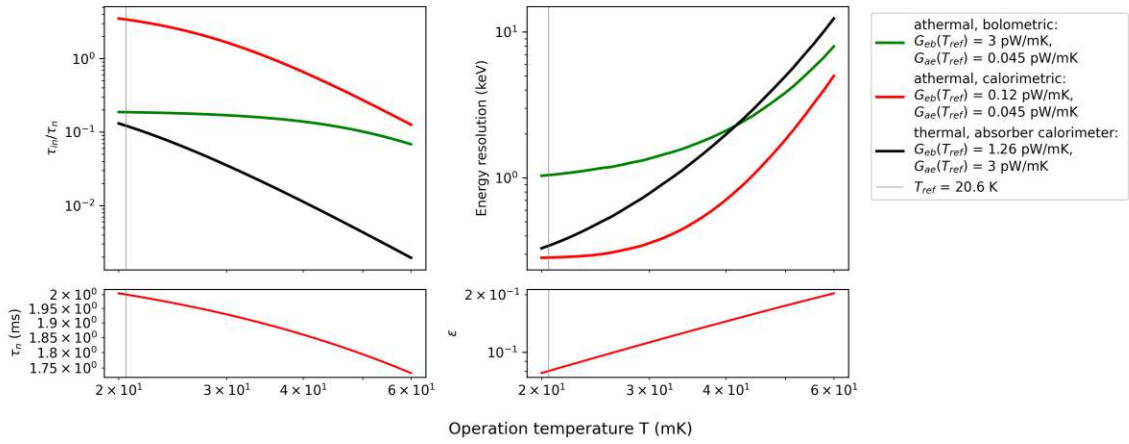
Figure 2.10: Impact of the operation temperature on the calorimeterness (upper left), the energy resolution (upper right), the phonon thermalization time (lower left), and the collection efficiency (lower right), for the three scenarios (established in Fig. 2.6) of an athermal bolometer (green), athermal calorimeter (red) and thermal absorber calorimeter (black). The quantities in the lower row are identical for the different scenarios and, therefore, only shown for the athermal calorimetric case. A higher operation temperature results in higher heat capacities and thermal links and generally degrades the energy resolution, dramatically impacting the thermal design.

capacity of the absorber is directly proportional to the decrease in energy resolution. For an athermal design, the same counts for the linearly lowered heat capacity of the TES. Furthermore, a lowered operation temperature causes a significant decrease in electron-phonon coupling strength, which at the lowest temperatures, effectively prevents the backflow of heat from the film into the absorber. Since this is crucial for athermal calorimetric designs, these lowest temperatures are crucial for the operation of modern, sensitive devices. The effect of a lowered operation temperature is shown in Fig. 2.10. The impact is weakest for the athermal, bolometric design. This detector cannot reach a calorimetric mode due to the strong coupling of its film to the heat bath and does, therefore, not benefit from the freeze-out of the electron-phonon coupling.

The final property that we shall study is the slope of the TES transition curve. The detector in Ref. [103] was manufactured with an iridium-gold bilayer TES, resulting in a transition curve with a width of almost 100 mK. Devices optimized for sensitivity use mostly tungsten TES with much steeper transition curves, often with a width of only ≈ 1 mK. From our simplified detector model, we would expect that the observed PH, which is proportional to the temperature-induced current change in the readout circuit, scales with the applied bias current and the derivative of the transition curve in the OP for small signals. The same behavior is expected for noise that is intrinsic

$\tau_n = 2$ ms, $C_a = 0.038$ pJ/mK, $C_e = 0.001$ pJ/mK, $G_{ab} = 0.7$ pW/mK, $\varepsilon = 0.08$, $G_{eb} = 3$ pW/mK, $G_{ae} = 0.045$ pW/mK
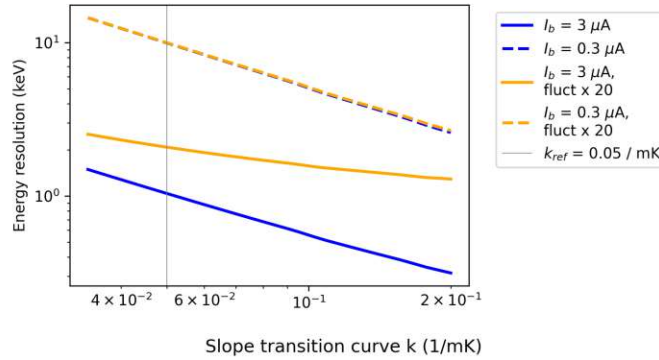
Figure 2.11: Impact of the slope of the transition curve on the energy resolution in scenarios with low (blue) and high (orange) flicker noise, and low (dashed) and high (solid) bias current. A steeper slope improves the resolution unless the TES-intrinsic noise sources dominate over other noise contributions.

to the TES and the magnitude of the ETF loop. For noise that is extrinsic to the TES, as e.g. Johnson noise, or SQUID noise, the noise is expected to be independent of the transition curve of the TES. Therefore we expect from our simplified model that the steepness of the curve positively impacts the energy resolution as long as the extrinsic noise sources are not negligible. This situation is depicted in Fig. 2.11, where we observe that the energy of the detector from Ref. [103] (in athermal bolometric mode) decreases with the slope of the transition curve in high and low bias scenarios, unless in a scenario with high bias current and high intrinsic noise. In this scenario, the intrinsic noise is already dominating over the Johnson noise. We can also observe in Fig. 2.11 that the same statement holds for the bias current: we can expect an improvement of the energy resolution from a higher bias current as long as the TES intrinsic noise sources are not dominating. In a more detailed model, we would expect the transition curve to change with the applied bias current, introducing additional effects.

For athermal designs, the collection efficiency for athermal phonons is a crucial quantity, with which the energy resolution scales linearly (see Eq. 2.42). A technology to increase the collection efficiency without impacting other properties, as e.g. the TES heat capacity, is the application of athermal phonon collectors. Those are aluminum films deposited on top of a part of the tungsten film. Aluminium has a higher transition temperature of 1.2 K (bulk) [121]. The aluminum is already fully superconducting at the device operation temperature of ≈ 15 mK. Due to the proximity effect of superconductors, the covered tungsten film is also pulled into a superconducting

state. Athermal phonons that scatter in the superconducting film break cooper pairs, which travel as Bogoliubov quasiparticles [122] diffusively through the superconducting film until they thermalize in the electronic system of the uncovered tungsten film in transition or in the phononic system of the crystal. This situation is depicted in Fig. 2.12.

The technology of phonon collectors was discussed at length in Ref. [104] under the name quasiparticle-trap ETF assisted TES (QET), where aluminum fins are used as collectors. Nowadays, it is used in most devices, including all CRESST-III designs, where two large area collectors are attached on both sides of the TES. There are several properties that affect the efficiency of a phonon collector. The collection area cannot be increased indefinitely since the broken cooper pairs have a finite lifetime, and due to their diffusive travel, they would not reach the TES from a large distance. Their lifetime was studied in Ref. [123]. Therefore, a too-large phonon collector would act as a parasitic phonon trap that prevents phonons from thermalizing in the TES. Additionally, the target's material can change the phonon collector's efficiency. The typical athermal phonon energy is the Debye energy of the crystal material. The phonon distribution is not monochromatic but a continuum in a more realistic picture. Only phonons that have higher energy than the superconducting band gap of the phonon collecting bilayer can break cooper pairs. A material with a lower Deybe temperature can, therefore, also negatively affect the athermal phonon collection efficiency through a mismatch of the band gap and Debye energy. A comprehensive theory for the optimal size and type of phonon collectors was not yet formulated.

Overall, the sensitivity of devices could also be improved by fostering a more profound understanding of the noise phenomena and phonon non-equilibrium dynamics. Many phenomena remain a matter of ongoing research and raise nontrivial questions. For physics searches, the energy resolution and the mitigation of backgrounds play a crucial role. Instrumental backgrounds that are hard or impossible to discriminate from particle recoils exist. We will discuss in the following such unexplained phenomena in more detail.

2.2.3 Unexplained phenomena

Physics at the lowest temperatures and energies keeps surprising experimentalists with observations of yet unexplored phenomena. Several of such phenomena are present in devices using superconducting thermometers, leading to excess noise and events above known noise and background sources. This especially includes excess phonon, flicker, Johnson, and burst noise. Recently devices have reached recoil energy thresholds
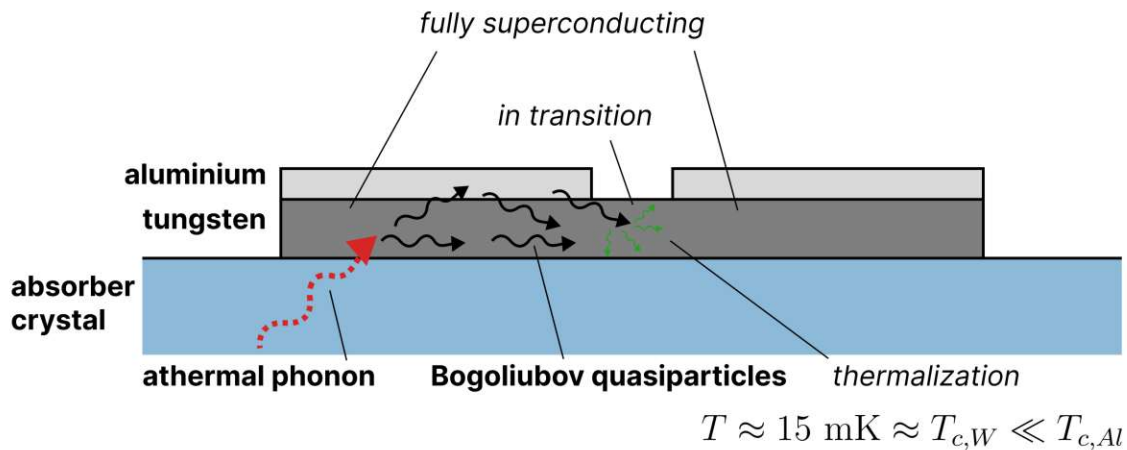
Figure 2.12: Working principle of athermal phonon collectors. The size of the tungsten film (dark grey) is increased, and a large part of it is covered with an aluminum layer (light grey). The higher transition temperature of aluminum pulls the underlying tungsten film into a fully superconducting state due to the proximity effect, while the uncovered tungsten remains in transition, acting as the thermometer. When athermal phonons (red arrow) from the crystal (blue) scatter in the fully superconducting part of the film, broken cooper pair quasiparticles (Bogoliubov quasiparticles, black arrows) are created and travel diffusively until they release their energy in the TES or the crystal. Fig. adapted from Ref. [117].

far below 100 eV, which lead to the observation of a yet unexplained event type: a so-called LEE, with a steeply rising event rate towards lower energies. It is not excluded that multiple of the observations at different energy scales could have similar origins, leading to individual low-energy events or contributing to the omnipresent sensor noise, depending on their rate and magnitude.

2.2.3.1 Intrinsic excess noise

The known and well-studied noise sources of TES devices were summarized in Sec. 2.1.2. Additional noise arises from effects that are not represented in our simplified model of a cryogenic detector. Three of these noise sources were studied in Ref. [124]. We treat the TES as a single thermal component, while in realistic conditions, the film consists of a multitude of microscopic heat capacities connected to each other. The thermal fluctuations between these components can induce internal thermal fluctuation noise (ITFN), observed with a frequency dependency similar to Johnson noise. Potential noise arising from the thermal link between the absorber crystal and the metallic film was neglected as well but might play a role in practice. Furthermore, we neglected the shape change of the transition curve induced by the bias current ($\beta = 0$). Including

them would introduce an additional nonlinear amplification of the Johnson noise. Finally, flux flow noise is a known phenomenon in superconductors arising from the movement of flux quanta across the metallic film. Ref. [124] attributes an inverse frequency dependency to these fluctuations. In most cases, noise from electromagnetic interference would be observed as a pole in the NPS and can usually be treated easily with frequency filters. Also, noise that arises from the digitizers or additional resistances in the bias line at temperatures higher than the operation temperature of the detectors can induce additional Johnson noise. Avoidance of such parasitic components is an incremental part of the experimental setup used to measure cryogenic detectors. Sub-threshold pulses that are indistinguishably intermixed with the noise can induce an additional component in the observed power spectrum, similar to the phonon noise.

Within the setting of our simplified model, we introduced several parameters heuristically to account for excess noise above the modeled noise sources: excess Johnson noise can be accounted for with the parameter E_J that scales the Johnson noise of the reference resistor, and flicker noise with the parameter $\frac{\Delta R_f}{R_{f0}}$. The origin of this additional noise intrinsic to the TES is not fully clarified, nor is it clarified what share of the observed excess noise could be accounted for with a detailed ITFN, nonlinear Johnson, and flux flow noise model. Recently, Ref. [125] showed that all their observed excess Johnson noise can be explained with ITFN. It is to be clarified if this explanation would hold up for all types of devices.

Besides this omnipresent noise, also burst noise (or telegraph noise), in the form of sudden, small jumps between discrete TES resistance levels, was observed. Fig. 2.13 shows this type of noise in a CRESST-III detector. This type of noise is typically connected to certain temperature and bias current points in the operation of the TES. Ref. [105] explains the jumps with transition in ordering the superconducting film's magnetic structure.

The bandwidth relevant for the sensitivity of devices, i.e. for reconstructing pulses, starts at about the frequency associated with the fast decay time constant of the pulse (τ_{in} in calorimeters) and stays constant towards lower energies. The flicker noise is, therefore, often the limiting noise process for the device sensitivity since it typically starts to rise significantly above the flat electrical noise at frequencies around $\approx 10 - 100$ Hz. Depending on the pulse shape, the associated frequency region is already within the region dominated by the steeply rising flicker noise. In the following, we discuss two interesting details about the observed excess flicker noise:
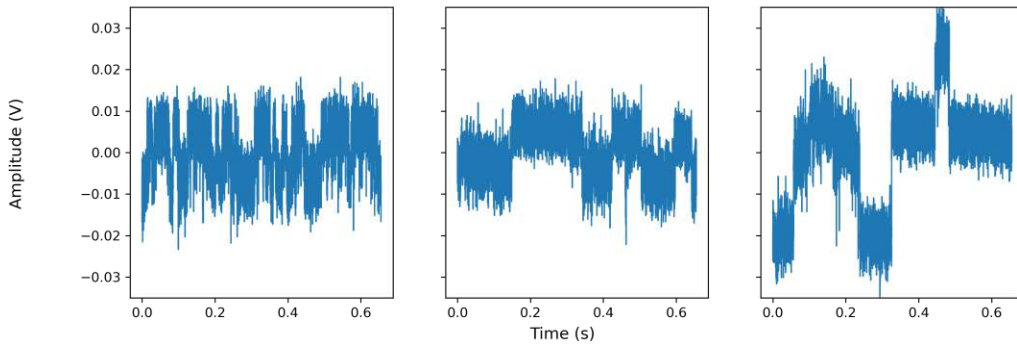
Figure 2.13: Telegraph noise events measured with a CRESST detector in run 36 of the experiment. The frequency and magnitude of the jumps change over time and might be related to the magnetic structure of the superconducting film.

- Every Gaussian noise process can be described as a OU process [109], using the common notation for stochastic calculus:

$$dX(t) = -\frac{1}{\tau}X(t)dt + \sqrt{c}dB(t), \tag{2.46}$$

where τ and c are characteristic time and distance scales of the system, called the relaxation time and diffusion constant, and B is a Brownian motion. The resulting stochastic process converges towards a standard deviation given by $\sqrt{c\tau/2}$ after an initial settling phase. The frequency spectrum of such a process can be written in closed form:

$$S_X(\nu) = \frac{2c\tau^2}{1 + (2\pi\tau\nu)^2}, \nu \geq 0. \tag{2.47}$$

It rises steeply towards lower frequencies, down to a cutoff frequency a $1/(2\pi\tau)$, below which it stays constant. Fig. 2.14 shows an OU process and the associated NPS. The cutoff frequency is the dominant frequency of the underlying process. Following this theory, a lower cutoff frequency of the flicker noise in TES devices must exist, but no systematic studies of this property of the system have been done so far. This cutoff frequency could still contain information about the origin of the noise process since it would determine a time scale of the causing process.

- Flicker noise occurs in many electrical devices and is often attributed to trapped charges. Ref. [126] studies flicker noise in gold-hydrogen junctions at low temperatures. The authors attribute it to electron scattering, which causes a
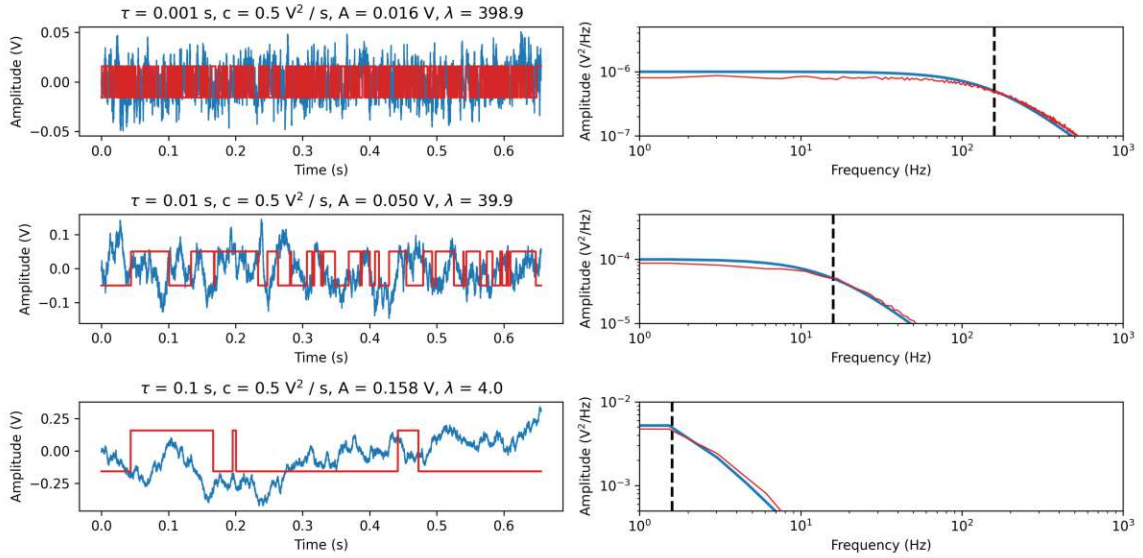
69

Figure 2.14: Comparison between noise traces (left column) and corresponding NPS (right column) following an OU process (blue) and telegraph noise (red). The relaxation time of the OU process τ and the expected number of jumps of the telegraph noise within the record window are set to realize cutoff frequencies in the NPS (black dashed line) of 159 Hz (upper row), 15.9 Hz (middle row) and 1.59 Hz (lowest row). The Euler-Majorana method was used to simulate the OU noise traces. The telegraph noise traces were simulated by uniformly distributing a number of jumps with magnitude A, where the number of jumps is a Poisson random variable with expectation λ.

high-frequent telegraph noise. It is further shown that the NPS of high-frequent, Poisson-distributed telegraph noise is similar to that of an OU process. Fig. 2.14 shows the comparison between noise produced by an OU process and telegraph noise. A potential connection between the observed telegraph and flicker noise in TES devices was not studied in the literature yet but would be compatible with current observations.

In summary, further studies and characterization are needed to fully understand all noise sources in TES devices and the relations between them. From the observation of the frequency spectrum alone, the harmful flicker noise could be explained by high-frequent telegraph noise and flux flow noise. The excess Johnson noise could be related to ITFN.

2.2.3.2 Low energy excess events

In recent years low-threshold experiments have reached energy thresholds far below 1 keV. Measuring these previously unexplored energy regions lead to the discovery of
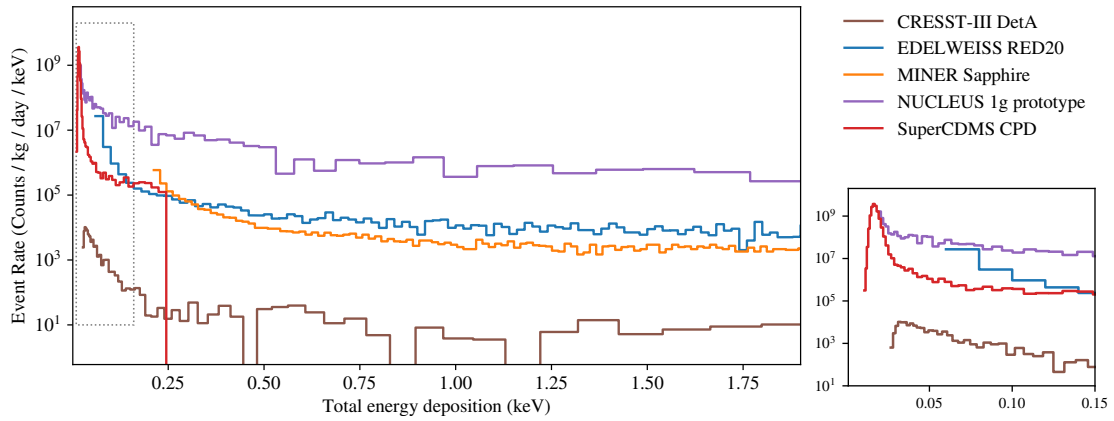
70

Figure 2.15: Energy spectra of measurements with units of total energy deposition. The apparent peaks in the CRESST (SuperCDMS CPD) data at 30 eV (20 eV) are caused by the trigger threshold. Fig. and caption also used in Ref. [7].

new types of backgrounds. Across multiple experiments and sensor types, these LEEs have in common that their spectral shape is steeply rising towards lower energies. These LEEs are a matter of intensive, ongoing research. However, most observations are not fully understood yet. We summarize here the presentations and conclusions from the first iteration of a dedicated workshop that was held in 2021, the EXCESS workshop[4], where we follow the workshop summary paper [7]. Furthermore, we discuss the observations made with TES devices by the CRESST experiment in more detail, which were published in Ref. [31]. Due to variations in the shape, rate, and time dependency of the signal among experiments, detectors, and measurements, a common particle explanation, as e.g. DM scattering, seems unlikely, even in non-standard scenarios. The same conclusion is drawn in Ref. [127], where observations from the EDELWEISS and SuperCDMS experiments were compared with a nuclear recoil form factor fitted to the data.

Ref. [7] reports on observations from the CRESST [30], DAMIC [128], EDELWEISS [129, 130], MINER [131], NUCLEUS [132], SENSEI [100], SuperCMDS [84, 88, 133], NEWS-G [134], CONNIE [135] and RICOCHET [136] collaborations. Collaborations shared the data of their excess observations in Ref. [137]. The data is shown in direct comparison in Figs. 2.15 and 2.16. A rise at low energies is clearly visible in the recoil spectra. The spectra are distributed across the two figures according to their energy units. The NUCLEUS, CRESST, EDELWEISS RED20, MINER,

---

[4]At the time of writing, already a second, third and fourth iteration of the EXCESS workshop were held.
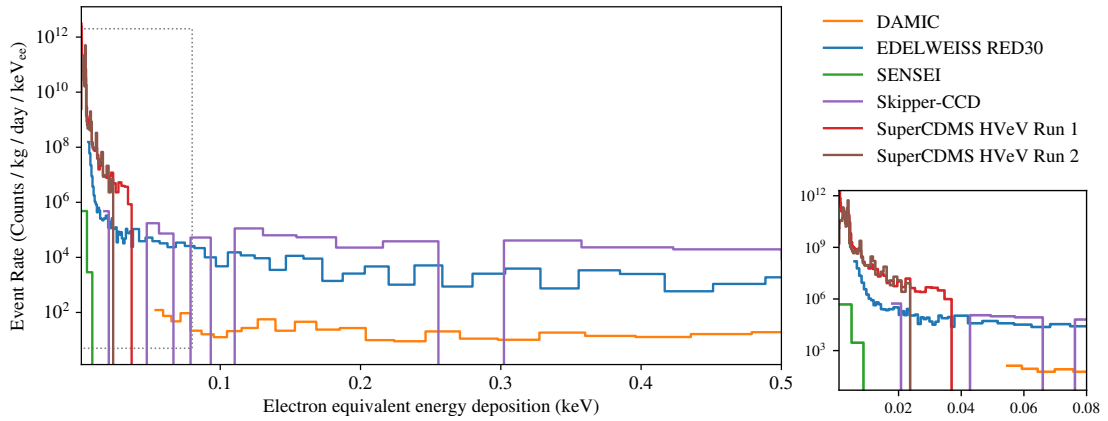
Figure 2.16: Energy spectra of measurements with electron equivalent energy deposition units. Note that this energy scale can only be approximated for SuperCDMS HVeV data (see main text for details). Fig. and caption also used in Ref. [7].

and SuperCDMS-CPD measurements are in units of total energy deposition, while the DAMIC, SuperCDMS-HVeV, Skipper-CCD, EDELWEISS RED30, and SENSEI measurements are in units of electron equivalent energy, i.e. assuming that all incoming particles scattered off electrons in the detector material. The conversion from electron equivalent to nuclear recoil or total energy deposition units can be model dependent. Since the origin of the excesses is possibly diverse and mostly not clarified, the stated recoil energies have to be interpreted carefully. Since the measured events are not discriminable from particle recoil signatures but might, in many cases, not be caused by particle recoils, the energy scales must be understood in terms of a "particle recoil equivalent" energy deposition.

In the following, we provide some key information about the measurements at which LEEs were observed. A detailed description was provided in the workshop presentations [138]:

- The NUCLEUS (NUCLEUS prototype 2017) measurement was performed with a sapphire cryogenic calorimeter equipped with a tungsten TES and aluminum phonon collectors. The operating temperature was 15-20 mK, in an above-ground laboratory at the Max-Planck institute for physics (MPP) in Munich. No veto systems or shielding were in place. An iron-55 source was used for the energy calibration. As a target, 0.49 g sapphire with dimensions of (5 x 5 x 5) mm^3 was held in place with bronze clamps, sapphire spheres, and copper plates. An energy resolution of 3.7 eV and an analysis threshold of 19.7 eV were reported. The measured energy range was 19.7 eV – 12 keV.

72

Measurement	Target	Sensor	Exposure (kg days)	Operation Temperature	Depth (MWE)
CRESST III DetA	23.6 g calcium tungstate	Tungsten TES	5.594	15 mK	3600 (LNGS)
EDELWEISS RED20	33.4 g germanium	NTD	0.033	17 mK	above ground
MINER Sapphire	100 g sapphire	QET	2.72	7 mK	above ground
NUCLEUS 1g prototype	0.49 g sapphire	Tungsten TES	0.0001	15-20 mK	above ground
SuperCDMS CPD	10.6 silicon	QET	0.0099	41.5 mK	above ground
DAMIC	40 g silicon	CCDs	10.927	140 K	6000 (SNO-LAB)
EDELWEISS RED30	33.4 g germanium	NTD, NTL amplification	0.081	20.7 mK	4800 (LSM)
SENSEI	1.926 g silicon	Skipper CCD	0.0955	135 K	225 (Fermilab)
Skipper CCD	0.675 g silicon	Skipper CCD	0.0022	140 K	above ground
SuperCDMS HVeV Run 1	0.93 g silicon	QET, NTL amplification	0.00049	33-36 mK	above ground
SuperCDMS HVeV Run 2	0.93 g silicon	QET, NTL amplification	0.0012	50-52 mK	above ground

Table 2.5: Key properties of the measurements presented at the EXCESS workshop. The first part contains the experiments shown in Fig. 2.15, and the second part corresponds to Fig. 2.16. The spectrum of the experiment in the third part is not shown in this work. Table also used in Ref. [7]. We omitted the NEWS-G experiment, since it is using gaseous targets.

- The CRESST (CRESST-III) measurement was performed with a calcium tungstate cryogenic particle detector equipped with a tungsten TES at LNGS, shielded by 3600 meters water equivalent (MWE). The operating temperature was 15 mK. An active muon veto system and polyethylene, lead, and copper shieldings were in place. A cobalt-57 source, stimulating the tungsten escape peak, and peaks from cosmogenic activation were used for the energy calibration. As a target, 23.6 g calcium tungstate with dimensions of (20 x 20 x 10) mm^3 was

held in place with calcium tungstate sticks. An energy resolution of 4.6 eV and an analysis threshold of 30.1 eV were reported. The measured energy range was 30.1 eV – 16 keV.

- The EDELWEISS RED20 (EDELWEISS-Surf) measurement was performed with a cryogenic germanium detector equipped with a germanium-NTD heat sensor of $(2 \times 2 \times 0.5)$ mm^3 size, glued on germanium. The operating temperature was 17 mK in an above-ground laboratory at IP2I Lyon. 10 cm lead with a 50-degree opening above the detector was in place. An iron-55 source was used for the energy calibration. As a target, 33.4 g germanium with cylindrical dimensions of 50 mm diameter and height 20 mm was held in place with teflon clamps on each side. An energy resolution of 17.7 eV and an analysis threshold of 60 eV were reported. The measured energy range was 60 eV – 10 keV.

- The EDELWEISS RED30 measurement was performed with a cryogenic germanium detector equipped with a germanium-NTD heat sensor $(2 \times 2 \times 0.45)$ mm^3 glued on germanium, together with aluminum electrodes for the ionization measurement. The operating temperature was 20.7 mK, at LSM (below ground, 4800 MWE). 15 cm lead and 60 cm polyethylene acted as shielding. Peaks from germanium-71 neutron activation were used for the energy calibration. As a target, 33.4 g germanium with cylindrical dimensions of 20 mm diameter and 20 mm held was held in place with teflon clamps on each side. An energy resolution of 1.58 eV$_{ee}$ (with 78 V bias, equivalent to 42.7 eV$_{nr}$) and an analysis threshold of 6 eV$_{ee}$ were reported. The measured energy range was up to 30 eV$_{ee}$.

- The MINER measurement was performed with cryogenic athermal phonon detectors. Multiple detectors were operated simultaneously. The sensors used are tungsten TES with aluminum fins. In an above-ground laboratory, the operating temperature was 7 mK as the base temperature, with T_c from 30-90 mK. An inner active veto (germanium detection with TES) was in place. An iron-55 and an external cobalt-57 source were used for the energy calibration. As a target, various masses between 100 and 700 g with germanium, silicon, and sapphire materials were held in place with standard SuperCDMS style mounting. An energy resolution of 15 eV and an analysis threshold of 50 eV were reported.

- The SuperCDMS-CPD measurement was performed with an athermal phonon detector equipped with a single channel of 1031 QETs. The operating temperature was 41.5 mK, in an above-ground laboratory. A minimal shielding of 5 cm

copper was in place. For the energy calibration, a collimated iron-55 source and a 38 um thick aluminum foil for aluminum fluorescence were used. As a target, 10.6 g silicon with 45.6 cm^2 surface area and 1 mm thickness was held in place with six clamps. An energy resolution of 3.86 eV and an analysis threshold of 16.2 eV were reported.

- The SuperCDMS-HVeV measurement was performed with an HVeV (NTL-gain phonon sensor, see Sec. 1.3) detector equipped with a SuperCDMS QET, together with a TES-based phonon readout, using the NTL effect at 100 V to produce quantized charge readout. The operating temperature was 50 mK, in an above-ground laboratory. No veto systems or shieldings were in place. A iron-55 source and a laser were used for the energy calibration. As a target, 0.93 g silicon with dimensions of (10 x 10 x 4) mm^3 was held in place with fiberglass, pressured on the corners of the chip.

- The DAMIC (DAMIC at SNOLAB) measurement was performed with a silicon semiconductor ionization detector and CCD sensors. The operating temperature was 140 K, at SNOLAB (6000 MWE). A 20 cm lead and 50 cm polyethylene shielding were in place. For the energy calibration, muons and a cobalt-60 source on the surface were combined with an in-situ LED calibration. As a target, 7 CCDs of 40 g silicon with dimensions of (62 x 62 x 0.68) mm^3 each were held in place with copper holders. A pixel white noise of 1.3 electrons (6 eV$_{ee}$) and a dark count rate of <0.5 electrons per pixel were reported. The WIMP search range is 50 eV$_{ee}$ - 6 keV$_{ee}$, the maximal pixel value 14 keV$_{ee}$.

- The Skipper-CCD (Skipper CCD with passive lead shield running at surface) measurement was performed using a silicon detector with single-electron charge resolution and a skipper charge-coupled device as a sensor. The operating temperature was 140 K in an above-ground laboratory. A shielding of 2 inches of lead was in place. An absolute calibration was done using charge discretization measurements on the same output data for the energy calibration. As a target, 0.675 g silicon with dimensions of (4.6 x 0.75 x 0.0675) cm^3 was held in place as a package with copper holders. An energy resolution of 0.17 electrons per pixel from readout noise and an analysis threshold of 5 electrons was reported. The measured range was 5 to 2000 electrons.

- The SENSEI measurement was performed using a silicon detector with single-electron charge resolution and a skipper charge-coupled device as a sensor. The

operating temperature was 135 K, in a shallow underground laboratory (MINOS cavern at Fermilab, 225 MWE). A thin non-hermetic lead shield was in place. For the energy calibration, the self-calibrating charge measurement was used. As a target, 1.926 g silicon with dimensions of (9.216 x 1.329 x 0.0675) cm^3 was held in place as a package with copper holders. An energy resolution of 0.14 elementary charges was reported, and independent analysis was performed for 1, 2, 3, and 4 electron events.

In summary, with solid and gaseous target materials, LEEs are observed at various energy scales, from ≈ 10 mK up to > 100 K. Some of the observed phenomena cause ionization, and others induce only crystal lattice movement. The observations are not restricted to individual sensor types. Thermistors and charge-counting devices both see them. The key data of the discussed observations are shown in Tab. 2.5. While Ref. [7] states, that a common origin of the observed excess signals is possible, this is in the light of more recent observations unlikely. The CRESST and SuperCDMS-CPD experiments established in Refs. [31, 139] that their observed LEEs are likely non-ionizing and caused by stress effects in the crystal, sensor or interface between them, while the backgrounds measured with CCD-like sensors are necessarily ionizing, as a pure heat signature would not be measured and therefore go unnoticed. The workshop also contained a presentation about a single-electron background observed by the NEWS-G experiments. However, this background is unlikely to be explained by stress effects due to their gaseous target, hence we omitted a detailed discussion. It is likely a phenomenon on its own, unrelated to the excesses in solid-state detectors. Ref. [81] showed that the excess electron-hole pairs measured in SuperCDMS HVeV are likely due to luminescence from the detector holding structure. Ref. [140] establishes that the excess seen by the DAMIC experiment is compatible with a DM interpretation. However, a DM signal observed by DAMIC would, under standard assumptions, be in tension with limits from multiple other experiments. Overall, there seems to be one common stress-related, non-ionizing instrumental background in cryogenic detectors at the coldest temperatures and possibly one common origin of ionizing backgrounds in CCD devices.

The excesses in the CRESST cryogenic TES-based detectors were studied in more detail in Ref. [31]. The authors report on a measurement campaign with a total duration of 670 days and six detectors with thresholds between 10 and 157 eV. The detectors use different target materials but are otherwise similarly manufactured. Five detectors have the CRESST-III default target size of (20 x 20 x 10) mm^3, while one
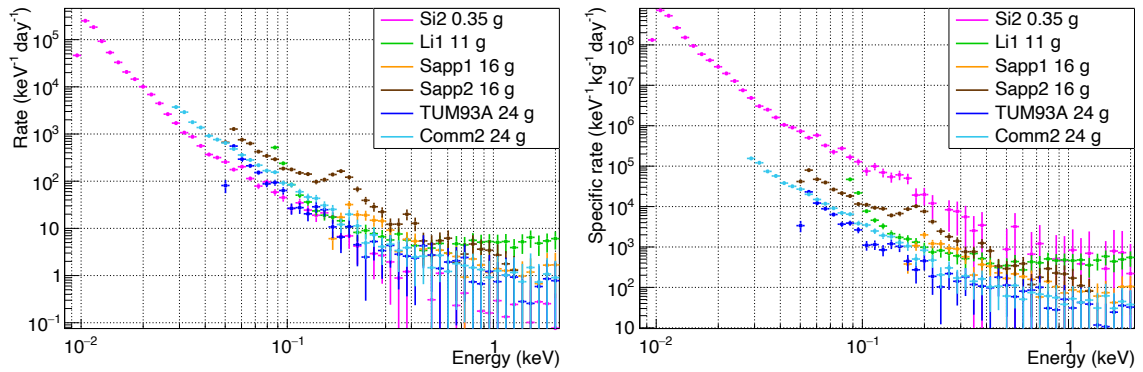
Figure 2.17: Energy spectra measured with six detectors in the CRESST setup in run 36 of the experiment. A strong rise in the event rate is observed toward lower energies in all modules. The excess rates agree better among the detectors when the spectra are scaled to the total measured rate (left) than to the specific rate (right). The origin of the bump in the Sapp2 spectrum (brown) around 200 eV is unknown but likely unrelated to the LEE. Figure taken from Ref. [31].

of the detectors, originally meant as a veto detector for photon scattering, has a wafer target with a size of $(20 \times 20 \times 0.5)$ mm^3. Since all detectors were operated in the same cryostat, the measurement allows for a direct comparison between different types of target materials and target sizes. The observed energy spectra in the energy region of the excess are shown in Fig. 2.17, scaled to units of total and specific rate, i.e. divided by the target mass. Scaling of the rate with the mass of the crystal would be expected for weak particle interactions, as for DM or neutrinos. We see in Fig. 2.17 (right) that the specific rate of the individual measurements varies significantly across the measurements. This is not a surprise since the LEE is likely not caused by particle scattering. The total rate scale shown in Fig. 2.17 (left) provides better comparability between the measurements. However, there are still large variations between the individual excess rates, of about half an order of magnitude. Most notably, the excess rate seen in the wafer detector is not significantly smaller than that of the larger detectors. This clearly excludes a bulk effect as the origin of CRESST's LEE.

Throughout the measurement, an approximately exponential decay of the excess rate is clearly visible, starting from the cooldown of the cryostat, and shown in Fig. 2.18. Several warm-ups of the setup to higher temperatures were performed throughout the measurement campaign, followed by cooldowns to operation temperature to study the effect on the excess rate. An effect is clearly visible for the warm-up to 60 K. The decay of the excess rate was observed on a time scale of (149 ± 40) days before the warm-up test and on a time scale of (18 ± 7) days after this warm-up. An independent effect of the successive two warm-ups to 600 mK and 200 mK is not visible.
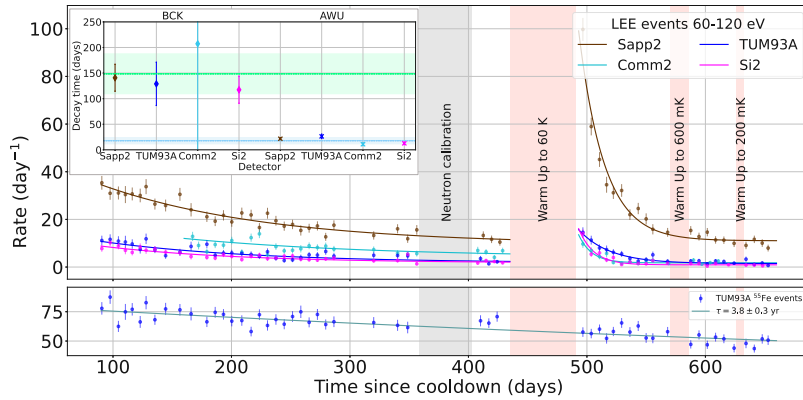
Figure 2.18: Rate of excess events measured with six detectors in the CRESST setup, in run 36 of the experiment, binned by time intervals since the cooldown of the experiment. The rate clearly decays with time, on an approximately exponential time scale, and seems to reset after a warm-up cycle of the experiment to 60 K. An exponential decay model was fitted to the time interval before and after the warm-up independently, and the resulting time scales of the decay are shown (inset). Fig. taken from Ref. [31].

A similar effect of warm-ups on superconducting thin films was discussed in Ref. [141]. The authors have measured quasiparticle-induced charge bursts across a tunnel junction between a superconducting charge island and a normal metal. They measure an overall comparable rate of charge bursts, as seen in the CRESST cryogenic detectors. The mechanism of quasiparticles that break cooper pairs in the charge island is likely the same mechanism used for collecting quasiparticles with aluminum collectors in cryogenic CRESST and SuperCDMS detectors. The measured burst rate decays with time after cooldown and resets after a warm-up and second cooldown. Therefore, the phenomenon's characteristics are compatible with the effect causing the LEE in CRESST detectors.

Understanding low-energy backgrounds has a top priority for the current low-threshold experiments since they coincide with the region of interest (ROI) for low-mass DM searches and CEvNS detection. The LEEs are, therefore, a matter of ongoing research, and it is expected that the coming years will bring insights into their nature. Solving excesses for superconducting thermometers might not only have an impact in the field of rare event searches but also on other superconducting devices used for photon measurements and quantum information processing. We discuss the connections to such devices briefly in the following section.
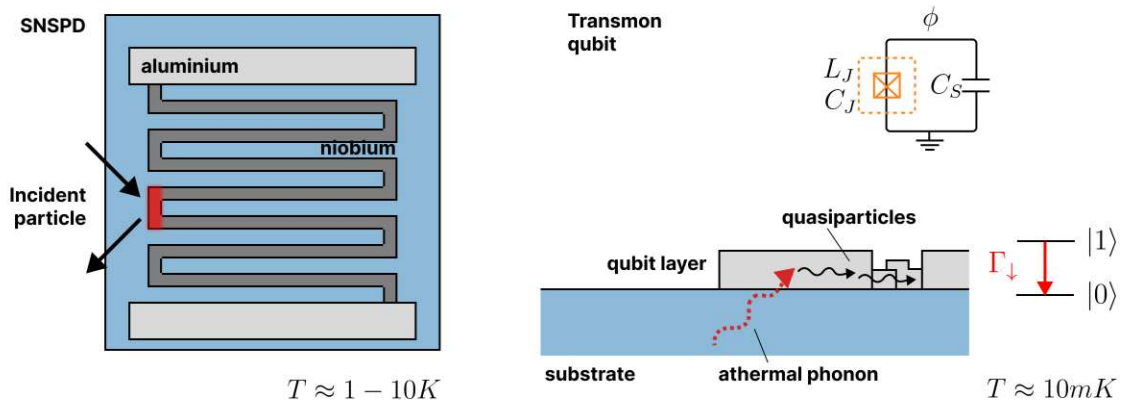
Figure 2.19: (left) Schematic figure of an SNSPD. The superconducting niobium wire (dark grey) is the target for photon scattering (black arrows), which causes the wire to become locally normal conducting (red). (right) Schematic figure of an electrical circuit of a transmon qubit (upper part) and its physical manifestation as a superconducting layer on a substrate (lower part). Athermal phonons can break cooper pairs in the superconducting layer, causing spontaneous state changes of the qubits (inset). Fig. adapted from Refs. [142, 143].

2.2.4 Related quantum technologies

Superconducting thermometers have similarities with other quantum technologies that are based on thin films. This includes other sensor types on the one hand, but also devices used for quantum information processing. In this section, we discuss the similarities with superconducting nanowire single-photon detectors (SNSPDs), briefly introduce metallic magnetic calorimeters (MMCs), and identify common challenges in background mitigation with superconducting qubits.

SNSPDs are particle detectors, similar to cryogenic detectors with TES but with a different use case [144]. Their working principle is visualized in Fig. 2.19 (left). While cryogenic detectors with TES are commonly optimized to measure particle scatterings through a weak interaction, requiring a certain target mass, SNSPDs are optimized to identify EM interactions. Instead of a thin superconducting film, they use a densely packed nanowire, which acts directly as the target for the particle interaction, instead of collecting phonons from the target crystal. The high cross-section for the EM interaction with the metallic wire is sufficient to achieve a high interaction probability. The wire is operated in a superconducting state slightly below its transition temperature. Energy deposited by a particle recoil in the wire heats the wire locally above the transition temperature. The bias heating of the wire causes in addition the surrounding region to transition into a normal conducting state, causing

a measurable resistance increase. The locally small heat capacity of the wire results in a low energy threshold and a fast relaxation time of the sensor. It is noteworthy that TES detectors can be optimized for photon detection by using the sensor as a low energy resolution target (see e.g. Ref. [145]). The idea of coupling a superconducting nanowire to an absorber for DM searches with improved energy threshold by collecting athermal phonons from the crystal was presented in Ref. [146].

Another type of sensors gaining in popularity are MMCs, reviewed e.g. in Ref. [147]. Those sensors measure the change in the magnetic structure of the metal by coupling it with a SQUID amplifier. The obtained energy resolutions with these sensors rapidly improved in the past years and reached today almost identical performance as TES-based detectors. Their strengths include a high dynamic range and fast relaxation time. The idea for a DM search with such devices was presented in Ref. [148] and called the MAGNETO-X experiment. Also, the planned DeLIGHT experiment intends to use them as sensors to measure the adsorption of helium atoms on a crystalline absorber [149].

The last quantum technology that we are discussing here are superconducting qubits. A schematic visualization of these devices and their electrical circuit is depicted in Fig. 2.19 (right). There are many realizations of superconducting qubits. We focus only on one of them, the so-called transmon qubits. A charge island is constructed using a Josephson junction, such that the circuit's lowest and first excited energy states are well separated from higher states, resulting in a two-state system. The technology allows to entangle multiple qubits with each other and can, therefore, be efficiently used as a fundamental building block for quantum computers. In recent years the time of error-free operation, called coherence time, of transmon qubits increased to the order of magnitude of milliseconds. This is required to operate many entangled qubits, which could be used to compute quantum algorithms. Compared to classical computers, a significant computational advantage for certain problems is expected from such operations. Several hundreds of qubits can be operated today, and further advancements are expected within the next decade. A good review of the technology is contained in Ref. [143].

Decoherence is caused by any interaction of the electrical circuit with its surroundings. The more energy deposited in the circuit in the interaction, the higher the error probability. Since the technology is also based on superconducting thin films deposited on a substrate, similar physics processes are expected to cause errors that are also seen in cryogenic detectors. What is seen as noise or pulses in a TES would potentially be observed as errors in qubits. Refs. [142, 150, 151] have shown

that qubits are indeed sensitive to astroparticle scattering. Methods to mitigate the error rate are suggested by operating devices with shielding and underground and implementing phonon-downconversion traps [152]. Ref. [141] studies quasiparticle poisoning, i.e. the existence of high energy phonons that contribute to the error rate, in a chip with a superconducting junction and charge island. The mechanism is likely the same mechanism that is used in aluminum phonon collectors. They measure a quasiparticle-induced error rate that is compatible with the rate of LEE events in TES. Furthermore, their error rate decreases on a time scale of weeks after the cool-down of their setup to Millikelvin, which is as well a common feature with the LEE in TES. This interesting similarity opens up several questions: is the LEE measured with TES a relevant error source for superconducting qubits? Could qubits also be used as sensitive particle sensors since they are sensitive to relatively low-energy interactions? The idea of using them as such was presented in Ref. [153]. Studying these questions is a matter of ongoing research.

We have introduced in this section the experimental realization of cryogenic particle sensors with superconducting thermometers, for which we built a theoretical model in Sec. 2.1. The scaling of relevant properties with temperature and the resulting impact on the energy resolution was studied through detector simulations. We introduced the concept of phonon collectors, a series of yet unexplained noise and excess rate phenomena, and the similarities with other quantum technologies. Cryogenic phonon detectors have not yet reached their fundamental limits. It is therefore expected that their sensitivity will, with advancements in technologies in the future, further improve. A better understanding of the limiting phenomena and their mitigation is required to realize this. Analyzing data from cryogenic detectors for physics searches is a non-trivial process, and Chapter 3 is dedicated to this discussion. We will in the following apply the built understanding of the detectors to realize a DM analysis with CRESST data in Chap. 5, and optimize the design choices for COSINUS detectors in Chap. 4. In Chap. 6, we discuss methods for the large-scale analysis of experiments and for automating the detector operation.

Chapter 3

Analysis techniques

Cryogenic detectors with superconducting thermometers are sensitive low-temperature devices that respond strongly to energy depositions from particle recoils and many other environmental influences. This is especially true when they are pushing for the lowest recoil energy thresholds. Obtaining reliable recoil energy spectra and particle identification requires a careful analysis of the recorded sensor signals, a process that is commonly called RD analysis. In a further step, the measured data have to be compared with the tested physics process and backgrounds to make statistical statements about the occurrence of rare events, e.g. DM scatterings. The canon of such methods is also called high-level data (HLD) analysis. In this chapter, we develop the techniques to perform such analysis. We start with the triggering, pulse shape processing, energy reconstruction, and data quality cuts of the RD analysis in Sec. 3.1. For this, the Cait Python package has recently been released and described in Ref. [5]. The package is dedicated to performing the RD analysis efficiently in the ecosystem of Python scientific computing frameworks, enabling the inclusion of modern data science and machine learning tools. In Sec. 3.2, we discuss typical methods for the derivation of cross-section upper limits on physics processes by showing the typical procedure and effects on toy models. We especially discuss the combination of experimental data from multiple experiments or detectors. This will be relevant for future searches with large-scale setups, as is the planned CRESST upgrade.

3.1 Raw data analysis

The process of RD analysis can roughly be separated into four major steps. When a measurement is done, the first step is to characterize the detector. This includes learning typical pulse shapes, noise conditions, and energy resolution. When fundamental knowledge about the typical event distribution is built, the events can
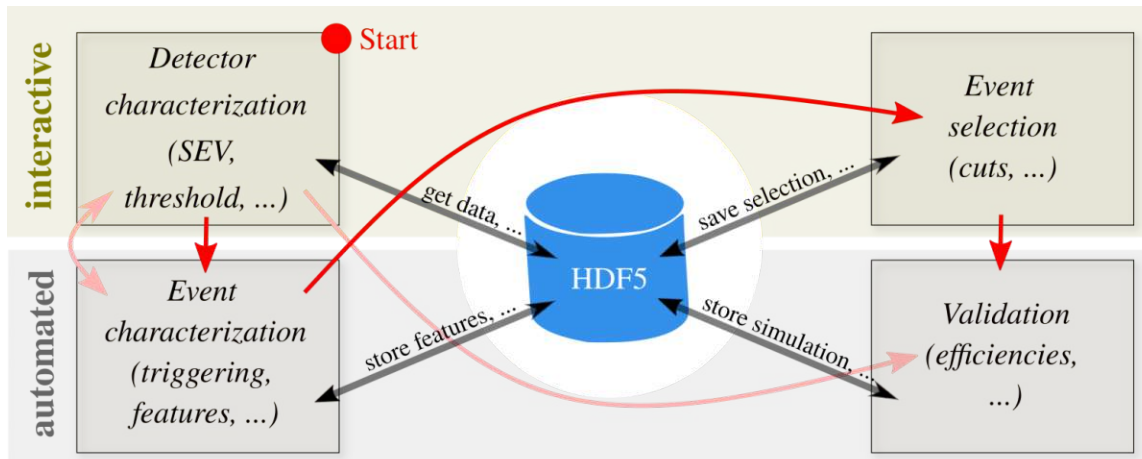
Figure 3.1: Visualization of the four steps of the analysis workflow: detector characterization, event characterization, event selection, and validation. The tasks that require significant run-time are automated, i.e. do not require user interaction. The red arrows indicate the usual order of tasks in the analysis. However, the time-intensive part of the validation (simulating events and calculating their features) can already be done in parallel with the event characterization. Also, the detector and event characterization have to go through at least two iterations: events need to be triggered to characterize the detector, and triggering with an optimized threshold is only possible after the detector characterization. Fig. and caption also used in Ref. [5].

be characterized one by one. This includes e.g. triggering with an optimized filter kernel, and calculation of pulse shape features that already include the knowledge from detector characterization. The calculation of such features can usually be decently automated. Based on this characterization, the selection of events must be decided according to the sought for physics and the data quality. The event selection is done by designing cuts on the data. In the standard workflow this involves to look at samples of the recorded data and decide on rejection and acceptance regions. However, automated procedures are, in principle, possible and will be discussed in this chapter and in Chap. 6. Finally, the analysis chain needs to be validated by using a data set of positive events, often simulated ones. This reduces the risk of a selection bias throughout certain steps of the analysis. The whole analysis procedure is visualized in Fig. 3.1. In the framework of the Cait Python package, the RD analysis is built around an HDF5 file, where the triggered voltage traces and all calculated features are stored. All methods described in this chapter are implemented in the Cait package, and most of them were also described in Ref. [5].

In measurements with cryogenic detectors, electrical heater pulses, called test and control pulses (TPs and CPs), are injected periodically to test the detector response.

The TPs are injected at different amplitudes to monitor the pulse shape and height for different energy depositions. The CPs are injected with maximal amplitude to drive the detector out of its superconducting transition and monitor by that the OP in the transition. Generally, heater pulses have a different shape compared to the pulses induced by particle scattering. Additionally, random noise triggers are forced to acquire empty voltage traces for calculating NPS and using them for event simulations. Typically events are triggered and recorded online, i.e. directly during the measurement, with a simple threshold trigger. In modern experimental setups, the continuous data taken from the sensors is additionally recorded to enable offline triggering in software at any later point with optimized filters and thresholds.

