Signal and Background Studies at 8 GeV for the Light Dark Matter eXperiment

Peter Gyorgy

Thesis Submitted for Degree of Bachelor of Science

Project Duration: 2 Months

Supervised by Ruth Pöttgen

LUND UNIVERSITY

Department of Physics
Division of Particle Physics
May 2020

Abstract

Light dark matter is a hypothesized form of dark matter in the 1 MeV to 1 GeV mass range. The Light Dark Matter eXperiment (LDMX) is an upcoming experiment to test the existence of light dark matter, by colliding an electron beam with a tungsten target and assembling a set of events that potentially indicate the existence of light dark matter. The three main detector components of the LDMX are a set of trackers, an electromagnetic calorimeter, and a hadronic calorimeter, each of which gives a collection of variables that can be used to assemble criteria to select potential dark matter events. The optimal selection criteria for a 4 GeV beam have already been established by the LDMX collaboration, but a higher beam energy will be used at later stages of the LDMX project, since higher energies provide better sensitivity. This thesis is the first step in exploring the creation of a viable set of selection criteria for the LDMX at an upgraded beam energy of 8 GeV. This was done by first achieving an understanding of the existing 4 GeV selection criteria, and then exploring new possibilities for additional selection criteria. A sample of ~17 million of the most difficult background events and ~1 million signal events was assembled, and a procedure was devised for creating a set of selection criteria that can reject all background events while maintaining a high signal efficiency. The final selection criteria for this exploration are: an electromagnetic calorimeter energy under 6274 MeV, an hadronic calorimeter energy under 15 MeV, a number of readout hits in the electromagnetic calorimeter less than 106, and a single recoil electron track. The signal efficiencies are: 83.77% for 1 MeV, 81.88% for 10 MeV, 71.26% for 100 MeV, and 67.26% for 1000 MeV dark matter mediator masses.

Acknowledgements

First and foremost, I would like to thank my supervisor Ruth Pöttgen for her extraordinary support, thorough feedback, and weekly meetings. I would also like to thank Lene for assembling the simulations that were the basis of my thesis. Additional thanks goes out to the LULDMX Zoom group for listening to my presentation. The computations and plotting were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at LUNARC partially funded by the Swedish Research Council through grant agreement no. 2016-07213.

List of Acronyms

LDMX Light Dark Matter eXperiment
LDM Light Dark Matter
DM Dark Matter
WIMPs Weakly Interacting Massive Particles
LHC Large Hadron Collider
LEP Large Electron–Positron Collider
LCLS-II Linac Coherent Light Source 2
LCLS-II HE Linac Coherent Light Source 2 High Energy
SLAC Stanford Linear Accelerator Center
Geant4 GEometry ANd Tracking (software framework)
Ecal Electromagnetic calorimeter
Hcal Hadronic calorimeter
$m_{A'}$ The mass of a dark matter mediator
EoT Electrons on Target
maxPE maximum number of photoelectrons
NReadoutHits Total number of Ecal cells that were hit
SummedTightIso The sum of energy from isolated Ecal cells.

Introduction

Dark matter is a theorized form of matter thought to make up 85% of all matter in the universe.[1] Despite its abundance, little is known of what it actually is, and humanity's quest to create and/or identify it has so far only yielded negative results. Most theories of dark matter postulate that it is composed of a yet undiscovered form of subatomic particles.[1] The majority of projects search for dark matter that can be classified as WIMPs - Weakly Interacting Massive Particles. Most of these projects aim to find dark matter with a mass between 1 GeV – 1 TeV. However, an expedition into a neighboring realm - the realm of Light Dark Matter, which lies in the mass range of 1 MeV – 1 GeV - has not really been done before.

The Light Dark Matter eXperiment (LDMX) is an upcoming multi-stage experiment in the design and development phase, intended to determine whether dark matter exists within this mass range.[1] It does so by colliding electrons from an electron beam with tungsten nuclei in a thin target, and then digitally reconstructing these collisions using detectors. This way, it hopes to find interactions where dark matter is created. The energy of the beam plays an important role in how the events evolve, with higher energies creating lower rates for some backgrounds, and making it harder to mimic dark matter signatures for some other backgrounds. In the first stage (Phase I), the beam energy will be 4 GeV. In the second stage (Phase II), the beam energy will be 8 GeV. The exact benefits of this energy increase will be investigated in this paper.

The data set of the experiment for Phase I is going to be a set of 4×10^{14} incident electrons on target (EoT).[1] For Phase II, this increases to up to 10^{16} EoT. This set will be sorted through, and events that closely resemble a dark matter creation event will be chosen. The criteria for selecting these events from a 4 GeV sample have already been optimised by the LDMX team.[2] This paper aims to find adequate selection criteria for an 8 GeV sample.

Background

The Standard Model and its limitations

According to the Standard Model of particle physics, matter is made up of three generations of quarks and leptons. Interactions between quarks and leptons are mediated by bosons. The Standard Model describes three fundamental forces: the strong force, which is mediated by gluons and acts only on gluons and quarks, the weak force, which acts on all quarks and leptons and is mediated by W and Z bosons, and the electromagnetic force, which acts on all

[1] Best estimates show the universe consists of 26.8 % dark matter and 4.9% regular matter,[3] so 26.8/(26.8+4.9) = 85%.

charged particles and is mediated by photons. The Higgs boson is responsible for giving particles mass, but does not belong to any force.[4]

While the Standard Model has been remarkably successful in providing experimental predictions,[5] it fails at explaining several important cosmological phenomena, such as the presence of hidden mass within galaxies, gravitational lensing effects, and structure formation processes in the universe.[6] The hidden mass within galaxies is a consistent observation that the stars in outer parts of most galaxies spin faster than the galaxy's apparent mass would allow, and that there must be more hidden mass towards their outer areas. The hidden mass in gravitational lensing effects comes from the observation that light is bent significantly more in certain mass clusters than their apparent mass would permit. The presence of hidden mass in galactic formation processes is attributed to the fact that, inside computer simulations of the universe, galaxies cannot form as early as they should.

Dark matter

To explain these phenomena, the existence of dark matter (DM) was proposed, which postulates that the hidden mass originates from yet unknown particles that do not interact electromagnetically, but do interact gravitationally. This way, the hidden mass can be explained, and the quantity of dark matter in the universe was estimated to be 27% of the universe's energy and 85% of the universe's mass.[3]

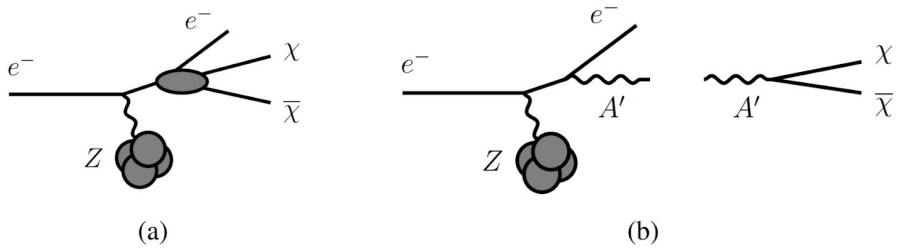
A promising theory for the origin of dark matter is that DM is a thermal relic formed during the early stages of the universe. In this theory, dark matter particles can have masses in the MeV to 100 TeV range and must have some small non-gravitational interaction with ordinary matter. The existence of any such interactions not only means that dark matter can be directly detected, but also that it can be produced in particle accelerators. Many experiments have been made to directly detect naturally-occurring dark matter. Other experiments have also been made that detect fragments made up of normal matter left over from cosmic dark matter annihilation. Even the LHC and LEP have attempted to create dark matter. However, these experiments were primarily tuned to finding DM with a mass above 1 GeV, and weren't particularly sensitive to masses below. So far, these experiments have only yielded stringent exclusion of possible dark matter models with over 1 GeV mass, and no confirmations of the existence of any dark matter.

Light dark matter and its production

While a large majority of possible WIMP models have been excluded, models for light dark matter (LDM), where "light" refers to the MeV to GeV region, remain mostly unexplored. Several new experiments have been created or are being designed to explore these models, such as NA64,[7] SHiP,[8] and LDMX.[1]

In the LDM scenarios considered for the LDMX, dark matter or a dark matter mediator is created

directly from Standard Model matter. In the main benchmark model of the LDMX, this happens when an electron undergoes a “dark bremsstrahlung” process with a heavy nucleus. There are two ways this could happen: directly, which produces a dark matter pair $\chi\chi$, or through the production and decay of a mediator particle, called a dark photon A'. The mass of dark matter is commonly denoted by m_{χ}, and the mass of a dark photon is denoted by $m_{A'}$, and will be used extensively in this thesis. Figure (1) shows the two possible processes in which this happens. The LDMX can also test other models with different DM creation processes that have similar results to the main benchmark model.



Figure(1): The creation of a dark matter pair $\chi\chi$ directly (a) and the creation of a dark matter $\chi\chi$ pair through a mediator A' (b)[14]

Much like regular photons, dark photons do not interact via the strong or weak force. However, unlike regular photons, dark photons have mass, and do not interact electromagnetically. This means that in a dark matter event, the dark photon will carry away at least its mass equivalent in energy, and usually more. And since it cannot be detected electromagnetically, or via the strong or weak force, it can't be practically detected at all. Therefore, looking at such an event, an observer would clearly see energy disappear into nothingness. This is called the missing energy. Additionally, since momentum is conserved in a collision, but the dark photon left the interaction with a momentum in a specific direction, the recoiling electron must have an equivalent and opposite momentum, which can be measured. The momentum carried away by the DM is called the missing momentum.

Design of the LDMX

The LDMX setup consists of an electron beam, a target, and several layers of detectors. While there are several potential beamlines the LDMX could be attached to, the primary candidate is the LCLS-II at SLAC.[1][9] For the LCLS-II, the electron beam has a relatively low current of 10^8 electrons/sec, and a relatively high bunch repetition rate of approximately 40 MHz. This is so that only one electron hits the target for each event. The LCLS-II can run at 4 GeV, but with the LCLS-II HE upgrade, this could be increased to 8 GeV. Figure (2) shows the entire LDMX setup.

