Implementation of Radius of Containment in Boosted Decision Tree for Light Dark Matter eXperiment (LDMX)

by

Harrison J. Siegel

A thesis submitted in partial satisfaction of the requirements for the degree of Bachelor of Science in Physics in the College of Letters and Science of the University of California, Santa Barbara

June 2019
The thesis of Harrison J. Siegel, titled Implementation of Radius of Containment in Boosted Decision Tree for Light Dark Matter eXperiment (LDMX), is approved:

Date __________

Date __________

Date __________

Date __________

University of California, Santa Barbara
Implementation of Radius of Containment in Boosted Decision Tree for Light Dark Matter eXperiment (LDMX)

Copyright 2019 by Harrison J. Siegel
Abstract

Implementation of Radius of Containment in Boosted Decision Tree for Light Dark Matter eXperiment (LDMX)

by

Harrison J. Siegel

Bachelor of Science in Physics

University of California, Santa Barbara

LDMX is a proposed fixed target missing momentum dark matter search that aims to probe light dark matter with exceptional sensitivity over a mass range far below that currently explored by most direct detection experiments. Much of the experiment’s power comes from the high granularity sampling electromagnetic calorimeter (ECAL) which the experiment will borrow from the HL-LHC upgrades to the CMS experiment at CERN. LDMX hopes to find dark matter by observing a “dark bremsstrahlung” process, wherein dark matter particles are produced with an electron beam in a similar manner to a normal bremsstrahlung process; the similarities mean that normal bremsstrahlung is one of the most crucial backgrounds to focus on rejecting. The most powerful tool currently used for background rejection in LDMX is the Boosted Decision Tree (BDT), a machine learning technique. By developing new features using the “radius of containment”, a characterization of the width of particle showers in each longitudinal layer of the ECAL, we demonstrate significant improvements in the efficiency of background rejection using the BDT. There is also room for improvement of the method beyond what is currently presented here.

This thesis aims to introduce a brief history of the dark matter problem before overviewing the LDMX experiment and the basics of BDT’s, and concludes with the most recent results of our radius of containment ECAL background rejection studies.
For Ant
Contents

Contents ii

List of Figures iii

1 Motivating Dark Matter Searches 1
  1.1 Why Dark Matter? ..................................................... 1

2 LDMX 5
  2.1 LDMX Design .......................................................... 6
  2.2 Identifying Dark Matter Signal .................................... 10
  2.3 Types of Backgrounds ................................................. 10
  2.4 Contemporary Accelerator-Based Dark Matter Experiments .... 13

3 Radius of Containment 14
  3.1 Boosted Decision Trees (BDTs) ...................................... 14
  3.2 New Method: Implementing Radii of Containment ............... 17

4 Discussion 26
  4.1 Conclusion .............................................................. 26
  4.2 Further Work .......................................................... 26

Bibliography 28
List of Figures

2.1 Compared to most direct detection experiments whose sensitivity to different dark matter scenarios varies over many orders of magnitude, LDMX is basically equally sensitive to many of the most likely dark matter scenarios due to the relativistic speeds of the particles it measures. The y axis indicates sensitivity. 6

2.2 This conceptual cartoon of the LDMX experiment shows the electron beam hitting the thin tungsten target, recoiling in a dark bremsstrahlung event and producing a shower in the ECAL, while the dark matter particles go through the detector unseen [1]. 7

2.3 This cutaway of LDMX shows the various detector subsections in detail [1]. 7

2.4 A cartoon depicting the process by which electromagnetic showers form by bremsstrahlung and pair production events [12]. 9

2.5 The depth of the shower in the ECAL depends on the energy of the particle that starts the shower as depicted above [9]. 9

2.6 A cartoon depicting the orientation of the 7 hexagonal sensors that make up a single layer of the ECAL. 10

2.7 Details the major backgrounds to consider in LDMX and their rates relative to the number of beam electrons incident on the target (y axis). For context, the relative rate of dark matter production that we are looking for is on the order of $10^{-14}$ [1]. 11