In the following, we discuss the four steps of the RD analysis. Special emphasis is put on the novel machine learning methods that can be used to separate pile-up events (see Sec. 3.1.2.6) and the automated event discrimination (see Sec. 3.1.2). They were for the first time presented in Refs. [5,8]. Furthermore, a method to determine a trigger threshold in the presence of high-frequent sub-threshold events (see Sec. 3.1.1.5) is discussed in detail. This method was introduced in an internal note of the CRESST collaboration [9].

3.1.1   Detector characterization

3.1.1.1   Standard events

The first step in the detector characterization is finding the typical pulse shape of particle recoils in the target, called the standard event (SEV). For this, often, the pulses that were triggered online are used. This event shape deviates for most standard cryogenic detectors only negligibly with the exact position of the recoil in the target, an effect that would be called position-dependence. Since targets are crystalline, and the initial phonon distribution homogeneously fills the crystal much faster than they are transferred to the sensor, no position dependence is usually observed. It would rather be an indication of some type of harmful phonon down-conversion, e.g. on impurities of the crystal structure. The height of the pulse is approximately proportional to the deposited recoil energy. The SEV is constructed by choosing a number of clean particle pulses from a similar energy region, e.g. from a peak in the PH spectrum induced by a calibration line. The pulses are then sample-wise averaged, leading to the reduction of noise fluctuations.

Special care has to be taken that the onset of the pulses in the record window is similar in all averaged traces since strong deviations in the onset would lead to a

smearing of the rise and decay of the SEV. Assuming that the online trigger algorithm works consistently, the onset of pulses with similar heights should be close enough to each other. Depending on the trigger algorithm, a systematic onset drift for pulses with different PHs can be visible, called the trigger walk. Furthermore, it's important that the SEV is built from an energy region where the detector response is still in good approximation linear. For higher energies, non-linear saturation effects are expected. They are discussed in more detail in Sec. 3.1.2. The equation developed in the previous chapter, Eq. 2.13, can be used as a fit model to extract the pulse shape's characteristic time constants and amplitudes. This fit usually leads to different values for the particle and TPs. For TPs, the underlying physics process is different from the process with particle pulses since the initial phonon distribution arriving in the crystal is already more of a thermal nature. It is a common procedure to construct SEVs also for the TPs and in case of multiple pulse shapes in the data, e.g. from recoils in different parts of the detector, for each of them individually. Multiple pulse shapes were seen e.g. due to carrier crystals [154] and holding structures [30].

3.1.1.2   Noise power spectrum

The typical noise conditions should be characterized after the typical pulse shape was obtained. This is done by constructing an NPS. For this, a selection of clean noise traces is made. A Fourier transform (FT) is applied to them, and they are sample-wise averaged in Fourier space. The fluctuating noise frequencies are smoothed out through this procedure, while the characteristic frequencies in the NPS remain at their typical amplitude. The phase information of the events is lost, which is no problem under the assumption that the frequencies are not correlated with each other. In practice, this assumption is certainly not satisfied since harmonics are correlated. For the set of methods we want to use in the further analysis, we would not benefit from the additional phase information and can neglect the correlation. The total scale of the NPS, and its normalization, is of relevance to make meaningful physics statements about the noise conditions and contributions. The proper normalization to retain interesting information is

$$NPS[\omega] = \frac{2}{N \cdot R \cdot F \cdot \xi} \sum_{i=1}^{N} \mathrm{rFFT}(x_i[t])[\omega], \tag{3.1}$$

where rFFT is the fast FT for real values, which contains only the frequencies larger or equal to zero to reduce redundancy. x_i are the noise baselines, N is the number of noise baselines used to compute the NPS, R is the record length, F is the sampling

frequency, t is the index of the record in time-space, and ω is the index of the record in Fourier space. The factor 2 is necessary since we use the rFFT operator instead of the full fast FT. The square brackets indicate that the functions worked with are discretely sampled arrays. The values need then still be converted from a SQUID output voltage to the current in the readout loop. For most use cases in the analysis workflow, the NPS units are unimportant since we only need their shape instead of their absolute values to build e.g. frequency filters. However, to learn information about the physics of the detector, we do need the proper physics units, and we will use them in Chap. 4 and Sec. 6.2. Usually, we build the SEV and NPS from a statistically representative subset of the measured data. If the observation is made that the event shape or noise conditions change significantly within the data set e.g. after interventions on the setup, such as adjustments of the OPs or warm-ups, then new SEVs and NPSs should be constructed for those time periods.

3.1.1.3 Optimum filter

Once we constructed the SEV and the NPS, we have not only learned something of the physics of our detector but can also use them to build an optimized frequency filter for detecting the pulse signature in a noisy data stream. This filter is called the matched or optimum filter (OF) [155], and the filter kernel that maximizes the SNR for Gaussian, colored noise with the frequency distribution of our NPS and a linearly superposed SEV. The OF is built as the ratio of the SEV and NPS in Fourier space:

$$H(\omega) = h\frac{SEV^*(\omega)}{NPS(\omega)}e^{-2i\omega\pi\tau},\tag{3.2}$$

where h is a normalization factor to conserve the PH of the SEV, the asterisk indicates the complex conjugation and τ is a constant between 0 and 1, indicating the position of the maximum in the record window of the SEV. The OF was discussed in more detail in Ref. [156]. Filtering a record window is typically done by transforming it into Fourier space, multiplying the filter kernel, and transforming it back. This methodology can be applied only for records of the fixed length of the noise traces that were used to construct the filter kernel. In Sec. 3.1.2, we discuss the procedure to filter a longer data stream by deconstructing it into partially overlapping records of the desired length. However, if filtering a longer window is necessary, a trick can be applied to extend the filter kernel. For this, the filter can be transformed back into time-space, and the record can be symmetrically padded with zeros, starting in the center of the record. By transforming the filter back into Fourier space, we

find a kernel with a longer shape, i.e., containing lower frequencies but an unspoiled frequency spectrum.

3.1.1.4   Energy resolution

Another essential information about the detector is its energy resolution, often abbreviated with σ. This value describes how well the detector can resolve the energy transfer from an incident particle. The energy resolution is an energy-dependent quantity. However, the most important information, and the one that is routinely extracted and stated for a cryogenic detector, is the baseline resolution, i.e. the energy resolution at zero or small recoil energies. This quantity is especially important since it is typically used to optimize a trigger threshold, as will be discussed later. Several methods exist to estimate this resolution, and in practice, multiple of them are applied to validate if their resulting values agree or obtain systematic uncertainties. Since these estimates usually change systematically beyond their statistical uncertainties and depend on the method used to calculate the estimate, it should be described which method was used to calculate the resolution.

The most widely applied method is to simulate artificial low-energy events by superposing the SEV, scaled to a fixed PH, on an empty noise baseline. These events can then be filtered with the OF, and their PH can be reconstructed. Ideally, the PHs would form an approximately Gaussian distribution around the true PH. The standard deviation of this distribution is an estimate of the energy resolution. Special care must be taken of the empty baselines used for this procedure. In practice, noise traces are typically not empty, but events often appear on the records that coincide with the random triggers. The choice of the baselines can therefore impose a strong negative or positive bias on the calculated energy resolution. In practice, data quality cuts are applied to the noise traces; ideally, these should be the same data quality cuts applied in the analysis of all events. We will discuss data quality cuts in more detail in Sec. 3.1.2. The advantage of this method is that the estimate of the energy resolution is very realistic since the same procedure is applied for the simulated events as for the measured ones. This comes at the price of a higher necessary effort as for alternative methods since a reasonably high number of events must be simulated, e.g. a thousand or more. The statistical uncertainty on the estimate is not straightforward to calculate since it is the standard deviation of a sample standard deviation. However, we can use the assumptions of a normal distribution and large sample size, which leads us to the estimate [157]

$$\text{var}\left(\sigma\right) \approx \frac{2\sigma^2}{n-1}. \tag{3.3}$$

Another method to estimate the energy resolution that does not require simulation of events is to use pulses from a source with known energy. This can be an X-ray source that provides a clean peak in the PH spectrum or induced heater pulses with constant injected amplitude. This method saves the effort of simulating events and reduces the uncertainty that could be induced by potential shifts in the pulse shape or height over time, which could lead to small bias effects in the previously explained method. However, it induces additional uncertainty if not all events from the chosen line in the PH spectrum have the true energy that is assumed for them. Such deviations are often not expected from X-ray or heater pulse lines, but they can occur e.g. when pulses from another background cannot be perfectly discriminated from the pulses with the desired energy. For heater pulses, uncertainties in the electronics or setup could spoil the process, even though this is expected to have a small effect. Typically a higher estimate of the energy resolution is reached with this method. Spectral lines from particle sources feature additionally all potential smearing effects from a position dependence of the recoil, while spectral lines from heater pulses do not show this effect.

The energy resolutions calculated with these two methods will generally have the physical units of a voltage difference since we are calculating the standard deviation of reconstructed PHs. The desired information is the energy resolution in units of the recoil energy. To translate between those units, an energy calibration has to be performed. We will discuss the procedure of the energy calibration in Sec. 3.1.2, and we will see that a linear calibration factor can be used for the conversion of the voltage value of the baseline resolution to energy units.

3.1.1.5   Trigger threshold

We mostly need the energy resolution in voltage units at this point of the analysis workflow to define an optimized trigger threshold. The threshold is typically chosen to balance a high detection efficiency at low energies with the appearance of accidental noise triggers. We often assume a Gaussian distribution of the noise samples, which is a reasonable approximation. The expected number of noise samples above the threshold, and therefore the noise trigger rate, can then be estimated from the tail of the Gaussian distribution of noise samples with mean zero and standard deviation of the energy resolution. We can extract the probability for noise samples to exceed a threshold chosen as a certain multiple of σ from the cumulative distribution function

(CDF) of the normal distribution. In a naive calculation, we can treat all noise samples as independent random numbers occurring with the sampling frequency of the recording. From this, we can estimate the number of noise triggers exceeding the threshold in a given measurement interval. Often the threshold is chosen between 5 and 6.5 σ or as one noise trigger in one kg day exposure. For the later estimate, the target mass is required. However, this procedure can be refined even better since not all noise samples are independent random numbers, but a finite correlation length exists, which is also represented in the nature of the noise as an OU process.

In a recent paper [158] the authors consider this correlation between noise samples and estimate an effective value of the independent noise samples in-situ by constructing a fit model for the maxima in empty noise traces. They propose a method to define a noise trigger rate for sensors with Gaussian noise. For this, they calculate the probability that one sample in a record window has a value of x_{max}, while all others have a lower value, with a binomial distribution:

$$P_d\left(x_{\max}\right) = \frac{d!}{1!(d-1)!}\left(P\left(x_{\max}\right)\right)\left(\int_{-\infty}^{x_{\max}} P(x)dx\right)^{d-1}. \tag{3.4}$$

Here, d is the number of independent samples in the record window, and P is the sample distribution. They assume that all samples follow Gaussian distributions with identical sample variance σ^2 and zero mean, which yields the distribution

$$P_d\left(x_{\max}\right) = \frac{d}{\sqrt{2\cdot\pi\cdot\sigma}}\cdot\left(e^{-\left(\frac{x_{\max}}{\sqrt{2}\sigma}\right)^2}\right)\cdot\left(\frac{1}{2}+\frac{\operatorname{erf}\left(x_{\max}/(\sqrt{2}\sigma)\right)}{2}\right)^{d-1}, \tag{3.5}$$

where σ is the standard deviation of the Gaussian distribution, and erf is the error function. This assumption is well-motivated and common in many signal processing frameworks, e.g. in the assumptions for optimum filtering. The noise trigger rate is then determined with a fit of the above model to the maxima of recorded noise baselines and the choice of a suitable quantile. This fit immediately determines a value for the energy resolution.

While most detector noise is very well modeled by a Gaussian distribution, especially after applying a frequency filter, the assembly of record windows containing only noise is often challenging. A high signal rate in the low energy region or non-trivial pollution might violate the assumption of a purely Gaussian noise sample distribution. A simple adaption to the distribution can be made to account for this fact, which was presented in Ref. [9].

A natural extension of the purely Gaussian, zero-mean model for the sample distribution is the inclusion of different distribution components. For the following

elaboration, we define pollution as a compact, non-oscillating object within a record window. This form of pollution is well motivated for temperature sensors because solutions of the heat equation typically have such a form. Furthermore, the sample distribution of oscillating polluting objects is typically well-modeled with the Gaussian sample distribution and would therefore be covered by the original model already. To include pollution in the model, we assume that the pollution governs only a single independent sample, i.e., the pollution is well described by a template allowing only for time- and scale variations. We extend P_d to the case, where one out of the d sampled follows a different distribution F:

$$P_d(x_{max}) = F(x_{max}) \cdot \mathcal{D}_P(x_{max})^{d-1} \tag{3.6}$$
$$+ \mathcal{D}_F(x_{max}) \cdot (d-1) P(x_{max}) \cdot \mathcal{D}_P(x_{max})^{d-2}. \tag{3.7}$$

Here $\mathcal{D}$ denoted the CDF of P and F respectively. Exemplary, we can assume pollution with the shape of a Gaussian peak. This type of pollution would e.g. occur if a known decay process contributed energy depositions in the temperature sensor with a discrete energy but smeared with an uncertainty. Also, frequently occurring vibrations with fixed amplitude would cause pollution of this kind. The correct model for peak-shaped pollution is then

$$P_d(x_{max}) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma_1} \cdot \left(e^{-\left(\frac{(x_{\max} - \mu_1)}{\sqrt{2}\sigma_1} \right)^2} \right) \cdot \left(\frac{1}{2} + \frac{\mathrm{erf}\left(x_{\max}/(\sqrt{2}\sigma) \right)}{2} \right)^{d-1} \tag{3.8}$$
$$+ \left(\frac{1}{2} + \frac{\mathrm{erf}\left((x_{\max} - \mu_1)/(\sqrt{2}\sigma_1) \right)}{2} \right) \cdot (d-1) \tag{3.9}$$
$$\cdot \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \left(e^{-\left(\frac{x_{\max}}{\sqrt{2}\sigma} \right)^2} \right) \cdot \left(\frac{1}{2} + \frac{\mathrm{erf}\left(x_{\max}/(\sqrt{2}\sigma) \right)}{2} \right)^{d-2}. \tag{3.10}$$

where σ_1 and μ_1 are the standard deviation and mean of the peak-shaped pollution, modeled as Gaussian. The second type of pollution is especially interesting for measurements that observe an exponentially rising low-energetic signal rate, as was discussed in Sec. 2.2.3. To model this signal contribution, we include an exponentially rising pollution term instead of the peak-shaped pollution:

$$P_d(x_{max}) = \lambda e^{-(\lambda x_{max})} \cdot \left(\frac{1}{2} + \frac{\text{erf}\left(x_{\max}/(\sqrt{2}\sigma)\right)}{2} \right)^{d-1} \tag{3.11}$$

$$+ \left(1 - e^{-\lambda x_{max}}\right) \cdot (d-1) \tag{3.12}$$

$$\cdot \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot \left(e^{-\left(\frac{x_{\max}}{\sqrt{2}\sigma}\right)^2} \right) \cdot \left(\frac{1}{2} + \frac{\text{erf}\left(x_{\max}/(\sqrt{2}\sigma)\right)}{2} \right)^{d-2}, \tag{3.13}$$

where λ is the rate parameter of the exponential distribution. For values of x_{max} smaller than zero, this model collapses to the original model for pure Gaussian noise but with only $d-1$ independent samples.

This method was used in Ref. [87], to define a trigger threshold in the presence of high frequent low energy events that were not separable from the empty baselines.

Finally, the last method to estimate an energy resolution is based on simulated events as well and is discussed in more detail in Sec. 3.1.4. It is usually done as a final cross-check of the events used for validating the analysis chain.

3.1.2 Event characterization

3.1.2.1 Trigger algorithm

In many modern low-threshold experiments, the continuous data stream of the sensor signals is recorded in addition to the online triggered events. This is especially useful since an optimized trigger threshold, and software algorithm can be used after data taking, and the whole data stream is available for event simulations for arbitrary time intervals within the measurement period. An optimized trigger algorithm uses the OF to maximize the SNR for particle pulses while decreasing the impact of other events. Since the OF can only be applied in Fourier space and only to record windows containing the number R samples, we need a non-trivial algorithm to perform the filtering that also takes care of any edge effects that can appear during the filtering. This algorithm works as follows:

1. Choose the first R sample from the stream and multiply them with a window function that forces the values to zero at the edges but leaves the inner half of the record unspoiled.

2. Transform the record into Fourier space, multiply if with the OF, and transform it back into time-space.

3. Evaluate the maximum of the filtered record and check if it surpasses the trigger threshold.

4. If the trigger threshold is not surpassed, jump to step 9.

5. If the trigger threshold is surpassed, we call the position of the maximum in that window the triggered sample.

6. Choose a new stream record with the triggered sample at 1/4 of its length and apply the same procedure as above (window function, filtering, evaluate maximum).

7. If the new triggered sample does not sit at 1/4 of the record window, repeat the procedure until the triggered sample is at 1/4 of the record window or a stopping condition is reached. This stopping condition can e.g. be a number of maximal iterations, to avoid a shifting of the window too far away from the original trigger.

8. Add the time stamp of the triggered sample to a list of triggers and store the record window with the trigger sample at 1/4 of its length.

9. Pick the next record half-overlapping, i.e. such that it starts $R/2$ samples after the triggered event, and start from the algorithm from the top.

This algorithm is optimized for a scenario where the stream's baseline is mostly a constant value only disturbed by Gaussian noise conditions, and only discrete, compact events rise significantly above the noise level. Several pitfalls could happen with this algorithm. However, in practice, they rarely have an observable impact: in a scenario with a very high event rate, small events are less likely to be recorded since higher pulses in the same record window would mask them. Pulses that exceed the record window by far can trigger a second time with their tail. These effects make it important to ensure a good efficiency of the trigger algorithm by first choosing record windows of a reasonable size that ideally contain the whole pulse and, second, by checking the consistency of the algorithm with simulated events. Latter practice is described in detail in Sec. 3.1.4. The algorithm works well in practice. However, its computational cost is non-negligible since it is auto-regressive. With this, we mean that the position of the record window on the stream depends on the content of the previous window. The algorithm can, therefore, not be fully parallelized. In practice, a suitable solution to that problem is data parallelism, i.e. splitting the measurement interval into chunks

of several hours, for which the algorithm takes only about a minute to trigger and launch multiple jobs for the triggering in parallel. The potential loss of the last and first record window around splits is typically negligible since the measurements taken are usually about hours, and record windows have lengths of fractions of a second. The same paradigm of data parallelism is applied in the whole event characterization and validation procedure. Parallel computing literature often argues that data parallelism, i.e. performing the same processing pipeline individually for different chunks of data, comes with a higher risk of bottlenecks than task parallelism, where one processing pipeline uses a large number of computing cores through specifically designed software. However, such bottlenecks arise only when common data is accessed by all pipelines. This is not the case for our analysis pipelines, and the proposed procedure is therefore suitable for a large-scale analysis workflow.

Once the list of time stamps of triggers is available, the list needs to be compared with the timestamps of injected test and CPs to identify the artificial pulses caused by the heating resistor and separate them from the particle event candidates. This is done by a proximity cut on the time stamps. In some data acquisition systems, the list of TP time stamps might not be accessible directly. In this case, the voltage values applied to the heating resistor can be continuously recorded and equivalently triggered instead to identify the injected pulses. If this data is unavailable, it is impossible to separate the lists of particle and heater pulses clearly. While this scenario should be avoided, it is still possible to discriminate between particle and heater pulses based on their pulse shape and discrete PHs.

In situations where multiple sensor signals were recorded from one detector, e.g. in a two-channel readout, we usually trigger all of them independently. If one of the channels is defined as dominant, i.e. exclusively used for the recoil energy reconstruction, one could only trigger the dominant channel. However, by doing so, one would potentially miss out on learning the characteristics of the sub-dominant channels, which can contain many interesting learnings about the detector.

3.1.2.2 Pulse height reconstruction

Once the list of triggered events is fixed, we need to identify their PH. The logical and most commonly applied method is to use the difference between the baseline level and the maximum of the filtered record window as PH. This has the advantage that the PHs used in the analysis are consistent with the record window maxima used during triggering. However, there are some situations where additional methods are useful: first, this PH, after applying the OF, always depends on the SEV that was
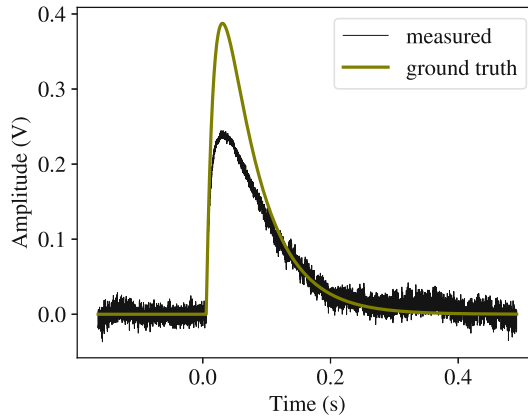
Figure 3.2: Simulated, saturated pulse shape, superposed with readout noise (black). Pulse without saturation effects and readout noise (olive), which is proportional to the true, underlying temperature increase in the thermometer originating from the particle recoil.

used to build the OF. If a measurement contains multiple pulse shapes, the others will therefore be biased towards lower values. Second, PHs scale only linearly with the energy deposition in a small signal approximation. For higher energy depositions, the pulses can reach or come close to the normal conducting region of the superconducting film, which causes non-linear saturation effects on the pulse shape.

A moving average filter can be applied for a scenario in which we want to mitigate the noise to some extent but not spoil the pulse shape too much. This filter assigns to each sample in the record the average of its N/2 - 1 preceding and N/2 successive samples. The filter efficiently mitigates high frequent noise above a wavelength of the moving window size and can be calculated efficiently as a convolution between the record window and a window of the desired filter length. Often a filter with length N=50 samples is used. This filter especially improves the visibility of compact shapes on a record and can be useful to compare PHs between events of different shapes quickly without building OFs for all event shapes. However, the estimate of the PH is certainly not comparable to the quality of the estimate from an OF. Furthermore, for fast pulses that rise and fall within the length of the filter window, the moving average filter will introduce a significant negative bias on the PH estimate.

High-energy events often exceed the approximately linear detector response range. This happens, when the temperature increase in the thermometer is large enough to drive the resistance close to or in the normal conducting range of the film (comp. Fig. 2.1). For such events, any of the above-discussed PH estimators will be negatively

biased since the peak of the pulse is deformed. The situation is depicted in Fig. 3.2. As long as the event is only deformed within the record window, and the voltage values return to the baseline level well within the record window, this bias can be mitigated by the use of a truncated pulse shape fit. In this procedure, the SEV pulse shape is fitted through the minimization of the mean squared error (MSE) between the SEV pulse shape and the record. Typically the free parameters of such a fit are the onset and height of the SEV, and superposed polynomials up to the third degree, to decently model the baseline and its long-range fluctuations. This fit can be truncated by excluding samples from the MSE where the record exceeds a certain voltage value, called the truncation limit. By this, the rigid shape of the SEV will reproduce the height of the unsaturated event while only fitting to the samples that are well within the linear region of the detector response. This fit was discussed in more detail in Ref. [156], and also its potential complications. While it provides a mostly unbiased estimate, which can be controlled for with the test and CPs, the obtained energy resolution is in the saturated region by far worse than that of the OF in the linear region. Furthermore, in scenarios where no high-energy TPs are available to control for bias effects, it is possible that strong bias could be introduced by the absence of ETF in the duration when the pulse is in the normal conducting region of the detector. The absence of ETF can lead to a slower release of the pulse back to the baseline. Another bias effect can be introduced by the temperature dependency of the thermal couplings. Since the couplings scale $\propto T^3$ this effect can lead to a faster release of the pulse back to baseline. The magnitude of these effects varies strongly between detectors. The SEV-fit procedure can also be used to evaluate the PH of unsaturated pulses. However, the computational cost is higher compared to the application of an OF, and it only rarely exceeds the quality of the OF PH estimate, e.g. in situations with unstable noise conditions.

The reconstruction of PHs is especially complicated when a detector has two channels that are worked on in parallel, e.g. a phonon and a light channel. In this case, the decision has to be made if one channel is dominant and will be used exclusively for energy reconstruction, such that the other channel is only used as additional information to veto events or for particle identification. In this case, the PH on the veto channels should be evaluated in correlation with the PH on the dominant channel. This is done by fixing the position at which the value of the record is taken in the subdominant channels, not as its maximum but as the position of the maximum in the dominant channel. While the dominant channel will generally be biased by noise upward fluctuations, the subdominant channel is, with this restriction, statistically

unbiased, even for low energies. This procedure will lead to negative PHs and negative energies if the secondary channel does not record a signal from the event and the record is evaluated at downward fluctuations of the noise. Although negative energies may seem counterintuitive, this is the desired result of the method and the only way to avoid bias in subsequent statistical analysis. It is important to remember that the reconstructed PHs are an estimate of the true physical quantity and not the true quantity itself.

In a situation where neither of the channels should be defined as dominant, but both are supposed to be included in the estimate of the energy, a function of the individual PHs has to be defined that acts as an energy estimator. This can e.g. be a weighted average of the values.

3.1.2.3 Energy calibration

The translation of PHs in estimates of the recoil energy is a crucial part of the RD analysis. Since one rarely knows all physics parameters of the detector precisely enough to derive a transformation that maps voltage values to recoil energies from first principles, it is necessary to do a calibration with a source of known recoil energies. Typically this is done with a radioactive source that produces photons with a high rate and known energy. The photon is then absorbed in the detector target and releases its whole energy in the crystal. By this, a line with a well-defined true energy appears in the PH spectrum. This line can be used to calibrate the energy scale. A naive approach would be to define a linear conversion factor between PH and recoil energies. This is an acceptable solution if the calibration line resides well within the linear region of the detector response (see e.g. Fig. 2.1, for small ΔT) and the measurement conditions and detector response stay stable over the time of the measurement. If these conditions are not met, the detector resolution will significantly degrade for non-zero energy depositions. This can be prevented by finding a better approximation that maps PHs non-linearly and time-dependent to recoil energies. Typically this is done by using the injected heater pulses. These pulses originate from heat inputs in the detector target, and as was discussed in Sec. 2.1, they scale similarly to energy depositions from particle recoils. The TPs can therefore be used to fit a function

$$f : (PH, t) \rightarrow TPA, \tag{3.14}$$

where t is the time stamp. Constructing this function is a two step process: First, for each of the discrete injected TPA values, a time-dependent approximator for the PH is fitted. This approximator is e.g. a linear, polynomial or spline regression. In

the second step, at the time stamp of each event individually, a linear or polynomial approximator is built to map the PH of the event to an equivalent TPA value, called test pulse equivalent (TPE) value. Alternatively, other function approximators (e.g. two-dimensional fits) could be used, as long as the applied procedure is controlled for overfitting and the gaps between the discrete TPA values are continuously filled. Once the TPE values for each event are found, a calibration factor, called the conversion pulse-energy (CPE) factor, can be defined between them and the true recoil energies. This is done by matching the average value of the calibration line TPE_{cal} with the true energy E_{cal} of the source:

$$CPE = \frac{E_{cal}}{TPE_{cal}}. \tag{3.15}$$

Multiplying the TPE values of events with this factor leads to an estimate of the true energies. Since voltage values at different time stamps now lead to different energy estimates, it is not trivial anymore to translate the baseline energy resolution from voltage to energy units. Since we want to state in practice a single value for the energy resolution, we build an average over the measurement for comparability among measurements and detectors. We do this by sampling time stamps uniformly in the measurement interval and building the average of the energy estimate. However, the systematic uncertainty from time-dependent effects is usually small.

3.1.2.4 Pulse shape features

The shape of the voltage trace of an event contains important information about the origin of the event. Several of the often-appearing events are shown in Fig. 3.3, simulated with the Cait package. While particle recoils create pulse shaped traces, not all pulses are necessarily target recoils. When particles recoil in e.g. holding structure or utility crystals, as carrier crystals, or in the TES directly, they induce a pulse with a shape that can vastly differ from the shape of target recoils. Other origins create voltage traces that do not have the shape of a pulse at all, and we call these events artifacts. The typical origins of such events are summarized in Tab. 3.1.

There are several typical features that contain information about the pulse shape and are used in the RD analysis. We will discuss them in the following, one by one. Since analysis chains are in the standard workflow handcrafted by each analyst and for each detector, it is likely that some analysts use features that are not described here.

The offset of a record is the mean value of the first 50 samples. These should usually be fully in the pre-trigger region and therefore provide an estimate of the

Event pulse	A pulse-shaped target recoil (see Fig. 3.3, i). We simulate them with a parametric model from Ref. [103].
Noise	An empty noise trace (see Fig. 3.3, ii).
Decaying baseline	The decaying part of a particle recoil, which happened before the start of the trigger record window.
Temperature rise	A sudden rise in temperature, producing a linear upwards drift of the baseline.
Spike	Glitches in the digitizer can cause a small number of consecutive samples to have a malicious value, effectively producing an upward- or downward-facing spike in the sensor signal (see Fig. 3.3, iii).
SQUID jump	A fast rise of the thermometer temperature, often due to a high-energy particle recoil, can cause the jump from a higher flux quantum state of the SQUID amplifier to a lower one, i.e., the loss of a flux quantum (see Fig. 3.3, iv).
Reset	After a certain number of flux quanta losses, the SQUID amplifier resets to a higher voltage baseline level.
Carrier event	Depending on the design of the cryogenic detector, there might be additional thermal components, other than the target crystal, in which particle recoils can happen. An example is a carrier crystal, a small separate crystal on which the TES is evaporated.
Tail event	A pulse-shaped event with an additional, slowly decaying component, usually caused by a feedback effect, e.g., the reabsorption of scintillation light.
Decaying baseline with event pulse	A decaying baseline event, which is coincident with a particle recoil (see Fig. 3.3, v).
Pile-up	Multiple particle recoils happening inside the same record window, seen as partially overlapping pulse shapes (see Fig. 3.3, vi).
Early or late trigger	In some scenarios, the maximum of the pulse can appear too early or too late inside the record window, e.g., when a time interval with a blocked trigger overlaps with a subsequent event.

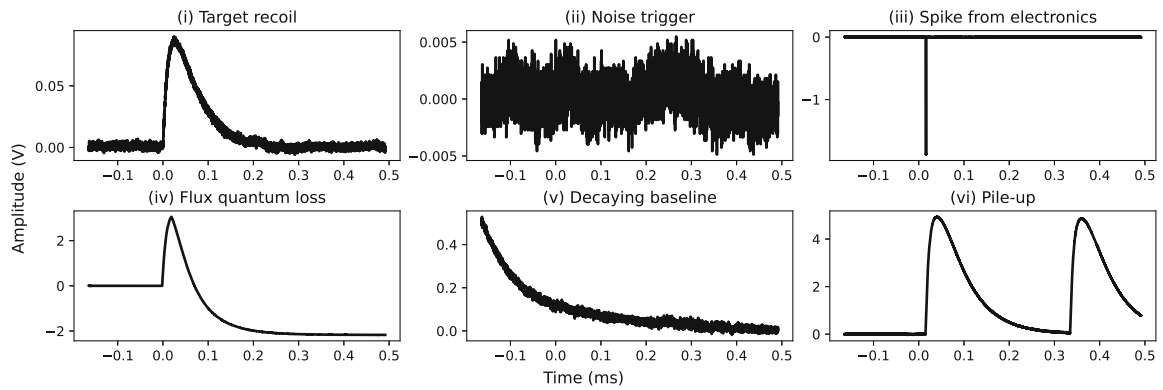Table 3.1: Typically appearing event and artifact classes. Table also used in Ref. [5].

Figure 3.3: Simulated voltage traces corresponding to events. Apart from pulse shaped target recoils (i) and noise triggers (ii), a set of typical artifacts occur frequently in measured data. The data acquisition electronics can introduce spikes or glitches (iii). SQUID readout systems can transition between individual flux quanta, causing discrete jumps in the voltage offset (iv). Pulses can pile up in various ways, fully or partially within the same recording window, leading to strongly decaying voltage baselines (v) or multiple pulse shapes within one window (vi). Fig. also used in Ref. [5].

SQUID amplifier's baseline level corresponding to the current in the SQUID branch of the readout circuit without a particular energy deposition in the detector. This estimate should be close to one value from a discrete set of values corresponding to the SQUID baseline voltage, shifted by a multiple of the voltage of flux quantum. The value can be biased for events triggered on the tail of a large pulse, where the shape of the voltage trace falls throughout the whole record window.

The onset of pulses is calculated by finding the sample with the maximal value in the pulse trace. The last preceding sample that has only 20 % of its height gives the time stamp of the onset. The offset of the record has to be subtracted beforehand. This value should be close to 1/4 of the record window for pulses. It can be far away from this supposed trigger point for artifacts or noise triggers. The rise time of pulses is the time difference between the onset and the maximum of the record. The decay time is the time difference between the maximum and the first subsequent sample that falls below $1/e$ of the maximum. These features are crucial for discriminating different pulse shapes, but they return nonsensical values for voltage traces that are not pulse-shaped.

The baseline difference is calculated as the average of the last 50 samples in the window minus the offset. This feature makes artifacts with rising or falling voltage traces, such as SQUID jumps, easily identifiable.

The sample mean, variance, skewness, and minimum are defined according to their

definition in statistics considering the values in the record window as a sample of random variables (see a suitable statistics textbook, e.g. Ref. [159]).

The maximal and minimal derivatives in the record window are calculated by building the differences between succeeding samples and finding their maximum and minimum. They are a good indication of the presence of spikes induced by electronic artifacts.

The (truncated) SEV fit that we used earlier to reconstruct saturated PHs can also be used to quantify the similarity between the event in question and the SEV. Small fit residuals generally point towards an event with a pulse shape that is similar to the SEV, while large residuals of the fit are typically observed for artifacts. This quantity can serve as a multipurpose discriminator. However, there are several potential pitfalls to consider. The fit residuals are, also for target recoil events, energy-dependent and individually different for pulse shapes. Furthermore, small artifacts are sometimes hard to spot according to their fit residuals, whereas other pulse shape features might be more sensitive.

The pulse shape fit can be taken one step further by fitting not the SEV but the full parametric pulse shape model that was introduced in Eq. 2.13. This fit would automatically produce estimates of the rise and decay times in the pulse shape and does not depend on the formerly produced SEV. The fit error can again be used as an estimate of the similarity between a pulse and the measured event. It is, however, much harder to implement a properly converging fit with this model since most of the parameters in the model induce nonlinear and potentially chaotic changes to the fit function. This method will be studied in more detail in the yet unpublished Ref. [160].

3.1.2.5 Veto yields

Veto channels are secondary channels recorded simultaneously with the target channel to monitor coincidences. These are in standard designs, mostly light channels for scintillating targets or ionization channels for semiconducting targets. Also, secondary TES placed on utilitarian structures surrounding the target, e.g. holders, can be veto channels. The energy deposition in veto channels is often quantified in terms of the strength of the coincident signal in the veto channel as a fraction of the signal strength in the target channel, called a "yield". In this work, we will use the light yield (LY) for detectors with phonon and secondary light channels, which is defined as

$$LY = \frac{E_{l,ee,target}}{E_p} \tag{3.16}$$

where $E_{l,ee,target}$ is the energy measured with the light channel, calibrated to EM interactions in the target crystal. E_p is the energy measured in the phonon channel, which is in a standard analysis workflow also used to estimate the total deposited energy[1]. Since the PH in veto channels is usually reconstructed in coincidence with the primary target channel (see Sec.3.1.2.2), the estimate of the energy deposition in the veto channel, and therefore also the yield value, can be negative.

3.1.2.6 Pile-up separation

In Ref. [8], a special treatment of pile-up events was studied. Since we observe above-ground and calibration runs large amounts of pile-up events due to the high trigger rates, we can get into a situation where discarding all pile-up events might remove a non-negligible share of exposure. On the other hand, all information necessary for an energy reconstruction of two overlapping pulses is contained in the voltage trace. The problem is that the information that two and not just one pulse are sitting on the trace must be included in the method. Ref. [8] approaches the problem with the use of long-short term (LSTM) neural networks (NNs) [162], which are trained to disentangle the two pulse shapes. Their good results are especially interesting since not only the linear region of detector response is addressed, where the two pulses would live in a simple, linear superposition, but also nonlinear saturation effects can be counteracted.

The data used for training for the LSTM model in Ref. [8] consist of simulated pile-up events. They were simulated similarly to the simulation discussed in Sec. 3.1.1.4. Instead of just superposing one pulse shape to a noise trace, a second one was superposed, with an onset that was uniformly sampled between the first onset and the end of the record window. Afterward, the simulated record was input into a generalized logistics function to mimic the detector saturation. The simulated record lengths consisted of 16384 samples. The maximal PH was 0.5 V, and significant saturation started at a value of 0.25 V. A data set of 60k events was simulated. The simulated events were prepared for the training process by normalizing the dataset such that the highest sample is 1 and the lowest 0. Furthermore, the records were downsampled such that they have a record length of 512.

The used model was a 3-layer LSTM with 8 input values, 16 output values, and 200 hidden nodes per layer. The record is input by processing it in 64 time steps.

---

[1]More precise methods for the energy estimation exist. In a readout scheme with phonon and light channel, the total deposited energy splits between the production of phonons and photons. For scenarios with high light output, the energy in the phonon channel should be corrected to account for the produced photons, to build a reasonable estimate of the total energy scale. This phenomenon is called "phonon anti-quenching" and was described in more detail in Ref. [161].
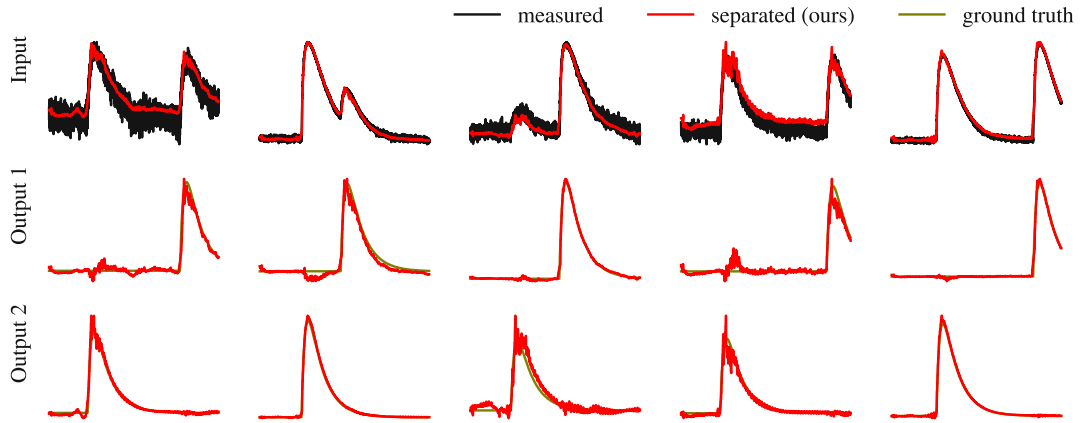
Figure 3.4: The record windows of the recorded particle recoils. The five columns correspond to five independent samples in the data set. The first row shows the time series that are input to the LSTM (black) and the summed outputs after the application of the non-linear transition function (red). We observe a good agreement in most samples, with minor distortions for events with a low SNR. The second and third rows show the two output traces of the LSTM (red), which are mostly in good agreement with the ground truth, plotted underneath (olive). Fig. and caption also used in Ref. [8].

NNs were built in PyTorch and PyTorch Lightning [163, 164]. The record windows containing the pile-up events (Fig. 3.4, black) were the input to the model, and the model was trained to reconstruct the ground truth of the underlying individual pulses (Fig. 3.4, olive).

The data set was split in the ratio 7/1/2 into a training, validation, and test set. The LSTM model was trained on the training set with the ADAM optimizer [165], a learning rate of 10^{-5} and a batch size of 32, for 100 epochs with an MSE loss function. Final loss values obtained were $2.4 \cdot 10^{-4}$ on the test set, and $1.9 \cdot 10^{-4}$ on the training set. Training took on a Tesla P100 GPU with 12 Gig RAM ≈ 0.35 minutes per training epoch and on an Intel Xeon Gold 6138 CPU @ 2.00GHz ≈ 5 minutes per training epoch. Once trained, the model could process more than 30000 events/second during inference on a single-core CPU.

The authors obtained good results in reconstructing the original pulse shapes (see Fig.3.4) and the shape of the simulated PH spectrum (see Fig.3.5). They reason that the impact of the remaining distortions on the pulse traces could be mitigated by using OFs of the resulting separated events, while they only used the maximum value of the record. The resolution of the separated PHs, i.e. the standard deviation of the reconstructed PH of separated events from their ground truth, show energy-dependent
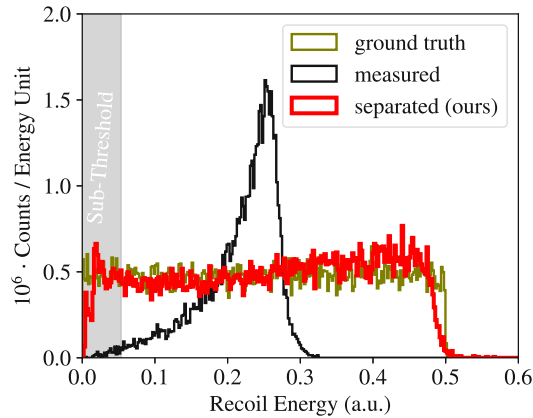
Figure 3.5: The results of the PH reconstruction on an independent test set. (a) The black histogram represents the maximal values of the piled-up and saturated traces. These strongly deviate from the uniformly sampled ground truth of the appearing PHs (two per record window, olive). We observe a much better agreement of the maximal values of the separated pulse traces (red), with only slight deviations for strongly saturated pulse shapes close to the upper end of the spectrum. Fig. and caption also used in Ref. [8].

effects by blowing up for intermediate and very high PHs and having almost as low values as events that were simulated without pile-up for low and high PHs. The resolution is expected to rise towards higher PHs due to the saturation. However, the drop at intermediate-high PHs seems confusing at first sight. The authors attribute it to a statistical effect. The reconstructed PHs at the upper end of the energy spectrum are visibly spoiled toward too-low values. The resulting overdensity at intermediate-high energies leads to a better estimate for pulses with ground truth in that value range, while it leads to worse estimates above and below that range.

The workflow of applying this pile-up separation method is the following:

1. Characterize the detector, extract the typical pulse shape, and model the saturation conditions with a generalized logistics curve.

2. Simulate a suitable data set of pile-up events.

3. Train the LSTM model to separate the events.

4. Apply the LSTM model to the whole data set. Events that are not pile-up events are negligibly affected by the model. The higher number of resulting sub-threshold events can be cut with quality cuts (see Sec. 3.1.3.2).
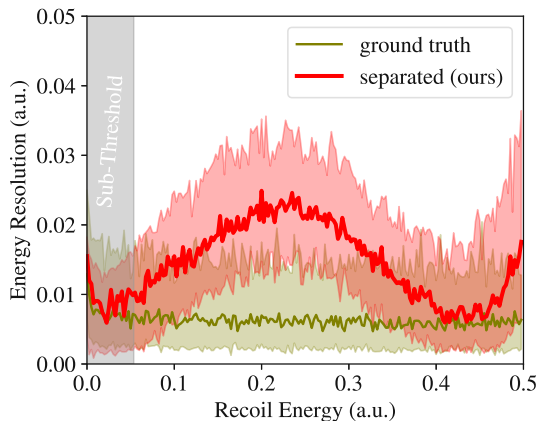
Figure 3.6: (b) The energy resolution of pulses without pile-up and saturation does not depend on the recoil energy but only on the noise conditions (olive). The reconstructed pulses' energy resolution depends on the energy, with local maxima at 0.25 and 0.5 and local minima at 0 and 0.45. This peculiar shape is discussed in the text. Fig. and caption also used in Ref. [8].

This method was not yet used in a production analysis. However, since the model performance was suitable for practical applications, we will mention this method again in Sec. 4, where we will identify a suitable application for it the upcoming COSINUS measurements. This method is currently the only method that can reconstruct pile-up events in the nonlinear region of the detector response of cryogenic TES-based detectors. It is likely that the method could also be applied to other types of detectors that record pulse traces as sensor signals.

In this section, we have introduced the event characterization process with special emphasis on our machine learning method that separates pile-up events. We continue with the next step in the RD analysis, the event selection, where we will discuss the criteria to be applied in a statistically sound analysis of a measurement.

3.1.3   Event selection

3.1.3.1   Data cuts

From all triggered and recorded events we need to identify those that correspond to target recoils, and for which we can reconstruct the recoil energy reliably. The goal of the event selection is to construct a data set of such events by applying cuts based on the data quality and the features of the triggered event. Here we discuss the former ones, and they especially include stability and rate cuts.

The stability cut is to exclude time intervals when the detector was not in stable operation, i.e. where its OP deviated. This can happen due to any environmental influence, such as vibrations or fluctuations in the cryostat's base temperature. The OP is monitored through the injected CPs (heater pulses with maximal TPA) that drive the TES out of its superconducting transition. The pulses therefore saturate at a certain PH. If the detector is consistently in the same OP, this maximal PH value should always be constant within the uncertainty of the detector resolution. If the OP fluctuates, these fluctuations are additionally visible in the measured CP PHs over time. The strategy to exclude unstable time periods is to calculate the mean and standard deviation of the CP PHs over time and exclude all regions where the CP PHs are not within a certain number of standard deviations of their mean value. For the exact procedure, it is important to consider several useful tricks that avoid cutting or leaving too many time intervals due to the possibly significantly non-Gaussian distribution of the CP PHs. First, before the mean and sigma values of the CP PHs are calculated, a rough cut should be applied to them. This can be done by plotting the distribution and cutting strong outliers by eye. Also, quantile-based statistics can be used, by e.g. flagging the lowest and highest several percent of the CPs as unstable. The exact values of these cuts need to be finetuned to the measurement conditions. The second useful choice is to ignore individual outliers. This means that all CPs outside the desired confidence region are flagged as unstable, but only events that are within two unstable CPs are excluded, while individual unstable CPs have no effect on the event exclusion. This is an allowed cut on the data, since real fluctuations in the operation conditions, as in the overall temperature, would happen on longer time scales than the spacing between two CPs. Individual outliers are therefore very likely due to e.g. pile-ups or other artefacts that affect one individual events.

The rate cut is to identify regions with changed noise conditions that can lead to highly frequent triggering of noise upward fluctuations. Such changes in the noise conditions can happen because of external magnetic fields or mechanical vibrations. We calculate the number of non-heater-induced event triggers for the rate cut and bin it into intervals of 10 minutes elapsed real time. Then we build again the mean and standard deviation and exclude outliers of the distribution. In this case we exclude only the bin that is an outlier, but leave the bins before and after. Here again, we need to exclude far outliers before we calculate the mean and standard deviation, to avoid spoiling the calculated values of the distribution. This can be done similarly as it was done in the stability cut.

3.1.3.2 Quality cuts

Within stable measurement periods, not every triggered event is a useful particle pulse, but many of the recorded events are artifacts that need to be cut from the data set (see also Fig. 3.3). Such cuts are then typically done based on the pulse shape in a procedure called pulse shape discrimination (PSD). Every reasonably justifiable cut can be done on the pulse shape, even if it takes out potential signal events, as long as the effect of the cut on the expected signal is taken into consideration (see Sec. 3.1.4 for details). Cuts can be applied on any of the pulse shape features that were discussed in Sec. 3.1.2.4, as well as on the veto yields discussed in Sec. 3.1.2.5. Often cuts are performed as one dimensional acceptance criteria on individual features. It is also possible to construct two-dimensional criteria to e.g. accept only events within a box of the parameter space. It is useful to have interactive tools available to define such cuts, such as the VizTool that comes with the Cait package and is shown in Fig. 3.7. The VizTool has a Plotly backend and allows to scatter events in two feature dimensions, zoom and select in this space. However, only cuts should be defined that can be automatically applied later in the validation step. Often this means they should be writable as a logical expression, which disfavors the use of a lasso tool to select and cut events, despite its availability in Cait.

The data quality cuts should generally be designed with the goal of discarding artifacts and keeping particle recoils in the target crystal. However, there are some individual choices that can be made depending on the goal of the analysis. Ideally, an analysis leads to a multi-purpose data set, with which all physics searches that can be done with the detector at hand are possible. The data selection in the RD analysis should be a mere data cleaning, and it is also sometimes called that. In practice, choices are often made to tune cuts more towards a certain recoil energy ROI, e.g. the low energy region for low mass DM searches. Cuts that are intended for PD, usually based on veto yields, have a special purpose in this sense since they, per definition, will discard target recoils if they do not satisfy the desired properties, e.g. EM recoils. Therefore cuts on the veto yields intended for PD are often kept for the high-level analysis to avoid restricting the feasible physics searches already in the data selection (see Sec. 3.2). The separation between PD and data cleaning is easier for our TES-based cryogenic detectors since PD is not feasible through PSD, as EM events and nuclear recoils have the same pulse shape, and the features can therefore be separated into two distinct sets according to their purpose for PD or data cleaning.
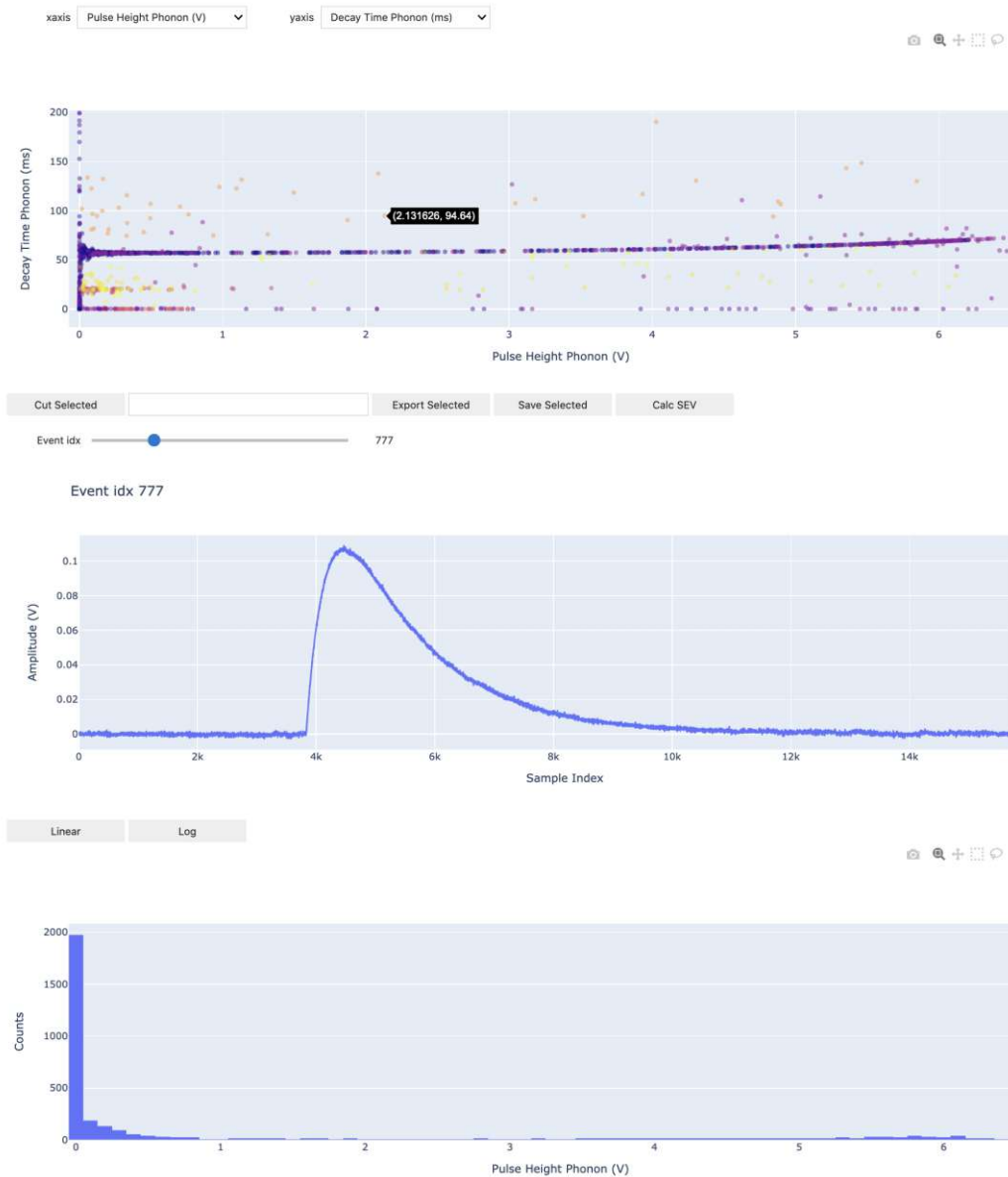
Figure 3.7: Screenshots of the VizTool from Cait. (upper panel) A clickable scatter plot of two feature dimensions allows for zooming and selecting regions of the feature space with a box and a lasso tool. Through the color of the markers, a third feature dimension can be made visible. (middle panel) The pulse shape of clicked events is shown when an event is clicked in the scatter plot. (lower panel) A histogram of the feature dimension on the abscissa is shown for the selected event population. Fig. and caption are also used in Ref. [5].