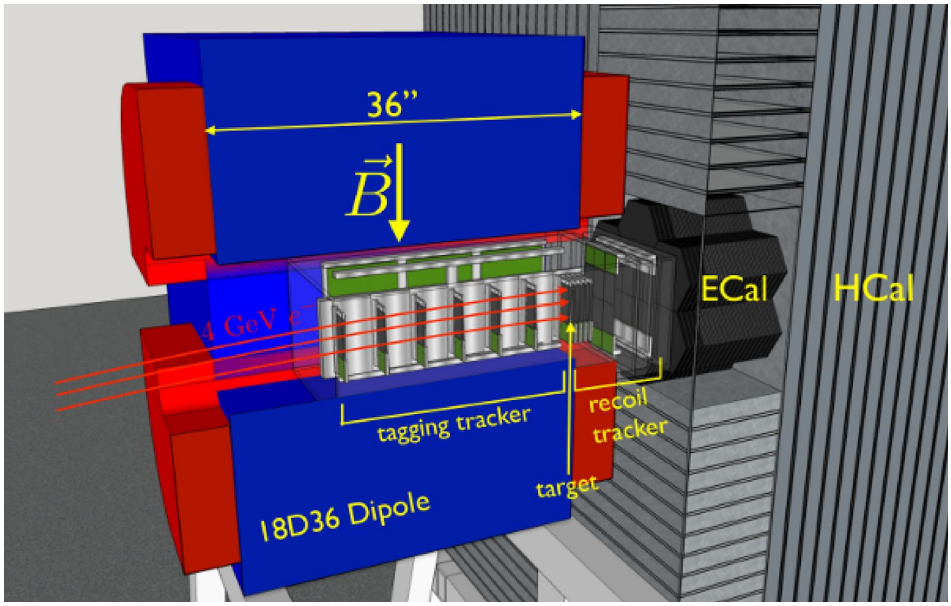


Figure (2): The LDMX setup.[15]

The first layer is the tagging tracker, which consists of six modules that are immersed in a 1.5 Tesla magnetic field. These modules are double-sided silicon microstrips, which have a horizontal resolution of 6 μm, and are arranged in 10 cm intervals. The tagging tracker is used to identify the exact momentum of incoming electrons, and filter out any off-energy electrons.

After passing the tagging tracker, the electrons hit the target, which is 350 micron thick and made of tungsten. This corresponds to 10% of a radiation length, which should provide a reasonable balance between maximizing the signal rate and minimizing multiple scattering. Nevertheless, the target is easy to change, so different materials and thicknesses may also be used.

After hitting the target, the recoil electron and other interaction remnants fly through the recoil tracker. The recoil tracker consists of several modules, the first 4 being identical to the tagging tracker modules, but spaced 7.5 mm apart. However, the last two recoil tracker modules are different, being larger and consisting of six standard p^+-in-n silicon microstrip sensors each. The recoil tracker is designed to best identify low-momentum 50 MeV to 1.2 GeV electrons, but it also helps to reject multi-particle backgrounds.

After the recoil tracker comes the electromagnetic calorimeter (Ecal), which detects mainly the electrons, photons, and most other charged particles that were created during the collision. It is a sampling calorimeter, and consists of 32 layers of detectors, with each layer having 7 hexagonal detector plates. Each plate consists of 4 component layers. The first component layer is a W-Cu baseplate. The second is with a gold-kapton insulator. Third is a high granularity 0.5 mm silicon sensor, with a granularity of 432 pads, each of area 0.52 cm^2. The final component layer is a printed circuit board that houses the electronics needed for the

functioning of each plate. The entire Ecal is 40 radiation lengths deep, so while runaway electromagnetic showers can happen, they are extremely rare.

Finally, the hadronic calorimeter (Hcal) detects any neutral hadrons, minimum ionising particles (such as muons), and the very rare electromagnetic showers that escape the Ecal. The Hcal itself is also a sampling calorimeter, and consists of close to 100 layers. Each layer has scintillator bars of dimensions 2 m x 50 mm x 15 mm, made up of doped polystyrene, attached to a 25 mm thick sheet of steel absorber. When a particle hits the absorber, it creates secondary charged particles, which then hit the scintillator. The scintillator produces photons as a result, which are collected via a wavelength-shifting fiber in each bar. These fibers route the photons into a silicon photo-multiplier, which creates photoelectrons that can be counted by the current they induce. While a certain design of the hadronic calorimeter is used in the simulations in this paper, the exact geometry and layout of the final Hcal is still being optimised. Figure (3) shows the entire LDMX, and a human for scale.

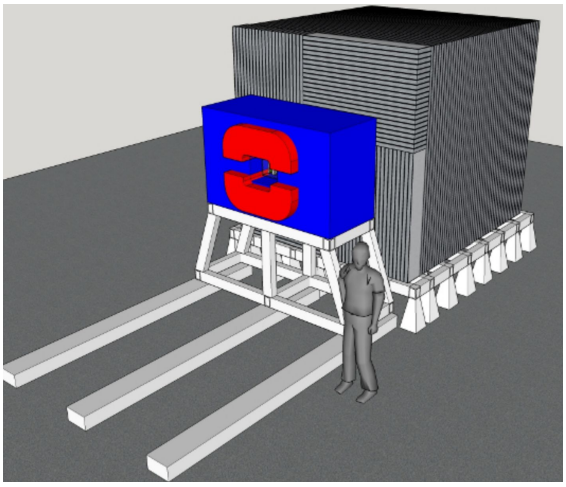


Figure (3): The LDMX, and a human for scale. The magnet is highlighted in red, and the striped grey cube is the Hcal. The trackers are inside the magnet, and the Ecal is fully encased by the Hcal.[16]

Simulating events in the LDMX

In order to understand the data that the LDMX will generate, a clear understanding of the particle interactions inside the LDMX must be achieved. Additionally, one must also carefully investigate how the device presents the data, and what can be inferred from it. To do this, simulations are run, and their output analysed. The simulations are done using an in-house program called LIGT, which stands for **L**DMX **I**nterface to **G**EANT4 **T**oolkit. As the name

suggests, it is primarily based on GEANT4,[2][10] but also includes features adopted from GDML,[2][11] Madgraph,[2][12] and ROOT.[2][13] The simulations consist of three stages.

In the first stage, the interaction between the incoming electron and the target is simulated. This is done in GEANT4 for non-dark matter events, and Madgraph for dark matter events. The events generated by Madgraph are then interfaced to GEANT4 using LHE files.[1]

In the second stage, the interaction between all the created particles and the detector is simulated. This is done entirely in GEANT4.

The third stage of the simulation is reconstructing the event based on what the detector can detect. At this stage, the detector hits are digitised, noise is added, and basic track finding is performed. This yields a result that closely matches the data that the physical detector would show.

The result of the simulation then has to be analysed. This is done using a ROOT enhanced Python program. The program identifies tracks and categorises photonuclear backgrounds, and also prepares the data for plotting. A final Python program applies the selection criteria, and plots the results. While both Python programs were written by the LDMX collaboration, the programs have been enhanced and modified as part of this thesis. New features include: the implementation of logging Hcal energy, calculation of efficiencies, plotting up to 5 variables, and streamlining the plotting process.

Signal events

Within the framework of the LDMX, a dark matter event occurs when a beam electron interacts with a target nucleus via a theorized "dark bremsstrahlung" process. This way, most of the energy of the electron is carried away by either a dark matter mediator particle or a dark matter particle pair. This allows a relatively clean detection of the single recoiling electron in the electromagnetic calorimeter. In the vast majority of cases, the dark matter will carry away more than half of the electron's initial energy.

The mass of a dark matter mediator in this experiment could be anywhere between 1 MeV and 1 GeV. This mass plays an important role in how the event evolves. Larger masses tend to deflect their mother electrons more than lower masses, which causes more deflected electron tracks. However, they also carry away more energy in the form of dark matter, so they deposit less energy into the detector overall. Therefore, simulations are run for multiple scenarios, with the mass of the DM mediator being: 1 MeV, 10 MeV, 100 MeV, and 1 GeV. This way, the behavior of DM in its whole mass range is sufficiently accounted for.

Background events

There are several main types of background events that have to be considered for the LDMX. These are: beam impurities, electrons that don't interact with the target, hard bremsstrahlung in the target or the Ecal, electro-nuclear interactions in the target or Ecal, and neutrino backgrounds.

The beam impurities can be effectively filtered by the tagging tracker. The non-target-interacting electrons can be easily filtered by the recoil tracker, and the Ecal. The neutrino backgrounds happen rarely enough to be insignificant at Phase 1 of the LDMX. The more difficult to reject background events fall into the category of hard bremsstrahlung and electro-nuclear interactions.

In hard bremsstrahlung, a high-energy photon is created when the incoming electron's path is heavily bent by a nucleus. This photon can then undergo five particularly noteworthy interactions: It can simply be detected by the electromagnetic calorimeter, it can undergo conversion in the target or Ecal and produce an e^-e^+ pair or a $e^-e^+e^-$ trident, it can undergo a photonuclear reaction, or convert to a $\mu^-\mu^+$ pair. The first interaction can be easily vetoed as long as both the photon and recoiling electron's energies are detected, which the LDMX is well equipped to do. The e^-e^+ and $e^-e^+e^-$ backgrounds should produce several tracks when created in the target, or deposit energy into the Ecal when they are created inside the Ecal, so in both cases they can be easily identified and vetoed. $\mu^-\mu^+$ pair conversions in the target can be rejected by capping the number of tracks.[1] If the $\mu^-\mu^+$ conversion happens in the Ecal, the Hcal can still veto the event, since even muons deposit a large number of hits well above the Hcal's noise level. Additionally, the Ecal has a high granularity, so the $\mu^-\mu^+$ would leave a distinct track in the Ecal.[2]

The most difficult background to veto is when the photon undergoes a photonuclear reaction inside the Ecal. Such interactions have a very large range of possible outcomes, and some of them are particularly hard to veto. Usually, free protons, neutrons, pions, kaons, and exotic nuclear fragments are produced. In the case of 4 GeV in Phase I, this event occurs at a relative rate of $1.7*10^{-5}$ per incident electron, so in a sample of 4×10^{14} electrons, there will be 6.8 billion background events. For the case of 8 GeV in Phase II, this is expected to occur at a smaller rate, but given a larger sample size, a few billion events are still anticipated. Therefore, billions of simulations of photonuclear reactions have been assembled by the LDMX team, and this paper only focuses on photonuclear backgrounds originating in the Ecal. Given the limited scope of this paper, a sample of 17 million photonuclear events is used, which represents about 1% of all the photonuclear events simulated.

Difference between signal and photonuclear background events

For signal events, the recoil electron has a significantly lower energy than the beam energy, and it has a noticeable transverse momentum. This manifests itself in the Ecal showing a lower energy than the beam energy, and the trackers noticing a significant transverse momentum.

Photonuclear events, on the other hand, will not necessarily have a significantly lower energy. The photonuclear events without a significant energy carried by the photon are trivial to veto, and therefore only photonuclear events with a significant amount of energy in the photon are considered and simulated. Nevertheless, even these events tend to have a lower transverse momentum than the signal events.