2.8 The distribution of transverse momentum for recoil e⁻’s created due to a variety of signal masses and for background. Note the discriminating power of $p_T$ comes from the extreme drop off in high transverse momentum events for background relative to all signal (the various masses are different possible A' masses). The background drop off in black on the figure goes as $1/p_T^3$ for high $p_T$ [1]. 12

3.1 A simple example of the ideas behind a decision tree. The datasets being used get broken into consecutively smaller subsets while an associated decision tree is simultaneously incrementally developed. Decision nodes, like Temperature for example, have branches (Hot, Mild, Cool). Leaf nodes, the only one of which here is Play Golf, represent a classification (Yes or No). The topmost decision node in the tree corresponds to the best predictor, in this case Outlook [13] . . . 15
3.2 The basics of ROC curves: the y axis is the true positive rate or TPR for classifications (for LDMX, how much the BDT correctly identifies a signal event), and the x axis is the false positive rate or FPR (for LDMX, how much the BDT mistakenly identifies backgrounds as signal events). The curve is made by selecting a continuous range of values of the discriminant produced by the BDT and plotting their performance. The optimal BDT has a normalized area under the curve (AUC) of 1, so the closer the curve gets to the top-left corner of the plot, the more efficient the BDT is. A diagonal line would represent a 50/50 chance of correct classification, i.e. the BDT is just guessing.

3.3 These cartoons illustrate the radius of containment concept. Fig. (a) illustrates the application of the radii to a single particle creating a shower: the trajectory of an incoming particle is projected through the ECAL (orange line symbolizing \(e^-\) trajectory), radii for each respective layer and concentric multiples of said radii are then drawn around the trajectory (green circles), and the hits in each radius of containment and concentric annulus are summed up through all the individual layers. Fig. (b) illustrates that the process is then repeated for the trajectory obtained via kinematics of a potential photon or \(A'\) (purple line).

3.4 The original CMS radius of containment plot [6]. The y axis is the radius \(\rho\) in mm.

3.5 Momentum vs angle with respect to the normal of the ECAL for electrons entering the detector in single electron signal events. The lightest yellow curve in the plot indicates the most frequent values.

3.6 The 4 different binnings used for the radii of containment, as follows: 1) \(p > 500, \theta < 10\) 2) \(p < 500, \theta < 10\) 3) \(10 < \theta < 20\) 4) \(\theta > 20\)

3.7 The various 68 percent binned radii of containment generated for LDMX. Note that in the back layers there is a large deviation in radii, the radii are all very large, and there is very little shower energy deposited in the back layers. Most of this is because the ECAL is designed to be longer than the shower, and thus the hits in the back layers at the end of the ECAL are very sporadic.

3.8 Because of the low statistics in the back of the ECAL, we opted for a linear fitting in the back layers. This should help the BDT perform optimally.

3.9 This ROC curve demonstrates the improvement of the BDT containing the radius of containment features (New) over the previous features (Nominal) for a range of possible dark photon masses. This ROC curve is the result of running on a sample of the order \(\mathcal{O}(10^{14})\) electrons incident on the target.

3.10 This ROC curve is the result of a BDT with features that are linear combinations of the outside and annulus variables: in addition to summing the energy, hits, etc. outside the largest annulus, we also include everything outside the fourth largest annulus, outside the third largest, etc. There is clear improvement, however this ROC curve was produced over a sample on the order of \(\mathcal{O}(10^6)\) events, and thus still requires more testing.
3.11 These figures demonstrate transverse momentum bias in the BDT with respect to signal efficiency - a distribution that is not flat or that rises with increasing transverse momentum ($p_T$) hints at the BDT making use of transverse momentum, which is undesirable. The orange line is for the New BDT, the blue line is for the Nominal BDT.
Acknowledgments

I would like to thank Prof. Joseph Incandela for giving me a chance, Valentina Dutta for putting up with my myriad of labyrinthine slide decks, Jack Bargemann for reminding me that the LDMX framework was designed, Alex Patterson for marching to the beat of his own drum, Melissa Quinnan for giving me faith that we could actually get something done, and Python for not being C++.
Chapter 1

Motivating Dark Matter Searches

"Let there be light [sic]."