3.1.3.3 Quality cuts with machine learning

The tuning of quality cuts is one of the steps in the RD analysis that can be poorly automated since the distributions dealt with are of no predictable shape. They are especially not Gaussian. Furthermore, sometimes quality cuts are tricky to design when the information contained in multiple features must be combined. The typical procedure is to engineer features that contain all relevant information, such that cuts can be done in one or maximally two dimensions. However, it was shown in Ref. [5], and by others e.g. in Refs. [166–174] that this handcrafted feature engineering procedure can be replaced by using classifiers that work in multiple dimensions. The challenge is then to have labeled data sets available that contain the desired types of events, e.g. only target recoils and only artifacts or a specific class of artifacts. This can be overcome by dedicated data sets from calibration runs, where it is known that a high rate of target recoils is contained, as was done e.g. in Ref. [175], or by simulating the desired event classes. The simulation procedure was studied in Ref. [5] and is very similar to the simulated pulses for the energy resolution evaluation. However, in this section, no measured noise traces were used for the simulation, but simulated noise (with the method from Ref. [108]) from artificially created NPS.

The Cait package and its contained methods discussed in Ref. [5] add new possibilities to the discrimination with machine learning classifiers. It contains methods to simulate not only pulse-shaped events but also artifacts and construct artificial events that can be used as negatively labeled sets for training classifiers dedicated to data cleaning. Ref. [5] contains a demonstration of such a classifier trained and evaluated on distinct training and test sets of simulated data, which we shall review and discuss in this section. The distributions of event types within their two data sets are shown in Tab. 3.2, and the event types were introduced in Tab. 3.1. There is one qualitative difference between the event class "event pulse" in the training and the test set: while in the test set, this is always a fixed pulse shape, in the training set, the pulse shape parameters are sampled from a set of reasonable pulse shapes parameter, such that every pulse has an individual shape. With this, the largest part of the reasonable parameters space of possible pulse shapes is contained in the training set. Event heights are sampled from a uniform distribution between zero and a reasonable maximal height.

A classifier was trained in Ref. [5] to discriminate between pulse-shaped events, the event types called "event pulse" and "carrier event", and other events that are all artifacts, i.e. to perform exactly the task of data cleaning. A convolutional NN (CNN) was used as a classifier with two convolutional layers performing feature extraction

Event type	Test set	Training set
Event pulse	8000	20000
Noise	3000	10000
Decaying baseline	500	1500
Temperature rise	500	1500
Spike	500	1500
SQUID jump	500	1500
Reset	500	3000
Tail event	500	5000
Decaying baseline with event pulse	500	5000
Decaying baseline with tail event	500	-
Pile-up	9000	10000
Early or late trigger	500	5000
Carrier event	8000	-

Table 3.2: Event classes in the simulated data sets. Table also used in Ref. [5].

from the raw voltage traces, followed by a logical unit of two feed-forward layers. The architecture of the network is summarized in Tab. 3.3. The PyTorch [163] framework was used for constructing the NN. Before, the events were normalized to a PH of one, such that the PH and, therefore, the energy cannot have an impact on the choice of whether the event survives or not. All time series were downsampled to a length of 512 samples.

The training set was again split in the ratio 7:2:1 into a subset on which the weights of the network are trained, one on which the hyperparameters (validation set) are trained, and one that is used to test the model performance (test set). The training was done in 70 epochs with the ADAM optimizer [165], a learning rate of 0.001, and a batch size of 64 and otherwise the default PyTorch options for the optimizer. Loss values of the used negative loglikelihood function were observed as 0.070 in the training, 0.081 in the validation, and 0.085 in the test set after the training was completed. The training was done within the PyTorch Lightning framework. The full training took 23 min on a 2 GHz Quad-Core 10th gen Intel Core i5 CPU. The survival rates of the two pulse classes, from which one is a longer and one a shorter pulse and the survival rate of noise events are shown in Fig. 3.8, in comparison to a threshold of five times the energy resolution. It can be seen that the survival rates are close to optimal above the threshold. Ref. [5] reports furthermore the model reaches an unweighted accuracy score of 0.941 across all event classes. The largest group that

Layer	Specifications
1D convolutional layer	1 input channel, 50 output channels, kernelsize 8, stride 4
1D convolutional layer	50 input channel, 10 output channels, kernelsize 8, stride 4
Feed forward layer	639 input nodes, 200 output nodes, ReLU activation function
Feed forward layer	200 input nodes, 2 output nodes, ReLU activation function

Table 3.3: The CNN model architecture. The rectified linear unit (ReLU) is used as activation function. Table also used in Ref. [5].

could automatically not perfectly be cut are pile-up events. A combination with the method studied in Sec. 3.1.2.6 could significantly improve the results. A distribution of the survival rate depending on the PH does not make sense for artifacts since the PH is not a reasonably descriptive quantity for events that are not pulse-shaped.
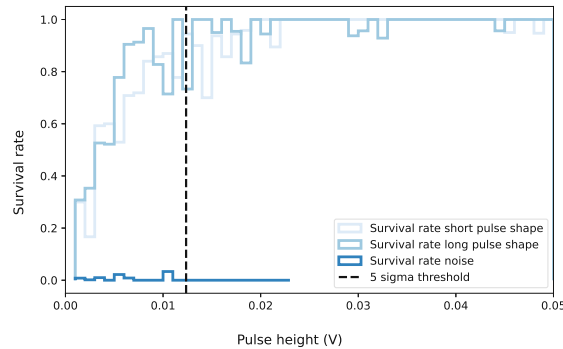

Figure 3.8: Survival rate of different event classes after application of the CNN as a quality cut. As a reference, a black dashed line indicates a typical threshold value at five times the noise resolution σ value. The long pulse shapes (medium blue) and the short pulse shapes (light blue) achieve both an efficiency of around 0.9 down to the threshold. No noise events (dark blue) survive the CNN cut above the threshold. Fig. and caption also used in Ref. [5].

A demonstration of the capability of such an automated data cleaning CNN is shown in Fig. 3.9. The output of the CNN is a scalar between 0 and 1 and can be interpreted as the belief of the network that the event is pulse-shaped. The network
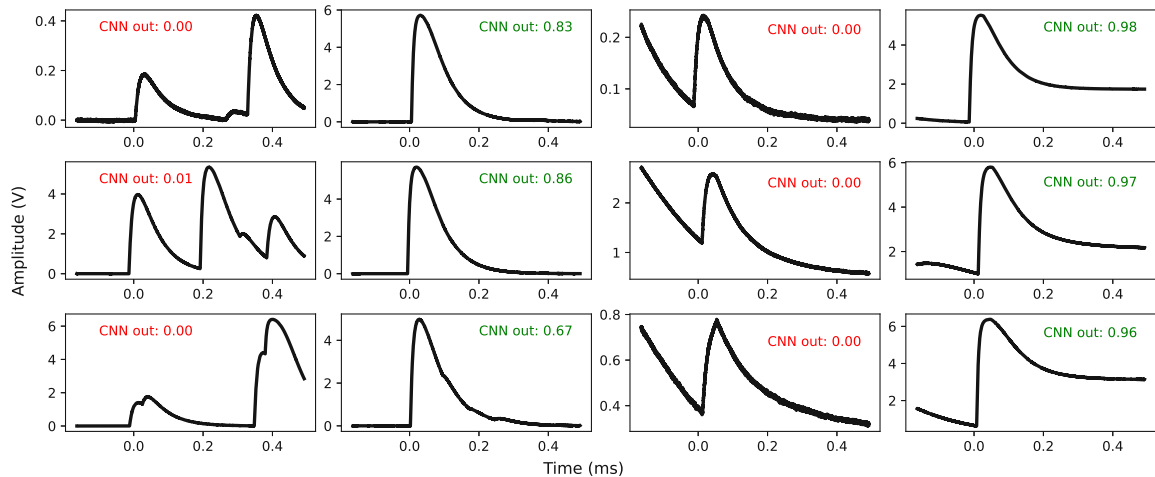
successfully discriminated against strong pile-ups of pulses or decaying baselines from pulse-shaped events. The network especially also interprets the event class "decaying baseline with tail event" as positive pulse-shaped events even though they were not contained in the training set. This choice resembles the choice an analyst would make with handcrafted cuts. The demonstration from Ref. [5] is uplifting and provides optimism that data cleaning could be automated with pre-trained NNs.

Figure 3.9: Exemplary "pile-up" (first and second columns) and "decaying baseline with tail event" (third and fourth column) events from the test set. For the first event class, only pulse shapes with small piled-up, secondary pulses survive the cut (CNN out > 0.5), while severe pile-ups are significantly rejected (CNN out << 0.5). The second event class was not included in the training set and serves, therefore, as a metric for the performance of the CNN on yet unseen event classes. Events with a strong decaying baseline get rejected, while events with a weak decaying baseline survive the cut. Fig. and caption also used in Ref. [5].

There are several limitations of the method as it is presented in Ref. [5]: if the validation of the analysis, which will be discussed in Sec. 3.1.4, is done with simulated data and the training of classifiers is also done on simulated data, is unclear how one would spot a difference in the effect of the model to measured and simulated data due to a systematic difference between simulation and measured data. Therefore the analyst must trust that the simulation resembles the measured data in the realistic reproduction of pulse shape and noise so closely that no difference in the effect of the classifier is possible.

To resolve this issue Ref. [4] applied a similar method, but with training on labeled data sets from measured data. Those results will be discussed in more detail in Sec. 6.1. One further limitation of this automated data cleaning method should be mentioned: the discrimination between different pulse shapes that all survive the classifier. This

cannot be fully automated since the analyst must decide which pulse shape corresponds to the target recoils. However, the procedure can be further simplified, e.g. with a PCA method that was used for PSD in Ref. [176], and is further discussed for our purposes in Ref. [4] and reviewed in Sec. 6.1.

3.1.3.4 Coincidence cuts

Many experiments operate multiple detector modules simultaneously in a joint experimental setup. In such scenarios, the coincidences between energy depositions in the detectors can be used to further discriminate between candidate events for physics searches and events with other origins. Weakly interacting events, such as DM events, are expected to interact only in one of the detector modules due to their low cross-section and interaction probability. EM events are expected to leave a trace of interactions along their flight path. This is often used to discriminate muon events with dedicated muon veto detectors. Also neutrons were in previous CRESST runs observed to leave a trace of coincidental interactions. This is possible due to the strongly interacting nucleus of the neutron, which provides it with a higher interaction probability with other nuclei than DM has. Also environmental impact, as vibrations, could cause events in coincidence in multiple detectors. The coincidence cut is then a cut that excludes events in temporal proximity of several milliseconds, depending on the temporal resolution of the system, with events in the veto channel.

In this section, we have discussed the selection criteria often used to discriminate target events from artifacts. This process can be called data cleaning and consists of elements of PSD, data cuts based on the operation conditions, and coincidence cuts. Large parts of this process can be automated using hard-coded criteria based on feature statistics or machine learning. The process of PD, where e.g. EM nuclear events are discriminated, is typically kept for the HLD analysis. The categorization between cleaning and PD is not strict, and in the literature used with different meanings in different references. Finally, the impact of the chosen cuts on the expected scattering rate of the sought-physics process must be estimated, which we discuss in the following section.

3.1.4 Validation

3.1.4.1 Cut and trigger efficiency

We discussed the data processing pipeline in the previous section, starting from the raw sensor signals to the final recoil energy estimates. A crucial step was the data cleaning, the discrimination between target events and artifacts. In a realistic experiment, the feature distributions of these positives and negatives do not cover distinct subsets of the parameter space but usually overlap in the tails of their respective distributions. In addition, the processing pipeline itself might be prone to errors which makes it necessary to validate the whole procedured, and to take systematic effects into account. This is, in practice, done by calculating the efficiency of the data selection cuts and the trigger algorithm by simulating target events and subjecting them to the whole data processing pipeline. The cut efficiency is an energy-dependent function that returns an estimate of the probability that target events with the given recoil energy survive all selection criteria. The trigger efficiency is analogously defined, but for the chance that those events are triggered. The simulation of target events is done similarly to the described procedure in Sec. 3.1.1.4 used to determine the energy resolution. The relevant differences are the choice of PH and pulse onset distribution. The onset distribution should be chosen as similarly as possible to the expected onset distribution in real target recoils. Usually, this requires covering a narrow interval around 1/4 of the record window. For the PH distribution, two different approaches exist: the best approach is to choose the energy distribution exactly as the expected DM signal, which means, in the case of elastic nuclear scattering, a roughly exponentially decreasing function towards higher energies (exact formula in Sec. 1.3), with its parameters determined from the DM mass that shall be tested. However, this procedure can be impractical since an individual simulation must be made for each tested DM mass. Alternatively, we simulate signal events with a uniform or a generic logarithmic PH distribution, fulfilling the requirement to cover energy regions densely where strong changes in the effect of the survival criteria are expected. Such a region is especially the trigger threshold, which needs to be sampled with a dense mesh grid. Since the simulated PH distribution is known, we can later adjust the expected DM scattering rate by multiplying it with the energy-dependent cut and trigger efficiency. For the simulation of veto channels, as e.g. the light channel, it is important that those are simulated according to the expectation from the DM scattering process, in this case with the relevant QFs.

The trigger probability of the simulated events is typically error-function shaped around the set trigger threshold. By fitting an error function to the histogram of triggered events, depending on their simulated PH, we can read an estimate of the energy resolution from the fitted width of the error function. This estimate is a final check of the resolution and should agree with the calculated values as described in Sec. 3.1.1.4.

Before we end this section, we want to motivate a thought experiment about the reliability of efficiencies. Assuming that we can simulate perfectly realistic events, their distribution fully overlaps with the distribution of signal events that would be observed in an experiment. In this scenario, we could simulate such events and train a classifier on the full set of event features, or even directly on the pulse trace, with the objective to discriminate (simulated) signal events and measured events. An ideal classifier would learn to reject measured events that do not fulfill the signal characteristics and retail the largest share of the simulated signal events and measured events that fulfill the signal characteristics. To validate the cut efficiency, we could simulate a second, distinct set of signal events, acting as a test set. While it sounds like a good strategy to design an ideal data selection, it comes at the same time at the price of the impact of systematic uncertainties in the data simulation. As an example: it was shown in Ref. [156] that simulated noise traces are clearly distinguishable from measured ones by NNs. Such an effect would effectively enable the classifier to fit the systematic differences between the simulation and measured data and score a high cut efficiency while actually rejecting a much larger share of measured events with signal characteristics. One has to be careful with the details of the data selection criteria that one applies, both when using learned or handcrafted cuts. Applying strong cuts is generally often beneficial for the results, but it should always be made sure that in the parameter regions where strong cuts are applied, the simulated and measured data (e.g. from calibration sources) are similar enough to estimate reliable cut efficiencies.

3.1.4.2 Blinding schemes

One could expect that the calculation and inclusion of an efficiency in the statistical verification of signals is alone a strong enough intervention to assume the validity of an analysis chain and the emerging results. This is not always the case. Consider one of the following two scenarios where a limit on DM interactions would be spoiled: first, let's extend the thought experiment from the previous section. An analyst trained a model to reject all measured events while retaining all simulated efficiency events. If they have enough capability to learn complex dependencies, classifiers can learn the

114

positive and negative events by heart, an effect called overfitting in machine learning literature. The second and equivalent scenario is that of a "human" classifier, an analyst that designs sophisticated enough cuts to reject all the measured events while the cut efficiency stays high. This procedure could be considered a manual form of overfitting. In both cases, the discrete distribution of measured events was used to design selection criteria instead of an estimate of the underlying, true probability distribution. Automatically, results on the data set at hand will be better than on an equivalent but freshly measured data set.

This problem is in physics searches dealt with by a blinding scheme. The idea of a blinding scheme is that a defined subset of the measured data, or events within a certain region of estimated recoil energies, are left untouched and are not looked at until the analysis chain and all selection criteria are defined. Using the analysis chain, which was designed on another data set, on the blinded subset prevents a series of cognitive bias effects that analysts can otherwise unknowingly be subject to.

Designing quality cuts in the presence of a blinding scheme has many parallels to the well-known problem of bias-variance trade-off in machine learning and function approximation literature. This problem formalizes the optimum between overfitting to training data, i.e. learning samples by heart, and underfitting it, i.e. not using all information that is available in the data. There are many strategies to address this problem. A widely used one is the application of a second data split of the unblinded data set in a training set, on which classifiers are trained, or cuts are designed, and a validation set, on which those are compared and cross-checked for consistency.

In physics searches, the common practice is sometimes to have multiple analysts in parallel and to decide the official analysis chain that shall be applied to the blind data based on the results that were obtained by the analysts on the training set. This is a reasonable strategy, as long as all analysis chains fulfill quality criteria sufficiently to avoid overfitting on the training data. However, we know from the problem of bias-variance trade-off, that it would be a much better strategy to compare results on a validation set, instead of the training set. Furthermore, reviewing the analysts' consistency and quality of decisions, instead of the obtained results, would likely provide better chances of a good generalization to the blind data set. This statement holds not only for decisions in the analysis of physics experiments but in all other situations of data-based decision-making as well.

In this section, we discussed the process of RD analysis in detail, with its four

subprocesses: the detector characterization, the event characterization, the event selection, and the validation of the analysis chain. The subprocesses are partially entangled with each other. In theory, they can be executed subsequently or independently in parallel. It's often a fourth and back between the individual steps. The event characterization and validation steps can be readily automated by choice of a few reasonable parameters. The detector characterization and event selection process are in the currently used analysis chain, not automated, but performed manually per hand. The contributions of Ref. [5] towards automation of the data cleaning are crucial and enable the automation of the full RD analysis procedure for well-behaved detectors that do not surprise with unseen event classes. The problem of validating event selections that were trained on simulated events remains and will be dealt with in Chap. 6.

The final result of a RD analysis is a list of the cleaned target events, including their recoil energies and auxiliary information that might be useful in the HLD analysis, such as veto yields and time steps. Additionally, the energy-dependent cut and trigger efficiencies are required, either as a binned histogram or as a list of simulated events with their energies, auxiliary information, and binary flags that mark if they were triggered and if they survived the selection criteria. The further information necessary for the HLD analysis is the measurement time and target weight to calculate the exposure to scale the expected DM scattering rate accordingly. Finally, information about the energy resolution and energy threshold is typically required. The full procedure of the HLD analysis is discussed in the following.

3.2 High-level data analysis

We have established a systematic workflow for the RD analysis in Sec. 3.1. Such analysis results in a spectrum of measured recoil energies, a corresponding cut efficiency, and auxiliary quantities, such as e.g. LYs or signals from veto detectors. The goal of physics searches is to make a statement about the existence of a certain physics signal in those data. Since this process is relatively detached from the exact detectors and experimental apparatus, we call it HLD analysis. We will start in this section by introducing HLD selection cuts, which aim to discriminate particle backgrounds from signal candidates, i.e. PD. Once we establish this procedure, we will investigate statistical methods to conduct signal searches: Yellin's maximum gap and optimum interval methods and the likelihood (LH) analysis. Finally, we will discuss challenges in the combination of multiple data sets by using simulated toy data from generic experiments.

3.2.1 Particle discrimination

We introduced the idea of experimental multi-channel readouts in Sec. 1.3. These experiments allow for collecting data on multiple energy scales. The energy scale from phonon readouts is typically described as total recoil energy since the largest share of the recoil energy is transferred into quasiparticles that successively induce heat in the target or sensor. This energy scale only negligibly depends on the type of recoil, nuclear or electron, massive particle, or X-ray[2]. The situation is different for light and ionization channels, which significantly depend on the type of recoil. Their energy scales are usually described as electron-equivalent or nuclear recoil-equivalent, depending on the source which was used for calibration.

We will consider for this section an experiment that has measured a total and an electron-equivalent energy scale through the measurement of a phonon and a light channel. This is the concept that is used in the CRESST and COSINUS experiments. The energies are typically plotted in a scatterplot or two-dimensional histogram against each other to characterize the particle types in the data set. We will consider, for now, that total energy depositions correspond to x-values and electron-equivalent energy depositions to y-values. For better visibility, these quantities are often visualized by scattering the the y/x ratio against the x values. A plot of these quantities will be

---

[2]As mentioned in the previous section, this statement is only true for crystals with relatively low light output. Otherwise, the energy scale has to be corrected for the energy that was emitted as photons, as was e.g. done in Ref. [161].

shown later in this work, in Fig. 5.6. Since both energy estimates are subject to noise fluctuations, the distributions of individual event types build band-like structures in the landscape of the two energy estimates. Depending on the energy resolutions in the channels, individual bands for recoils on individual nuclei appear and for the scattering of α, β and γ particles. The commonly used analysis method to determine the LH of individual events belonging to a certain distribution is a bandfit. A band B is a weighted Gaussian function $\mathcal{N}(\cdot|\mu, \sigma)$, where the mean value μ is a function of the total energy deposition:

$$B(x, y) = b(x)\mathcal{N}(y|\mu(x), \sigma). \tag{3.17}$$

The Gaussian distribution is justified since most readout noise, which determines the baseline energy resolution, is approximately Gaussian. A fit with a mixture of such functions allows the description of any band-like structure in the two-dimensional energy plane, where the explicit choice of the parametrization of μ determines the flexibility of the fit. It is possible also to parametrize the standard deviation σ. Since this leads to a more complicated fit, it is often not done in a standard analysis. Furthermore, this fit allows in the presence of both electron and nuclear recoils in the data for a determination of the light-quenching factors and their energy dependence, i.e. the difference in the light output of the target between nuclear and electron recoils. This allows for in-situ measurements of the QFs, which is a strong advantage of two-channel readouts compared to single-channel experiments.

For an individual event with energy values (x_0, y_0), the LH of belonging to one of the bands can be determined by normalizing the mixture of bands along the y-axis while fixing x_0. The Gaussian mixture can then be interpreted as a probability distribution, and their evaluations at (x_0, y_0) are the LHs of the event belonging to the specific bands.

Typically a ROI is defined in the two dimensions plane, which is expected to contain the signal-like events. This can e.g. be chosen as the lower half of one of the nuclear recoil bands[3] to search for DM-nucleus scattering. Alternatively, to the definition of an ROI, the bands can also be used for a full statistical LH analysis by including specific backgrounds and the sought-after signal in the coefficients of the bands. Statistical methods for both searches with and without ROI are discussed in the following section.

---

[3]Each nucleus has an individual QF and recoil band.

3.2.2   Statistical methods for a signal search

Searches for evidence of a physics process in a collected data set follow statistical procedures to ensure the validity of the signal or limit claim. These procedures are typically designed to compare an expected signal distribution S with the measured data set in the presence of expected background distributions B that would inflict events in the signal region while accounting for random fluctuations. We will call samples from our collected data set in this section x for simpler notation.

The statistical method for performing this analysis is a hypothesis test. A hypothesis H is a certain statement that can make a difference in the observed data. In our case of testing the parameter space of a specific signal model, the hypothesis would be the statement that a parameter of the signal model equals a certain value, e.g. the interaction cross section that causes signal events σ_S. There are generally two types of hypothesis tests. A test can be designed to test the LH of an alternative hypothesis H_1 against a null hypothesis H_0. The second design is a test that seeks to reject the null hypothesis without an alternative hypothesis. Statistical tests are always bound to a significance level α. The significance level is the probability that the null hypothesis is rejected when it is true, which is also called a type 1 error. A second type of error, the type 2 error, is that an alternative hypothesis H_1 is falsely rejected. The probability of this error is denoted with β, and the value $1 - \beta$ is called the power of the test. A type 2 error can only occur in the first test design introduced above. The ideal test has a low probability of both type 1 and type 2 errors, i.e. a low significance level α and a high power $1 - \beta$. The feasibility of this goal is determined by the available data. In physics and everywhere else, the goal of experiment design is to collect data with a distribution that would be as different as possible if the null or alternative hypothesis were true.

In many branches of science, the test result is quantified with a p-value, which is the probability of obtaining the observed data, assuming the null hypothesis is true. The null hypothesis should be rejected once the p-value falls below the significance level $p < \alpha$. Since we want to make general statements about nature in physics, we usually work with very low significance levels to not risk a chance of wrongly rejecting established rules of physics. The significance level is typically stated not as a p-value but at the equivalent outlier probability of a random number drawn from a standard normal distribution. The probability of observing this random variable further than "1 σ" (one standard deviation) away from its expected value is 0.317. For higher values of σ, see Tab. 3.4. A typical significance level of hypothesis tests in physics is 5 σ, which is equivalent to $\alpha = 5.73 \cdot 10^{-7}$. Luckily we can conduct many types of experiments

in reproducible lab environments, different from other fields e.g. medicine or social sciences, which enables us to collect relatively large amounts of high-quality data.

Number of Sigmas (k)	Probability ($2(1 - \Phi(k))$)
1	0.317310507863
2	0.045500263896
3	0.002699796063
4	0.000063342484
5	0.000000573303

Table 3.4: Probability of a random number from a standard normal distribution being further than k sigmas away from the expected value. $\Phi(k)$ is the CDF of the standard normal distribution.

In signal searches, our alternative hypothesis H_1 is that the data distribution contains signal and background components:

$$H_1 : x \sim \mathcal{S}(\sigma_\mathcal{S}) + \mathcal{B}. \tag{3.18}$$

We want to test this hypothesis against the null hypothesis H_0, which is that the data contains only background components:

$$H_0 : x \sim \mathcal{B}. \tag{3.19}$$

Since the signal models which we treat in this work scale linearly with an interaction cross section parameter σ_S, the above-stated hypothesis is equivalent to the hypothesis $H_1 : \sigma_S > 0$ and $H_1 : \sigma_S = 0$. We have no predicted value for σ_S a-priori in many searches. We therefore want to treat σ_S as a free parameter. The methodology used is that of a LH ratio test, which we will introduce in the following.

3.2.2.1 The likelihood ratio test

The LH function $\mathcal{L}(\sigma_S, \lambda, \theta)$ is a function of the model parameters and equivalent to the evaluation of the data distribution $\mathcal{P}(x)$ predicted by the model with the observed data points x_i:

$$\mathcal{L}(\sigma_S, \lambda, \theta) = \prod_i \mathcal{P}(x_i | \sigma_S, \lambda, \theta). \tag{3.20}$$

The evaluation of the probability distribution of events can be multiplied since the events are assumed to be independent of each other. Here λ is a placeholder for other parameters of the signal model $\mathcal{S}(\sigma_S, \lambda)$, and can in general be a vector of real

numbers. In the case of testing DM-nucleus elastic scattering, the mass of the tested DM particle would be such a parameter. θ corresponds to free parameters of the background model $\mathcal{B}(\theta)$, which are also called nuisance parameters. θ is in general a vector of real numbers as well. Note that for this framework, the alternative hypothesis must supersede the null hypothesis $H_1 \supset H_0$.

The LH function requires per design that the distribution $\mathcal{P}$ must be a probability distribution, it especially must be normalized. Since the normalization of the predicted data distribution usually depends on the parameters, adjustments to the parameters could have unpredictable effects on the LH. We therefore use a simple extension to the LH function that was introduced in Ref. [177]:

$$\mathcal{L}(\sigma_S, \lambda, \theta) = e^{-\nu} \prod_i \mathcal{P}(x_i|\sigma_S, \lambda, \theta). \tag{3.21}$$

This function is called the extended LH function. Here ν is the norm of the predicted data distribution over the observation space, e.g. the ROI:

$$\nu = \int_{\text{ROI}} \mathcal{P}(u|\sigma_S, \lambda, \theta) du \tag{3.22}$$

This adaption is based on the assumption that the number of observed events is a Poissonian distributed random number, with the expected value ν. When we adjust the parameters of the signal model, the extended LH cannot rise indefinitely, while the normal LH function could.

Maximizing the LH function w.r.t. the parameters of the model (σ_S, θ) leads to the so-called maximum LH estimator for the model parameters $(\hat{\sigma_S}, \hat{\theta})$. We use this estimator to describe the collected data set. Note that we consider a scenario where we only test the cross-section of the signal model while we keep the other signal parameters λ fixed. This corresponds to testing a specific theory, in our case, a specific DM mass. The procedure can then be applied to all theories of interest.

Since a numerical optimization procedure for the product of many probabilities can be cumbersome, one usually maximizes the log-LH, where the product is naturally replaced by a sum. Furthermore, since most optimization literature follows the convention that parameter optimization problems are minimization problems, we minimize the negative logarithmic LH function:

$$(\hat{\sigma_S}, \hat{\theta}) = \arg \min_{\sigma_S, \theta} \left(-\ln \mathcal{L}(\sigma_S, \lambda, \theta) \right). \tag{3.23}$$

Our alternative hypothesis H_1 can now be stated as $\sigma_S = \hat{\sigma_S}$. The LH ratio test is based on the comparison of the value of the LH function with the maximum

LH estimator, the alternative hypothesis H_1, and the LH function under the null hypothesis H_0. The null hypothesis restricts the signal cross section $\sigma_S = 0$, while the nuisance parameters are fitted alone, resulting in θ_{null}. The test statistic D that we shall use for our hypothesis test is

$$D = -2 \ln \left(\frac{\mathcal{L}(0, \lambda, \theta_{null})}{\mathcal{L}(\hat{\sigma_S}, \lambda, \hat{\theta})} \right). \tag{3.24}$$

Wilk's theorem [178] states that D is asymptotically distributed according to the χ^2 distribution with k degrees of freedom if the null hypothesis is true. Here k is the number of fitted signal parameters minus the number of fitted background parameters, i.e. $k = (\dim(\sigma_S) + \dim(\nu)) - (\dim(\nu)) = 1$ in our case. The null hypothesis should then be rejected if the quantity $1 - \Phi(D)$ is smaller than the required significant level α, where Φ is the CDF of the χ^2 distribution. Since the χ^2 distribution with k degrees of freedom is the sum of the squares of k independent standard normal random variables, we can, for our case with one degree of freedom, compare the quantity $\sqrt{D}$ with the CDF of the standard normal distribution.

When the sign of a signal is observed in a DM data set, we would typically perform a so-called positive analysis, to confirm the statistical significance of the signal. This procedure is equivalent to the procedure explained above, only that also the theory, i.e. the λ parameters, are fitted to the data. In this scenario, the degrees of freedom of the χ^2 distribution increase by the number of signal parameters $\dim(\lambda)$. Fixing the theory already a-priori is usually done to construct signal islands in the parameter space, or to set upper limits on the cross section. The procedure for setting a limit is explained in the following.

In a scenario where no convincing signal is observed, we want to quantify which regions of the signal parameter space are incompatible with our observations by a certain significance level. We keep our previous alternative hypothesis H_1 as the maximum LH parameters for this scenario. Instead of fixing the null hypothesis, we fix the significance level. For testing which parameter space is excluded by 90%, a commonly used significance level for exclusion limits, we compare the quantity $\sqrt{D(\sigma_{\text{excl}})}$ with the value 1.282, which is the corresponding quantile of the standard normal distribution. The procedure works equivalently to produce signal-islands, by using different significance levels, and a symmetric CDF. The test quantity has in this setting a free parameter, the free parameter of the null hypothesis, which is the cross-section σ_{excl} that can be excluded with the chosen significance. Since we know that higher cross-sections would lead to a stronger signal signature in the data, we

know that any cross-section higher than this value is excluded as well, with at least similar significance.

Technically the calculation of exclusion limits requires a two-step process of numerical optimization. The problem we want to solve has the character of a root search:

$$\text{Find } \sigma_{\text{excl}} \text{ s.t. } 0 = -\frac{Z^2}{2} - \ln\left(\mathcal{L}(\sigma_{\text{excl}}, \lambda, \hat{\theta_{\text{excl}}})\right) + \ln\left(\mathcal{L}(\hat{\sigma}_S, \lambda, \hat{\theta})\right). \qquad (3.25)$$

For every evaluation of the LH function with a new parameter, we need to solve a minimization problem again when obtaining the best fitting background parameters $\hat{\theta}_{\text{excl}}$ for the new setting of the cross-section. Calculating an exclusion limit through the LH ratio test can be costly, especially considering that the process needs to be repeated for every signal parameter λ of interest. We will explore in the next section another method for calculating exclusion limits that is cheaper in terms of computational cost and has other advantages.

3.2.2.2 The maximum gap and optimum interval methods

We have discussed in Sec. 2.2.3 excesses above known backgrounds that are currently observed in data sets from cryogenic detectors. Usually, suitable excesses could be interpreted as a signal. However, for these excesses, it is known that they are incompatible with signatures of new physics. This information is not reflected in the data in a way that is accessible to our default statistical methods for signal searches. Therefore, we need to account for this knowledge in how we use our methods. Typically this is done by interpreting the excesses as signal events but setting only exclusion limits. This way, we reach a conservative limit, and by not performing a positive analysis, we witness that the excesses can and should not be interpreted as signal. Such excesses are not only harmful since they could inflict a wrong signal claim, but the situation is even more complicated since no proper or justified parametrization exists for many of them. They cannot be included in a background model. For such situations, we require a method that can set limits on a signal model without explicit knowledge of the background distribution. Therefore, we have only a null hypothesis H_0 and seek to reject it.

The simplest method of such kind would be a cut-and-count analysis: the total number of events in the ROI is the test statistic. We can test this quantity against a Poissonian distribution where the mean value is the number of events predicted in the ROI by the signal model. This method is likely the most conservative since all
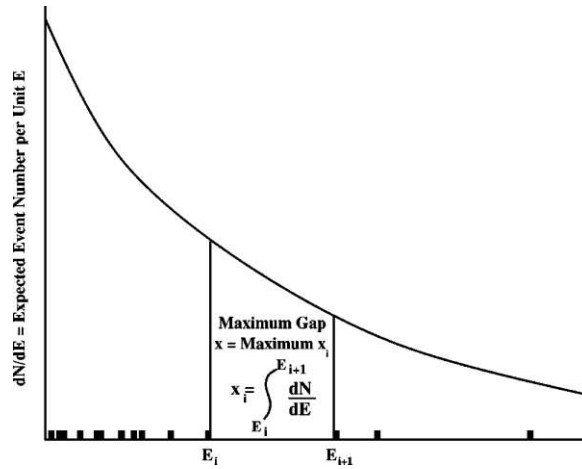
Figure 3.10: Illustration of the maximum gap method. The horizontal axis is some parameter "E" measured for each event. The smooth curve is the signal expected for the proposed cross section, including any known background. The events from signal, known background, and unknown background are the small rectangles along the horizontal axis. The integral of the signal between two events is "xi". Fig. and caption taken from Ref. [179]. The quantity $\frac{dN}{dE}$ corresponds to $\mathcal{S}$ in the text.

background events in the ROI are automatically interpreted as signal events. However, we would discard the information about the expected distribution of signal events within the ROI.

Another set of methods can use the expected signal shape to discriminate between this and background distributions. These are the maximum gap and optimum interval methods, introduced in Ref. [179] for DM searches and often called Yellin methods, after the author.

We consider here the case of a one-dimensional observation space, namely the scenario of interest, where this data dimension is the measured recoil energy E. For the maximum gap method, we define the gap sizes x_i as the values [4]:

$$x_i = \int_{E_i}^{E_{i+1}} \mathcal{S}(E)dE. \tag{3.26}$$

Here $\mathcal{S}$ is again the signal model. A gap is visualized in Fig. 3.10. Ref. [179] motivates an injective transformation of the data, such that the expected signal distribution would be uniformly distributed with unit density in the interval 0 to μ, the total number of expected events. Ref. [179] further derives that the probability C_0 of the maximum gap size being smaller than a particular value of x is given by:

---

[4]Be aware that in this discussion, x_i does not correspond to the individual events as in the previous section, but to the gaps in between.

$$C_0(x, \mu) = \sum_{k=0}^{m} \frac{(kx - \mu)^k e^{-kx}}{k!} \left(1 + \frac{k}{\mu - kx}\right), \qquad (3.27)$$

where m is the greatest integer $\leq \mu/x$. Since μ is linearly dependent on the signal cross section σ_S, we can now increase σ_S until C_0 reaches the required significance level for the exclusion.

The optimum interval method is based on the same idea, but instead of measuring gaps, we measure intervals between data points:

$$x_{ij} = \int_{E_i}^{E_j} \mathcal{S}(E) dE, \qquad (3.28)$$

where $j > i$. The interval x_{ij} covers $j - i - 1$ events. We can then equivalently define probabilities $C_n(x, \mu)$ for any $n > 0$, that all intervals that cover $\leq n$ events have their expected number of events $\leq x$. The C_n functions are not accessible analytically but must be tabulated numerically with Monte Carlo methods. The largest C_n for a given data set and signal model is called C_{Max}. Ref. [179] defines further a function $\overline{C_{\text{Max}}}(C, \mu)$ to be the value such that fraction C of random experiments with that μ, and no unknown background, will give $C_{\text{Max}} < \overline{C_{\text{Max}}}(C, \mu)$. Also, this function has to be computed with Monte Carlo. The exclusion limit with probability $0 < C < 1$ is then the smallest value of σ_S where C_{Max} of the data equals $\overline{C_{\text{Max}}}(C, \mu)$. The optimum interval method is today widely used for setting exclusion limits for low threshold experiments.

With Yellin's methods it is generally only possible to test a specific theory, i.e. the signal parameters λ, e.g. the DM mass, introduced in the previous section, need to be fixed a-priori. Several extensions of the Yellin method exist. Ref. [180] introduces an asymptotic function for C_n that can be used for data sets with high statistics. It further introduces a generalization of the procedure for more than one data dimension. Ref. [181] discusses methods to combine limits from multiple data sets that were derived with the optimum interval method. The combination of limits will be the subject of the following section, where we discuss the potential and risks of combining data sets with both Yellin's methods and the LH ratio test in detail.

We have introduced the LH ratio test for a statistical signal search and limit setting with known background distributions and Yellin's methods for setting limits in the presence of unknown backgrounds. We have not explicitly included the time information of events in our elaboration. However, it can be included in the same way

as other observables of the events, as energy information, given a decent model of the expected event distribution in time. This is particularly the case when looking for an annual modulation in the event rate. We refer to Ref. [95], where such a search was conducted with data from the DAMA/LIBRA experiment and signal islands were projected in the landscape of DM nucleus scattering (see Fig. 1.7). Intuitively we would expect that more data are always better for a signal search. However, this is only the case if the used method is capable of extracting all the information from the data. Otherwise, the combination of data sets can even lead to results that allow only for weaker statements as one of the individual data sets. Exactly this question is discussed next in this chapter.

3.2.3 Challenges for the joint analysis of multiple data sets

This section closely follows the discussions that will be put forward in Ref. [1] (unpublished), and part of this section will be part of the publication. The DM community is interested in performing joint analyses of experiments to possibly gain stronger limits, or hints towards signals, by using even more of the information in the data. The first part of this section deals with the necessary ingredients for a combined analysis, and the later parts discuss complications in the statistical methods applied.

The results of a DM search depend strongly on the statistical procedures used and the preparation of the measured data. An open data policy is explicitly encouraged to enable comparability between experiments and reproducibility of results. This also enables independent searches for non-standard DM models by individual scientists and the joint analysis of measurements from multiple experiments. The information and data necessary are discussed in the following. This chapter's subsequent subsections describe examples of analysis techniques that are popular in the community to compare data between experiments and derive combined results.

First, the observables used for deriving the DM result are necessary on a suitable level of abstraction. Experiments measuring particle recoils in a detector material concern the measured recoil energies and additional information, e.g. time, stamps, if relevant to the analysis. Auxiliary observables should be included if they are used directly in the extraction of DM results, e.g. in combined LH fits on recoil energies and LYs. They can be absorbed in the overall cut efficiency if they are merely used to perform veto cuts (see below). Unbinned data formats are generally preferably over binning of physical quantities. However, there are scenarios where the latter is sufficient, especially for experiments that cannot resolve energies beyond natural binning (counting experiments) or face very high interaction rates. The recoil energies

should be specified as total recoil energy, nuclear-recoil equivalent energy, or electron equivalent energy, depending on the procedure of energy calibration.

The tested DM signal model should be described as a function mapping the experimental observables to the magnitude of the interaction rate density per exposure depending on the tested physics parameters. These are typically effective interaction strengths and variable parameters of the DM model. In the example of elastic DM nucleus scattering, the former would refer to a mapping from particle recoil energy to an expected count density and the latter to the DM mass. Results are only comparable and combinable under the identical underlying assumptions made in the derivation, they therefore need to be stated.

An understanding of applied data selection cuts on signal events is crucial for the derivation of reliable physics statements. The common procedure to develop such an understanding is the simulation of signal events with all features that are relevant for the selection criteria, e.g. recoil energies, LYs, pulse shapes, or SNRs. The simulation should be tuned to the individual measurement, e.g. by using calibration or training data, under consideration of the applied blinding scheme. After applying equivalent selection criteria to the data set of simulated signal events, their survival probability can be estimated as a function of the experimental observables used in deriving physics results. This typically results in an energy-dependent cut efficiency for recoil-based experiments. The derived efficiency should ideally be shared separately from the signal model to allow for independent modifications.

Setting an upper limit on the interaction strength of a scattering process is possible solely based on the absence of events and without knowledge of the origin of appearing background contributions. Many DM searches are conducted this way, as backgrounds appearing for detectors using novel technologies are sometimes not fully characterized or of an instrumental nature with varying, hardly predictable count rates and features. While this lies in the nature of such experiments and will likely not change for future searches, there are clear benefits in the knowledge of background contributions that can be used to improve upper limits and are especially important to claim discoveries reliably. A description of known backgrounds should therefore be included.

A special challenge for the data analysis is, when a background component resembles the spectral shape of the expected signal. At the time of writing, this harmful situation is present for various LEEs observed by low threshold experiments, described e.g. in Refs. [31, 139] and summarized in Ref. [7]. This case has to be treated with special caution: such observations can be declared as backgrounds, even if their origin is not fully understood, if they feature qualitative properties that are incompatible with

Experiment identifier	"A"	"B"	"C"
Exposure (kg days)	7	5	100
Rate (1 / (kg days))	300	200	20
Number surviving events	2100	1000	2000
Threshold (keV)	0.1	0.035	1
Upper end of ROI (keV)	20	10	400
Energy resolution (eV)	15	5	150
Cut efficiency above threshold	0.8	0.65	0.5

Table 3.5: Summary of the characteristics of the simulated data sets. See Sec. 3.2.3.1 for details.

the sought-for signal. This especially includes incompatibilities in the count rates across different measurements or in the time dependency. Ideally, such information is represented in the statistical method chosen to derive the physics result. If this is not feasible, certain pitfalls can occur both for individual and combined analysis of such data, which shall be discussed later in this section.

Finally, the derived physics statement should be included in a tabulated format that is usable in a data processing pipeline or for plotting routines. For this, the information in the published file naturally depends on the method used to obtain the result. Generally, combining results that were derived with different of the above named methods can be problematic. A more detailed discussion of these methods and the types of combined analysis they allow for follows in subsequent sections.

In full presence and knowledge of the above-summarized items and information, released data can reach their maximal impact. We showcase exemplary procedures to combine results in the following sections: in Sec. 3.2.3.2, we discuss counting experiments and the optimum interval and maximum gap methods for the scenario of unknown backgrounds, in Sec. 3.2.3.3 we discuss counting experiments and the profile LH method with background models. We use several simulated data sets for our examples that we describe in Sec. 3.2.3.1.

3.2.3.1   Characterization of the simulated data sets

Several effects of the methods of combined data analysis shall be shown on exemplary data in the following sections. To keep the examples simple, we simulate data that is qualitatively realistic but simplified and is sampled from known and analytically tractable distributions. We simulate data sets from three mock experiments A, B, and C. All of them share the same parametrization of an expected signal model, depending on a parameter β. The experiments have individual backgrounds, count

rates, and measurement characteristics. We generally simulate background-only data, but experiments A and B are designed to have background components that resemble signals for certain choices of β. The relevant observables are exclusively the recoil energies, and the energy calibration is done such that it is independent of the type of recoil. The corresponding cut efficiencies are flat but below unity, above the energy threshold, and smeared out by the baseline energy resolution around the threshold as they rapidly approach zero. The same applies to the upper end of the ROI. The global characteristics of the simulated measurements are summarized in Tab. 3.5. The common signal model for all three experiments has the form:

$$\mathcal{S}(E|\beta, w_0) = w_0 \cdot \mathcal{E}(E|\beta), \tag{3.29}$$

where E is the recoil energy, β is a parameter of the DM model, and w_0 is a parameter describing the effective interaction strength between DM and the detector material. $\mathcal{E}(E|\beta)$ describes an exponential distribution with shape parameter β. We can think of this model as a simplification of the model for elastic DM-nucleus scattering, with β taking the role of the DM mass. However, a realistic model for a scattering rate would depend on additional parameters, especially the detector material, and would have an upper bound, i.e. a highest possible energy that can be inflicted in a recoil, due to the upper bound of the velocity distribution of DM in the Milky Way. These refinements would not change the qualitative statement for which we aim in the following sections. The advantage of our simplified model is that the parameter β immediately corresponds to the expectation value of the recoil energy of signal events. The individual background events for our experiments A, B, and C are sampled from the following probability distributions:

$$\mathcal{B}_A(E) = 0.4 \cdot \mathcal{U}(E) + 0.3 \cdot \mathcal{N}(E|6, 0.5) + 0.3 \cdot \mathcal{E}(E|0.3) \tag{3.30}$$

$$\mathcal{B}_B(E) = 0.35 \cdot \mathcal{U}(E) + 0.3 \cdot \mathcal{N}(E|6, 0.5) + 0.35 \cdot \mathcal{E}(E|0.1) \tag{3.31}$$

$$\mathcal{B}_C(E) = 0.5 \cdot \Gamma(E|0.375, 933.33) + \sum_i w_i \cdot \mathcal{N}(E|\mu_i, \sigma_i) \tag{3.32}$$

with $w_i = (0.08, 0.12, 0.05, 0.15, 0.1)$, $\mu_i = (45, 75, 120, 200, 300)$ and $\sigma_i = (2, 3, 5, 10, 20)$. Here $\mathcal{U}(E)$ describes a uniform distribution in the ROI, $\mathcal{N}(E|\mu, \sigma)$ describes a normal distribution with mean μ and standard deviation σ and $\Gamma(E|k, \theta)$ a gamma distribution with shape and scale parameters k and θ.

We will in the following sections proceed to determine signal exclusion limits with our data sets. As experiments A and B contain exponentially decaying background

129

components, we first want to compute the signal parameters matching their backgrounds' shape and rate. For experiment A(B), the choice $\beta = 0.3(0.1)$ keV leads to identical signal and background shapes. For the rate parameter, we need first to disentangle the prefactor of the corresponding background component: 30% of the observed events are sampled from the exponential background component, which gives a rate of 100(70) counts / (kg days) for experiment A(B). We want to note that all background parameters are for our data sets chosen such that the choice of ROI and the cut efficiency change the distribution of events between the individual background components only negligibly. Therefore the distribution of observed background events resembles the distribution of simulated background events.

3.2.3.2   Combining upper limits with unmodelled backgrounds

Methods for statistical statements about the presence of a signal in a measurement were introduced in Sec. 3.2.2. Measurements with unmodelled backgrounds that inflict events in the ROI can easily be used to establish upper limits on the signal interaction rate. However, establishing signal discoveries under the presence of unmodelled backgrounds is generally not possible, as the statistical method has to make an implicit signal-only assumption.

Combining limits derived with a cut-and-count analysis is straightforward: the result of a cut-and-count analysis contains all information in the number of observed events, and the probability distribution is determined by the assumption of Poissonian statistics. Therefore, the observed event rates can be summed, and the signal rate corresponding to the smallest excluded hypothesis with the desired significance can be derived. That being said, this simple merging strategy will improve the limits for two experiments with similar background count rates, but if the background count rate of one experiment is significantly higher than the other, the combined limit will come out to be weaker than the individual limit of the experiment with lower count rate. This is not the result that one wants to achieve by combining information from measurements, and a similar issue was treated in detail for the maximum gap and optimum interval methods in Ref. [181]. This effect is shown for the maximum gap method in Fig. 3.11. Upper limits for the signal interaction rate are derived with the maximum gap method for experiments A, B, and C. The kinks in the limits from A and B around the parameter space where the background is compatible with the signal model are clearly visible. Merging the data and deriving combined limits does not lead to stronger limits: the combined limits are positioned between individual
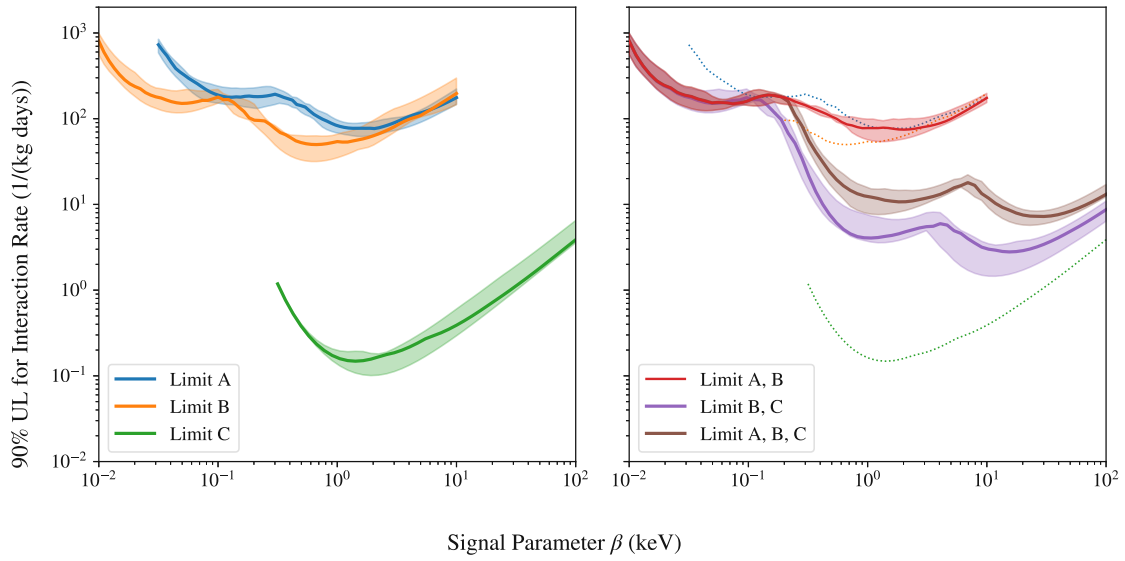
Figure 3.11: Upper limits on the signal interaction rate of experiments A, B, and C (left) and their combinations (right), depending on the signal shape parameter β. The error bounds are derived by the independent sampling of ten datasets from the background densities described in Sec. 3.2.3.1. The mean values of the individual limits are shown as dotted lines in the right plot. The combination of experiments does not improve the limits, which is fully expected as all experiments are limited by background and not exposure.

ones. This is fully expected due to the fact that the background rate is the limiting factor in our experiments instead of the exposure.

In Ref. [181] several alternative methods are studied to combine limits derived with the maximum gap and optimum interval methods, with the objective to a) improve limits from experiments with similar background rates due to the gain in exposure and b) to avoid a degradation of a limit far below the individual limits by using the information that all experiments would have to have observed a DM signal separately, within bounds of statistical fluctuation. It is concluded that methods exist that do well in both of the stated objectives individually, but no method exists that achieves good performance in both objectives when combining limits derived with the maximum gap or optimum interval method. An individual decision on which methods should be used has to be made based on previous knowledge about the compatibility of the measurement background rates. An interesting fact has to be noted about the methods presented in Ref. [181]: there are several methods that allow for combining maximum gap/optimum interval limits without accessing the data that was used to derive the limits directly, but by only accessing the calculated limits. This is interesting for large-scale joint analysis projects, where accessing all data sets individually would

131

create some technical overhead.

Generally, limits that are conservatively derived with different methods but identical confidence levels are comparable but not necessarily combinable without extra effort. Counting and profile LH methods automatically include LH functions that can easily be combined, while for the maximal gap and optimum interval methods, a probability distribution for the appearance of the gap/interval size has to be designed first, which can then be redefined as LH. This extra step enables the combination of maximum gap and optimum interval limits with limits from counting experiments and profile LH methods, and it is discussed as well in Ref. [181].