The most notable difference, however, is that only the photonuclear background has deposits in the Hcal, and signal events should have no Hcal interactions at all. Figure (4) shows a simplified design of the LDMX, and how a signal event behaves inside the LDMX compared to certain photon-background events.

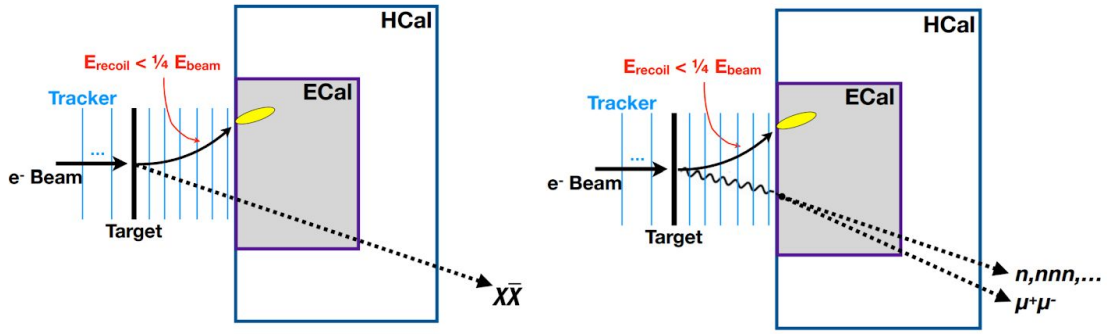


Figure (4): The conceptual design of the LDMX and how it behaves during an event where a dark matter pair is created (left), and an event where a photonuclear and muon pair background occurs (right). [17]

Assembling a set of signal candidates

After Phase I of the LDMX, a total of 4×10^{14} events will have been collected. After Phase II, up to 10^{16}. For each phase, these events will be sorted through, and the easily recognisable background events will be quickly discarded. What will remain depends on whether LDM that the LDMX is designed to detect exists or not. If it exists, there will be a set of events that consist almost entirely of signal and photonuclear events. If it doesn't, then only photonuclear events will remain. For most of the events in either scenario, it will not necessarily be certain whether they are signal or photonuclear events. However, statistical tendencies will still apply, so applying a certain set of criteria will remove the majority of background events

while not removing as many signal events - if they exist. These criteria are called the selection criteria. The selection criteria are obtained by creating a mixed sample of signal simulation events and photonuclear simulation events, trying out a wide combination of criteria, and finally selecting the one that yields the best signal/background efficiency ratio. The selection criteria for 4 GeV has already been established by the LDMX collaboration,[2] and the selection criteria for 8 GeV will be investigated for this paper.

Analysis

The aim of this paper is to establish a specific set of selection criteria that can veto most photonuclear backgrounds, while maintaining a good signal efficiency for an upgraded 8 GeV electron beam. Therefore, in this section, comparisons between 4 and 8 GeV backgrounds will be made, followed by 4 and 8 GeV signals. Then, the procedure for finding the best selection criteria will be explained, followed by an explanation of each selection criteron. After this section, the results will be presented and discussed

The quality of a set of selection criteria can be measured using its signal and background efficiency, and the signal-to-background efficiency ratio. The reason the number of signal events is not compared directly to the number of background events is because it is unknown how many signal events will occur - if any. Signal efficiency gives the percentage of signal events that remain after discarding all signal events that do not fit the criteria. The same logic applies to background efficiency. Signal-to-background efficiency ratio is simply the signal efficiency divided by background efficiency. In cases where there are no background events, only signal efficiency can be used.

While there is only one type of background to consider, the mass of the dark matter mediator ($m_{A'}$) could be anywhere between 1 MeV to 1 GeV. Therefore, to cover the full range at an appropriate resolution, there are 4 scenarios considered: a $m_{A'}$ of 1 MeV, 10 MeV, 100 MeV, and 1 GeV.

The histograms that are used have been normalized, so that the integral of each plot is equal to one. This allows for visual comparison between different sized samples, since there are 17 million background events, but only 200 to 400 thousand signal events for each m_A mass.

Comparison between 4 GeV and 8 GeV Backgrounds

The increase in energy leads to a change in the observed background. As expected, the 8 GeV events penetrate deeper into the calorimeters, more energy is deposited into the Ecal, and the particles themselves have proportionately more kinetic energy. Some less obvious consequences are that for higher energies, the particles disperse at a smaller angle, more photonuclear reactions occur, and more kinds of particles are created. Figure set (5) shows each of these differences in the order they were mentioned.

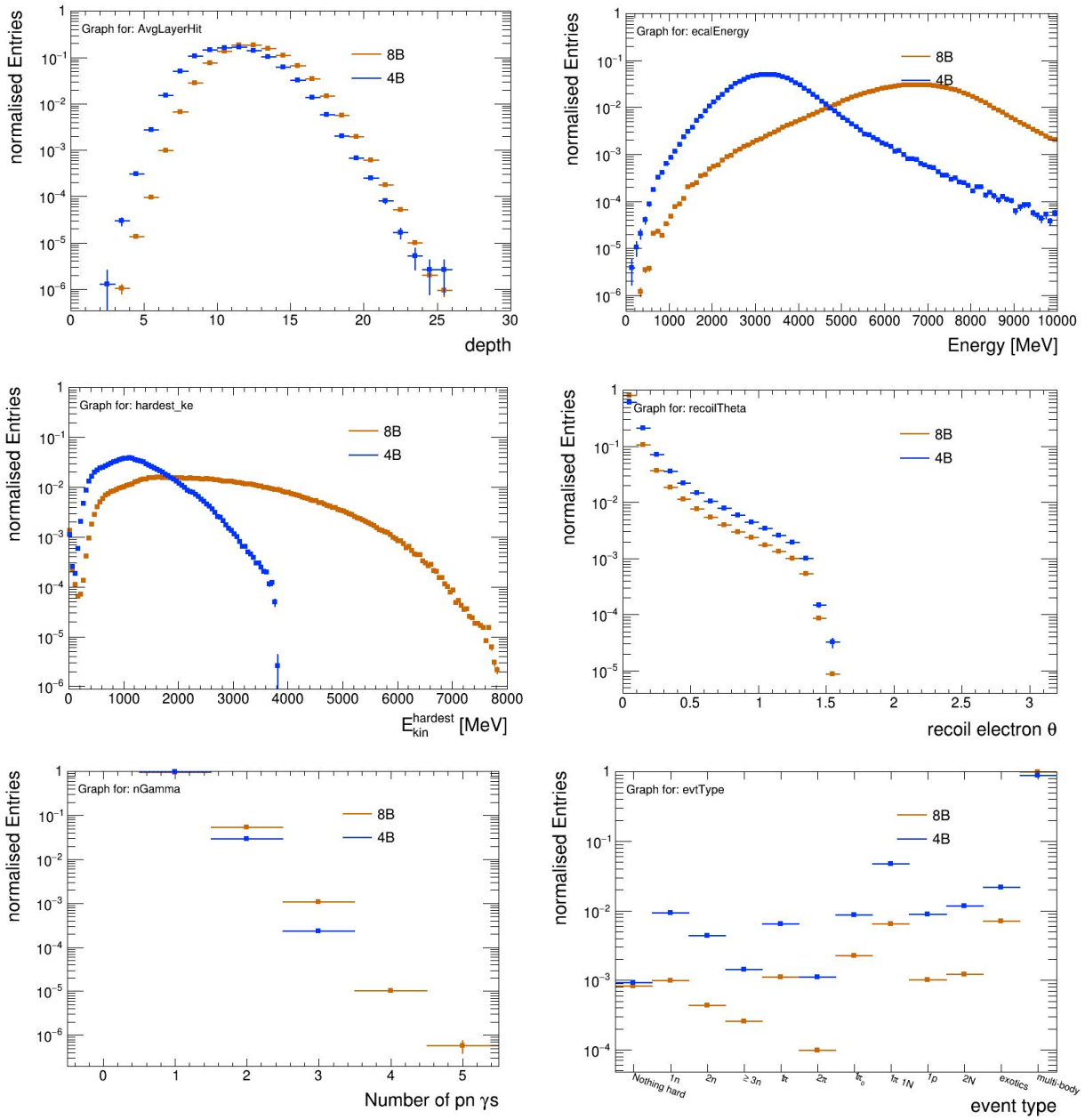


Figure set (5): Orange signifies background events for 8 GeV beam electrons, and blue signifies background events for 4 GeV beam electrons. Top left figure shows the average layer depth the Ecal showers reach. Top right shows the energy deposited in the Ecal. Middle left figure shows the kinetic energy for the most energetic particle. Middle right shows the recoil electron's angular dispersion θ. Bottom left shows the number of photons undergoing a photonuclear reaction. Bottom right shows the types of final states occurring in each event. "Nothing hard" indicates that no final states had a kinetic energy over 200 MeV. Every other state denotes particles with over 200 MeV. 1n, 2n, and ≥ 3n denotes only neutrons being created. 1π and 2π denote positive pions being created, while $1\pi_0$ denotes a neutral pion being created. 1π 1N denotes one positive pion and one nuclide, 1p dentoes one proton, and 2N denotes 2 nuclides. Exotic denotes events with particles that are neither nuclides nor positive and neutral pions. Multi-body denotes events where a combination of these types occur.

Comparison between 4 GeV and 8 GeV Signals

Since the selection criteria have already been chosen for 4 GeV, and the selection criterion for the Ecal energy is 1.5 GeV, the 4 GeV signal generation software has been modified so as to not generate recoil electrons above 2 GeV, since they would be discarded anyway. Therefore, there is a cutoff present at 2 GeV in the 4 GeV simulation samples, but no cutoff is present in the 8 GeV simulation samples. Nevertheless, a comparison can still be made, and the results have a similar conclusion to the background comparisons. Namely, the higher energy leads to deeper hits, more energy deposited into the Ecal, and a lower recoil angle. Figure set (6) shows this.

