LUND UNIVERSITY
Faculty of Science

Rejection of the Photon Induced Background in LDMX at 8 GeV

Erik Wallin

Thesis submitted for the degree of Master of Science
Project duration: 9 months

Supervised by Ruth Pöttgen and Torsten Åkesson

Department of Physics
Division of Particle Physics
June 2022

Abstract

The Light Dark Matter eXperiment (LDMX) is a proposed fixed target experiment at SLAC's LCLS-II electron beamline that will attempt direct production of sub-GeV dark matter, measurable through a missing momentum and missing energy signature.

An initial run of LDMX is planned at 4 GeV beam energy, for which a study on LDMX's capabilities to separate dark matter production events from the standard model background was previously performed. Later phases of data taking at LDMX will after an upgrade of the beamline be performed at 8 GeV beam energy, which is expected to improve both the dark matter production rate and background rejection capabilities. Using a dark photon mediator between standard model and dark matter as a benchmark model with a missing momentum and missing energy signature, a comparative study on the performance of LDMX at 8 GeV is done with the same methodology as in the 4 GeV study.

A variant of the complete background rejection sequence is presented, optimized for 8 GeV beam energy. The sequence is shown to reject each of 2×10^{14} simulated standard model backgrounds events, focusing on photo-nuclear and $\gamma \to \mu\mu$ processes in the target or forward electromagnetic calorimeter. Improvement in signal to background discrimination is seen in parts of the dark photon mass range, compared to 4 GeV beam energy.

Populärvetenskaplig Sammanfattning

Astronomiska tecken tyder på att det mesta av materian i universum inte är av den vardagliga typ vi möter på jorden, som noggrant beskrivs av standardmodellen för fysik. Att introducera denna nya substans är användbart för att lösa astrofysiska problem genom den stora gravitationella dragningskraft som denna materia har. Namnet mörk materia får den genom att det inte verkar påverkas av ljus och andra elektromagnetiska fenomen, utan kan mest påvisas indirekt genom den gravitationella effekt den har på sin omgivning.

Partikelfysikens uppgift blir att lyckas beskriva den mörka materians natur och vilka egenskaper den har, som bland annat dess massa. Ett starkt bevis för mörk materia hade varit att i ett experiment på jorden, direkt producera mörk materia. The Light Dark Matter eXperiment (LDMX) är ett planerat experiment för direktproduktion av mörk materia, fokuserat på att utforska modeller för mörk materiapartiklar som har en liten massa. Energirika elektroner från en partikelaccelerator ska träffa en måltavla av volfram, där vissa modeller postulerar att mörk materia kan skapas i ovanliga fall. Restprodukterna (strålningen) från denna kollision uppmäts av en detektor som ska kunna urskilja om mörk materia skapades, eller om det var en process som redan är känd från standardmodellen.

Produktionen av mörk materia vid LDMX är, om möjlig, en mycket ovanlig reaktion. Det är därför viktigt att kunna separera mätningar där vanliga processer händer, från de få där mörk materia har skapats. Denna uppsats analyserar möjligheterna LDMX har för att separera ut enskilda mätningar där mörk materia skapats, i en planerad andra datatagninsperiod med högre energi i partikelacceleratorn.

Acknowledgements

Thanks to Lene whose simulations and insights made this project happen, and thanks to my supervisors Ruth and Torsten who have patiently looked at the minute details of the analysis. This project builds on the studies by Leo Östman, as well as many collaborators at UC Santa Barbara, whose hard work I am thankful for. The test of a prototype hadronic calorimeter at CERN was a great experience for me personally, and I am grateful to Lennart, Geoffrey, Andrew and all others there for trying to teach me how to be an experimentalist.

Abbreviations

- AUC - Area Under the Curve
- BDT - Boosted Decision Tree
- DM - Dark Matter
- ECal - Electromagnetic Calorimeter
- EoT - Electrons on Target
- HCal - Hadronic Calorimeter
- LDMX - Light Dark Matter eXperiment
- MIP - Minimally Ionizing Particle
- PE - Photoelectron
- PN - Photo-nuclear
- ROC - Receiver Operator Characteristic
- SM - Standard Model

Contents

Chapter 1

Introduction

1.1 Analysis Goals

The Light Dark Matter eXperiment (LDMX) is a proposed experiment to be built at SLAC, attempting to detect direct production of dark matter via its missing energy and missing momentum signatures. Chapter 2 presents the general context for dark matter (DM) searches, as explored by the Light Dark Matter eXperiment, with further details of the experimental design in chapter 3.

To measure a rare dark matter production process requires an effective procedure to discriminate between events with DM production from standard model background events. An initial study at 4 GeV beam energy was done, where it was shown that (an equivalent of) 2×10^{14} background events could all be rejected.

An accelerator upgrade will let LDMX run at 8 GeV beam energy, after a shorter 4 GeV data taking period. This study explores two questions: First whether an as efficient background rejection procedure is possible for 8 GeV beam energy. Secondly, whether there are any improvements in background rejection efficiency by doubling the beam energy. To answer the second question, the background rejection procedure in this study follows the same principles and methods as in the 4 GeV study as a comparative study. This helps disentangling the physical improvements from doubling the beam energy from improvements on the background rejection procedure itself.

A background rejection sequence will be presented, optimized for 8 GeV, focused on distinguishing DM events from background mainly based on the granular forward electromagnetic calorimeter at LDMX. The background rejection procedure is evaluated on simulated data equivalent to 2×10^{14} background events at LDMX, with the goal of reaching zero background events remaining. The efficiency for DM signal events to survive the extensive background rejection procedure will also be explored.

1.2 Author's Contribution

Large scale simulations of dark matter production and standard model background processes were produced by Lene Kristian Bryngemark, with assistance from the LDCS project for distributed computing [1].

This thesis work begins with the qualitative description of the simulated data in chapter 4. All of the background rejection procedure in chapter 5 is part of this thesis project except when noted otherwise, while using the same analysis framework as in the study at 4 GeV beam energy [2]. Beyond simulation of physics and particle-matter interactions in the detector, this study is based on a new implementation of all the higher level event reconstruction variables and background rejection steps that will be discussed in chapter 5.

The analysis project that this thesis is a major part of is subject to possible publication, and all results in this text should be considered preliminary and may differ from future final results.

Chapter 2

Dark Matter Searches

2.1 The Dark Sector

Albert Michelson once claimed, mistakenly attributing the quote to Lord Kelvin: "An eminent physicist has remarked that the future truths of Physical Science are to be looked for in the sixth place of decimals." [3] This sentiment might be considered a correct prophecy for most of standard model physics, a massively successful and precise theory in its claims and predictions. As the standard model (SM) however does not propose a viable candidate for dark matter (DM), missing more than 80% of matter in the universe [4] may be considered a flaw beyond just disagreements in precision measurements. Beyond the standard model extensions are therefore considered to correct for dark matter.

Current astronomical measurements favor particle dark matter as an explanation of various astronomical phenomena, evident from its interaction with SM matter through gravitation, but without any electromagnetic interaction (thus the *darkness* of it). For particle physicists wishing to measure dark matter, only *discoverable* dark matter may be probed, where there is some non-gravitational interaction with standard model matter. Dark sector DM are particles lacking strong, weak and EM interactions, forces understood by the standard model, but still acting gravitationally and through some new interaction with SM matter. One could extend the standard model to introduce a *mediator* particle, acting on both standard model particles and on dark matter to facilitate a bridge between DM and standard model matter [5].

Past experiments searching for Weakly Interacting Massive Particles (WIMPs), that have masses above a few GeV, and various searches at colliders have been lacking to probe dark sector DM with masses below one GeV, a range for which there is now considerable interest [6]. Many dark sector models have gotten insufficient experimental attention, partly due to that such precision searches do not necessarily benefit from the collider high energy frontier [5]. A particular interest is now in the region with *light* mediator masses between 1 MeV and 1 GeV, i.e. roughly between the electron and proton masses. One possible such mediator portal between SM matter and DM is to introduce a new vector (photon-like) boson A', or through a Higgs or neutrino portal. The A', also called the dark photon, will be used in this study as a benchmark model, and is described in section 2.2.

2.1.1 Thermal Relics

It is easy to imagine that in the early hot universe, dark matter and other particles were in thermal equilibrium, as long as there is some small non-gravitational interaction [6], with the particle density of DM kept in equilibrium through constant pair-production and annihilation. As the universe expands and cools, this process might not be able to be sustained anymore, ending the thermal equilibrium between DM and SM matter. From this point onward, the abundance of (stable) DM in the universe is unchanged, freezing the amount of DM that still remains to the present day.[1] We call DM in this scenario a *thermal relic*.

To explain the current day abundances of dark matter, the DM annihilation cross section $\langle \sigma v \rangle$, averaged over the kinetic distribution of DM, is constrained by the dynamics of DM leaving the thermal equilibrium. This cross section has a lower bound in the thermal dark matter scenario:

$$\langle \sigma v \rangle \approx 10^{-26} \text{ cm}^3\text{s}^{-1} \tag{2.1}$$

to produce the right amount of DM observed today [6], as a lower cross section would lead to overabundance of DM today. This non-zero cross section necessitates that dark matter production is possible.

Assuming the thermal relic scenario and the lower bound on the dark matter annihilation cross section, it may be possible to directly produce and probe the properties of dark matter in accelerator experiments. Dark matter produced in an experiment would practically be invisible to any particle detector, but it could have a missing momentum or missing energy signature. One such experiment that searches for the missing energy signature of dark photons is NA64, that attempts dark photon production from 100 GeV electrons dumped on a fixed target [8].

A proposed model of dark matter in the thermal relic scenario must be able to explain the measured $\langle \sigma v \rangle$, which constrains the free parameters of the theory. Then only the subset of the parameter space that can adequately explain this cross section needs to be explored, the so called *thermal target*. Searches for dark matter then have concrete goals for the reach of the experiments, focusing on searches that at least reach the thermal targets. An example of such thermal targets will be later given in figure 2.2.

The rest of this analysis will be dedicated to a search probing the thermal targets in the MeV to GeV range of *light dark matter*.

2.2 Benchmark Model

As a benchmark dark sector model that captures the kinematics of direct production of sub-GeV light DM, a new vector particle A', charged under a $U(1)$ gauge symmetry, called the *dark photon*, is introduced as mediator to the hidden sector in this study. The Feynman diagram 2.1 shows the interaction between two standard model fermions f and dark matter χ, through the A' channel. This model introduces a few free parameters,

[1] For a standard introduction to freeze-out of a particle species and out-of-equilibrium dynamics in cosmology, see chapters 5.1 to 5.2 of The Early Universe, by Kolb and Turner [7].

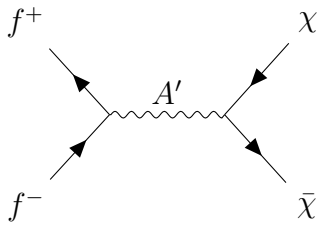


Figure 2.1: Standard model fermions f annihilating to some form of dark matter χ through a vector mediator A'.

the A' mass $m_{A'}$, the DM mass m_χ, the coupling constant between the A' and standard model matter, ϵ, and the coupling constant between the A' and DM, α_D.

The specific type of dark matter χ that the A' can decay into will not be specified, as these particles are invisible to the detector. A discussion about the experimental signatures of dark matter must only make assumptions about the properties of the A'. The nature of χ is however important for the production cross sections and in consequence, to determine the reach of the physics program.

The *light* in the Light Dark Matter eXperiment comes from that DM masses m_χ below one GeV will be considered. The experimental signature of dark matter production is dependent on $m_{A'}$ as it would change the kinematics of the direct production process. The other free parameters, α_D and ϵ, of the model will not affect the experimental signature of direct production, so the choice of them can be considered arbitrary for this study, although ϵ has significance for the cross section of dark matter production.

The ratio $m_{A'}/m_\chi = 3$ will be assumed henceforth. With $m_{A'}/m_\chi > 2$ the A' will most often decay into the $\chi\chi$ pair of interest (since $\alpha_D >> \epsilon$), and not back into standard model matter, increasing the sensitivity of the search. The regime where $m_{A'}/m_\chi < 1$ is already well constrained by cosmic microwave background measurements [6].

The dark matter to A' coupling constant α_D is assumed to be of order $\mathcal{O}(1)$ for naturalness reasons, and with $m_{A'}$, $m_{A'}/m_\chi$ also chosen, the standard model to A' coupling constant ϵ will be constrained by the thermal target. These relations are shown in figure 2.2, where the thermal targets appear as black lines, for a few different types of dark matter χ. The grey region is already excluded by previous experiments, that don't quite reach the thermal targets in this mass range. The area above the red lines shows the limits of what a missing momentum experiment, like LDMX, can probe, where it is evident that a lot of the thermal target is covered by the strongest limit of missing momentum experiments.

Note that this reach plot is for a certain set of free parameters. The experimental reach of experiments depends on the $m_{A'}/m_\chi$ ratio, with an alternative $m_{A'}/m_\chi = 10$ shown in dashed lines. The thermal targets depend on $m_{A'}/m_\chi$ as well, shifting them upwards in the figure for $m_{A'}/m_\chi > 3$.

The process of direct production of an A' relevant for this analysis, is a bremsstrahlung like process, seen in figure 2.3, where an A' is emitted from an electron recoiling on a nucleus, similar to the standard model process with a bremsstrahlung photon. The A' can then decay into some form of $\chi\bar{\chi}$, but which does not effect the kinematics of the scattered electron that is the only visible by-product of the A' production.

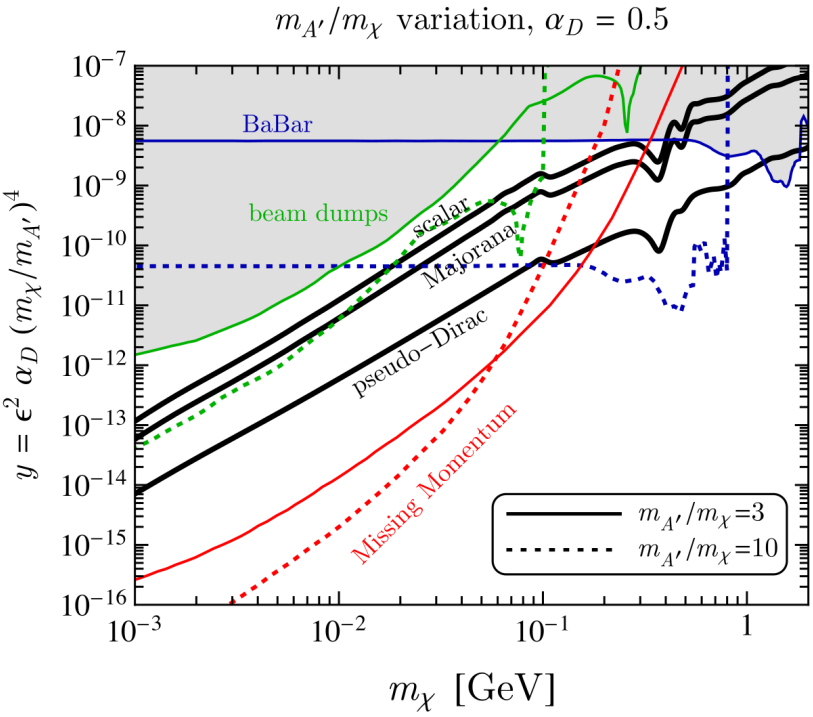


Figure 2.2: Thermal targets in the dark photon model with different types of dark matter χ, in black. Note: this is for a certain free parameter $\alpha_D = 0.5$ and dark photon to dark matter mass ratio $m_{A'}/m_\chi = 3$ or 10. The maximum reach of missing momentum experiments is shown in red, together with the already explored parameter space by existing experiments such as BaBar in blue and various beam dump experiments in green. Beam dump experiments include the Liquid Scintillator Neutrino Detector, MiniBooNE and E137. The interaction strength y is related to the thermally average cross section with a $\langle\sigma v\rangle \sim y/m_\chi^2$ behavior. Image from: [6]

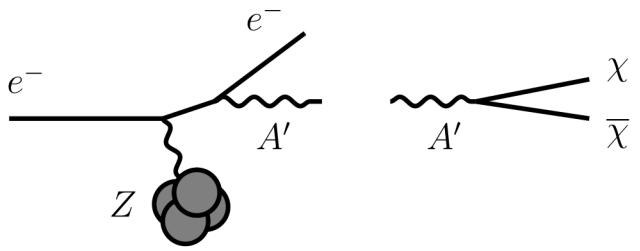


Figure 2.3: Production of a dark photon, A', from an electron recoiling on a nucleus. Image from: [6]

Chapter 3

The Light Dark Matter eXperiment

3.1 Experimental Goals

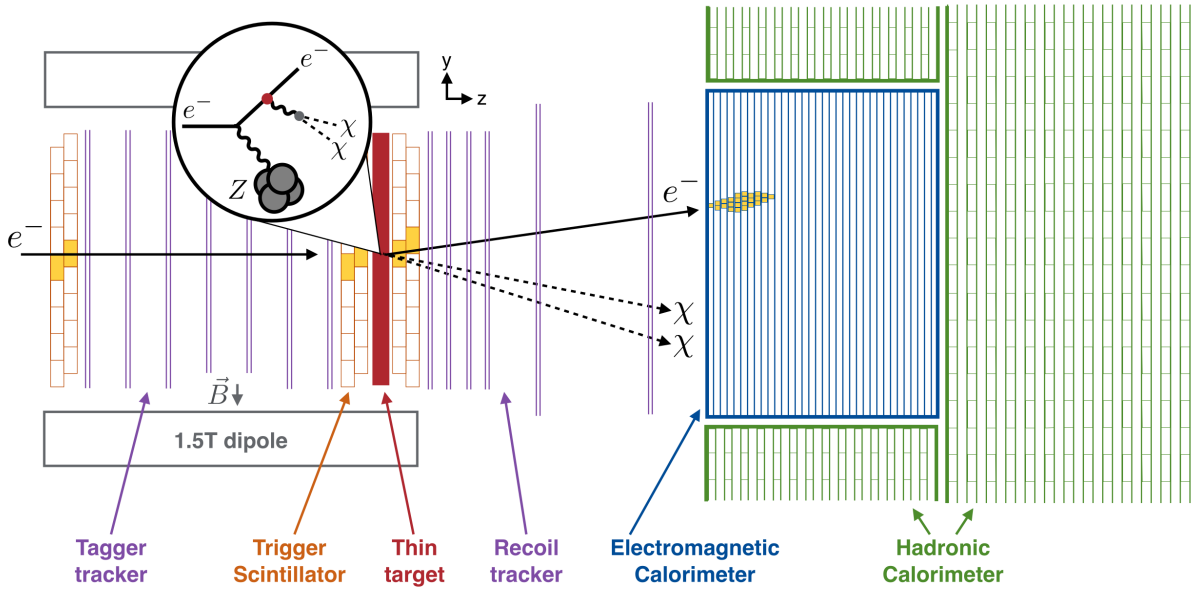


Figure 3.1: Schematic view (not to scale) of the LDMX detector showing the detector signature of invisible dark matter being produced. Image source: [9]

The Light Dark Matter eXperiment is a fixed-target experiment, searching for new physics in the scattering of few-GeV electrons from the LCLS-II linear accelerator at SLAC [10]. Electrons will scatter in a thin tungsten target and the downstream LDMX detector will make a precision measurement of the scattered particles, sensitive to missing energy and missing momentum of the scattered electron due to the possible production of an invisible signal [6].