A blinding scheme for a combined analysis has to consider the included data sets to avoid a bias due to including only sets with beneficial realizations of statistical fluctuations. In practice, a fully blind analysis that combines several experiments is certainly challenging, as none of the used data sets may be unblinded for an individual analysis first. However, it is realistic that one experiment would wish to combine data sets from several detectors, and for that, the choice of included detectors must be made before the data is unblinded or otherwise justified with arguments that are statistically independent of the sought-for physics process, e.g. technical problems of detector operation.

3.2.3.3 Combining results with full background models

Models of expected background count rates are an advantage for any physics search. Such rates can be included in LH functions, which can be used to compare combined signal and background hypothesis and background-only hypothesis to derive upper limits and discoveries. Even more, the observed background in a combined analysis of several data sets can be attributed to the background components of individual data sets by using a global fit where only the signal parameters are shared. This procedure formalizes the information in the statistical method that a signal would necessarily be seen by all experiments, while backgrounds can differ. For this, special caution to a complete background model is required. The absence or a poor model of individual background components can not only lead to wrong physics statements but also lead to the same effect that was discussed in Sec. 3.2.3.2 where one data set deteriorates the individual results of another. Therefore it is generally preferable that contributions that were clearly identified as backgrounds but for which no model derived from the underlying origin is available are included with heuristic parametrizations in the background model.
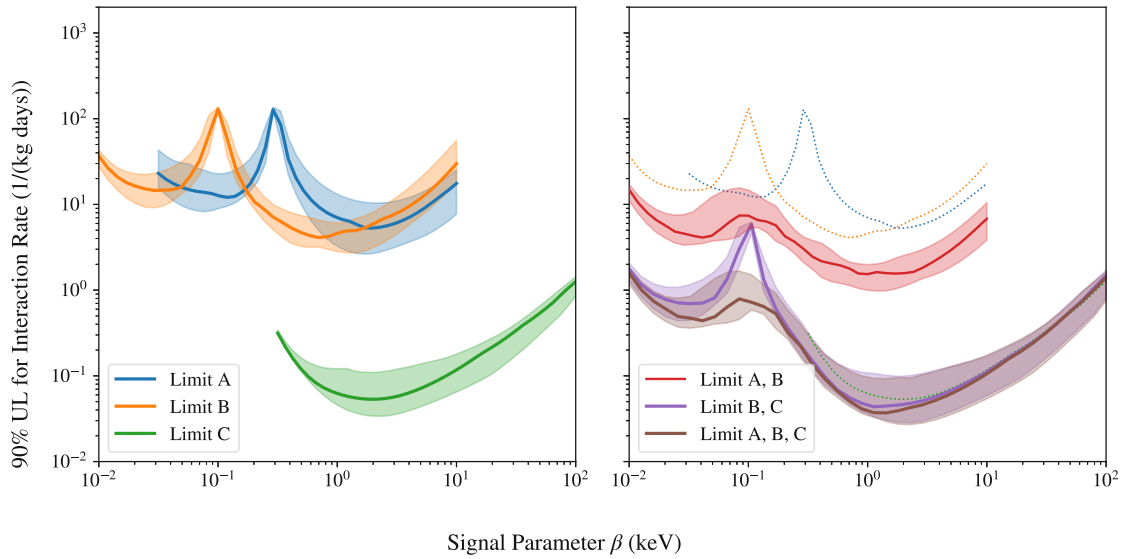
Figure 3.12: Same plot as shown in Fig. 3.11, but with upper limits derived with the profile LH method and full background models. Combining experiments always leads to an improvement in this setting, provided there are no unmodelled or poorly modelled backgrounds left, as the correlated fit can attribute events to background contributions from different experiments individually. The signal-like background of experiment A(B) for $\beta = 0.3(0.1)$ is also excluded in the combined fit.

As an example, we derived limits with the profile LH method for our experiments A, B, and C. For this, we use the full background models of the corresponding data sets and the signal model and perform a fit of all contributions' prefactors (weight parameters) by minimizing the extended negative log-LH. The extended LH method has the advantage of not requiring the LH to be normalized to one, i.e. it does not require a probability density, but any density function [177]. To derive upper limits, we compare the LH of the "best fit" hypothesis with a hypothesis with increased signal contribution. The signal contribution of the second hypothesis is increased, starting from the best-fit parameters, until the desired confidence level of exclusion is reached. Results for all data sets and their combinations are shown in Fig. 3.12. It is significant to note that the combined results are better in all cases, and the parameter space corresponding to the backgrounds compatible with the signal hypothesis in experiments A and B can be excluded in the combined fit.

A noteworthy detail is, that the LH functions of the individual experiments can be tabulated, only depending on the parameters of the signal model. This allows for combined fits and analysis without evaluating fits again depending on the individual events of the data sets, with the same results as are shown in our mock experiments.

A similar procedure can be applied for deriving signal discoveries. The hypotheses that are compared in that case are background-only and the combined signal and background scenario. As per definition, the signal component is missing from the background-only scenario, the significance of a signal verification with combined data sets will always be in between the significance of the individual data sets. This already represents the statement that all experiments would necessarily see a signal.

This chapter was dedicated to the analysis techniques for cryogenic detectors. We first introduced the methodology of the RD analysis. This procedure can be categorized into the detector characterization, the event characterization, the event selection, and the validation of the analysis chain. We especially emphasized machine learning methods for data cleaning and pile-up separation. We will mention the former method again in Chap. 4 and come back to the latter method at length in Chap. 6, where we shall further refine it. The RD analysis results in the clean recoil energy spectrum of the measurement, the corresponding cut efficiency, and auxiliary quantities. Using this data, we can conduct a HLD analysis consisting of PD and statistical signal searches, for which we introduced the LH ratio test and Yellin's methods. Finally, we discussed the challenges when combining data sets for signal searches. The last studies will become relevant for future large-scale setups, such as the CRESST upgrade and the final setup of the COSINUS experiment. These experiments are the subject of the following chapters, where we will discuss the COSINUS experiment, its current status, and optimized detector designs in Chap. 4. In Chap. 5, we discuss the latest measurements and DM results from the CRESST experiment, and in Chap. 6 we discuss automation strategies for the detector setup and analysis of the CRESST upgrade.

Chapter 4

Characterization of remoTES prototypes for COSINUS

Experiments have achieved stronger and stronger sensitivity to low DM masses and interaction rates in the past decades. More than once, excesses appeared close to the detection thresholds [182]. Most of them turned out to be detector effects or unmodelled backgrounds. Some of these excesses are the subject of ongoing research, with continuously new insights published (see Sec. 2.2.3). However, one excess is especially puzzling: the claimed DM signal of the DAMA/LIBRA experiment. With high statistical significance, the experiment measures an annual modulation of its event rate compatible with that expected from DM scattering. That was considered for a long time a "smoking gun" evidence for a DM signal. The problematic fact about their results is that the parameter space compatible with their observations is severely excluded by other experiments. Currently no DM model exists that would explain a signal observed by the DAMA experiment, but not by any other experiment [39]. However, it is not excluded that DM could have non-trivial interactions with certain nuclei, which could not satisfyingly be modeled in our assumptions (see Sec. 1.3). Therefore, an independent verification of the DAMA/LIBRA excess, using the same target material, is one of the recommendations of the APPEC committee report [38].

Several experiments are dedicated to this mission. They were discussed in Sec. 1.4. Almost all of them are using single-channel readouts from scintillating sodium iodide targets. This leaves questions about the compatibility of light-QFs open, since not all experiments use identically grown crystals. The available measurements suggest that discrepancies between DAMAs QF and that of other experiments exist [46]. While calibration of these setup factors is possible, there is also an even more elegant way to resolve the question of QFs while acquiring more beneficial capabilities on the way: using a two-channel readout. The COSINUS experiment [45] is set to

instrument sodium iodide targets with TES while simultaneously measuring the produced scintillation light with a beaker-shaped light detector, fully surrounding the target. This design choice enables both an in-situ measurement of the QF and PD on event-by-event basis through the two energy scales (see also Sec. 3.2). Therefore the COSINUS experiment can reach much lower nuclear recoil energy thresholds, and in case of a verification of the DAMA/LIBRA excess, determine if it is of nuclear or electron recoil origin. In the case of nuclear recoil events, it would even be possible to determine whether the signal originated from scattering off sodium or iodine.

Manufacturing cryogenic detectors with sodium iodine introduces interesting questions in the detector design. This target material is naturally hygroscopic, which makes it impossible to evaporate films on the target directly. Furthermore, the low Debye temperature of sodium iodide requires fine-tuning the thermal properties of the sensor to the target. The originally planned detector design for COSINUS features unexpected problems due to the re-absorption of schintillation light in the carrier crystal, which was manufactured with another material than the target. The collaboration recently published a dedicated alternative design and works intensively on R&D to optimize their energy thresholds and reach their performance goal. Furthermore, the achievable threshold depends for phonon readouts strongly on the crystal size, which makes it harder to collect high exposure. The COSINUS experiment plans to operate ten identical detectors in a first run, and 20 in a second run. This is about the maximum number of detectors that can be operated in parallel without exceeding the available resources. We provide projections of the achievable energy threshold in this chapter, depending on the target mass, based on recent measurements. We show in our electrothermal response simulation (CryoEnv), also used in Sec. 2.2.2, which design choices can lead to achieving the performance goals of the experiment.

We discuss the necessary sensitivity to validate the DAMA/LIBRA claim and the performance goal of the experiment in Sec. 4.1. A conservative and preliminary analysis of two current measurements is discussed in Sec. 4.2. Our propositions for the final detector design that can reach the COSINUS performance goal are calculated in detail in Sec. 4.3.

4.1 Physics case and performance goal

The DAMA/LIRBA experiment, operated in the LNGS, has been taking data in its second phase since 2011, previously in its first phase, and previously under the name DAMA NaI. They use 250 kg sodium iodide target mass, distributed to 25 detectors,
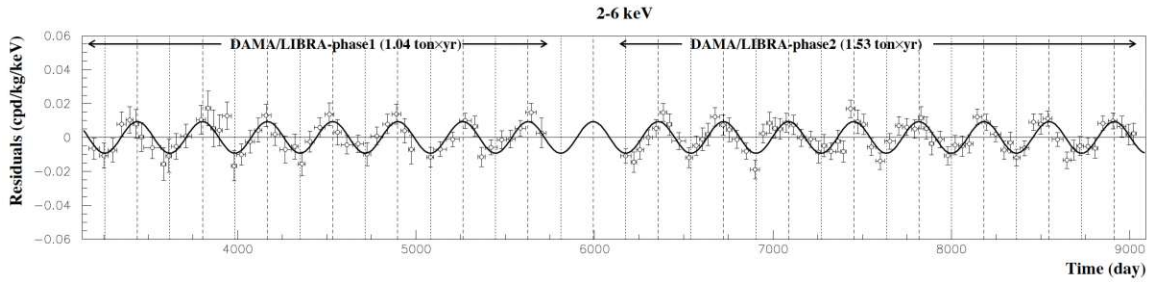
2-6 keV

Figure 4.1: Experimental residual rate of the single-hit scintillation events measured by DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 in the (2–6) keV$_{ee}$ energy intervals as a function of the time. The superimposed curve is the cosinusoidal functional forms $A\cos\omega(t-t_0)$ with a period $T=\frac{2\pi}{\omega}=1$ yr, a phase $t_0=152.5$ day (June 2nd) and modulation amplitude, A, equal to the central value obtained by best fit. Fig. and caption from Ref. [38].

and measure the produced scintillation light with PMTs. Their energy threshold was 2 keV$_{ee}$ in the first phase with a total exposure of 1.04 tonne years and 1 keV$_{ee}$ in their second phase with an exposure of 1.53 tonne years. In their single hit events, they observe an annually modulating signal between 2-6 keV$_{ee}$, peaking close to June 2nd. This phase of the annual modulation is consistent with what would be expected from Earth orbit through a DM halo. The rate of the part of their energy spectrum that is attributed to the modulation is shown in Fig. 4.1, with the constant event rate subtracted. Their total event rate is not publicly available.

An extensive sensitivity study for the COSINUS experiment was performed in Ref. [48]. They study three different scenarios: the standard assumptions of elastic DM-nucleus scattering with a Maxwell-Boltzmann distributed DM velocity, a generic falling DM spectrum, and finally, an arbitrary DM-nucleus scattering recoil spectrum. The necessary sensitivity to test the DM signal region is visualized in Fig. 4.2. In this work, we do not study the official COSINUS performance goal, corresponding to a event rate of 0.1 counts/kg/day in the signal region. Instead, we study the potential of the experiment in a background-free scenario. We include a note on how realistic this assumption is further below. This means we require less than 2 counts in the signal region and the whole measurement interval considered, which is compatible with a zero count hypothesis under Poissonian assumptions. We perform our projections for the assumption of a 10 GeV/c² DM particle mass, which is the mass corresponding to the lower DAMA island, from a hypothesis of DM scattering elastically on sodium. The statements hold qualitatively for higher and lower WIMP masses (see also corresponding plots in Ref. [48]). The threshold projections are based on the analysis done in Sec. 4.2, and are calculated in Sec. 4.3.
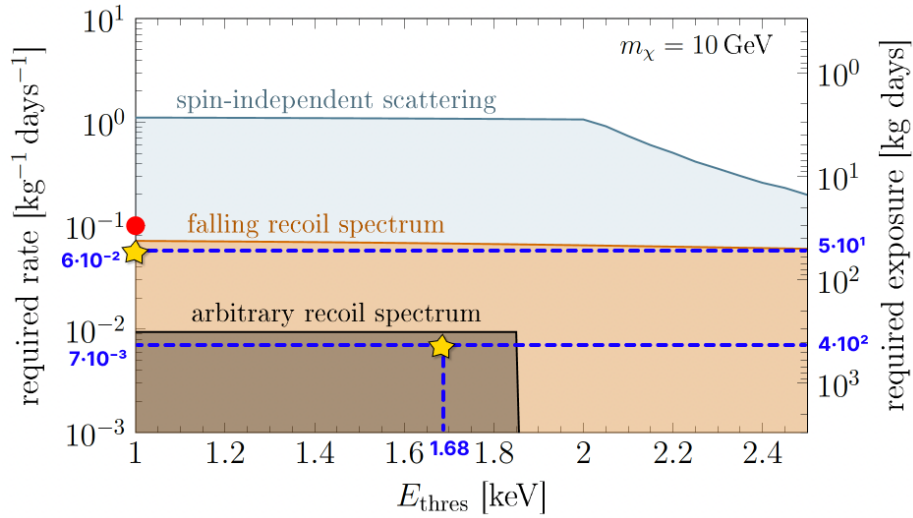
137

Figure 4.2: COSINUS exclusion power, defined as the bound on the total rate (or equivalently the total exposure with zero observed events) that COSINUS must achieve for excluding DAMA in a halo-independent way, as a function of the assumed threshold in COSINUS for different DM masses. More general assumptions correspond to weaker exclusion power, meaning that stronger bounds are necessary to achieve an exclusion. All bounds are derived in Ref. [48]. The red dot indicates the design sensitivity of COSINUS. We added the golden stars and blue lines, they visualize the achievable threshold and sensitivity that we project in two scenarios in this work. Fig. and caption from Ref. [48], with adaptions.

We assume the COSINUS experiment will collect data in two run, that will last for one year each. The goal of the first run is to test the standard spin-independent scattering scenario, the second run should be sensitive to arbitrary recoil spectra.

We make conservative assumptions about the technical performance of the experiment, namely that 7(14) out of 10(20) detectors that are built in the final setup of run 1(2) can operate with an overall efficiency in the signal region of 80 % (accumulating effects of cut efficiency, interruptions of the measurement, etc.). We assume that energy thresholds are set at 6.5 times the energy resolution, ≈ 1 noise triggers/kg/day are expected from this threshold. However, noise upward fluctuations of this magnitude are typically easily identifiable with standard pulse-shape cuts. We therefore assume that no noise triggers survive in the signal region.

According to our projections, the experiment should achieve a nuclear recoil threshold of 1 keV or lower in the first run, using 29.4 g crystals for each of the ten detectors. With this target mass and the assumed efficiency, an exposure of 50 kg days could be collected in 11 months. Assuming that the count rate in the signal region stays below the required 3 counts (equivalent to 0.06 counts/kg/day), this scenario

would reach the performance goal for the first run.

For the second run, we project a threshold of 1.68 keV, using a 99.1 g target for each of the 20 detectors. A measurement time of one year would lead to a net exposure of 400 kg days, which is sensitive to arbitrary recoil spectra, assuming a signal candidate count rate below 0.007 counts/kg/day.

Ref. [48] only considers cut and count experiments, neglecting any spectral information within the ROI. Including this information could potentially lead to stronger sensitivity. Also, analysis of the modulation, not only the total rate, can lead to much stronger results. Our performance projections are, in that sense, the most conservative analysis.

There is one strong limitation about this type of analysis: non-discriminable backgrounds. In this work, we do not treat the expected EM interaction rate. Due to the strong discrimination power of the two-channel readout for EM events in the signal region, we can reasonably assume that no electron scattering events would be interpreted as signal candidates. The muon flux and their decay products are the subjects of simulation studies [183], where the ambient neutron flux inside the passive shielding was calculated to be ≈ 0.035 counts/year. An active muon veto will identify muon events, reducing the cosmogenic neutron flux to ≈ 0.11 counts/year [183]. The potential cosmogenic activation of the crystals and the cryostat materials are the subject of ongoing screening and simulation campaigns. Generally, known and previously measured backgrounds can be included in a more sophisticated analysis procedure, see e.g. 3.2 for methodology.

4.2 Performance of sodium iodide remoTES prototypes

The initially planned COSINUS design was composite: the TES was evaporated on a small carrier crystal made from a different material such as calcium tungstate, which was then glued to the target sodium iodide crystal. This detector concept reached suitably low thresholds in the earlier runs of CRESST-II [154], where the same carrier and target materials were used. However, this concept turned out to be problematic since the usable carrier crystal materials are not fully transparent for the scintillation photons from sodium iodide, with a maximum of its emission spectrum at $\lambda = 415$ nm / ≈ 3 eV. The scintillation light from the target was to a large extend reabsorbed in the carrier crystal, degrading the discrimination power and energy resolution. The
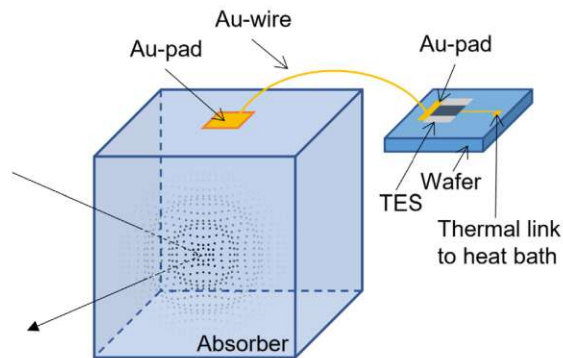
Figure 4.3: Schematic view of a remoTES detector. The wafer where the TES is fabricated onto (right) is separated from the absorber crystal (left). This crystal is equipped with a gold pad which is linked directly to the TES via a gold bonding wire. Fig. and caption from Ref. [47].

phonon signal in the composite design was already relatively small and masked by the reabsorbed light signal. This problem was intensively studied in Ref. [184].

An alternative and suitable design choice was proposed in Ref. [114]. The TES is placed on a remote wafer and is connected to a gold pad on the target through a gold bond. A particle recoil in the target induces a temperature increase in the gold pad. Ideally, the electron system in the gold pad is strongly coupled to the electron system in the TES. This would lead to a comparable temperature increase in the TES as in the gold pad. Gluing a gold pad on the sodium iodide crystal is a much simpler procedure than evaporating or sputtering directly on the crystal. COSINUS performed the first successful measurement of this design, and we refer to it under the name "remoTES" [47]. An image of the situation is shown in Fig. 4.3.

Recently several results were published from remoTES detectors with 3.7 g sodium iodide targets. Ref. [49] contains the report from a measurement in an above-ground laboratory in Munich at the MPP in December 2021, collecting a data set of 9.06 g days and an analysis threshold of 15 keV for nuclear recoils. This measurement showed PD in a sodium iodide target on event-by-event basis for the first time. At the same time, Ref. [50] reports from a measurement in the CRESST test cryostat in the Gran Sasso underground laboratory (3800 MWE), internally referred to as run 376, in July 2022. They collected 11.6 g days of data with an energy resolution of 441 eV and an analysis threshold of 4 keV. They further performed the first DM search with a cryogenic sodium iodide detector. Under standard assumptions, the resulting limit on the cross-section is less than two orders of magnitude above the DAMA-compatible region. This result bears evidence of the potential of the COSINUS experimental
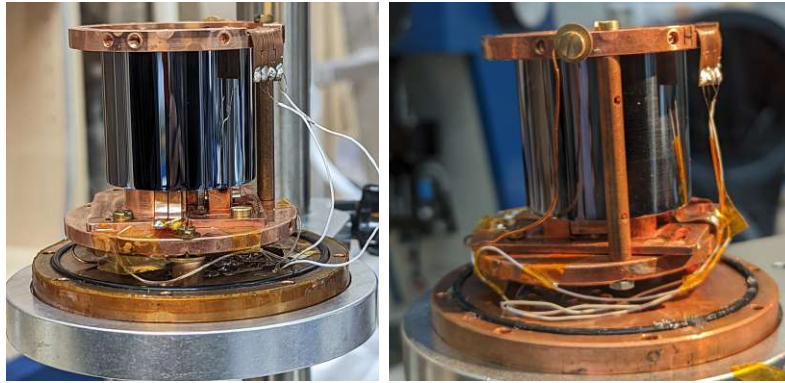
Figure 4.4: Images of the detectors used to measure COSINUS remoTES prototypes in run 376 of the Munich R&D lab in spring 2023 (left) and the run 599 of the CRESST test cryostat in the LNGS underground laboratory (right). The black beaker is the light detector, fully surrounding the remoTES phonon detector.

design. Assuming the limit to scale approximately linearly with the collected exposure, the sensitivity could already, with a few kg days of data, fully exclude the DAMA-compatible region under standard assumptions. Both measurements were done with a two-channel readout and PD.

For this work, we performed a dedicated and conservative analysis of two measurement with sodium iodide remoTES. First, we do an individual analysis of the phonon channel of run 376. Second, we analyze another measurement that was done in the MPP above-ground laboratory, dubbed run 599, in spring 2023. The latter one was also performed with a 3.7 g sodium iodide target. We show images of the operated detector in Fig. 4.4. In both measurements (run 376 and run 599), a lead castle was built around the cryostat, shielding from EM particle interactions, e.g. induced by muons and radioactivity.

The goal of our analysis was to learn the physical properties of the detector. We want to extract SEVs, NPSs, an energy resolution, and a rough energy spectrum that enables us to identify strong features. In the following (see Sec. 4.3) we will use this information to simulate detectors with larger targets and optimized components and operation conditions. We chose these two measurements because they had comparably good results, but with different choices in the detector design. While run 376 featured a relatively good athermal phonon collection efficiency, the detector used in run 599 was optimized towards measuring the temperature of the target crystal. This was realized by a thicker gold pad and a thick glue layer below the pad. More details of the detectors are discussed in Sec. 4.3. We will learn in that section, that the thermal readout scheme, which is realizable with current fabrication methods, is suitable for

achieving the performance goal of the first COSINUS run, while the athermal readout scheme is required for the performance goal of the second COSINUS run.

We took for our analysis exclusively the online triggered data from a measurement time of 63 h (run 376) and 21 h (run 599). The sampling frequency was set to 50 kHz (run 376) and 200 kHz (run 599), and the length of the record window to 32768 samples (both runs). An iron-55 source was mounted inside the detector housing in both runs, shining on the target. We performed quality cuts on the pulse shape, which rejected the majority of artifacts and strongly saturated pulses. After cuts, the surviving number of events is 4385 (run 376) and 4219 (run 599). The SEVs of target hits and TPs after quality cuts are shown in Fig. 4.5.

The PHs were estimated with the OF for particle and TPs, respectively. The energy calibration was done by interpolating between TPs' PH and TPA values (see Sec. 3.1). The CPE factor was extracted from the position of the iron peak in the TPE spectrum. The spectra are plotted in Fig. 4.6. The iron line is for both data sets close to the threshold and in an approximately linear region of detector response. The pulses in the upper part of the energy spectrum show slight saturation effects, which introduce an uncertainty of up to $\approx 30\%$ on their energy estimate. No such uncertainty is expected for small pulses and the threshold and resolution estimates.

We estimate the baseline resolution by applying the fit model of Gaussian noise trigger maximal that was discussed in Sec. 3.1.1.5. The number of available clean, empty baselines is 916 (run 376) and 1721 (run 599). The resulting fit, including the estimate of a trigger threshold for 1 noise trigger/kg/day, is shown in Fig. 4.7. The voltage units of the PHs are converted to recoil energies with the interpolator function from the energy calibration. This procedure finds energy resolutions of 0.33 mV / 344 eV (run 376) and 1.33 mV / 410 eV (run 599). The estimate from run 376 agrees, within the expected systematic errors from different methods of calculating the resolution, with the estimate from Ref. [50].

We did not simulate a cut efficiency since we don't intend to state any physics results with these data sets for which the rate would be of importance.

In summary, we have found comparable energy resolutions with two data sets, which furthermore agree with the resolutions calculated in previous literature. In the pulse shapes of the SEVs, we see that several choices in the detector design were different enough to impact the thermal processes. While in run 376 the athermal part of the pulse shape dominated, the sensitivity of run 599 was driven by the thermal part. In the following section, we will use these two measurements to project achievable
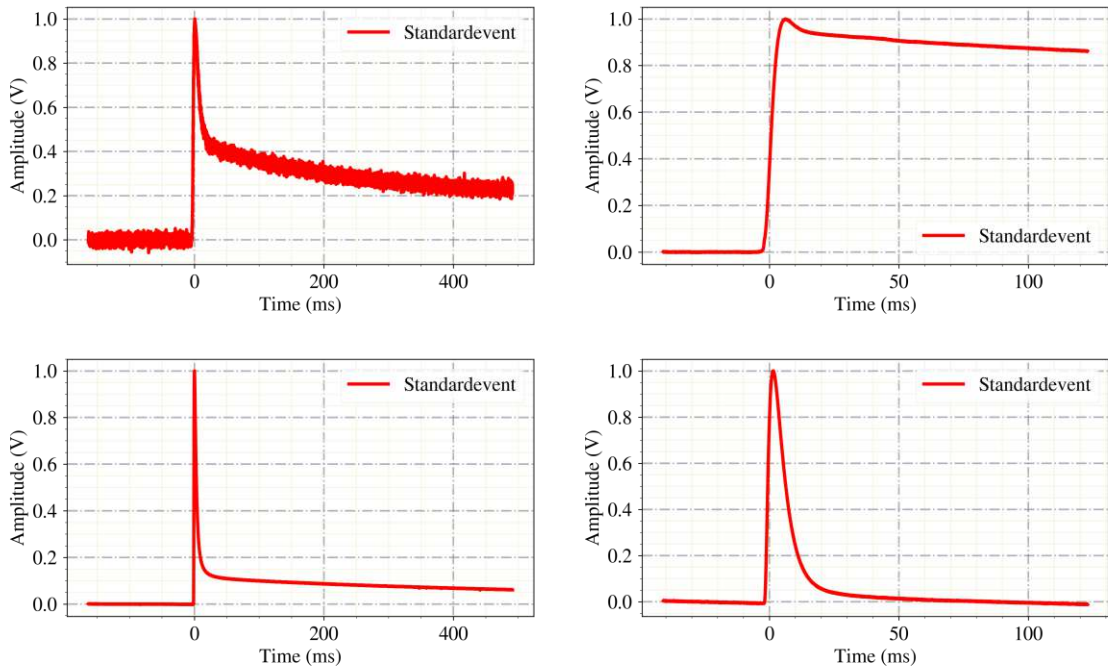
Figure 4.5: The SEVs for target hits (upper row) and TPs (lower row) for run 376 (left) and run 599 (right). The athermal part of the pulse shape is more pronounced in run 376, while the pulse shape of run 599 is clearly dominated by the thermal part of the signal. The heater was placed on the wafer, close to the TES, suppressing the thermal part of the TP shape, which measures the temperature of the absorber crystal.
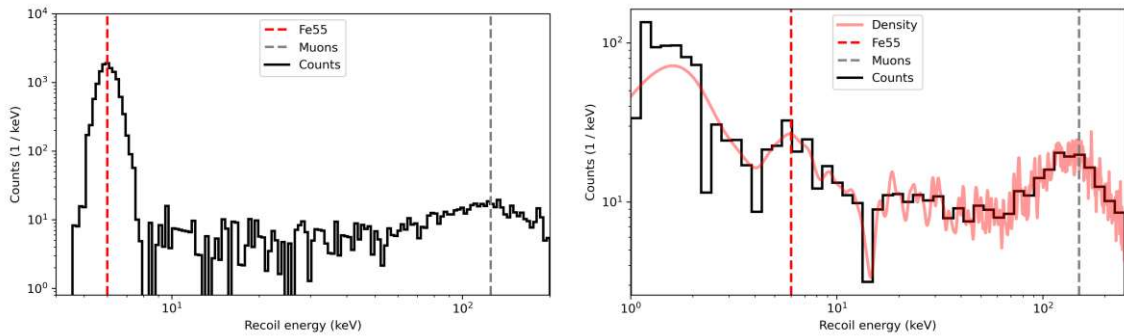

Figure 4.6: The energy spectra of run 376 (left) and run 599 (right), with highlighted positions of the iron lines (red dashed) and the typical bump-ish feature induced by muons (grey dashed). The rising event distribution below the iron line in the run 599 spectra are majorly noise triggers. Due to a higher trigger threshold, such noise events are not visible in run 376. Additionally to the histogram of recoil energies of run 599, we show a density plot that highlights the peak induced by X-ray events from the iron source.
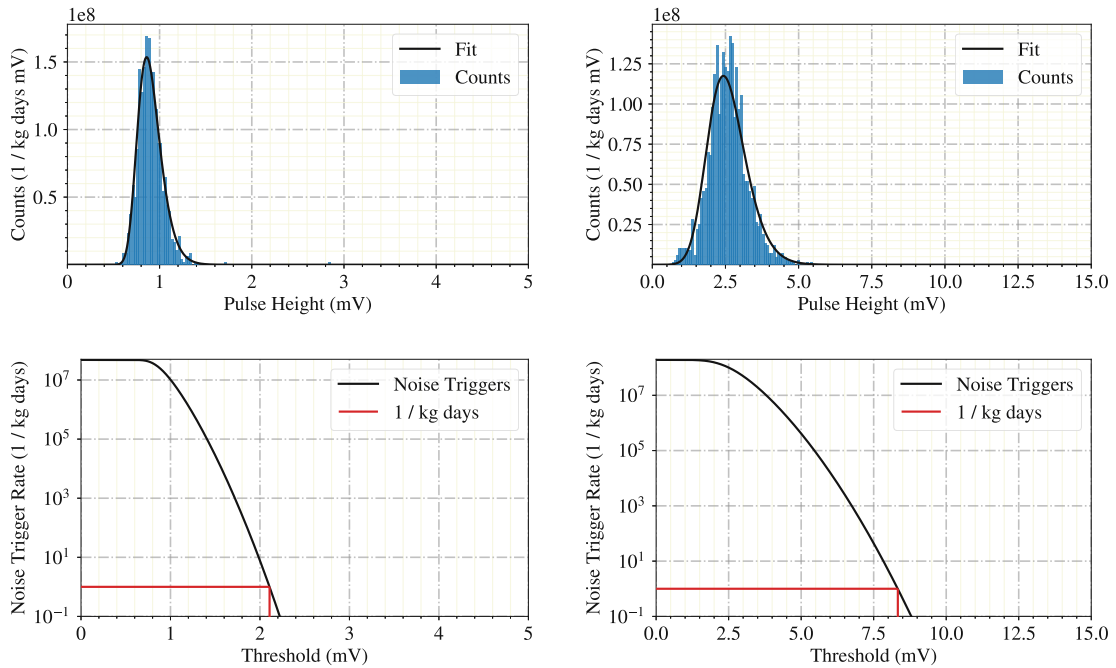
Figure 4.7: Fit to the noise trigger maxima after application of the OF for run 376 (left) and 599 (right). The top panel shows a histogram of the noise trigger maxima and function with the maximum LH estimates from an unbinned fit. The lower panel shows the expected rate of noise triggers, depending on the chosen energy threshold. The black line in the lower panel is the integral of the black line in the top panel, from the respective PH and to the end of the window, and scaled to the exposure. The recommended trigger threshold for 1 noise trigger/kg/days exposure is highlighted (red).

thresholds for two different choices of detector design optimized for the thermal or athermal signal components.

4.3 Propositions for the final detector designs

The remoTES design consists of more components than a standard detector design, as we introduced in Chap. 2. Therefore, a full electrothermal model should individually contain components for the wafer phonons and the electrons in the gold pad, bond, and TES. Furthermore, the heat inputs from the heater and target recoils have more differences than in the standard design, as the heater is placed on the remote wafer. However, the thermal couplings between these components would be hard to determine from measurements, and the vastly larger number of parameters would not contain more actionable intelligence than a simple 2-component model. For the purpose of this work, we introduce several simplifications:

1. we assume that the electron systems of the gold pad and tungsten TES, and the bond connecting them, are much stronger thermally coupled to each other than to the phononic systems of the wafer, absorber, and the thermal link between TES and heat bath. We assume that the temperature is uniform in all these metallic components. While this assumption is certainly not fulfilled in practice, the electron temperature can also be interpreted as an effective temperature, determining the resistance of the TES and absorbing all geometric effects in the other constants.

2. we ignore the wafer as a thermal component by assuming that the thermal link between TES and wafer is either frozen out or can be absorbed in an effective coupling between TES and the heat bath.

3. we assume the same base temperature of crystal and wafer. Since the wafer is heated, this is very unlikely. However, the absolute temperature of the crystal has no impact on the results, only its rise due to pulses.

Due to these assumptions, we can again use the CryoEnv simulation that was already used in Sec. 2.2.2. In the first step, we use the simulation to determine several of the thermal parameters. Other parameters we know a-priori or from theory. We summarize the determination of the parameters in the following. Once found, we use the simulation to scale the targets to the larger size that is required to reach the exposure goal of COSINUS and optimize the components of low thresholds. We will see that the performance goal is indeed achievable.

Many of the values necessary for a proper detector response simulation are known or can be reasonably estimated. The values are summarized in Tab.4.1. Most of them

are explained in Chap. 2. The value V_{set} is the parameter for a proportional-integral-derivative (PID) controller, regulating the DAC value such that the CPH stays close to this value. If the CPs drive the TES in its normal conducting region, the OP in the transition curve can be determined from the knowledge of the V_{set} value, the shunt resistance R_S, the SQUID conversion factor η and the bias current I_b. The rise and decay time of heater pulses is not the observed rise and decay time of the voltage trace corresponding to TPs in the data but the time constants of the injected pulse signal driven through the heater. The decay time of the injected pulse is expected to match the rise time of observed TPs.

Quantity	Run 376 (summer)	Run 599 (spring)
size sodium iodide	(10 x 10 x 10) mm³	(10 x 10 x 10) mm³
size gold pad	1 μm x 1.77 mm²	8 μm x 1 mm²
size gold bonds	1 bond, 2.3 mm, 17 μm diameter	2 bonds, 1 cm, 17 μm diameter
size tungsten TES	(400 x 100 x 0.08) μm³	(400 x 100 x 0.24) μm³ (est)
size thermal link	(1 x 200 x 100) μm³ (est)	(1 x 200 x 100) μm³ (est)
bath temperature T_b	15 mK (est)	7 mK
shunt resistance R_s	40 mΩ (est)	40 mΩ
coil inductance L	1.3 μH (est)	1.3 μH
heater pulse rise time	0.5 ms	0.5 ms
heater pulse decay time	3.0 ms	3.0 ms
SQUID conversion factor η	6.67 V/μA (est)	16.67 V/μA (est)
bias current I_b	1 μA	1 μA
V_{set}	0.9 V	2 V
PH iron line	0.0057 V	0.02 V
PH TPs	0.23 V ($TPA = 1$ V)	0.35 V ($TPA = 0.6$ V)

Table 4.1: Table of known values of the cryostat, electronics, and detector components. Some of the values are known precisely, others are reasonable estimates, marked with (est).

Further parameters were extracted from the data. This is a step-wise process and not trivial. Some parameters are strongly correlated. In such cases, we tried to determine conservative estimates.

We calculated the heat capacities of the TES and the gold pad with the Sommerfeld model and the absorber heat capacity with the Debye model using the physical constants from Tab. 2.3 and Tab. 2.4.

The time constants of the pulse shape were determined by fitting the analytic pulse shape model 2.13 to the SEVs.

An initial guess of the thermal couplings can be extracted from numerically inverting the equations for the pulse shape time constants derived in Sec. 2.1. They need manual adjustment to reproduce the observed pulse shape.

The transition temperature and the steepness of the transition curve were adjusted to match the measured data. Measurements of the transition curve steepness are known to have strong systematic uncertainties due to bias heating and systematic uncertainties in the measurement setup. The steepness is, therefore, also adjusted to match the measured NPS. This parameter is also correlated with the SQUID conversion factor η. Further parameters extracted from the NPS are the magnitudes of the adjustable noise components.

ϵ, β, DAC, and R_h were determined to match the PH of particle and TPs and the OP in the transition curve.

The resulting values are contained in Tab. 4.2.

We use the determined parameters to simulate the response of detectors. The transition curve, a pulse from a 6 keV X-ray event from an iron-55 source, and the NPS of the simulated and measured data are shown in Figs. 4.8 (run 376) and 4.9 (run 599). While the simulated and measured data agree mostly, there are deviations in the steepness of the transition curve and the athermal part of the pulse shape from run 599. Also, the OPs calculated from the V_{set} did not agree very well with the otherwise observed data and were adjusted to match the measurements. Since these quantities do not strongly impact the purpose of this work, projecting the changes in the sensitivity when the target size is increased, we can neglect these deviations.

As discussed in Sec. 4.1, the first COSINUS run requires to collect 50 kg days of exposure, and the second 400 kg days of exposure. To reach this goal, larger targets are needed than in the R&D runs. A challenge in lowering the threshold is the heat capacity of the gold pad on the crystal, which directly decreases the observed PH for a recoil with a given energy. A thinner gold pad would lead to a lower heat capacity, and a thickness of < 100 nm is a common value for phonon collectors and TES e.g. in the standard CRESST detector design. However, the COSINUS collaboration has currently not developed a technique to sputter or evaporate directly on the sodium iodide crystals, and the gold pad is typically glued to the crystal. This gives a practical lower limit for the thickness of the pad of ≈ 1 μm. We will therefore consider two scenarios.

Quantity	run 376 (summer)	run 599 (spring)
$C_{f,nc}$ (pJ/mK)	$1.53 \cdot 10^{-5}$	$6.02 \cdot 10^{-5}$
C_{au} (pJ/mK)	$3.55 \cdot 10^{-3}$	$1.87 \cdot 10^{-2}$
C_a (pJ/mK)	0.23	0.37
G_{eb} (pW/mK)	0.46	0.3
G_{ab} (pW/mK)	0.15	0.04
G_{ea} (pW/mK)	0.51	18
τ_n (ms)	2.42	2.44
τ_{eff} (ms)	4.44	2.80
τ_t (s)	0.6	1.2
τ_{TP} (ms)	3	3
ϵ	0.04	0.15
R_H (Ω)	10	10
$R_{f,op}$ (Ω)	0.055	0.12
R_{f0}	70 mΩ	290 mΩ
T_c (mK)	27.85	32.6
k (1/mK)	2.86	3.1
β	0.1	0.1
I_H (μA)	2	2
i_{sq} (pA/$\sqrt{\text{Hz}}$)	1.2	1.2
E_J	2	3
$\frac{\Delta R_{f,flicker}}{R_{f0}}$ (pJ)	$9 \cdot 10^{-3}$	$1 \cdot 10^{-2}$
α	2.5	1.

Table 4.2: Physics parameters for the sodium iodide remoTES detectors operated in run 376 and 599 that were calculated from theory or extracted from the data.

First, we will show that the required threshold is achievable with a (2 x 2 x 2) cm³ crystal, with the already accessible techniques (gluing the gold pad), by optimizing for the thermal part of the pulse shape. We will use the measured data from run 599 and our simulation to optimize the component sizes and measurement conditions. This detector design is sufficient to reach the goal of 50 kg days of exposure within the experiment's first run, with a planned measurement time of one year.

Second, we assume that R&D efforts continue during the experiment's first run and that depositing a gold pad with a thickness of ≈ 100 nm on the sodium iodide crystal becomes feasible for the second run of the experiment. We show then that the required threshold is achievable with a (3 x 3 x 3) cm³ crystal by optimizing for athermal phonon collection. With this target size, the exposure goal of 400 kg days for the second run of the experiment is achievable within 2 years of operation. We use
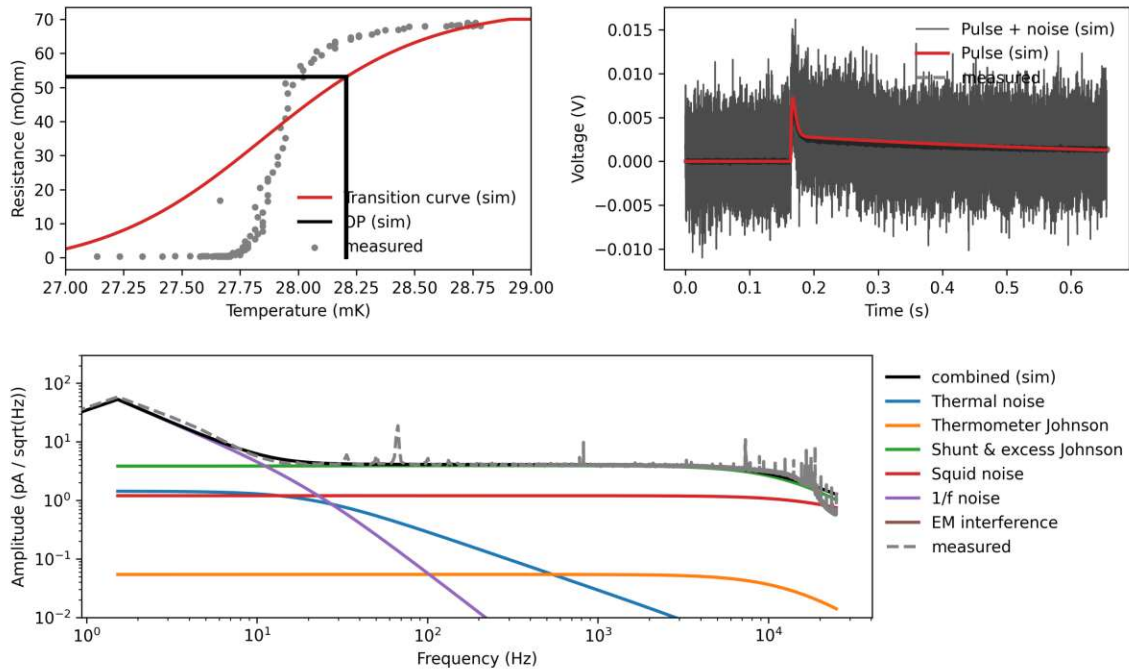
Figure 4.8: Comparison between simulated and measured 6 keV X-ray event in the run 376 remoTES sodium iodide detector. (upper left) Measured (grey dotted) and simulated (red) superconducting transition curves and the OP within them. The discrepancy between the curves is due to the measured curve's peculiar shape and does not impact our projections. (upper right) A measured SEV, scaled to the height of measured 6 keV events (grey dashed), almost perfectly overlapped by the simulated pulse (red) and the simulated pulse with superposed noise (grey). (lower panel) Simulated noise contributions (colored) and total NPS (black) overlapped by the measured NPS (grey dashed). We neglect the EM interference contributions.

the measured data from run 376 for these projections.

To account for systematic uncertainties in the determination of the physical quantities from runs 376 and 599, and in our simulation, we make several conservative assumptions: we assume that the coupling between absorber and bath, which is purely parasitic for the sensitivity, scales with the surface of the absorber. Furthermore, we assume, similarly to Sec. 2.2.2, that the absorption efficiency of athermal phonons in the gold film scales with the temperature. While this is very unlikely to be a real effect, we stick to the assumption to account for other systematic uncertainties, which might only become apparent once the final designs are measured.
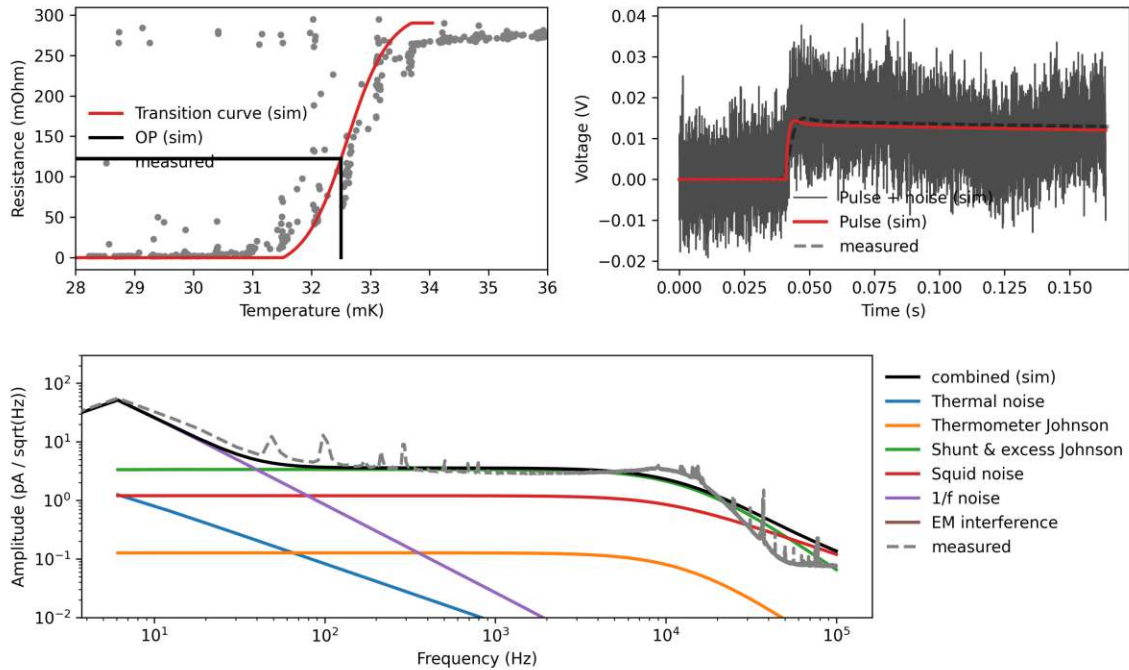
Figure 4.9: Comparison between simulated and measured 6 keV X-ray event in the run 599 remoTES sodium iodide detector. (upper left) Measured (grey dotted) and simulated (red) superconducting transition curves and the OP within them. The grey dots far outlying from the curve are measurement artifacts. (upper right) A measured SEV, scaled to the height of measured 6 keV events (grey dashed), overlapped by the simulated pulse (red) and the simulated pulse with superposed noise (grey). The discrepancy in the rise time is due to a smear-out when creating the SEV. We will be using mostly the thermal component of this measurement, and any uncertainty in the athermal part and rise has little impact on our projections. (lower panel) Simulated noise contributions (colored) and total NPS (black) overlapped by the measured NPS (grey dashed). We neglect the EM interference contributions.

4.3.1 Optimizing for a temperature measurement of the absorber

Our simulation using the physics parameters from run 599 gave a resolution after optimal filtering of 482 eV. This is slightly worse than the measured value of 410 eV and, therefore, suitable for a conservative estimate.

We optimized the detector components for scenarios with target sizes of 1, 2 and 3 cm^3, within our CryoEnv simulation. To optimize the threshold, we successively performed the following steps in the simulation. We first lowered the operation temperature. A lower operation temperature is generally better for the sensitivity of the detector. It can practically be challenging to achieve lower transition temperatures of the TES but transitions as low as 10 mK were measured before for CRESST detectors. Since it is not yet clear is the dry COSINUS cryostat of the final setup will reach as low base temperatures as the wet CRESST cryostat, we made the conservative assumption that transition temperatures of 17 mK can be achieved. With the updated physics parameters of the detector, which scaled with the temperature as introduced in Sec. 2.2, we then optimized the dimensions of the gold pad. A larger gold pad adds parasitic heat capacity but also increases the collection efficiency of phonons and prolongs the relaxation time of the TES τ_{in}, which contributes to a calorimetric operation mode (see Sec. 2.2.2).

Finally, we argue that in run 599 the record window was chosen very short. We performed a simulation with Gaussian, white noise and showed that a lower record window would allow for better filtering with the optimal filter and a further decrease in the energy resolution. This can only be tested in an underground measurement since the high background rate above ground would lead to significant losses due to pile-up.

The impact of lowering the temperature and increasing the gold pad size is shown in Figs. 4.10 and 4.11 for multiple possible target sizes. The energy resolutions achieved for the target size of interest with 29.4 g are 0.85 keV after lowering the temperature and 0.34 keV after additionally matching the gold pad size.

The optimal dimensions of the gold pad, depending on the dimensions of the cubic target crystal, are an interesting question. One could assume that the best choice is to scale them proportionally to each other. However, the thermal dynamics of the system are more complicated. We show in Fig. 4.12 the optima of the gold pad dimensions found in our simulation for the three target weights, corresponding to side lengths of the cube of 1, 2, and 3 cm. It is clearly visible that the points deviate strongly from a linear scaling behavior. The reason is that the size of the gold pad changes multiple
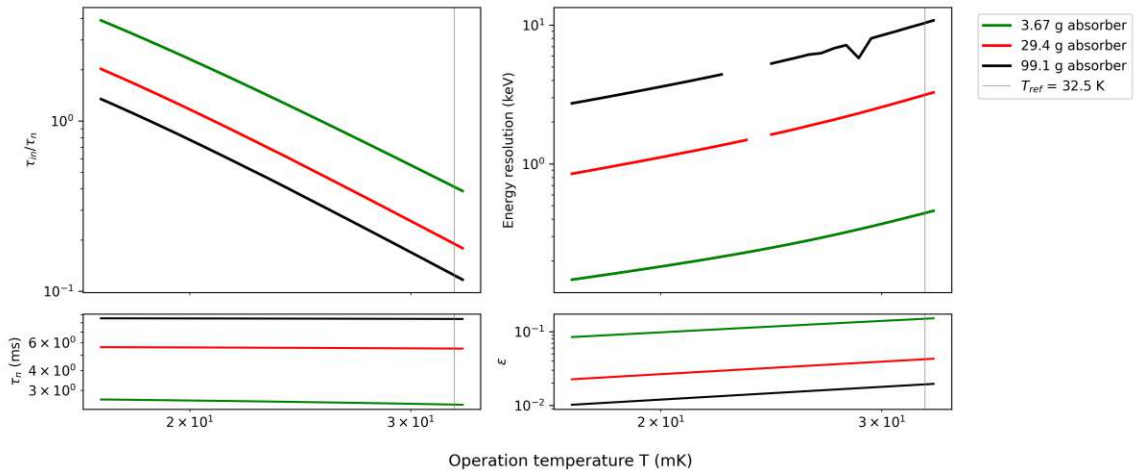
151

Figure 4.10: Impact of a changing operation temperature on the calorimeterness (upper left, see Sec. 2.2.2 for Def.), the energy resolution (upper right), the phonon thermalization time (lower left), and the collection efficiency (lower right), for the three scenarios of target sizes (green, red, black). The missing points and the outlier in the energy resolution curves are non-physical artifacts of the simulation and can be ignored. Based on measured data from run 599.
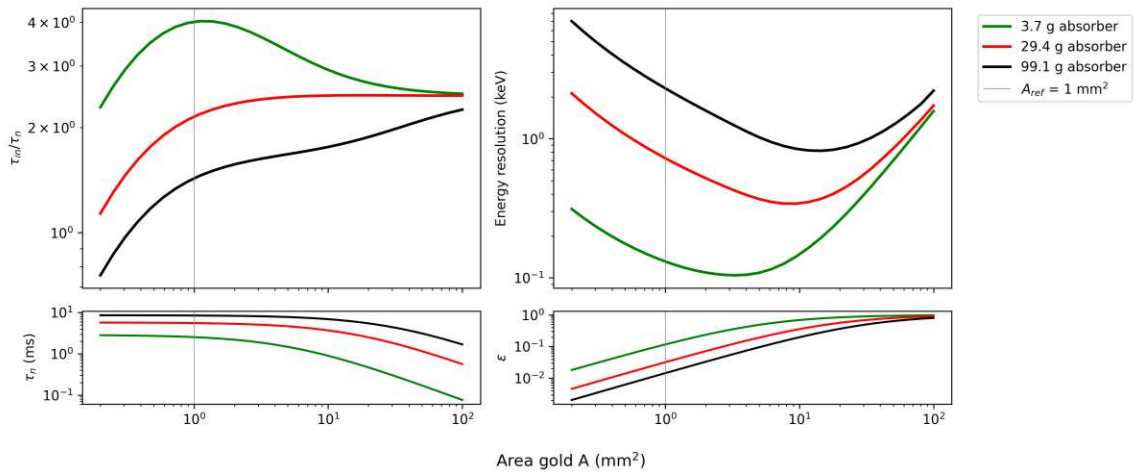

Figure 4.11: Impact of a changing area of the TES on the calorimeterness (upper left), the energy resolution (upper right), the phonon thermalization time (lower left), and the collection efficiency (lower right), for the three scenarios of target sizes (green, red, black). Based on measured data from run 599.

physics parameters in the model: the athermal collection efficiency is a function of the competing thermalization of phonons in the pad and the crystal; the relaxation time of the TES and, with that, the calorimeterness of the detector; finally, the heat capacity of the pad. The effects are partially counteracting and generally not linear.