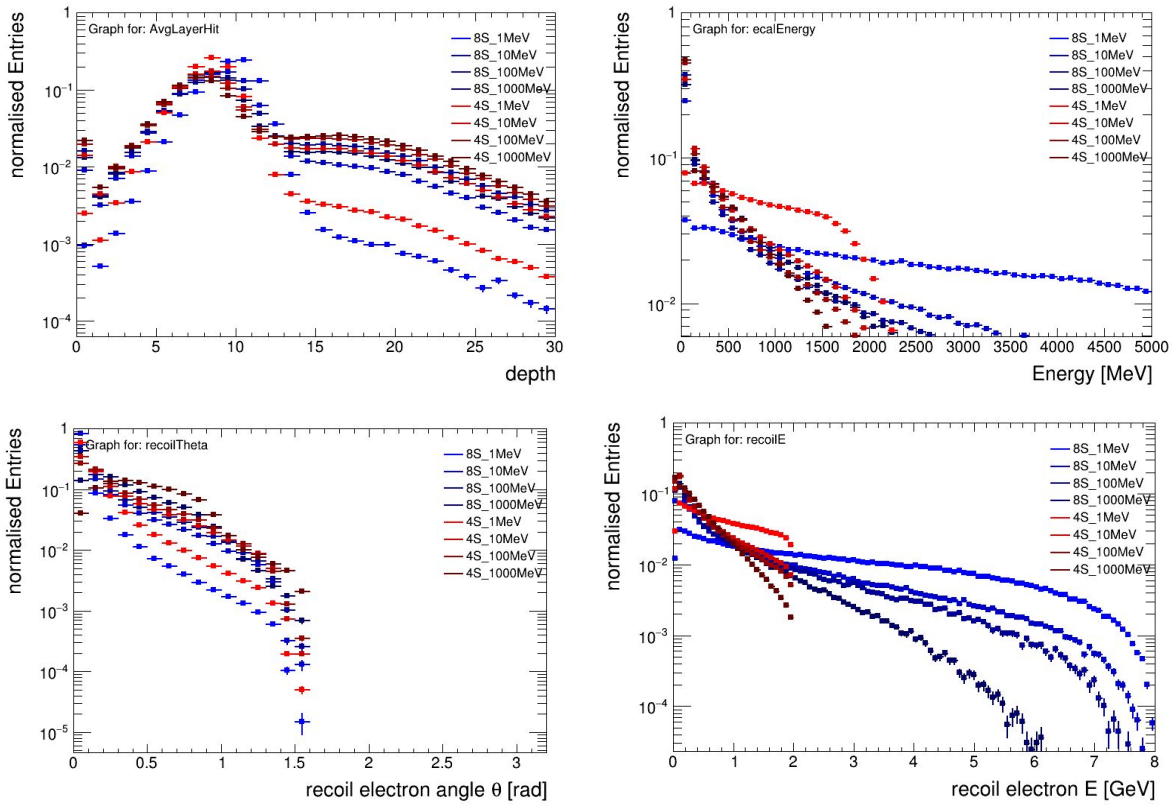


Figure set (6): Red signifies signal events for 4 GeV beam electrons, and blue signifies signal events for 8 GeV beam electrons. Darker colors signify heavier $m_{A'}$. Top left figure shows the average layer depth the Ecal showers reach. Top right shows the energy deposited in the Ecal. Bottom left figure shows the kinetic energy for the most energetic particle. Bottom right shows the energy of the recoil electron, with the 2 GeV cutoff being especially apparent.

Selection criteria

The selection criteria for 4 GeV have already been established.[2] There are 4 main selection criteria: the energy deposited in the electromagnetic calorimeter, the number of tracks in the event, the maximum number of photoelectrons produced in a single scintillator bar in the hadronic calorimeter, and a confidence value assigned by a boosted decision tree (which is a machine learning program).

Selection criteria optimisation

While the very best selection criteria could only be obtained by trying out every possible variant and choosing the one with the best result, this is computationally not feasible within the scope of this project. The second best option is to use a boosted decision tree (BDT), which is unfortunately unavailable for this project. The remaining option is to optimize using human decision making.

This procedure consists of choosing an initial loose set of selection criteria, and adding more criteria based on what appears to improve the efficiency ratio the most, with the final added criteria almost or completely eliminating all the backgrounds. This method is applied several times with different initial criteria. The most successful attempt is then chosen. In the case that complementary criteria exist, both are applied, and balanced so as to reach best results. Finally, each one of the selection criteria is loosened as much as possible so as to still reject all backgrounds. The combination of criteria that result in the highest 1 MeV signal efficiency is then chosen as the final criteria. The reason for optimizing for 1 MeV signals is that 1 MeV $m_{A'}$ events are the most difficult to separate from background events, and they are also the most sensitive to the selection criteria that can be optimized.

Energy deposited in the electromagnetic calorimeter

In a signal event, part of the energy of the original beam electron is carried away by dark matter. The more energy it carries away, the greater the certainty that dark matter was created. Additionally, the most frequent photonuclear background events deposit relatively high levels of energy into the Ecal. This is illustrated in Figure (7)

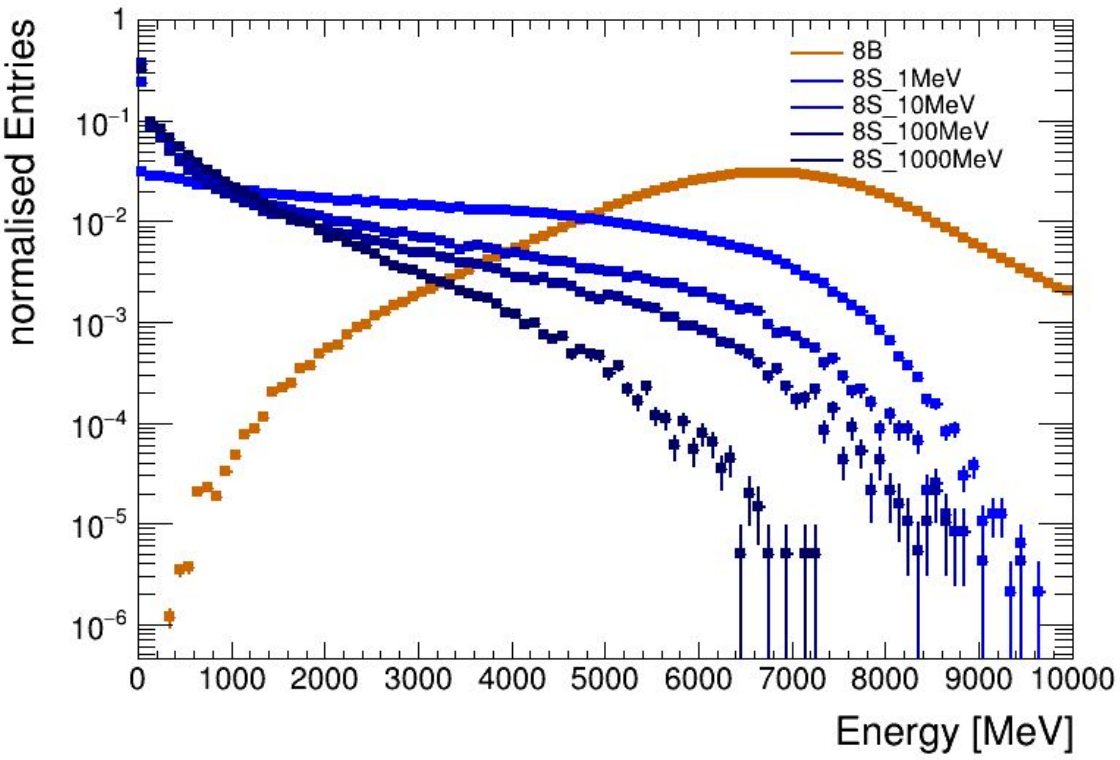


Figure (7): The energy deposited into the Ecal, with background events shown in orange, and signal events shown in blue, for the cases of 1, 10, 100, and 1000 MeV DM mediator at an 8 GeV beam. Darker blue represents a heavier $m_{A'}$.

The Ecal energy selection criteria is so far the only trigger in the LDMX. This means that any events that do not pass it are automatically discarded without even being recorded, which eases the load on the data acquisition buffer on the LDMX.

After thorough optimisation, taking into account multiple possible DM masses, an energy deposit of at most 1.5 GeV was chosen as a trigger criterion for the 4 GeV beam by the LDMX collaboration.

Adapting this to an 8 GeV beam, but keeping the criterion at 1.5 GeV, a signal efficiency of 87.1% is achieved for 1 GeV DM, but only 35.1% for 1 MeV DM, while having a background efficiency of 0.0584%.[Appendix A.1]

Using a 3 GeV criterion - which means the ratio between the beam energy and the selection criterion's energy is kept the same - we get a signal efficiency of 97.1% for 1 GeV DM, but only 59.3% for 1 MeV DM, while having a background efficiency of 1.29%.[Appendix A.1]

In order to help find the optimal Ecal energy selection criterion, the efficiencies can be plotted as a function of the Ecal energy criterion. Figure (8) shows the full range of efficiencies as a function of the Ecal energy selection criterion. Unfortunately, the plot has limited use when

also applying additional selection criteria, but it is informative of the rejection power of the Ecal trigger veto. A 5 GeV cut, for example, would automatically filter about 90% of photonuclear background events, while still preserving at least 80% of signal events.

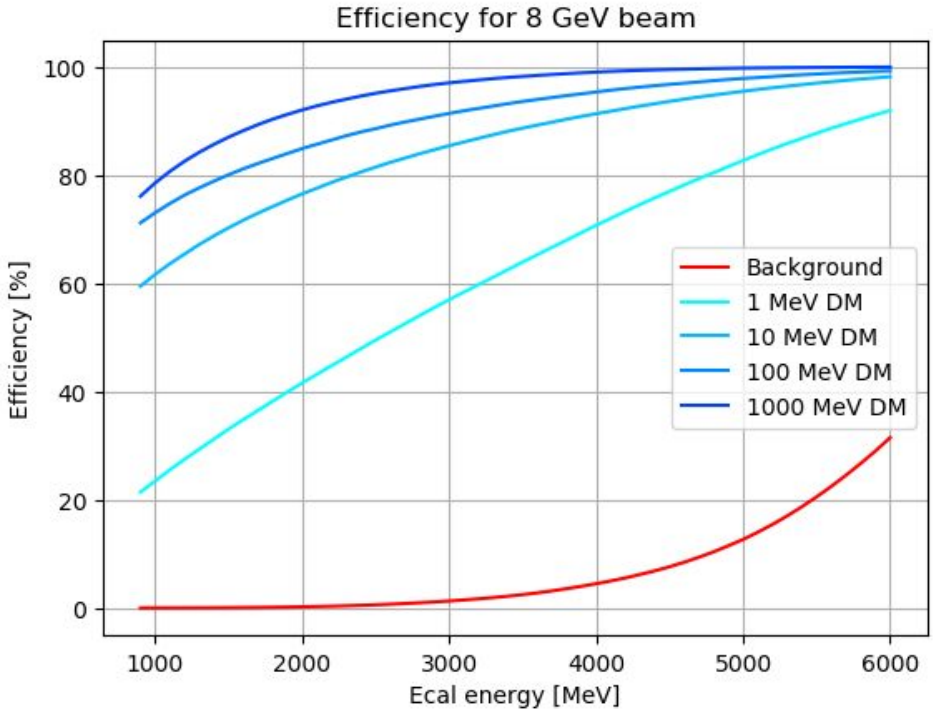


Figure (8): The background and signal efficiency for various $m_{A'}$, as a function of the selection criterion of energy deposited into the electromagnetic calorimeter, for an 8 GeV beam.

