University of California

Santa Barbara

Exploring Visible Decay Scenarios In The Light Dark Matter eXperiment

A dissertation submitted in partial satisfaction of the requirements for the degree of

Bachelor of Science

in

Physics

by

Juan Manuel Lazaro Ruiz

Committee in charge:
Professor Joseph Incandela

2021

Acknowledgements

Contents

Abstract

The Light Dark Matter eXperiment (LDMX) will use a fixed-target approach to search for dark matter in the sub-GeV mass range. LDMX has demonstrated the effectiveness of an ECal-based boosted decision tree (BDT) for achieving sensitivity to the invisible signal of a dark photon. I propose to study the effectiveness of expanding our current machine learning models to use the HCal and other subdetectors to extend sensitivity to visibly decaying dark photons and axion-like particles, as well as to improve invisible signal efficiency.

This document outlines the first steps towards these goals. I begin with introductions to the standard model, dark matter, and LDMX. I cover the generation of the new visible signal samples needed for this study, followed by descriptions of the extended BDTs designed to separate said signal from its backgrounds. An analysis of these BDTs is carried out for invisible dark photon signal as well. Results are promising for invisible signal, as one new BDT increases the efficiency of the lowest signal mass point from $\sim$88% with an ECal BDT to over 99% at the same, 10^{-4}, background efficiency. Similarly, signal efficiencies for visible signal at a background efficiency of 10^{-4} are above 99% in the phase space surveyed thus far.

Chapter 1

Introduction

We will begin with a brief introduction to the Standard Model (SM) of particle physics. In combination with some astrophysical and cosmological observations, this will motivate the search for a new form of matter, dark matter, which has yet to be directly observed due to its very weak or null coupling to the SM. I will then briefly discuss some proposed extensions to the SM. This will be followed by an overview of the Light Dark Matter eXperiment which will be searching for dark matter of the kind in the aforementioned extensions. Finally, we introduce boosted decision trees as a tool for rejecting difficult SM backgrounds to invisible dark matter signals, and how existing BDTs could be extended to improve signal efficiency and cover visible signals.

1.1 The Standard Model

The SM is well known for its mixture of complexity, elegance, and precision. These features are a direct result of its mathematical formulation and the interpretation of that formulation. The SM is a quantum field theory which means that what we call fundamental particles are excitations from a vacuum state of fields which pervade the universe. The behavior of the fields are governed by the principle of least action, which states that the integral over some time interval of the Lagrangian, or Lagrangian density $\mathscr{L}$ (this is typically what is meant by Lagrangian), must be a local extremum. The fields, and hence particles, will simply do what it takes to make this true. For example, the terms of the SM Lagrangian relating to the electron an photon, excluding mass-related terms, are

$$\mathscr{L} \supset \overline{f_e} i \not{D} f_e - q_e \overline{f_e} \gamma^\mu f_e A_\mu - \frac{1}{4} F^{\mu\nu} F_{\mu\nu}. \tag{1.1}$$

where f_e is the electron field, A_μ is the photon field, and $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. For this discussion, it is the second term which will be of most interest. Since it involves both fields, it determines the interactions between the electrons and photons. In other words, the electron's non-zero electromagnetic charge q_e allows the electron to (couple) influence and be influenced by the photon which we refer to as the quantum electrodynamic (QED) mediator. Similar terms exist for the muon, tau lepton, and quarks. Fig 1.1 shows all the particles of the SM in their standard categorizations.
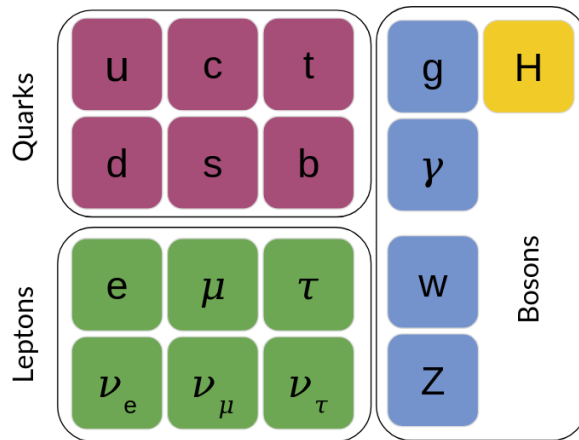
Figure 1.1: A summary of the particles in the Standard Model.

1.2 Dark Matter

1.2.1 Why dark matter is necessary

Despite its success, there are several questions that the SM leaves unanswered. Two of these questions relate directly to Fig 1.1. Recently, the Muon g-2 experiment at Fermi National Accelerator Laboratory has reinforced the anomaly in the muon g-2 value found at Brookhaven National Laboratory [1]. This value characterizes the influence of virtual particles, which are those that arise from the vacuum as allowed by the Heisenberg uncertainty principle, on a real muon. The SM correctly predicts a non-zero value, but the larger experimental value suggests that there is something unaccounted for in the theory.

Another question, which originates in astrophysics and becomes relevant to particle physics, has to do with the rotational velocity curves of spiral galaxies. As seen in Fig 1.2 (Left), the outer edges of galaxies are rotating much faster than what is predicted by Newtonian gravity, which is an appropriate approximation on this scale, given the distribution of observable matter. There have been attempts to explain these curves with modifications to gravity. However, these modifications have trouble explaining other astrophysical phenomena, such as the bullet cluster shown in Fig 1.2 (Right). The bullet cluster is the result of a collision between two galaxies. Using gravitational lensing, which is predicted by general relativity, and infrared telescopes, we can see that most of the mass from the two galaxies does not coincide with most of the observable matter which has been heated by the collision [4].

Particulate dark matter (DM), so named due to it's lack of a strong enough QED coupling for easy direct observation, could be the solution to all of these problems. Just as virtual electrons and positrons contribute to the muon g-2 value, the same could be true of virtual DM particles, and they could thus be the cause of the g-2 anomaly. The rotation curves of galaxies could be explained by DM halos which speed up the matter on their outer edges. Finally, while the bullet cluster may set constraints on the DM self-interaction strength, particulate DM does not suffer from the same fine-tuning problems that modified gravity does when faced with similar phenomena.

The broad definition of dark matter means that there are many avenues to explore. The most familiar possibilities for dark matter are massive compact halo objects (MACHOs) which may include brown dwarf stars, rogue planets, and primordial black holes. Such massive objects may be hard to detect through conventional astronomical surveys, and if they are much more common than current estimates, they could account for some or all dark matter. Some theorized classes of fundamental dark matter, such as weakly interacting massive particles (WIMPs), have received a lot of attention for decades due to their connection to supersymmetry and the relative accessibility of conducting relevant searches. In comparison, light dark matter, so named for its low mass relative to WIMPs' masses, has only recently begun receiving increased attention due to persistent negative results for WIMPS and new technologies making searches for this class of dark matter feasible. In particular, the Light Dark Matter eXperiment makes use of old technologies in a new way. This is expanded upon in Sec. 1.3.
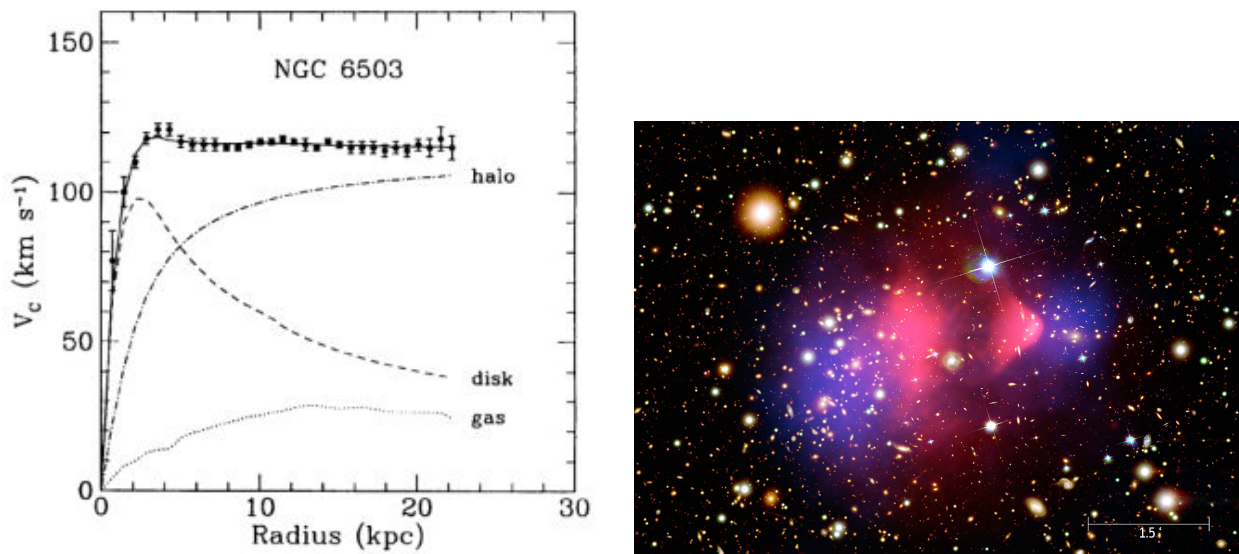

Figure 1.2: Left: Rotations curves for the galaxy NGC 6503 showing that the sum of contributions from the the gas, disk, and a dark matter halo fit the observed data [2]. This assumes Newtonian gravity which is a good approximation to general relativity at this scale. Right: An X-ray map (pink) showing the heated matter as a result of the colliding clusters that now make up galaxy cluster 1E0657-56 and gravitational lensing map (blue) indicating the distribution of mass superimposed on a visible light image of the same part of the sky [3, 4].

1.2.2 Light dark matter (LDM)

Henceforth, the focus of this document will be on light dark matter (LDM) since, as previously mentioned, it is relatively unexplored on the experimental end of high energy physics. The lower bound on LDMX's search comes from considering the intersection between LDM and thermal relic dark matter. Thermal DM describes a class of DM which reached thermal equilibrium in the early hot universe, being created and self-annihilating at equal
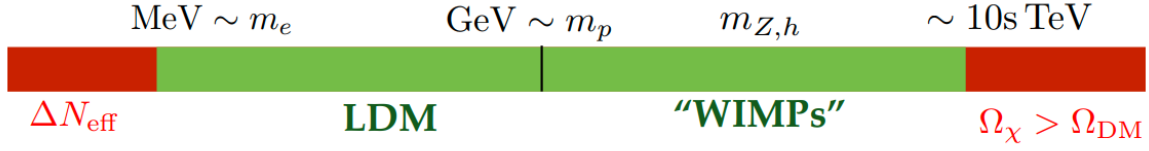
Figure 1.3: Thermal DM mass ranges. This highlights that one reason for interest in LDM is that it is in the same range as much of stable SM matter. Upper and lower bounds are imposed by cosmological observation [6].

rates. As the universe expanded and cooled, production processes would become negligible due to particles lacking the required energy. Annihilation would become negligible when the number density of dark matter dropped such that DM collisions became exceedingly rare. From this idea, thermal dark matter mass is constrained such that these processes occur at the right times to produce what we observe today [7]. The same processes that allowed dark matter to thermalize in the early universe could be also used to discover it today.