- Genesis 1:3

1.1 Why Dark Matter?

What do you see when you look out into space? It depends on what you look with. Using the naked eye on a clear night, you may see the Milky Way’s swath of light and the brightest and closest stars and planets sitting in a sea of darkness. Using an optical telescope, the Milky Way’s splotches of light reveal themselves to be a litany of stars, and the planets in our Solar System are suddenly joined by orbiting moons. Using a radio telescope, the sea of darkness spews forth a faint and almost isotropic glow, the Cosmic Microwave Background which is one of the strongest pieces of evidence for the Big Bang.

Clearly then, if different tools have allowed us to see features of the universe that would otherwise be invisible, it should come as no surprise that there might be some matter in the universe that we still cannot see, even with the most sensitive of instruments at our disposal. How would we even know that these invisible objects were there? In the event that they exert any influence on the visible matter around them, that would be a clear indication of their existence. In the most general sense, this is what we are referring to when we speak of "dark matter".

Experimental Evidence

In the early 1900’s astronomers began to make some of the first estimates of the ratio of luminous to non-luminous matter in galactic clusters. Fritz Zwicky’s famous 1933 paper on the redshift of extragalactic nebulae was the first to definitively state that dark matter should far outweigh visible matter. Using the virial theorem to predict the velocity of the bodies in the Coma Cluster, he found that the observed redshift would require an average
mass density up to 400 times larger than that attributed to the luminous matter alone [16]. Nonetheless, astronomers were not even sure what the nature of the dark matter might be. Many proposed less exotic theories to explain the phenomenon such as the dark matter being intergalactic remnants of stars, all of which would be ruled out decades later.

The revolution in our understanding of the scope of dark matter’s existence came in the 1970’s, with Kent Ford’s innovative image tube spectrograph. He and Vera Rubin used the spectrograph to make optical measurements of galaxies and produce highly accurate rotation curves, which plot the orbital speeds of visible matter in a galaxy as a function of radial distance from the galaxy’s center. Vera Rubin, Morton Roberts, and many others began noticing that the rotation speeds of many galaxies remained constant with increasing radial distance, rather than dropping off as would be expected given the amount of luminous matter present in the galaxies. Comparison between radio and optical observations showed that these flat rotation curves extended out beyond the optical size of the galaxies. It became evident that this behavior could only be explained by a significant mass density in individual galaxies at large radial distances that could not be due to luminous matter alone [2].

A New Particle?

Even into the 1970’s, the possibility of dark matter having particle origins was not acknowledged as a prevailing theory. Unlike today, there was a large gulf between particle physicists and astrophysicists: each side simply could not see how the other’s work would be relevant to their respective fields. It can be argued that studies of neutrinos began to change this trend. Theorists began to place constraints on the mass range of neutrinos by incorporating their influence on cosmic expansion after they would have achieved thermal equilibrium in the early universe [15, p. 161-5]. Neutrinos are special particles in that they are long-lived and interact only through the weak and gravitational forces. It began to dawn on physicists that a weakly interacting particle like the neutrino might be the culprit in the galactic mass density problem.

The advent of numerical simulations in the 1980’s would allow different particle dark matter models to be rigorously tested. The main use of numerical simulations was in probing the effects of whether a given dark matter particle candidate would be ”hot” or ”cold” during the time in the early universe when large scale structures were forming. Hot dark matter would have relativistic speeds during this time, while cold dark matter would have non-relativistic speeds. Simulations of hot dark matter tend to form very large structures first that later break down into galaxy sized halos. Meanwhile cold dark matter tends to do the opposite, forming small halos that coalesce over time. It quickly became clear through comparison of these simulations with galaxy surveys that hot dark matter would not produce galaxy clusters of the correct scale to match observations. As Standard Model neutrinos are light and would thus have highly relativistic velocities during the early universe, this largely ruled them out as dark matter particles: with that, there were no Standard Model particles left that fit the bill [2].
From the 1980’s up until now, many theories on the nature of dark matter have come and gone, such as MACHOs (Massive Astrophysical Compact Halo Objects), primordial black holes, and even MOND (Modified Newtonian Dynamics), which is a modification of the gravitational force. Nonetheless, the most convincing idea that has stood the test of time is that dark matter is a non-baryonic particle that at the very least interacts with Standard Model particles via the gravitational force. This is quite an astonishing possibility if one stops to consider it: we are relatively certain now that dark matter is about 5 times more prevalent than standard matter, so that means that our entire knowledge of particle physics, the Standard Model, currently accounts for a measly fraction of the totality of the universe’s contents [2]. And with so much about dark matter remaining a mystery, fleshing out the Standard Model to include it is far from straightforward, as evidenced by the host of candidate dark matter particles that have been considered.