The inset in figure 3.1 shows the production of a dark photon from an electron in the target, which subsequently decays into dark matter that is invisible to the detector. The signature of this process is then a single electron that has recoiled in the target, with missing energy and momentum that escapes with the dark photon. The design of LDMX

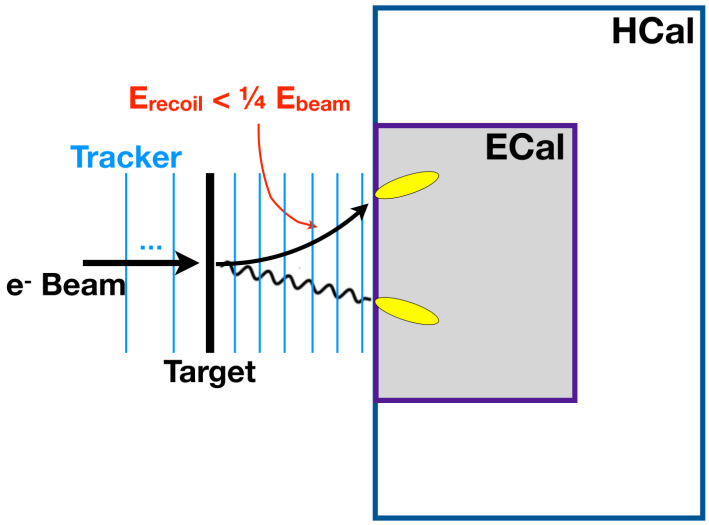


Figure 3.2: Background process with production of a bremsstrahlung photon, that induces a second shower separated from the recoil electron in the ECal. Image adapted from: [2]

is driven by the need to distinguish this characteristic dark matter production process, so called signal events, from the much more frequent standard model background processes.

One common background process that will be discussed in detail is where a photon is emitted instead of a dark photon in the scattering process, see figure 3.2, a much more common standard model process. Chapter 4 will describe several challenging cases where this background may imitate the characteristic of signal events.

The recoil electron will shower in an electromagnetic calorimeter (ECal) placed in the forward-region after the target, expecting a significant amount of energy to be missing in signal events. The ECal is also granular enough to separate energy deposits from individual showers, allowing the background process in figure 3.2 to be distinguished from signal events. Surrounding the ECal is a hadronic calorimeter (HCal) for detecting possible secondary particles induced by the photon interacting with detector material.

This study only considers one incident electron on the target at a time, which is unrealistic. A small bunch of electrons come in Poisson distributed numbers, creating pile-up, but increasing the chances of DM signal production. The trigger scintillator in figure 3.1 is responsible for determining the multiplicity of incoming electrons, which would then change the trigger's missing energy condition. The trigger scintillator is not present in the simulation geometry used in this thesis.

A run of LDMX will initially happen with 4 GeV beam energy, where the performance has been studied in detail [2]. An upgrade of the accelerator to 8 GeV is expected and is the nominal energy at which LDMX will operate after the initial commissioning run. Running at 8 GeV has expected increases in the dark matter production rate in the model considered, and possible improvements in background rejection efficiency. Chapter 5 presents the performance of the experiment at 8 GeV, and the improvement from 4 GeV the doubling of the beam energy brings.

3.2 The LDMX Detector

3.2.1 Target and Tracking

The few-GeV electrons scatter in a tungsten target 0.1 radiation lengths[1] thin. There are at least two design considerations to motivate this choice of target: A material such as tungsten with high atomic number Z is needed to stimulate the A' production process in figure 2.3. Additionally for good momentum resolution in the recoil tracker, a thin target with few radiation lengths is desired to reduce multiple scattering, a process where the electron or other secondary particles would scatter several times in the target at small deflection angles.

A series of tracking layers downstream of the target, forming the recoil tracker, provide the direction and momentum of recoil electrons and possibly other charged particles produced in the target. The first four layers closest to the target consist of two arrays of silicon microstrips, rotated 100 mrad relative to each other. These slightly rotated *stereo* layers provide better measurements of the particle's direction, which aids with reconstructing the track, and the position where the reaction took place in the target. Further downstream are two so called *axial* layers, that instead provide a better determination of the particle's momentum, as the recoil tracker is tasked with all missing momentum measurements [6].

A 1.5 T dipole magnet leads electrons to the target, while they are tracked with a series of silicon microstrip tracking layers in order to ensure that incoming beam electrons have the correct momentum and to determine where they hit the target. The fringe fields of the dipole magnet extend into the recoil tracker, to facilitate momentum determination in that region.

3.2.2 ECal

The LDMX electromagnetic calorimeter (ECal) is silicon-tungsten sampling calorimeter[2] consisting of 34 layers, spaced as in figure 3.3, a design adapted from the High-Granularity Calorimeter in the CMS experiment at CERN [6]. The layers near the front of the ECal are spaced more compactly to increase the longitudinal resolution in this region, where most of the ECal activity will be contained. Each layer is split into 7 hexagonal segments, each with 170 mm between opposite parallel edges, as seen to the right in figure 3.3. Furthermore, each hexagonal module is split into granular individual cells, a few mm in size, that each makes an individual measurement of the energy deposited inside it. This resolution is needed to separate individual calorimeter showers, which may be highly collimated and overlapping, as well as for the possibility to track individual minimally ionizing particles through the calorimeter [6].

[1] A radiation length X_0 is defined as the distance for which an electron would lose $1/e$ of its energy while travelling through the material.

[2] A sampling calorimeter makes energy measurements in active regions that are separated by layers of dense inactive material, where the energy loss dE/dx is higher. Significant energy is then lost in the inactive region, which requires indirectly reconstructing the original particle's energy, but will limit the size and extent of the particle shower in the detector.

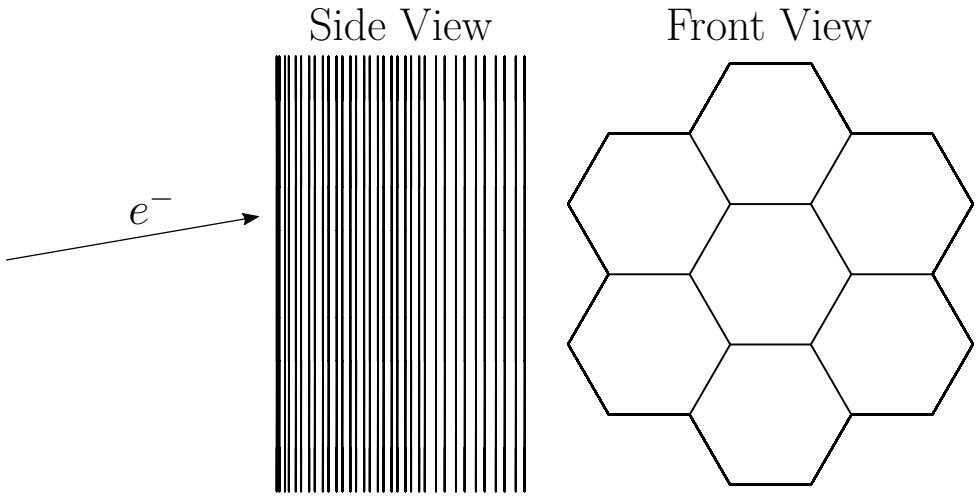


Figure 3.3: Shape of the ECal, with the side view showing the uneven spacing of the 34 layers, and the front view showing contour of the 7 honeycomb modules as seen from the front.

The first 20 layers of the ECal are designated as the *front ECal*, and are also used for triggering. A shower induced by the recoil electron will be well contained in the front ECal, such that for a signal process with a missing energy electron, the front ECal would be sufficient for seeing that energy is in fact missing. Only events with a small amount of energy in the front ECal can be from a signal process, and events with a larger deposit will not be triggered on, keeping the event from being saved to disk.

The ECal contains 2779 individual cells in each layer that make independent measurements of the energy deposit inside the cell. Any cell with an energy deposit recognizable above electronic background noise will be called an *ECal hit*, with a typical event having few hundred such hits. In simulation, electronic noise is added to all ECal cells. The noise has a Gaussian distribution with a standard deviation of 5.93 keV. To remove ECal hits that are purely noise, all hits with energy less than $4 \cdot 5.93 = 23.7$ keV, beyond four standard deviations, are removed completely. This means that very small energy deposits down to 10^1 keV scale can be measured in the ECal.

To be able to separate several showers in the ECal, the showers should be contained in as small a region of the ECal as possible to minimize the risk of several different energy deposit features overlapping in the same cells. One measure of the development of showers is the Molière radius. If one imagines a cylinder that on average contains 90% of the shower energy, placed around the central axis of the shower, its radius is the so called Molière radius. This radius is 27.8 mm in the ECal [6]. Further discussion about the shape and development of showers, for both electron and photon showers, is covered in section 5.2.1.

3.2.3 HCal

Secondary interactions in the target and ECal can create more complicated final states, with for example muons or various hadrons. Neutral hardons especially, like neutrons and neutral kaons, may register poorly in the ECal or even escape undetected. A large

hadronic calorimeter (HCal) with a much coarser segmentation of individual cells than the ECal, is placed around the ECal to tag any such escaping secondary particles [6]. As signal events expect no such escaping secondary particles, the HCal merely has to detect them without reconstructing the event with accurate detail.

The HCal is a sampling calorimeter with alternating layers of steel absorbers and layers of polystyrene scintillator bars with a 15×50 mm cross section. These scintillator bars are read out by silicon photo-multipliers (SiPM) on both ends of the bar. As the occupancy is expected to be low in the HCal, this bar structure and the coarse resolution, is sufficient.[3]

The HCal is separated into two parts: The deep *back HCal* is behind the ECal in the forward region, to detect hard particles from secondary interactions in the target and ECal. The smaller *side HCal* surrounds the ECal in order to catch softer particle transverse to the beam direction.

Photons created in the scintillating material induced by a passing particle will enter a fiber in the center of the bar, where they can travel unhindered to the readout ends of the bar where the SiPM is mounted. A SiPM is a matrix of solid-state pixels that may detect individual photons in each pixel, and the electronic response of one photon causing one single pixel to activate will be called a photoelectron (PE). A minimally ionizing particle (MIP) passing through a bar creates, on each readout end, an average response equal to 13.5 PE in the HCal. A few photo-electrons is a small physical quantity to measure accurately, considering electronic noise, and in order to allow some leeway in the HCal electronics this study considers HCal bars with up to 8 photo-electrons to be inactive. In the 4 GeV paper this limit was only 5 photo-electrons, and this study will later show that a looser cut does not significantly effect the physics results.

[3]Since the bar is read out on both ends, it is possible to infer where in the bar a particle passed through using the time difference it takes for the scintillating light to travel to the two ends.

Chapter 4

Signal and Background Characteristics

4.1 Signal Signature

In the production of a dark photon A' and its subsequent decay into dark matter, illustrated in figure 3.1, the only visible signature is that of the recoil electron, with the A' and dark matter being assumed to be completely invisible to the detector. An ideal *signal* event therefore expects one single track in the recoil tracker and a shower in the ECal stated by the electron, and nothing else.

The recoil tracker should determine significant missing energy in the recoil electron that escapes with the dark photon, and the missing-energy electron shower will be contained within the front ECal.

One motivation for increasing the beam energy is the positive effects it has on signal production rate. Figure 4.2 shows that there is a higher signal production rate at 8 GeV, almost reaching a factor 10 more than the yield at 4 GeV on the higher end of the $m_{A'}$ spectrum. As the number of background events outnumber signal events by many orders of magnitude, this enhancement is desirable. 8 GeV is currently the beam energy proposed for high energy upgrade of the LCLS-II accelerator [11], at which LDMX can operate at a second phase of data taking.

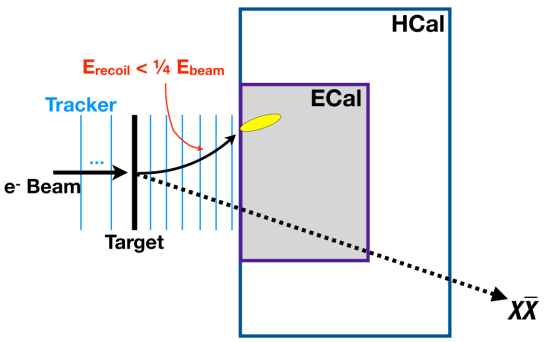


Figure 4.1: Signature of invisible dark matter production at LDMX, with only an missing momentum electron showering in the ECal. Image from: [2]

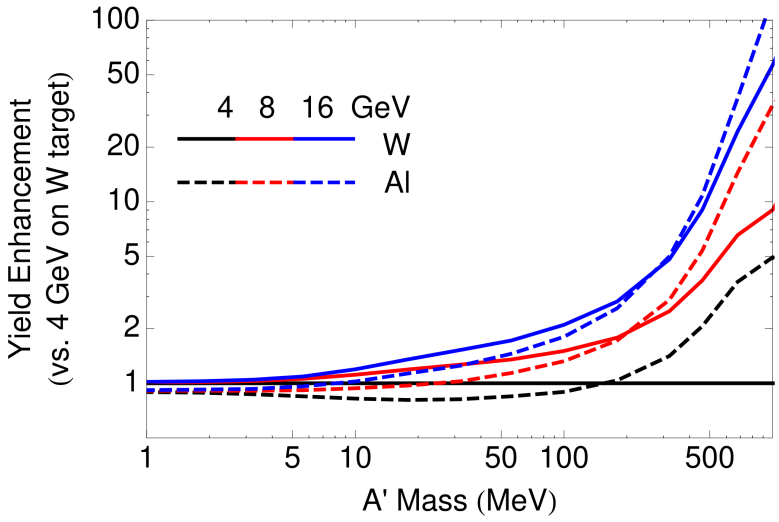


Figure 4.2: Enhancement of A' signal event rate, for events accepted by the detector, compared to the scenario with a 0.1 radiation lengths tungsten target and 4 GeV electrons. The black, red and blue lines show the enhancements at 4, 8 and 16 GeV beam energy. The solid and dotted lines show the enhancement with a tungsten or aluminium target respectively. The aluminium target scenario will not be considered in this text. Image from: [6]

4.2 Background Ladder

As LDMX aims to veto every kind of standard model process there is a diverse set of background that needs consideration. The relative rate of backgrounds is shown in figure 4.3, and they are separated into *visible* background that the experiment can reject, and rare *invisible* background that there is no experimental handle to separate from signal events on the rate of $< 10^{-16}$.

Figure 4.3 shows that most often, the electron will simply pass through the target without a significant interaction. The non-interacting electron would shower in the ECal without much energy missing. The electron shower would be well contained by the ECal and won't exhibit the missing energy as expected from a signal event.

Among the more interesting backgrounds is where the electron scatters in the target, emitting a bremsstrahlung photon in the process. If the recoiling electron and bremsstrahlung photon both shower in the ECal, there is no significant missing energy. But the photon can undergo secondary interactions producing more complicated final states that may imitate missing energy. Four classes of such photon induced backgrounds will be considered in detail in this study, namely where a hard bremsstrahlung photon undergoes a photo-nuclear (PN) interaction or conversion into a muon pair, either in the ECal, or in the target. Special attention will be given to the ECal photo-nuclear class, for which most background rejection steps will be optimized for.

Neutrinos produced in the target would give rise to missing energy in a signal-like way, as the neutrino is invisible to the detector. These irreducible neutrino events will be misclassified as signal, but the number of such events is only expected, at the very most,

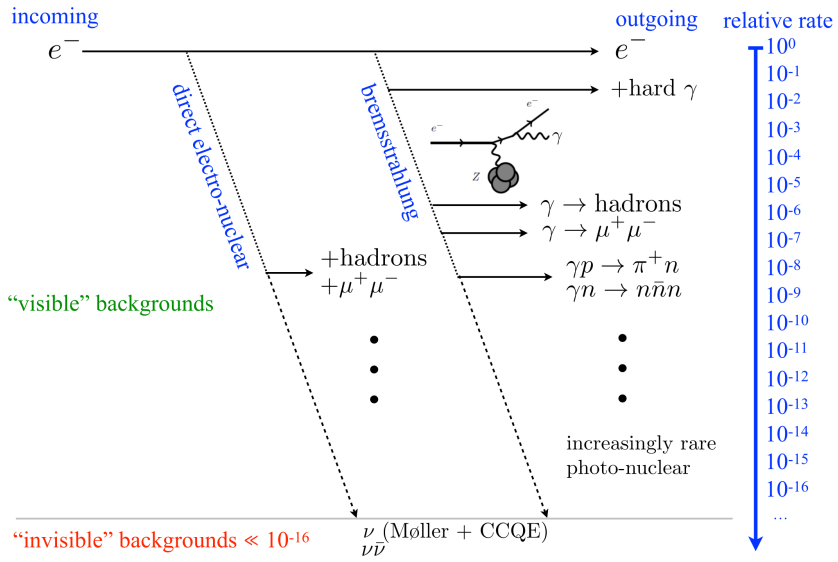


Figure 4.3: Relative rates of common background processes at LDMX. Image from: [6]

to be at the order of $\mathcal{O}(1)$ when considering 10^{14} to 10^{16} total events.