1.2.3 Dark photons

There are many dark photon models which all share the defining feature of a new guage boson (A') with a new broken $U(1)_D$ symmetry, hence the similar name to the QED mediator. This is appealing as QED is simple and well understood relative to the rest of the SM. A broken symmetry would not be new to the SM either. Minimal models simply couple A' to the SM via kinetic mixing as shown below, while more complex models also introduce a dark fermion (χ) which interact through the A' field. The mass and density of dark matter are of great cosmological importance. The χ-A' mass ratios are also important for understanding all the processes allowed by these models. Nonetheless, it is the mediator that demands more attention for experimental considerations.

The dominant term we are concerned with in this family of models is

$$\mathscr{L} \supset -\epsilon q_e \overline{f_e} \gamma^\mu f_e A'_\mu \tag{1.2}$$

where epsilon is a kinetic mixing parameter. This allows the dark photon to interact with the SM electron without giving it a $U(1)_D$ charge or giving the A' an electromagnetic charge.

1.2.4 Axion-like particles (ALPs)

Axion-like particles are an extension of QCD (quantum chromodynamics) axions which were originally conceived to solve the strong CP (charge-parity symmetry) problem. The models described by

$$\mathscr{L} \supset \frac{1}{4\Lambda_\gamma} a F^{\mu\nu} \widetilde{F}_{\mu\nu} + \frac{\partial_\mu a}{\Lambda_e} \overline{e} \gamma^\mu \gamma^5 e \tag{1.3}$$

were the original motivation for the hadron calorimeter based machine learning described in Sec. 2.4.1.
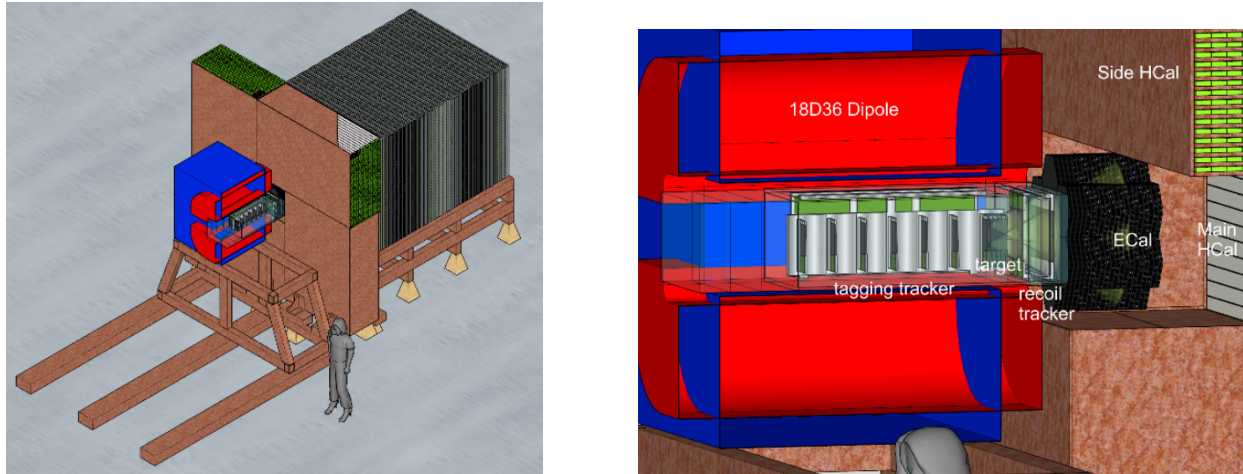
Figure 1.4: Left: A rendering of the complete LDMX detector [8]. Right: A vertical cross section of the subdetectors upstream up the ECal and Side HCal [8]. Note that the ECal will be rotated 90° from what is shown to gain some acceptance of recoil electrons, which will be deflected right by the downward magnetic field.

1.3 Light Dark Matter eXperiment (LDMX)

LDMX will be a fixed-target experiment beginning operation with a 4 GeV electron beam which will later be raised to 8 GeV. The full detector is shown in Fig. 1.4. It will focus on invisible signals from minimal dark photon models. This includes those where the dark photon is stable or tends to decay into DM, as well as models that are in a region of phase space where the dark photon decay length is longer than the LDMX detector. For this reason, it is designed to be able to precisely identify missing momentum.

A brief description of the expected processes in a signal event in the invisible scenario is as follows. A beam electron would radiate a dark photon in a bremsstrahlung-like process allowed by Eq. 1.2. The dark photon and recoil electron would then proceed through the recoil tracker, electromagnetic calorimeter (ECal), and hadron calorimeter (HCal), but only the electron would shower. As the ECal is highly granular and several radiation lengths long, it is expected that it would only fail to reconstruct the beam momentum when an A' has been produced, or in the case of a few rare backgrounds. These backgrounds and their veto handles are thoroughly documented in [8].

1.4 LDMX machine learning

The LDMX collaboration has used the XGBoost boosted decision tree (BDT) algorithm to develop a veto for photo-nuclear background that only uses ECal information [8]. BDTs are an ensemble of decision trees, which themselves are a series of binary classifiers, as shown in Fig. 1.5, which assign a likelihood to the input. The second and subsequent trees seek to reduce a loss function of the pseudo residuals from previous trees. The final output is the sum of individual outputs and ranges, in our use case, from 0 to 1 for background and signal respectively. For further explanation, see the original XGBoost paper and documentation
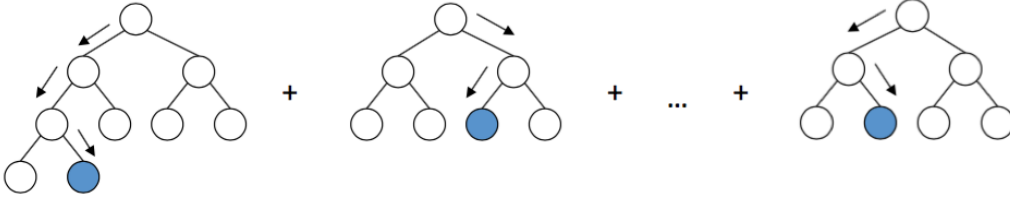
Figure 1.5: A boosted decision tree (BDT) schematic. This illustrates how inputs to the model are evaluated by small decision trees, which all contribute to a final discriminator value indicating the likelihood of said input belonging to one class versus another.

[10, 11]. BDTs offer a few benefits over other machine learning models. They are relatively simple and rigid models. This means that although they can be out-performed by models such as neural networks, if care is taken in the selection of variables, it is much easier to intuitively understand how they "make decisions." Also, due to their simplicity, they are less computationally expensive to develop.

1.4.1 Bias considerations

An important feature of this ECal BDT is that it aims to minimize transverse momentum ($|\vec{p_T}|$) bias. As seen from Fig. 1.6 (Left), heavier A's tend to recoil more strongly. By avoiding $|\vec{p_T}|$ bias, we can preserve the shape of the $|\vec{p_T}|$ distribution of any candidate events. $|\vec{p_T}|$ bias would shift these distributions closer together, which implies a loss of information. At the end of the experiment, we can then fit the observed distribution to the distribution of simulated events with known mass, therefore determining the m'_A that would most likely lead to the observed $|\vec{p_T}|$ distribution. Fig. 1.6 (Right) shows the results of this as done for invisible signal.

The kinematics of visibly decaying A's are similar, so $|\vec{p_T}|$ bias will still be a concern for BDTs made for visible signal. However, for visibly decaying A' and ALPs, we are also concerned with decay length ($\gamma c\tau$) bias as $\gamma c\tau$ would also carry information about $m_{A'}$ and ϵ. Predictions of visible signal $m_{A'}$ and ϵ will not be included in this document, but chapter. 3 does include bias plots, which will be useful to be aware of in the future when we do need to make predictions. Variables that are independent of both $|\vec{p_T}|$ and $\gamma c\tau$, mostly related to shower shape, will be useful in both scenarios. This suggests a symbiotic development process is possible for an extension to the existing ECal BDT for invisible signal and a new BDT focused on a visible signal search.

1.4.2 Current state

In a recent LDMX collaboration paper [8], we demonstrated the performance of an ECal BDT. The variables used can be grouped into two broad categories. First, there are global variables such as the number of readout hits, standard deviation of hit positions in x and y,
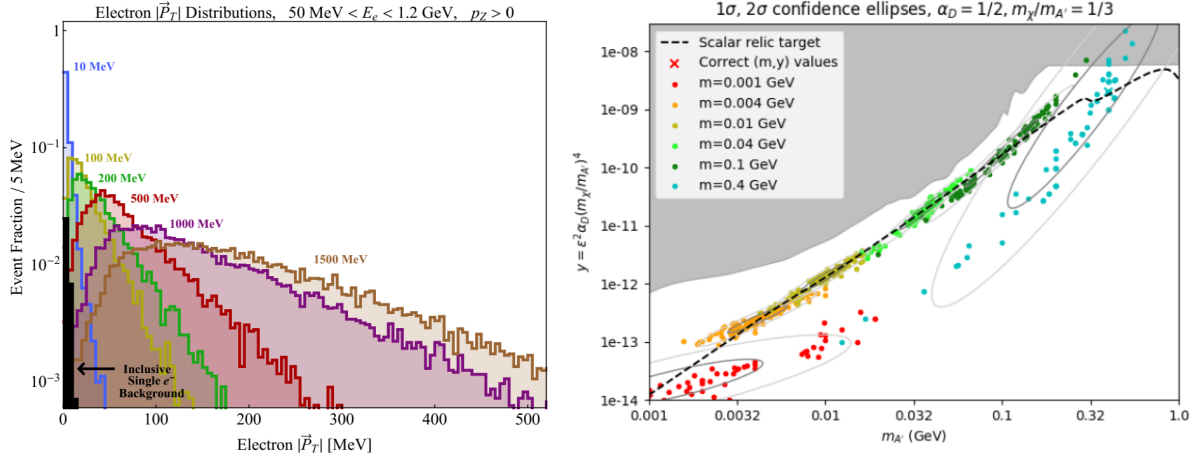
Figure 1.6: Left: $|\vec{p_T}|$ distributions for a few $m_{A'} \in [10 \text{ MeV}, 1000 \text{ MeV}]$ and inclusive single electron background [8]. Right: Mass and coupling predictions from fitting a sample of unknown $m_{A'}$ with its $|\vec{p_T}|$ and total ECal energy distributions. ϵ is determined from the number of events in the sample.

and the average layer number. In addition, the ECal BDT uses variables based on radii of containment (RoC) around the recoil electron trajectory and the projected photon trajectory. RoC are defined at each layer of the ECal such that they contain, on average, 68% of an electromagnetic (EM) shower's energy. Variables such as the energy within the first radius of containment around the electron give information about the presence of an EM shower where we would expect one in the case of difficult to reject background. The performance of this BDT is shown in Fig. 1.7 via receiver operating characteristic (ROC) curves.