Number of tracks

In a signal event, there should only be a single track present - one left by the recoiling electron. If there are more, that means that additional interactions took place during the event. These interactions pollute a clear picture, and are therefore discarded. In some cases, no track is detected at all, which means that the recoil electron's path remains unknown. These are also discarded, since essential information is missing. Therefore, the selection criterion for the number of tracks is "1 track." Applying this selection criterion leaves a signal efficiency of 94% for 1 MeV, 85% for 10 MeV, 73% for 100 MeV, and 68% for 1000 MeV $m_{A'}$[Appendix A.2]. This is because heavier dark matter deflects its mother electron at a larger angle, and will more frequently deflect at an angle extreme enough to not be picked up by the recoil tracker, leaving no electron track. The tendency of heavier $m_{A'}$ events to deflect at a larger angle can be seen well in the bottom left figure of Figure set (6). Figure (9) shows the distribution of the number of tracks for the background and each $m_{A'}$. Since having a single recoil electron track is such a

fundamental selection criterion, the rest of the analysis will be carried out with the single track criterion as a basis.

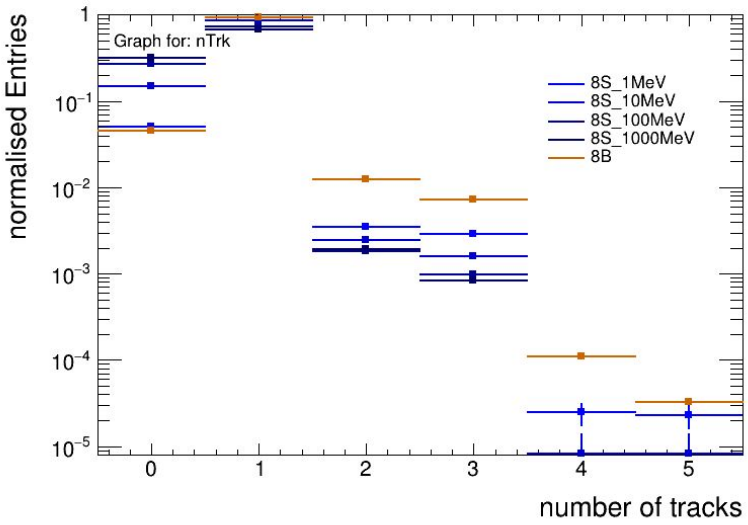


Figure (9): The number of tracks for each $m_{A'}$ and background. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$.

Hadronic calorimeter veto

In theory, the hadronic calorimeter should only signal an energy deposit when it is hit by particles that got past the Ecal. This would mean that a signal event where only dark matter is created should not have any energy deposited in the hadronic calorimeter. Unfortunately, calorimeters are affected by random noise, so a few photoelectrons will be registered, even when nothing hits the calorimeter.

For a 4 GeV beam, it was found that a limit of a maximum of 4 photoelectrons emerging from any bar is optimal.[2] This selection is called the maximum number of photoelectrons, or maxPE for short. Using an 8 GeV beam with the same selection criteria, an interesting development can be seen: no background events remain with an Ecal energy under 5 GeV. In fact, the same phenomenon can be seen up until a maximum of 7 photoelectrons. Figure set (10) shows how no background events remain under Ecal energy 5 GeV when plotted with these selection criteria.

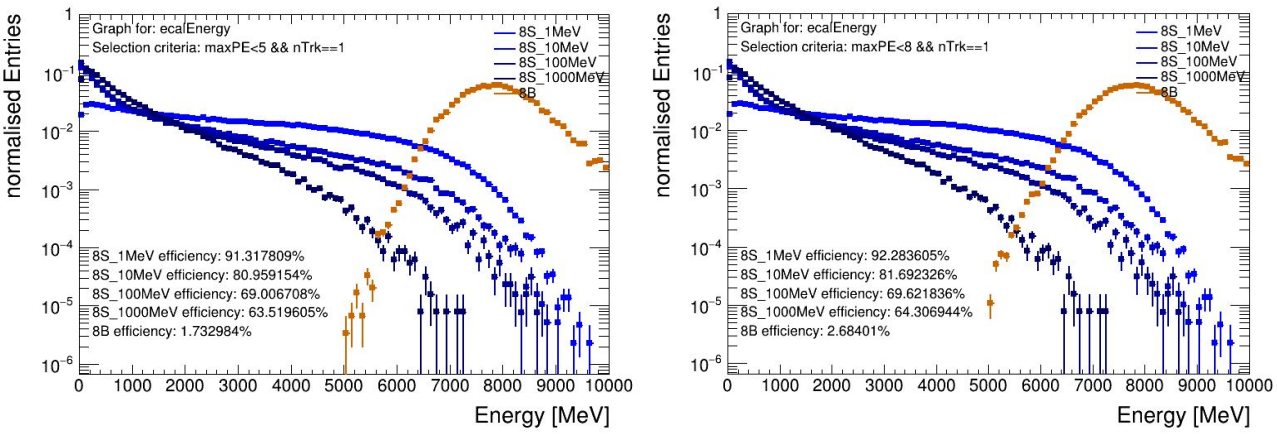


Figure set (10): the Ecal energy for events with a single electron track, at maximum photoelectron selection criteria <5 (left) and <8 (right). Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$. Of special interest is the lack of background events under 5 GeV.

Applying the selection criteria of only one recoil track, a maximum Ecal energy of 5 GeV, and less than a maximum of 8 photoelectrons, a signal efficiency of 77.6% for 1 MeV, 77.7% for 10 MeV, 67.8% for 100 MeV, and 64.1% for 1000 MeV $m_{A'}$ is obtained[Appendix A.3]. This is contrasted with a complete elimination of the background for the sample this paper is working with. While this is a possible solution, it is possible to increase the signal efficiency while still maintaining no background by using additional selection criteria.

There is yet another useful measurement that the Hcal can make. This is the overall energy deposited in the Hcal. While it is more susceptible to noise, it is distinctly different from the maximum number of photoelectrons, and thus gives a different set of information. Setting a selection criteria of a single track and energy deposited in the Hcal under 10 MeV, a similar observation emerges as with filtering by the maximum number of photoelectrons (maxPE). Namely, it is that events under 5 GeV Ecal energy disappear, as can be seen in Figure (11). While the background efficiency is clearly worse in this case, with 4.00% as opposed to 1.73% in the case of under 5 photoelectrons, the selection criterion of Hcal energy will be beneficial when choosing more complicated selection criteria. This is because the Hcal energy criterion has a complimentary effect to the maxPE criterion when it has larger values. This can be seen really well in Figure (12) with the Hcal energy plot for maxPE < 14, where it is possible to make the cut Hcal energy < 13 MeV without any loss to signal while discarding a good chunk of background events.

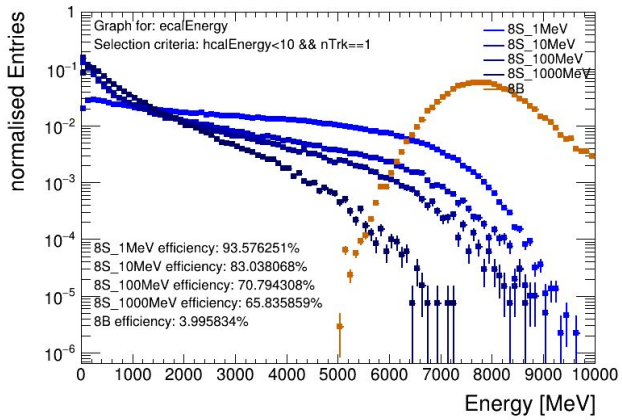
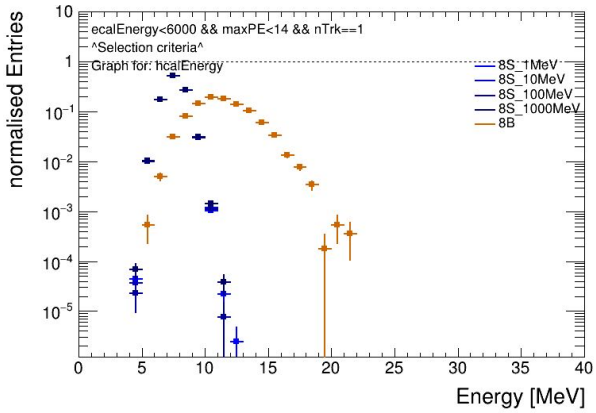


Figure (11): the Ecal energy for events with less than 10 MeV deposited in the Hcal. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$.



Figure(12): The Hcal energy distribution for events with less than 6 GeV deposited in the Ecal, a single electron track, and a maximum number of 13 photoelectrons. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$. Of special interest is the possibility to make an Hcal energy cut at 13 MeV, which allows discarding a good chunk of background events without discarding any signal events.

Boosted decision tree

For the 4 GeV phase, a boosted decision tree was created, and thoroughly trained through machine learning to recognise signal events and discard background events. It based its decisions entirely on 12 variables obtained by the Ecal. The result of this program is a single confidence value that lies between 0 and 1 for each event. A value of 0.99 was chosen as a selection criteria. Unfortunately, a boosted decision tree could not be trained in time for this project, so it is currently not included as a selection criterion for the 8 GeV phase. However, the 12 variables that it based its decisions on can still be used to manually improve the selection criteria. A full list and plots of these variables can be found in Appendix B. Out of these, 2 make a very visible difference:

1. NReadoutHits: The total number of cells that were hit.
2. SummedTightIso: The sum of energy from isolated Ecal cells. Isolated cells are cells that show a readout, but their neighbors do not.

   For both variables, photonuclear events tend to have a higher value.

The benefit of using these variables can be seen by starting out with a relatively loose set of selection criteria, such as a single recoil electron track, paired with an Ecal energy under 6 GeV and maximum number of photoelectrons being under 20. Plotting the remaining events from these criteria, we can observe the potential of new selection criteria in Figure set (13). Of particular note is NReadoutHits, which when set to 145, allows for the exclusion of the vast majority of backgrounds at no loss of signal, and SummedTightIso, which when set at 1500 MeV, allows for a very good exclusion of backgrounds at almost no signal loss.