**Standard Model Problems and Solutions**

Even ignoring the evidence for dark matter, the Standard Model has ample inconsistencies that suggest new physics exists. The main issue that motivates dark matter candidate proposals is the gauge hierarchy problem. The simplest possible explanation of the problem is that we expect the Higgs boson mass $m_h$ to be much larger than it actually is. The Planck mass $M_{Pl} \equiv \sqrt{\hbar c / G_N} \simeq 1.2 \times 10^{19}$ GeV is the unit of mass in natural units (natural units here refers to the Planck units, which are made from just Planck’s constant $\hbar$, the speed of light $c$, Newton’s gravitational constant $G_N$, the Coulomb constant $\frac{1}{4\pi\epsilon_0}$, and Boltzmann’s constant $k_B$, and are arguably the most ”natural” natural units because they are based solely on the properties of free space), and we’d expect the Higgs mass $m_h$ to be on this scale. But $m_h$ is actually on the order of 100 GeV, 17 orders of magnitude lower than $M_{Pl}$. Quantum corrections $\Delta m_h$ to the SM Higgs mass are proportional to the energy scale at which the SM no longer correctly describes the universe:

$$\Delta m_h^2 \sim \frac{\lambda^2}{16\pi^2} \int \frac{d^4p}{p^2} \sim \frac{\lambda^2}{16\pi^2} \Lambda^2$$

(1.1)

The integral is over the momenta of particles in loop diagrams contributing to the mass correction, $\lambda$ is a dimensionless coupling, and $\Lambda$ is the energy scale at which the SM breaks down.

The only clear solutions to the gauge hierarchy problem are to have $\Lambda \leq 1$ TeV instead of $\sim M_{Pl}$ as expected based on the Standard Model, or to have the Higgs boson not be a fundamental scalar boson (boson with zero spin), but either option requires new physics at the weak scale around 10 GeV - TeV. Particle candidates that are mainly motivated by the gauge hierarchy problem include WIMPs (Weakly Interacting Massive Particles) and superWIMPs [7].

The reason that new physics should appear at the previously mentioned weak scale of 10 GeV - TeV is tied to the "WIMP miracle". The WIMP miracle is based on the idea that if stable WIMPs were created after the Big Bang, the number (referred to as the thermal
CHAPTER 1. MOTIVATING DARK MATTER SEARCHES

relic density) of WIMPs that were not annihilated after the Universe sufficiently cooled and expanded would be directly related to the WIMP annihilation cross section, which is in turn related to the WIMP mass, and by factoring in the structure of the Universe that we have observed this constrains the possible mass range of the WIMP to somewhere between 100 GeV and 1 TeV. Because this coincidentally is the same energy scale where we would expect to see new physics based on the gauge hierarchy problem, and because of the generality of the WIMP’s properties, the WIMP is favored as one of the most likely dark matter candidates [7].

Light Dark Matter

LDMX is focused on light dark matter, whose mass range is below the traditional WIMP mass that most experiments have looked for so far. The light dark matter mass range of ~10 keV - GeV is well-motivated by "hidden sector" or "dark sector" models wherein dark matter is a particle with its own unique forces and interactions completely removed from the Standard Model, but with sufficient coupling to visible matter that it achieves thermal equilibrium in the early Universe. This range is also independently motivated by asymmetric dark matter scenarios where dark matter carries a net particle number, analogous to the matter-antimatter asymmetry of visible matter. Most crucially, this mass range is simply one that has yet to be explored, and the mass on the scale expected for the WIMP miracle has been largely ruled out by direct detection experiments that have come up empty-handed to date.