There has since been two enhancements under development. The first of these was incorporating the minimum ionizing particle (MIP) tracking, which was previously a separate cut, into the ECal BDT. Although all background in a 2.1×10^{14} EoT equivalent sample was already rejected with this method [8], including it in the BDT may increase the overall signal efficiency. Further, although RoC are defined at each layer, the related quantities were summed over all layers, destroying all z-related shower information. To address this, RoC variables are now calculated in three segments: layers 1 to 6, layers 7 to 17, and layers 18 to 34.
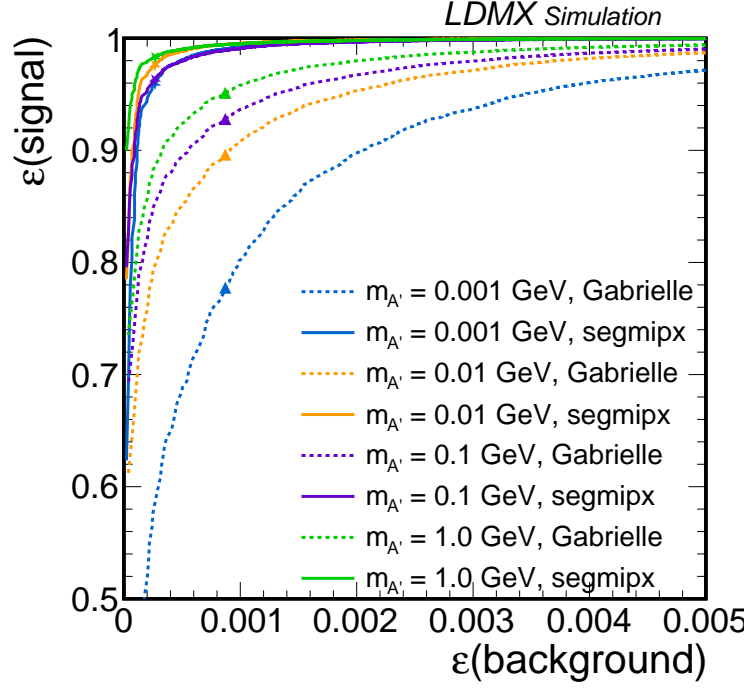
Figure 1.7: Receiver operating characteristic (ROC) curves for invisible A' signal and ECal PN efficiencies (ε) produced by making cuts at a range of BDT discriminator values. The markers indicate the efficiencies at discriminator values of 0.99 which was used in the final veto sequence [8]. Gabrielle is the ECal BDT used in [8] and segmipx is the BDT with longitudinal segmentation and MIP Tracking discussed in Sec. 1.4.2. Note that these samples have already passed the trigger, which cuts events with greater than 1.5 GeV deposited energy in the first twenty layers of the ECal.

Chapter 2

Methods

We begin this chapter by determining the model, $A' \rightarrow e^+e^-$, and regions of phase space on which to carry out a complete analysis based on initial reach predictions. We then describe the production method for samples of this model. Similarly, this chapter contains a review of the relevant backgrounds and their production methods. With samples prepared, we proceed with the development of a new trigger which is more suitable to our chosen signal than the existing trigger for an invisible signature. Finally, we will review the development process of a new BDT for this signal that will utilize other subdetectors in addition to the ECal. The extension of the ECal BDT for invisible signal will receive a shorter review.

2.1   Signal

We will use the minimal A' decaying to e^+e^- model as our point of entry to a full visible signal analysis. Physically, this is more simple than any model involving a DM fermion and therefore is also easier to simulate. The same is true of the comparison to ALP models, which have multiple decay channels. As shown in Fig. 2.1, the branching fraction of A' to e^+e^- is 1 over the entire $m_{A'}$ reach of LDMX. This figure also shows that, unfortunately, LDMX will not be able to explore any new phase space in its first run with a beam energy of 4 GeV. Accordingly, we start this analysis for visible signal with a beam energy of 8 GeV.

The production process for samples of this model is currently done in three steps. The kinematics of the dark bremsstrahlung and decay are handled using a custom MadGraph5 model, and are output to LHE (Les Houches Event) files. In order to alleviate rounding errors for low A' masses in MadGraph5, all input parameters with dimensions of energy are scaled up by a factor of 10. The LHE files are then rescaled to their physical values and split into two files, one containing the bremsstrahlung vertex and one containing the decay vertex, which are both generated at this stage using the momenta from MadGraph5. Though it is possible to generate the decay with a physically accurate decay length distribution for a given ϵ, in practice, we sample decay lengths from a uniform distribution in a portion of the detector to maximize the utility of the LHEs from MadGraph5. Fig. 2.3 illustrates the result of this simulation process. This idea is expanded upon in sections 2.3 and 2.4.2. Each LHE then initializes its particles as primaries in a custom Geant4 framework for the full detector simulation [8]. In the future, this will be simplified and improved by moving the choice of

the decay vertex into the custom framework, thus eliminating the file splitting.

2.2   Background

All background samples were generated through the custom LDMX version of Geant4 [8]. An inclusive sample was made for the development of the trigger as discussed in Sec. 2.3 and for quantitatively determining backgrounds that will require specialized attention.

A common difficult background for visible and invisible signals are events with a hard bremsstrahlung in the target followed by a photo-nuclear (PN) process in early in the ECal. This can mimic the displaced shower of a visible A', or ALP, decay since the hadronic products would deposit very little energy in the ECal. The energy loss of charged hadrons through matter as they approach a stop is proportional to the inverse square of their speed [15], so the majority of the energy that these hadronic products deposit will be near the back of the ECal or in the HCal. The result is the data collected by the detector looking very similar to the data collected in the signal event depicted in Fig. 2.2.

Another background of concern is bremsstrahlung in the target where the photon converts to an electron-positron pair late in the ECal. Though the ECal is 40 radiation lengths (X_0) long [6, 8], the rate of these events means it is a visible background at 10^{16} electrons on target (EoT). Further, the photon will produce an EM shower, just as we expect from the visible A' signal. This contrasts the hadronic showers of the ECal PN background, which have different profiles in each of the subdetectors and may therefore be easier to identify.

2.3   Trigger

The standard LDMX trigger for invisible A' signal cuts events that deposit greater than 1.5 GeV in the first twenty layers of the ECal. This is problematic for a visible signal search as, if we wanted to consider signal in the entire ECal, it would create a clear bias for higher decay lengths, as earlier decays would deposit more energy in the layers considered by the trigger and would consequently have lower efficiencies. One potential workaround is to use information from the recoil tracker to reject events with more tracks than the number of electrons indicated by the tagging tracker. A similar cut is part of the invisible signal veto sequence. However, this could require upgrades to the tracker readout to use as a trigger. Thus, for this analysis, we will use a trigger similar to the one in [8] and consider only signal with decay lengths beyond layer 20 of the ECal. This equates to shrinking the potential reach by raising z_{min} in the following equation from [5].

$$N_{\text{sig}} = N_{A'} \times \left(e^{-z_{\min}/\gamma c\tau_{A'}} - e^{-z_{\max}/\gamma c\tau_{A'}} \right) \tag{2.1}$$

$N_{A'}$ is the total number of A' produced and $\gamma c\tau_{A'}$ is the typical decay length for the given $m_{A'}$, ϵ, and beam energy. Expressions for these are also given in [5]. All z values are meaured from the target. Fig. 2.4 (Left) shows the reduction that results from raising z_{min} from 32 cm, which was used in Fig. 2.1 to 48 cm, between layers 20 and 21 of the ECal. Note that this reduced reach with an 8 GeV beam is still greater than the reach with a 4 GeV beam using $z_{\min} = 32$ cm. Fig. 2.4 (Right) shows the efficiencies of a few signal masses for events, where
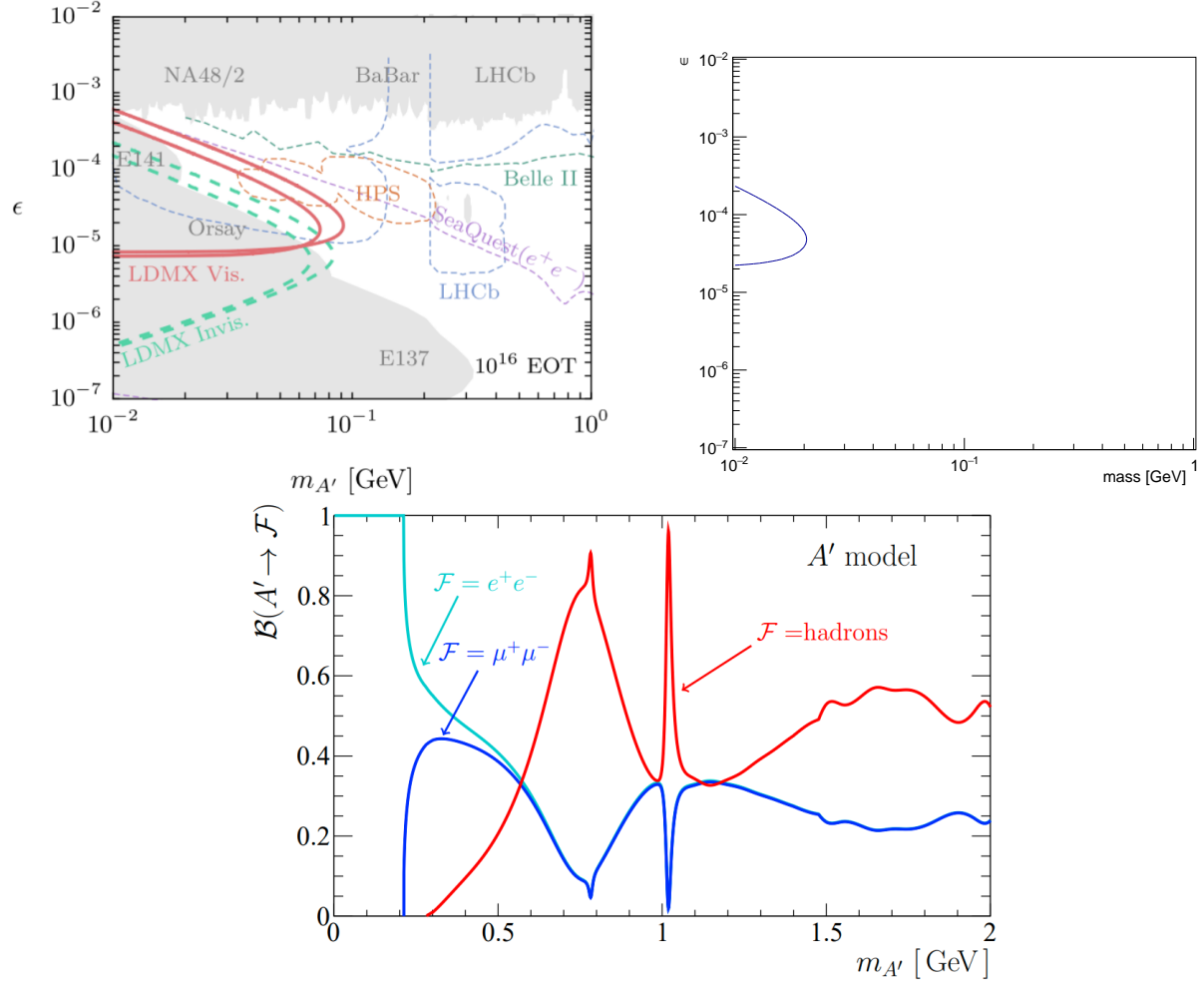
Figure 2.1: Top Left: Projected reach for LDMX and other dark photon searches in $m_{A'}$-ϵ space. The shaded regions have been previously excluded. The smaller red contour corresponds to an 8 GeV beam, while the larger contour corresponds to a 16 GeV beam [5]. Top Right: Projected reach for LDMX in Phase I, using a 4 GeV electron beam and having 10^{14} EoT. Comparing with the 10^{16} EoT plot, we see that there is no new coverage of phase space. Both top plots require the A' decay to happen within 32 cm and 480 cm downstream of the target. Bottom: Branching fractions of A' to SM fermions [14]. For $m_{A'} < 0.1$ GeV, where LDMX may have sensitivity, A' is assumed here to always decay to e^+e^-.
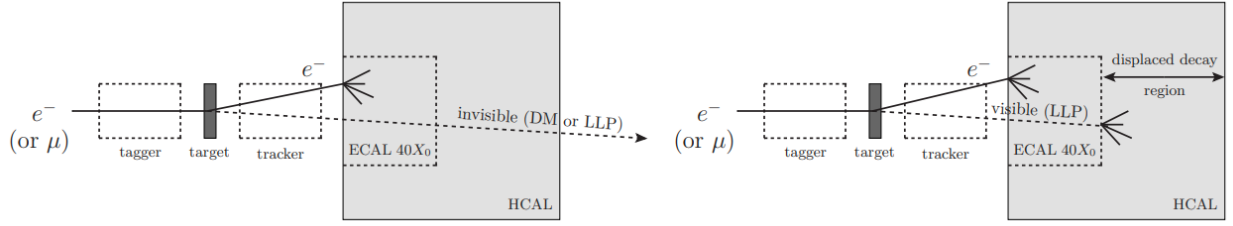
Figure 2.2: Cartoons of invisible (Left) and visible (Right) DM signatures [5]. For our purposes in this paper, the LLP (long lived particle) is an A'. Future analyses may consider ALPs. The μ suggests a beam option for other potential similar experiments, motivated by the possibility of the LLP having a stronger coupling to heavier leptons.
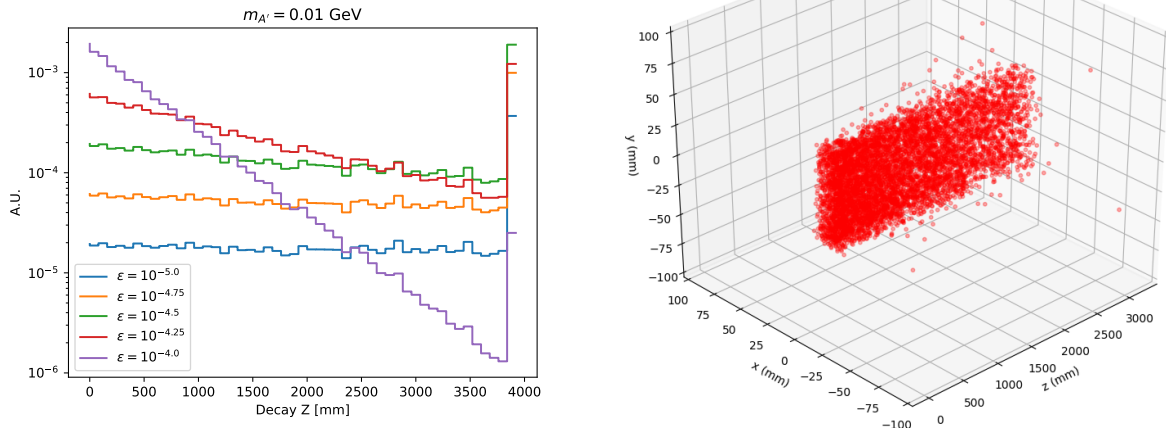
Figure 2.3: Left: Decay z distributions for $m_{A'} = 0.01$ GeV. The distribution for $\epsilon = 10^{-4.25}$ is fully simulated, and the distributions for other epsilons are obtained by weighting using the mean A' energy at the given beam energy and mass. Right: A' decay vertices. Note that the density decreases with increasing z due to exponential decay, and the shape of the beam spot is preserved fairly well throughout the length of the ECal and HCal considered in this section. The latter is because the typical polar angle of the A' momentum with the beam axis is less than $1°$.