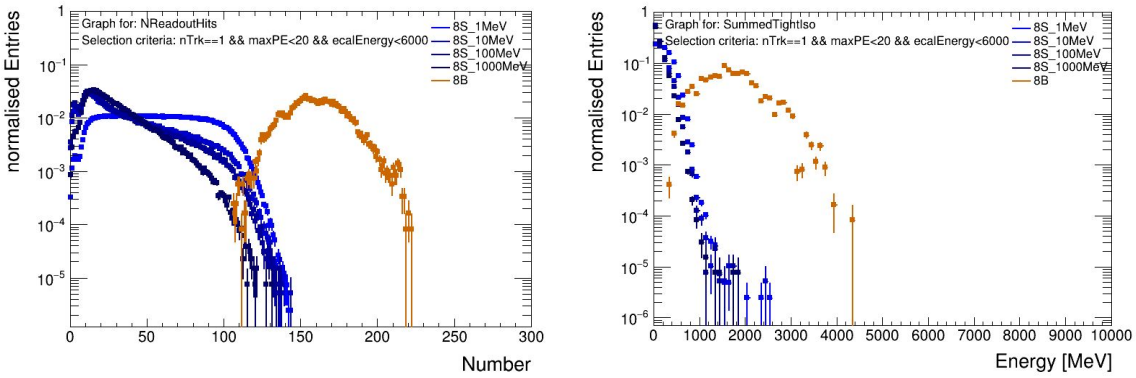


Figure set (13): NReadoutHits and SummedTightIso, for selection criteria Ecal energy < 6 GeV, maximum number of photoelectrons < 40, and number of tracks = 1. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$.

After applying both cuts, only 1 in 10 of the previous background events remained. Unfortunately, the other 10 BDT variables are rendered very inefficient for making additional cuts, some examples of which can be seen in Appendix B Figure (B.2). Therefore, they won't be used.

Results

The procedure outlined in the selection criteria optimisation section was followed.

The first step of the procedure was to start with a loose set of criteria. Five loose sets of initial criteria were chosen, and each one had selection criteria added until no apparent improvement was left to be made. The results can be seen in Table (1). As trial #3 was the one that eliminated all the backgrounds while maintaining a good efficiency, that one was chosen for further optimisation.

Trial #	Initial selection criteria	Final selection criteria	1 MeV $m_{A'}$ Signal efficiency (%)	Background efficiency (%)	Efficiency ratio
1	maxPE<20, ecalEnergy<6000, nTrk==1	NReadoutHits<125, hcalEnergy<14, maxPE<20, ecalEnergy<6000, nTrk==1	86.16	0.00077	132000
2	ecalEnergy<5000, nTrk==1	hcalEnergy<11, maxPE<14, ecalEnergy<5000, nTrk==1	78.06	0	NaN
3	maxPE<10, ecalEnergy<6000, nTrk==1	NReadoutHits<105, SummedTightIso<1300, maxPE<10, ecalEnergy<6000, nTrk==1	81.52	0	NaN
4	ecalEnergy<7000, nTrk==1	NReadoutHits<110, hcalEnergy<12, maxPE<10, ecalEnergy<7000, nTrk==1	84.94	0.00018	547000
5	maxPE<10, ecalEnergy<6500, nTrk==1	NReadoutHits<115, hcalEnergy<12, maxPE<10, ecalEnergy<6500, nTrk==1	86.37	0.00013	791000

Table (1): Initial and final selection criteria for each attempt, and resulting efficiencies.

The next step is implementing complementary variables. Since maxPE and hcalEnergy can complement each other, and the selection criteria did not have hcalEnergy, maxPE was loosened in small steps while hcalEnergy was tightened. This was done incrementally until it stopped yielding additional signal efficiency while still maintaining no background. The final step is loosening each selection criteria while still maintaining zero background efficiency. Several criteria were loosened, and it was found that the criteria for maxPE and SummedTightIso could be loosened into infinity, and therefore eliminated. The final selection criteria were then reached. A breakdown of these steps can be seen in Table (2)

Optimization phase	Selection criteria	1 MeV $m_{A'}$ Signal efficiency (%)	10, 100 and 1000 MeV $m_{A'}$ Signal efficiency (%)
Initial	NReadoutHits<105, SummedTightIso<1300, maxPE<10, ecalEnergy<6000, nTrk==1	81.52	79.10 68.66 64.55
MaxPE vs hcalE	maxPE<20, hcalEnergy<100, NReadoutHits<105, SummedTightIso<1300, ecalEnergy<6000, nTrk==1	82.07	79.79 69.24 65.42
	maxPE<30, hcalEnergy<22, NReadoutHits<105, SummedTightIso<1300, ecalEnergy<6000, nTrk==1	82.33	80.15 69.62 65.86
	maxPE<40, hcalEnergy<22, NReadoutHits<105, SummedTightIso<1300, ecalEnergy<6000, nTrk==1	82.50	80.56 70.03 66.25
	maxPE<100, hcalEnergy<14, NReadoutHits<105, SummedTightIso<1300, ecalEnergy<6000, nTrk==1	82.82	81.45 70.95 67.07
Loosening /eliminating	hcalEnergy<14, NReadoutHits<105, SummedTightIso<1300, ecalEnergy<6000, nTrk==1	82.82	81.46 70.96 67.09
	hcalEnergy<14, NReadoutHits<105, ecalEnergy<6000, nTrk==1	82.83	81.46 70.96 67.09
	hcalEnergy<14, NReadoutHits<106, ecalEnergy<6000, nTrk==1	83.29	81.61 71.04 67.10
	hcalEnergy<14, NReadoutHits<106, ecalEnergy<6274, nTrk==1	83.73	81.73 71.09 67.10
Final	hcalEnergy<15, NReadoutHits<106, ecalEnergy<6274, nTrk==1	83.77	81.88 71.26 67.26

Table (2): The steps taken after selecting trial 3. Each row represents a successive step, and shows the phase of the optimisation procedure they were in, the selection criteria, and the resulting efficiency for each $m_{A'}$.

The final selection criteria are: an Ecal energy under 6274 MeV, an Hcal energy under 15 MeV, a number of readout hits in the Ecal less than 106, and a single recoil electron track. The signal efficiencies are:
83.77% for 1 MeV DM, 81.88% for 10 MeV DM, 71.26% for 100 MeV DM, and 67.26% for 1000 MeV DM. The result of this method can be seen in Figure (14).

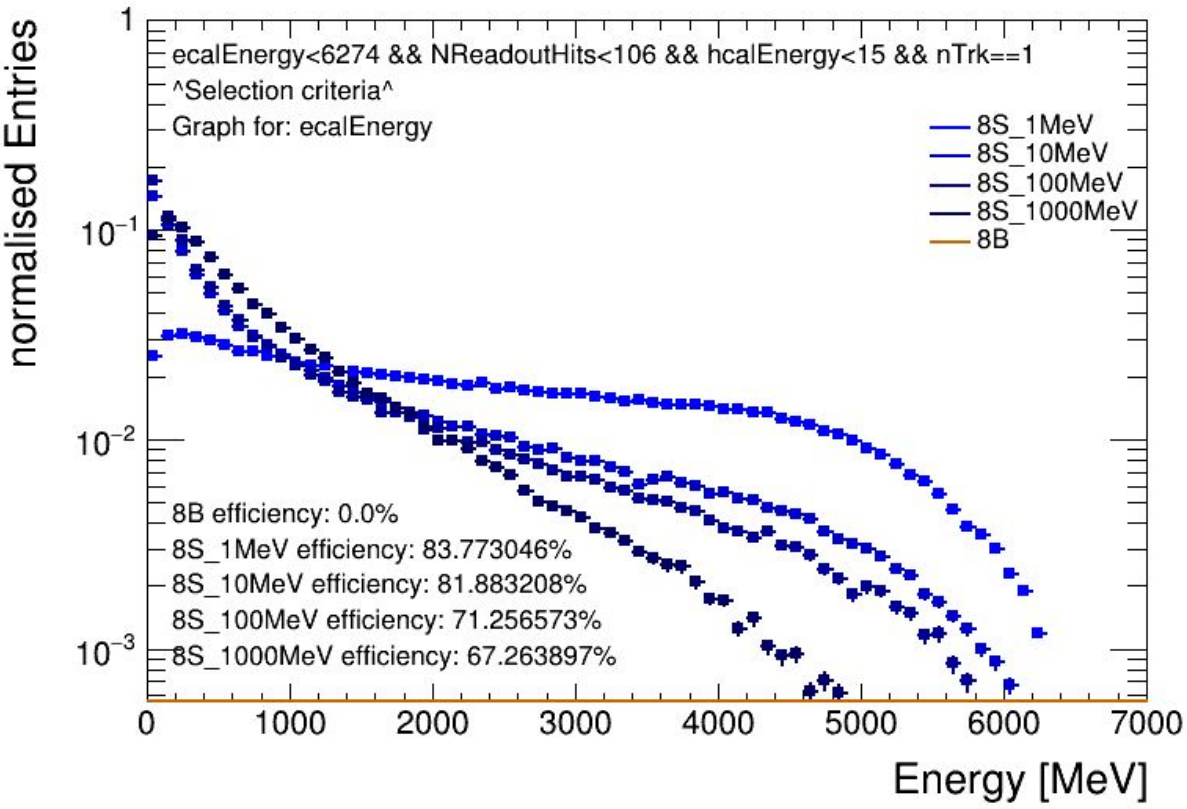


Figure (14): the Ecal energy distribution of the final selection criteria. Signal events are in blue, with darker colors indicating a heavier $m_{A'}$. No background events remain.

Discussion of final criteria

A very interesting development that arose from the final step of optimization - namely loosening the selection criteria - was that the criteria for the maximum number of photoelectrons and the sum of energy in isolated cells was made entirely redundant, and thus eliminated. This is probably because the effects that made the number of readout hits, number of tracks, Hcal energy, and Ecal energy relevant fully account for those criteria.

Another interesting development is that every step of optimizing for 1 MeV $m_{A'}$ also increased the efficiency for all other masses. This is because once a single electron recoil track is required, the signal samples behave rather similarly, in a way that increasing the efficiency for one leads to an increase in efficiency in all others. While for the loosening steps this is a mathematical certainty, this need not necessarily have been the case for balancing maxPE with Hcal energy. Since efficiency increases are so related across $m_{A'}$ masses, it is reasonable to assume that optimising for one $m_{A'}$ with the procedure in this paper leads to a "good enough for the scope of this paper" optimisation for all other $m_{A'}$.

The final selection criteria in this paper should not be used as the selection criteria for the LDMX, but rather a stepping stone. This is due to its numerous limitations and shortcoming.

First of all, the photonuclear background sample used only has about 1% of the number of background events that the LDMX will record during its Phase II run. This means that there will be several extreme events that were not accounted for in this sample, and hence not covered by these selection criteria. The selection criteria for a sample size of 100% that still filters all backgrounds will likely be more strict, and have a lower signal efficiency. Additionally, the Ecal energy criterion of exactly 6274 MeV is background sample specific, and would vary somewhat even for different 1% background samples. The other sample criteria might also be somewhat sample-specific, albeit to a lesser extent.