For the case of LDMX, the main theoretical assumption underpinning the nature of light dark matter is very general: the dark matter particle $\chi$ is defined to have a $U(1)$ gauge boson mediator $A'$, colloquially referred to as a "dark photon". $\chi$ can be a Majorana fermion (its own antiparticle), a pseudo Dirac fermion, or a complex scalar particle with elastic or inelastic coupling to $A'$. Two distinct annihilation scenarios are possible: if $m_{A'} < m_\chi$ then $\chi$ annihilates predominantly into $A'$ pairs, whereas if $m_{A'} > m_\chi$ then $\chi$ annihilates into a virtual $A'^*$ which itself converts into standard model fermions. [1] This generality makes LDMX an attractive experiment, as it allows us to cast a larger net in our search than most other dark matter experiments.
Chapter 2

LDMX

LDMX is a uniquely powerful dark matter experiment for a number of reasons: it aims to probe a scale of “light” dark matter in the mass range of keV to GeV, which is orders of magnitude below the most sensitive direct detection experiments, and because its missing momentum approach involves relativistic particles it is sensitive to an especially broad range of dark matter scenarios in comparison to most traditional direct detection experiments (Fig. 2.1) and even other accelerator-based experiments. LDMX is in the final stages of submitting proposals for funding, and could be slated for construction as soon as the early 2020’s.

The experiment’s design is fairly simple: a 4 GeV (or possibly higher energy) electron beam gets shot at a fixed target, behind which lies an electromagnetic calorimeter (ECAL) and hardonic calorimeter (HCAL) and ”recoil” and ”tagging” trackers, which respectively measure the energy of the particles coming out of the target and reconstruct those particles’ trajectories. The electrons in the beam are theorized to be able to undergo a ”dark bremsstrahlung” process, whereby when they pass through the electric field of the nucleus of an atom in the target they produce a dark matter particle or dark matter mediator particle, instead of a photon as they would in a normal bremsstrahlung process. Dark matter particles leave no energy in the calorimeters, and this means that the electron which produced the dark matter leaves significantly less energy than it normally would in the calorimeters. This missing momentum is the telltale sign of successful dark matter production. An event where dark matter is produced is referred to as ”signal”. Events where this does not happen are referred to as ”background”. The main experimental challenge is to identify the background events and throw them away while saving as many signal events as possible, and the most sophisticated tool we implement for this task is the Boosted Decision Tree (BDT).

Every study of LDMX is currently based entirely on simulations, as nothing has been built physically yet. GEANT4, a framework for simulating physics events using specific detector geometries, is used to create individual files containing events to be analyzed. An ”event” constitutes an instance where one or more electrons hit the target and pass through the ECAL together. Thus everything shown in this thesis is a validation study of the detector’s design, made with realistic simulated data.
Figure 2.1: Compared to most direct detection experiments whose sensitivity to different dark matter scenarios varies over many orders of magnitude, LDMX is basically equally sensitive to many of the most likely dark matter scenarios due to the relativistic speeds of the particles it measures. The y axis indicates sensitivity.

2.1 LDMX Design

"With a little bit of imagination, anything is possible."

- MacGyver

LDMX will look for a dark matter event in an initial sample of $4 \times 10^{14}$ electrons incident on a target made of tungsten. The target is one tenth as long as its radiation length, to best ensure that it produces bremsstrahlung events and little else.

Because the signal events involve very low energy electrons which have experienced a recoil from the dark bremsstrahlung process, it is important to eliminate stray low-energy particles from the electron beam halo upstream from the target. This is done with a strong B field placed before the target.