18

the A' decays occur uniformly between 48 cm and 320 cm downstream of the target, against the efficiency of inclusive background. A uniform distribution of $z_{A'\text{decay}} \in (z_{\min}, z_{\max})$ is used here because many events produced by MadGraph5 would be outside this region of interest if any ϵ or $\gamma c\tau$ was chosen, and it represents the approximate efficiencies for a range of $\gamma c\tau$s for each mass.
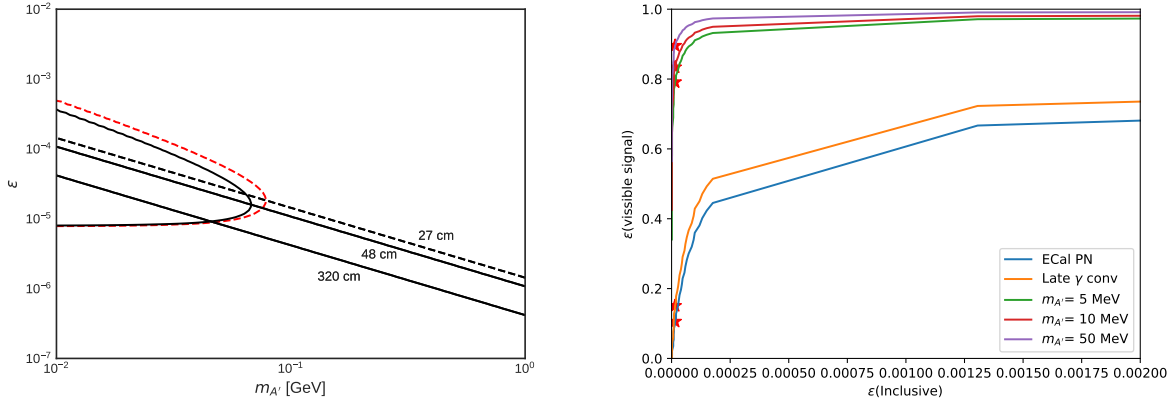
Figure 2.4: Left: Projected reach with an 8 GeV electron beam, where the A' is required to decay between 27 cm and 320 cm (red) or between 48 cm and 320 cm (black) downstream of the target. The corresponding decay length contours for the $z_{\min}$ and $z_{\max}$ values are also shown for reference. Right: ROC curves for the efficiencies of visible signal, PN, and late photon conversion against the efficiency of inclusive background after a cut on the energy in the first 20 layers of the ECal. The red star markers indicate a cut on events with greater than 3 GeV in the first 20 layers of the ECal, which is used as a trigger in this initial study of visible signal. Note that the $z_{A'\text{decay}}$ distribution for each mass is uniform. The sharp bends in the PN and late photon conversion samples are an artifact of the number and distribution of energies used to calculate efficiencies.

2.4 Extended BDTs

This subsection is further divided to provide context about an HCal BDT and to clearly separate the new BDTs for invisible and visible decays. As previously stated, the focus of this document will be on the visible A'-oriented BDTs. The invisible A'-oriented BDT is included as it is a natural extension of the work done for the former and because it is the focus of the LDMX collaboration at large. The naming convention for BDTs using information from the ECal and HCal is {ECal BDT name}_{HCal BDT name}. Additional variables or BDTs from the same detector are seperated by "AND." Some BDTs with the same name will appear in sections 2.4.2 and 2.4.3. The difference between these are the samples used for training, and of course, evaluation. This is done both for completeness and to give additional insight into how the variables respond to the different processes.

2.4.1 HCal BDT

Development on the HCal BDT began before obtaining the resources for producing phenomenologically accurate simulations. Thus, anticipating PN backgrounds to an EM signal, toy samples were made as described below. Though this would not allow for a complete optimization process, it provides an idea of the HCal's potential for separating EM and hadronic showers, which is transferable to new BDTs made with more accurate samples. Note that, except where noted otherwise, "HCal" will henceforth refer to the back HCal. The side HCal may play a bigger role in later developments related to wide-angle backgrounds; however, this simplification is currently used in light of the fact that the trajectory of A' particles tends to be within 1° of the beam axis.

In place of visible A' or ALP samples, photons were shot from various positions in the HCal with a Geant4 particle gun, through the LDMX framework. This would emulate a non-interacting LLP undergoing conversion and subsequently beginning an EM shower at the point from which it was fired. Hadronic backgrounds were represented by neutrons coming from the region were the ECal should be. The ECal was removed in these simulations so that the neutrons would be guaranteed not to interact until they got at least as far as the face of the back HCal. The position and the direction of the particle guns were chosen such that the particles would appear to come from the center of the target. Within this restriction, they were given a range of polar angles, measured from the beam axis, azimuthal angles, and energies.

The variables used in this BDT are limited to simple statistical quantities. The primary reason for this is to avoid over-fitting to the toy samples used in its development. However, this is also a consequence of the different designs of the ECal and HCal. The x and y granularity and reliability of the HCal is much less than that of the ECal due to the long bars in comparison to cells. Thus, although we have variables similar to those of the ECal BDT such as number of readout hits, total energy deposited, and average energy per hit, the HCal BDT has no analog to the ECal radii of containment. Fortunately, a form of z-segmentation is still possible. Rather than using fixed values, we begin by finding the energy-weighted z mean ($\hat{z}_{\mathrm{avg}}$) and unbiased standard deviation estimate ($\hat{\sigma}_z$) of hits. Hits are then put into regions defined by $z_{\mathrm{hit}} < \hat{z}_{\mathrm{avg}}$, $\hat{z}_{\mathrm{avg}} \leq z_{\mathrm{hit}} < \hat{z}_{\mathrm{avg}} + \hat{\sigma}_z$, and $z_{\mathrm{hit}} > \hat{z}_{\mathrm{avg}} + \hat{\sigma}_z$. Using these relative region definitions, we can get information about the shower shape with minimal $\gamma c\tau$ bias. Note that despite sharing similar names to some variables in the ECal BDT, HCal variables such as $\hat{\sigma}_x$ and $\hat{\sigma}_y$ are calculated relative to the centroid of the shower, whereas in the ECal BDT, they are calculated relative to the center of the beam axis.