4.3 ECal Photo-Nuclear Interactions

A standard model process much more common than the production of an A' would have the electron scattering in the target and producing a recoiling bremsstrahlung photon, see Fig. 4.4. If this bremsstrahlung photon showers in the ECal, well separated from the recoil electron's shower, then it is evident that this is not a signal event by the fact that the ECal's granularity can separate the two showers. The photon may however have secondary interactions in the front layers of the ECal, possibly carrying away energy into hadronic final states from photo-nuclear interactions, and one has to then rely on identifying less energetic ECal features induced by the photon. The veto process for background rejection will in later sections mostly be focusing on this challenging ECal photo-nuclear background.

Figure 4.5 shows what type of particles the bremsstrahlung photon ends up in through the PN reaction, as simulated by Geant4 [12]. Notably, many types become more rare when going to 8 GeV beam energy. Here, all other event categories are grouped together as *multi-body reactions* in the last bin. Note that this histogram only bins events that pass the experimental trigger, which will be further explained in section 5.1.1. A refers to any nucleon and *exotics* to any particle not named in the other categories. The kaon bins consider exactly one kaon and an arbitrary amount of other particles.

Figure 4.5 does not show a significant change in charged kaon events, a background that will be discussed in detail in section 5.3. While strangeness production could intuitively be thought to be more favored at higher energies, the non-trivial relation between enhanced kaon production to the trigger must also be considered. Neutral kaons, that are inefficiently measured by the ECal and therefore require a capable HCal, only show minor changes when doubling the beam energy.

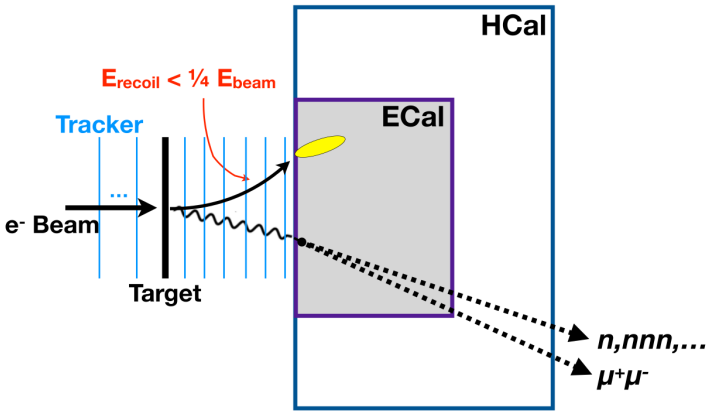


Figure 4.4: A bremsstrahlung photon can induce photo-nuclear interaction or undergo muon conversion in the front layers of the ECal, so that the photon does not fully shower. Image from: [2]

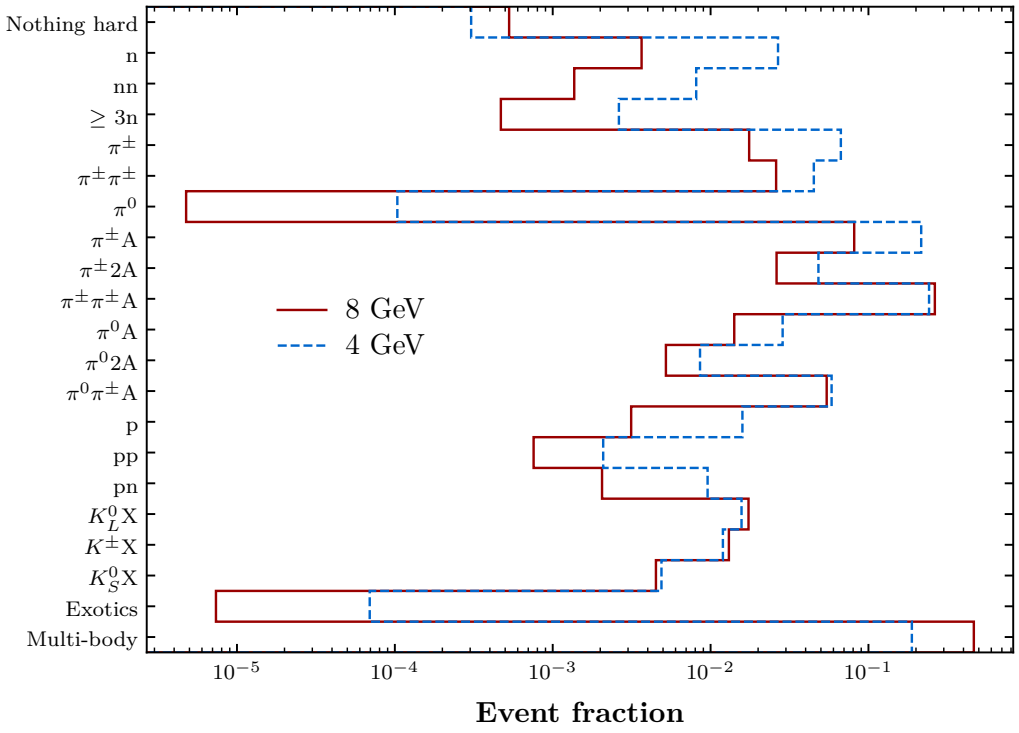


Figure 4.5: Classification of daughter particles from PN reactions from the bremsstrahlung photon. *Soft* particles below 200 MeV are not considered, and events with no hard, energetic, particles fall into the top-most bin. The kaon bins represent exactly one kaon and arbitrarily more non-kaon particles. Other types of particles are classified as *exotics* and more complicated final-states go into the multi-body bin.

If the ECal can not detect the hadronic final states from photo-nuclear reactions, the particles may deposit energy into the surrounding HCal. This is especially important for neutral hadrons that will not effectively interact with the electromagnetic ECal. Any activity in the surrounding HCal is a sign that it is not a signal event, as will be discussed in section 5.1.3. The HCal is purposefully designed to have very low inefficiencies for detecting single neutrons [6], and similarly very low inefficiencies for detecting single neutral kaons is desired.

The ECal has high enough granularity to resolve a track from individual minimally ionizing particles (MIPs). Charged kaon pairs from a photo-nuclear reaction may leave track like features to recognize them. At 4 GeV beam energy, the production of one hard and one soft charged kaon was an especially challenging background, and section 5.3 will be devoted to an algorithm for finding MIP tracks. Finding tracks becomes an especially important handle to recognize these events, when the hard kaon decays while still inside the ECal, giving most energy into an invisible neutrino, as the most common decay channel for charged kaons is with a neutrino. Such an event could have little activity in the ECal, and missing energy disappearing with the neutrino.

4.4 Additional Photon Induced Backgrounds

4.4.1 ECal Muon Conversion

Muon pair-production may be induced by the bremsstrahlung photon in the front layers of the ECal, as shown in figure 4.4. At this energy scale this process happens mostly in γ-nuclei elastic scattering, rather than scattering on the atomic electrons, and assuming that all energy goes into the pair-produced muons and not into the nuclei well describes the process [13]. The total cross section for this process increases with the photon energy, so understanding this background process becomes even more important at 8 GeV.

Muons will often not fully deposit their energy into the ECal, often escaping through the sides. Then, the process will be recognized by deposits from the muons near the trajectory of the bremsstrahlung photon, and not solely by missing energy. Like the charged kaon pair in ECal PN events, minimally ionizing muons can also leave longer well-defined track like features, and may also decay into neutrinos while still inside the ECal.

4.4.2 Target Interactions

While the bremsstrahlung photon will most often interact in the front layers of the ECal, it can also induce a photo-nuclear or muon conversion interaction in the target. As the tungsten target is thin, only 0.1 radiation lengths, the rates for this process is suppressed compared with the same reaction in the ECal as the photon will often simply leave the target without reacting.

In the case of muon conversion in the target, the recoil tracker has a good opportunity to see not only the track of the recoil electron, but either of the two muons produced. At 4 GeV beam energy the recoil tracker could reduce this background by roughly three orders of magnitude [6].

Background	**EoT**	**Biasing Factor**
ECal PN	1.97×10^{14}	550
ECal $\gamma \to \mu\mu$	2.00×10^{15}	25000
Target PN	1.65×10^{14}	550
Target $\gamma \to \mu\mu$	1.08×10^{16}	10000

Table 4.1: Electron on Target equivalent sizes of simulated background datasets at 8 GeV.

4.5 Simulated Samples

Samples of simulated events are produced to evaluate the response to both standard model backgrounds and direct dark matter production in the LDMX detector, as well as separate samples for optimizing the signal-background discrimination procedure. The particle-matter interaction and detector response is simulated using the particle transport framework Geant4 [12], together with a complete model of the target, recoil tracker, ECal and HCal detector geometry. This study uses the same geometry as in the 4 GeV veto study [2] in order to make a direct comparison.

For a certain background process, the size of a simulated dataset is measured with respect to how many electrons on the target (EoT) would produce a certain amount of such events, in a subset of all possible background events. The goal is to reject background equivalent to at least 2×10^{14} EoT, but it is computationally unfeasible to simulate and store that many events. For a background process with a relative rate of 10^{-6} among all possible interactions, then only 10^8 events need to be simulated, as a subset of the 10^{14} EoT. Some further computational tricks are used in the simulation: To generate just a specific class of background, uninteresting events can be thrown away before the full simulation is completed. Additionally, as very rare backgrounds are considered that would require throwing away many such uninteresting events, Geant4 can artificially increase the cross section for the process of interest, and do so in a specific volume such as the ECal or the target. The factor which a cross section is increased by is called the *biasing factor*. As increasing a cross section is not merely a computational trick, but actually a change to the physics, the biasing factor should be increased no more than necessary to complete the simulation in a reasonable time.

The first step of background rejection, as will be described in section 5.1.1, is the experimental trigger that expects missing energy in the ECal. Any event not passing the trigger will be immediately discarded and not stored on disk, bringing down the total number of stored events by several orders of magnitude. As this analysis is not concerned with trigger studies, this is not an issue.

The results of the background rejection procedure will be evaluated on large samples of the four previously described backgrounds, as seen in table 4.1, where there is roughly a 2×10^{14} EoT dataset for each background class. The target photo-nuclear background only reaches 1.65×10^{14} EoT, but it will be evident later that this is not expected to be a difficult background to veto.

The biasing factor for ECal $\gamma \to \mu\mu$ is 25000, meaning that the cross section for the photon to pair-produce muons is artificially and significantly increased. This could skew where in the calorimeter the muon conversion happens, making the pair-production

$m_{A'}$ [MeV]	Events
0.2	41295
0.5	35704
1	353528
18	89496
32	122565
56	38969
100	135590
180	54777
320	169426
460	50619
680	69559
1000	88472
Σ	1250000

Table 4.2: Distribution of A' masses in the signal events used for training the BDT in section 5.2

happen closer towards the front of the ECal than it would have with the true physical cross section. As a shower induced by the photon would happen early in the ECal, any muon conversion process must happen early as well before the photon shower. Therefore, this study will accept the physical consequences of the large biasing factor.

Signal samples do not have an EoT equivalent as the background does, as there are several free parameters determining the number of expected DM events for a run of 10^{14} EoT. As this study makes no particular choice for these parameters other than for $m_{A'}$, signal events will always be given in absolute numbers. Between 4.5×10^5 and 1.1×10^6 signal events for evaluating signal efficiencies were simulated for mediator masses $m_{A'} =$ 1, 10, 100, 1000 MeV.

In section 5.2 some statistical learning will be done, with the method of so called Boosted Decision Trees. This requires separate *training* data sets, as to not optimize the background rejection on the same data on which the results are evaluated. 1.25×10^6 ECal PN events were produced for such a purpose, together with 1.25×10^6 signal events distributed between 0.2 MeV and 1 GeV, with a distribution of mediator masses skewed towards the low end that are more difficult to veto. Table 4.2 shows the full distribution of mediator masses. The same training data is used to optimize free parameters in the tracking in section 5.3, but only using the 1 MeV signal samples.

Chapter 5

Background Rejection

LDMX aims to be a zero-background experiment, meaning that every standard model background event is to be vetoed, such that in principle every event passing the background rejection is from dark matter production. Idealistically, a background rejection procedure to reject each of 2×10^{14} SM background events is desired. This chapter will present a series of background rejection steps on LDMX simulation data, optimized to remove such a 2×10^{14} EoT background.

To simulate very rare background events that occur with a relative rate of 10^{-13} requires excellent modeling of the background physics processes, which pushes the capabilities of Geant4 to its limits [6], requiring careful background studies once LDMX is built. But even if the cross sections of rare background processes are uncertain, the simulated detector response to such events is interesting for estimating the background rejection efficiency at LDMX, and in turn the reach of LDMX's dark matter searches.

Four classes of background events are simulated, as described in chapter 4. The background rejection procedure will be optimized on only the most challenging of these backgrounds, the ECal photo-nuclear class, expecting it to generalize to the three other classes.

A study on the performance of LDMX at 4 GeV beam energy [2] was performed, where the zero-background goal was reached. The following comparative study will follow the very same principles for background rejection, to see the benefits in performance purely by increasing the beam energy.

5.1 Trigger, Recoil Tracker and HCal Vetos

5.1.1 Trigger

The signal characteristic expects only the recoil electron to shower in the ECal. The first veto, and most significant in absolute numbers, is therefore the trigger that expects only a small energy deposit in the ECal from the recoil electron, whereas if a bremsstrahlung photon also showered completely then the total energy would be close to 8 GeV. Setting such a condition must not only take into account preserving the signal efficiency, but to consider the two technical limitations for the trigger strategy.

The first is about how to aggregate the energy from the large number of ECal cells

quickly in order to make a split-second decision for the trigger. As common recoil electron or photon showers, i.e. with no complications such as photo-nuclear reactions, will be well contained by the front of the ECal, the following naive trigger strategy will simply be summing all energy in the 20 first layers. The second technical consideration is that of the data bandwidth the experiment can process, targeting nominally a 5 kHz event rate [6]. At 4 GeV, the trigger strategy also discussed here was deemed to have a sufficient rate reduction in principle [2], leaving the technical details for further studies.

At 8 GeV, the condition for the trigger is to keep events with strictly less than 3160 MeV in the 20 first ECal layers[1]. This is roughly double the number found for 4 GeV beam energy, 1500 MeV, such that the cut has doubled together with the beam energy. The 3160 MeV value of the cut was determined such that the signal efficiency of the trigger is the same as it was at 4 GeV beam energy.

If the 3160 MeV cut is too loose with regards to some technical limitation such as data bandwidth, that would require it to be lowered, is not crucial for this analysis. If the zero background goal can be reached with this trigger cut (which will be seen in section 5.4.1), then a tighter cut will not further improve the results. If instead the cut was proposed to be lowered for reasons of improving the background rejection capabilities of this particular veto step, the same procedure of summing the energy in the ECal can be done in what will be described in section 5.2, complementing any deficiencies in the trigger in the offline analysis instead.

5.1.2 Track Veto

A signal event expects a single track in the recoil tracker induced by the electron. Events with zero or more than one track will therefore be vetoed. We will not consider a realistic tracking algorithm to infer the number of particles and their momenta from their interaction in the recoil tracker, but instead take their simulated paths for granted, the *simulation truth*. The momentum of the tracked particle as it enters the face of the ECal is therefore taken directly from simulation truth. Although in a small step towards realism, we will impose a condition on whether the track is *findable* or not, ignoring the track if it does not hit enough layers in the recoil tracker as a naive way of saying whether it could or could not be reconstructed with a real tracking algorithm in place.

We let a particle's track be findable if it hits at least four layers in the recoil tracker. Two more conditions affect the findability of tracks: Firstly, neutral particles will never be considered findable. Secondly, a 1% inefficiency for hits to be registered in a layer is introduced, to introduce a chance for tracks to be missed. An event will only pass the track veto if it has exactly one such findable track (that is most often, but not necessarily, the recoil electron).

The production of an invisible signal should carry away some momentum of the beam electron, such that the recoil electron has some missing momentum. The particle responsible for the single findable track should enter the ECal with momentum less than 2400 MeV/c (taken from simulation, and not inferred from the path through the recoil tracker). This is simply twice the 1200 MeV/c cut used at 4 GeV beam energy.

[1]Calculation courtesy of Leo Östman

5.1.3 HCal Veto

Any secondary particle from photo-nuclear reactions that reaches the HCal is a sign that it is a background event, as no energetic secondary particles are expected from the electron shower in signal events. If the SiPMs on any HCal bar see 8 or more photoelectrons (PEs), the event is rejected as background. In the 4 GeV study a cut of 5 PEs was applied [2], but slightly increasing it will give some leniency to more noise in the HCal readout electronics. The discussion of signal efficiencies in section 5.4.2 will show that this veto condition will only reject a few percent of signal events at most.

The HCal veto complements the ECal especially for neutral hadrons produced in photo-nuclear reactions, which may escape the ECal undetected. The HCal geometry is therefore optimized to efficiently detect single neutrons and neutral kaons. Inefficiencies in this veto become evident on the scale of the simulations here, and in the results section 5.4.1, one such event with a hard K_L challenges the kaon efficiency of the HCal, entering the HCal without depositing significant energy into any bar.

5.2 ECal BDT

For signal to background discrimination in the ECal and the high-resolution information it provides, a more intricate statistical learning method is introduced. If we were to fully utilize the granularity of the ECal in our veto sequence, there is no intuitive single value constructed from features in the ECal that fully captures all fine details. We will instead present a large set of variables that each capture some aspect of the ECal activity. Introducing a long series of cuts on these input variables is possible, but would ignore any mutual information that the variables can provide together. Instead, the statistical method of boosted decision trees (BDT) is used to aggregate all input variables into a single *BDT score* determining its classification as background or signal. The 42 input variables to the BDT are discussed in section 5.2.1 and all of them enumerated in appendix A.