Secondly, the use of a BDT would likely yield a better signal efficiency and smarter selection criteria, while still rejecting all backgrounds. This is because a BDT evaluates each event individually, whereas the method used here applies broad cuts to every event. However, the extent of the benefit of better selection criteria is limited, given the requirement of a single track already cuts the signal efficiencies down significantly. Contrasting this with the final selection criteria's signal efficiencies, we get a best-case scenario improvement shown in Table (3).

$m_{A'}$ (MeV)	Signal efficiency of nTrk = 1 (%)	Signal efficiency of final selection criteria (%)	Maximum possible improvement (%)
1	94.37	83.77	10.6
10	84.88	81.88	3
100	72.71	71.26	1.45
1000	68.07	67.26	0.81

Table (3): The maximum possible improvement of alternative selection criteria. This is calculated by subtracting the signal efficiency of the final selection criteria from the signal efficiency of the criterion nTrk = 1.

The third limitation of the selection criteria is rooted in how realistic the simulations are. There might be subtle physical phenomena the simulation has not accounted for, or hidden instrumental errors, or limitations to calibrations, or any number of other things that could go wrong. These will have to be addressed with thorough testing once the LDMX is built and operating, followed by a newly optimised set of selection criteria once the specifics are known.

A final limitation to the selection criteria is non-photonuclear backgrounds. While they should be significantly less challenging to weed out, and hence why they were not targeted in this paper, they will still need to be filtered from the final set of signal candidates. This requires a change in the selection criteria. Most notably, it is quite likely that the selection criterion for Ecal energy will be significantly lower than 6 GeV, since the vast majority of non-photonuclear background events lie at energies above 4 GeV, and the trigger based veto makes the rejection of these events very efficient and economical.

Conclusion

The aim of this thesis was to establish a set of selection criteria for an 8 GeV beam energy at the LDMX. This consisted of exploring the already established 4 GeV selection criteria, and adding new and alternative ones, such as the energy deposited in the hcal and the use of individual BDT variables instead of a BDT confidence value. Then, different selection criteria were procedurally applied to a sample of ~17 million simulated events to obtain a final set of selection criteria. The selection criteria were: an Ecal energy under 6274 MeV, an Hcal energy under 15 MeV, a number of readout hits in the Ecal less than 106, and a single recoil electron track.

Outlook

While this paper establishes a set of selection criteria for an 8 GeV electron beam LDMX, this set of selection criteria should only be used as guidance to create a better set. To establish a final set of selection truly fit for finding light dark matter, several things will have to be done. First, the LDMX must be built and tested, and the simulations adjusted to these results. Second, the simulated photonuclear background sample size must be at least the same size as the total expected number of photonuclear events during Phase II of the LDMX. Third, a boosted decision tree must be implemented to efficiently filter backgrounds. Finally, the exclusion of non-photonuclear background events must also be implemented into the selection criteria. Once all of that is established, the identification of dark matter at the Phase II of LDMX can finally begin.

References

[1] Åkesson, Torsten, et al. "Light Dark Matter eXperiment (LDMX)." *arXiv preprint arXiv: 1808.05219* (2018).

[2] Åkesson, Torsten, et al. "A High Efficiency Photon Veto for the Light Dark Matter eXperiment." *arXiv preprint arXiv:1912.05535* (2019).

[3] Dunbar, Brian. “Planck Mission Brings Universe Into Sharp Focus.” NASA, NASA, 7 June 2013, www.nasa.gov/mission_pages/planck/news/planck20130321.html.

[4] Kane, Gordon. Modern Elementary Particle Physics. 2nd Ed., Cambridge Univ Press, 2017.

[5] Shears, Tara. “The Standard Model.” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 370, no. 1961, 2012, pp. 805–817., doi:10.1098/rsta.2011.0314.

[6] “Dark Energy, Dark Matter.” NASA, NASA, 18 Mar. 2018, science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy.

[7] Kirsanov, Mikhail. “Recent Results of the NA64 Experiment at the CERN SPS.” CERN Document Server, 15 Nov. 2019, cds.cern.ch/record/2701581/.

[8] Lanfranchi, Gaia. “SHiP Physics Reach .” CERN Document Server, October. 2017, https://indico.cern.ch/event/644961/contributions/2714850/attachments/1538989/2412733/SHiP_Physics_Reach_Lanfranchi.pdf.

[9] Duston, Adam, and Tim Nelson. “Light Dark Matter Experiment - Light Dark Matter Experiment.” SLAC Confluence, SLAC, 26 Sept. 2019, confluence.slac.stanford.edu/display/MME/Light+Dark+Matter+Experiment.

[10] Agostinelli, S., et al. “Geant4-a Simulation Toolkit.” Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, North-Holland, 11 June 2003, www.sciencedirect.com/science/article/pii/S0168900203013688.

[11] Chytracek, Radovan & McCormick, Jeremy & Pokorski, Witold & Santin, Giovanni. (2006). Geometry Description Markup Language for Physics Simulation and Analysis Applications. Nuclear Science, IEEE Transactions on. 53. 2892 - 2896. 10.1109/TNS.2006.881062.

[12] Alwall, J. et al. “The Automated Computation of Tree-Level and Next-to-Leading Order Differential Cross Sections, and Their Matching to Parton Shower Simulations.” Journal of High Energy Physics 2014.7 (2014): n. pag. Crossref. Web.

[13] Rene Brun and Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/

[14] Figure 2 in Åkesson, Torsten, et al. "A High Efficiency Photon Veto for the Light Dark Matter eXperiment." *arXiv preprint arXiv:1912.05535* (2019).

[15] Figure 17 in Åkesson, Torsten, et al. "Light Dark Matter eXperiment (LDMX)." *arXiv preprint arXiv: 1808.05219* (2018).

[16] Figure 16 in Åkesson, Torsten, et al. "Light Dark Matter eXperiment (LDMX)." *arXiv preprint arXiv: 1808.05219* (2018).

[17] Figure 4 in Åkesson, Torsten, et al. "A High Efficiency Photon Veto for the Light Dark Matter eXperiment." *arXiv preprint arXiv:1912.05535* (2019).

Appendix A

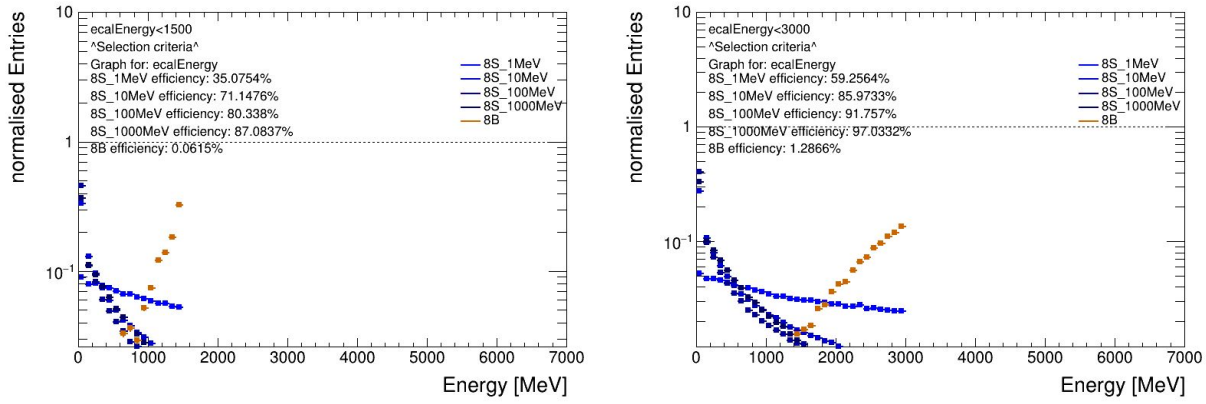


Figure set (A.1): Ecal energy plots for 1.5 GeV (left) and 3 GeV (right) ecalE selection criteria for 8 GeV beams. Efficiencies for each signal $m_{A'}$ variant and background shown in top left of each diagram. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$.

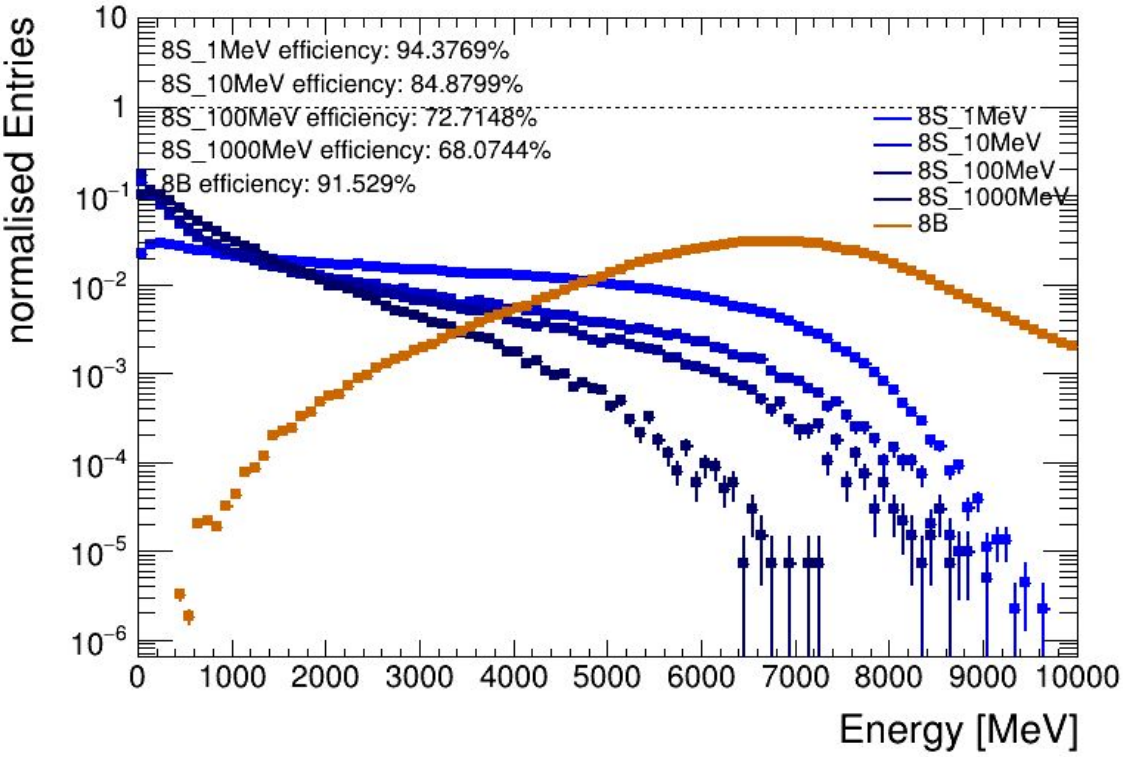


Figure set (A.2): Ecal energy plots for nTrk = 1 selection criteria for 8 GeV beams. Efficiencies for each signal $m_{A'}$ variant and background shown in top left. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$.