For training the HCal BDT, all neutron samples were combined into one set and photon samples were combined into another set. For evaluation, background remained as a single set while photon samples were separated by energies. In total, the training sets each contained 10^6 events and the evaluation sets contained 250,000 events, split evenly amongst the different energies for the photon samples. The resulting discriminator values and ROC plots are shown in Fig. 2.5. Note that the rejection of neutron initiated showers approaches $\mathcal{O}(10^{-3})$ which is the theoretical limit of what is achievable with the LDMX HCal [5]. Thus, performance is not a great concern in this area. It may, however, be possible to reduce the number of variables used in the BDT while preserving an acceptable performance. To this end, the set of variables from the original HCal BDT, henceforth referred to as "backv1" is split into
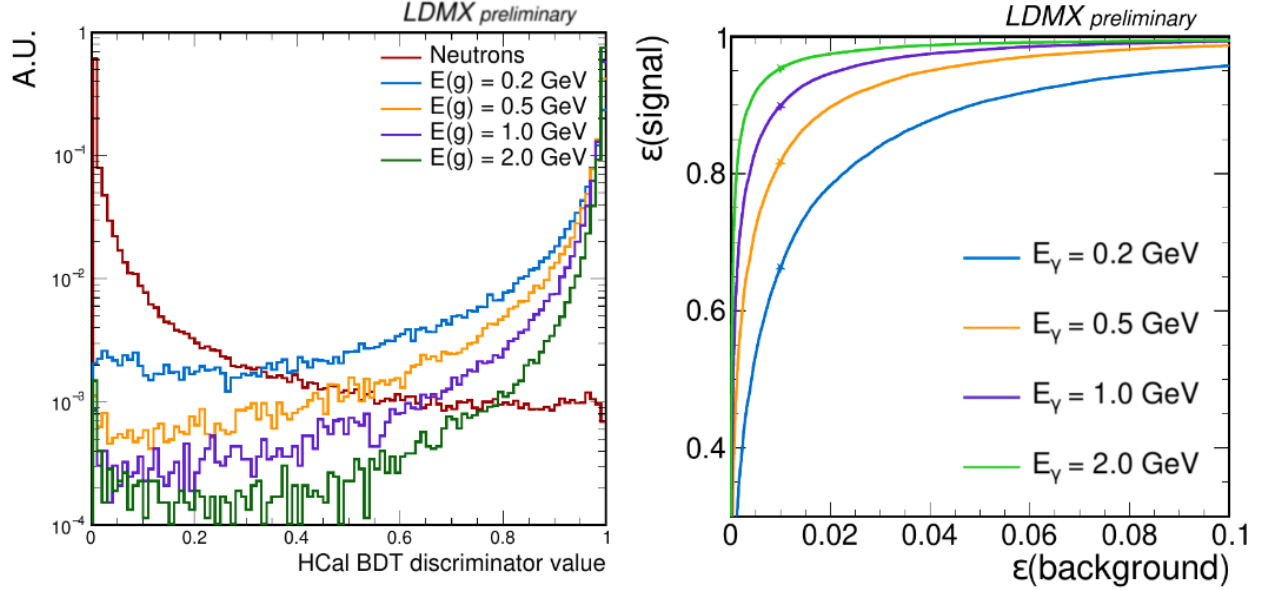
Figure 2.5: Left: Discriminator values produced by the version of the HCal BDT without the global maximum photo-electrons variable. $E(g)$ in the legend denotes the energy of the photons (γ). Right: The ROC curves corresponding to the discriminator values on the left where the neutron sample is considered background and the photon samples are considered signal. The markers are at a discriminator value of 0.9.

smaller sets. Variables calculated using the entirety of the back HCal form the "backv1all" BDT, and those calculated in the longitudinal segments form the "backv1seg," BDT so "backv1" = "backv1allANDbackv1seg." The maximum number of photoelectrons from a hit in entire back HCal and side HCal, refered to as "maxPE", appears in following sections and differs from the all other varibales as it is currently the only one that includes the side HCal.

2.4.2 Combination BDT for visible A' decay

We begin by noting that although z_{min} does not greatly affect reach, it does greatly affect N_{sig}. These are both results of exponential decay (i.e. $\exp(-t/\tau_{A'})$ or $\exp(-z/\gamma c\tau_{A'})$). The effect on the former was acknowledged and quantified in the discussion about the trigger (Sec. 2.3). Since the final sensitivity of LDMX can be deduced from prior reach plots and efficiencies, we only need to guarantee that the N_{sig} produced is statistically significant. Thus, for training the following BDTs, we continue to use samples with uniform distributions of $z_{A'\text{decay}}$ and we will weight events in the evaluations of the BDTs and subsequent analyses to get physical distributions without losing any events produced by Madgraph5. This is further motivated as training on signal samples set with nonuniform $z_{A'\text{decay}}$ would lead to bias in that variable.

Since the variables in backv1 were made with a visible ALP signal in mind, they can be directly applied to BDTs for visible A' signal. Anticipating decay length bias from the absolute longitudinal segmentation of the current ECal BDT (segmipx), a related set
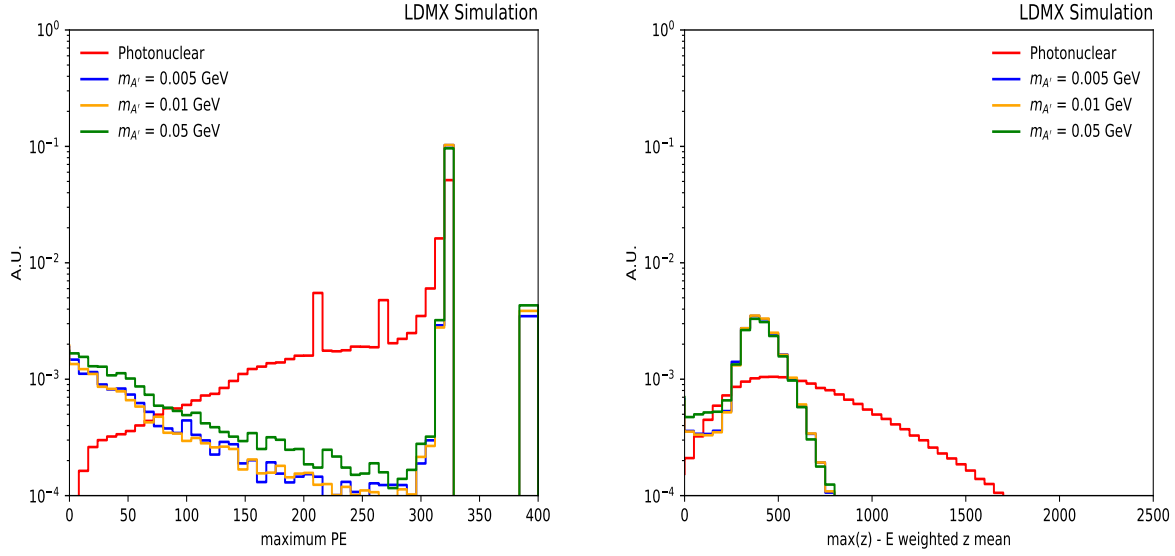
Figure 2.6: Distributions of the maximum number of photoelectrons (PE) from a hit in any section of the HCal (Left) and $z_{\mathrm{max}} - \hat{z}_{\mathrm{avg}}$ in the back HCal (Right) for simulated visible signal and ECal PN events with an 8 GeV beam. In both, the events for each signal mass are weighted to correspond to exponential decays with a characteristic length of 1 m.

of variables was made and labeled relative segmipx or "rsegmipx." In this set, variables specifying absolute z positions (such as the average layer of hits) are removed, and the segments are defined as in backv1 ($z_{\mathrm{hit}} < \hat{z}_{\mathrm{avg}}$, $\hat{z}_{\mathrm{avg}} \leq z_{\mathrm{hit}} < \hat{z}_{\mathrm{avg}} + \hat{\sigma}_z$, and $z_{\mathrm{hit}} > \hat{z}_{\mathrm{avg}} + \hat{\sigma}_z$). Figures 2.6 and 2.7 show some of the variables that go into each of these BDTs.

2.4.3   Combination BDT for invisible A'

For invisible signal, we have three broad categories of BDTs that we could expect to perform reasonably well. This is in contrast to invisible signal, for which the additional constraint of limiting decay length bias prevents the usage of some of the following BDTs. Of course, the current BDTs, Gabrielle and segmipx, are options. The next most straight forward category would be to combine the current BDTs with the HCal BDTs. Although rsegmipx discards information about the absolute z-position of showers, it may better capture the shape of the showers, so combinations of rsegmipx with HCal BDTs are also of interest.
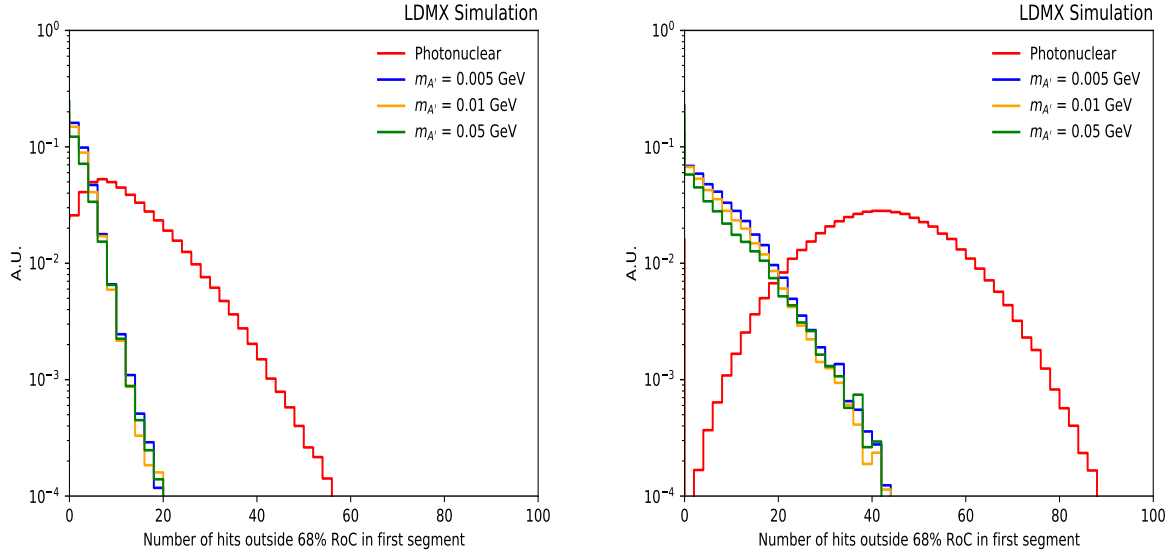
Figure 2.7: Distributions of the number of hits outside the radius of containment and in the first longitudinal segment, as defined by segmipx (layerID$_{\text{hit}} < 7$) (Left) and rsegmipx ($z_{\text{hit}} < \hat{z}_{\text{avg}}$) (Right). While the signal distributions reach higher numbers of hits, the ECal PN distribution is shifted to the right more significantly with the relative segmentation, leading to better separation between signal and background. Signal samples are weighted to correspond to exponential decays with a characteristic length of 1 m.
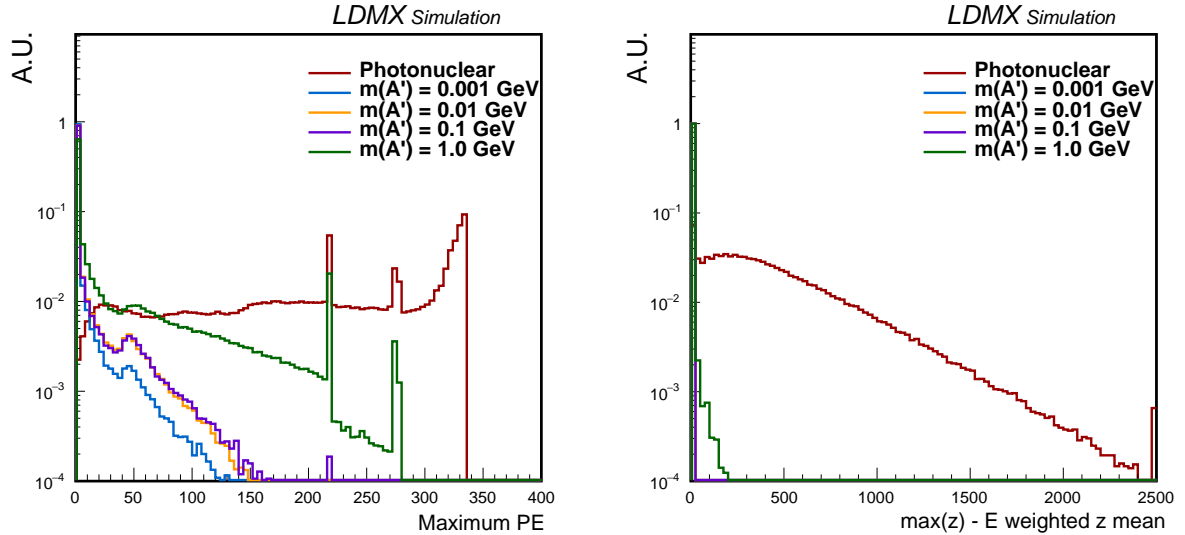
Figure 2.8: Distributions of the maximum number of photoelectrons (PE) from a hit in any section of the HCal (Left) and $z_{\text{max}} - \hat{z}_{\text{avg}}$ (Right) in the back HCal for simulated invisible signal and ECal PN events with a 4 GeV beam. These variables provide much better separation from PN events for invisible signal than they do for visible signal
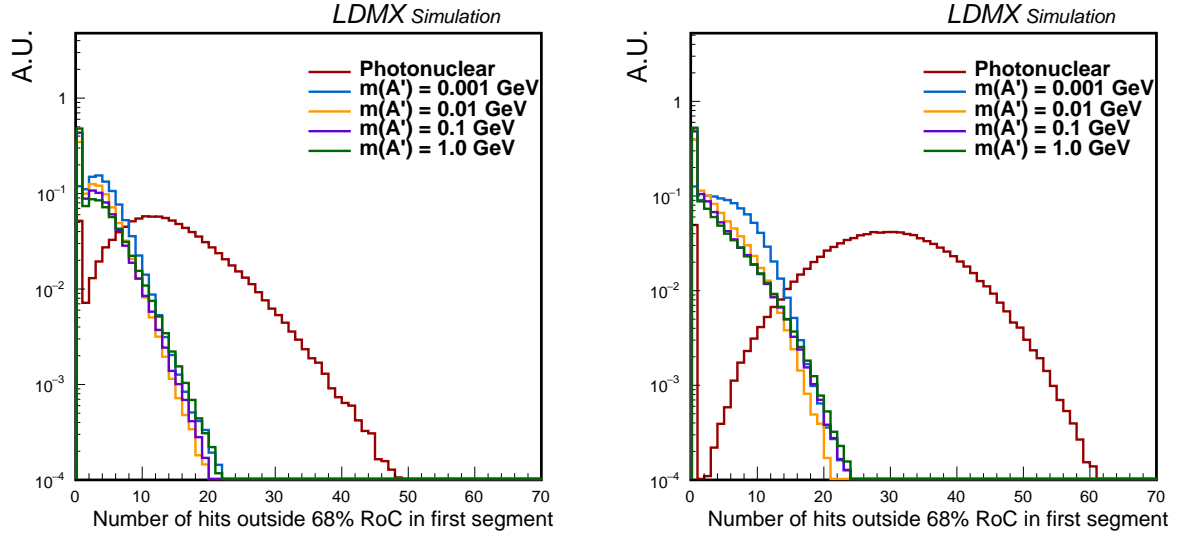
Figure 2.9: Distributions of the number of hits in outside the radius of containment and in the first longitudinal segment, as defined by segmipx (layerID$_{\text{hit}}$ < 7) (Left) and rsegmipx ($z_{\text{hit}} < \hat{z}_{\text{avg}}$) (Right). As with visible signal, the separation of signal and PN improves.