The basic building block of a BDT is a decision tree, illustrated in figure 5.1. A decision tree is a function $f(x_n)$ that through a tree of decisions assigns a prediction score to the input values. In the example, the tree takes two input values A and B and walks down the branches consisting of single cuts, *splits*, then returns the prediction score and the final leaf it descends to.

The modeling capability of the trees is improved when considering that many trees can be used simultaneously, summing prediction scores from a larger *ensemble* of trees. The sum of predictions from each leaf will from now on will be called the *BDT score*. A background event is defined to have a score of 0, and a dark matter signal event of 1. With the objective of the BDT properly defined, what remains is to construct an ensemble of trees that does this classification. Without delving into detail, using a *training dataset* of events where the truth of them being signal or background events is known, a *gradient boosting* algorithm [14] optimizes an ensemble of trees to perform this signal to background discrimination[2].

[2]The implementation used in this study comes from the Python library XGBoost [15].

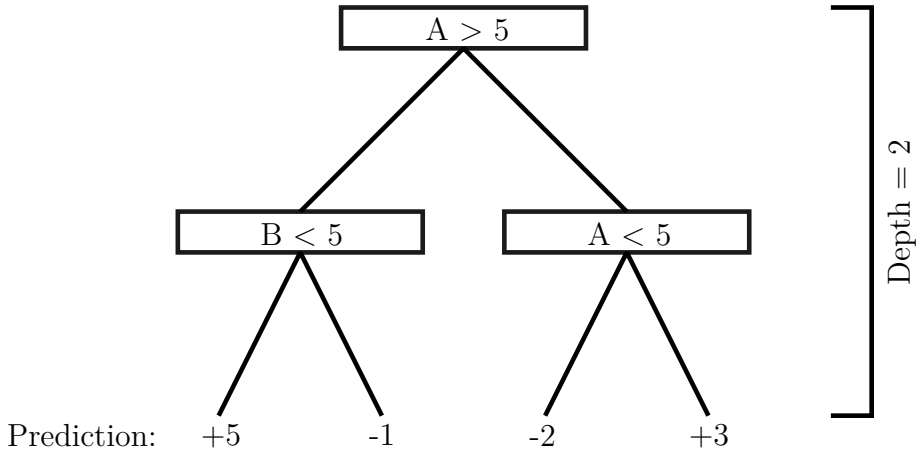


Figure 5.1: A single decision tree taking two input variables A and B, and with three *splits* (decisions), assigns a prediction score at each leaf in the bottom.

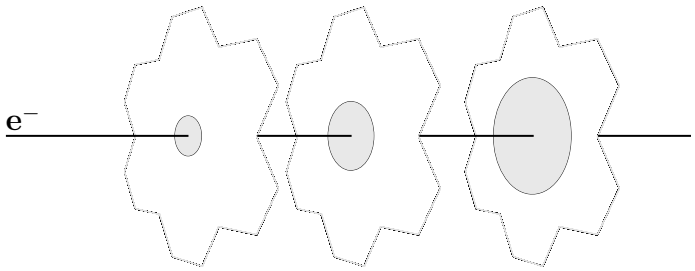


Figure 5.2: Schematic view of containment radii placed around the projected path of the incoming electron, as a disk on each individual ECal layer that should on average contain 68% of the electron shower energy in that layer. Separate sets of shapes exist depending on the electron's momentum and incoming angle to the first ECal layer.

The training samples in section 4.5 with 1.25×10^6 ECal photo-nuclear background events and 1.25×10^6 signal events of various mediator masses is used for the training the BDT, all of which pass the trigger but not necessarily any other veto.

5.2.1 Input Variables

42 BDT input variables are calculated from activity in the ECal, all with the goal of seeing whether there is any activity in the ECal outside the recoil electron shower. Some are global features of the hits and energy in the whole ECal, some variables try to describe the longitudinal (along the beam direction) distribution of energy, and the largest subset of variables describe the transverse distribution of energy.

The recoil electron's momentum and expected path through the ECal is measured by the recoil tracker. A standard volume that contains most of the electron shower energy is constructed around the expected path the electron would take if it did not shower. Each layer is given a *containment radius* that on average contains 68% of the energy depositions of a single signal electron in that layer[3], schematically depicted in figure 5.2.

[3]Inspired by a similar procedure used in the CMS Endcap HGCAL [16].

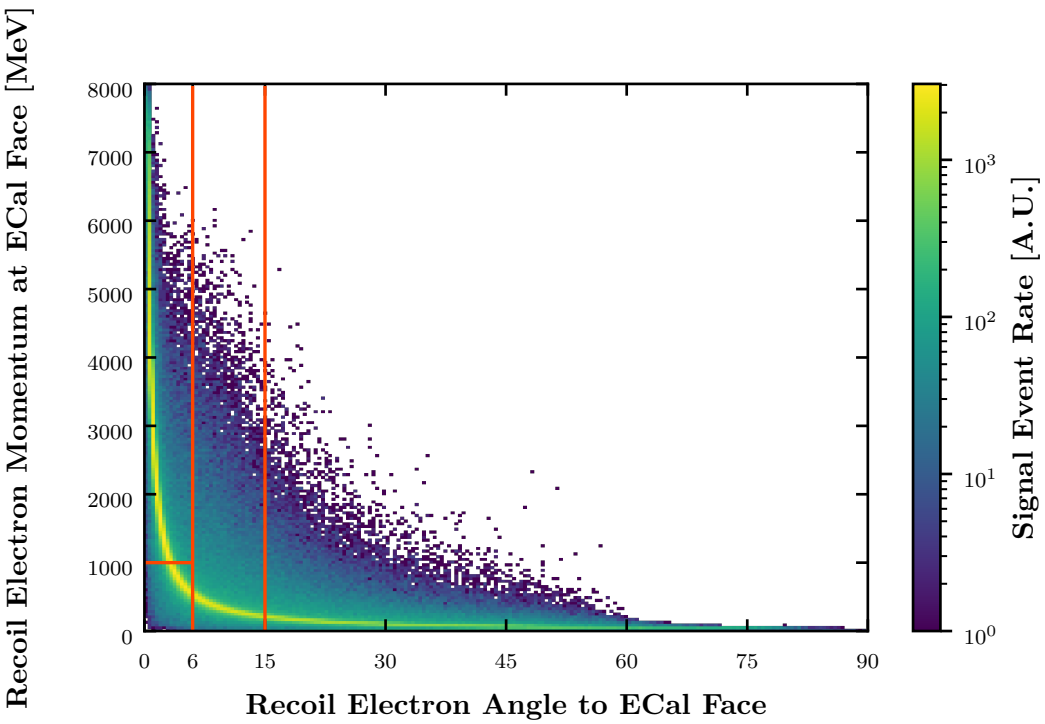


Figure 5.3: Distribution of momenta and incoming angles of signal electrons. The angle is between the beamline axis direction and the face of the ECal, as the recoil electron hits the ECal face. The lines split the space into four bins, as used by the containment radii.

The energy in the ECal contained inside or outside this volume becomes a rough metric for the transverse distribution of energy around the electron trajectory. The shower shape in the ECal is dependent on the electron's momentum and incoming angle. E.g. with an incoming angle, the ECal planes are not necessarily transverse to the electron's path any longer. Figure 5.3 shows the distribution of recoil electron kinematics, where it is shown that electrons may enter the ECal with a large angle. This phase space is split into four bins that will be treated separately, to better describe the shower shapes, which should especially differ between low and high incoming angles. A finer binning of the electron p and θ phase space could be imagined, or more complicated models for the shower volume altogether.

The containment radii for each ECal layer are given in figure 5.4, separating the four parts of the electron momentum and incoming angle phase space. It is seen that showers

are collimated in the first few layers of the ECal, meaning that only a narrow radius is needed to contain the shower, before developing in the transverse direction in deeper layers. The amount of energy deposited in a layer peaks at around layers 6-8 in the front ECal, with only small amounts of energy in the back ECal. The containment radii are somewhat similar for all electrons with $\theta^\circ < 15$, but a large change is seen in the $15 < \theta^\circ$ case where much wider radii are needed to contain the shower.

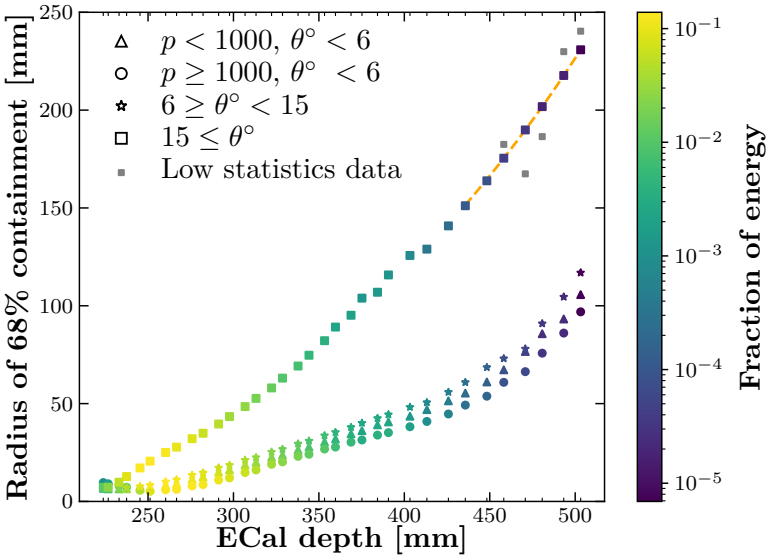


Figure 5.4: Radius of disks in each ECal layer that contains on average 68% of the energy in an electron shower. These are split into four cases depending on the kinematics, by the electron's momentum p at the ECal face and the electron's angle θ° to the ECal face as it enters. In the $15 < \theta^\circ$ case, a fourth-degree polynomial fit is used instead in the last 5 ECal layers, due to low statistics.

The energy deposits from photon showers around the inferred bremsstrahlung photon path can be contained similarly, though always using the smallest set of electron radii $p < 1000, \theta^\circ < 6$. The hits near the photon trajectory are especially important, because no activity is expected there in signal events in contrast to background events! The electron shower and photon induced hits in the ECal may be partly or completely overlapping however, complicating the original problem of finding activity from the photon that is separate from the electron shower.

The electron kinematics is inferred from simulation truth information and no tracking or momentum determination algorithm is used. The bremsstrahlung photon trajectory is inferred from the electron's path through conservation of momentum, using the fact that the incoming electron has exactly 8 GeV/c momentum. The kinematics of the electron through the recoil tracker is taken from truth information, meaning that it must not necessarily be deemed findable by the tracker. As such, the single track veto in section 5.1.2 can be finding a track that is not the recoil electron.

Figure 5.5 shows some of the BDT input variables based on containment radii, showing the total energy inside the radii around the electron, around the photon, or outside both

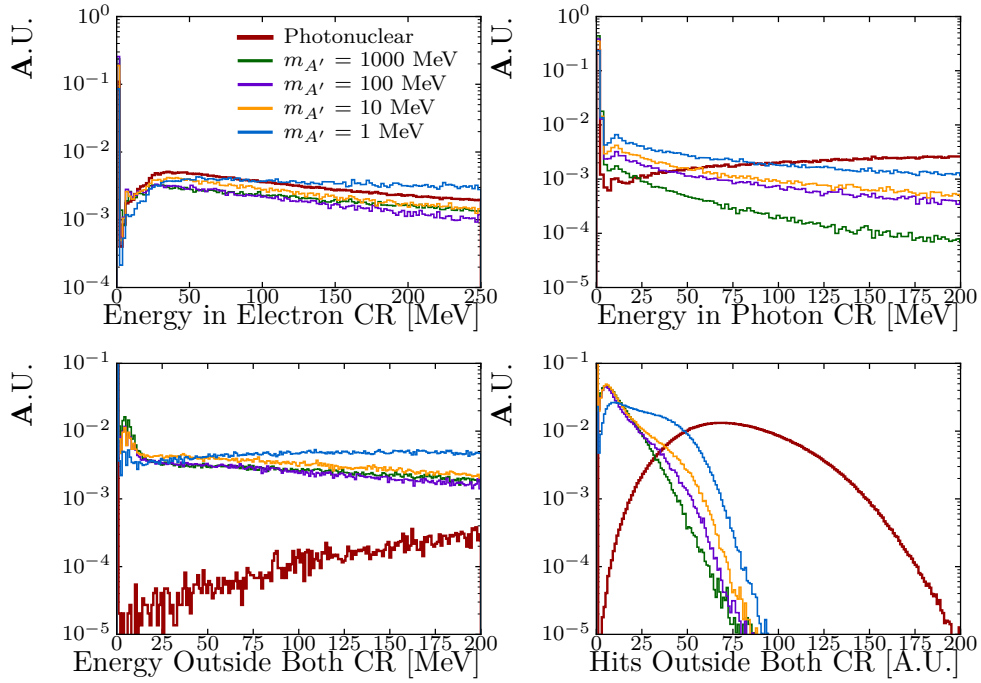


Figure 5.5: Selection of BDT variables calculated using the containment radii.

the radii around both the electron and the photon trajectories. The bottom right figure shows the number of hits outside both the electron and photon containment radii, which shows a significant separation between signal and background when there is more than around 100 hits. Variants of these, with the containment radii increased by an integer factor 2 to 5 are also used, as enumerated in appendix A. This set of BDT variables in total describes, coarsely, the transverse shower distributions around the electron and photon path.

A set of global input variables calculated on the whole ECal volume are additionally used, independent of the containment radii, all of which can be found in appendix A. A discussion about the "importance" of input variables will be covered in section 5.2.2. Figure 5.6 shows a selection of important variables.

The simplest global features are the total energy in the ECal (shown in the top left of figure 5.6), and the total energy in the back 14 layers, both of which have important discriminatory power. The most important input variable is the sum of hit energies, only considering ECal cells where there is no hits in neighbouring cells in the same layer, shown in the top right of the figure. This would intuitively remove hits from large showers, but keep small features further away from the large electron shower. This distribution shows a clear separation between signal and background using this variable, with close to zero signal events having a total isolated energy above 1000 MeV.

The average layer that hits occupy, in the bottom left of figure 5.6, shows that signal events seem to shower earlier in the ECal, than the deposits of photo-nuclear secondaries.

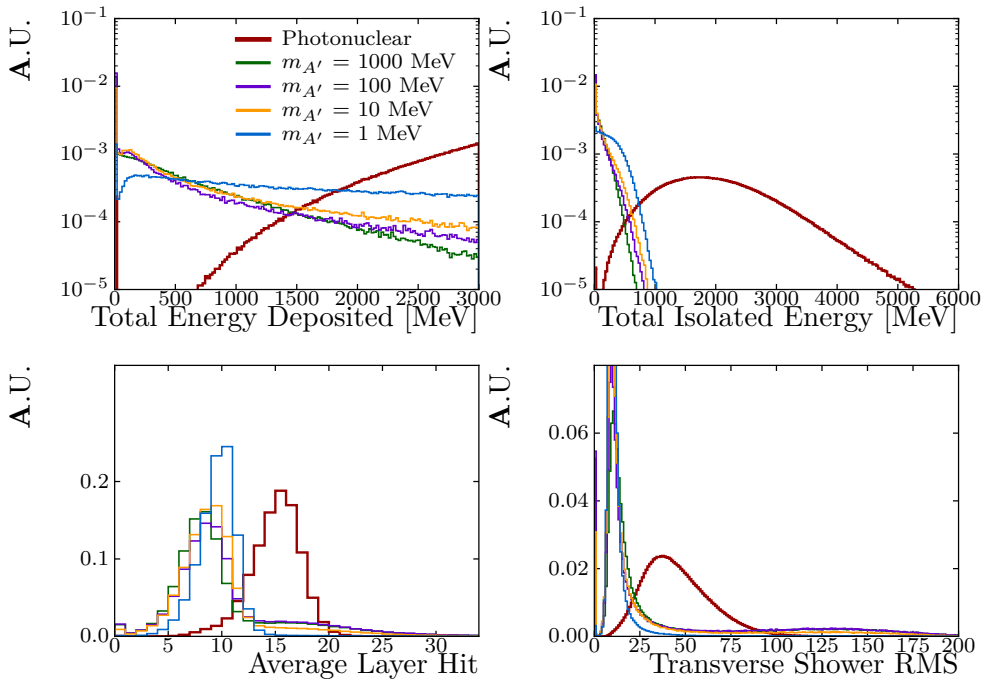


Figure 5.6: Selection of BDT variables not dependent on containment radii.

The transverse spread of hits, in the bottom right of the figure, also shows some discriminatory power, with photo-nuclear events depositing hits further away from the otherwise collimated electron showers, as is expected.