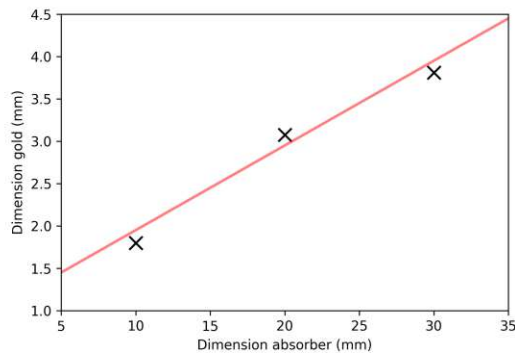Our final intervention is prolonging the record windows. Very short record windows

Figure 4.12: Optimal dimensions of a quadratic gold pad, all other parameters kept constant depending on the dimensions of a cubic crystal. The linear regression is a guide for the eye, visualizing the non-linear dependency. Based on measured data from run 599.

were used in run 599 since the high rate above ground would have created significant losses due to pile-up. The energy resolution after optimal filtering can intuitively be understood to scale with the area of the pulse covered within the record window. To test this hypothesis, we made a simulation of pulses as they were measured in run 599 and superposed with Gaussian noise. We test how sensitivity changes with the length of the record window by densely simulating such pulses and measuring the energy resolution. The results are shown in Fig. 4.13. We observe that the energy resolution for run 599 could have been a factor 0.4 better if the record windows were chosen much longer than the relaxation time of the crystal. We conservatively assume the same factor for the optimized design we calculated above. The improvement would likely be even better since we measured the energy resolution with the record window length from run 599, despite the pulses in the optimized design being longer. With the conservative assumption, we still reach an energy resolution of 136 eV after all interventions for the 29.4 g target crystal.

A simulated pulse with the optimized design is shown in Fig. 4.18. The time constants of the pulse shape are $\tau_n = 4.7$ ms, $\tau_{in} = 11.5$ ms, and $\tau_t = 2.78$ s.

In summary, we projected a conservative estimate of the achievable energy resolution with a value of 136 eV, using a 29.4 g crystal and measuring its temperature. It is, therefore, very realistic that the performance goal, assuming a threshold of 6.5 times the resolution, for the first run of the COSINUS experiment can be reached with already accessible and tested methods. An interesting detail for the reproducibility of the run 599 measurement is that a thick glue layer was used to increase the thermal coupling between the gold pad and the crystal. Differently from that in run 376, a
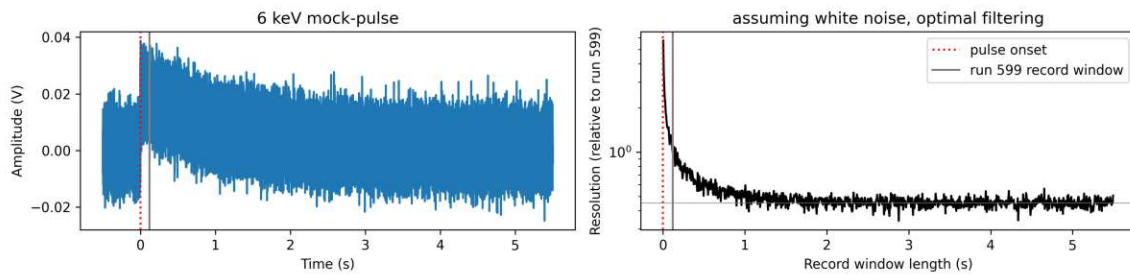
Figure 4.13: Dependency of the energy resolution on the record window. (left) A simulated 6 keV X-ray pulse with the parameters from run 599, with white, Gaussian noise superposed. The end of the used record window (grey dotted) is close to the onset of the pulse (red dotted). (right) Energy resolution after optimum filtering, depending on the length of the record window, normalized to the energy resolution from run 599. The data is based on simulated and optimum filtered pulses. The run 599 record length is marked with a horizontal grey line. The best achievable energy resolution with optimum filtering is a factor 0.4 lower and marked with a grey horizontal line.

thin glue layer was used.

There are several uncertainties that will potentially need to be dealt with. It is possible that electron recoils and events in the wafer crystal can induce a high pile-up rate, also below ground. For this case, we recommend using the dedicated method presented in Ref. [8] and summarized in Sec. 3.1.2.6 to separate pile-up events.

Another uncertainty stems from the behavior of low frequent noise, which is hard to predict. We believe that we overestimated the steepness of the NPS's rise towards lower frequencies in our analysis of run 599. This phenomenon is commonly occurring for short record windows and is likely an edge effect. It is also possible that sub-threshold events are mixed in the baselines we used to calculate the NPS, creating a steeper rise since the cutoff frequency of the pulse frequencies is very low. Comparing the steepness of the rise with NPS from other sodium iodide detectors, we expect that the lower frequencies do not rise too strongly, such that our estimates of the sensitivity increase due to a longer record window are legitimate. Furthermore, is was observed in previous measurements that a lower operation temperature also lowers the low frequent noise contribution. We did not account for this effect, since there is no suitable model to simulate it.

However, both these uncertainties would be eliminated if pulses were shorter, as in a design optimized for athermal phonon collection. Such a design will be discussed in the next section. A summary of the energy resolutions after interventions in our simulation is contained in Tab. 4.3.
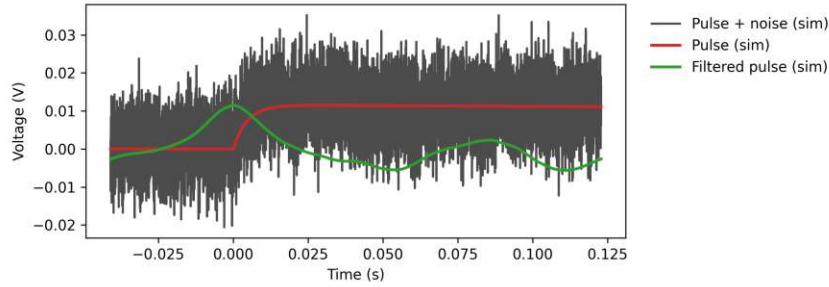
Figure 4.14: Simulated 6 keV X-ray pulse with the optimized design for measuring the absorber temperature. The thermal signature is shown in red, with superposed noise in grey. The pulse after optimum filtering is shown in green. The waves at the end of the record window are artifacts from the edge of the record window. A longer record window (as discussed in the text) would prevent such edge effects.

Energy resolutions (keV) with ...	3.7 g target	29.4 g target	99.1 g target
...lower temperature ($T_{op} = 17$ mK)	0.15	0.85	2.72
...optimal gold pad size	0.1 (d = 1.8 mm)	0.34 (d = 3.1 mm)	0.82 (d = 3.8 mm)
...longer record window ($>> \tau_t$)	0.04	0.136	0.328

Table 4.3: Energy resolutions of a remoTES optimized to measure the temperature of the absorber crystal after certain interventions. Based on data from run 599. The values in each row also include all interventions from the rows above.

155

4.3.2 Optimizing for athermal phonon collection

Modern thermometer-based low-temperature detectors do not measure the temperature of the target but collect the initial wave of athermal phonons, causing a stronger temperature increase in the sensor with lower heat capacity. The sensor must be weakly thermally coupled to the target for this concept to work, such that the heat does not easily flow back from the sensor to the crystal. This is done by freezing out the electron-phonon interaction that scales $\propto T^5$. The collection of athermal phonons works despite a weak electron phonon coupling since the athermal phonons can cause quasiparticles in the metal. We projected energy thresholds for such a detector design based on the data from run 376. The parameters from run 376 provide an energy resolution of 397 eV in our simulation. This value is higher than that from our analysis (344 eV) and closer to the official value from Ref. [50]. The target size required for reaching the performance goal of the second run of the COSINUS experiment is a cube with a 3 cm side length that weighs 99.1 g.
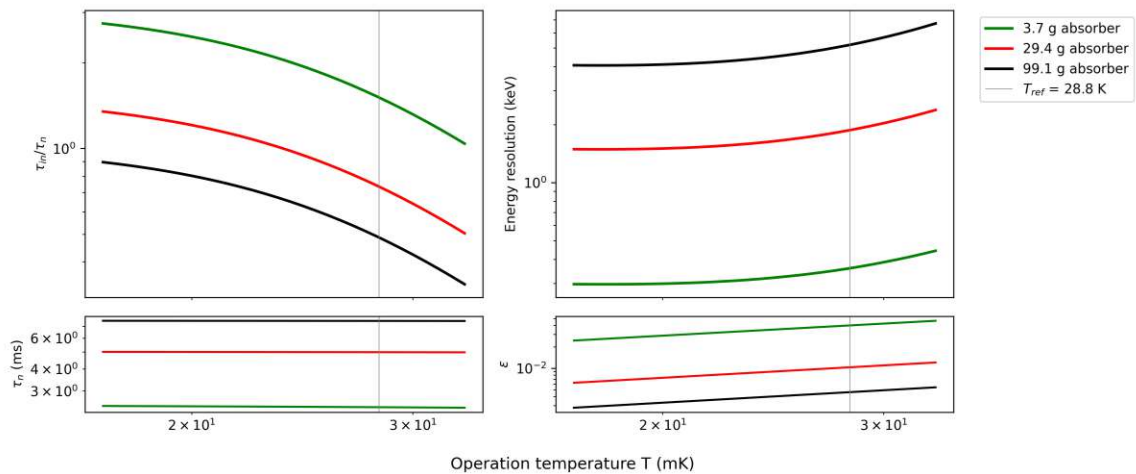

Figure 4.15: Impact of a changing operation temperature on the calorimeterness (upper left, see Sec. 2.2.2 for Def.), the energy resolution (upper right), the phonon thermalization time (lower left), and the collection efficiency (lower right), for the three scenarios of target sizes (green, red, black). Based on measured data from run 376.

We performed the first two steps of the optimization procedure as in the previous section by decreasing the operation temperature and optimizing the dimensions of the gold pad. After the first intervention, we observed a resolution of 4.06 keV for the large crystal. With an optimized gold pad size, the resolution improved to 0.844 keV. Figs. 4.15 and 4.16 show the effects of the optimization process. We can
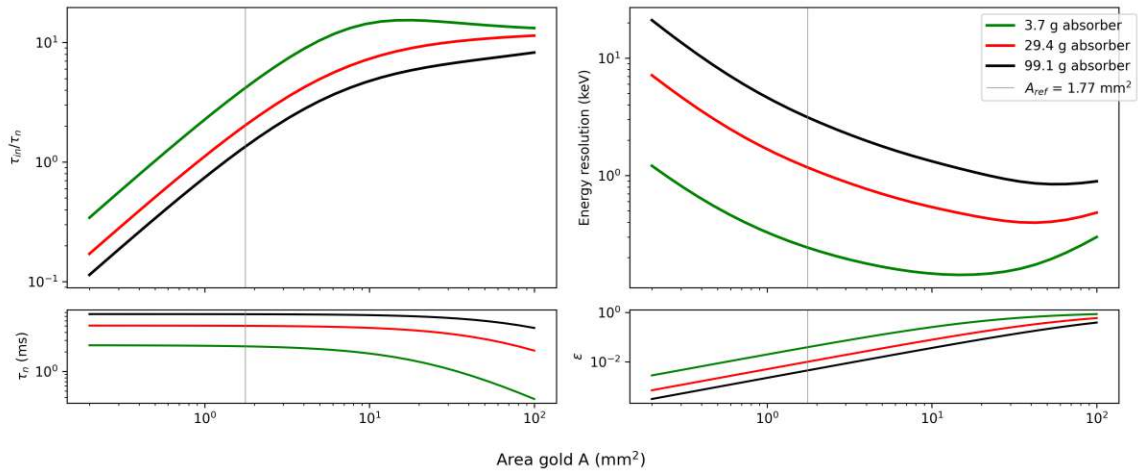
Figure 4.16: Impact of a changing area of the TES on the calorimeterness (upper left), the energy resolution (upper right), the phonon thermalization time (lower left), and the collection efficiency (lower right), for the three scenarios of target sizes (green, red, black). Based on measured data from run 376.

clearly observe that the lower operation temperature has less impact compared to the thermally dominated signal discussed in the previous section.

We assume for this scenario that the COSINUS collaboration will find a way to deposit thin films on sodium iodide crystals (e.g. by sputtering). We therefore optimized the thickness of the gold film to $w = 100$ nm. The optimized thickness and size of the gold film are close to the sizes used in Ref. [185] for a similar design by the RICOCHET experiment. The energy resolution of the 99.1 g crystal with lowered thickness is 0.26 keV. The impact of the changing thickness is shown in Fig. 4.17. The transition from a thermally dominated signal to a signal dominated from the collection of athermal phonons is clearly visible at the point where a lower thickness of the TES also lowers the energy resolution.

A pulse from the final, optimized design for athermal phonon collection is shown in Fig. 4.18. The time constants of the pulse shape are $\tau_n = 6.98$ ms, $\tau_{in} = 16$ ms, $\tau_t = 3.66$ s.

The resolutions after the several interventions are summarized in Tab. 4.4. Interestingly enough, the final resolution is not much better than that which we projected for a thermal detector in Sec. 4.3.1. This is likely due to the reason that also, in the scenario that was planned as thermal, the athermal signal takes over for the largest crystal size. Furthermore, our overly conservative assumption that the collection efficiency decreases with the temperature as well had more impact in the athermally optimized scenario.
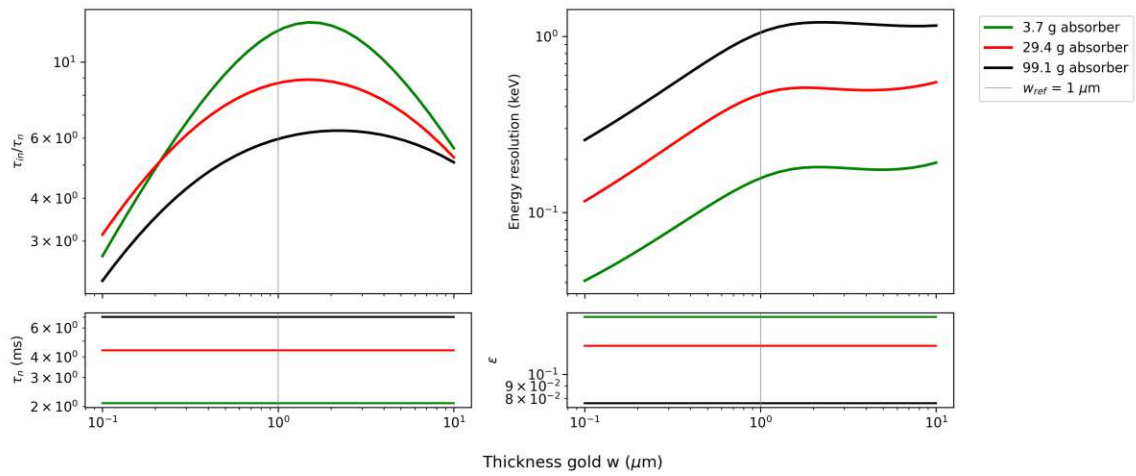
Figure 4.17: Impact of changing the thickness of the TES on the calorimeterness (upper left, see Sec. 2.2.2 for Def.), the energy resolution (upper right), the phonon thermalization time (lower left), and the collection efficiency (lower right), for the three scenarios of target sizes (green, red, black). Based on measured data from run 376.
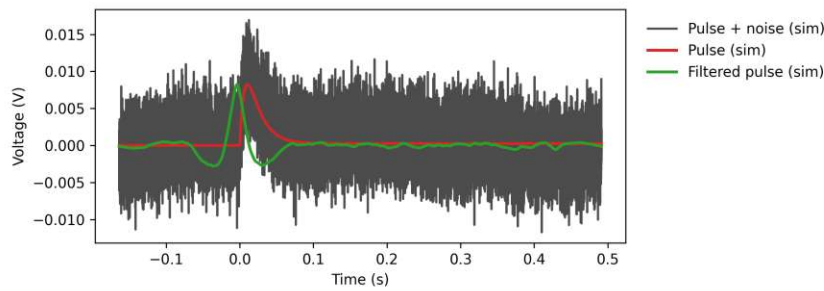

Figure 4.18: Simulated 6 keV X-ray pulse with the optimized design for collecting athermal phonons. The thermal signature is shown in red, with superposed noise in grey. The pulse after optimum filtering is shown in green.

An uncertainty in this scenario is that it is not precisely known how the absorption probability of phonons in the film changes with the thickness of the film. However, many experiments are using films of the proposed thickness.

Furthermore, experience with above and below-ground measurements provides optimism that we have not discussed yet. It was typically observed that noise conditions improve because measurements are performed in shielded below-ground and for low noise conditions optimized environments. Also an improvement of the excess noise conditions with lower temperature was observed.

We kept the transition curves from the runs that provided the data in all projections. In the case of run 376, where the simulated transition curve was less steep than regularly

Energy resolution (keV) with . . .	3.7 g target	29.4 g target	99.1 g target
. . . lower temperature (T_{op} = 17 mK)	0.3	1.49	4.06
. . . optimal gold pad size	0.14 (d = 3.8 mm)	0.4 (d = 6.5 mm)	0.84 (d = 7.3 mm)
. . . thinner gold pad (w = 100 nm)	0.04	0.12	0.26

Table 4.4: Energy resolutions of a remoTES optimized to collect the athermal phonons from the absorber crystal after certain interventions. The values in each row also include all interventions from the rows above. Based on data from run 376. The optimization process is done one after the other, not in parallel.

observed in measurements, we could also hope for a curve that allows for picking better OPs in the final detector design. We did not account for this in our projections and can therefore hope for an even better final result.

Overall, we have shown that the performance goals of the two runs of the COSINUS experiment are very realistic: excluding arbitrary falling recoil spectra within the first year of measurement time and arbitrary recoil spectra after an additional year of measurement time. We based our projections on trigger thresholds of 6.5 times the energy resolution and assumed that we could discriminate noise triggers by their pulse shape. While this is a reasonable assumption, we could also choose a higher trigger threshold. A threshold of 7.5 times the energy resolution would lead to ≈ 1 noise trigger in 1000 kg days. We used for this estimate the approximation of the CDF of the standard normal distribution $\Phi(n)$, depending on the number of standard deviations n, with

$$1 - \Phi(n) \approx \frac{e^{-n^2/2}}{n\sqrt{\pi/2}}. \tag{4.1}$$

In our projection, this would lead to a threshold of 1.94 keV for the second run, slightly exceeding the required 1.85 keV for the exclusion of arbitrary recoil spectra. However, we did not consider that the energy resolution of our detector would smear recoil spectra, effectively making a threshold of $\approx$ 1-2 energy resolution higher than the target threshold still suitable to calculate exclusions.

Finally, it should be noted that the projections from Ref. [48], on which the performance goals we have discussed are based, are for a DAMA threshold of 2 keV$_{ee}$. Recently, results have been resealed with a lower threshold [38], and the additional,

lowest bins show a strong modulation amplitude. It would be possible to compute exclusions based on these new results that require much less exposure, given a lower nuclear recoil threshold. Since the COSINUS detectors could easily be tuned to a lower threshold for a smaller target mass, an even more convenient exclusion scenario is conceivable. We have not used this information in our work.

In this chapter, we used our simulation of cryogenic detector response to project energy thresholds with optimized detector designs for sodium iodide targets to be used in the COSINUS experiment. We have shown that the performance goals of the experiment are feasible, provided a low enough background rate. We will continue in the following Chap. 5 by using similar methods for light DM searches with the CRESST experiment. In the final Chap. 6 we discuss automation techniques based on machine learning for future large-scale detector setups.

Chapter 5

Spin-dependent results from lithium aluminate targets in CRESST-III

The CRESST experiment spearheaded the technology of two-channel readouts for particle identification with phonon and light channels in the past decade. The experiment is in its third phase, CRESST-III, focusing on the detection of light DM. They hold currently the strongest results for sub-GeV/c^2 DM-nucleus interactions under standard assumptions [30]. In recent runs, the low energy sensitivity of the CRESST experiment is limited by the LEE, as introduced in Sec. 2.2.3. This problem is not unique to the CRESST experiment but a joint challenge of the low-threshold cryo-detector community. In a scenario that is not background-limited, there are two strategies to increase the sensitivity of the CRESST experiment further: lower nuclear recoil thresholds would make the parameter space of even lighter DM masses testable; larger target masses would make the parameter space of even lower cross sections accessible. We will treat the second of these strategies in Chap. 6. Concerning the achievability of lower recoil energy thresholds, we discuss here an interesting alternative.

We introduced in Sec. 2.2.2 the scaling properties of cryogenic detectors with TES at length. One of the key messages was that, in a design-optimized limit, a trade-off between target mass and low threshold must be made as a design choice of the operated detector. However, with a fixed energy threshold, sensitivity towards lower DM masses could also be achieved by using lighter target nuclei (see Fig. 1.5). The CRESST experiment traditionally used calcium tungstate as the target material due to its strong scintillation light output, and the heavy tungsten nuclei, since the DM-nucleus cross section increases $\propto A^2$. The oxygen in its molecular composition, with proton number 8, realizes the sensitivity to light DM. Lithium aluminate is an even better material for this purpose, which contains oxygen and lithium, with proton

number 3, and a significant abundance of the lithium-6 isotope. The odd proton and neutron number of lithium-6 makes it additionally sensitive to spin-dependent scattering processes.

The CRESST collaboration has carried out an intensive R&D campaign using lithium aluminate as the target material. Refs. [186–188] report on the first tests of lithium aluminate targets in above and below-ground settings. A framework for calculating spin-dependent DM interaction limits was developed, and the first limits on spin-dependent DM-nucleus interactions with lithium aluminate targets were presented.

This work reviews Ref. [6], where the first limits on spin-dependent DM-nucleus interactions from a blinded data-taking period in the main CRESST setup at LNGS were reported. We summarize the performed RD analysis in Sec. 5.3 and the DM results in Sec. 5.4. Additionally, we extract a set of physics parameters with which we can simulate detectors in CryoEnv that reproduce the performance of the three detectors (two phonon detectors, one light detector) used in Sec. 5.2. These simulations were initially presented in Ref. [3]. Before that, we specify the experimental setup of the CRESST experiment in Sec. 5.1. We conclude with a discussion of the obtained results and potential directions for future CRESST measurements in Sec. 5.5.

5.1 Experimental setup and data taking

The main CRESST experimental setup is located in the LNGS underground laboratory, shielded from environmental radiation and astroparticle impact by a rock overburden of 3600 MWE. A dilution refrigerator cools the setup to ≈ 5 mK. The detector modules are mounted inside a copper carousel, which is thermally linked to the cryostat through a copper rod called the "cold finger". Layers of polyethylene, lead, copper, and a second inner polyethylene layer shield the detectors from radiation in the lab and neutrons, produced by the environment or decay chains in outer shielding layers. Scintillator panels are built around the setup to build an active muon veto. The muon flux inside the LNGS underground lab was measured to ≈ 1 count/m^2/hour in Refs. [189–191], which is 10^6 times lower than at sea level.

The data set used for Ref. [6] included data from two detector modules, "Li1" and "Li2". The first one has a working phonon and light channel, while the second one has only a working phonon channel since the TES on its light detector never showed a superconducting transition. Later we will refer to the channels individually as Li1P, Li1L, and Li2P. The targets of the modules are crystals of size (1 x 2 x 2) cm^3, which
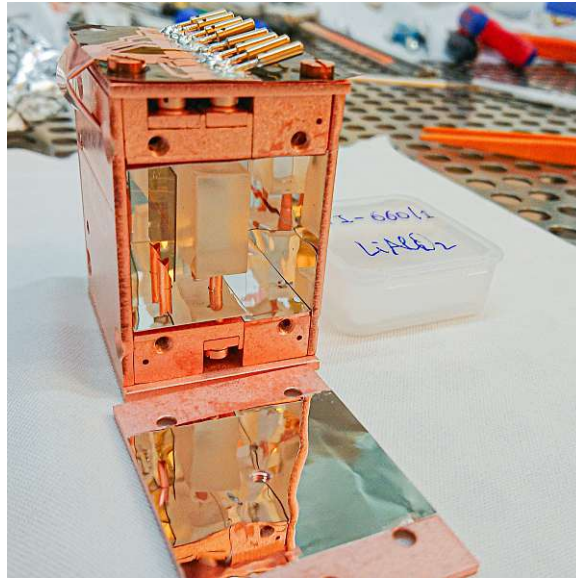
Figure 5.1: The Li1 detector module. Inside the copper housing, a LiAlO2 crystal (right, transparent) as a target for particle scattering is held by three copper sticks as a target for particle scattering. Next to the crystal an SOS light detector (left, grey) is mounted. The inside of the housing is covered with reflective foil, best visible on the detached side of the module (centre, lower part of the picture). Fig. and caption from Ref. [6].

amounts to 10.46 g of lithium aluminate target mass. The light detectors target is of size (0.04 x 2 x 2) cm^3 and made of a silicon-on-sapphire (SOS) substrate. The lithium aluminate crystals are held by copper sticks and the light detectors with plastic sticks. The modules are individually housed in copper boxes, fully lined with reflective and scintillating foil on the inside. The foil maximizes the amount of scintillation light that arrives from the lithium aluminate crystals at the light detector, and enables to veto surface backgrounds.

The data were continuously recorded between February and August 2021 with a sampling frequency of 25 kHz and triggered in software. Triggered records were stored with a record length of 16384 samples. The data taking consisted of a background data set and a neutron calibration during which an AmBe source with activity of ≈35.5 MBq was deposited outside the shielding. Throughout the measurement, an iron-55 source with activity of ≈1 mBq was mounted inside the housing of the detector modules.

5.2 Characterization of the detector modules

The same detectors that were used for Ref. [6] were characterized with their physics parameters in Ref. [3]. This was done by adjusting a CryoEnv simulation to reproduce the detector response of the three detectors Li1P, Li1L, and Li2P. The procedure was described in Sec. 4.3 and the simulation in Sec. 2.1. The resulting parameters are summarized in Tab. 5.1. It is important to note that some of the parameters are strongly correlated in their effects on the detector response, such as R_H and the DAC values and the pulser strength β, the parameter δ, and δ_H.
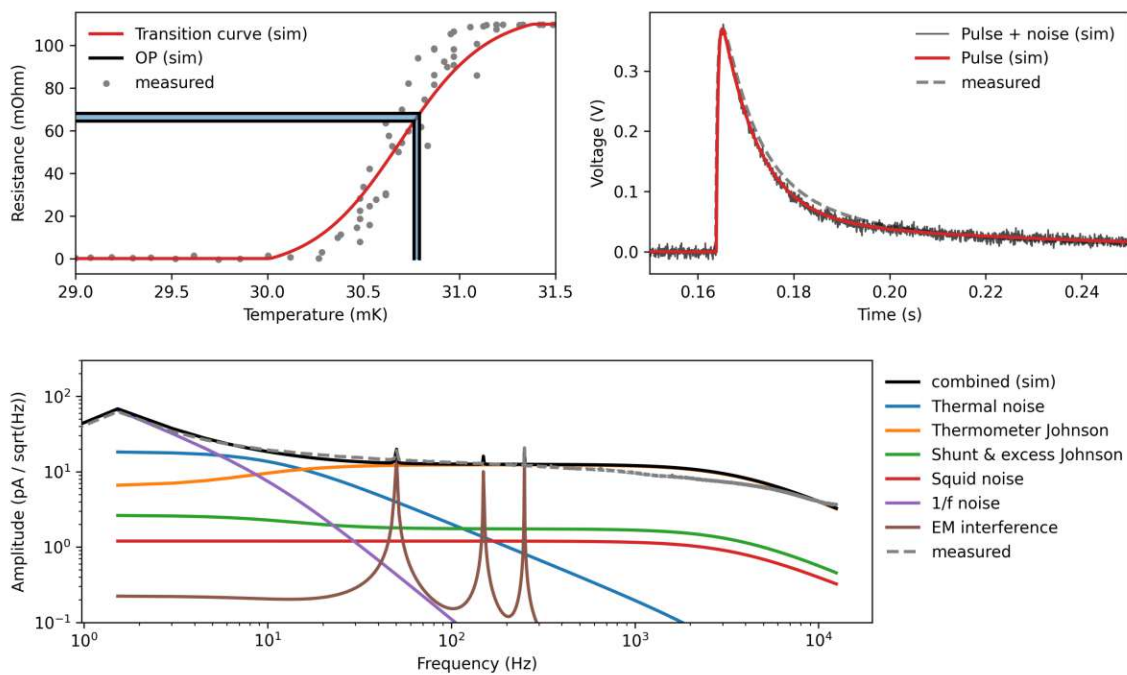

Figure 5.2: Simulation and measurement of a 5.95 keV X-ray event induced by a calibration source in the Li1P detector. (upper left) The OP (black/blue lines) within the simulated transition curve of the TES (light red line). A measurement of the transition curve is shown for comparison (grey dots). (upper right) The voltage pulse induced in the simulated SQUID amplifier without noise (red) and overlayed with noise generated from the simulated NPS (black). A measured voltage pulse is shown for comparison (grey dashed). (lower part) The simulated NPS (black) has individual noise contributions (colored). The 1/f, excess Johnson, and EM interference noise components were adjusted to fit the measured NPS (grey dashed). Fig. and caption also used in Ref. [3].

It is especially positive to note that the parameters agree with the expected value ranges of their designs. The plastic sticks realize a much weaker coupling between

absorber crystals and heat bath than the copper sticks. The electron-phonon couplings of Li1P and Li2P and much larger due to the higher temperature and the larger tungsten area than in the light detector. The collection efficiency of the sensors on the large crystals is larger since the TES-to-crystal surface is also much larger. However, the scaling is not linear. The phonon collectors on the large crystal have a surface of 16.53 mm^2, and those on the light detector have a surface of 1.01 mm^2. The collection efficiency on the large crystal is a factor of 2 smaller. The less fortunate shape of the light detector, with a larger surface-to-volume ratio, also has an impact on the collection efficiency. All tungsten and aluminum films have a thickness of roughly 200 nm. The thermal links made of gold have a size of (40 x 610 x 0.08) μm^3 on the large crystals and (40 x 2000 x 0.08) mm^3 on the light detector. Differently to the data we were working with in Sec. 4.3, the simulation can reproduce the noise power spectra of Li1P, Li1L, and Li2P without additional excess Johnson noise.

We show exemplarily the detector response of Li1P in Fig. 5.2. A suitable agreement between measured data and simulation is visible. The small deviations do not significantly impact the difference in performance. The fitted time constants of the measured pulse shapes were $\tau_n = 0.38$ ms, $\tau_{in} = 9.4$ ms and $\tau_t = 72$ ms for Li1P; $\tau_n = 0.094$ ms, $\tau_{in} = 3.7$ ms and $\tau_t = 17.8$ ms for Li1L; and $\tau_n = 0.4$ ms, $\tau_{in} = 12$ ms and $\tau_t = 79.2$ ms for Li2P.

The PHs of the 5.9 keV iron-55 events were 0.37 V (Li1P), 3.3 V (Li1L), and 0.523 V (Li2P). The highest observed pulses were 3.35 V (Li1P), 2.24 V (Li1L), and 9.28 V (Li2P). The CPs of the detectors did not fully saturate the TES. The V_{set} values were much lower than the values reported above.

Quantity	Li1P	Li1L	Li2P
V_f (mm^3)	$4.08 \cdot 10^{-4}$	$1.44 \cdot 10^{-5}$	$4.08 \cdot 10^{-4}$
V_a (mm^3)	$4 \cdot 10^3$	$2 \cdot 10^2$	$4 \cdot 10^3$
C_e (pJ/mK)	$2.11 \cdot 10^{-3}$	$3.5 \cdot 10^{-5}$	$2.5 \cdot 10^{-3}$
C_a (pJ/mK)	0.113	$1.61 \cdot 10^{-4}$	$9.7 \cdot 10^{-2}$
G_{eb} (pW/mK)	0.123	$2.66 \cdot 10^{-2}$	0.138
G_{ab} (pW/mK)	1.565	$9 \cdot 10^{-3}$	1.16
G_{ea} (pW/mK)	0.214	$2.27 \cdot 10^{-3}$	0.1
τ_n (s)	$3.82 \cdot 10^{-4}$	$9.4 \cdot 10^{-5}$	$4 \cdot 10^{-4}$
τ_{TP} (s)	$4.97 \cdot 10^{-3}$	$2.98 \cdot 10^{-3}$	$4.97 \cdot 10^{-3}$
ϵ	0.115	0.056	0.104
δ	0.144	0.26	0.056
δ_H	0.9	$0.$	0.8
R_s (Ω)	$4 \cdot 10^{-2}$	$4 \cdot 10^{-2}$	$4 \cdot 10^{-2}$
R_H (Ω)	6.75	7.25	5.25
R_{f0} (Ω)	0.11	0.115	0.1
L (H)	$3.5 \cdot 10^{-7}$	$3.5 \cdot 10^{-7}$	$3.5 \cdot 10^{-7}$
T_c (mK)	30.7	23.0	29.4
k (1/mK)	4.4	13.5	5.52
β	$2.25 \cdot 10^{-2}$	$6.25 \cdot 10^{-3}$	$2 \cdot 10^{-2}$
I_H (μA)	4.8	0.904	8.27
i_{sq} (pA/$\sqrt{\text{Hz}}$)	1.2	1.2	1.2
E_J	1	1	1
$\frac{\Delta R_{f,flicker}}{R_{f0}}$ (pJ)	$8 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$8 \cdot 10^{-5}$
α	2	1	2
p_0	$1.5 \cdot 10^{-5}$	$3 \cdot 10^{-5}$	$3 \cdot 10^{-5}$
p_1	$1 \cdot 10^{-5}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-5}$
p_2	$1.5 \cdot 10^{-5}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-5}$
IB_{min} (μA)	0.5	0.5	0.5
IB_{max} (μA)	17.9	17.9	17.9
η (V/μA)	5.77	5.77	5.77

Table 5.1: Values used in the simulation of the detectors Li1P, Li1L and Li2P. See text, Sec. 2.1 for definitions. Tab. and caption also used in Ref. [3].

5.3 Raw data analysis

The RD analysis performed in Ref. [6] follows the procedure introduced in Sec. 3.1. The measured background data set was split into a training data set of 0.153 kg days for Li1 and Li2 each and a blind data set of 2665 h / 1.161 kg days (Li1) and 2716 h / 1.184 kg days (Li2). The data set from the neutron calibration period amounted to 0.178 kg days for Li1 and Li2 each.

SEVs, NPSs, OFs, pulse shape features, PHs, and recoil energies were calculated as explained in Ref. 3.1. The trigger thresholds were tuned to 1 noise trigger/kg/day, assuming Gaussian noise, using a fit on the noise maxima as discussed in Sec. 3.1.1.5.

For the analysis of the Li1 module and a cross-check analysis of the Li2 module, performed in Ref. [6] and this work, the Cait package was used [5]. The official analysis chain for the Li2 module, and a cross-check analysis for the Li1 module, were done with the collaboration-internal CAT software package.
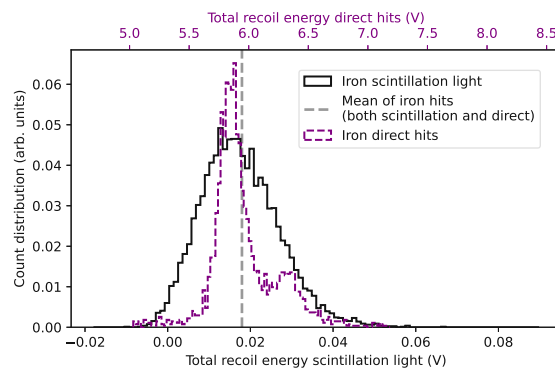

Figure 5.3: Overlay of the normalized Li1 light detector energy count distribution of the iron line from scintillation light (black, bottom x-axis) and the iron line from direct hits (purple dotted, top x-axis). The two x-axes are shifted and scaled such that the average value of the two iron lines overlap. Their ratio determines the collected light of the target (see text). Fig. and caption also used in Ref. [6].

Only a relatively small share of the recoil energy in the absorber is converted into scintillation photons, and not all of them are collected in the light detector. The collected scintillation light is useful for estimating the discrimination power of the two-channel readout. Lithium aluminate is a worse scintillator than the traditionally used calcium tungstate. The collected scintillation light can be measured by comparing the position of the iron peak from direct hits and that from scintillation light events in the energy spectrum of the light detector. They are shown in Fig. 5.3. Ref. [6] reports that $(0.302 \pm 0.001)\%$ of the recoil energy in the absorber is collected in the light detector as scintillation light, for an e^-/γ-event.

The data were cleaned with a stability and rate cut, where outliers of 3 sigma and more were rejected. The average trigger rate, excluding heater pulses and random triggers, was 0.4 events/min. Furthermore, pulse shape cuts were performed on the events to reject artifacts. Events with higher energy as 32 keV are cut. The total number of surviving events is 12487. The distribution of decay times versus PHs of all events, and only the events that survived the pulse shape cuts, are shown in Fig. 5.4, for the phonon and light channel individually. The event band with the characteristic target recoil pulse shape is clearly separated from event bands that are artifacts, direct hits of the light detector, or SQUID resets. One peculiar event distribution is visible whose origin was attributed to the scintillating foil in Ref. [6]: while the light channel shows a strong energy deposition and the pulse shape of a direct hit, the phonon detector shows only a low energy recoil, intermixed with the LEE. These foil events could be identified and removed in Li1 due to the coincidence between phonon and light channel, but not in Li2, where they remain and make up a large share of the low energy events. They cannot be interpreted as the sole origin of the low-energy event distribution because their spectral shape is significantly different from the known LEE events.
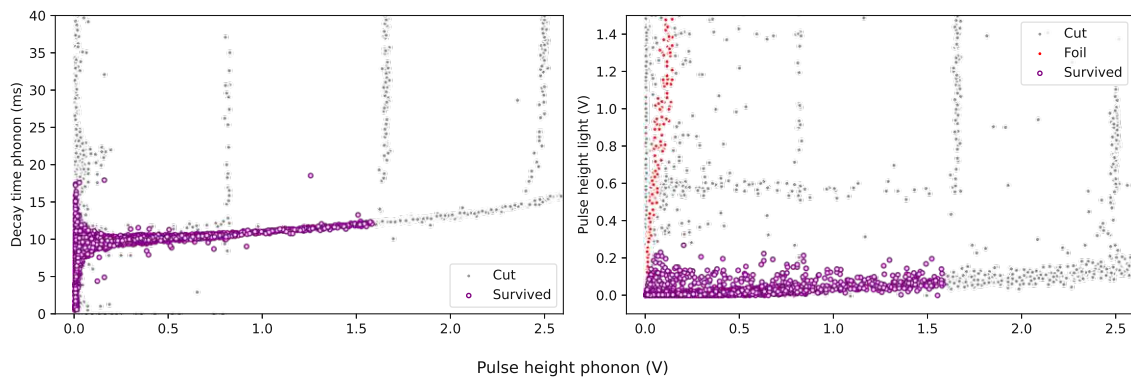

Figure 5.4: Visualisation of the surviving (purple) and cut (grey) events in the Li1 DM data set. (left) The distribution of decay times in the phonon channel over PHs in the phonon channel. The band of recoil events is clearly visible and mostly distinct from the artifact events. For PHs below 0.2 V, the band widens, which degrades the discriminating power of quality cuts. (right) The distribution of PHs in the light channel versus the corresponding PH in the phonon channel. Again, the band of particle recoils is clearly visible. For low phonon PHs the event class of foil events appears: due to their high PH in the light channel, higher than for regular target recoils, these events can be rejected as background. In both pictures, the vertical and secondary horizontal event bands are SQUID resets caused by high-energy recoils. Fig. and caption also used in Ref. [6].

Additional coincident cuts were applied for the data sets that were later used for a DM analysis. Events that showed a coincidence with muon veto triggers in a time window of -5 to +10 ms w.r.t. the trigger time stamp of the muon veto were rejected. The muon veto triggered with 4.52 Hz, which reduced the exposure by 6.79 %, and 6.82 % of the events that survived all previous cuts were rejected. The expected rate of events removed with non-muon induced origin is (6.79 ± 0.23) %. The error was calculated with the binomial formula for the variance of the proportion of wins in N Bernoulli trials: $\sqrt{(p(1-p)/N)}$, where N is the number of events that survived the previous cuts, and p is the percentage of removed exposure. In a window of -10 to +10 ms, events that were coincidental with other channels were removed. This removed 0.93 h runtime, corresponding to 0.0387 % of the exposure. In Li1, 2 events were removed by this cut, and none in Li2.
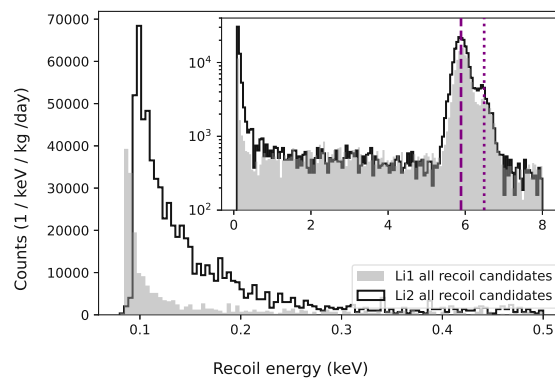

Figure 5.5: Recoil energy spectrum for the Li1 (grey) and Li2 (black) modules. (inset) The energy region up to 8 keV. The most prominent event clusters are the LEE and the two iron lines (purple dashed - K-α, purple dotted - K-β). (main window) The energy region up to 0.5 keV, dominated by the LEE. The Li1's LEE is less prominent due to the cut based on light channel information, which removes the foil events. Fig. and caption also used in Ref. [6].

The recoil energies are calibrated with the prominent spectral line from iron-55. The surviving recoil candidate events are shown in Fig. 5.5. There are three dominant contributions visible: the iron-55 line, an approximately flat distribution and the steeply rising event rate close to thresholds. The low energy event distribution contains the known LEE events (discussed in Sec. 2.2.3) in both modules and in Li2 additionally foil events. The flat background is majorly from the beta spectrum of tritium. Lithium-6 is subject to the reaction lithium-6(n, α), where tritium nuclei are a product.
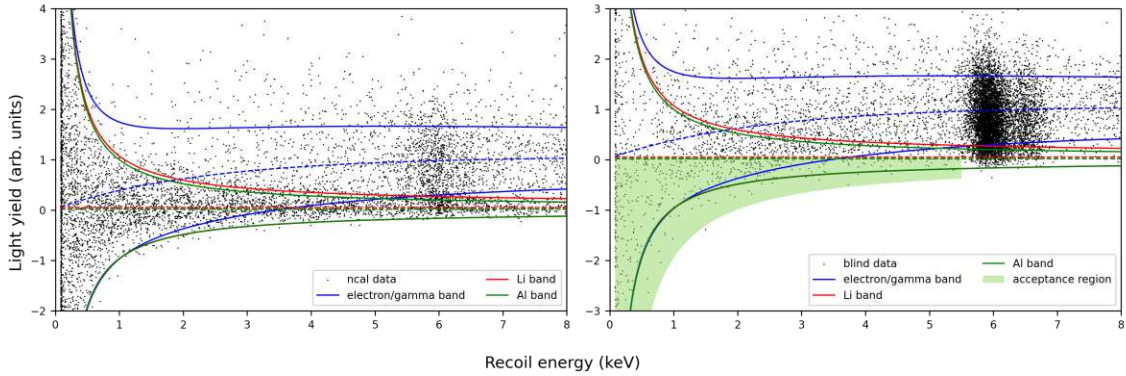
Figure 5.6: Fitted LY bands as a function of the recoil energy in the Li1 neutron (left) and blind (right) data sets, after application of the selection criteria. Electron/γ (blue) and nuclear recoils off the nuclei with odd proton number (lithium red, aluminium green) cluster in band-like structures and are fitted with Gauss distributions, with energy dependent means and standard deviations. The acceptance region for DM candidates (light green) is chosen as the lower half of the lithium and aluminium bands, mitigating the EM background. Fig. and caption taken from Ref. [6].

A band fit, as introduced in Sec. 3.2, was performed on the recoil energies and the LYs. The LY is due to the definition of target-electron-scattering equivalent energies in the light channel automatically one for events from the iron source. The resulting bands are shown in Fig. 5.6. The drawn lines correspond to the 80 % inclusion interval of the Gaussian distributions along the LY axis. For the bandfit, the collaboration-internal software "Limitless" was used. Individual bands for all recoils, and for electron recoils were fitted with the data from the neutron calibration. The nuclear recoil bands are clearly visible in the neutron calibration data, while the electron recoil band is distributed over a large range of LYs. This is due to the low light output of the lithium crystals. The oxygen band overlaps with the aluminum band and is not resolved individually.

For Li2, no LY data were available, and a bandfit could therefore not be performed. Particle identification could not be used for the discrimination of electron recoils.

For the DM search, the ROI was defined as the lower half of the lithium nuclear recoil band and up to recoil energies of 5.5 keV.

The LEE is the subject of many ongoing studies, and the comparability of energy spectra is an important question to draw conclusions from experiments. In Ref. [6] a heuristic, parametric description of the LEE was fitted to the low energy event spectrum of Li1. The equation that was fitted to the histogram is

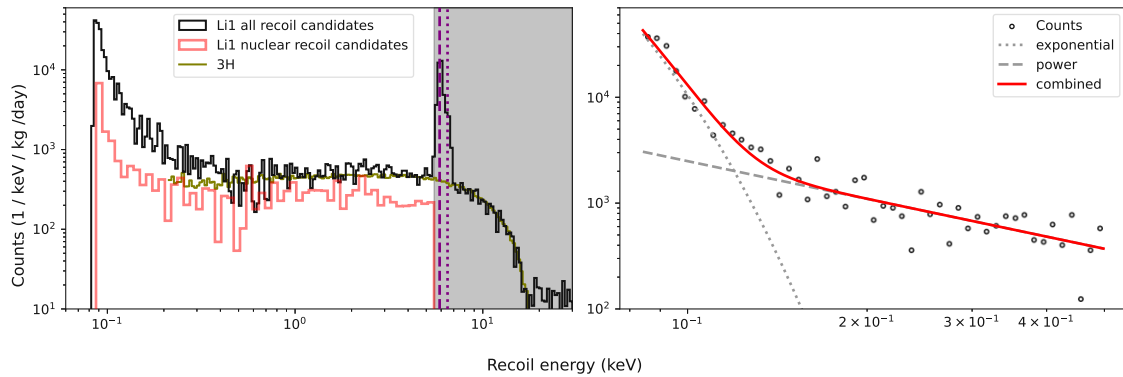$$f(x, a, b, c, d) = a \exp(-bx) + c x^{-d}, \tag{5.1}$$

Figure 5.7: Recoil energy spectrum of particle events inside the target of Li1. (left) The recoil spectrum up to 30 keV (black), the ROI for the DM search ends at 5.5 keV, indicated by the grey shade. The choice of the ROI is motivated in the text. The three prominent contributions are the clearly visible iron line (purple dashed - K-α, purple dotted - K-β), the tritium background (olive line to guide the eye) and the LEE. The events within the acceptance region are considered nuclear recoil candidates (red). (right) The region below 0.5 keV, which is dominated by the LEE. The recoil energy spectrum of all recoil candidate events (black dots) can be fitted with the sum of an exponential (grey, dotted) and a power law component (grey, dashed). Fig. and caption also used in Ref. [6].

where x is the recoil energy variable, and a, b, c, and d are free fit parameters. The results from the fit are shown in Tab. 5.2. It was especially important to use two components instead of one exponential or one power law to achieve a meaningful spectrum description. The fitted low energy spectrum is shown in Fig. 5.7 (right). The exponential component describes a steeper rise close to the threshold, and the power law component describes an event population at higher energies. In Fig. 5.7 (left) the spectrum of Li1 with all recoil candidates and only the events inside the ROI are shown. The flat background is significantly reduced by the particle identification procedure. A spectral component of the tritium beta spectrum, adjusted to the height of the flat background, is shown in comparison, as a guide for the eye, to pronounce the characteristic beta spectrum.

To validate the analysis chain were 2 M events were simulated, and the identical analysis chain was used on them as on the measured data. The share of triggered and surviving events is visualized in Fig. 5.8 for Li1 and Li2 individually. An error function fit to the triggered events was used to define the energy threshold, with the method discussed in Sec. 3.1.4. The fit resulted in an energy threshold of (83.60 $\pm$ 0.02) eV for Li1 and (94.09 $\pm$ 0.13) eV for Li2. The baseline energy resolution was estimated for the phonon channels from the width of the error function, which results

	value	uncertainty	units
a	$4.7 \cdot 10^8$	$\pm\ 7.3 \cdot 10^8$	$(\text{keV} \cdot \text{kg} \cdot \text{day})^{-1}$
b	84	$\pm\ 16$	$(\text{keV})^{-1}$
c	162	$\pm\ 41$	$(\text{keV}^{(1 - \text{d})} \cdot \text{kg} \cdot \text{day})^{-1}$
d	1.2	$\pm\ 0.2$	

Table 5.2: The parameters obtained from a χ^2 fit of Eq. (5.1) to the binned spectrum for the Li1 LEE. The errors were calculated from the covariance matrix of the fit parameters. Tab. and caption also used in Ref. [6].

in (13.10 ± 0.02) eV for Li1 and (15.89 ± 0.18) eV for Li2. The errors were calculated from the covariance matrix of the fit parameters. The resolutions were also calculated by superposing events on empty baselines, and the results agreed within the error bounds. For the light channel of Li1 only the second method was used and resulted in a baseline resolution value of (748 ± 7) eV$_{ee}$, for target recoils, and (2.26 ± 0.02) eV for direct hits. The errors were calculated as described in Sec. 3.1.1.4.
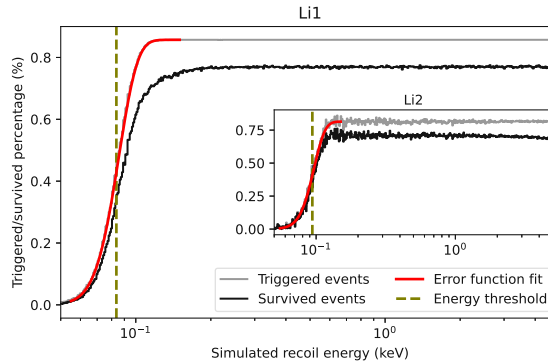

Figure 5.8: The normalized trigger rate (grey) and survival rate (black) of simulated Li1 events (Li2 events in inset), as a function of the simulated recoil energy. The latter provides a realistic estimate of the survival probability. The energy threshold (olive, dashed) is the recoil energy at which the fitted error function (red) drops below 0.5 times the constant triggered fraction above threshold. The constant trigger efficiency for Li1 is $(85.71 \pm 0.01)\%$ and the trigger energy threshold (83.60 ± 0.02) eV. For Li2 the trigger efficiency is $(81.26 \pm 0.08)\%$ and the trigger energy threshold (94.09 ± 0.13) eV. Fig. and caption also used in Ref. [6].

5.4　Dark matter results

The data sets of Li1 and Li2 were used in Ref. [6] to calculate exclusion limits on the spin-dependent interaction of DM with nuclei. For this, all events inside the ROI defined in Sec. 5.3 were interpreted as signal candidates. This especially includes the largest share of the LEE event population. This is a conservative analysis. However, an analysis of this kind is limited to setting limits on the interaction rate. A LH based on these data and assumptions could not reasonably be combined with results from other experiments.