Chapter 3

Results

In this chapter, we present the performance of a few pre-existing and new BDTs on visible and invisible signal, along with their biases.

3.1   Visible A' oriented BDT Performances

In the following results, all samples have been trigger skimmed and signal samples are weighted to represent $\gamma c\tau = 1$ m for each $m_{A'}$. Fig. 3.1 shows the performance of three BDTs of particular interest: segmipx, backv1, and rsegmipx_backv1. Tables 3.1 and 3.2 summarize the performance of more BDTs at fixed discriminator values and fixed ECal PN efficiencies. The aforementioned triplet is selected to represent the best performing ECal-based BDT, HCal-based BDT, and combination BDT respectively. BDTs involving the maximum number of photoelectrons in any HCal section are not considered for representing the best combination BDT, as there is much more room for development of variables involving the side HCal.

As anticipated, segmipx, which has information about z positions in the ECal, slightly outperforms rsegmipx, which lacks this information. Interestingly, it was also found to have less decay length bias. This is because we have only considered signal decaying in the last segment defined by segmipx or beyond, which means there is effectively no segmentation of the ECal for these signal events. From Fig. 3.2 we see that of the $|\vec{p_T}|$ bias of backv1 and segmipx are similar while the $|\vec{p_T}|$ bias of rsegmipx_backv1 is much less. Recalling that A' trajectories typically have a small angle with the beam axis, decay length bias is approximated in Fig. 3.3 by $z_{A'\text{decay}}$ bias. Clearly, HCal BDTs lack discrimination power for $z_{A'\text{decay}}$ shorter than the distance from the target to the front of the back HCal. ECal BDTs have discrimination power within the ECal and far beyond it, as this would look like invisible signal from the perspective of the ECal, but they struggle at the end of the ECal and the region soon after it.
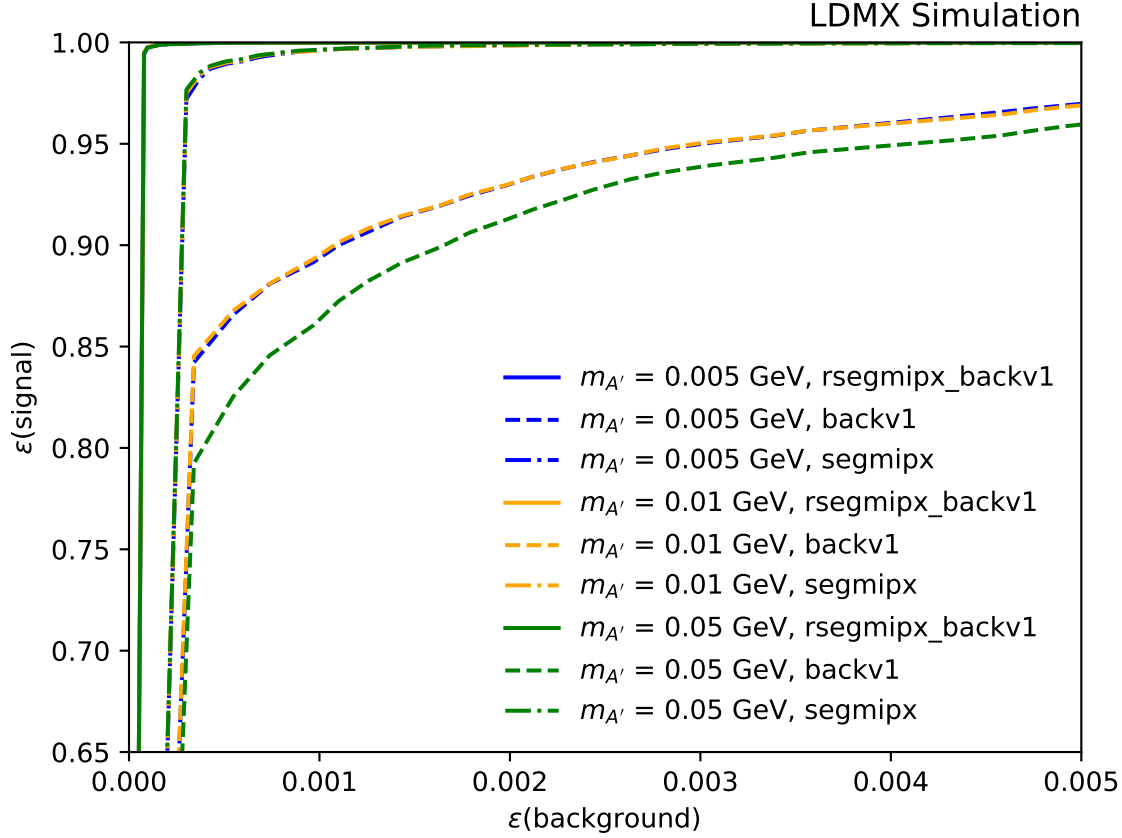
Figure 3.1: ROC curves for segmipx, backv1, and rsegmipx_backv1, showing the difference in veto power from an ECal-based BDT, an HCal-based BDT, and a BDT that combines them.

BDT Name	Discriminator value > 0.99			
	$\varepsilon(\text{PN})$	$\varepsilon(0.005)$	$\varepsilon(0.01)$	$\varepsilon(0.05)$
backv1all	4.1×10^{-4}	0.8101	0.8155	0.7521
backv1ANDmaxPE	3.7×10^{-4}	0.8561	0.8586	0.8087
backv1seg	3.5×10^{-4}	0.8140	0.8191	0.7505
backv1	3.4×10^{-4}	0.8418	0.8451	0.7919
rsegmipx	3.0×10^{-4}	0.9662	0.9707	0.9741
segmipx	3.0×10^{-4}	0.9720	0.9748	0.9767
rsegmipx_backv1	8.0×10^{-5}	0.9941	0.9949	0.9947
rsegmipx_backv1ANDmaxPE	$< \mathcal{O}(10^{-5})$	$< \mathcal{O}(10^{-5})$	$< \mathcal{O}(10^{-4})$	$< \mathcal{O}(10^{-4})$