5.2.2 BDT Variable Importance

The *gain* of a split in a decision tree can be roughly defined as the change in prediction score that split contributes. For a certain feature A, the average gain of all decisions concerning the feature A, a measure of how much it effects the BDT score, is the *importance* of A. Figure 5.7 ranks the importances for all 42 input variables, and compares them to the BDT's importances in the 4 GeV study. It should be noted that this importance metric is inherent to the BDT, and not to one particular signal or background dataset. It is therefore not apparent, for example, how the importance ranking changes for different mediator masses. Explanations for the here abbreviated variable names are described in appendix A.

summedTightIso, the sum of isolated hit energies, is the most important variable for both BDTs, followed by the number of hits *nReadoutHits* or the energy in the back of the ECal *ecalBackEnergy*. Features about the longitudinal distribution of showers, the energy in the last 14 layers *ecalBackEnergy*, the standard deviation of the layer hits are in *stdLayerHit* and the deepest layer a hit is in*deepestLayerHit* are more important at 8 GeV, which is consistent with that longer showers and more activity in the back of the

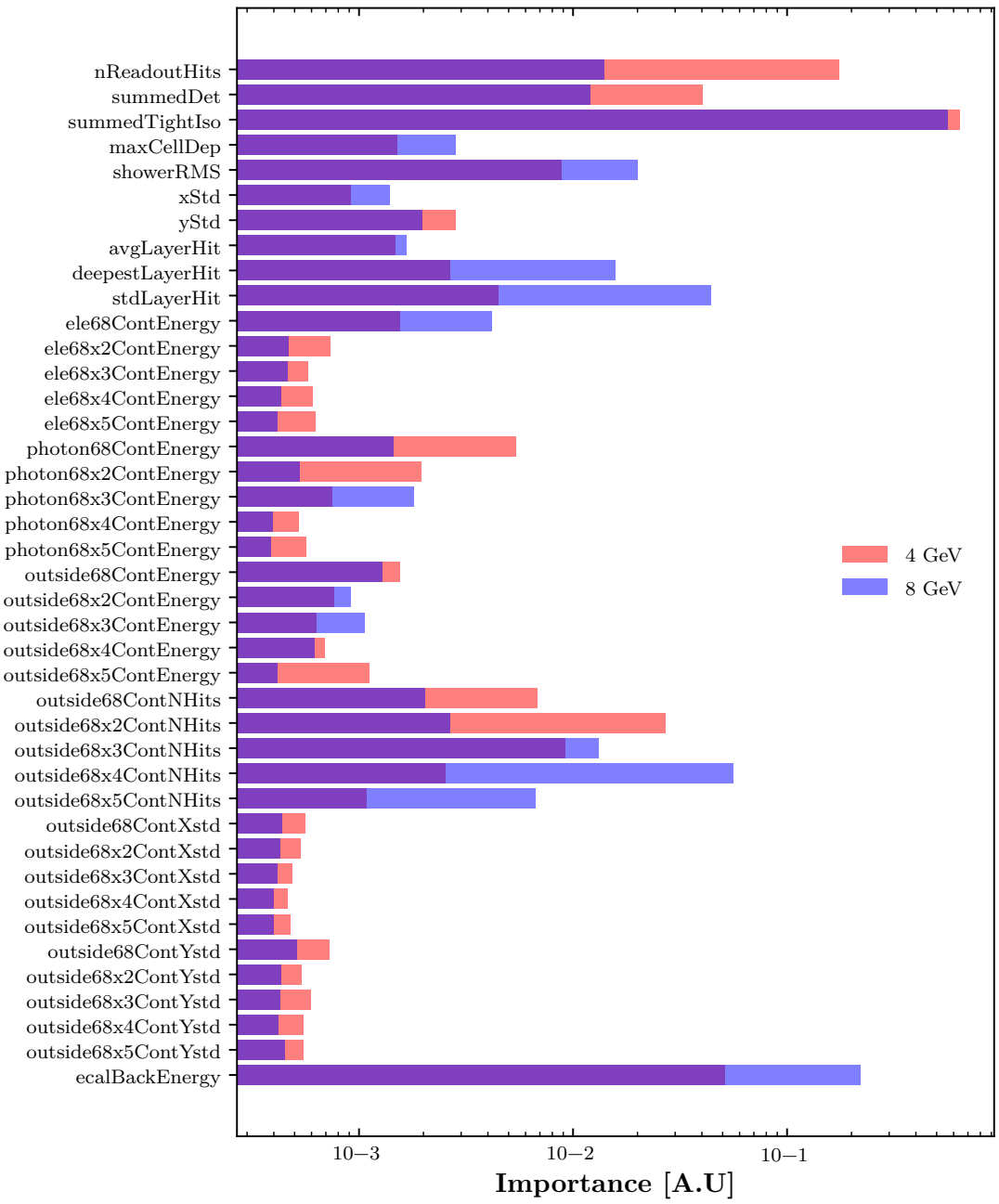


Figure 5.7: Importances of BDT input variables, normalized to one, for both the 4 GeV and new 8 GeV BDT. The importance is calculated as the change in BDT score a certain split in a tree contributes, averaged over all splits concerning a specific input variable. Definitions for variable names are given in appendix A.

ECal is expected at 8 GeV.

It is evident that many variables have low importance, which could make them targets for removal in order to assist the BDT training process. However, for two highly correlated input variables, boosted decision trees will only favor the importance of one, pushing down the importance of the other. Meaning that removal of input variables can not only be based on this importance metric.

5.2.3 BDT Performance

The BDT separates signal and background as in figure 5.8, showing a several order of magnitude separation near the, signal-like, BDT score of 1. The signal curves are similar across the A' mass spectrum.

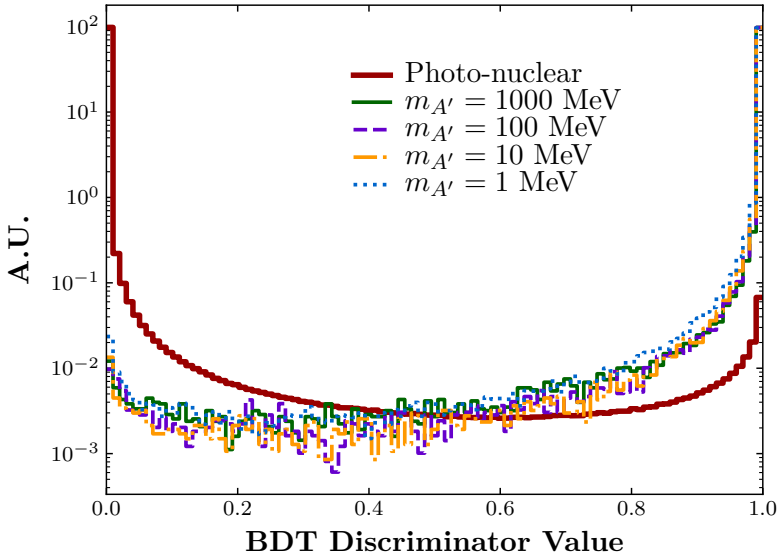


Figure 5.8: BDT score given to ECal photo-nuclear and signal events at 8 GeV, showing the separation between signal and background.

While the BDT assigns a score to each event, it does not propose a suitable cut on this value. This BDT score cut can be chosen arbitrarily, to achieve a certain signal and background rejection efficiency. Choosing a high BDT cut would result in high background rejection efficiency but also misclassifying more signal as background. In this study, a veto cut on the BDT score is chosen to preserve 70% of signal events across the whole mass spectrum, yielding a cut at 0.99989, which rejects about four orders of magnitude of ECal PN background events.

The relationship between signal and background rejection efficiency is seen in the Reciever Operating Characteristic (ROC) curve, seen in figure 5.9[4], made by varying the value of the BDT cut. An optimally trained BDT would approach the top left corner of this plot, while a completely untrained BDT that is simply guessing would approach

[4]4 GeV background simulations courtesy of Piotr Yartsev.

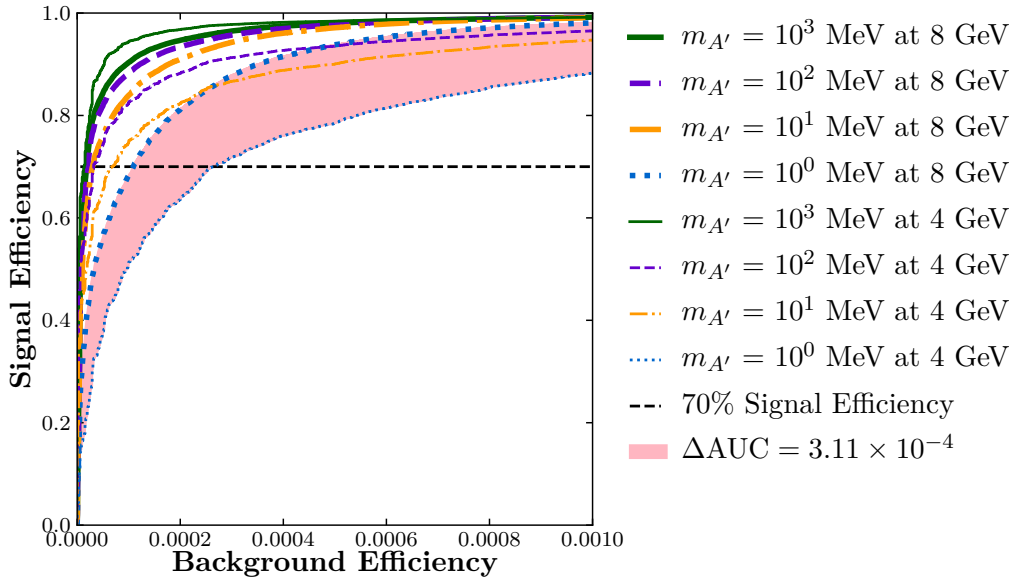


Figure 5.9: Receiver operator characteristic for the 8 GeV BDT, compared with the 4 GeV study's BDT, evaluated on different signal masses against the ECal photo-nuclear background.

a straight $y = x$ line. A measure of the performance is therefore the area under curve (AUC) of the ROC curve, which is 0.5 at worst and 1.0 at best.

Figure 5.9 shows that for 70% signal efficiencies, the background is rejected by about four orders of magnitude. The curves at higher mediator masses tend to perform better than lower masses, with $m_{A'} = 1$ MeV being the worst performing and the one which the BDT score cut has to be chosen for. In the figure the ROC curves in the 4 GeV study are given as well, where they are outperformed expect for at $m_{A'} = 1000$ MeV, where the 4 GeV BDT is better. However, the largest improvement of going to 8 GeV is seen in the $m_{A'} = 1$ MeV curve, which is what ultimately affects the BDT cut and the background rejection efficiency.

Because the cut on the BDT value is based on the worst-performing mediator mass to achieve the desired signal efficiency across all masses, the 1 MeV mass in this case, the improvements in background rejection capability is best seen in how the 1 MeV ROC curve changes. The pink area in figure 5.9 shows the difference in AUC in the 1 MeV mass between 4 GeV and 8 GeV beam energy, with an area of 3.1×10^{-4}. For a signal efficiency of 70% there is roughly a factor two in background rejection between the two curves 1 MeV curves, but the ratio would greatly increase if instead the BDT cut was based on e.g. a 90% signal efficiency.

A table of the ROC curve's AUCs is given in table 5.1. The AUC of the blue $m_{A'} = 1$ MeV curve at 8 GeV outperforms the yellow $m_{A'} = 10$ MeV and purple $m_{A'} = 100$ MeV curve at 4 GeV, but figure 5.9 show that they actually intersect. At lower signal efficiencies the 4 GeV curves have better background rejection but above roughly 80% and 90% signal efficiency respectively the 8 GeV curve is more efficient, showing that the

curves are not symmetric around the $y = 1 - x$ diagonal.

	$m_{A'} = 1$ MeV	$m_{A'} = 10$ MeV	$m_{A'} = 100$ MeV	$m_{A'} = 1000$ MeV
4 GeV	46.6×10^{-5}	22.0×10^{-5}	15.5×10^{-5}	4.7×10^{-5}
8 GeV	15.4×10^{-5}	7.5×10^{-5}	6.1×10^{-5}	5.6×10^{-5}

Table 5.1: 1-AUC for ROC curves at different mediator masses, comparing the 4 and 8 GeV beam energy BDTs.

5.2.4 BDT Score and HCal Activity

The BDT and HCal together have complementary background rejection power. Figure 5.10 shows the maximum photo-electrons in all HCal hits, together with the BDT score for ECal photo-nuclear events. Many events with BDT scores above the cut are rejected by the several hundred PEs in the HCal. One ECal photo-nuclear event passes into the signal region in the top left corner, more clearly shown in the inset, but all other events deemed signal-like by the BDT have significant activity in the HCal. Beneath the scatter plot, the distribution for $m_{A'} = 1$ GeV signal events is shown, where they are densely contained in the top-left signal region of the figure.

The similar figure 5.11 shows instead ECal muon conversion events. No such events pass into the signal region, but some have BDT scores close to the cut and low HCal activity, as seen in the inset. With several events being so signal-like, the ECal muon conversion background can not be regarded as a trivial background that is easily rejected by a BDT trained on ECal PN events.

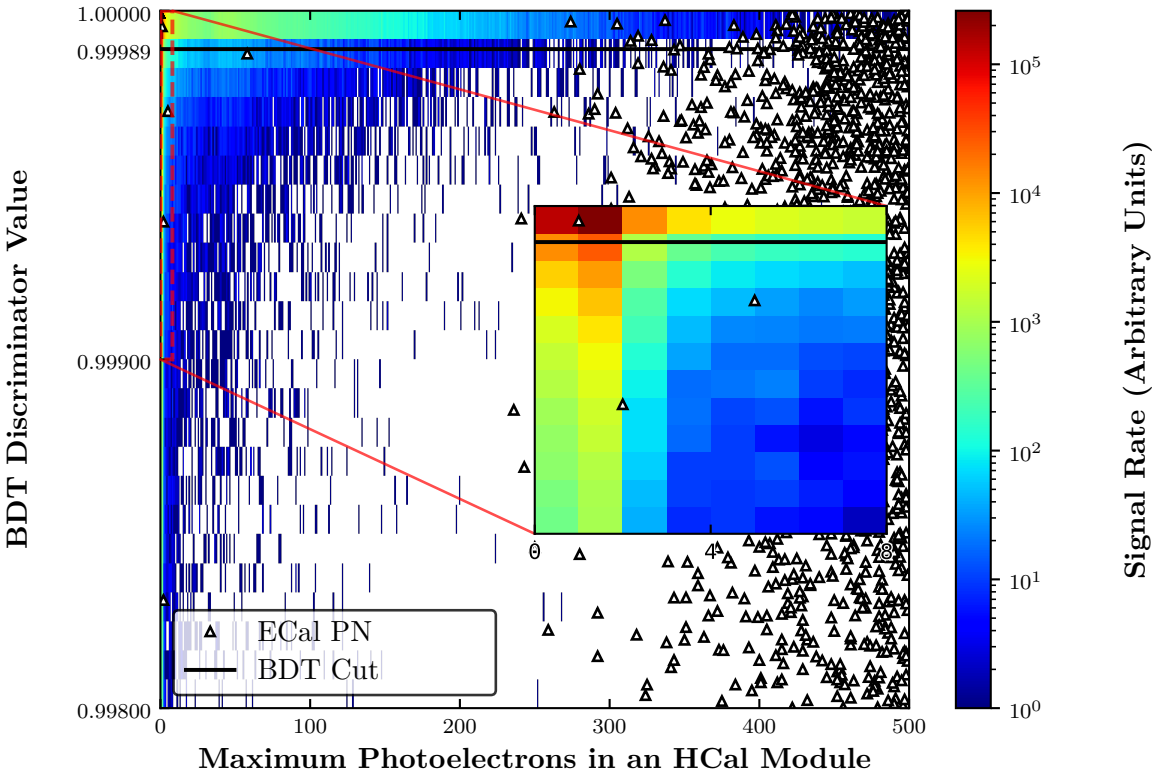


Figure 5.10: BDT Score and maximum HCal photo-electrons in ECal photo-nuclear events. The distribution of $m_{A'} = 1$ GeV signal events is shown underneath, most densely in the very top-left corner of the signal region.

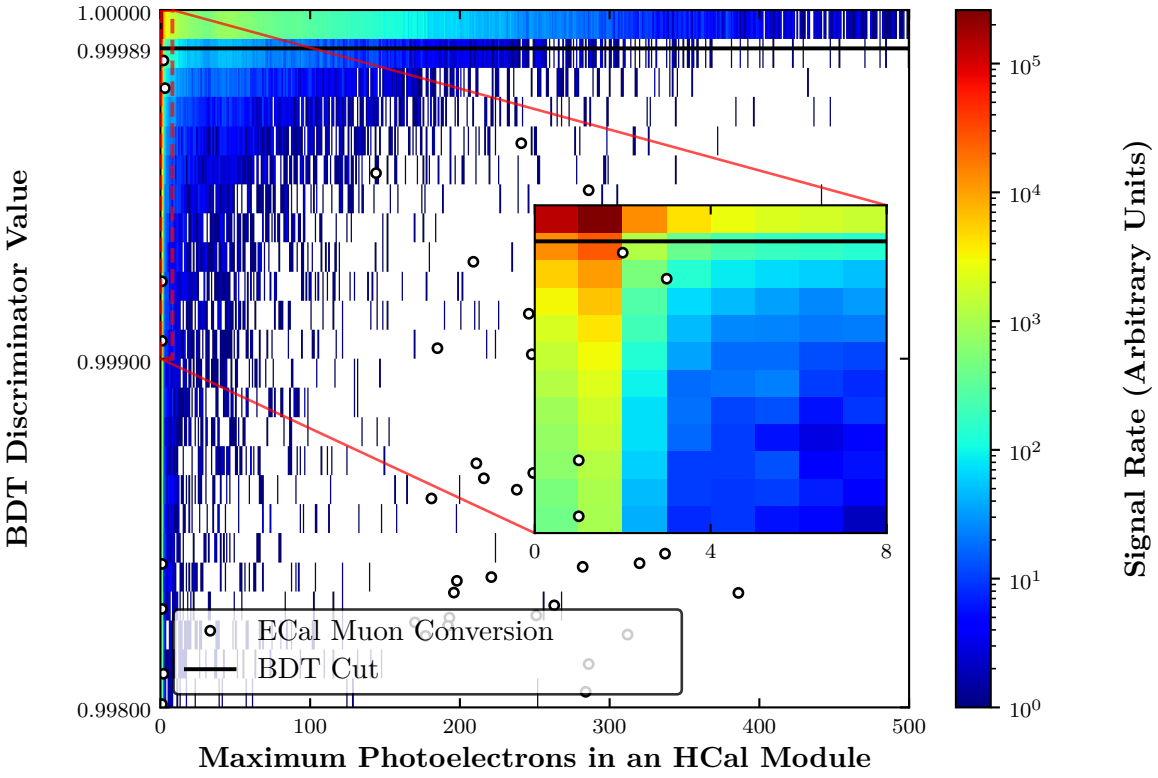


Figure 5.11: BDT Score and maximum HCal photo-electrons in ECal muon conversion events. The distribution of $m_{A'} = 1$ GeV signal events is shown underneath.