For the DM analysis, the natural abundances of the lithium isotopes were used: 92.41 % of the lithium nuclei were assumed to be lithium-7 and 7.49 % lithium-6. All aluminum nuclei were assumed to be the aluminum-27 isotope. The oxygen nuclei were not included in the calculation since they would only negligibly raise the expected signal for spin-dependent scattering. The DM analysis is done as introduced in Sec. 3.2, using the Yellin optimum interval method and the astrophysical assumptions introduced in Sec. 1.1.3. The neutron- and proton-only limits were calculated independently. For the rate formulas the following variables were used: $\langle S_n \rangle = \langle S_p \rangle = 0.472$ for lithium-6 according to Ref. [192]; $\langle S_p \rangle = 0.497$ for lithium-7 according to Ref. [193]; and $\langle S_n \rangle = 0.0296$, $\langle S_p \rangle = 0.343$ for aluminum-27 according to Ref. [194].

The spectra of expected scattering rates for given DM masses are adjusted to the realistic expected observed spectra. For this, the event spectrum used for the cut efficiency simulation was reweighted to resemble the expected DM scattering spectrum. The spectrum of surviving events automatically is then the expected DM spectrum observed in the detector. Strong up- and downward noise fluctuations are not allowed since they likely correspond to artifacts or unphysical effects of the analysis. They are removed by excluding simulated events for which more than 3 times the energy resolutions are between the simulated and the reconstructed recoil energy.

In Fig. 5.9 the resulting 90 % confidence upper limits for spin-dependent scattering on protons and neutrons are shown for DM masses between 0.16 and 0.6 GeV/c^2, and in comparison to leading limits from other experiments. The limits from Li1 are lower since the rate of signal candidates is reduced by the PD through the LY.

5.5　Discussion

We reviewed in this chapter the analysis and results from Ref. [6], obtained with two detector modules with lithium aluminate targets, and the detector characterization
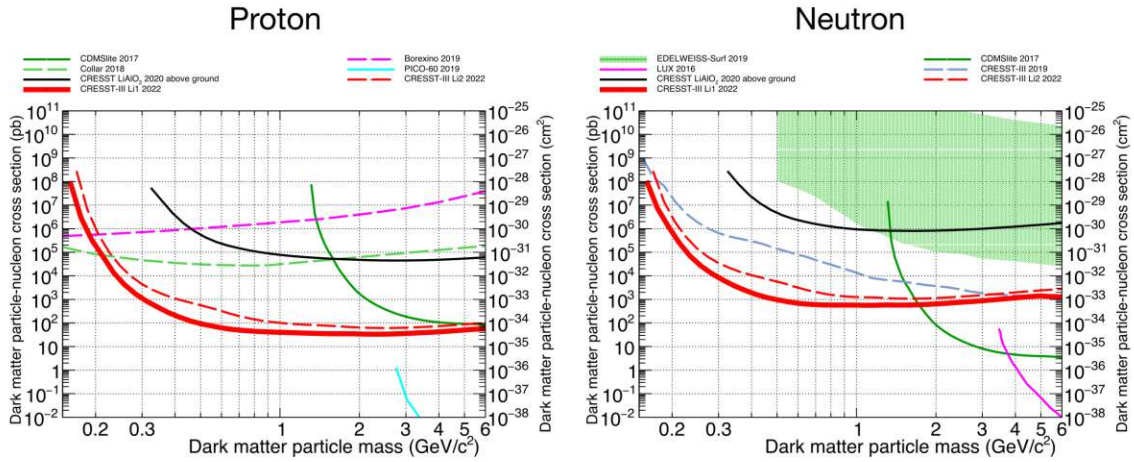
Figure 5.9: The exclusion limits for proton-only (left) and neutron-only(right) spin-dependent DM-nucleus cross sections versus DM particle mass set by various experiments compared with the two lithium modules described in this work with lithium-6, lithium-7 and aluminum-27. This work gives the most stringent limits between 0.25 GeV/c^2 - 2.5 GeV/c^2 for proton-only and between 0.16 GeV/c^2 - 1.5 GeV/c^2 for neutron-only interactions. The solid red line shows the Li1 limits which includes the scintillation light information and the dashed red line shows the Li2 limits where no light information was available (hence worse). The previous above ground results from CRESST using the same detector material and procedure with higher threshold and lower exposure is also shown with solid black line [188]. Also, CRESST-III 2019 results for neutron-only interactions using oxygen-17 are shown also in dashed light-blue line(right) [30]. Additionally, we show the limits from other experiments: EDEL-WEISS [195] and CDMSlite with germanium-73 [196], PICO with fluorine-19 [197], LUX [198] which use xenon-129 and xenon-131, J. I. Collar with hydrogen-1 [199] and the constrain derived in [200] from Borexino. Fig. and caption from Ref. [6].

of the same detector modules reported in Ref. [3]. The resulting cross-section upper limits for proton- and neutron-only spin-dependent scattering are world-leading in the mass region 0.25 to 2.5 GeV/c^2 for proton-only and 0.16 to 1.5 GeV/c^2 for neutron-only interactions. Overall, lithium aluminate was confirmed as a promising target material for further DM search projects. The obtained energy thresholds were relatively high, compared to other CRESST detector modules. However, it's unlikely that this is due to the target material since lithium aluminate has a reasonably high Debye temperature. The thresholds could very likely be much improved by using TES with a lower transition temperature. With a similar threshold, a lithium aluminate target leads to a much stronger sensitivity for low-mass DM compared to tungsten. Comparing isotope masses of lithium-6 and oxygen-16 shows, that the lithium aluminate target could detect a factor $16/6 \approx 2.7$ lighter DM, compared

with a calcium tungstate crystal of the same energy threshold. Lithium aluminate is therefore also highly interesting for spin-independent light DM searches.

The tritium background needs to be accounted for in future analysis with lithium aluminate targets. This can easily be done by including its spectral shape in a LH model as background. Also, the prominent background enhancement at low energies from the iron source should be avoided in the future. Other calibration methods must be found. For this, the fixed endpoint of the beta spectrum from tritium could serve as a useful calibration point. However, collecting the necessary exposure for a suitable fit to obtain an energy calibration might be impractical.

The same reaction, which causes the tritium background, could also be an appealing property to monitor the neutron flux in the CRESST setup [187]:

$$^{6}\text{Li} + \text{n} \rightarrow \alpha + {}^{3}\text{H} + 4.78\text{MeV}.$$
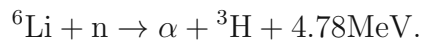
The incident muons are expected to produce secondary neutrons in the outer shielding layers. This process has never been measured in the CRESST setup, and the neutron flux produced by this reaction and reaching the detector is not well known, although it is expected to be very small. The 4.78 MeV γ produced in the neutron capture process would far exceed the range of recoil energies that can reasonably be reconstructed in the phonon channel. However, the scintillation light produced could be measured and reconstructed in the light channel. A coincidence analysis of recoils in the light channel in the favourable energy range with the muon veto trigger could therefore lead to an estimate of the neutron flux produced.

In the remainder of this section, we discuss other possible actions for the CRESST experiment to improve its sensitivity. The LEE is one of two major challenges for the CRESST experiment. A major measurement campaign to establish more insights about the phenomenon was carried out in the past year, and the major learnings are summarized in Ref. [31]. An open question is whether the excess is a process located in the superconducting film or in the absorber crystal. We would expect that a process in the film induces pulses with a faster rise time than absorber recoils and that they would have a suppressed thermal part of the pulse shape. However, the SEVs built from low-energy and intermediate-energy events do not look significantly different. It is possible that the limited slew rate of the SQUID, the limited sampling frequency, and the imperfect overlap of events when building an SEV could smear out the rise time. Furthermore, the thermal part of the pulse shape is also subdominant for absorber recoils. A potential strategy to clarify the questions of varying rise times with existing detectors would be to fit the two-component pulse shape model to all

pulses individually and perform a statistical comparison of the rise times of events at the lowest and higher energies.

Another test that we have performed is to consider the change in the energy scales when events would originate in the TES directly, compared to an origin in the absorber. A TES event is expected to appear with a much larger phonon collection efficiency, probably $\epsilon \approx 1$. To test if the excess spectra overlap better under this hypothesis, we multiplied the low energy spectra of Li1P, direct hits in Li1L and Li2P with their individual collection efficiencies and show them in Fig. 5.10. The disagreement between the spectra is smaller than without this correction. However, it remains large. The low energy spectrum of Li2P is also in this scenario spoiled by the foil events and therefore only shown for the sake of completion, not to be interpreted as LEE.
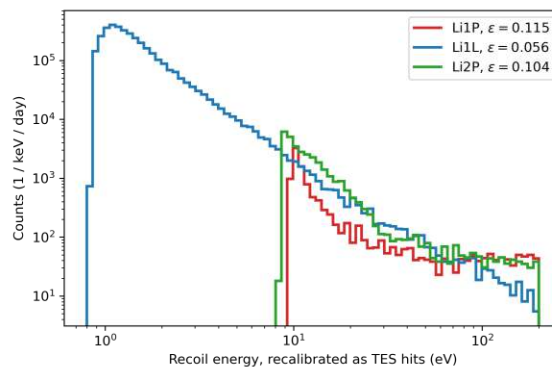

Figure 5.10: Low energy spectra of direct hits in Li1P, Li1L and Li2P, interpreted as hits in the TES. The energy scale was multiplied by the individual phonon collection efficiencies. The low energy spectrum of Li2P is visibly spoiled by the foil event background. The spectra of Li1P and Li1L do not agree, where the rate in the smaller TES of the light detector is significantly larger. The hypothesis of TES scattering is therefore not more plausible than the standard hypothesis of absorber scattering.

The CRESST phonon collectors are another subject for future studies. We observe a decrease in the phonon collection efficiency of factor two by comparing the light and phonon detectors studied in this chapter. This difference is already expected from the different geometries of the crystals. The fact that the light detector's phonon collectors are a factor 16 smaller does seemingly not impact the collection efficiency too much. Interesting questions to study would be the movement of produced quasiparticles in the aluminum film and if this movement can be directed towards the TES, e.g. through applied current or fields. A systematic study, maybe supported by simulations, of the collectors' optimal shape and size could also be of interest. Phonons that produce

quasiparticles in the films that do not reach the TES are lost, and a too-large phonon collector can therefore be seen as a parasitic phonon trap.

The second major challenge for the CRESST experiment is the collection of larger amounts of exposure. This requires the simultaneous operation of a large number of detectors, likely hundreds. We address this challenge in the following chapter, where we propose methods to automate the control and crucial parts of the analysis with machine learning, such that fewer manual interventions are necessary. Operation of a large number of detectors is planned for CRESST in the future in a major hardware upgrade. As proposed in the following chapter, methods will become essential to handle the larger workload for analysts and operators.

Chapter 6

Automating detector optimization and analysis for large-scale detector setups

Cryogenic calorimeters with TES and single crystals as targets are currently the most sensitive devices to test interactions of sub-GeV/c^2 DM with target nuclei. This is mostly due to their low nuclear recoil energy thresholds and their strong discrimination power for EM interactions. Accessing the parameter space of even lighter DM masses and lower cross sections requires optimization of the detector design and the operation of larger target masses. While the former requirement was discussed in Chap. 5, we focus on the operation of larger target masses in this chapter.

In Chap. 2 we established a relationship between target mass and threshold for detectors using TES. Using a larger target would necessarily lead to a higher threshold. The strategy of low threshold experiments using solid targets is to operate many detectors with small targets simultaneously instead of a single, large target. In an upcoming hardware upgrade the CRESST experiment plans to realize this by operating more than 100 detector modules [97]. Also other experiments such as the SuperCDMS experiment plan to scale up their operated target mass.

The operation of a large-scale multidetector setup brings several challenges. Assuming that the necessary detector modules and electronics can be mass-produced in a suitable way, the detector still needs to be analyzed, and its measurement conditions need to be optimized. Both processes require manual interventions, and the available human workforce resources are already exhausted by the current operation of ≈ 10 detector modules in the latest CRESST runs. Therefore, the processes need to be streamlined and automated so that a handful of human operators and analysts can reasonably handle a detector setup of hundreds of detectors.

We have introduced a method for automating the event selection in Sec. 3.1.3.3, also reported on in Ref. [5]. In the same section, we emphasized that there are potential issues with the method when training and validation are done exclusively with simulated data. In Sec. 6.1 we review the follow-up work Ref. [4], where these uncertainties were mitigated by using measured and augmented data for the training of the event classifiers. The authors efficiently generalized previously performed pulse shape cuts on a large set of historic CRESST data to new and unseen detectors.

The optimization of the control parameters of a TES-based detector is a task that is currently performed by human experts, for each device individually. An optimum of the SNR needs to be found, depending on the bias current run through the TES and the temperature governed by the heating resistor. Per detector, the optimization task typically takes several hours to be completed. In Ref. [3] it was proposed to automate the task with RL. The authors achieve human-level performance both in a virtual environment and live on the CRESST experiment. We review their method and results in Sec. 6.2.

Finally in Sec. 6.3 we discuss the impact of these automation methods with machine learning on large-scale detector setups. We elaborate on which necessary manual interventions remain and what a realistic workflow could look like.

6.1 Automated data cleaning with deep learning

One of the tasks that requires the most manual interventions in the RD analysis is the data cleaning process. The methodology was introduced in Sec. 3.1: particle pulses are jointly triggered with artifacts, and a data pre-selection has to be made to achieve useful energy calibrations and avoid unnecessary instrumental backgrounds. Examples of such artifacts and pulses are shown in Fig. 6.1.
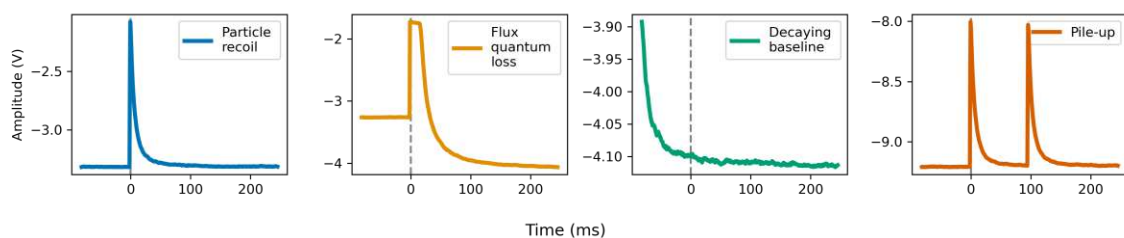

Figure 6.1: Particle recoils produce a pulse-shaped record (blue). Flux quantum losses of the SQUID amplifier in the read-out circuit are caused by fast magnetic field changes, e.g. from high energy recoils (orange). Decaying baselines are residuals from earlier high-energy pulses (green). Pile-up originates from multiple particle recoils within the same record window (red). Fig. and caption also used in Ref. [4].

The quality cuts, which are defined as rejection regions in the feature space of the events, are different for each operated detector since pulse shapes and artifacts look different. They are hardly predictable a-priori but designed by using a training set of the data. Most cuts are performed in one dimension, leading to overlaps between artifact and signal candidate populations.

Ref. [4] interprets the discrimination between pulse shapes and artifacts as a binary classification problem. NNs show good performance in image recognition tasks, and the authors argue that the data structure of the triggered records of events can be interpreted as one-dimensional "images" of the events or as time series. They build the case that it should be possible to pre-train a binary classification model to decide if given events are pulses or artifacts without the necessity for a training set from the detector being worked at. Their work builds directly on the NN-based data classification method introduced in Ref. [5] and discussed in Sec. 3.1.3.3.

In Ref. [4] the authors created a large data set to train classifiers for the discrimination between pulse shapes and artifacts. The authors labeled the data set per hand by performing pulse shape cuts on the detectors individually, as it is done in a standard analysis chain. The data labeling was done using the Cait analysis package [5].

Machine learning was previously used for the event selection in similar cryogenic detectors. The discrimination between different pulse shapes from different particle recoil types or detector components was studied in Refs. [156, 167–172], unsupervised data cleaning with autoencoders was studied in Refs. [156, 174] and with a principal component analysis (PCA) in Ref. [176]. While the method proposed in Ref. [4] is trained in a supervised fashion, it works on previously unseen detectors, where the shape of the pulse shapes is yet unknown. This is accomplished by pre-training classifier models on historical data. Therefore, all types of pulse shapes are positives, while all shapes deviating from a pulse-like shape are negatives.

Ref. [4] also discusses potential combinations between the proposed pre-trained discrimination method and the PCA method from Ref. [176] to cluster different types of pulse shapes for further analysis.

The fact that their method does not require a training set from the detector at hand is convenient and prevents potential selection bias induced by subjective choices from a human analyst.

This section is structured as follows: in Sec. 6.1.1 we describe the used data set for training the classifier models; in Sec. 6.1.2 the training process and model architectures models are summarized, and in Sec. 6.1.3 we report on the final performance metrics obtained with the pre-trained models on a test set. An extensive summary and

discussion of this method, jointly with another one for automation in the detector optimization process, follows later in this chapter, in Sec. 6.3.

6.1.1 Large-scale historical data set

The data set used in Ref. [4] was put together from several tens of hours from all measurements that were performed in the main CRESST setup between 2013 and 2019. Since CRESST traditionally uses detector modules with multi-channel readouts, the individual channels, mostly phonon and light channels, were interpreted as individual detectors for the data set. Additional TES are used as veto channels on holding structures in several detector modules. They were also treated as individual detectors. The data set consisted of 68 detectors and more than a million triggered records. Only calibration data from the periods where the detectors were exposed to strong neutron or γ flux from americium beryllium and cobalt-57 sources was used to prevent potential bias effects if the trained models would later be applied to background data from the same detectors. Seven detectors were chosen to make up a test set, and five percent of the remaining data was randomly selected to build a validation set. The sizes of the resulting data sets are summarized in Tab. 6.1.

Data set	# of records
Training	930,368
Validation	49,024
Test	78,084

Table 6.1: The sizes of the used data sets. Tab. and caption also used in Ref. [4].

The rejection regions on the pulse shape features that define the positive and negative labels (see also 3.1.2.4) were set with the rectangle and lasso selection tools of Cait's VizTool (see also Sec. 3.1). The objective of the applied cuts is described in Ref. [4] as:

- reject all jump, drift, spike, glitch, and pile-up artifacts that deviate significantly from a recoil-type pulse shape;

- reject all pulse shapes that rise far away from the trigger position at 1/4 of the record window or do not decay within the window;

- accept all pulse shapes that fit the above criteria, not only those from target recoils, even if they show saturation effects typical for high energy recoils;

- accept empty noise traces if their slopes are within the typical slope of noise traces for the corresponding detector.

A small share of events was mislabeled by the applied cuts since rejection regions defined in one- or two dimensions cannot perfectly separate artifacts from pulses. Overall, 83.6 % of the events in the data set are positives. The voltage traces of all events were sampled with a frequency of 25 kHz, and the records have a length of 8192 or 16384 samples, depending on the measurement campaign.

The data was preprocessed to convert it to the format that is most useful for training classifier models. The voltage traces were downsampled to 512 samples by averaging over 16/32 samples in the record window. All records were normalized individually such that every record's minimal value is 0 and maximal value is 1. These manipulations reduce the precision of the time resolution and the information about the energy of the event. However, this information is not useful for the classification task anyway and discarding it significantly reduces the computational cost of training classifiers.

Data augmentation was used to counteract three potential bias effects:

1. Pile-up events are especially tricky to discriminate. A share of the positive events was superposed with a time-shifted copy of themselves to introduce more pile-up events into the data set artificially. The resulting events were relabeled as negatives.

2. Gaussian noise was superposed on the voltage trace to reduce the SNR and increase the robustness of the classifiers to high-noise conditions.

3. The record windows were randomly shifted by up to 26 samples to increase the robustness of the classifiers against small time shifts.

The augmentations were applied on the fly when mini-batches of data were pulled from the data set. The three types of augmentations were applied with chances of 20%, 20%, and 100%, respectively. A mini-batch of data from the training set is shown in Fig. 6.2.
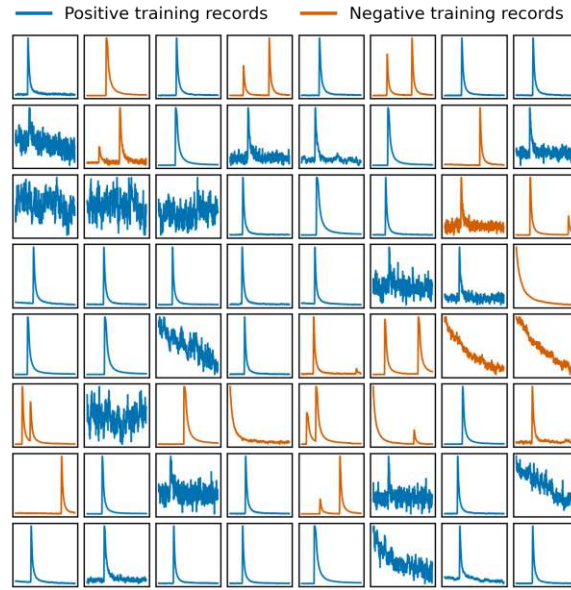
Figure 6.2: A mini-batch of 41 positive (blue) and 23 negative (red) records from the training set, all from the same detector. About half of the negative records are created from positive ones, with a data augmentation technique (see text). At least one record (first row, second column) is wrongly labeled as negative. Fig. and caption also used in Ref. [4].

6.1.2 Models and training

Classifiers were trained to map the raw time series to a number between 0 and 1, corresponding to the degree of belief that the time series shows a pulse. Four different models were tested in Ref. [4]: two different sizes of CNNs, a bidirectional LSTM [201], and a transformer. Since the architectures are dedicated to time series we call the larger CNN a time series convolutional network (TSCN) and transformer a time series transformer (TST). The transformer architecture was previously used in Ref. [202] and uses linear embedding layers for building the key, value, and query functions that govern the internal self-attention layer. The details of the model architectures are shown in Tabs. 6.2, 6.3, 6.4, and 6.5. All models were implemented in PyTorch [163].

The binary cross entropy loss function was used for training. The loss from negative records was weighted such that all positive records had the same impact as all negative ones. If we did not weight the loss values it would be beneficial for the classifier to preferably classifying events as positive.

All models were trained for 15 epochs. The mini-batch size used during the training was set to 64, and the order of the mini-batches was randomized. Training was performed on a Tesla P100 GPU on the Vienna CLIP cluster, using the ADAM

Layer	Specifications
1D convolutional	kernel size 8, 4 output channels, pooling 4, stride 1, ReLU activation
1D convolutional	kernel size 8, 16 output channels, pooling 4, stride 1, ReLU activation
1D convolutional	kernel size 8, 64 output channels, pooling 4, stride 1, ReLU activation
Fully connected	320 input nodes, 200 output nodes, ReLU activation function
Fully connected	200 input nodes, 84 output nodes, ReLU activation function
Fully connected	84 input nodes, 1 output nodes, sigmoid activation function

Table 6.2: The details of the CNN architecture. Tab. and caption also used in Ref. [4].

Layer	Specifications
Bidirectional LSTM	3 layers, input size 8, hidden size 200
Fully connected	12800 input nodes, 1 output node, sigmoid activation function

Table 6.3: The details of the LSTM architecture. Tab. and caption also used in Ref. [4].

optimizer [165]. The used learning rates are summarized in Tab. 6.6. They were optimized with the cyclic learning rate finder technique described in Ref. [203].

The training loss for each mini-batch and the validation loss after each epoch are shown in Fig. 6.3. The typical elbow shape of the training curve is clearly visible, where the TSCN reaches the lowest values, followed by the LSTM and the CNN. The TST is relatively far behind and catches up towards the end of the training. The validation loss decreases except for the TST where it increases towards later episodes, a clear sign of overfitting. The models from the episode where the lowest validation losses were achieved are saved and used for the tests reported later in Sec. 6.1.3. The validation loss, usually expected to be higher than the training loss, is in our case lower than the training loss, since data augmentations are only applied during training.

Layer	Specifications
1D convolutional	kernel size 3, 16 output channels, stride 1, padding 1, ReLU activation
1D convolutional	kernel size 3, 16 output channels, pooling 2, stride 1, padding 1, ReLU activation
1D convolutional	kernel size 3, 32 output channels, stride 1, padding 1, ReLU activation
1D convolutional	kernel size 3, 32 output channels, pooling 2, stride 1, padding 1, ReLU activation
1D convolutional	kernel size 3, 64 output channels, stride 1, padding 1, ReLU activation
1D convolutional	kernel size 3, 64 output channels, pooling 2, stride 1, padding 1, ReLU activation
1D convolutional	kernel size 3, 128 output channels, stride 1, padding 1, ReLU activation
1D convolutional	kernel size 3, 128 output channels, pooling 2, stride 1, padding 1, ReLU activation
Fully connected	4096 input nodes, 1024 output nodes, ReLU activation function
Fully connected	1024 input nodes, 512 output nodes, ReLU activation function
Fully connected	512 input nodes, 1 output nodes, sigmoid activation function

Table 6.4: The details of the TSCN architecture. Tab. and caption also used in Ref. [4].

Hyperparameter	Value
n. blocks	1
n. heads	8
dim. model	64
dim. linear	256

Table 6.5: The hyperparameters of the TST architecture. Tab. and caption also used in Ref. [4].

185

Model	Learning rate
CNN	$5 \cdot 10^{-4}$
LSTM	10^{-4}
TSCN	$5 \cdot 10^{-5}$
TST	10^{-5}

Table 6.6: The learning rates used for the training process of the models. Tab. and caption also used in Ref. [4].
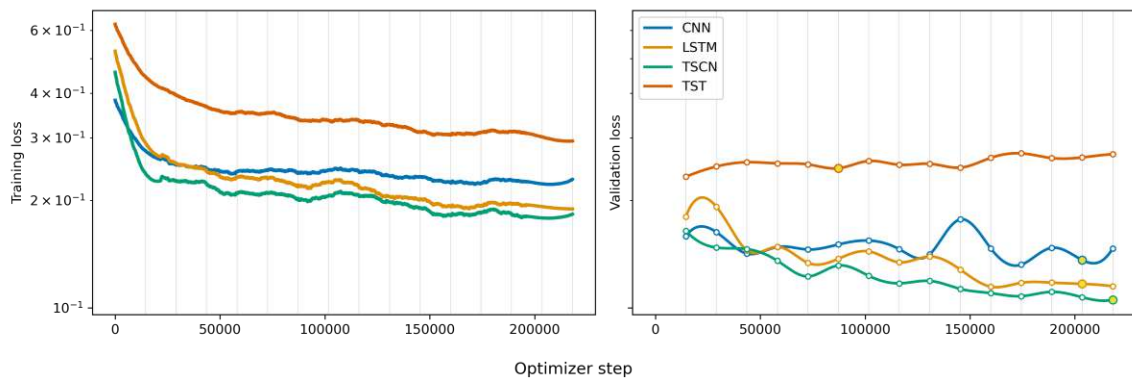

Figure 6.3: Progression of loss values throughout the training process for the four considered models. (left) Loss on the training set, recorded for each optimizer step. (right) The loss on the validation set is evaluated at the end of each epoch. The spline interpolation is a guide for the eye. The yellow dots indicate the point in the training process, where the model reached the best agreement between labels and predictions (accuracy) on the validation set. The bumps in the validation loss, clearly visible for the CNN around 150 k steps, are a typical artefact of stochastic optimizers. Fig. and caption also used in Ref. [4].

6.1.3 Performance on unseen detectors

The performance of the classifiers was evaluated on the test set, and several metrics were calculated in Ref. [4]. Those were defined as:

Recall	$R := TP/T,$
Selectivity	$S := TN/N,$
Balanced Accuracy	$BA := (R + S)/2,$
Precision	$P := TP/(TP + FP),$
Integral Over Recall	$IOR := \int_\Omega R(\mu)d\mu,$
Integral Over Selectivity	$IOS := \int_\Omega S(\mu)d\mu,$

where T are positive records, N are negative records, TP are true positive predictions, and TN are true negative predictions. $R(\mu)$ and $S(\mu)$ are the recall and selectivity as

functions of pulse shape features. Ω is the region over which the recall or selectivity was integrated, and μ is a placeholder for one or multiple pulse shape features (e.g. the PH) w.r.t. which the integration was performed.
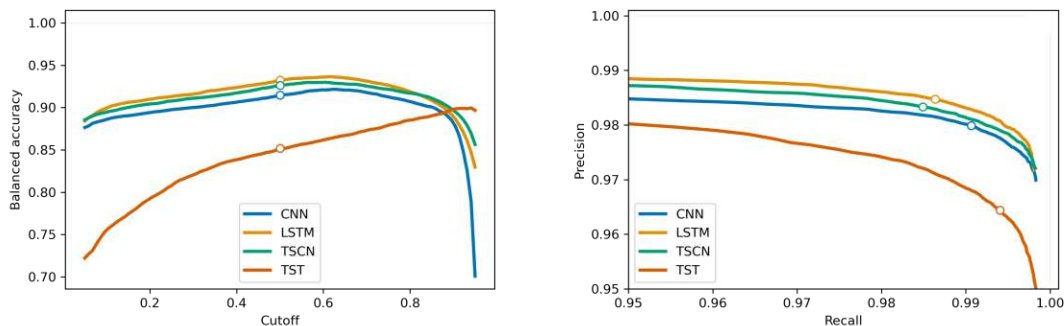

Figure 6.4: Metrics of all classifier models, under varying cutoff values, evaluated on the test set. The white dot marks the default cutoff value of 0.5. (left) The balanced accuracy w.r.t. the cutoff value. (right) The precision vs. recall curves, for cutoff values between 0.05 and 0.95. Fig. and caption also used in Ref. [4].

Model	Accuracy (b.)	Recall	Precision	IOR	IOS	Runtime (s)
CNN	.915	.991	.98	.970	.905	**1.9 (1)**
LSTM	**.932**	.986	**.985**	.993	.97	12.5 (6.5)
TSCN	.926	.985	.983	.987	**.976**	3.0 (1.6)
TST	.852	**.994**	.964	**.998**	.689	23.2 (12.1)

Table 6.7: The metrics of the trained models, evaluated on the test set and simulated data. (col. 2-4) The balanced accuracy, recall and precision on the test set with a cutoff value of 0.5. (col. 5, 6) The IOR score for the simulated positive particle recoils, and the IOS score for the simulated negative pile-up events. The values are defined in the text. (col. 7) The runtime for predicting the whole test set. The runtime divided by the lowest runtime is given in brackets. Tab. and caption also used in Ref. [4].

A cutoff value on the degree of belief of the classifier was defined. Above it, the event is accepted and below it, the event is rejected. The evaluated metrics as a function of the cutoff value are shown in Fig. 6.4. A natural trade-off between recall and precision exists and is clearly visible for the trained classifiers in Fig. 6.4 (right). This curve is also called the receiver operating characteristic (ROC) curve, and the its are is called the area under the ROC curve (AUC). Generally, classifiers are interpreted as reaching a higher quality if their AUC is higher (shown in Fig. 6.4 right). The LSTM

classifier achieved the largest AUC. This model's accuracy is relatively insensitive to the cutoff value for values close to 0.5. A generic choice of 0.5 as a cutoff is therefore suitable. The fact that a value of 0.5 is not fine-tuned is clearly a further asset of the choice.

The metrics for the generic cutoff value are shown in Tab. 6.7 (col. 2-4). The best performance is reached by the LSTM model, only closely ahead of the TSCN classifier. However, the latter one is a factor ≈ 4 faster in making predictions since it does not have to iterate through the record window recursively compared to the LSTM model. The runtime was evaluated with a batch size of 32, and the total time required for making predictions on the whole test set in Tab. 6.7.
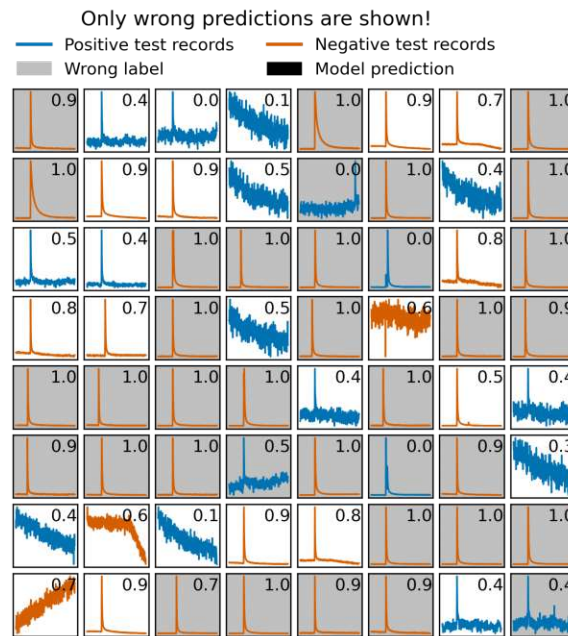

Figure 6.5: A batch of events from the test set that were wrongly predicted by the LSTM. The grey color indicates wrong labels. Some records, among them the tilted baselines, can hardly be flagged as positive or negative without additional context, namely the distribution of the remaining data of the corresponding detector. Fig. and caption also used in Ref. [4].

Further evaluations of the model performance were conducted. This includes an event-by-event analysis of the wrong predictions to understand the internal mechanics of the model's decision-making. An exemplary detector from the test set was chosen, from which for 70 out of 8422 records the LSTM predictions disagreed with the human-made cuts (labels). A visual inspection of the 70 events (also shown in Fig. 6.5) revealed that 39 of them were wrongly labeled, i.e. the human-made cuts rejected them despite the fact that they were reasonably normal pulses, or the other way around.

188

For noise traces the ground truth is detector-dependent. Therefore a disagreement on such events should not be interpreted as too harmful. In most practical analysis chains they would not be triggered anyway and are therefore irrelevant.
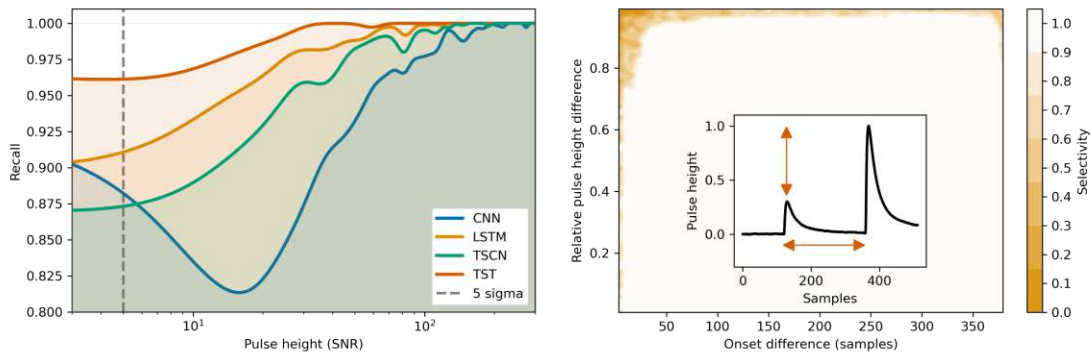

Figure 6.6: Metrics of the classifier models, evaluated on simulated data. (left) The recall values w.r.t. the SNR of simulated events. The recall drops towards lower values of the SNR, but is still reasonably high around a typical trigger threshold value of 5 baseline noise resolutions (grey, dashed). The reason for the local minimum of the CNN curve above 10 SNR is not cogently clarified. The most likely hypothesis is the absence of many low energy pulses in the training set, which can introduce a bias in models predictions. The simultaneous dip in the recall of multiple models around 80 SNR is a small sample effect of the simulation: it could be connected to two simulated events with similar energies, with relatively strongly tilted baselines. (right) The selectivity values for the LSTM model on simulate pile-up events, featuring two pulses, w.r.t. the difference in onset and relative difference in PH. Only pile-up events with large relative PH differences or very small onset difference are not rejected by the model. (right, inset) An example of a simulated pile-up event. The area that is covered by the inset holds only selectivity values of one. Fig. and caption also used in Ref. [4].

A further evaluation of the classifier performance was conducted using simulated events. These events have the advantage of making a ground truth without noisy labels accessible. Since the classifiers were trained with measured instead of simulated data, the evaluation of simulated data is more realistic and trustworthy, and overfitting of the systematics of the simulation can be excluded. The events were simulated as was done in Secs. 3.1.4 and 3.1.2.6 for single pulse shapes and pile-up events. Noise traces were simulated from the NPS of an exemplary detector from the test set, using the methods discussed in Ref. [108].

A data set of 50,000 events was simulated for pulses and pile-up events each. The PHs were distributed uniformly between zero and 300 times the SNR of the corresponding detector. The second pulse onset was uniformly distributed for pile-up

events in the record window, while the first pulse was put at 1/4 of the window. The recall and IOR were calculated with the simulated pulses depending on the PH in units of the SNR. For pile-up events the selectivity and IOS were calculated depending on the difference in PH and onset of the two pulses. The results are shown in Fig. 6.6 and values are also reported in Tab. 6.7. The IOR and IOS were calculated for the whole plotted range.
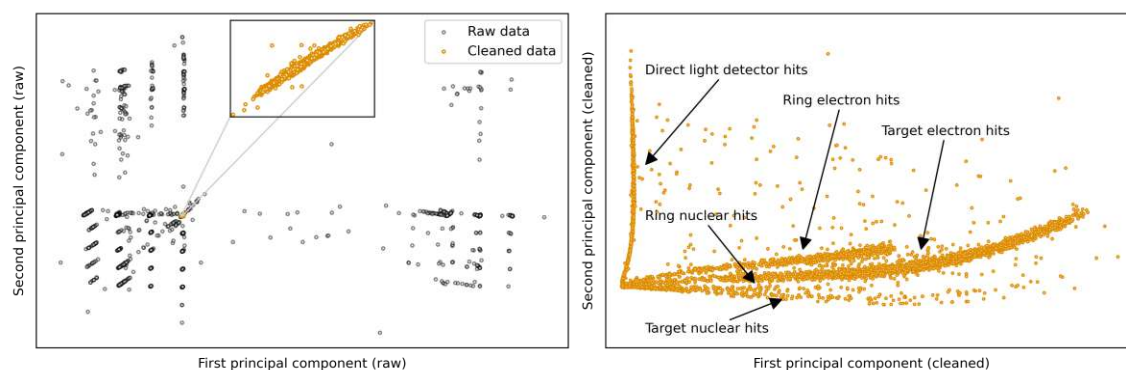

Figure 6.7: The data manifold visualized with the first two principal components. (left) The RD, without cleaning (black) and the cleaned data (orange), both projected to the first and second principal components of the RD matrix. (right) The cleaned data projected to the first and second principal components of the cleaned data matrix. The lines corresponding to the individual event types are clearly visible. The PH spectrum of the target channel is shown in Fig. 6.8. Fig. and caption also used in Ref. [4].

An exemplary detector module was chosen from the test set for the final evaluations. This detector module consists of three individual detectors and three readout channels. Two of them were used as phonon channels, with one TES on a holding structure called a ring and another TES on the target. One channel was a light channel, where the TES was placed on a beaker-shaped light detector that fully surrounded the target. The data for this module was triggered in a multi-channel readout mode. Therefore for each event voltage traces exist for each of the tree channels. The voltage traces corresponding to events of all three channels were cleaned with the LSTM model, and only events where the records of all three channels survived are considered clean.

For a visualization of the data distribution the records were concatenated and a PCA was performed on the data matrix of these sample trajectories that were then 1536 samples long. The first and second principal components of this data matrix were extracted for scenarios where the data matrix only included clean events as well as where it included all triggered events. The projections to these components are shown

in Fig. 6.7, with the components from the RD matrix on the left and the components from the cleaned data matrix on the right. The components from RD mostly identify SQUID jumps, leading to a raster-like structure of artifacts in the data manifold, while all pulses are collapsed to a small region of the transformed space.

For the cleaned principal components the meaningful data structure of different types of pulses is clearly visible. Bands corresponding to specific particle recoil types can be identified. The second principal component correlates strongly with the light channel, while the first principal component includes information from both phonon channels. Electron and nuclear hits can be discriminated by the strength of their light signal, and the same counts for direct light detector hits and ring and target hits. The ring structure is thermally coupled to the light-detecting beaker. Therefore, a stronger signal in the light channel is expected for hits in the ring. Finally, another event distribution is visible in Fig. 6.7 (right), which also appeared in Chap. 5. While this distribution features a strong signal in the light channel, it also shows a non-negligible signal in the phonon channel, leading to a widespread cloud of these events above the ring and target hit bands. While these events were identified as foil events in Chap. 5, we cannot assume the identical origin for this different detector design.
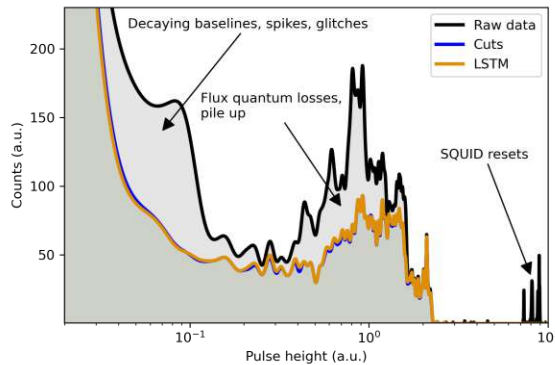

Figure 6.8: The PH spectrum of an exemplary detector without cleaning (black), with the cut analysis that we used as labels (blue) and the LSTM predictions (LSTM). The blue and orange curves almost fully overlap due to the strong agreement between cuts and LSTM. The data manifold of the corresponding 3-channel detector module is visualized in Fig. 6.7. Fig. and caption also used in Ref. [4].

The PH of the voltage trace is approximately proportional to the recoil energy released into the detector. The PH distribution for the target channel of the identical detector module as was used for Fig. 6.7 is shown in Fig. 6.8, in comparison for cleaning with human-made cuts and the events accepted by the LSTM. The two distributions

show a very good agreement. In comparison, the uncleaned PH distribution is also shown, which features significant bumps that are identifiable as various artifact classes.

We summarized the work originally presented in Ref. [4] on automated data cleaning. It was shown that NNs can be trained to event classifiers and perform the task of data cleaning with the same accuracy as humans using quality cuts. The pre-trained event classifiers can then especially be used instead of creating individual new cuts for each detector. The runtime of the trained models was in the order of several thousand predictions per second. A balanced accuracy score of 0.932, a recall of 0.986, and a precision of 0.985 were recorded for the best-performing LSTM network on a data set of historic, measured data and labeled by human-made cuts. A notable half of the wrongly predicted events were identified as wrongly labeled events. The IOS and IOR scores on simulated data were almost unity.

For the differences between the trained models we want to note that the trained transformer model could perform better with more training data or a dedicated, pre-trained feature extraction model, as often used for audio data [204].

The impact of this method on large-scale detector setups and their analyses is discussed in more detail in Sec. 6.3.

We continue in the following Sec. 6.2 with a second machine learning method for the automation of the initial optimization process of the detector control parameters.

6.2 Detector setup with reinforcement learning

The operation conditions of a cryogenic detector with an attached TES are mostly dependent on three influence variables impacting the state of the superconducting film. The most crucial of these variables is the temperature, which should be stable throughout the measurement but also tuned to find an opportune OP in the superconducting transition curve. The temperature is mostly controlled by the voltage applied to the heating resistor, governed by a parameter called DAC. The second parameter is the bias current IB lead through the film. A higher bias current emphasizes film-internal noise over external noise in the readout circuit and has a noticeable impact on the shape of the transition curve. Finally, the magnetic field has a noticeable impact on the shape of the transition curve and the noise conditions.

Two of these quantities are actively controlled in a default setup, the bias current and the temperature. They need to be placed in suitable settings in an optimization process at the beginning of measurements, often called the detector setup process. The magnetic field is ideally canceled by an active veto or superconducting shielding. It is therefore not included in a standard optimization procedure.

The default process for the optimization is to inject TPs with different TPA values through the heating resistor to observe the detector response in the current state. Usually, sweeps of the DAC value are performed for a small set of constant IB values. This procedure is not ideal for multiple reasons: the noise conditions are mostly neglected, only the observed PH is optimized for, and the procedure takes several hours and requires a significant amount of manual interventions. Optimal values for the control parameters are not predictable from theory due to the complexity of setup and detector response. Furthermore, the TES accumulates an internal state due to Joule heating, also called self-heating, when the film is not fully superconducting. This state-dependency generally confuses standard optimizers due to possibly different detector responses despite the same control settings. Therefore a state-dependent optimization algorithm is required, or the problem should be considered as optimizing a trajectory.

Ref. [3] proposes to automate this process using RL. RL is a framework to model decision-making problems that depend on a state and potentially have time-delayed results. In RL, an agent receives a state and a reward by taking actions in an environment. The objective of the agent is to maximize its returns, which are the sums of rewards over time. RL has already been used to optimize control settings in

physics in Refs. [205–207] for particle beams, in Ref. [208] for nuclear fusion reactors and in Ref. [209] for superconducting quantum bits.
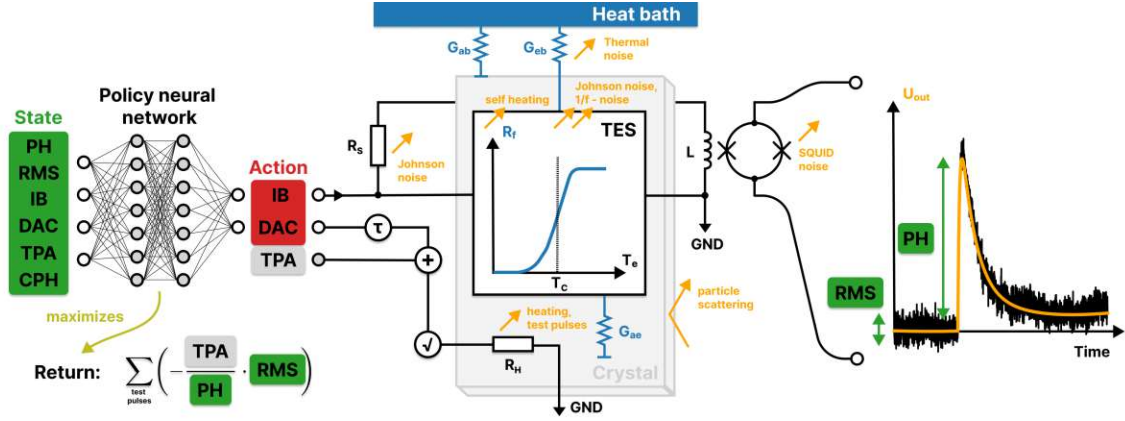

Figure 6.9: Schematic drawing of the detector environment. (center) The detector can be described as an electrothermal system, where the readout and heater electronics and the temperatures in the crystal and sensor interact with each other. (right) The recorded observable from particle recoils is a pulse-shaped voltage signal. (left) A NN is trained with RL to control the measurement settings w.r.t. the SNR of the observed pulses. See text in Sec. 2.1 and 6.2.1 for details. Fig. and caption also used in Ref. [3].

The situation of optimizing the detector with RL is depicted schematically in Fig. 6.9. The detectors used for the experiments in Ref. [3] were operated in the main CRESST setup at LGNS and are the same as reported on in Chap. 5: Li1P, Li1L, and Li1L. For the purpose of the experiments, they were treated as three independent detectors.

Virtual twins of the detectors were built, using the physics parameters determined in Sec. 5.2, to test the approach. The training procedure and results are summarized in Sec. 6.2.3. Furthermore, the system was tested live on the CRESST setup on the real-world version of the detectors. Results are summarized in Sec. 6.2.4. In preparation for the experiments, in Sec. 6.2.2 the reward function is defined, and in Sec. 6.2.1 we provide a general introduction to RL.

6.2.1   Reinforcement learning

RL is used to find optimal policies to decision-making problems or trajectories optimizing state-dependent processes. A long introduction can be found in Ref. [210], and a concise introduction will be provided here.

The original formulation is based on Markov decision processes (MDPs). An MDP is a 4-tuple of a state space $\mathcal{S}$, an action space $\mathcal{A}$, a dynamics function p, and a reward
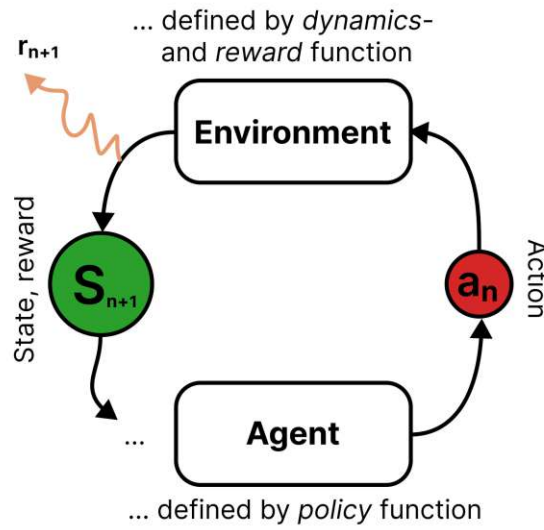
Figure 6.10: The mechanics of RL: an agent follows a policy function to interact with an environment. The environment, defined by its dynamics and reward function, responds to the agent's actions with a reward and observable state. Fig. adapted from Ref. [210]. Fig. and caption also used in Ref. [3].

function r. The dynamics function maps action-state pairs to transition probabilities:

$$p\colon \quad (A, S) \mapsto \{\text{probabilities for } S'\}, \quad A \in \mathcal{A}, \quad S, S' \in \mathcal{S}.$$

The reward function defines a scalar for each state transition and action combination:

$$r\colon \quad (S, A, S') \mapsto R \in \mathbb{R}.$$

The agent follows a policy function that provides action probabilities for a given state:

$$\pi\colon \quad S \mapsto \pi(A \mid S) = \{\text{probabilities for } A \text{ given state } S\}.$$

The agent searches for an optimal policy function that maximizes returns. The framework is depicted in Fig. 6.10. Dynamics and reward function are not observable for the agent and must be learned from experience, i.e. the data observed from the environment. This formulation of an MDP satisfies the Markov property automatically, which states that the system's dynamics depend only on its current state and not on prior history. However, while most theorems for convergence derived for RL assume MDPs, the framework also works well in many settings where the Markov property is violated.

For using RL as a state-dependent optimizer for cryogenic detectors, Ref. [3] introduced definitions of the state and action spaces. The procedure is to repeatedly

and alternatingly inject TPs and CPs, and the spaces are defined depending on observables from those pulses and the control parameters that are currently set:

$$\mathcal{S} \coloneqq \{PH, RMS, IB, DAC, TPA, CPH\},$$
$$\mathcal{A} \coloneqq \{DAC, IB\},$$

Here PH and RMS are the PH and the pre-trigger region root mean square (RMS) value of the TP, respectively. TPA is the TPA of the TP, and CPH is the PH of the CP. Several of these quantities scale, among other dependencies, proportionally with the IB value. To reduce these unnecessary correlations in the state space, the PH, RMS, and CPH values were divided by the IB value. Furthermore, all values were normalized such that they are contained in the interval (-1, 1). The used reward function is derived in Sec. 6.2.2.

The exact RL algorithm used in Ref. [3] was the Soft Actor-Critic (SAC) algorithm. The algorithm was originally introduced in Ref. [211], and is one of the state-of-the-art algorithms for deep RL in continuous state and action spaces. It showed good performance on real-world robotics tasks in Ref. [211]. Actor-critic algorithms use an additional value function that estimates the future returns for given state-action pairs:

$$q\colon \quad (S, A) \mapsto Q \in \mathbb{R},$$

In the context of Actor-Critic algorithms the policy function is also called the actor, and the value function is the critic. Both are realized as NNs π_ϕ, q_θ and their weights ϕ, θ are trained with gradient descent algorithms. The actor is additionally parameterized as a Gaussian function with the dimensionality of the action space to have an accessible probability distribution for the actions. The NN outputs are then the mean and standard deviation values for each dimension in the action space.

The data collected from the environment is stored in an experience replay buffer that consists of all past transitions and the corresponding rewards (S, A, R, S').

The NNs are trained on the collected data, where the loss function critic for the value function is the soft Bellman residual

$$J_q(\theta) \propto \left(q_\theta(S, A) - \left(R + \gamma \left(q_{\overline{\theta}}(S', a') - \alpha \ln \pi_\phi(a'|S') \right) \right) \right)^2,$$

with $a' \sim \pi_\phi(\cdot|S')$ and α the temperature, a parameter of the algorithm. The so-called discount factor γ emphasizes immediate rewards over the long-term future. The weights $\overline{\theta}$ are those of the target critic network, explained at the end of this section.

The loss function used for the policy network is

$$J_\pi(\phi) \propto \alpha \ln \pi_\phi(a|S) - q_\theta(S, a),$$

with $a \sim \pi_\phi(\cdot|S)$. It balances entropy of the action with exploitation.

There are several more details about the SAC algorithm that contribute to a stable and fast convergence: two critic functions are trained in parallel, their minimum is used for inference; the target critic's weights are exponentially smoothed $\bar{\theta}$ versions of the critic's network weights; in every step, the α parameter is updated with the goal that the target entropy is retained.

In Ref. [3] an additional adjustment to the algorithm was introduced. The authors lower the target entropy with each gradient step. They argue that this enables the Gaussian actor to converge towards smaller details in the control parameter space, which would otherwise be smeared out. A detailed discussion of this effect is given in Sec. 6.2.5.2, and the results of this effect are mentioned again in Sec. 6.2.4.