LDMX borrows technologies from a number of other experiments. It will take its tracking systems from the Silicon Vertex Tracker (SVT) being used in the Heavy Photon Search (HPS) at Jefferson Lab. SVT has high-precision tracking for low-momentum electrons (around 3 or 4 GeV, the energy targeted for LDMX's beamline) and hit-time resolution down to 2 ns. Arguably the most cutting-edge technology that LDMX borrows is the high granularity layered sampling ECAL being developed for the CMS HL-LHC upgrade. It has excellent
Figure 2.2: This conceptual cartoon of the LDMX experiment shows the electron beam hitting the thin tungsten target, recoiling in a dark bremsstrahlung event and producing a shower in the ECAL, while the dark matter particles go through the detector unseen [1].

Figure 2.3: This cutaway of LDMX shows the various detector subsections in detail [1].
resolution for electromagnetic showers, the potential for implementing MIP tracking, and time resolution well-suited to LDMX. The HCAL could potentially use technology from numerous other experiments including CMS or Mu2e.

The studies performed for this thesis are solely focused on the ECAL, and thus we will place emphasis on the ECAL among the various subsystems of the LDMX detector.

**ECAL**

The basic concept behind the operation of an electromagnetic calorimeter is fairly straightforward. When an electron passes through matter it has a chance of producing a photon through bremsstrahlung, and individual photons traveling through matter have a chance of pair production, creating an electron and positron. Thus photons or electrons traveling through matter go through a series of pair production and bremsstrahlung events to create a "shower" as in Fig. 2.4, which widens as the particles move deeper into the calorimeter. A naive shower model is that bremsstrahlung only happens if an electron has energy \( E > E_c \). Each electron with \( E > E_c \) travels an average distance \( X_0 \) called a radiation length before giving up \( 1 - \frac{1}{e} \) of its energy to a photon, which then splits that energy among an electron and positron in pair production after traveling \( \frac{9}{7}X_0 \) on average. The value of \( X_0 \) varies by material. Once the energy of the electrons in the shower is \( E < E_c \), they lose the remainder of their energy to ionization. This behavior means that particles with more energy create larger showers as in Fig. 2.5, and that particles that travel through more material lose their energy quicker: thus, particles that come into the ECAL at large angles relative to the normal of the layers produce showers that are wider and don’t penetrate as far into the ECAL as particles with the same energy moving straight through [12, 10, 9].

The LDMX ECAL is a sampling calorimeter, meaning that it has one material that generates the shower through bremsstrahlung and pair production, called the "absorber", and another material that measures the energy and position of the particles. The LDMX ECAL consists of 34 layers each composed of seven hexagonal modules as in Figs. 2.3 and 2.6. Each module has a maximum possible granularity of 432 pads, each with a surface area of 0.52 cm², that have individual channels recording the amount of energy deposited in them by charged particles passing through the ECAL.

Within each module, a layer of tungsten acts as the absorber which produces the shower. Silicon is used for the readout device which measures the energy. Silicon is the key ingredient that allows for the ECAL’s performance: it is radiation-hard and can also be manufactured at very small scales, allowing for the high granularity of the ECAL sensors. The silicon is doped to make a pn junction, which is reverse biased to increase the size of the depletion zone. Particles traveling through the depletion zone ionize electron-hole pairs, which move to the edges of the pn junction and generate electrical signals corresponding to the energy of the measured particles in the shower and simultaneously recording the particle’s trajectory. The silicon must operate colder than room temperature to avoid thermal effects of the electron beam and noise generated by excessive heat. The high granularity in particular allows the ECAL to track charged hadrons that cross multiple layers and to identify isolated charged
Figure 2.4: A cartoon depicting the process by which electromagnetic showers form by bremsstrahlung and pair production events [12].

Figure 2.5: The depth of the shower in the ECAL depends on the energy of the particle that starts the shower as depicted above [9].
hadrons in individual layers that deposit charge at the scale of a minimum ionizing particle (MIP), which produces the smallest possible electric signals.

2.2 Identifying Dark Matter Signal

The kinematics of an electron producing dark matter through a dark bremsstrahlung process depend on the A’ mass. For masses of the A’ or $\chi\chi$ pair produced that are near the electron mass, the differential cross-section for dark matter production is peaked where the dark matter carries the majority of the energy from the beam and the electron loses most of it. Usually normal bremsstrahlung events have the exact opposite behavior, meaning that LDMX mostly just has to contend with a rather small though still tricky set of normal bremsstrahlung backgrounds, simply by design.