5.3 MIP Tracking

5.3.1 ECal Tracks

Some PN events deemed very signal-like by the BDT have similar topology. These are events with photon-nuclear production of a charged kaon K^+K^- pair, with one energetic, *hard*, kaon and one not so energetic, *soft*, that is easily lost in the ECal. If the hard kaon escapes the ECal and is registered in the HCal there is no problem in vetoing such events. However, the charged kaon has a lifetime of $\tau = 12.4$ ns [17], and it can therefore decay while still inside the ECal, losing a lot of energy into some neutrino final state, e.g. $K^+ \to \mu^+\nu_\mu$. These events would register as having missing energy, but leave a short track in the ECal, a geometric feature with a lot of fidelity that the BDT is not well equipped to handle.

Although the 8 GeV BDT does veto these rare signal-like K^+K^- background events, they pass close to the signal region. Originally at 4 GeV, an algorithm to find short tracks from minimally ionizing particles was successfully introduced to veto the mentioned kaon background. This algorithm will here be optimized to reject ECal PN events at 8 GeV. As signal events expect no MIP tracks, any event with one or more tracks found will be immediately rejected.

Although the ECal has high granularity, if the electron and bremsstrahlung photon are very collimated, there is no possibility to distinguish the electron and photon shower in the calorimeter, nor seeing tracks of photo-nuclear daughters close to the electron shower. Even more often at 8 GeV, the kinematics will challenge the granularity of the ECal.

5.3.2 Tracking Algorithm

Not all ECal hits are considered in the tracking. Normally, all hits within the electron shower containment radii are discarded in order to not, very easily so, find any features within the electron shower. If the electron and bremsstrahlung photon are very colinear however, no hits will be discarded. If the electron and photon trajectories have an angular separation of less than 14° and enter the ECal with a separation less than 17.2 mm, then the trajectories are considered colinear. Any photon induced features in such an event will likely be overlapping with the electron shower. In practice, the following algorithm will almost always find a fake track within the electron shower, making the colinearity condition almost a veto cut in itself.

The MIP tracking algorithm is split into two independent steps. In the first, straight tracks perpendicular to the face of the ECal are searched for. For a minimum number of layers, m_{stl}, if there is a hit in the same ECal cell in several consecutive layers, it will be considered a track. m_{stl} is a free parameter of the algorithm. This is depicted in the left cluster of hits in figure 5.12. Any hits in a track will be removed from further iterations of the algorithm, which continues until no more straight tracks can be found.

A second pass is done on hits unaffected by the first straight track finding procedure, to find tracks with an angle to the ECal face by a linear fit. Groups of three hits will be considered, that should be line-like by a least-squares fit. Not all triplets of hits are considered, as there are n hits this would have a $\mathcal{O}(n^3)$ complexity. Instead, tracks are seeded by first finding an isolated hit, with no neighbouring hits in the same layer.

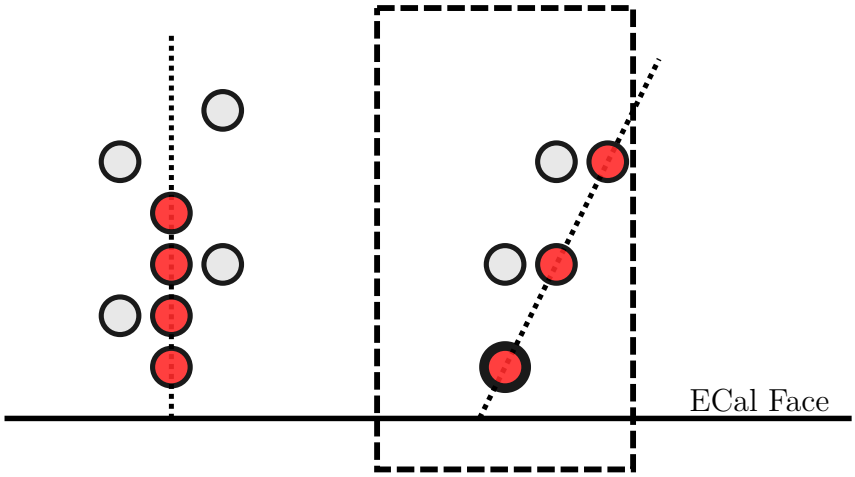


Figure 5.12: Schematic view of the MIP tracking algorithm on two clusters of hits in the ECal, near the ECal (downstream) face. The first pass of the algorithm finds a straight track, perpendicular to the ECal face in the left cluster, marked in red. The second pass finds an isolated track seed, marked with a thick border, and searches for tracks in a cylindrical volume around the seed. A linear regression between a triplet of hits finds a track, marked in red.

Around the track seed, parallel to the beamline, a cylinder with radii 17.4 mm is formed, discarding all hits outside the cylinder and therefore forces tracks to be vaguely in the forward direction, as in the right cluster of hits in figure 5.12.

A linear regression through a simple least squared procedure is done on the track seed and all other pairs of hits inside the cylinder. If the correlation coefficient r of the fit is lower than the free parameter r_{min}, the fit is linear enough to promote the triplet of hits to a track candidate. I.e. if the three hits do not lie on a straight line, the algorithm will not consider it a track.

A track candidate has one final condition to become a track, which is that it should be near the bremsstrahlung photon trajectory, but far from the electron trajectory. 8.7 mm for the former and 1.5×8.7 mm for the latter. If this is satisfied a track is found, and the three hits are removed from further repetitions of this algorithm. The linear regression phase is repeated until all such tracks have been found.

5.3.3 Tracking Efficiencies

The signal efficiency of lower A' masses is reduced most severely by the MIP tracking. Therefore, the two free parameters of the tracking, m_{stl} and r_{min}, are optimized to improve specifically the 1 MeV mediator mass efficiency. In such a way that, if S 1 MeV signal events and B ECal PN background events pass the MIP tracking, the free parameters maximize S/B. As very few background events in the training dataset passes the BDT veto, S/B is optimized on events that pass the trigger veto but the other previously discussed vetos are not considered. Figure 5.13 shows the separation between $m_{A'} = 1$ MeV signal and the ECal photo-nuclear background for events passing the trigger, with signal events predominantly having fewer tracks than background events.

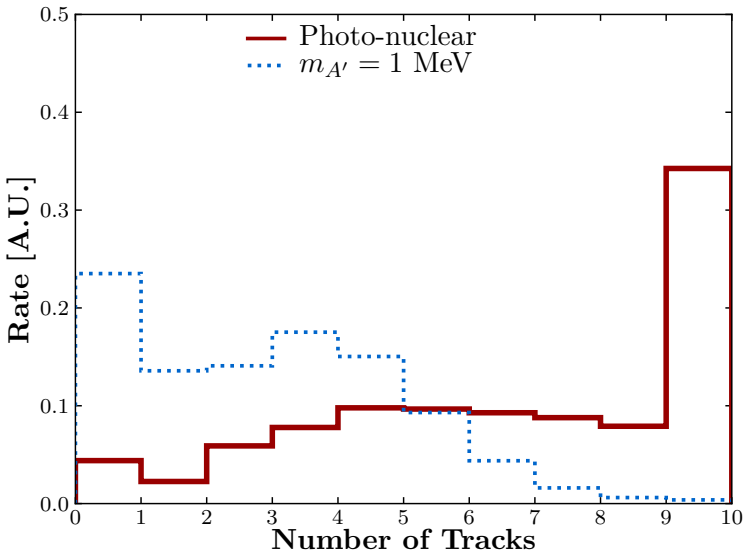


Figure 5.13: Number of tracks in $m_{A'} = 1$ MeV events and in the ECal photo-nuclear background, for events passing the trigger. Only events with zero tracks pass the MIP tracking veto, yielding this separation between signal and background when optimizing the free parameters of the tracking.

Evaluating the signal efficiency of adding the MIP tracking algorithm should be done in conjunction with the other veto steps. Figure 5.14 shows the distribution of counted MIP tracks in signal events that pass the trigger, recoil tracker, BDT and HCal vetos, where ideally there should ideally be no MIP tracks found. While the signal efficiency is above 60% for the 10, 100 and 1000 MeV mediator masses, it suffers greatly at 1 MeV, going down to 28%. Notably, the MIP tracking is the veto step where the signal efficiency is impacted most negatively, even more than all other veto steps together for $m_{A'} = 1$ MeV.

It is natural for the low mediator masses to perform worse in the MIP tracking, due to their kinematics more often leading to colinear electron and photon trajectories. The granularity of the ECal is therefore the limiting factor in the low mass region.

5.4 Results

5.4.1 Cut Flow

The rejection power of the cuts on the ECal PN background may now be evaluated, together with seeing whether it generalizes for the other backgrounds the cuts were not specifically optimized for. Applying the cuts sequentially, the reduction of the ECal PN background is shown in the *cut flow* table 5.2.

Table 5.2 shows that the zero-background goal is reached for all simulated background samples, equivalent to almost 2×10^{14} EoT on all classes of background. The zero-

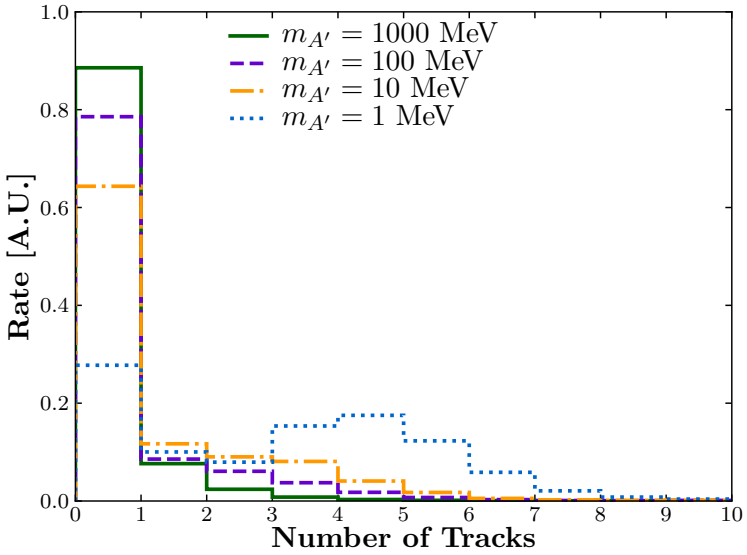


Figure 5.14: Number of MIP tracks in signal events passing all other cuts. As events with one or more tracks are vetoed, the zero tracks bin shows the signal efficiency, from 90% at best to 30% at worst depending on the mediator mass.

	Photo-nuclear		**Muon conversion**	
	Target-area	**ECal**	**Target-area**	**ECal**
EoT Equivalent	1.76×10^{14}	1.97×10^{14}	1.08×10^{16}	2.00×10^{15}
$<$ 3160 MeV Trigger	3.41×10^{8}	4.47×10^{8}	1.12×10^{9}	8.12×10^{8}
One track, p $<$ 2400 MeV/c	2.97×10^{7}	4.15×10^{8}	3.31×10^{6}	7.57×10^{8}
ECal BDT ($>$ 0.99989...)	3.47×10^{5}	2.35×10^{4}	< 1	< 1
HCal max PE < 8	< 1	1.00	< 1	< 1
ECal MIP tracks $= 0$	< 1	< 1	< 1	< 1

Table 5.2: *Cut flow* table of the four backgrounds of interest at 8 GeV beam energy, with all veto cuts applied in sequence. The EoT equivalent size of the data sets are given, with all being around 2×10^{14} or larger. Notably after the complete veto sequence, no simulated background event passes all cuts.

	Photo-nuclear		**Muon conversion**	
	Target-area	**ECal**	**Target-area**	**ECal**
EoT Equivalent	4×10^{14}	2.1×10^{14}	8.2×10^{14}	2.4×10^{15}
$<$ 1500 MeV Trigger	1×10^{8}	2.63×10^{8}	1.6×10^{7}	1.6×10^{8}
One track, p $<$ 1200 MeV/c	2×10^{7}	2.34×10^{8}	3.1×10^{4}	1.5×10^{8}
ECal BDT (> 0.99)	9.4×10^{5}	1.32×10^{5}	< 1	< 1
HCal max PE < 4	< 1	10	< 1	< 1
ECal MIP tracks $= 0$	< 1	< 1	< 1	< 1

Table 5.3: Cut flow table from the study at 4 GeV: [2]

background goal is therefore reached with the same statistics as in the 4 GeV beam energy study.

The track veto effectively cuts the target-area $\gamma \to \mu\mu$ background by three orders of magnitude. The target-area photo-nuclear background in comparison decreases by just an order of magnitude, with the reactions daughter particles being more difficult to track than muons.

The BDT rejects four orders of magnitude for the ECal PN background it was optimized for. The background with charged kaons decaying inside the ECal, discussed in the context of MIP tracking, is completely removed after the BDT and HCal cuts. The BDT also generalizes to reject more than 10^6 target $\gamma \to \mu\mu$ events and 10^8 ECal $\gamma \to \mu\mu$ events, completely rejecting the simulated muon conversion events.

One single event, in the ECal photo-nuclear background, is deemed signal-like by both the BDT and HCal veto. The 5.8 GeV bremsstrahlung photon in this event produces a hard 5.6 GeV K_L that does not leave any clear features in the ECal. With the K_L being a neutral hadron that the HCal is responsible for vetoing, the HCal's efficiency to veto neutral kaons is crucial to veto such a rare background (as would be a similar event with a neutron instead). While there is no MIP track for the tracking to discover, the event is vetoed due to the angle between the recoil electron and the bremsstrahlung photon being small, as discussed in section 5.3. For every veto step except the last the event is seemingly completely signal-like.

Beam Energy Comparison

The cut flow in the 4 GeV study is given in table 5.3, which is similar to the 8 GeV cut flow. Although both tables reach the zero-background goal, the cuts are differently strict in the two studies, making absolute numbers difficult to compare between the two. Importantly, the cutflow table suggests an order of magnitude improvement in the BDT cut between 4 and 8 GeV beam energy, while part of the improvement stems from that a higher signal efficiency was required at 4 GeV.

It is overall not straightforward to compare the BDTs trained at the two different beam energies, as the results do not disentangle a performance boost due to a more well-trained BDT from that the underlying physics allowing for better discrimination between background and signal. Nevertheless, the red area in figure 5.9 shows the difference between the worst performing mediator mass, namely the 1 MeV, between 4 and 8 GeV

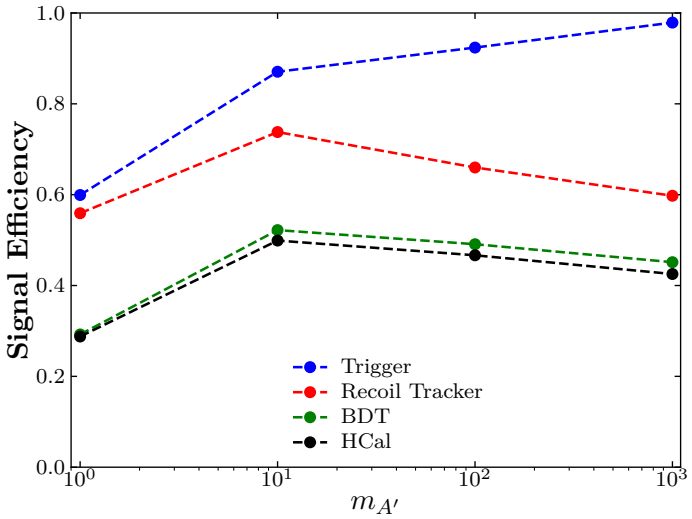


Figure 5.15: Fraction of signal events passing a sequential series of veto cuts, in the order used in the cutflow table 5.2.

beam energy. As the BDT score cut is chosen to preserve a certain signal efficiency across the whole A' mass spectrum, the background rejection is strictly dependent on the $m_{A'} = 1$ MeV performance. For a 70% signal efficiency, there is roughly a factor 2 between the two blue $m_{A'} = 1$ MeV curves, while if higher signal efficiencies were desired this ratio is even larger.

The $m_{A'} = 1$ GeV curves in figure 5.9 shows a slight performance decrease. While the performance in the cut-flow table will not be effected by this, as has been discussed, it does mean a few percent decrease in signal efficiency. This is acceptable and much smaller than e.g. the ∼20% loss from the following MIP tracking. Furthermore, figure 4.2 shows a large yield enhancement for this mediator mass that would make a few percent signal inefficiency inconsequential. The $m_{A'} = 1$, 10, 100 MeV curves at 8 GeV are above their 4 GeV counterparts in the ROC curves, and therefore with higher signal efficiencies.

5.4.2 Signal Efficiency

After a veto process that reduces the background by 14 orders of magnitude, one should not have vetoed all the signal events in the process. Figure 5.15 shows the signal efficiency, the fraction of signal events remaining with the veto steps applied sequentially. Around 30% to 40% of signal events are kept across the whole $m_{A'}$ spectrum.

The veto step that by far has the most impact on cutting the signal efficiency is the MIP tracking, as seen separately in figure 5.14 for events passing all previous veto steps. For the 1 MeV mediator less than 30% of signal events are preserved, as impactful as the other veto steps together. The effect of the MIP tracking is less severe for higher mediator masses, preserving 60% to 80% of signal events.

While the low $m_{A'}$ range has the worst signal efficiency, the difference in rate of signal production in absolute events should be considered. At 4 GeV beam energy, the inclusive

cross section for A' production is several orders of magnitude higher at $m_{A'} = 1$ MeV than at $m_{A'} = 1000$ MeV. Assuming that it is similar at 8 GeV beam energy, it is apparent that improvement in signal efficiency on the higher end of the mass spectrum is also important.

Chapter 6

Outlook

6.1 Summary of Results

The cutflow table 5.2 confirms that the zero-background goal was reached with roughly 2×10^{14} EoT at 8 GeV, as was the case at 4 GeV. For low mediator masses the BDT performs better than the previous 4 GeV BDT, which ultimately gives better background rejection efficiency for a chosen BDT score cut. This is achieved with a 30% to 50% signal efficiency across the A' mass spectrum, up until the HCal veto.