6.2.2 A metric for low thresholds

For using RL to optimize control parameters of cryogenic detectors an objective function needs to be defined first. This function needs to balance several objectives. High-energy pulses tend to saturate when the corresponding temperature increase in the sensor reaches the upper end of the transition curve, and low-energy pulses can drown in sensor noise. Both effects should ideally be avoided. As a third objective, an optimal control parameter setting should provide stable operation conditions. Any setting will be a trade-off between dynamic range, noise conditions, and stability.

It was shown in Ref. [3] that these objectives are equivalent to minimizing the energy threshold of a detector

$$\underset{DAC,\ IB}{\arg\min}(E_{th}). \tag{6.1}$$

Using a small-signal approximation the energy thresholds can be written with a linear conversion constant γ as a voltage threshold U_{th}[1]:

$$\cdots = \underset{DAC,\ IB}{\arg\min}(\gamma U_{th}). \tag{6.2}$$

The voltage threshold is typically defined as a multiple of the noise resolution:

---

[1]It should be noted that γ appears twice in this section, with different meanings: in the context of RL-algorithms, γ is the discounting factor. In the context of this derivation, γ is the linear PH-energy conversion factor.

$$\cdots \propto \operatorname*{arg\,min}_{DAC,\ IB}(\gamma RMS). \tag{6.3}$$

Finally, the conversion constant γ can be estimated from the ratio between an injected TPA_i and the observed PH_i :

$$\cdots \propto \operatorname*{arg\,min}_{DAC,\ IB}\left(\sum_{i=1}^{N}\frac{TPA_i}{PH_i}RMS_i\right). \tag{6.4}$$

In the last step we switched from a formula for a single injected TP to that calculated from a number of injected TPs. Multiplying Eq. (6.4) with -1 provides a return function suitable for maximization:

$$\operatorname*{arg\,max}_{DAC,\ IB}\left(-\sum_{i=1}^{N}\frac{TPA_i}{PH_i}RMS_i\right). \tag{6.5}$$

The rewards can be additionally weighted by the inverse of the TPA to emphasize the search for best noise conditions over the objective of a large dynamic range, as for many physics searches this poses the most relevant objective:

$$\operatorname*{arg\,max}_{DAC,\ IB}\left(-\sum_{i=1}^{N}\frac{1}{PH_i}RMS_i\right). \tag{6.6}$$

This function has several convenient properties:

1. The TPA value is not explicitly contained anymore in the equation. It could therefore also be used for triggered particle events.

2. The function can be evaluated on an event-by-event basis and therefore be used as a reward.

3. The value range of the function is restricted to the interval from -1 to 0, which makes it convenient to use.

This function realizes the balance between the three objectives as follows: the main objective is the optimization of the energy threshold. The saturation of pulses would appear as a violation of the small signal approximation. Flattening the pulses has the same impact on the reward function as worse noise conditions. Since the target of the optimization problem is the sum of the rewards, stable operation conditions are intrinsically contained in the target.

6.2.3 Tests in virtual environment

For testing the proposed method, virtual twins of Li1P, Li1L, and Li2P (discussed in Sec. 5) were built in the CryoEnv detector simulation. The CryoEnv simulation was then wrapped as an OpenAI Gym environment. The detector response was tested by injecting TPs with TPA values from 1 to 10 in integer steps, and 0.1 and 0.5, in successive and cyclic order, and CPs in between. The action that the agent could take after each TP and CP combination was to change the control parameters, as defined in Sec. 6.2.1.

The training was performed in episodes of 60 environment steps, and the control parameters were reset to an edge of the parameter space after each episode. Each step in the environment takes 10 seconds of equivalent measurement time since the system needs to equilibrate before control parameters can be readjusted. A total number of 315 agents were trained, equally distributed over the three detectors. Among these agents, 3 different scenarios of the environment settings were tested, as well as 7 different sets of hyperparameters of the agent. For each of these combinations, 5 agents with different random seeds were trained. The physics parameters of the detectors were randomized by 20 % for each version of the agents to ensure that the performance of the method does not degrade too much for adjustments in the detector design. Each training consisted of 40 episodes. The three scenarios considered were:

1. The first scenario did not involve any adjustments to the explained procedure.

2. In the second scenario a fast sweep of the control parameter space was performed at the beginning of the training. This sweep consisted of 120 environment steps, where the bias current oscillates between its minimal and its maximal values, while the DAC value was gradually decreased. This intervention is used to help the agent find the position of the superconducting transition in the parameter space.

3. The third scenario also started with a sweep of the parameter space. Additionally, the effect of adjusting the DAC value was delayed on an exponential time scale of 20 seconds to simulate the behavior of either slowly responding electronics or long thermal relaxation processes.

The agents were trained on the Vienna CLIP computing cluster as single CPU jobs that took about two hours to complete. Half of the runtime was spent in the simulation, which currently can be run on CPU only. Using GPUs would therefore not significantly

improve the runtime. Policy and value functions were both NNs with 2 layers and 256 nodes, ReLU activation functions, implemented in PyTorch [163]. For training the ADAM [165] optimizer was used with a weight decay of 10^{-5}. For the SAC algorithm the τ value, governing the exponential smoothing of the weights of the target networks, was set to 0.005, and the initial entropy α was set to 0.2. The training started once one minibatch of data was collected in the replay buffer from the environment. Gradients were clipped at 0.5.

The target entropy was set to the entropy of a multivariate Gaussian distribution with a given standard deviation in each dimension. At the beginning of the training this standard deviation was set to the value 0.088 and gradually decreased to 0.05 times its original value at the end of the training.
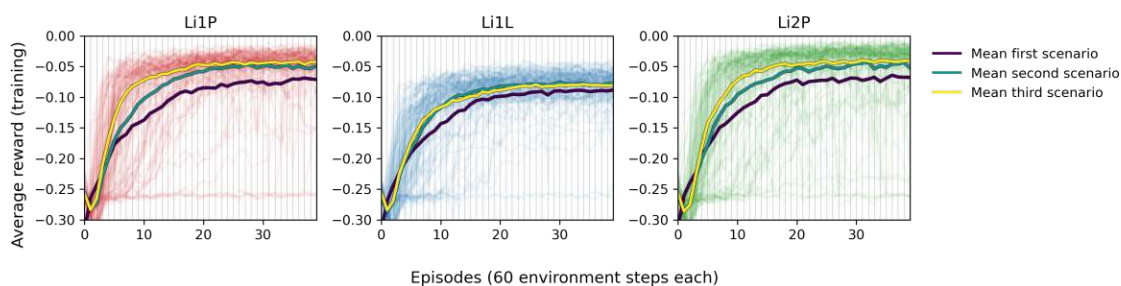

Figure 6.11: The average rewards per episode during training for all 105 versions of the three detectors Li1P (red, left), Li1L (blue, center), and Li2P (green, right). The thick lines are the mean values of all curves corresponding to the first/second/third scenario (violet/turquoise/yellow). The mean values rise close to the apex of the curves after 15 to 20 episodes. The second and third scenarios reach convergence significantly faster than the first. During the first 5 to 10 episodes, only little return is collected. The distribution of curves is clearly not normal distributed around the mean value, which is due to the different hyperparameter settings in the training of the individual detector versions. Fig. and caption also used in Ref. [3].

The average rewards obtained per episode during training are shown in Fig. 6.11. After 15 to 20 episodes, the agent converged to a high plateau of rewards. This corresponds to an equivalent measurement time of 2-3 hours, which is about the same time a human expert needs to perform the task. Generally, any intervention that shortens the exploration period speeds up the convergence but increases the risk of failure. The plot contains all trajectories from all hyperparameter settings and scenarios. According to the median values of the scenarios, it is clearly visible that the initial sweep of the control parameters has a positive impact on the speed of convergence and the number of agents that found an optimal setting after training concluded. A slight positive
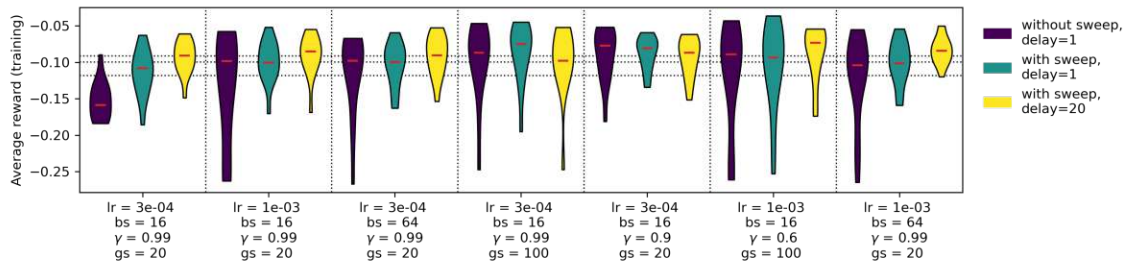
Figure 6.12: Average reward during all training episodes of the different versions of the virtual detectors, grouped into violins for each scenario (violet, turquoise, yellow) and setting of the hyperparameters learning rate (lr), batch size (bs), γ and gradient steps (gs). Each violin includes five versions of Li1P, Li1L, and Li2P each, sampled and trained with individual random seeds. The red bars indicate the mean values of the violins and the violins thickness the density of the represented return distribution. The dotted horizontal lines indicate the mean values of the collected returns of the first (lowest), second, and third (highest) scenarios of detector versions. The dotted vertical lines separate the violins with different hyperparameter settings. The values for the hyperparameters are written in the ticks on the abscissa. Fig. and caption also used in Ref. [3].

effect of the delayed DAC setting is visible. This is generally counterintuitive and will be discussed in detail later.

A detailed study of the used hyperparameters, performed in Ref. [3], is visualized in Fig. 6.12. In this plot, faster training and a higher final reward are collapsed variables and cannot from this plot alone be studied independently. However, it is also clearly visible that the initial sweep positively impacts the average rewards obtained. The authors of Ref. [3] contribute this effect to the beneficial covering properties of grid-based methods over Monte Carlo sampling for a parameter space with less than four dimensions. This argument is based on the fact that before the agent finds the region of control parameters that leads it into the superconducting transition, it can only randomly sample the parameter space without useful feedback about its choices. Once this region is found, the systematics of the temporal difference method take over, leading to a much faster convergence.

The delayed effect of the heating also causes slightly higher rewards. The origin of this effect is not fully understood. It is believed that the relaxation decreases the size of jumps the agent makes in the parameter space, which can lead to a better understanding of the system's mechanics. Also in the choices of hyperparameters, it can be seen that hyperparameters that emphasize a shorter exploration period can lead to a higher risk of overall failure. This especially includes the number of gradient

steps and the learning rate.

The performance of the trained models was tested by running inference trajectories on the detectors they were trained on. In an inference trajectory, the expected value of the Gaussian actor is taken instead of a random number drawn from the distribution. The behavior of the agents was generally that they moved directly to the found optimal control parameters. However, the agent changed the control parameters after each environment step. This is generally not beneficial for physics data taking, where stable and constant operation conditions are needed. This behavior can be explained by one of two possibilities: first, it is likely that different OPs in the superconducting transition are optimal for different injected TPA values. Second, the random initialization of the networks can lead to stochasticity in the choices of the network when different choices are equally good. The first of these phenomena is discussed in Sec. 6.2.4 using real-world data from the live CRESST experiment. The second phenomenon is studied in detail with a toy model in Sec. 6.2.5.1. A regularization term was added to the reward to mitigate the magnitude of jumps between control parameters. This term is the Euclidean distance between the current and the new control settings, weighted by a factor $\omega = 0.1$. The choice of ω is also discussed in Sec. 6.2.5.1. For larger values, the performance of the models degraded, while for smaller values, no effect was visible.
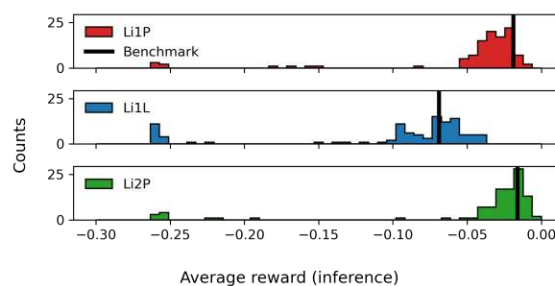

Figure 6.13: Histogram of the average reward achieved during inference trajectories with the trained agents for the 105 versions of Li1P (red, top), Li1L (blue, center), and Li2P (green, bottom) each. While rewards from versions with opportune choices of hyperparameters cluster around a benchmark value (black line), achieved by a human expert, other versions exhibit features in the histogram at lower rewards. The results of Li1L surpass the benchmark value, since high pulses saturate more strongly in this detector, which can be accounted for with the machine learning method. Fig. and caption also used in Ref. [3].

To arrive at a quantitative statement of the agent's performance, the average rewards obtained throughout the inference trajectory are shown for all trainings in Fig. 6.13. The values are compared with benchmark values that were recorded in a previous period of physics data taking with human-optimized control settings.

This comparison is subject to uncertainties: first, the randomization of the physics parameters changes the highest achievable rewards. Second, the figure includes all scenarios and hyperparameters; some were clearly more opportune for the results and optimization process than others. Third, the benchmark value originates from a stable period of data taking, where control parameters were not changed over a long time.

However, despite these uncertainties, the SAC agent performs similarly to the human-optimized benchmark value for good choices of hyperparameters. Furthermore, the optimization with SAC agents is about as fast as a human expert but can be parallelized for all detectors, while a human expert cannot reasonably divide their attention onto more than 1-2 detectors. These results confirm that RL is a promising candidate for automating the detector optimization process. Before we review the real-world operation results reported in Ref. [3], we discuss the optimization of more complicated detector designs with larger state and action spaces in the following section.

6.2.3.1 Detector designs with multiple components

An interesting question is the behavior of the RL optimization method for more complex detector designs. Recently designs with multiple TES, where heat signals are seen in both simultaneously, were used, and their optimization with the default methods creates significant overhead in the required time. Often only sub-optimal settings can be found. To test the performance of the automated method Ref. [3] built versions of Li1P, Li1L, and Li2P with 2 TES. The CryoEnv simulation can also accommodate much more complex designs since the system of ODEs describing the current and heat flow can easily be generalized into a matrix equation of an arbitrary number of thermal and electrical components:

$$
\dot{\underline{T}}(t) = \mathrm{diag}(\underline{C})^{-1}\bigg(\underline{P}\Big(t, \underline{T}(t), \underline{I_f}(t) \Big)
$$

$$
+ \mathrm{diag}\left(\underline{G_b}\right)\left(\underline{T_b} - \underline{T}(t)\right) + (\underline{\underline{G}} - \mathrm{diag}(\underline{G1}))\underline{T}(t)\bigg), \tag{6.7}
$$

$$
\dot{\underline{I}}_f(t) = \mathrm{diag}(\underline{L})^{-1}\bigg(\mathrm{diag}\left(\underline{R_s}\right)\underline{I_b}
$$

$$
- \mathrm{diag}\left(\underline{I_f}(t)\right)\left(\underline{R_f}(\underline{T}(t)) + \underline{R_s}\right)\bigg), \tag{6.8}
$$

where underlined (double underlined) quantities are vector (matrix) valued, and $\underline{\underline{G}}$ describes the symmetric matrix of thermal couplings between components. All

other quantities are equivalently generalized to vectors from Eq. (2.7). The noise is independently simulated for each TES and readout circuit since there are no correlations between the expected noise sources.

The optimization of the correlated heaters is harder since the parameter spaces are a factor of 2 larger. The physics parameters of the simulation were adjusted to the parameters of Li1P, Li1L, and Li2P. The parameters were changed to the expectation from a two-TES design. Specifically, the TES size, thermal couplings, and collection efficiency were divided into two.
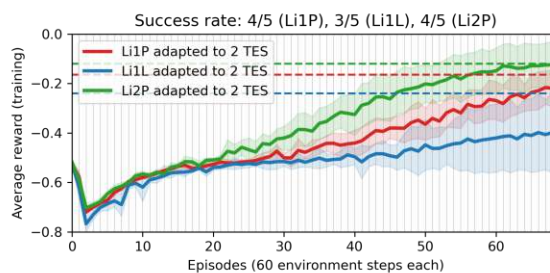

Figure 6.14: Return per episode for Li1P (red), Li1L (blue) and Li2P (green) adjusted to two TES. The thick lines represent the mean of five trained versions of the detectors, sampled with different random seeds. The shaded regions show the upper and lower standard deviations. We show benchmarks (dashed lines) for all three detectors. These benchmarks were calculated by taking the average reward in the last episode of the training for all versions of the single-TES detectors that were trained in Sec. 6.2.3, and multiplying it by two. The benchmark is reached by Li1P and Li2P, but not by Li1L. Fig. and caption also used in Ref. [3].

Five versions of each detector were trained, and the training took 6-8 hours each, as single CPU jobs. A five times higher target entropy was used than for the single TES detectors. For all trainings, the first scenario introduced in the previous sections was used. In Fig. 6.14 the obtained average rewards are shown. The training was successful for 4/5 versions of Li1P, 3/5 versions of Li1L, and 4/5 versions of Li2P. The others found only one of two superconducting transitions. A plausible explanation for this is that a good local optimum was found that was further separated in the parameter space from the global optimum. This could be solved by choosing the hyperparameters of the algorithm in a way that encourages more exploration.

Finding optimal control settings takes about twice as much equivalent measurement time for the detectors with two TES. The scaling behavior for larger setups is not predictable from these data. In a worst-case scenario, it could scale exponentially with the number of dimensions. However, NNs are known not to suffer from this curse of dimensionality in many other tasks. They are generally good at disentangling complex

dependencies and the identification of underlying, uncorrelated variables. There is therefore reason to believe that the scaling behavior would be better than exponential.

6.2.4 Live operation on the CRESST setup

Having established the applicability of the RL method in a virtual environment, the ultimate test is the training and operation on a live setup and in the real world. For this, a measurement spot of 12 days was dedicated to the main CRESST underground setup in February 2023, where the real-world versions of Li1P, Li1L, and Li2P were operated. These measurements have originally been reported on in Ref. [3], and are reviewed here.

The first week of operation was dedicated to the implementation of the technical setup and debugging. The communication between the Python-based RL control and the control and data acquisition (DAQ) software of the CRESST setup was implemented via the internet of things (IoT) protocol MQTT [212]. The DAQ system acted as the RL environment and pulse shape parameters were calculated on the environment side and broadcast via the MQTT broker to the RL model. The RL control was run on the Vienna CLIP cluster, and messages were sent through an SSH tunnel to the MQTT broker running in the LNGS network. The RL control side had simultaneously two processes running: one process to receive messages, calculate rewards, write to the buffer, query the policy model, and reply with the new control parameters. A second process continuously trained the models on the buffer. The replay buffed was implemented as memory-mapped arrays. This setup is also visualized in Fig. 6.15. With the current CRESST electronics, no parallel training on detectors is possible. All operations and training were therefore performed in succession.

In the given time frame a total of 48 experiments were performed, where the majority was dedicated to debugging of the system. Finally, 6 runs were performed to benchmark the performance of the method. Of these, one run was done with Li1P, two with Li1L, and three with Li2P. The training started in all runs with a sweep of the action space, and long-term thermal relaxation processes were observed. The real-world setting is therefore best compatible with the third scenario from Sec. 6.2.3.

The same NN architecture was used as in the runs in the virtual environment. The ADAM optimizer with a batch size of 16 was used, and 20 gradient steps with a learning rate of $3 \cdot 10^{-4}$ were taken after each environment step. Otherwise, the default settings of the optimizer were used. Gradients were clipped at the value of 0.5, and the initial entropy of the SAC algorithm was set to 0.2. The τ update parameter
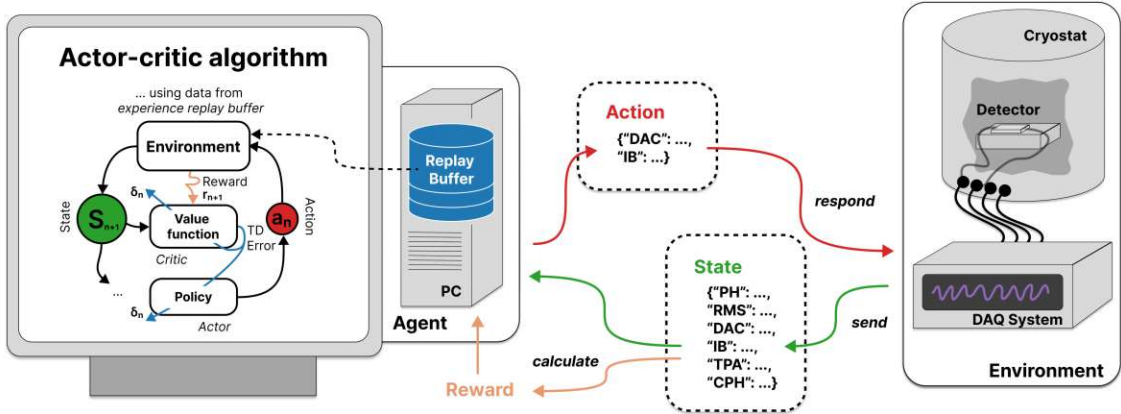
Figure 6.15: Schematic visualization of the implemented setup to optimize CRESST detector control. (right side) The detectors are operated in a cryostat and read out by a DAQ system. The parameters of recorded TPs are sent via an MQTT broker to a client as state. (left side) The client calculates the reward from the state, stores the data in an experience replay buffer, and responds to the DAQ system with new control parameters. An independent process trains the AC agent on the buffer. This is a symbolic visualization, the algorithm we are using is the SAC algorithm. Fig. and caption also used in Ref. [3].

Detector run	Li1P 1	Li1L 1	Li2P 1	Li2P 2	Li2P 3	Li1L 2
γ	0.9	0.99	0.99	0.99	0.99	0.99
Reward	Eq. (6.5)	Eq. (6.5)	Eq. (6.5)	Eq. (6.5)	Eq. (6.6)	Eq. (6.6)
TP int. (s)	20	20	10	10	10	10
DAC_{max}	10	10	5	5	5	10
IB_{min}	0.5	0.1	0.5	0.5	0.5	0.1
IB_{max}	5	3	5	5	5	3
ADC range	± 10	± 0.3	± 1	± 1	± 1	± 0.3
TPA in state	yes	no	no	no	no	yes
CPH in state	no	no	no	no	yes	yes
ADCs/IB	no	no	yes	no	yes	yes
ω	0	0	0	0	0.01	0.01

Table 6.8: Hyperparameter and settings of the RL problem used for the six performance runs on the CRESST underground setup. See the text for explanations. Tab. and caption also used in Ref. [3].

was set to $5 \cdot 10^{-3}$. The target entropy was set to the entropy of a 2-dim Gaussian function with a standard deviation of 0.088.

Individual adaptions to the hyperparameters, the configuration of the state space, and the number and length of training episodes were made in all trainings; they

are summarized in Tab. 6.8. The settings were chosen to prioritize fast and reliable convergence over optimality of the final configuration. For this, the ω values were generally chosen smaller, and the target entropy was fixed to its initial value and not reduced during training.

The unweighted reward function, Eq. 6.5, was used in some runs, whereas the reward weighted with the inverse TPA Eq. 6.6 was used in some other runs. A higher waiting time between TPs was tested to control the impact of longer relaxation processes.

The normalization intervals and value range of the analog-digital converter (ADC) were individually adjusted for the runs and detectors. Some runs contained the TPA and CPH in the state space, which enables to control for the usefulness of the contained information in these values. Furthermore, the ADC values were divided by IB for some of the runs.
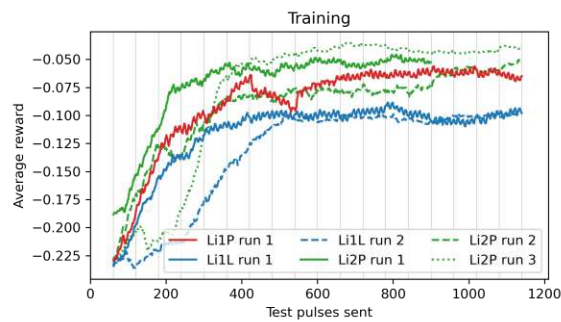

Figure 6.16: Average rewards per TP sent during the live training on the CRESST setup, smoothed with a moving average of 60 TPs. Results from six runs with different training settings are shown for Li1P (red), Li1L (blue, blue dashed), and Li2P (green, green dashed, green dotted). For this comparison, we re-calculated the rewards after training with Eq. 6.6, while during training for some of the runs, the unweighted reward function was used. Fig. and caption also used in Ref. [3].

The average obtained reward depending on the number of TPs sent since the start of the run are shown in Fig. 6.16. Convergence has been in all runs achieved before 600 pulses were sent or before 1.5 hours of measurement time elapsed. The total run time was dominated by the measurement time, not the time required for computing. Convergence was therefore reached faster than on average in the virtual environment, and also faster than a human expert.

It is generally hard to disentangle all effects of the adapted parameters with the limited amount of data. However, we believe that no impact of the longer waiting time between pulses was observed.

A larger state space generally needed more time to converge but provided a better responsivity to the environment. This is especially the case for the adjustments to different injected TPA values, which will be discussed in detail further below.

A faster convergence can also be attributed to the weighting of the reward function, the regularization factor ω, and the division of the ADC values by the IB value.

However, it cannot be claimed that the effects are significant compared to random fluctuations given the complexity of the system and the small data set.

It should also be noted that not all learnings from the virtual environment could be used for choosing the hyperparameters on the live setup since the studies in the virtual environment were not concluded before the measurement slot on the experiment.
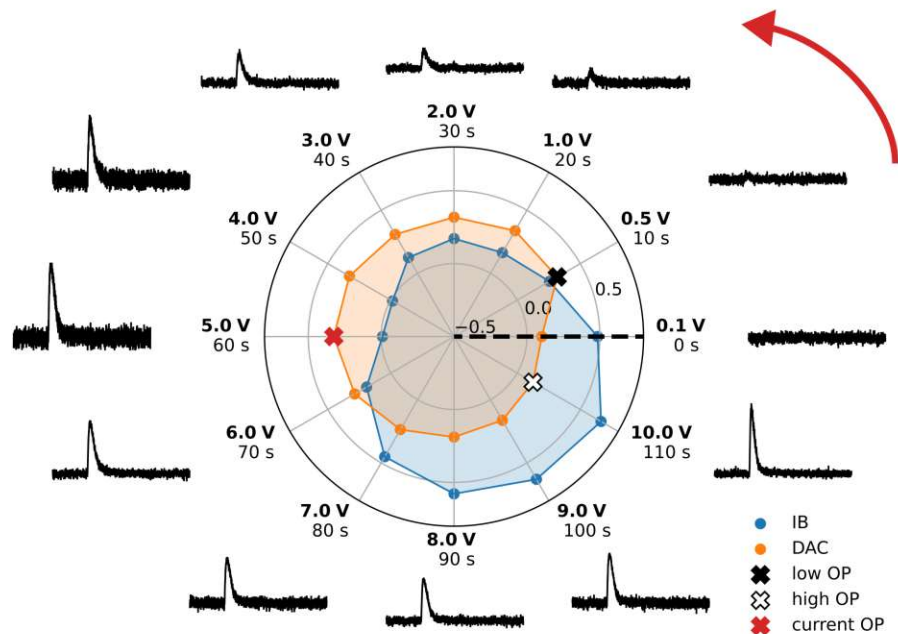

Figure 6.17: Visualization of the cyclic adjustment of the control parameters during an inference trajectory on Li1L, run 2. The ascending trajectory of injected TPs is visualized in the circle in anti-clockwise direction. The voltage traces of the observed pulses (black) are normalized to a fixed voltage interval. The pulses are normalized to the applied bias current, leading to smaller pulses and noise for higher IB. The TPA values (bold) and measurement time since the start of the TP trajectory are written next to the voltage traces. The polar plot includes the IB and DAC values that were set while the corresponding pulse was recorded. Three OPs are marked with black, red, and white crosses, corresponding to OPs that were chosen for low, intermediate, and high TPA values. Fig. and caption also used in Ref. [3].

After training was concluded, inference trajectories were run with all six trained agents. All agents find control parameters and OPs in the superconducting transition.

The jumping between control parameters after each environment step was also visible in this setup, and it is exemplarily visualized in Fig. 6.17 for run2 of Li1L. The agent prefers small DAC and high IB values for small TPA values and the other way around. This provides a strong hint that the agent switches between two OPs, where one is better for small pulses due to better noise conditions, and the other is better for large pulses due to a higher dynamic range. The OP corresponding to the energy of interest can be fixed for physics data taking.

While the jumping between OPs seemed to be a malicious feature in the beginning, it turned out that the agent had learned even more information than expected. It learned not only to find optimal control parameters, but it found a function that provides optimal control parameters for a given recoil energy or TPA value. This can practically be used to tune detectors for different objectives without the necessity of repeating the initial setup process.
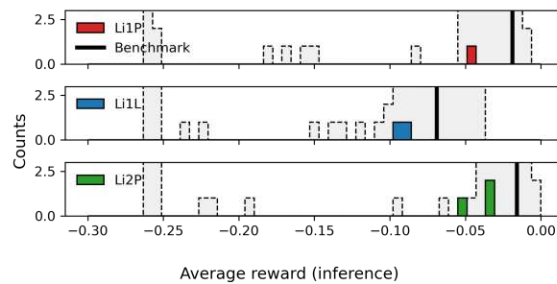

Figure 6.18: Histogram of the average reward obtained during inference trajectories with the trained agents on the real-world versions of Li1P (red, top), Li1L (blue, center), and Li2P (green, bottom) each. The rewards obtained in the simulation (grey, dotted histogram) and the human-optimized benchmark value (black line) are shown for comparison. The obtained rewards are worse than the benchmark value but correspond to our expectations from the simulation. For a discussion of the achievable optimality see also Fig. 6.19. Fig. and caption also used in Ref. [3].

We show the average reward obtained in the inference trajectories to quantify the optimality of the found control parameters in Fig. 6.18. For comparison, the rewards obtained in the virtual environment and the benchmark value from a human-optimized OP are also shown. The comparison is subject to the same uncertainties as discussed in Sec. 6.2.3. Furthermore, it is unclear if the noise conditions of the worked-on detectors have changed after the previous data-taking period, from which the benchmark value and the physics parameters were extracted. In the meantime, several warm-up tests on the CRESST setup were performed and reported in Ref. [31].

The optimality of the found OPs is worse than that of the benchmark value. However, this does not come as a surprise since the hyperparameters of the agents

were not tuned for optimality but for fast and reliable convergence. The results lie well within the expected distribution from the simulation. Therefore, tuning the parameters for optimality of the control parameters is expected to reach the human expert level. The technical details for this expectation are discussed below.
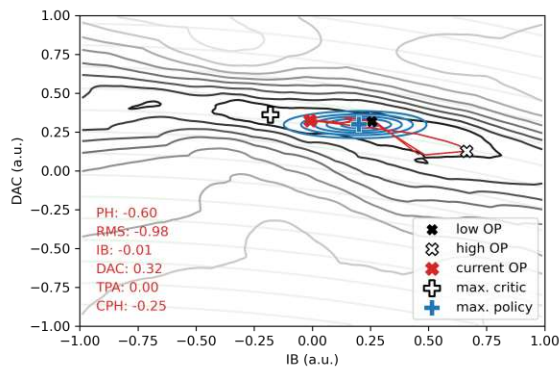

Figure 6.19: Visualization of the Gaussian policy probability distribution (blue) and the critic function (grey-black) over the two-dimensional action space, for a fixed "current" state (red text, lower left) and Li1L run 2. The maximum of the critic function is marked with a white plus. The current control parameters are marked with a red cross, that of OPs that were chosen by the agent for high/low TPA values with a white/black cross. These crosses correspond to the OPs marked with similar crosses in Fig. 6.17. The trajectory of actions that are chosen by the agent in inference is drawn with a red line, partially covered by the blue policy function. We can clearly see a mismatch between the actions preferred by the policy function and the maximum of the critic function. The reason for this mismatch is discussed in the text and in 6.2.5.2. The expected lines of equivalent heating caused by the DAC through the heating resistor and the IB through Joule heating are shown in the background (light, transparent green). As expected, the island of actions that are preferred by the critic stretches along the equivalent heating lines. The state values are normalized to the interval -1 to 1. The original value ranges are written in Tab. 6.8. Fig. and caption also used in Ref. [3].

The value function and policy function for a fixed (current) state are studied in Fig. 6.19. These functions are from a state in the trajectory of the Li1L, run 2. The current state is also marked in the trajectory in Fig. 6.10. A severe mismatch between the chosen actions and the maximum of the critic function is observed. This is generally not expected since after successful training, the policy function should choose actions previously identified as beneficial by the critic function. However, the policy function acts intrinsically stochastic. Therefore it is optimized to position its probability distribution in the parameter space such that the expected value is maximized, and not that the value at the mean value of the Gaussian function is

maximal. Simultaneously, the target entropy forces a minimum width of the Gaussian function, preventing it from converging to features on smaller scales in parameter space than its own width. However, the decreasing target entropy used in the virtual environment can enable the agent to converge on arbitrarily small scales. This issue is discussed at length using a toy model in Sec. 6.2.5.

In summary, a proof of principle was provided for the applicability of the RL method for optimizing control parameters of cryogenic detectors in the real world. With a time to convergence of 1.5 hours the time required for training was faster than the typical human expert. The optimality of the found control parameters was as expected from simulations and slightly worse than the benchmark value. This was due to the deliberate choice of the hyperparameters, which can be tuned in follow-up studies for optimality.

Before we reach the overall conclusion of this chapter, we continue with two toy model studies of two interesting phenomena that were observed and discussed in this section. The profound understanding of these effects, which we will develop in the following Sec. 6.2.5, can be useful for similar RL implementations to control experiments.

6.2.5 Toy-model studies

We study several effects we observed while working with RL on cryogenic detectors in this section. These effects seemed to be interesting enough to dedicate an individual discussion to them. The first is the jumping parameters in inference trajectories. We learned that these jumps are mainly caused by the changing optimality of OPs for different TPAs and discussed this in Sec. 6.2.4. However, we also expect an effect from the initialization of the NNs, and we study the magnitude of this effect in Sec. 6.2.5.1. The second effect we dedicate a discussion to in Sec. 6.2.5.2 is the mismatch between the policy function and value function observed in the live training. This is fully expected from the enforced entropy of the Gaussian policy function.

The studies in this section are performed with small toy experiments to isolate the effects from other physics of the environment.

6.2.5.1 Regularization of jumps in inference

The impact of the random initialization of the network parameters on the chosen actions was studied in Ref. [3] in a toy environment. This environment is a two-dimensional box, corresponding to the value range -1 to 1 in each dimension. In the

box, a target line is drawn that changes its position, and the agent has to try to get close to this line.

The available actions are to jump to a specific position in the box. The state space consists of the current position in the box and a third value, which cyclically changes between -1, 0, and 1, independently of the actions taken.

Also, the position of the target line changes cyclically between three states. The reward is the Euclidean distance from the target line, depicted in Fig. 6.20.
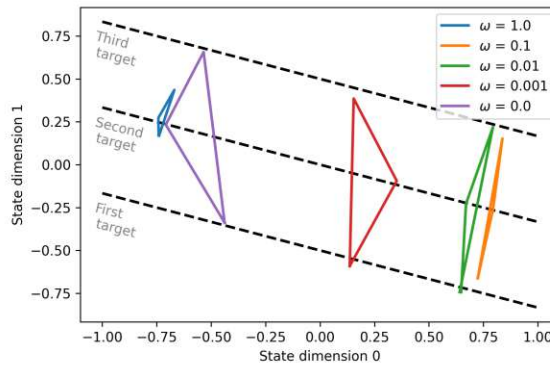

Figure 6.20: Movement of a SAC agent in a two-dimensional toy box environment. The goal of the agent is to jump close to the cyclically changing target lines (black dashed). The paths taken by an agent in inference trajectories are drawn with colored lines, for different magnitudes of the jump regularization parameter ω. Fig. and caption also used in Ref. [3].

The optimal policy for the agent is to jump directly to the next target line, which requires the agent only to understand the order in which the target lines appear. The point that the agent chooses along the target line to jump to can be arbitrary. There is therefore not a unique optimal policy function.

The situation is different when a regularization term $-\omega(A_1 - A_0)^2$ is introduced, where A_1 is the current and A_0 is the previous action. Now the optimal policy is to choose a trajectory of movement orthogonal to the target lines. The optimal policy is still not unique but severely restricted compared to the original scenario.

The results from training a SAC agent in this environment in the two scenarios are shown in Fig. 6.20. In the first scenario, the chosen point along the target lines depends only on the initialization of the network and the weight decay regularization of the ADAM optimizer. Competing effects from the regularizers are visible for the second scenario, and small values of ω. For larger values of ω the jump regularizer clearly overrules the weight regularizer but also starts competing with the main objective of jumping close to the target lines.

The choice of the ω parameter of the regularizer is therefore delicate, and an optimum likely depends on the exact environment and NNs used.

6.2.5.2 Entropy mismatch for action fine-tuning

We observed a mismatch between the actions taken by the actor and the actions preferred by the critic in the experiments reviewed in Sec. 6.2.4. The reason for this mismatch was attributed to the target entropy for the Gaussian actor, which kept it from converging towards small-scale features in the value distribution over the parameter space. This is generally a desired feature, as it keeps the policy from collapsing early in the training. Furthermore, the automatic tuning of the entropy ensures a certain width of the Gaussian policy function and prevents that we need to make an educated guess for the α parameter that would stay in a trade-off with the a-priori unknown returns. However, it can lead to a degradation of the final model performance in situations such as ours.

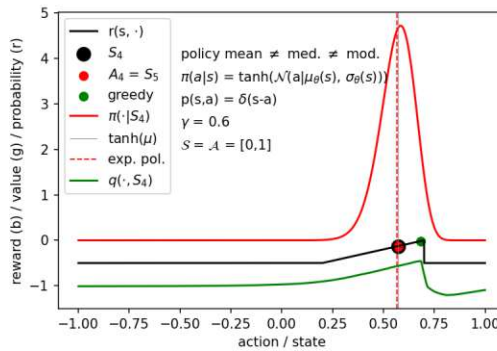Using a toy model, this effect is visualized in Fig. 6.21.

Figure 6.21: Toy environment of an agent that climbs a mountain. The reward function (black) has the shape of a side view of the mountain that ends in a cliff on one side. The critic function (green) learns the shape of the reward function sufficiently well. The policy function (red) learns to overlap with the mountain, instead of placing its expected value on top of the mountain (see text for details). To keep all actions in the interval between -1 and 1, the actions sampled from the Gaussian are in the SAC algorithm again input to a hyperbolic tangent function. This leads to a deviation between the mean, the median, and the mode of the resulting probability distribution. We show by comparing the expected/mean value of the policy (red dashed) with its median (grey) and its mode (peak of the red Gaussian) that this effect does not play a significant role in our experiment. Fig. and caption also used in Ref. [3].

The toy model corresponds to the side view of a mountain with a cliff on one side. The reward is the height of the agent on the mountain. The state is one-dimensional,

a parametrization of the distance between the agent and the mountain top, and the action is a position the agent can jump to on this axis. The optimal policy of the agent is to jump to the peak of the mountain and stay there.

The SAC agent learns to jump immediately towards the peak but does not aim directly for the peak, but to a point offset from it (Fig. 6.21, red dot). The maximum of the value function does agree with the position of the peak (Fig. 6.21, green dot).

The mismatch is owned to the fact that the SAC agent is trained to maximize the return for actions sampled from the actor function, not the return when the expected value of the actor function is taken.

The enforced width via the target entropy prevents the actor from moving closer to the peak because otherwise, the agent would step with a certain probability over the cliff, earning a very low reward.

This issue can generally be resolved by tuning the target entropy to the length scale of interest in the parameter space. In Sec. 6.2.3 we used a decreasing target entropy throughout the training to accomplish this.

6.3   Discussion

We have reviewed Refs. [4] and [3], in which studies were presented for automating the data cleaning process and the optimization of control parameters for cryogenic detectors.

In Sec. 6.1 the training of several deep learning models was reported for the task of discriminating artifacts and pulse shapes. For this, a large-scale data set from historic CRESST data was labeled. The best-performing model, a bidirectional LSTM, reached a balanced accuracy score of 0.932, a recall of 0.986, and a precision of 0.985 on a test set. About half of the wrongly predicted events turned out to be wrongly labeled. The recall and selectivity were practically unity for a data set of simulated pulses and pile-up events. With a runtime of several thousand predictions, the model performs considerably faster than for instance fit-based methods. The methods can be applied to new detectors without fine-tuning. Therefore, the amount of manual workforce necessary is considerably reduced compared to a standard cut-based data cleaning procedure.

In Sec. 6.2 the training of deep RL models was reported to find optimal control parameters for cryogenic detectors. The first experiments were performed in a virtual environment, where 315 trainings were performed on randomized versions of three CRESST detectors. The agents converged after less than 3 hours of equivalent

measurement time and reached with suitable hyperparameters the same performance as human-optimized benchmark values. Experiments on the live CRESST setup led to convergence within less than 1.5 hours of measurement time. The optimality of the found control parameters was inferior to the benchmark value but was well within the expectation from the simulation. In follow-up experiments, the hyperparameters of the SAC algorithm can be further tuned to reach the benchmark values, and the necessary measures for that were discussed.

Future experiments with RL to find control parameters could also include the possibility to tune the magnetic field. Also, both above summarized methods could be combined to reach a better stability and convergence speed by validating that in fact pulses are visible on the considered event traces.

A higher TP rate could be used in stable conditions to speed up the process. Also, a sine wave in the relevant frequency range could be used instead of the pulse, and the signal and noise frequencies could then be separated with a notch filter. With this, potentially much shorter record windows could be used. Finally, the change of the noise conditions that are observed when the phase transition is crossed could be used to understand where the transition is in the control parameter space.

By using the two newly introduced automation methods and all tools discussed in Chap. 3, the whole analysis could be put into a script and performed with almost no manual interventions required. The structure of such a script could follow the following steps:

1. Estimate a conservative, high threshold and trigger events with that threshold.

2. Clean all events with the LSTM model.

3. Build SEVs, NPS, OFs.

4. Estimate the threshold and resolution with a fit to the filtered noise triggers (see Sec. 3.1.1.5).

5. Trigger the recorded stream again with an optimized trigger threshold and an OF.

6. Clean again the events from all channels with the LSTM, or only the dominant channel, depending on the detector concept.

7. Perform additional automated data cuts based on statistics of the feature distribution, e.g. rates and stability cuts.

8. Calibrate the energy with the TPs to find the TPE values for each particle event.

9. Create the plot of the projections to the principal components as was shown in Sec. 6.1.3.

10. Plot the cleaned TPE spectrum.

11. Simulate a large number of signal candidate events and apply the same analysis chain to them to obtain trigger and cut efficiencies.

12. A human operator decides which peak in the TPE spectrum is the calibration peak with known energy.

13. A human operator controls that all the surviving structures in the data manifold visualized with the principal components belong to events that were meant to survive the quality cuts.

In a script like this only the last two steps require manual interventions, and these can be executed for many channels in rapid succession.

Steps that are not automated yet are to decide which pulse shape corresponds to target events in a scenario where multiple pulse shapes exist in the data. However, detectors can be designed to majorly feature one pulse shape, namely that from target recoils. This would be realized by avoiding large holding structures or carrier crystals, with which particles could interact. For a large-scale setup, a simple detector concept would therefore be beneficial.

Finally, the HLD analysis also needs to be streamlined in the future. As discussed in Sec. 3.2.3.2, there can be difficulties if the HLD sets from multiple detectors have different background rates but should be combined. Detectors should therefore ideally be manufactured as identically as possible to obtain similar background and artifact rates.

Conclusion

In this work we discussed several contributions to the quest of measuring light dark matter-nucleus interactions with superconducting thermometers.

In Chap. 1 the observational evidence for the existence of dark matter and the challenges for dark matter direct detection were summarized. The key ingredients for next-generation dark matter searches with superconducting thermometers were identified to be optimizing detector designs, scaling up the number of simultaneously operated detectors, and mitigating instrumental backgrounds.

The details of the technology of superconducting thermometers for dark matter direct detection were introduced in Chap. 2. We especially established a mathematical description of the detector response and used historical data to show the scaling properties of a standard detector design in a dedicated response simulation called "CryoEnv". We furthermore discussed the results from the first iteration of a dedicated workshop series for low energy excesses, the EXCESS workshop.

We explained the standard methods for raw data and high-level data analysis in Chap. 3. Furthermore, we discussed our recent release of "Cait", a software package for a Python-based raw data analysis workflow that can include modern machine learning methods and a method to reconstruct pile-up events based on deep learning. In Sec. 3.2.3 we show that the combined analysis of data sets taken with different detectors can lead to complications and that one has to account for individual backgrounds of the data sets in a likelihood framework to satisfactorily exploit the information of the combined data.

The introduced CryoEnv detector simulation was used in Chap. 4 to project achievable thresholds for proposed COSINUS detector designs with optimized components and operation conditions, under the constraint of a given target mass. The results are based on measured data from recent R&D runs with smaller target crystals.

In Chap. 5 we used the framework of the CryoEnv simulation to extract the physics parameters of several CRESST detectors, dubbed Li1P, Li1L, and Li2P, that were operated in the CRESST experiment. These detectors were grouped into two detector

modules Li1 and Li2, which both have a target made of lithium aluminate. This material is especially useful for dark matter searches since the isotope lithium-6 has an especially light nucleus and odd proton and neutron numbers. A data set measured with these detectors was used to set the currently strongest limits on spin-dependent sub-GeV/c^2-nucleus interactions. In this analysis, we used the Cait software package.

The final Chap. 6 treats the challenges of large-scale detector operation and analysis. Solutions to automate the required manual interventions in data cleaning and the initial optimization process of control parameters are proposed using deep and reinforcement learning. For this, a data set of events from historic CRESST runs was assembled, and all events were labeled as pulses or artifacts. It was then shown that deep learning models could learn to classify unseen data with similar accuracy as human analysts would achieve with quality cuts. Furthermore, the physics parameters extracted from Li1P, Li1L, and Li2P were used in Chap. 6 to build an reinforcement learning environment that simulated the optimization process of detector control parameters. The reinforcement learning environment is built as a wrapper around the CryoEnv simulation. Within this environment, Soft Actor-Critic agents were successfully trained to perform the optimization with similar quality, but faster than a human expert. The identical method was then tested live and in the real world on the actual detectors, achieving promising results.

In summary, we contributed to the key quests for measuring dark matter in the following ways: first, the CryoEnv detector response simulation is used to extract physics parameters from measurements and optimize detector designs. Second, the Cait software package lays the foundation for machine learning methods to become part of the standard analysis workflow. Using Cait and our method for automated data cleaning, the raw data analysis can be fully automated for standard detector designs. Together with our automated control parameter optimization method, the presented methods enable the operation of large-scale cryogenic detector setups by only a small number of analysts and operators. Finally, the summarized results from the EXCESS workshop are an essential contribution to the understanding and mitigation of low energy excesses.

Specifically for the COSINUS experiment, we have provided the first projections of achievable thresholds, including a full detector response model. Our projections provide actionable intelligence for planning the experiment's runs. Especially important is the recommendation that a first run can be performed with detectors optimized for thermal signals, while for the second run with larger targets, the realization of thinner films and an athermal readout scheme is necessary. Our method for separating pile-up

events can increase the efficiency of cuts significantly in the first run, where pulses are expected to be several seconds long.

For the CRESST experiment, this work summarized the analysis and the results from the first physics data taking in the main CRESST setup of detectors with lithium aluminate targets. The extraction of the physics parameters of the modules identified promising opportunities for improving the detector design. Most crucial among them is a lower transition temperature; component sizes of the transition-edge sensors and phonon collectors could be further optimized to achieve lower thresholds as well.

Several questions remain that would provide an interesting research program for follow-up studies. Assuming that raw data analysis and detector operation are sufficiently automated using the results of this work, the mass production of cryogenic detectors remains a challenge. Automation may require a very different set of methods. Furthermore, while we have simulated optimized detector designs in this work, experimental verification of these results would be very interesting and would directly lead to improved sensitivity in the search for dark matter.

Potential connections between the low energy excesses observed in cryogenic detectors and unexplained phenomena in superconducting qubits can be studied for future applications. The two major unexplained phenomena observed in superconducting qubits are two-level systems and excess quasiparticles in superconducting films. A 1/f-noise is also observed in superconducting qubits, where it could potentially be explained with a large ensemble of two-level systems [213]. The two-level systems most harmful for coherence and gate operations live in the oxide layer of the superconducting tunnel junctions. Interestingly, their frequency changes on a time scale of several days after the cool-down of the setup, and their frequency shifts also with stress in the silicon substrate on which the junctions are sputtered. The excess quasiparticle density in the superconducting film of which the qubit consists drops with time after the cooldown as well [141], and it was also observed to scale with stray infrared radiation. The latter observation confirms that energy depositions on the eV-scale in the qubit can cause an increased quasiparticle density. Since the low energy excess could also cause comparable energy depositions, a connection between these phenomena is possible. For all these phenomena different types of defects in crystal and film could explain all observations. However, to date, no sufficient model of the microscopic mechanisms exists.

Finally, the biggest open question is the question of existence of dark matter particles in the mass range of MeV/c²-GeV/c². However, this question can only

ultimately be answered once dark matter is experimentally characterized. Until then, every effort towards an initial measurement of such particles is valuable.

In conclusion, significant progress in the quest of testing sub-GeV/c^2 dark matter was made in the past years, but the journey remains challenging. The operation of large-scale setups will enable fine-grained studies of instrumental backgrounds in early runs. With the knowledge gained from there, the following runs will allow us to test dark matter masses on MeV/c^2-scale and down to the neutrino fog. While these are certainly ambitious outlooks, they may realistically be realized within the coming years.

Appendix A

Table of abbreviations

Abbr.	Definition	Abbr.	Definition
ΛCDM	Lambda cold dark matter	ADC	analog-digital converter
AUC	area under the ROC curve	BCS	Bardeen–Cooper–Schrieffer
CCD	charge coupled device	CDF	cumulative distribution function
CDMS	cold dark matter search	CEvNS	coherent elastic neutrino nucleus scattering
CMB	cosmic microwave background	CNN	convolutional neural network
COSINUS	cryogenic observatory for signals seen in next-generation underground searches	CP	control pulse
CP symmetry	charge-parity symmetry	CPE	conversion pulse energy
CRESST	cryogenic rare event search with superconducting thermometers	DAC	digital-analog converter
DAQ	data acquisition	DD	direct detection
DM	dark matter	EM	electromagnetic
ESA	European Space Agency	ETF	electro thermal feedback
FT	Fourier transform	HLD	high level data
ITFN	internal thermal fluctuation noise	IoT	internet of things
KID	kinetic inductance detector	LA	longitudenal
LEE	low energy excess	LH	likelihood
LHC	large hadron collider	LNGS	laboratori nazionali del Gran Sasso
LSP	lightest supersymmetric particle	LSTM	long short term memory
LY	light yield	MDP	Markov decision process
MMC	metallic magnetic calorimeter	MPP	Max-Planck-Institut für Physik
MSE	mean squared error	MWE	meters water equivalent
NFW	Navarro-Frenk-White	NN	neural network
NPS	noise power sprectrum	NTD	neutron transmutation doped
NTL	Neganov-Trofimov-Luke	ODE	ordinary differential equation
OF	optimum filter	OP	operation point
OU	Ornstein-Uhlenbeck	PCA	principal component analysis
PD	particle discrimination	PH	pulse height
PID	proportional–integral–derivative	PMT	photomultiplier tube
PSD	pulse shape discrimination	QCD	quantum chromodynamics

Abbr.	Definition	Abbr.	Definition
QF	quenching factor	RD	raw data
RL	reinforcement learning	RMS	root mean squared
ROC	receiver operating characteristic	ROI	region of interest
ReLU	rectified linear unit	SAC	Soft Actor-Critic
SEV	standard event	SM	standard model
SNR	signal-to-noise ratio	SNSPD	superconducting nanowire single-photon detector
SOS	silicon-on-sapphire	SQUID	superconducting quantum interference device
SUSY	supersymmetry	TA	transversal
TES	transition edge sensor	TP	test pulse
TPA	test pulse amplitude	TPC	time projection chamber
TPE	test pulse equivalent	TSCN	time series convolutional network
TST	time series transformer	WIMP	weakly interacting massive particle

References

[1] D. Baxter, I. M. Bloch, D. Durnford, *et al.*, "White paper: Complications to comparing results from low-threshold direct dark matter searches." To be published in 2024.

[2] F. Wagner, "Optimal operation of cryogenic phonon detectors through deep reinforcement learning," Master's thesis, 2023. `https://doi.org/10.34726/hss.2023.110322`.