Table 3.1: Efficiencies of ECal PN and visible signal samples after requiring a discriminator value greater than 0.99 for various BDTs. They are ordered principally by decreasing ECal PN efficiencies and subsequently by increasing efficiencies of $m_{A'} = 0.01$ GeV signal. As seen from table 3.2, rsegmipx_backv1ANDmaxPE has acceptable signal efficiencies at more appropriate discriminator values, around 0.6.
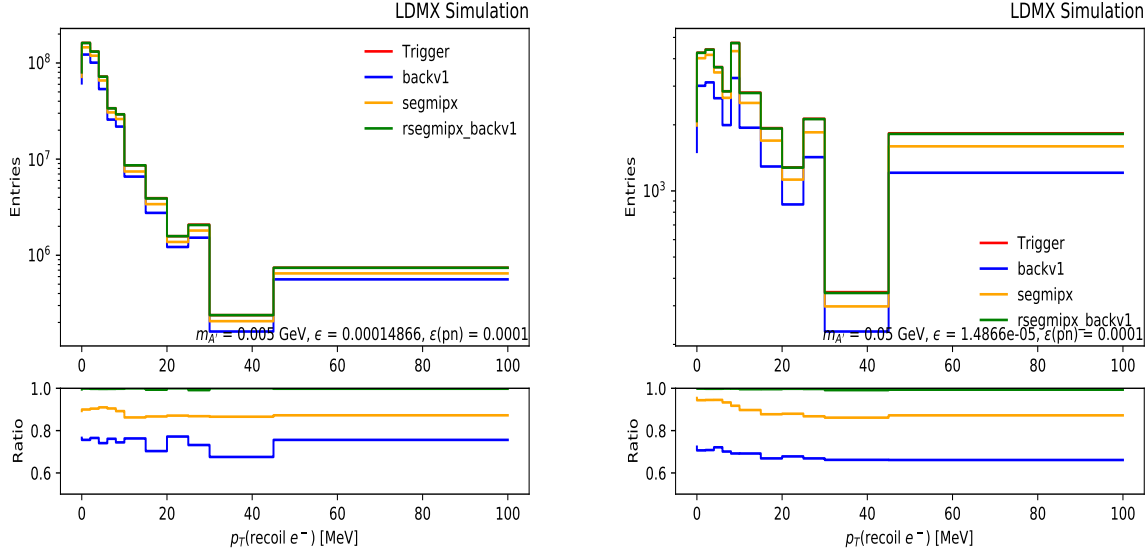
Figure 3.2: Distributions of transverse momentum of the recoil electron in $m_{A'} = 0.005$ GeV (Left) and $m_{A'} = 0.05$ GeV (Right) samples before and after cuts on the discriminator value, such that ECal PN efficiency is 10^{-4} for each BDT. Sloped ratios indicate transverse momentum bias.
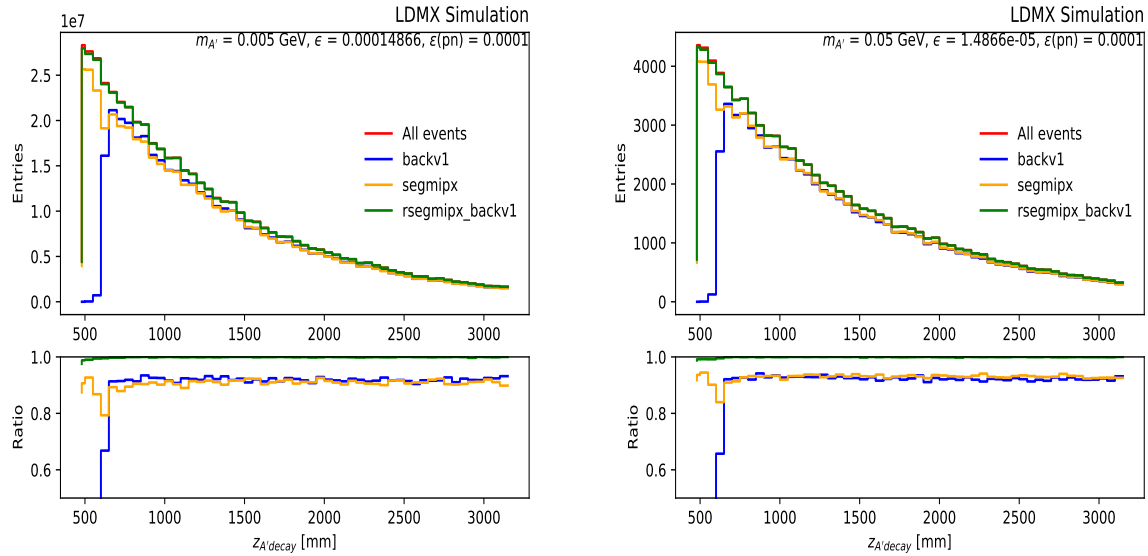
Figure 3.3: Distributions of the z position of the A' decay in $m_{A'} = 0.005$ GeV (Left) and $m_{A'} = 0.05$ GeV (Right) samples before and after cuts on the discriminator value, such that ECal PN efficiency is 10^{-4} for each BDT. Sloped ratios indicate decay length bias. The dip in ratio of segmipx and the drop-off in the ratio of backv1 around $z_{A'\text{decay}} = 700$ mm correspond to the insensitive region between the ECal and HCal.

27

BDT Name	$\varepsilon(\text{PN}) = 0.0001$			
	Disc. value	$\varepsilon(0.005)$	$\varepsilon(0.01)$	$\varepsilon(0.05)$
backv1all	0.999	0.7038	0.7135	0.6496
backv1seg	0.999	0.7340	0.7420	0.6730
backv1	0.999	0.7571	0.7648	0.6967
backv1ANDmaxPE	0.999	0.7641	0.7717	0.7046
rsegmipx	0.999	0.8662	0.8817	0.8945
segmipx	0.999	0.8999	0.9101	0.9194
rsegmipx_backv1ANDmaxPE	0.639	0.9744	0.9750	0.9743
rsegmipx_backv1	0.967	0.9974	0.9974	0.9974

Table 3.2: Discriminator values required to achieve ECal PN efficiency of 10^{-4} and efficiencies of visible signal samples after requiring the same discriminator value or greater. They are ordered by increasing efficiencies of $m_{A'} = 0.01$ GeV signal.
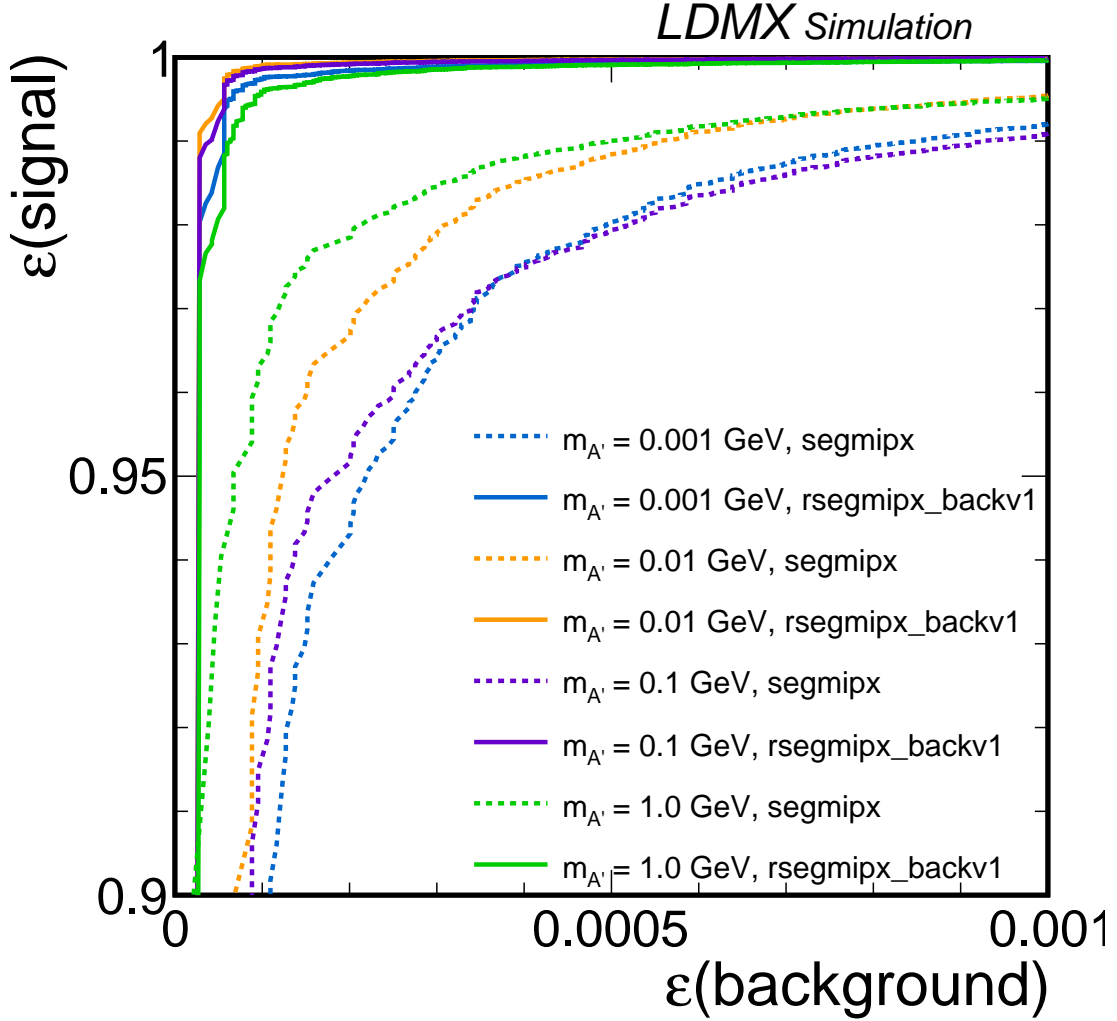
Figure 3.4: ROC curves for segmipx and rsegmipx_backv1 trained and evaluated on invisible signal.

3.2 Invisible A' oriented BDT Performances

In this section, we continue to highlight segmipx and rsegmipx_backv1. Note however, from table 3.3, that the variants with relative segmentation outperform those with absolute segmentation. From figures 2.7 and 2.9 we see that the improvement in separation in these segment variables is greater for invisible signal than for visible signal. Since there is never any activity from the A' in the invisible scenario, using relative segmentation affects background more than signal, and this offsets the loss of absolute z information. Still, the difference in performance between the variants is small.

Fig. 3.4 illustrates that using HCal information is especially helpful for the lower mass signal efficiencies. This is consistent with Fig. 2.8 which shows that the $m_{A'} = 1.0$ GeV signal produces the most activity in the HCal.
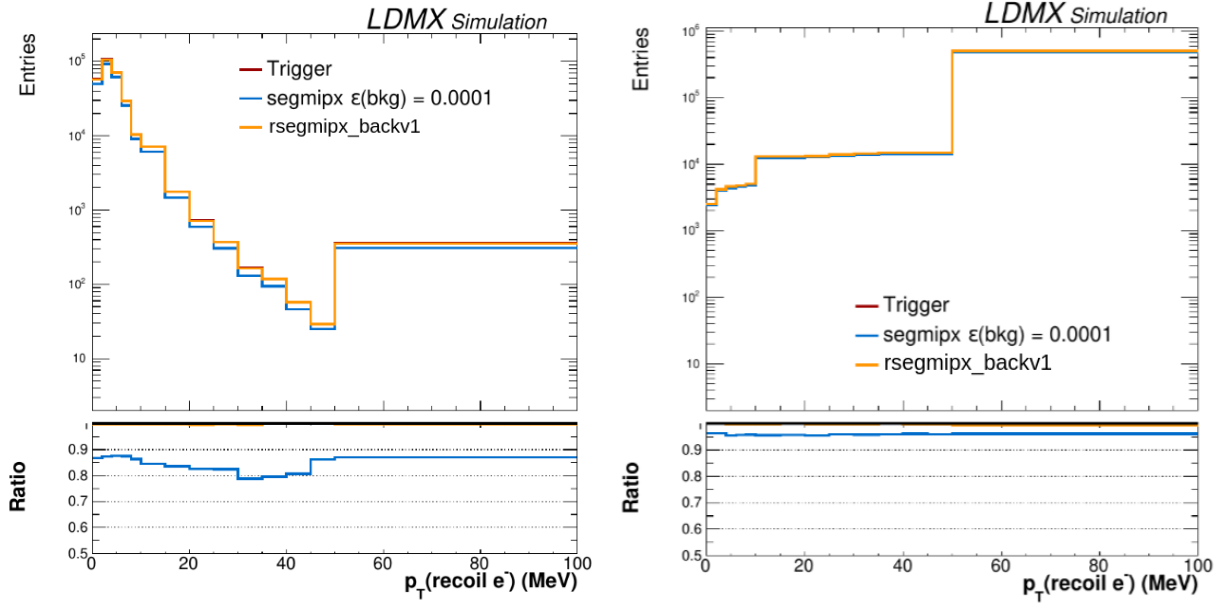
Figure 3.5: Distributions of transverse momentum of the recoil electron in $m_{A'} = 0.001$ GeV (Left) and $m_{A'} = 1.0$ GeV (Right) samples before and after cuts on the discriminator value, such that ECal PN efficiency is 10^{-4} for each BDT. Sloped ratios indicate transverse momentum bias. In both cases, the histogram for rsegmipx_backv1 is mostly overlapping with the trigger histogram and the ratio line is therefore overlapping with the thick border at 1.

| BDT Name | $\varepsilon(\text{PN}) = 0.0001$ | | | | |
	Disc. value	$\varepsilon(0.001)$	$\varepsilon(0.01)$	$\varepsilon(0.1)$	$\varepsilon(1.0)$
gabrielle	0.999	0.4016	0.6797	0.7536	0.7953
segmipx	0.998	0.8791	0.9309	0.9151	0.9631
segmipx_backv1	0.651	0.9958	0.9984	0.9984	0.9945
rsegmipx_backv1	0.651	0.9974	0.9990	0.9986	0.9959
segmipx_backv1ANDmaxPE	0.644	0.9980	0.9993	0.9994	0.9954
rsegmipx_backv1ANDmaxPE	0.673	0.9987	0.9996	0.9996	0.9980

Table 3.3: Discriminator values required to achieve ECal PN efficiency of 10^{-4} and efficiencies of invisible signal samples after requiring the same discriminator value or greater. They are ordered by increasing efficiencies of $m_{A'} = 0.001$ GeV signal.