One photo-nuclear K_L event passed all veto steps up until the MIP tracking, where it was rejected due to the electron and inferred bremsstrahlung photon being colinear. Further studies on simulated kaon events and the kaon inefficiencies in the HCal should be conducted to better understand this rare background.

The BDTs in the 4 and this 8 GeV study are trained with different distributions of signal mediator masses, meaning that they are not perfectly comparable, and what improvement comes from difference in training or difference in underlying physics is not disentangled. The improvement for low $m_{A'}$ at 8 GeV and worse performance at $m_{A'} =$ 1 GeV may be due to the training samples being weighted towards the low masses. Even with similar training data sets however, the stochastic nature of training BDTs would still hinder making direct connections with the improvements in results to the physics. Instead, more rigorous studies on understanding the impact of BDT input variables, and their correlations, is needed beyond the simple importance ranking in figure 5.7.

6.2 Further Studies

To only simulate one incoming electron at a time is an unrealistic simplification. In reality, each event would have several, Poisson distributed, electrons. This would introduce challenges to all detector subsystems. The trigger scintillator and tracker must recognize the number of incoming electrons, where each possibility would need a unique missing energy trigger condition. Furthermore, the ECal will have more difficulties in separating showers and finer features like PN interactions, when several electrons shower simultaneously. Although more difficult, a future multi-electron background rejection study will be more representative of the conditions where LDMX will operate. The positive reason

for running with more electrons is to increase the EoT, increasing the number of possible DM production interactions.

Speculation about the performance of the background rejection sequence is done on a small number of events on the rate of 10^{-13} per beam electron, for which the absolute number of such events is subject to large fluctuations by the "law of small numbers". For a study that expects 2×10^{14} EoT it would be desirable to simulate more events than this, maybe on the 2×10^{15} scale, to better understand the veto capabilities.

6.2.1 BDT Improvements

The current set of BDT input variables is not complete and could be further improved. One natural extension would be introduce more variables that captures the longitudinal distribution of energy throughout the ECal, similar to how the current containment radii try to capture the transversal distribution of energy. Furthermore, not only simple ECal features can be given as BDT inputs, but also the results from the MIP tracking, HCal activity and recoil tracker measurements. Such a BDT could potentially find correlations that won't be found with veto cuts on different detector subsystems that work completely independently.

The variable importances in figure 5.7 shows that many ECal features do not significantly contribute to the discrimination. *Pruning* input features to include only a subset of important variables may be beneficial for training the BDT. As correlations between variables is not reflected in this importance metric, and what supporting role the low-importance variables have, pruning variables should be made with more consideration than just the importance.

BDTs are widely used and are comparably easy to interpret compared to many other statistical learning methods. It may be worthwhile to explore other machine learning methods that utilize the full granularity of the ECal, instead of aggregating them into a small set of BDT input variables. Possible, more geometrically aware, methods include convolutional neural networks [18] or graph-based neural networks [19] that have been successfully applied in calorimetry before. This could especially be interesting for finding smaller and finer details in the ECal hits, such as MIP tracks, that the current BDT is not equipped to handle.

6.2.2 Tracking Improvements

The somewhat heuristic MIP tracking algorithm may be considered too tailored to the problem, and may be too constrained to find more of the diverse possible tracks that MIPs can leave in the ECal. A more general tracking algorithm could be considered, either one among many existing tracking algorithm that exist or possibly with some statistical learning based approach. Alternatives for the tracking algorithm should also be considered in view of the signal efficiencies, that takes a strong hit with the current MIP tracking algorithm.

The current MIP tracking algorithm is optimized using events that pass the trigger, meaning that most such events will be vetoed before the MIP tracking is applied. Instead, if a larger sample with only the rare backgrounds of interest for the tracking was produced,

the free parameters of the algorithm could be optimized on the events it will actually be applied on. As the current ECal PN simulation only yielded a handful of such events, some computational tricks would have to be introduced to feasibly generate more of these very rare background events.

Bibliography

[1] L. K. Bryngemark *et al.*, "Building a distributed computing system for LDMX," *EPJ Web of Conferences*, vol. 251, C. Biscarat *et al.*, Eds., p. 02 038, 2021. DOI: 10.1051/epjconf/202125102038. [Online]. Available: https://doi.org/10.1051%2Fepjconf%2F202125102038.

[2] T. Åkesson *et al.*, *A High Efficiency Photon Veto for the Light Dark Matter eXperiment*, Journal of High Energy Physics, vol. 4, 2020. DOI: 10.1007/JHEP04(2020)003. arXiv: 1912.05535 [physics.ins-det].

[3] U. of Chicago, *Annual Register*. 1893-1930., 1896, LCCN: 08011821.

[4] N. Jarosik *et al.*, "SEVEN-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: SKY MAPS, SYSTEMATIC ERRORS, AND BASIC RESULTS," *The Astrophysical Journal Supplement Series*, vol. 192, no. 2, p. 14, Jan. 2011. DOI: 10.1088/0067-0049/192/2/14.

[5] J. Alexander *et al.*, *Dark sectors 2016 workshop: Community report*, 2016. DOI: 10.48550/ARXIV.1608.08632. [Online]. Available: https://arxiv.org/abs/1608.08632.

[6] T. Åkesson *et al.*, *Light Dark Matter eXperiment (LDMX)*, Aug. 2018. DOI: 10.48550/ARXIV.1808.05219. arXiv: 1808.05219 [hep-ex].

[7] E. Kolb and M. Turner, *The Early Universe*, ser. Frontiers in physics. Addison-Wesley, 1990, ISBN: 9780201116038.

[8] D. Banerjee *et al.*, "Dark matter search in missing energy events with NA64," *Physical Review Letters*, vol. 123, no. 12, Sep. 2019. DOI: 10.1103/physrevlett.123.121801.

[9] T. Åkesson *et al.*, "Current Status and Future Prospects for the Light Dark Matter eXperiment," in *2022 Snowmass Summer Study*, Mar. 2022. DOI: 10.48550/ARXIV.2203.08192. arXiv: 2203.08192 [hep-ex].

[10] J. Stohr, "Linac coherent light source ii (lcls-ii) conceptual design report," Nov. 2011, Report number: SLAC-R-978. DOI: 10.2172/1029479.

[11] T. Raubenheimer *et al.*, "DASEL: Dark Sector Experiments at LCLS-II," Jan. 2018, Report number: SLAC-PUB-17225. DOI: 10.48550/ARXIV.1801.07867. arXiv: 1801.07867 [physics.acc-ph].

[12] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. DOI: 10.1016/S0168-9002(03)01368-8.

[13] Geant4 Collaboration, *Geant4 Physics Reference Manual 10.4*, Release 11.0, Rev6.0: December 10th, 2021. [Online]. Available: https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsReferenceManual/fo/PhysicsReferenceManual.pdf.

[14] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Annals of Statistics*, vol. 29, pp. 1189–1232, 2001. DOI: 10.2307/2699986.

[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, San Francisco, California, USA: ACM, 2016, pp. 785–794, ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939785.

[16] "The Phase-2 Upgrade of the CMS Endcap Calorimeter," CERN, Geneva, Tech. Rep., Nov. 2017. DOI: 10.17181/CERN.IV8M.1JY2. [Online]. Available: http://cds.cern.ch/record/2293646.

[17] P. Zyla *et al.*, "Review of Particle Physics," *PTEP*, vol. 2020, no. 8, p. 083C01, 2020. DOI: 10.1093/ptep/ptaa104.

[18] Östman, Leo, *Imaging Using Machine Learning for the LDMX Electromagnetic Calorimeter*, LUP Student Papers, 2020.

[19] H. Qu and L. Gouskos, "Jet tagging via particle clouds," *Physical Review D*, vol. 101, no. 5, Mar. 2020. DOI: 10.1103/physrevd.101.056019. [Online]. Available: https://doi.org/10.1103%2Fphysrevd.101.056019.

[20] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007. DOI: 10.1109/MCSE.2007.55.

[21] J. Pivarski *et al.*, *Scikit-hep/uproot4: 4.1.9*, version 4.1.9, Dec. 2021. DOI: 10.5281/zenodo.5767911. [Online]. Available: https://doi.org/10.5281/zenodo.5767911.

[22] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. DOI: 10.1038/s41586-020-2649-2. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2.

[23] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011, ISSN: 1532-4435.

[24] R. Brun and F. Rademakers, "Root — an object oriented data analysis framework," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1, pp. 81–86, 1997, New Computing Techniques in Physics Research V, ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00048-X.

Software Acknowledgements The Python side of this analysis made frequent use of the plotting library matplotlib [20], the ROOT-file I/O library uproot [21], and for numerical computations both numpy [22] and scikit-learn [23].The training and evaluation of boosted decision trees was done with the XGBoost [15] Python library.

Data serialization and offline analysis was performed using the ROOT [24] framework. Detector simulations were made with Geant4 [12] and version 1.7.1 of the LDMX offline software ldmx-sw courtesy of the LDMX collaboration.

Appendix A

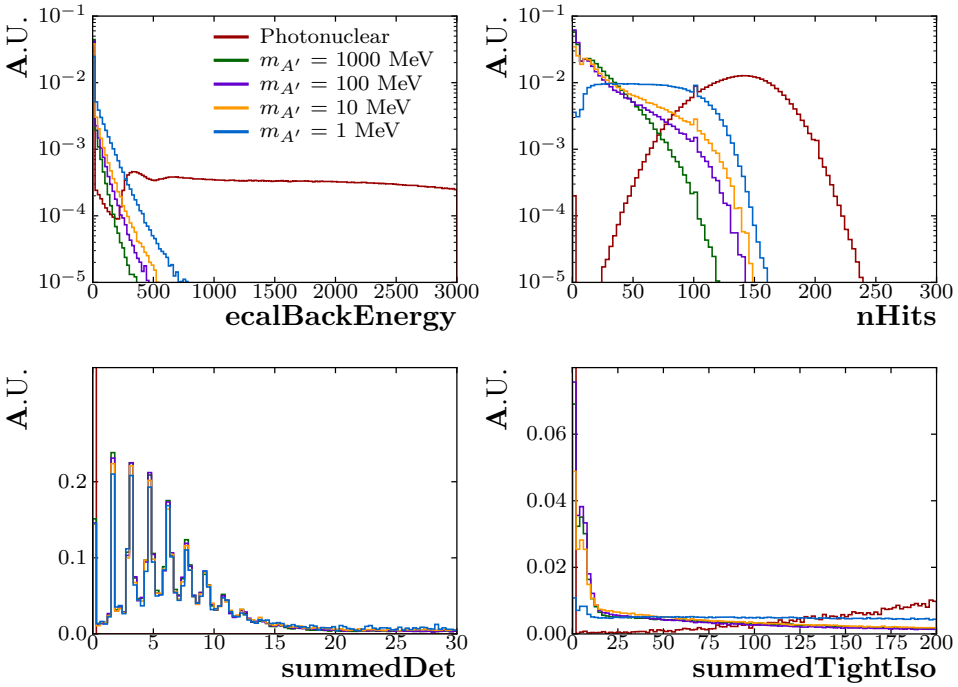
BDT Variables

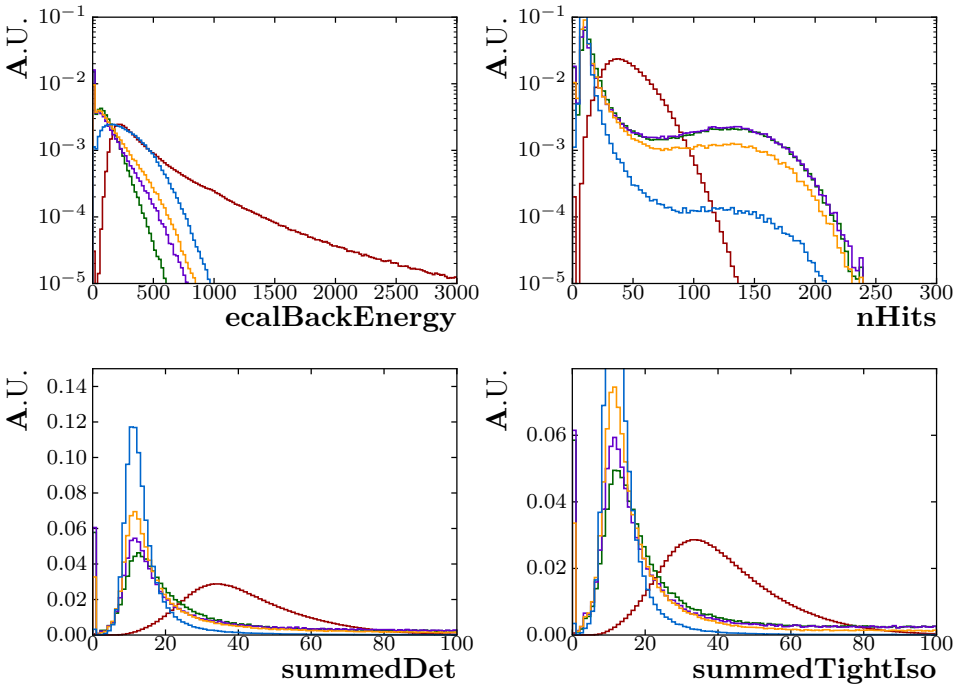
Variable Definitions

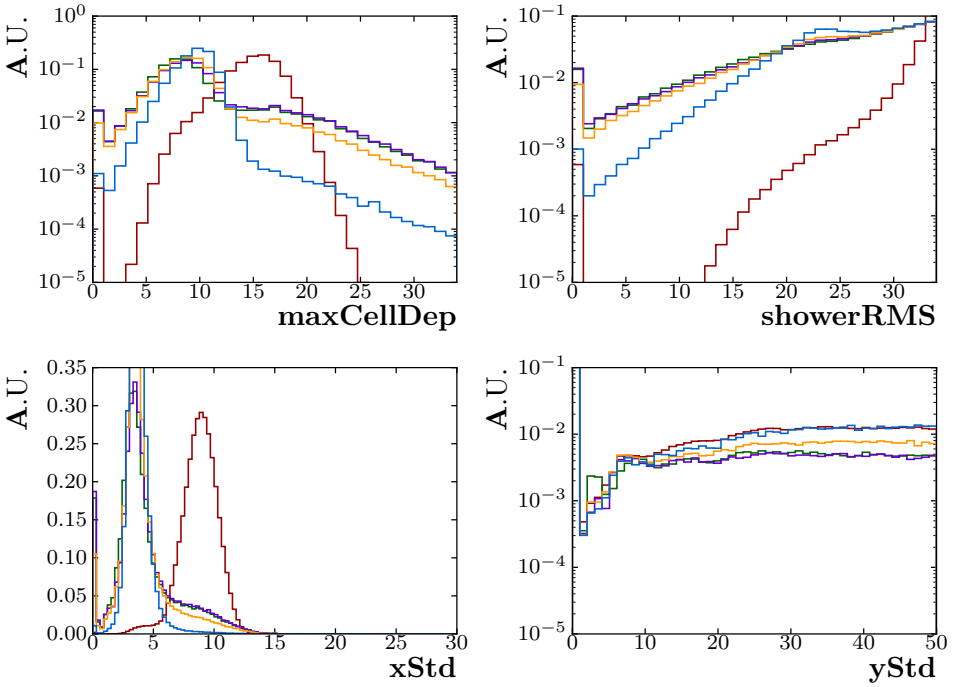
Name	Description
ecalBackEnergy	Total energy in ECal layers 20 and deeper.
nHits	Total number of ECal cells with an energy deposit.
summedDet	Total energy in ECal.
summedTightIso	Total energy of *isolated* ECal hits, where no neighbouring ECal cell has fired.
maxCellDep	Energy of cell with largest energy deposit.
showerRMS	The *centroid* is the ECal cell closest to the center of energy for all hits. showerRMS is the root mean square of the value $E \cdot R$, for all hits with energy E and euclidean distance R from the centroid.
xStd	Standard deviation of hit X-positions.
yStd	Standard deviation of hit Y-positions.
avgLayerHit	Average layer index (from 0 to 33) of all hits.
deepestLayerHit	Deepest ECal layer with a hit.
stdLayerHit	Standard deviation of the layer index for all hits.
ele68ContEnergy	Sum of energy contained within one containment radii around the electron trajectory.
ele68x2ContEnergy	As above, but within two times the radii.
ele68x3ContEnergy	As above, but within three times the radii.
ele68x4ContEnergy	As above, but within four times the radii.
ele68x5ContEnergy	As above, but within five times the radii.
photon68ContEnergy	Sum of energy contained within one containment radii around the photon trajectory.
photon68x2ContEnergy	As above, but within two times the radii.