[3] G. Angloher, S. Banik, D. Bartolot, *et al.*, "Optimal operation of cryogenic calorimeters through deep reinforcement learning." To be published in 2023.

[4] G. Angloher, S. Banik, D. Bartolot, *et al.*, "Towards an automated data cleaning with deep learning in CRESST," *The European Physical Journal Plus* **138** no. 1, (Jan., 2023) 100. `https://doi.org/10.1140/epjp/s13360-023-03674-2`.

[5] F. Wagner, D. Bartolot, D. Rizvanovic, *et al.*, "Cait: Analysis Toolkit for Cryogenic Particle Detectors in Python," *Computing and Software for Big Science* **6** no. 1, (Dec., 2022) 19. `https://doi.org/10.1007/s41781-022-00092-4`.

[6] G. Angloher, S. Banik, G. Benato, *et al.*, "Testing spin-dependent dark matter interactions with lithium aluminate targets in cresst-iii," *Phys. Rev. D* **106** (Nov, 2022) 092008. `https://link.aps.org/doi/10.1103/PhysRevD.106.092008`.

[7] P. Adari, A. Aguilar-Arevalo, D. Amidei, *et al.*, "EXCESS workshop: Descriptions of rising low-energy spectra," *SciPost Phys. Proc.* (2022) 001. `https://scipost.org/10.21468/SciPostPhysProc.9.001`.

[8] F. Wagner, "Nonlinear pile-up separation with lstm neural networks for cryogenic particle detectors," 2021. arXiv:2112.06792.

[9] F. Wagner, "A parametric fit model for gaussian noise maxima in the presence of pollutions." CRESST-internal note., 2020.

[10] F. Iocco, M. Pato, and G. Bertone, "Evidence for dark matter in the inner Milky Way," *Nature Physics* **11** no. 3, (Mar., 2015) 245–248. `https://doi.org/10.1038/nphys3237`.

[11] D. Clowe, M. Bradač, A. H. Gonzalez, *et al.*, "A Direct Empirical Proof of the Existence of Dark Matter," *The Astrophysical Journal* **648** no. 2, (Sept., 2006) L109–L113. `https://iopscience.iop.org/article/10.1086/508162`.

[12] Planck Collaboration, Aghanim, N., Akrami, Y., *et al.*, "Planck 2018 results - VI. Cosmological parameters," *A&A* **641** (2020) A6. `https://doi.org/10.1051/0004-6361/201833910`.

[13] G. Jungman, M. Kamionkowski, and K. Griest, "Supersymmetric dark matter," *Physics Reports* **267** no. 5-6, (Mar, 1996) 195–373. `https://doi.org/10.1016%2F0370-1573%2895%2900058-5`.

[14] K. Garrett and G. Duda, "Dark Matter: A Primer," *Advances in Astronomy* **2011** (Dec., 2010) 968283. `https://doi.org/10.1155/2011/968283`. Publisher: Hindawi Publishing Corporation.

[15] N. Bernal, M. Heikinheimo, T. Tenkanen, *et al.*, "The dawn of fimp dark matter: A review of models and constraints," *International Journal of Modern Physics A* **32** no. 27, (2017) 1730023, https://doi.org/10.1142/S0217751X1730023X. https://doi.org/10.1142/S0217751X1730023X.

[16] A. M. Green and B. J. Kavanagh, "Primordial black holes as a dark matter candidate," *Journal of Physics G: Nuclear and Particle Physics* **48** no. 4, (Feb, 2021) 043001. https://dx.doi.org/10.1088/1361-6471/abc534.

[17] M. Milgrom, "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis.," *The Astrophysical Journal* **270** (July, 1983) 365–370.

[18] A. Boveia and C. Doglioni, "Dark matter searches at colliders," *Annual Review of Nuclear and Particle Science* **68** no. 1, (2018) 429–459, https://doi.org/10.1146/annurev-nucl-101917-021008. https://doi.org/10.1146/annurev-nucl-101917-021008.

[19] J. M. Gaskins, "A review of indirect searches for particle dark matter," *Contemporary Physics* **57** no. 4, (2016) 496–525, https://doi.org/10.1080/00107514.2016.1175160. https://doi.org/10.1080/00107514.2016.1175160.

[20] E. Aprile, K. Abe, F. Agostini, *et al.*, "Search for new physics in electronic recoil data from xenonnt," *Phys. Rev. Lett.* **129** (Oct, 2022) 161805. https://link.aps.org/doi/10.1103/PhysRevLett.129.161805.

[21] E. Aprile, J. Aalbers, F. Agostini, *et al.*, "Dark matter search results from a one ton-year exposure of xenon1t," *Phys. Rev. Lett.* **121** (Sep, 2018) 111302. https://link.aps.org/doi/10.1103/PhysRevLett.121.111302.

[22] B. W. Lee and S. Weinberg, "Cosmological lower bound on heavy-neutrino masses," *Phys. Rev. Lett.* **39** (Jul, 1977) 165–168. https://link.aps.org/doi/10.1103/PhysRevLett.39.165.

[23] C. Bœhm, P. Fayet, and J. Silk, "Light and heavy dark matter particles," *Phys. Rev. D* **69** (May, 2004) 101302. https://link.aps.org/doi/10.1103/PhysRevD.69.101302.

[24] D. Hooper and K. M. Zurek, "Natural supersymmetric model with mev dark matter," *Phys. Rev. D* **77** (Apr, 2008) 087302. https://link.aps.org/doi/10.1103/PhysRevD.77.087302.

[25] J. L. Feng and J. Kumar, "Dark-matter particles without weak-scale masses or weak interactions," *Phys. Rev. Lett.* **101** (Dec, 2008) 231301. https://link.aps.org/doi/10.1103/PhysRevLett.101.231301.

[26] K. M. Zurek, "Multicomponent dark matter," *Phys. Rev. D* **79** (Jun, 2009) 115002. https://link.aps.org/doi/10.1103/PhysRevD.79.115002.

[27] D. E. Kaplan, M. A. Luty, and K. M. Zurek, "Asymmetric dark matter," *Phys. Rev. D* **79** (Jun, 2009) 115016. https://link.aps.org/doi/10.1103/PhysRevD.79.115016.

[28] A. Falkowski, J. T. Ruderman, and T. Volansky, "Asymmetric dark matter from leptogenesis," *Journal of High Energy Physics* **2011** no. 5, (May, 2011) 106. https://doi.org/10.1007/JHEP05(2011)106.

[29] L. Einfalt, S. Kulkarni, M. Procura, and F. Reindl, "Dark biportals at direct detection," *Phys. Rev. D* **106** (Aug, 2022) 035031. https://link.aps.org/doi/10.1103/PhysRevD.106.035031.

[30] A. H. Abdelhameed, G. Angloher, P. Bauer, *et al.*, "First results from the cresst-iii low-mass dark matter program," *Phys. Rev. D* **100** (Nov, 2019) 102002. https://link.aps.org/doi/10.1103/PhysRevD.100.102002.

[31] G. Angloher, S. Banik, G. Benato, *et al.*, "Latest observations on the low energy excess in CRESST-III," *SciPost Phys. Proc.* (2023) 013. https://scipost.org/10.21468/SciPostPhysProc.12.013.

[32] F. J. Kerr and D. Lynden-Bell, "Review of galactic constants," *Monthly Notices of the Royal Astronomical Society* **221** no. 4, (Aug., 1986) 1023–1038. https://doi.org/10.1093/mnras/221.4.1023. _eprint: https://academic.oup.com/mnras/article-pdf/221/4/1023/9404848/mnras221-1023.pdf.

[33] Salucci, P., Nesti, F., Gentile, G., and Frigerio Martins, C., "The dark matter density at the Sun\'s location," *A&A* **523** (2010) A83. https://doi.org/10.1051/0004-6361/201014385.

[34] S. K. Lee, M. Lisanti, S. Mishra-Sharma, and B. R. Safdi, "Modulation effects in dark matter-electron scattering experiments," *Phys. Rev. D* **92** (Oct, 2015) 083517. https://link.aps.org/doi/10.1103/PhysRevD.92.083517.

[35] K. Freese, M. Lisanti, and C. Savage, "Colloquium: Annual modulation of dark matter," *Rev. Mod. Phys.* **85** (Nov, 2013) 1561–1581. https://link.aps.org/doi/10.1103/RevModPhys.85.1561.

[36] F. Donato, N. Fornengo, and S. Scopel, "Effects of galactic dark halo rotation on WIMP direct detection," *Astroparticle Physics* **9** no. 3, (1998) 247–260. https://www.sciencedirect.com/science/article/pii/S0927650598000255.

[37] R. Bernabei, P. Belli, F. Cappella, *et al.*, "First results from DAMA/LIBRA and the combined results with DAMA/NaI," *The European Physical Journal C* **56** no. 3, (Aug., 2008) 333–355. https://doi.org/10.1140/epjc/s10052-008-0662-y.

[38] R. Bernabei, P. Belli, F. Cappella, *et al.*, "Dark matter: Dama/libra and its perspectives," 2022. arXiv:2209.00882.

[39] F. Kahlhöfer, "Annual modulation of dark matter signals: Experimental results and new ideas." Conference talk at IDM2022, accessed on 12.10.2023, 2022.

[40] G. Adhikari, N. Carlin, J. Choi, *et al.*, "An induced annual modulation signature in COSINE-100 data by DAMA/LIBRA's analysis method," *Scientific Reports* **13** no. 1, (Mar., 2023) 4676. https://doi.org/10.1038/s41598-023-31688-4.

[41] J. H. Davis, "Fitting the annual modulation in dama with neutrons from muons and neutrinos," *Phys. Rev. Lett.* **113** (Aug, 2014) 081302. https://link.aps.org/doi/10.1103/PhysRevLett.113.081302.

[42] J. Amaré, S. Cebrián, D. Cintas, *et al.*, "Annual modulation results from three-year exposure of anais-112," *Phys. Rev. D* **103** (May, 2021) 102005. https://link.aps.org/doi/10.1103/PhysRevD.103.102005.

[43] G. Adhikari, E. B. de Souza, N. Carlin, *et al.*, "Strong constraints from COSINE-100 on the DAMA dark matter results using the same sodium iodide target," *Sci. Adv.* **7** no. 46, (Nov., 2020) eabk2699. https://www.science.org/doi/10.1126/sciadv.abk2699. Publisher: American Association for the Advancement of Science.

[44] E. Barberio, T. Baroncelli, L. J. Bignell, *et al.*, "Simulation and background characterisation of the SABRE South experiment," *The European Physical Journal C* **83** no. 9, (Sept., 2023) 878. https://doi.org/10.1140/epjc/s10052-023-11817-z.

[45] G. Angloher, P. Carniti, I. Dafinei, *et al.*, "COSINUS: Cryogenic Calorimeters for the Direct Dark Matter Search with NaI Crystals," *Journal of Low Temperature Physics* **200** no. 5, (Sept., 2020) 428–436. https://doi.org/10.1007/s10909-020-02464-9.

[46] D. Cintas, P. An, C. Awe, *et al.*, "Quenching factor consistency across several nai(tl) crystals," *Journal of Physics: Conference Series* **2156** no. 1, (Dec, 2021) 012065. https://dx.doi.org/10.1088/1742-6596/2156/1/012065.

[47] G. Angloher, M. R. Bharadwaj, I. Dafinei, *et al.*, "First measurements of remoTES cryogenic calorimeters: Easy-to-fabricate particle detectors for a wide choice of target materials," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1045** (2023) 167532. `https://www.sciencedirect.com/science/article/pii/S0168900222008245`.

[48] F. Kahlhoefer, F. Reindl, K. Schäffner, *et al.*, "Model-independent comparison of annual modulation and total rate with direct detection experiments," *Journal of Cosmology and Astroparticle Physics* **2018** no. 05, (May, 2018) 074. `https://dx.doi.org/10.1088/1475-7516/2018/05/074`.

[49] C. Collaboration, G. Angloher, M. R. Bharadwaj, *et al.*, "Particle discrimination in a nai crystal using the cosinus remote tes design," 2023. arXiv:2307.11066.

[50] T. C. Collaboration, G. Angloher, M. R. Bharadwaj, *et al.*, "Deep-underground dark matter search with a cosinus detector prototype," 2023. arXiv:2307.11139.

[51] A. Macías and A. Camacho, "On the incompatibility between quantum theory and general relativity," *Physics Letters B* **663** no. 1, (2008) 99–102. `https://www.sciencedirect.com/science/article/pii/S0370269308003936`.

[52] M. K. Gaillard, P. D. Grannis, and F. J. Sciulli, "The standard model of particle physics," *Rev. Mod. Phys.* **71** (Mar, 1999) S96–S111. `https://link.aps.org/doi/10.1103/RevModPhys.71.S96`.

[53] G. Bertone and D. Hooper, "History of dark matter," *Rev. Mod. Phys.* **90** (Oct, 2018) 045002. `https://link.aps.org/doi/10.1103/RevModPhys.90.045002`.

[54] A. Friedman, "Über die Krümmung des Raumes," *Zeitschrift für Physik* **10** no. 1, (Dec., 1922) 377–386. `https://doi.org/10.1007/BF01332580`.

[55] A. Einstein, "Die grundlage der allgemeinen relativitätstheorie," *Annalen der Physik* **354** no. 7, (1916) 769–822, `https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702`. `https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702`.

[56] R. H. Brandenberger, "Formation of structure in the universe," 1995. arXiv:astro-ph/9508159.

[57] A. A. Penzias and R. W. Wilson, "A Measurement of Excess Antenna Temperature at 4080 Mc/s.," *Astrophysical Journal* **142** (July, 1965) 419–421.

[58] E. S. Agency, "Planck's view of the cosmic microwave background," 2018. `https://www.esa.int/ESA_Multimedia/Images/2018/07/Planck_s_view_of_the_cosmic_microwave_background`. [Online; accessed 21-April-2020].

[59] K. Jedamzik and M. Pospelov, "Big bang nucleosynthesis and particle dark matter," *New Journal of Physics* **11** no. 10, (Oct, 2009) 105028. `https://dx.doi.org/10.1088/1367-2630/11/10/105028`.

[60] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln," *Helvetica Physica Acta* **6** (Jan., 1933) 110–127.

[61] K. C. Freeman, "On the Disks of Spiral and S0 Galaxies," *The Astrophysical Journal* **160** (June, 1970) 811.

[62] Wikipedia, The Free Encyclopedia, "Galaxy rotation curve." `https://en.wikipedia.org/wiki/Galaxy_rotation_curve`. [Online; accessed 19-April-2020].

[63] E. S. Agency, "The bullet cluster." `https://www.esa.int/ESA_Multimedia/Images/2007/07/The_Bullet_Cluster2`. [Online; accessed 12-Sept-2023].

[64] J. F. Navarro, C. S. Frenk, and S. D. M. White, "The Structure of Cold Dark Matter Halos," *The Astrophysical Journal* **462** (May, 1996) 563, `arXiv:astro-ph/9508025 [astro-ph]`.

[65] I. Arnquist, N. Avalos, D. Baxter, *et al.*, "Search for daily modulation of mev dark matter signals with damic-m," 2023. arXiv:2307.07251.

[66] F. Reindl, *Exploring Light Dark Matter With CRESST-II Low-Threshold Detectors*. PhD thesis, Technical University Munich, 2016. `https://mediatum.ub.tum.de/?id=1294132`.

[67] M. L. Perl, E. R. Lee, and D. Loomba, "Searches for fractionally charged particles," *Annual Review of Nuclear and Particle Science* **59** no. 1, (2009) 47–65, `https://doi.org/10.1146/annurev-nucl-121908-122035`. `https://doi.org/10.1146/annurev-nucl-121908-122035`.

[68] F. Wilczek, "Problem of strong p and t invariance in the presence of instantons," *Phys. Rev. Lett.* **40** (Jan, 1978) 279–282. `https://link.aps.org/doi/10.1103/PhysRevLett.40.279`.

[69] R. D. Peccei, "The Strong CP Problem and Axions," in *Axions: Theory, Cosmology, and Experimental Searches*, M. Kuster, G. Raffelt, and B. Beltrán, eds., pp. 3–17. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. `https://doi.org/10.1007/978-3-540-73518-2_1`.

[70] L. Hui, "Wave dark matter," *Annual Review of Astronomy and Astrophysics* **59** no. 1, (2021) 247–289, `https://doi.org/10.1146/annurev-astro-120920-010024`. `https://doi.org/10.1146/annurev-astro-120920-010024`.

[71] A. Boyarsky, M. Drewes, T. Lasserre, *et al.*, "Sterile neutrino Dark Matter," *Progress in Particle and Nuclear Physics* **104** (2019) 1–45. `https://www.sciencedirect.com/science/article/pii/S0146641018300711`.

[72] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, "Mechanism for thermal relic dark matter of strongly interacting massive particles," *Phys. Rev. Lett.* **113** (Oct, 2014) 171301. `https://link.aps.org/doi/10.1103/PhysRevLett.113.171301`.

[73] Y. Hochberg, E. Kuflik, H. Murayama, *et al.*, "Model for thermal relic dark matter of strongly interacting massive particles," *Phys. Rev. Lett.* **115** (Jul, 2015) 021301. `https://link.aps.org/doi/10.1103/PhysRevLett.115.021301`.

[74] M. W. Goodman and E. Witten, "Detectability of certain dark-matter candidates," *Phys. Rev. D* **31** (Jun, 1985) 3059–3063. `https://link.aps.org/doi/10.1103/PhysRevD.31.3059`.

[75] R. H. Helm, "Inelastic and elastic scattering of 187-mev electrons from selected even-even nuclei," *Phys. Rev.* **104** (Dec, 1956) 1466–1475. `https://link.aps.org/doi/10.1103/PhysRev.104.1466`.

[76] M. Ibe, W. Nakano, Y. Shoji, and K. Suzuki, "Migdal effect in dark matter direct detection experiments," *Journal of High Energy Physics* **2018** no. 3, (Mar., 2018) 194. `https://doi.org/10.1007/JHEP03(2018)194`.

[77] N. F. Bell, J. B. Dent, J. L. Newstead, *et al.*, "Migdal effect and photon bremsstrahlung in effective field theories of dark matter direct detection and coherent elastic neutrino-nucleus scattering," *Phys. Rev. D* **101** (Jan, 2020) 015012. `https://link.aps.org/doi/10.1103/PhysRevD.101.015012`.

[78] R. Essig, M. Fernandez-Serra, J. Mardon, *et al.*, "Direct detection of sub-gev dark matter with semiconductor targets," 2016. arXiv:1509.01598.

[79] A. Aguilar-Arevalo, D. Amidei, D. Baxter, *et al.*, "Results on Low-Mass Weakly Interacting Massive Particles from an 11 kg-day Target Exposure of DAMIC at SNOLAB," *Phys. Rev. Lett.* **125** no. 24, (Dec., 2020) 241803.

[80] B. A. Cervantes-Vergara, S. Perez, J. C. D'Olivo, *et al.*, "Skipper-CCDs: Current applications and future," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1046** (2023) 167681. `https://www.sciencedirect.com/science/article/pii/S0168900222009731`.

[81] M. F. Albakry, I. Alkhatib, D. W. P. Amaral, *et al.*, "Investigating the sources of low-energy events in a supercdms-hvev detector," *Phys. Rev. D* **105** (Jun, 2022) 112006. `https://link.aps.org/doi/10.1103/PhysRevD.105.112006`.

[82] H. Lattaud, E. Armengaud, Q. Arnaud, *et al.*, "Sub-MeV dark matter searches with EDELWEISS: results and prospects," in *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2021)*. Sissa Medialab, Jan, 2022. `https://doi.org/10.22323%2F1.398.0153`.

[83] P. K. Day, H. G. LeDuc, B. A. Mazin, *et al.*, "A broadband superconducting detector suitable for use in large arrays," *Nature* **425** no. 6960, (Oct., 2003) 817–821.

[84] D. W. Amaral, T. Aralis, T. Aramaki, *et al.*, "Constraints on low-mass, relic dark matter candidates from a surface-operated supercdms single-charge sensitive detector," *Phys. Rev. D* **102** (Nov, 2020) 091101. `https://link.aps.org/doi/10.1103/PhysRevD.102.091101`.

[85] S. A. Hertel, A. Biekert, J. Lin, *et al.*, "Direct detection of sub-gev dark matter using a superfluid ^{4}He target," *Phys. Rev. D* **100** (Nov, 2019) 092007. `https://link.aps.org/doi/10.1103/PhysRevD.100.092007`.

[86] E. Aprile, K. Abe, F. Agostini, *et al.*, "First dark matter search with nuclear recoils from the xenonnt experiment," *Phys. Rev. Lett.* **131** (Jul, 2023) 041003. `https://link.aps.org/doi/10.1103/PhysRevLett.131.041003`.

[87] G. Angloher, S. Banik, G. Benato, *et al.*, "Results on sub-gev dark matter from a 10 ev threshold cresst-iii silicon detector," *Phys. Rev. D* **107** (Jun, 2023) 122003. `https://link.aps.org/doi/10.1103/PhysRevD.107.122003`.

[88] I. Alkhatib, D. W. P. Amaral, T. Aralis, *et al.*, "Light dark matter search with a high-resolution athermal phonon detector operated above ground," *Phys. Rev. Lett.* **127** (Aug, 2021) 061801. `https://link.aps.org/doi/10.1103/PhysRevLett.127.061801`.

[89] SuperCDMS Collaboration, R. Agnese, T. Aralis, *et al.*, "Search for low-mass dark matter with CDMSlite using a profile likelihood fit," *Phys. Rev. D* **99** no. 6, (Mar., 2019) 062001. `https://link.aps.org/doi/10.1103/PhysRevD.99.062001`.

[90] R. Agnese, A. Anderson, M. Asai, *et al.*, "Search for Low-Mass Weakly Interacting Massive Particles with SuperCDMS," *Phys. Rev. Lett.* **112** no. 24, (June, 2014) 241302. `http://link.aps.org/doi/10.1103/PhysRevLett.112.241302`.

[91] E. Aprile, J. Aalbers, F. Agostini, *et al.*, "Light Dark Matter Search with Ionization Signals in XENON1t," *arXiv:1907.11485 [astro-ph, physics:hep-ex]* (July, 2019) . `http://arxiv.org/abs/1907.11485`. arXiv: 1907.11485.

[92] P. Agnes, I. F. M. Albuquerque, T. Alexander, *et al.*, "Low-mass dark matter search with the darkside-50 experiment," *Phys. Rev. Lett.* **121** (Aug, 2018) 081307. `https://link.aps.org/doi/10.1103/PhysRevLett.121.081307`.

[93] J. Aalbers, D. S. Akerib, C. W. Akerlof, *et al.*, "First dark matter search results from the lux-zeplin (lz) experiment," 2022. `https://arxiv.org/abs/2207.03764`. arXiv:2207.03764.

[94] L. Hehn, E. Armengaud, Q. Arnaud, *et al.*, "Improved EDELWEISS-III sensitivity for low-mass WIMPs using a profile likelihood approach," *Eur. Phys. J. C* **76** no. 10, (Oct., 2016) 548. `https://doi.org/10.1140/epjc/s10052-016-4388-y`.

[95] C. Savage, G. Gelmini, P. Gondolo, and K. Freese, "Compatibility of dama/libra dark matter detection with other searches," *Journal of Cosmology and Astroparticle Physics* **2009** no. 04, (Apr, 2009) 010. `https://dx.doi.org/10.1088/1475-7516/2009/04/010`.

[96] A. Gütlein, G. Angloher, A. Bento, *et al.*, "Impact of Coherent Neutrino Nucleus Scattering on Direct Dark Matter Searches based on CaWO$_4$ Crystals," *Astropart. Phys.* **69** (Sept., 2015) 44–49. `http://arxiv.org/abs/1408.2357`.

[97] J. Billard, M. Boulay, S. Cebrián, *et al.*, "Direct detection of dark matter—appec committee report*," *Reports on Progress in Physics* **85** no. 5, (Apr, 2022) 056201. `https://dx.doi.org/10.1088/1361-6633/ac5754`.

[98] R. Anthony-Petersen, A. Biekert, C. L. Chang, *et al.*, "Applying superfluid helium to light dark matter searches: Demonstration of the herald detector concept," 2023. arXiv:2307.11877.

[99] B. von Krosigk, K. Eitel, C. Enss, *et al.*, "Delight: a direct search experiment for light dark matter with superfluid helium," 2022. arXiv:2209.10950.

[100] L. Barak, I. M. Bloch, M. Cababie, *et al.*, "Sensei: Direct-detection results on sub-gev dark matter from a new skipper ccd," *Phys. Rev. Lett.* **125** (Oct, 2020) 171802. `https://link.aps.org/doi/10.1103/PhysRevLett.125.171802`.

[101] I. Arnquist, N. Avalos, P. Bailly, *et al.*, "The damic-m experiment: Status and first results," 2022. arXiv:2210.12070.

[102] A. Aguilar-Arevalo, F. A. Bessia, N. Avalos, *et al.*, "The oscura experiment," 2022. arXiv:2202.10518.

[103] F. Pröbst, M. Frank, S. Cooper, *et al.*, "Model for cryogenic particle detectors with superconducting phase transition thermometers," *Journal of Low Temperature Physics* **100** no. 1, (July, 1995) 69–104. `https://doi.org/10.1007/BF00753837`.

[104] K. D. Irwin, S. W. Nam, B. Cabrera, *et al.*, "A quasiparticle-trap-assisted transition-edge sensor for phonon-mediated particle detection," *Review of Scientific Instruments* **66** no. 11, (11, 1995) . `https://www.osti.gov/biblio/124855`.

[105] K. Irwin and G. Hilton, "Transition-Edge Sensors," in *Cryogenic Particle Detection*, C. Enss, ed., pp. 63–150. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. `https://doi.org/10.1007/10933596_3`.

[106] M. Tinkham, *Introduction to Superconductivity*. Dover Publications, 2 ed., June, 2004. `http://www.worldcat.org/isbn/0486435032`.

[107] M. Galeazzi and D. McCammon, "Microcalorimeter and bolometer model," *Journal of Applied Physics* **93** no. 8, (Mar., 2003) 4856–4869. `https://doi.org/10.1063/1.1559000`. _eprint: https://pubs.aip.org/aip/jap/article-pdf/93/8/4856/10626979/4856_1_online.pdf.

[108] M. Carrettoni and O. Cremonesi, "Generation of noise time series with arbitrary power spectrum," *Computer Physics Communications* **181** no. 12, (2010) 1982–1985. `https://www.sciencedirect.com/science/article/pii/S0010465510003486`.

[109] D. T. Gillespie, "The mathematics of Brownian motion and Johnson noise," *American Journal of Physics* **64** no. 3, (Mar., 1996) 225–240. `https://doi.org/10.1119/1.18210`. _eprint: https://pubs.aip.org/aapt/ajp/article-pdf/64/3/225/11851228/225_1_online.pdf.

[110] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Holt-Saunders, 1976.

[111] B. B. Triplett, N. E. Phillips, T. L. Thorp, *et al.*, "Critical field for superconductivity and low-temperature normal-state heat capacity of tungsten," *Journal of Low Temperature Physics* **12** no. 5, (Sept., 1973) 499–518. `https://doi.org/10.1007/BF00654953`.

[112] C. Kittel, *Introduction to Solid State Physics*. Wiley, 8 ed., 2004. `http://www.amazon.com/Introduction-Solid-Physics-Charles-Kittel/dp/047141526X/ref=dp_ob_title_bk`.

229

[113] M. Sisti, O. Meier, M. Bühler, *et al.*, "Massive cryogenic particle detectors with low energy threshold," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **466** no. 3, (2001) 499–508. https://www.sciencedirect.com/science/article/pii/S0168900201008014.

[114] M. Pyle, E. Figueroa-Feliciano, and B. Sadoulet, "Optimized designs for very low temperature massive calorimeters," 2015. arXiv:1503.01200.

[115] A. Jain, S. P. Ong, G. Hautier, *et al.*, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials* **1** no. 1, (July, 2013) 011002. https://doi.org/10.1063/1.4812323. _eprint: https://pubs.aip.org/aip/apm/article-pdf/doi/10.1063/1.4812323/13163869/011002_1_online.pdf.

[116] S. Kim, J. Chen, T. Cheng, *et al.*, "PubChem 2023 update," *Nucleic Acids Research* **51** no. D1, (Oct., 2022) D1373–D1380. https://doi.org/10.1093/nar/gkac956. _eprint: https://academic.oup.com/nar/article-pdf/51/D1/D1373/48441598/gkac956.pdf.

[117] E. Pantic, *Performance of Cryogenic Light detectors in the CRESST-II Dark Matter Search*. PhD thesis, Technische Universität München, 2008.

[118] T. F. Harrelson, I. Hajar, and S. M. Griffin, "Theoretical investigation of decoherence channels in athermal phonon sensors," 2021. arXiv:2109.10988.

[119] V. I. Khvesyuk, B. Liu, and A. A. Barinov, "Development of acoustic and diffuse mismatch models for predicting the kapitza resistance," *Journal of Physics: Conference Series* **1382** no. 1, (Nov, 2019) 012155. https://dx.doi.org/10.1088/1742-6596/1382/1/012155.

[120] R. Strauss, J. Rothe, G. Angloher, *et al.*, "Gram-scale cryogenic calorimeters for rare-event searches," *Phys. Rev. D* **96** (Jul, 2017) 022009. https://link.aps.org/doi/10.1103/PhysRevD.96.022009.

[121] J. F. Cochran and D. E. Mapother, "Superconducting transition in aluminum," *Phys. Rev.* **111** (Jul, 1958) 132–142. https://link.aps.org/doi/10.1103/PhysRev.111.132.

[122] N. E. Booth, "Quasiparticle trapping and the quasiparticle multiplier," *Applied Physics Letters* **50** no. 5, (Feb., 1987) 293–295. https://doi.org/10.1063/1.98229. _eprint: https://pubs.aip.org/aip/apl/article-pdf/50/5/293/7762797/293_1_online.pdf.

[123] M. Loidl, S. Cooper, O. Meier, *et al.*, "Quasiparticle diffusion over several mm in cryogenic detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **465** no. 2, (2001) 440–446. https://www.sciencedirect.com/science/article/pii/S0168900201006210.

[124] M. Galeazzi, "Fundamental noise processes in tes devices," *IEEE Transactions on Applied Superconductivity* **21** no. 3, (2011) 267–271.

[125] N. A. Wakeham, J. S. Adams, S. R. Bandler, *et al.*, "Thermal fluctuation noise in Mo/Au superconducting transition-edge sensor microcalorimeters," *Journal of Applied Physics* **125** no. 16, (Apr., 2019) 164503. https://doi.org/10.1063/1.5086045.

[126] O. Shein-Lumbroso, J. Liu, A. Shastry, *et al.*, "Quantum flicker noise in atomic and molecular junctions," *Phys. Rev. Lett.* **128** (Jun, 2022) 237701. https://link.aps.org/doi/10.1103/PhysRevLett.128.237701.

[127] P. Abbamonte, D. Baxter, Y. Kahn, *et al.*, "Revisiting the dark matter interpretation of excess rates in semiconductors," *Phys. Rev. D* **105** (Jun, 2022) 123002. https://link.aps.org/doi/10.1103/PhysRevD.105.123002.

[128] A. Aguilar-Arevalo, D. Amidei, D. Baxter, *et al.*, "Results on low-mass weakly interacting massive particles from an 11 kg d target exposure of damic at snolab," *Phys. Rev. Lett.* **125** (Dec, 2020) 241803. https://link.aps.org/doi/10.1103/PhysRevLett.125.241803.

[129] E. Armengaud, C. Augier, A. Benoît, *et al.*, "Searching for low-mass dark matter particles with a massive ge bolometer operated above ground," *Phys. Rev. D* **99** (Apr, 2019) 082003. https://link.aps.org/doi/10.1103/PhysRevD.99.082003.

[130] Q. Arnaud, E. Armengaud, C. Augier, *et al.*, "First germanium-based constraints on sub-mev dark matter with the edelweiss experiment," *Phys. Rev. Lett.* **125** (Oct, 2020) 141301. https://link.aps.org/doi/10.1103/PhysRevLett.125.141301.

[131] G. Agnolet, W. Baker, D. Barker, *et al.*, "Background studies for the MINER Coherent Neutrino Scattering reactor experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **853** (May, 2017) 53–60. https://www.sciencedirect.com/science/article/pii/S0168900217302085.

[132] J. Rothe, G. Angloher, F. Ardellier-Desages, *et al.*, "NUCLEUS: Exploring Coherent Neutrino-Nucleus Scattering with Cryogenic Detectors," *Journal of Low Temperature Physics* **199** no. 1, (Apr., 2020) 433–440. https://doi.org/10.1007/s10909-019-02283-7.

[133] R. Agnese, T. Aralis, T. Aramaki, *et al.*, "First dark matter constraints from a supercdms single-charge sensitive detector," *Phys. Rev. Lett.* **121** (Aug, 2018) 051301. https://link.aps.org/doi/10.1103/PhysRevLett.121.051301.

[134] Q. Arnaud, D. Asner, J.-P. Bard, *et al.*, "First results from the NEWS-G direct dark matter search experiment at the LSM," *Astroparticle Physics* **97** (2018) 54–62. https://www.sciencedirect.com/science/article/pii/S0927650517301871.

[135] A. Aguilar-Arevalo, X. Bertou, C. Bonifazi, *et al.*, "The CONNIE experiment," *Journal of Physics: Conference Series* **761** (Oct, 2016) 012057. https://doi.org/10.1088/1742-6596/761/1/012057.

[136] J. Billard, R. Carr, J. Dawson, *et al.*, "Coherent neutrino scattering with low temperature bolometers at chooz reactor complex," *Journal of Physics G: Nuclear and Particle Physics* **44** no. 10, (Aug, 2017) 105101. https://doi.org/10.1088/1361-6471/aa83d0.

[137] EXCESS workshop, "Data repository of the excess workshop." Https://github.com/fewagner/excess, 2021. https://github.com/fewagner/excess.

[138] EXCESS workshop, "Indico page of the excess workshop." EXCESS workshop, 2021.

[139] R. Anthony-Petersen, A. Biekert, R. Bunker, *et al.*, "A stress induced source of phonon bursts and quasiparticle poisoning," 2022. arXiv:2208.02790.

[140] A. Aguilar-Arevalo, I. Arnquist, N. Avalos, *et al.*, "Confirmation of the spectral excess in damic at snolab with skipper ccds," 2023. arXiv:2306.01717.

[141] E. T. Mannila, P. Samuelsson, S. Simbierowicz, *et al.*, "A superconductor free of quasiparticles for seconds," *Nature Physics* **18** no. 2, (Feb., 2022) 145–148. https://doi.org/10.1038/s41567-021-01433-7.

[142] M. McEwen, L. Faoro, K. Arya, *et al.*, "Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits," *Nature Physics* **18** no. 1, (Jan., 2022) 107–111. https://doi.org/10.1038/s41567-021-01432-8.

[143] P. Krantz, M. Kjaergaard, F. Yan, *et al.*, "A quantum engineer's guide to superconducting qubits," *Applied Physics Reviews* **6** no. 2, (June, 2019) 021318. https://doi.org/10.1063/1.5089550. _eprint: https://pubs.aip.org/aip/apr/article-pdf/doi/10.1063/1.5089550/16667201/021318_1_online.pdf.

[144] F. Marsili, V. B. Verma, J. A. Stern, *et al.*, "Detecting single infrared photons with 93% system efficiency," *Nature Photonics* **7** no. 3, (Mar., 2013) 210–214. https://doi.org/10.1038/nphoton.2013.13.

[145] K. Hattori, T. Konno, Y. Miura, *et al.*, "An optical transition-edge sensor with high energy resolution," *Superconductor Science and Technology* **35** no. 9, (Jul, 2022) 095002. https://dx.doi.org/10.1088/1361-6668/ac7e7b.

[146] H. Lattaud, E. Armengaud, Q. Arnaud, *et al.*, "Sub-MeV Dark Matter Searches with EDELWEISS: results and prospects," *PoS* **EPS-HEP2021** (2022) 153.

[147] S. Kempf, A. Fleischmann, L. Gastaldo, and C. Enss, "Physics and Applications of Metallic Magnetic Calorimeters," *Journal of Low Temperature Physics* **193** no. 3, (Nov., 2018) 365–379. https://doi.org/10.1007/s10909-018-1891-6.

[148] G.-B. Kim, "Magneto-x: Fast athermal phonon sensing with magnetic quantum sensors for dm and cevns detection," in *7th Symposium on Neutrinos and Dark Matter in Nuclear Physics (NDM22)*. 2022. https://indico.phy.ornl.gov/event/142/contributions/836/.

[149] B. von Krosigk, K. Eitel, C. Enss, *et al.*, "DELight: A Direct search Experiment for Light dark matter with superfluid helium," *SciPost Phys. Proc.* (2023) 016. https://scipost.org/10.21468/SciPostPhysProc.12.016.

[150] A. P. Vepsäläinen, A. H. Karamlou, J. L. Orrell, *et al.*, "Impact of ionizing radiation on superconducting qubit coherence," *Nature* **584** no. 7822, (Aug., 2020) 551–556. https://doi.org/10.1038/s41586-020-2619-8.

[151] L. Cardani, F. Valenti, N. Casali, *et al.*, "Reducing the impact of radioactivity on quantum circuits in a deep-underground facility," *Nature Communications* **12** no. 1, (May, 2021) 2733. https://doi.org/10.1038/s41467-021-23032-z.

[152] V. Iaia, J. Ku, A. Ballard, *et al.*, "Phonon downconversion to suppress correlated errors in superconducting qubits," *Nature Communications* **13** no. 1, (Oct., 2022) 6425. https://doi.org/10.1038/s41467-022-33997-0.

[153] A. Das, N. Kurinsky, and R. K. Leane, "Dark matter induced power in quantum devices," 2022. arXiv:2210.09313.

[154] G. Angloher, A. Bento, C. Bucci, *et al.*, "Results on light dark matter particles with a low-threshold CRESST-II detector," *The European Physical Journal C* **76** no. 1, (Jan., 2016) 25. https://doi.org/10.1140/epjc/s10052-016-3877-3.

[155] E. Gatti and P. F. Manfredi, "Processing the Signals From Solid State Detectors in Elementary Particle Physics," *Riv. Nuovo Cim.* **9N1** (1986) 1–146.

[156] F. Wagner, "Machine learning methods for the raw data analysis of cryogenic dark matter experiments," 2020. https://repositum.tuwien.at/handle/20.500.12708/15023.

[157] M. R. C. (https://stats.stackexchange.com/users/11032/michael-r chernick), "How can i find the standard deviation of the sample standard deviation from a normal distribution?" Cross validated. https://stats.stackexchange.com/q/27980. URL:https://stats.stackexchange.com/q/27980 (version: 2012-05-08).

[158] M. Mancuso, A. Bento, N. F. Iachellini, *et al.*, "A method to define the energy threshold depending on noise level for rare event searches," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **940** (2019) 492–496. https://www.sciencedirect.com/science/article/pii/S0168900219308708.

[159] S. Kokoska and D. Zwillinger, "Crc standard probability and statistics tables and formulae, student edition," 1999. https://api.semanticscholar.org/CorpusID:64249427.

[160] D. Schmiedmayer, *Likelihood fits and parametric event descriptions in cryogenic detectors.* PhD thesis, Technische Universität Wien. to be published in 2024.

[161] G. Angloher, A. Bento, C. Bucci, *et al.*, "Results on low mass WIMPs using an upgraded CRESST-II detector," *The European Physical Journal C* **74** no. 12, (Dec., 2014) 3184. `https://doi.org/10.1140/epjc/s10052-014-3184-9`.

[162] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation* **9** no. 8, (1997) 1735–1780.

[163] A. Paszke, S. Gross, F. Massa, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[164] W. Falcon et al., "Pytorch lightning," *GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning* **3** (2019) .

[165] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017. arXiv:1412.6980.

[166] D. Rizvanovic, "Convention versus convolution: Two methods of low-energy event classification in cresst," 2022. `https://repositum.tuwien.at/handle/20.500.12708/20435`.

[167] P. Holl, L. Hauertmann, B. Majorovits, *et al.*, "Deep learning based pulse shape discrimination for germanium detectors," *The European Physical Journal C* **79** no. 6, (May, 2019) 450. `https://doi.org/10.1140/epjc/s10052-019-6869-2`.

[168] C. K. Khosa, L. Mars, J. Richards, and V. Sanz, "Convolutional neural networks for direct detection of dark matter," *Journal of Physics G: Nuclear and Particle Physics* **47** no. 9, (Jul, 2020) 095201. `https://doi.org/10.1088/1361-6471/ab8e94`.

[169] A. Abdulaziz, J. Zhou, A. Di Fulvio, *et al.*, "Semi-supervised gaussian mixture variational autoencoder for pulse shape discrimination," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3538–3542. 2022.

[170] A. J. Zöller, *Artificial Neural Network Based Pulse-Shape Analysis for Cryogenic Detectors Operated in CRESST-II*. Dissertation, Technische Universität München, München, 2016. `http://mediatum.ub.tum.de/?id=1303343`.

[171] S. Delaquis, M. Jewell, I. Ostrovskiy, *et al.*, "Deep neural networks for energy and position reconstruction in EXO-200," *Journal of Instrumentation* **13** no. 08, (Aug, 2018) P08023–P08023. `https://doi.org/10.1088/1748-0221/13/08/p08023`.

[172] C. Mühlmann, "Pulse-shape dicrimination with deep learning in CRESST," 2019. `http://hdl.handle.net/20.500.12708/14865`.

[173] G. Fantini, A. Armatol, E. Armengaud, *et al.*, "Machine Learning Techniques for Pile-Up Rejection in Cryogenic Calorimeters," *Journal of Low Temperature Physics* (May, 2022) . `https://doi.org/10.1007/s10909-022-02741-9`.

[174] Y. Ichinohe, S. Yamada, R. Hayakawa, *et al.*, "Application of Deep Learning to the Evaluation of Goodness in the Waveform Processing of Transition-Edge Sensor Calorimeters," *Journal of Low Temperature Physics* (Apr., 2022) . `https://doi.org/10.1007/s10909-022-02719-7`.

[175] I. Coarasa, J. Apilluelo, J. Amaré, *et al.*, "Improving anais-112 sensitivity to dama/libra signal with machine learning techniques," *Journal of Cosmology and Astroparticle Physics* **2022** no. 11, (Nov, 2022) 048. `https://dx.doi.org/10.1088/1475-7516/2022/11/048`.

[176] R. Huang, E. Armengaud, C. Augier, *et al.*, "Pulse shape discrimination in CUPID-mo using principal component analysis," *Journal of Instrumentation* **16** no. 03, (Mar, 2021) P03032. `https://doi.org/10.1088/1748-0221/16/03/p03032`.

[177] R. Barlow, "Extended maximum likelihood," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **297** no. 3, (1990) 496–506. https://www.sciencedirect.com/science/article/pii/0168900290913348.

[178] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *The Annals of Mathematical Statistics* **9** no. 1, (1938) 60 – 62. https://doi.org/10.1214/aoms/1177732360.

[179] S. Yellin, "Finding an upper limit in the presence of an unknown background," *Phys. Rev. D* **66** (Aug, 2002) 032005. https://link.aps.org/doi/10.1103/PhysRevD.66.032005.

[180] S. Yellin, "Extending the optimum interval method," 2007. arXiv:0709.2701.

[181] S. Yellin, "Some ways of combining optimum interval upper limits," 2011. arXiv:1105.2928.

[182] G. B. Gelmini, "Light weakly interacting massive particles," *Reports on Progress in Physics* **80** no. 8, (Jun, 2017) 082201. https://dx.doi.org/10.1088/1361-6633/aa6e5c.

[183] G. Angloher, I. Dafinei, N. D. Marco, *et al.*, "Simulation-based design study for the passive shielding of the COSINUS dark matter experiment," *The European Physical Journal C* **82** no. 3, (Mar., 2022) 248. https://doi.org/10.1140/epjc/s10052-022-10184-5.

[184] V. Zema, *Unveiling the nature of dark matter with direct detection experiments.* PhD thesis, Chalmers University Of Technology, 2020. https://www.dissertations.se/dissertation/9b8d4e4321/.

[185] R. Collaboration, C. Augier, G. Baulieu, *et al.*, "Results from a prototype tes detector for the ricochet experiment," 2023. arXiv:2304.14926.

[186] A. H. Abdelhameed, G. Angloher, P. Bauer, *et al.*, "First results on sub-GeV spin-dependent dark matter interactions with ^{7}li," *The European Physical Journal C* **79** no. 7, (July, 2019) 630. https://doi.org/10.1140/epjc/s10052-019-7126-4.

[187] A. H. Abdelhameed, G. Angloher, P. Bauer, *et al.*, "Cryogenic characterization of a lialo$_2$ crystal and new results on spin-dependent dark matter interactions with ordinary matter," *The European Physical Journal C* **80** no. 9, (Sept., 2020) 834. https://doi.org/10.1140/epjc/s10052-020-8329-4.

[188] G. Angloher, G. Benato, A. Bento, *et al.*, "Probing spin-dependent dark matter interactions with ^{7}li," *The European Physical Journal C* **82** no. 3, (Mar., 2022) 207. https://doi.org/10.1140/epjc/s10052-022-10140-3.

[189] M. Ambrosio, R. Antolini, G. Auriemma, *et al.*, "Vertical muon intensity measured with macro at the gran sasso laboratory," *Phys. Rev. D* **52** (Oct, 1995) 3793–3802. https://link.aps.org/doi/10.1103/PhysRevD.52.3793.

[190] M. Aglietta, B. Alpat, E. D. Alyea, *et al.*, "Muon "depth-intensity" relation measured by the lvd underground experiment and cosmic-ray muon spectrum at sea level," *Phys. Rev. D* **58** (Oct, 1998) 092005. https://link.aps.org/doi/10.1103/PhysRevD.58.092005.

[191] M. Schumann, "Direct detection of WIMP dark matter: concepts and status," *Journal of Physics G: Nuclear and Particle Physics* **46** no. 10, (Aug, 2019) 103003. https://doi.org/10.1088/1361-6471/ab2ea5.

[192] A. Gnech, M. Viviani, and L. E. Marcucci, "Calculation of the ^{6}Li ground state within the hyperspherical harmonic basis," *Phys. Rev. C* **102** (Jul, 2020) 014001. https://link.aps.org/doi/10.1103/PhysRevC.102.014001.

[193] A. F. Pacheco and D. Strottman, "Nuclear-structure corrections to estimates of the spin-dependent wimp-nucleus cross section," *Phys. Rev. D* **40** (Sep, 1989) 2131–2133. https://link.aps.org/doi/10.1103/PhysRevD.40.2131.

[194] J. Engel, M. T. Ressell, I. S. Towner, and W. E. Ormand, "Response of mica to weakly interacting massive particles," *Phys. Rev. C* **52** (Oct, 1995) 2216–2221. https://link.aps.org/doi/10.1103/PhysRevC.52.2216.

[195] E. Armengaud, C. Augier, A. Benoît, *et al.*, "Searching for low-mass dark matter particles with a massive ge bolometer operated above ground," *Phys. Rev. D* **99** (Apr, 2019) 082003. https://link.aps.org/doi/10.1103/PhysRevD.99.082003.

[196] R. Agnese, A. J. Anderson, T. Aralis, *et al.*, "Low-mass dark matter search with cdmslite," *Phys. Rev. D* **97** (Jan, 2018) 022002. https://link.aps.org/doi/10.1103/PhysRevD.97.022002.

[197] C. Amole, M. Ardid, I. J. Arnquist, *et al.*, "Dark matter search results from the complete exposure of the pico-60 c_3f_8 bubble chamber," *Phys. Rev. D* **100** (Jul, 2019) 022001. https://link.aps.org/doi/10.1103/PhysRevD.100.022001.

[198] D. S. Akerib, S. Alsum, H. M. Araújo, *et al.*, "Limits on spin-dependent wimp-nucleon cross section obtained from the complete lux exposure," *Phys. Rev. Lett.* **118** (Jun, 2017) 251302. https://link.aps.org/doi/10.1103/PhysRevLett.118.251302.

[199] J. I. Collar, "Search for a nonrelativistic component in the spectrum of cosmic rays at earth," *Phys. Rev. D* **98** (Jul, 2018) 023005. https://link.aps.org/doi/10.1103/PhysRevD.98.023005.

[200] T. Bringmann and M. Pospelov, "Novel direct detection constraints on light dark matter," *Phys. Rev. Lett.* **122** (May, 2019) 171801. https://link.aps.org/doi/10.1103/PhysRevLett.122.171801.

[201] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation* **9** no. 8, (1997) 1735–1780.

[202] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, *et al.*, eds., vol. 30. Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[203] L. N. Smith, "Cyclical learning rates for training neural networks," 2015. https://arxiv.org/abs/1506.01186. arXiv:1506.01186.

[204] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "Wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20. Curran Associates Inc., Red Hook, NY, USA, 2020.

[205] J. St. John, C. Herwig, D. Kafkes, *et al.*, "Real-time artificial intelligence for accelerator control: A study at the fermilab booster," *Phys. Rev. Accel. Beams* **24** (Oct, 2021) 104601. https://link.aps.org/doi/10.1103/PhysRevAccelBeams.24.104601.

[206] V. Kain, S. Hirlander, B. Goddard, *et al.*, "Sample-efficient reinforcement learning for cern accelerator control," *Phys. Rev. Accel. Beams* **23** (Dec, 2020) 124801. https://link.aps.org/doi/10.1103/PhysRevAccelBeams.23.124801.

[207] F. M. Velotti, B. Goddard, V. Kain, *et al.*, "Towards automatic setup of 18 mev electron beamline using machine learning," *Machine Learning: Science and Technology* **4** no. 2, (Apr, 2023) 025016. https://dx.doi.org/10.1088/2632-2153/acce21.

[208] J. Degrave, F. Felici, J. Buchli, *et al.*, "Magnetic control of tokamak plasmas through deep reinforcement learning," *Nature* **602** no. 7897, (Feb., 2022) 414–419. https://doi.org/10.1038/s41586-021-04301-9.

[209] H. P. Nautrup, N. Delfosse, V. Dunjko, *et al.*, "Optimizing quantum error correction codes with reinforcement learning," *Quantum* **3** (Dec, 2019) 215. https://doi.org/10.22331%2Fq-2019-12-16-215.

[210] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction.* MIT press, 2018.

[211] T. Haarnoja, A. Zhou, K. Hartikainen, *et al.*, "Soft actor-critic algorithms and applications," *ArXiv* **1812.05905** (2018) . arXiv:1812.05905.

[212] R. A. Light, "Mosquitto: server and client implementation of the mqtt protocol," *Journal of Open Source Software* **2** no. 13, (2017) 265. https://doi.org/10.21105/joss.00265.

[213] R. McDermott, "Materials origins of decoherence in superconducting qubits," *IEEE Transactions on Applied Superconductivity* **19** no. 1, (2009) 2–13.

Acknowledgements

The past three years have been quite intense at times, but I have been fortunate to have a large number of people to turn to for advice or who have otherwise helped me along the way. I would like to mention a few in particular.

I would like to thank the rare event search group at HEPHY, where I have always felt welcome and which has been my safe space to ask dump questions since I started as a master's student in 2019. This atmosphere kept me productive during the Covid lockdowns, where we often had coffee breaks over ZOOM, which I usually abused to discuss physics.

My biggest supporters within this group were Jochen Schieck and Florian Reindl, which is not surprising as they were my academic supervisors. However, the amount of questions I pestered them with, not only about physics but also about organisational and career advice, and the amount of support I received in return, far exceeded what is expected from a PhD supervisor. I really appreciate this effort, it has been the single most important component in the successful completion of all my plans. Although my path will be away from HEPHY for the next few years, I very much hope that we can work together again in the future.

In the same vein, I would like to thank Wolfgang Waltenberger, my third supervisor. He was my first person to turn to for data science topics, helped me from my first day on at HEPHY, and gave me the chance to stay at HEPHY for my PhD with our joint machine learning project. Since three supervisors are not enough, I would also like to thank Clemens Heitzinger, who supervised my second master's thesis in mathematics and was a great help in getting our reinforcement learning project off the ground.

The CRESST and COSINUS collaborations provided the foundation on which I could build my work. In particular, I would like to thank Franz Pröbst, Karoline Schäffner and Vanessa Zema for extensive discussions on detector physics, without which I would never have reached the understanding necessary to complete my thesis. I would also like to thank Margarita Kaznacheeva for co-organising the EXCESS workshop, which was an adventure in itself.

Finally, I would like to pay special tribute to my parents, who have given me indescribable support over the past three decades, and to my brother, who has been an inspiration to follow the academic path. To my friends, whose company kept me sane. And finally, to my partner. Your patience, understanding and encouragement have been invaluable and I cannot express my gratitude enough.