Chapter 4

Discussion

4.1   Conclusion

In summary, we have explored a new dark matter model for LDMX. It required the consideration of a new constraint, decay length bias, on the machine learning models used to reject difficult backgrounds. This motivated inclusion of information from the HCal to make up for the information from the ECal that could not be used, as it is for the longer studied case of invisible signal. The decay length bias constraint also motivated the creation of a variant on the current ECal BDT used for visible signal that outperforms its predecessor. By combining this variant with the HCal BDT and training on the appropriate signal samples, we achieve greater than 99% efficiency of signal in the phase space considered for both scenarios at the reference ECal PN background efficiency of 10^{-4}. This is a substantial improvement from the ECal BDTs used primarily to optimize the ECal design, which on their own achieved signal efficiencies from $\sim$88% to $\sim$96%.

4.2   Future work

The novelty of the visible signal analysis means there is a lot of room for more studies. First, we would want to address matters of samples production such as integrating visible signal production into LDMX software and test more methods for producing late photon conversion samples. A similar analysis should be done for late photon conversion background as it was included in the trigger section (2.3), but left out of the BDT analysis due to time constraints. As mentioned in that section, the trigger should be optimized, if not completely redesigned. One option that was not mentioned was using a clustering algorithm, which was also decided against for this initial study due to time constraints. The set of variables in the HCal BDT and its derivatives remain unoptimized, and should be optimized before also doing the same for hyperparameters.

Despite the improved performance of the combined BDTs, they cannot reject all background, so if MIP tracking cannot reject all the remaining background, the HCal veto using the maximum number of photoelectrons would still be necessary. Further, the main search will likely be done with some form of machine learning using information from multiple sub-detectors, but the independent ECal and HCal vetoes will continue to be studied to provide

unbiased samples for the other to use in its ongoing design study.

Appendix A

Further BDT comparisons

This appendix contains ROC curves for BDTs that were not included in the body of this document for brevity and taking a conservative approach to BDTs involving maxPE.
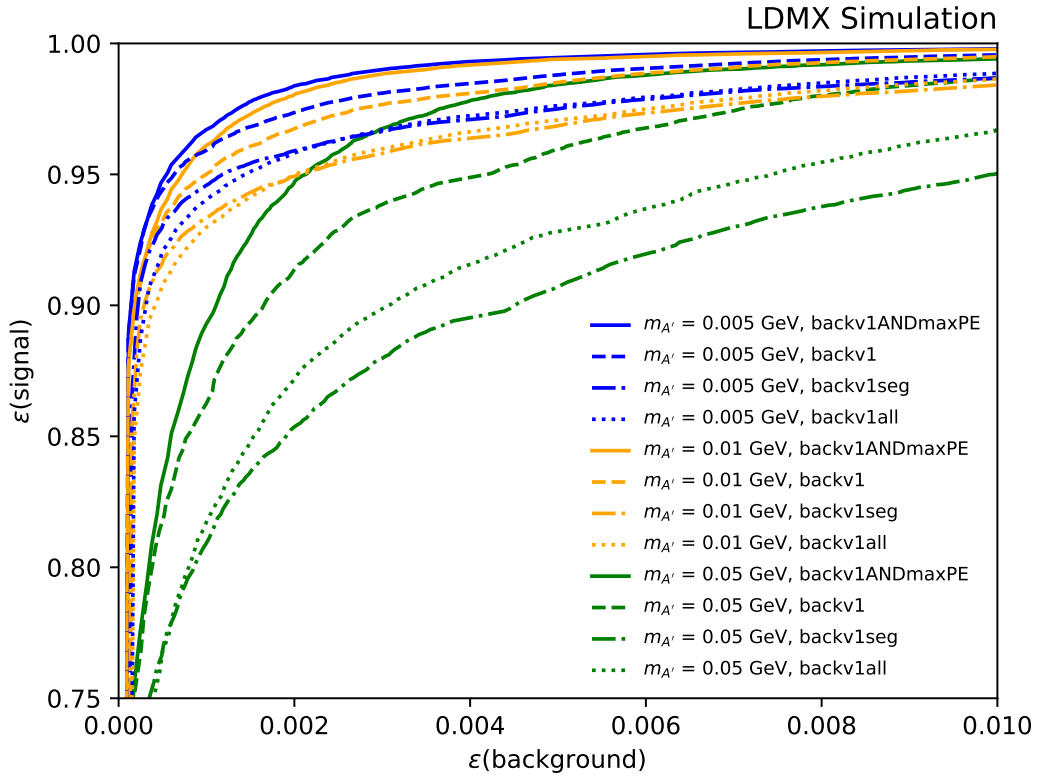
Figure A.1: ROC curves for HCal BDTs trained and evaluated on visible signal. This demonstrates the power of the segmented variables, global (within the back HCal) variables, and their combined power. A version with the maximum number of photoelectrons in any section of the HCAL is also included, though it requires more study than the others, as this is currently used as a separate veto for invisible signal. In addition, it is the only information taken from the side HCal and there is likely more to be gained from other side HCal variables.
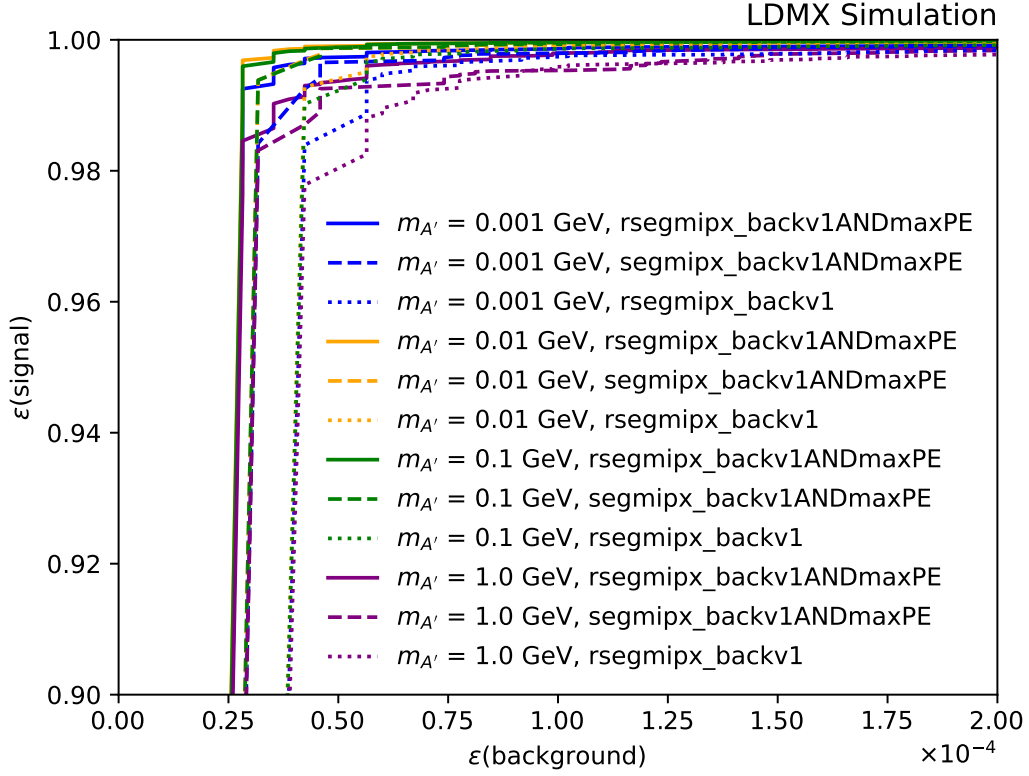
Figure A.2: ROC curves for BDTs trained and evaluated on invisible signal. As for backv1ANDmaxPE presented in Fig. A.1, rsegmipx_backv1ANDmaxPE and segmipx_backv1ANDmaxPE require further study. Rsegmipx_backv1 also provides an indirect comparison between the aforementioned BDTs and segmipx shown in Fig. 3.4.

List of Figures

List of Tables

References

[1] C. Lu, R. Ramos, Y. Tsai, Shedding light on dark matter with recent muon (g-2) and Higgs exotic decay measurements, arXiv:2104.04503v2 [hep-ph]

[2] K. Freese, Status of dark matter in the universe, https://ned.ipac.caltech.edu/level5/Sept17/Freese/Freese2.html

[3] Harvard-Smithsonian Center for Astrophysics, https://chandra.harvard.edu/photo/2006/1e0657/index.html

[4] M. Bradač *et al.*, *Strong and Weak Lensing United. III. Measuring the Mass Distribution of the Merging Galaxy Cluster 1ES 0657–558**, arXiv:astro-ph/0608408

[5] A. Berlin, N. Blinov, G. Krnjaic, P. Schuster and N. Toro, Dark Matter, Millicharges, Axion and Scalar Particles, Gauge Bosons and Other New Physics with LDMX, Phys. Rev.D 99 (2019) 075001

[6] T. Åkesson *et al.* [LDMX Collaboration], *Light Dark Matter eXperiment (LDMX)*, arXiv:1808.05219 [hep-ex].

[7] T. Binder, *et al.* Dark matter relic abundance beyond kinetic equilibrium. Eur. Phys. J. C 81, 577 (2021). https://doi.org/10.1140/epjc/s10052-021-09357-5

[8] The LDMX collaboration., Åkesson, T., Blinov, N. et al. A high efficiency photon veto for the Light Dark Matter eXperiment. J. High Energ. Phys. 2020, 3 (2020). https://doi.org/10.1007/JHEP04(2020)003

[9] A. Rogozhnikov, *Gradient Boosting explained [demonstration]*, https://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html

[10] T. Chen, C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, arXiv:1603.02754v3,

[11] XGBoost Developers, *XGBoost Documentation*, https://xgboost.readthedocs.io/en/latest/

[12] Alwall, J., Frederix, R., Frixione, S. et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. J. High Energ. Phys. 2014, 79 (2014). https://doi.org/10.1007/JHEP07(2014)079

[13] GEANT4 collaboration, GEANT4: A simulation toolkit,
     https://doi.org/10.1016/S0168-9002(03)01368-8

[14] P. Ilten, Y. Soreq, and M. Williams, Serendipity in dark photon searches, Serendipity
     in dark photon searches, https://arxiv.org/pdf/1801.04847.pdf

[15] Brookhaven Science Associates, Bragg Curves and Peaks,
     https://www.bnl.gov/nsrl/userguide/bragg-curves-and-peaks.php