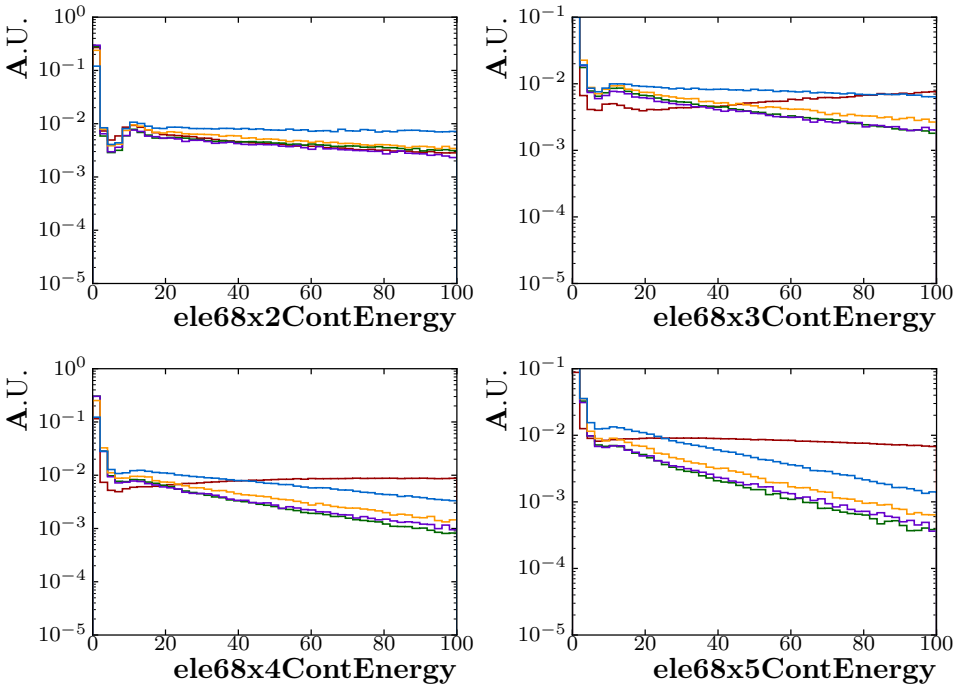
photon68x3ContEnergy	As above, but within three times the radii.
photon68x4ContEnergy	As above, but within four times the radii.
photon68x5ContEnergy	As above, but within fives times the radii.
outside68ContEnergy	Sum of energy outside one containment radii around the electron and photon trajectories.
outside68x2ContEnergy	As above, but outside two times the radii.
outside68x3ContEnergy	As above, but outside three times the radii.
outside68x4ContEnergy	As above, but outside four times the radii.
outside68x5ContEnergy	As above, but outside five times the radii.
outside68ContNHits	Number of hits outside one containment radii around the electron and photon trajectories.
outside68x2ContNHits	As above, but outside two times the radii.
outside68x3ContNHits	As above, but outside three times the radii.
outside68x4ContNHits	As above, but outside four times the radii.
outside68x5ContNHits	As above, but outside five times the radii.
outside68ContXStd	Standard deviation of hit X-positions outside one containment radii around the electron and photon trajectories.
outside68x2ContXStd	As above, but outside two times the radii.
outside68x3ContXStd	As above, but outside three times the radii.
outside68x4ContXStd	As above, but outside four times the radii.
outside68x5ContXStd	As above, but outside five times the radii.
outside68ContYStd	Standard deviation of hit Y-positions outside one containment radii around the electron and photon trajectories.
outside68x2ContYStd	As above, but outside two times the radii.
outside68x3ContYStd	As above, but outside three times the radii.
outside68x4ContYStd	As above, but outside four times the radii.
outside68x5ContYStd	As above, but outside five times the radii.

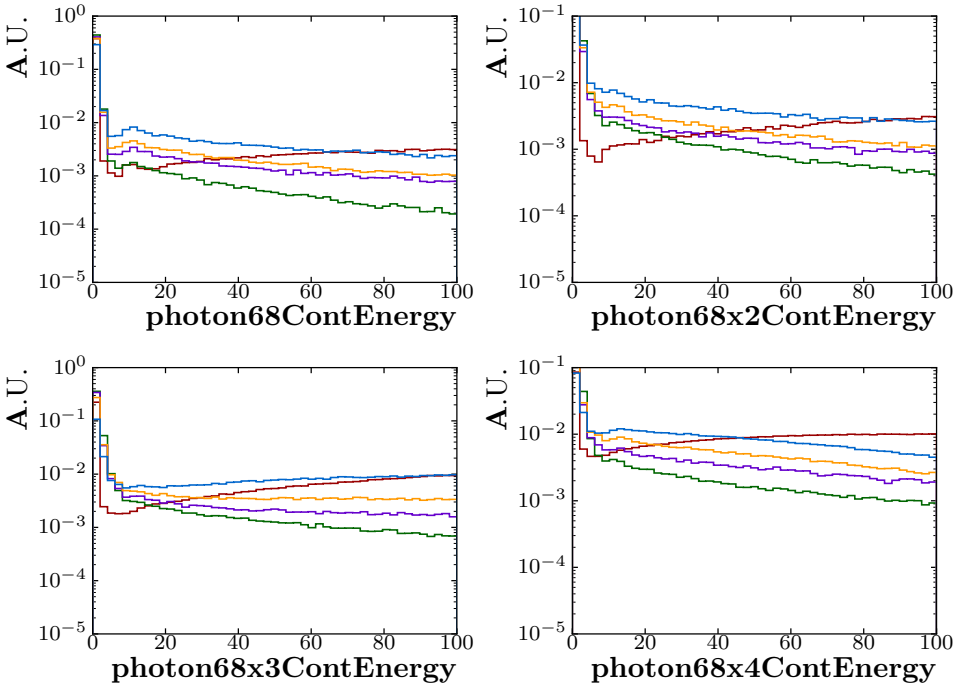
Variable Distributions

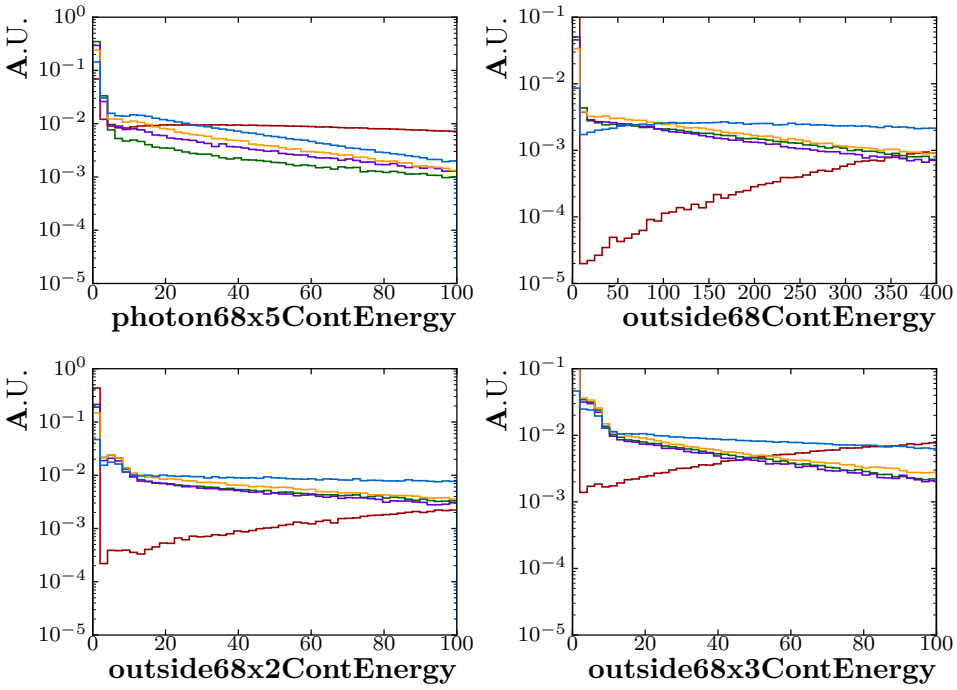

A.U.
Photonuclear
$m_{A'}$ = 1000 MeV
$m_{A'}$ = 100 MeV
$m_{A'}$ = 10 MeV
$m_{A'}$ = 1 MeV
ecalBackEnergy
nHits
summedDet
summedTightIso

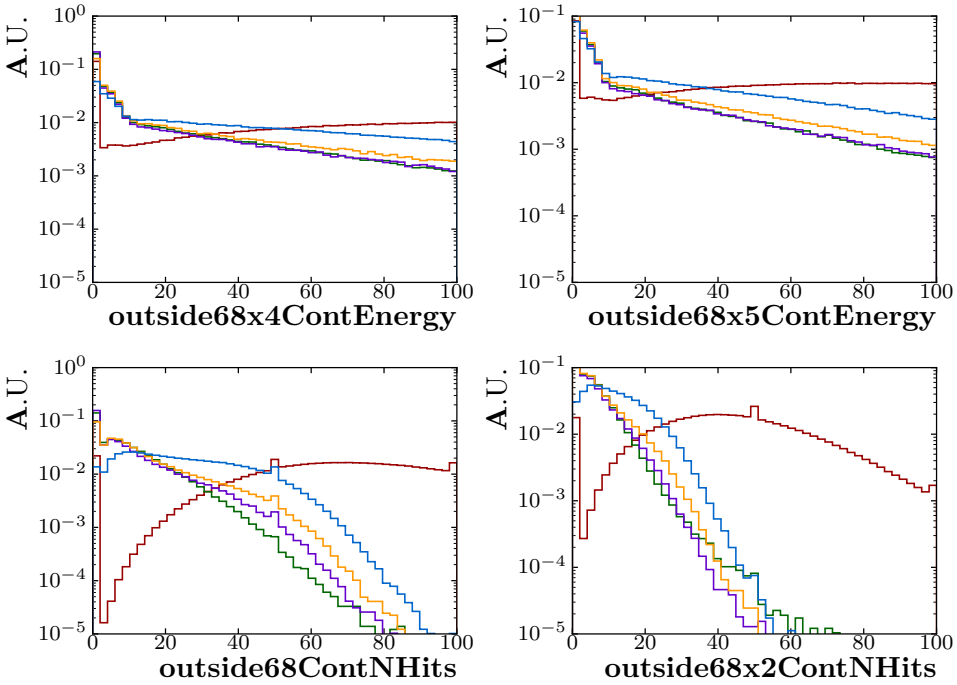

A.U.
ecalBackEnergy
nHits
summedDet
summedTightIso

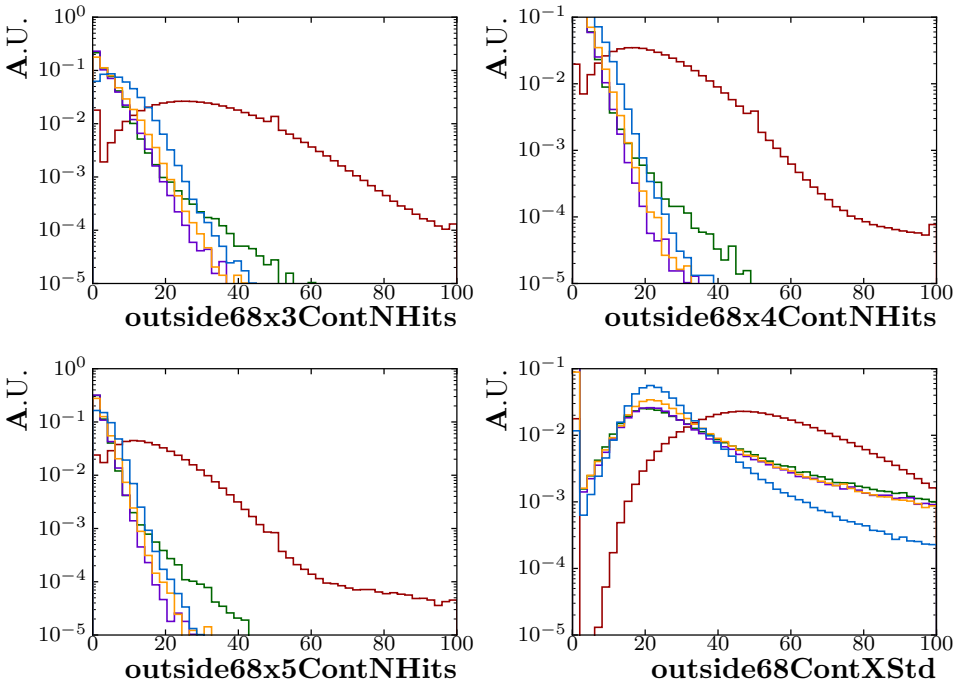

A.U.
maxCellDep
A.U.
showerRMS
A.U.
xStd
A.U.
yStd

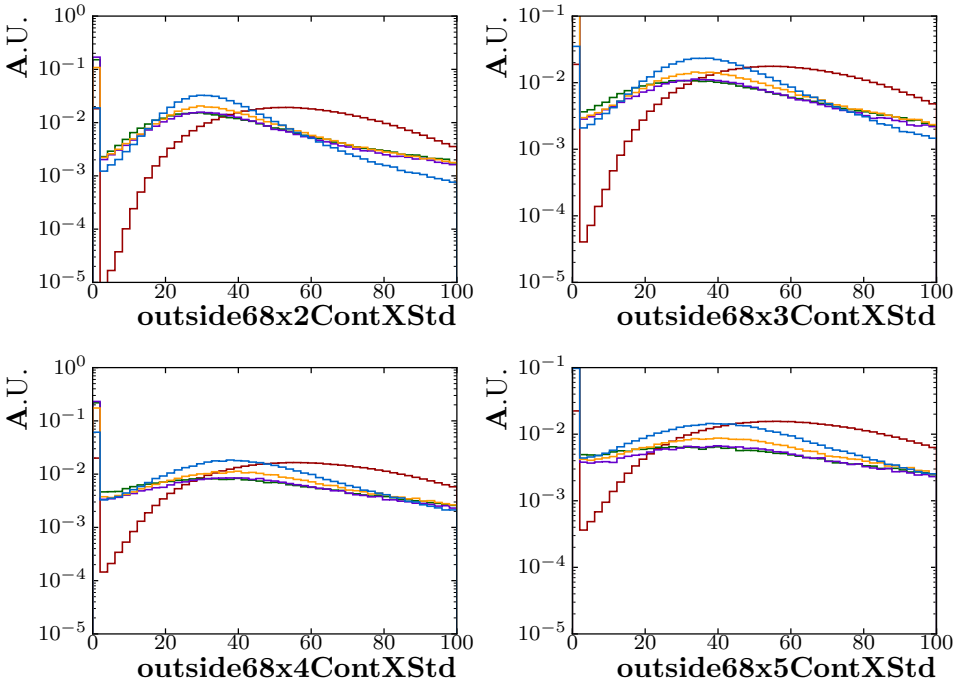

A.U.
ele68x2ContEnergy
A.U.
ele68x3ContEnergy
A.U.
ele68x4ContEnergy
A.U.
ele68x5ContEnergy

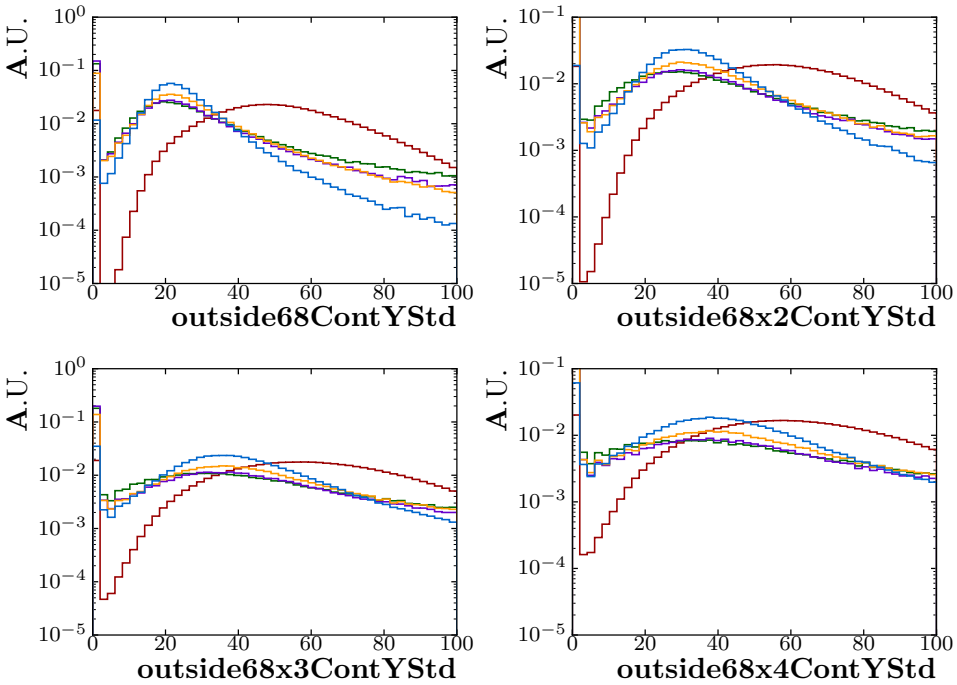

A.U.
photon68ContEnergy
photon68x2ContEnergy
photon68x3ContEnergy
photon68x4ContEnergy

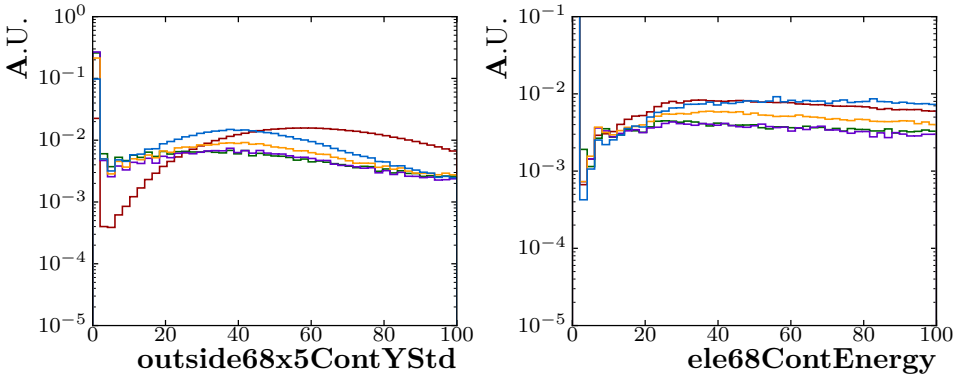

A.U.
photon68x5ContEnergy
outside68ContEnergy
outside68x2ContEnergy
outside68x3ContEnergy


A.U.
outside68x4ContEnergy
A.U.
outside68x5ContEnergy
A.U.
outside68ContNHits
A.U.
outside68x2ContNHits


A.U.
outside68x3ContNHits
A.U.
outside68x4ContNHits
A.U.
outside68x5ContNHits
A.U.
outside68ContXStd


A.U.
outside68x2ContXStd
outside68x3ContXStd
outside68x4ContXStd
outside68x5ContXStd


A.U.
outside68ContYStd
outside68x2ContYStd
outside68x3ContYStd
outside68x4ContYStd


A.U.
outside68x5ContYStd
A.U.
ele68ContEnergy