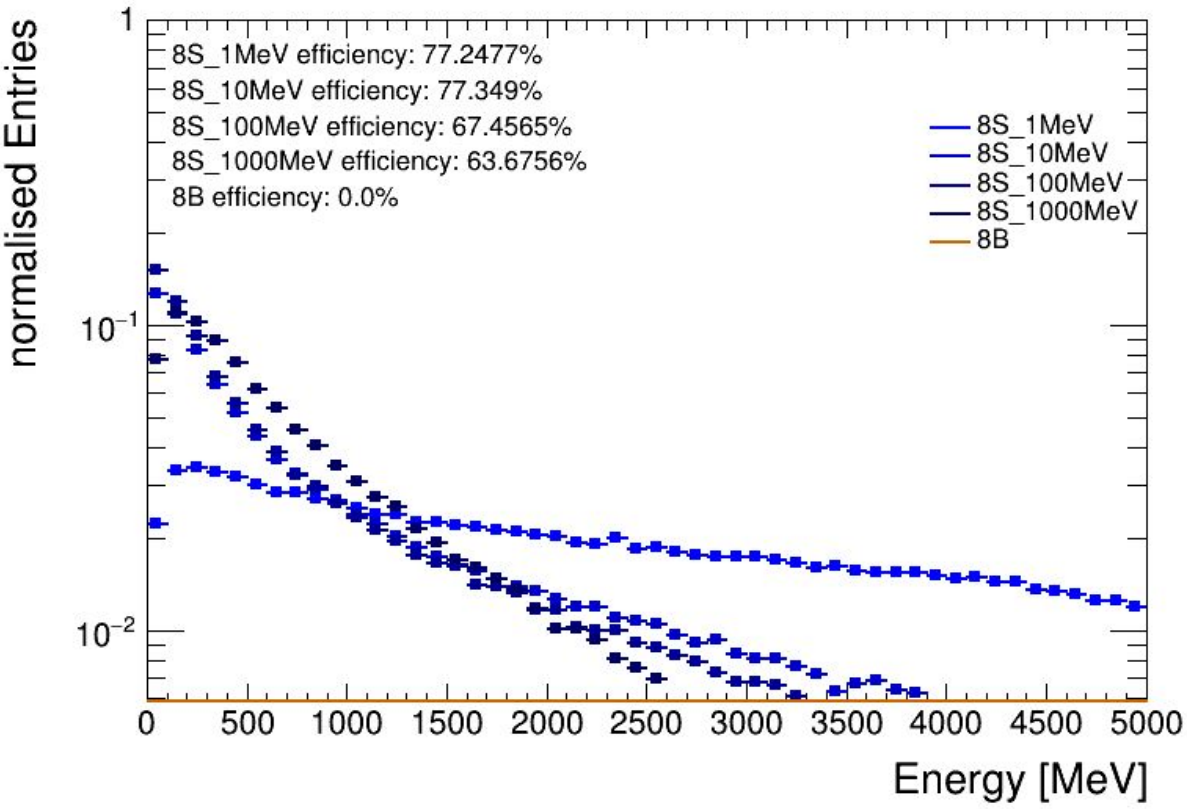


Figure set (A.3): the resulting Ecal energy distributions for nTrk = 1, MaxPE < 6, EcalE < 5 GeV selection criteria for an 8 GeV beam. Efficiencies for each signal $m_{A'}$ variant and background shown in top left. Signal events are in blue, with darker colors indicating a heavier $m_{A'}$. No background events remain.

Appendix B

Full description and plots BDT of variables

1. NReadoutHits: The total number of cells that were hit.
2. SummedTightIso: The sum of energy from isolated Ecal cells. Isolated cells are cells that show a readout, but their neighbors do not.
3. ShowerRMS: Represents the general width of the showers in the Ecal.
4. XStd: Represents the horizontal width of the showers in the Ecal.
5. YStd: Represents the vertical width of the showers in the Ecal.
6. DeepestLayerHit:
7. MaxCellDep: The maximum depth that each shower goes into the Ecal?
8. AvgLayerHit: The average depth that each shower goes into the Ecal?
9. StdLayerHit:
10. SummedDet: This is just the Ecal energy.
11. passesVeto: Indicates whether or not an event passed the first three 4 GeV selection criteria, namely an Ecal energy of less than 1.5 GeV, a single electron recoil track, and a maximum of 4 photoelectrons created in each bar of the Hcal. This is irrelevant to 8 GeV events.
12. Disc: The confidence value that the BDT gives after taking the other 11 variables into account. This is not implemented for 8 GeV.

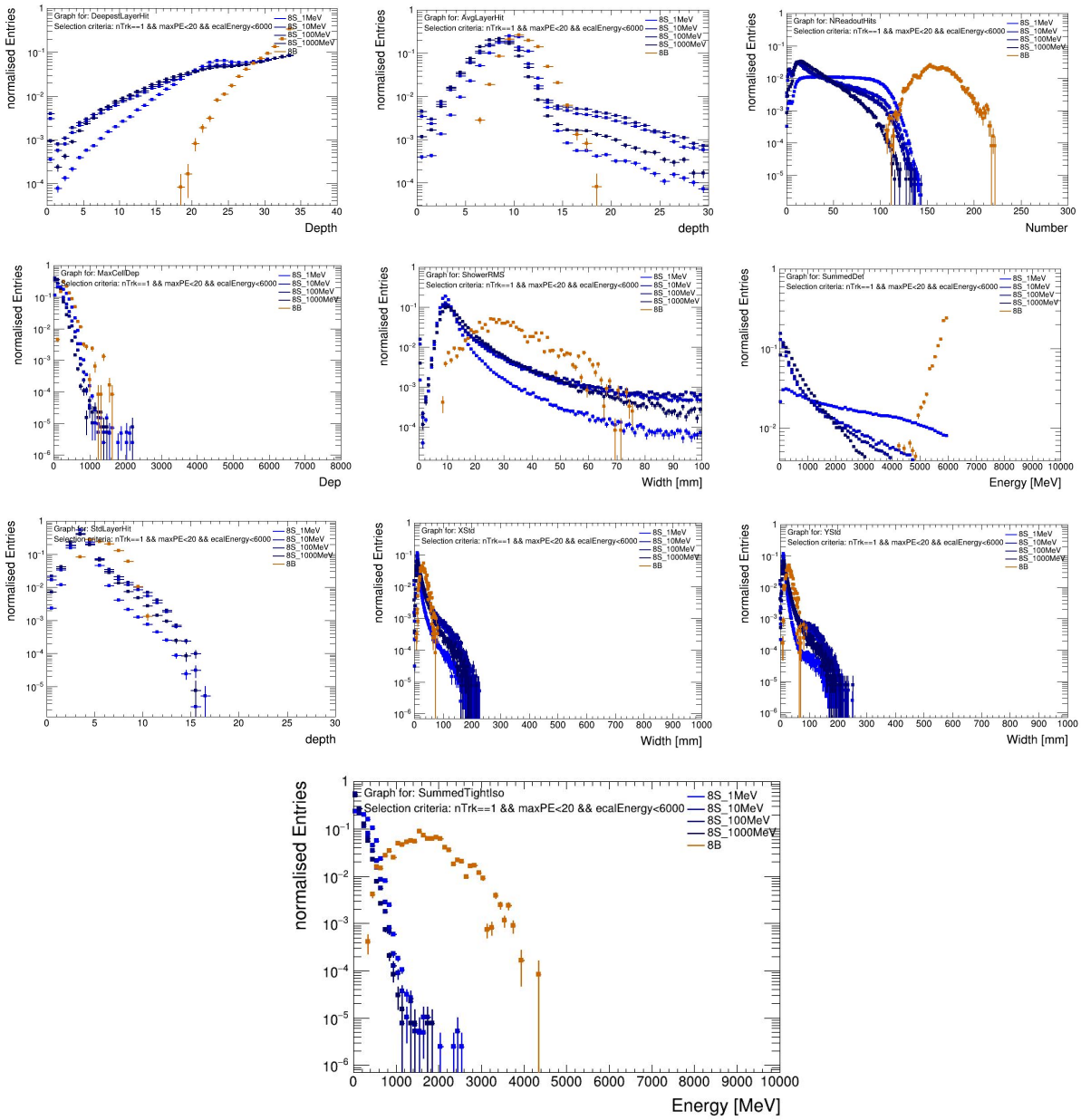


Figure set (B.1): A plot for each of the BDT variables (indicated top left) after applying the loose set of selection criteria nTrk = 1, maxPE < 20, ecalEnergy < 6000 MeV. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$. Disc and PassesVeto are not included due to missing implementation and irrelevance, respectively.

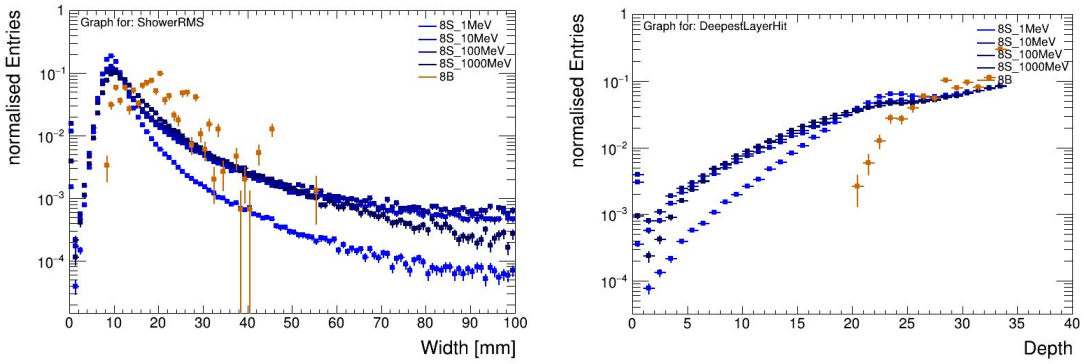


Figure set (B.2): A plot for some of the BDT variables (indicated top left) after applying the selection criteria NReadoutHits < 145, SummedTightIso < 1500 MeV, nTrk = 1, maxPE < 20, ecalEnergy < 6000 MeV. Background events are in orange, while signal events are in blue, with darker colors indicating a heavier $m_{A'}$.