The clearest indicator of dark matter production is the recoil electron’s transverse momentum ($p_T$) because it turns out that there is a large discrepancy in its distributions for signal and background (Fig. 2.8). Due to the power of $p_T$, the aim of much of the LDMX analysis work is to discriminate between signal and background without making use of $p_T$, so that it is in our back pockets to provide extra validity to any dark matter discoveries.

2.3 Types of Backgrounds

"I coulda been a contender."

- Terry Malloy

Background rejection is vital for the successful operation of the experiment, because without background rejection any potential dark matter discoveries made by LDMX could
Figure 2.7: Details the major backgrounds to consider in LDMX and their rates relative to the number of beam electrons incident on the target (y axis). For context, the relative rate of dark matter production that we are looking for is on the order of $10^{-14}$ [1].

just as easily be normal physics processes that look similar to the expected dark matter signal described in Sec. 2.2. It’s worth understanding exactly what the different types of backgrounds are to get a better sense of why some are hard to throw out and how the radius of containment methods we describe in Chapter 3 might be useful. Again, this discussion is largely based on [1]. Fig. 2.7 details the nature of the most important backgrounds.

Low-energy Particles and Beam Impurities and Electrons with No Interactions in the Target

In the case of an electron from the beam with very low energy, similar to a recoil e-, that makes it through the target and the B fields, the tagging tracker is used to measure its trajectory and distinguish it from a 4 GeV electron that had a recoil after the target.

As for full 4 GeV beam electrons that go through the target, they tend to produce a very large shower in the ECAL and are easily vetoed. However, as we will see below, photo-nuclear or electro-nuclear interactions in the ECAL during shower development can sometimes make these particles have low energy showers anyway, in which case they are more of a concern.
Figure 2.8: The distribution of transverse momentum for recoil e-’s created due to a variety of signal masses and for background. Note the discriminating power of $p_T$ comes from the extreme drop off in high transverse momentum events for background relative to all signal (the various masses are different possible A’ masses). The background drop off in black on the figure goes as $1/p_T^3$ for high $p_T$ [1].

**Hard Bremsstrahlung**

"Hard" simply refers to a photon being produced through bremsstrahlung with high energy, which usually occurs in the target. These events drop the recoil e- energy below the threshold expected for a dark matter recoil e- (this energy varies depending on the models used, but for our studies we denote hard bremsstrahlung as anything with a recoil e- below 1.5 GeV). These events occur at a relative rate of $3 \times 10^2$ per incident electron on the target. The result of a hard bremsstrahlung is generally two showers in the ECAL that add up to 4 GeV, one shower from the electron and one from the photon. If the showers are resolved poorly (they overlap nearly completely, or one shower has very little energy deposited in the ECAL), this can create a false positive.

This problem is exacerbated in the rare case that the bremsstrahlung photon under-
goes a photonuclear reaction (essentially the photon colliding into the nucleus of an atom and interacting with subatomic particles), or a conversion to muons or pions. The hard bremsstrahlung with photonuclear interaction events are at the center of the motivation and design of the radius of containment variables.

**Electro-nuclear Interactions in Target**

Electro-nuclear interactions have a similar composition to photo-nuclear reactions but occur at lower rates and have a wider $p_T$ distribution. Because of their similarities to photo-nuclear interactions, they can be rejected in essentially the same way.

**Neutrino Backgrounds**

Neutrino backgrounds are unique among the other backgrounds in that they are not caused by detector mismeasurements that lead to false positives, but rather are fundamental physics processes that can produce low energy electrons that have no second shower counterpart since the neutrinos pass through matter essentially undetected. Their rates are incredibly small, and as such they are not a significant background concern, but they are irreducible physics background in some cases. Because of this, neutrino backgrounds set limits on the sensitivity of missing momentum experiments like LDMX.

**2.4 Contemporary Accelerator-Based Dark Matter Experiments**

”No man is an island.”

- John Donne

Many different approaches to dark matter searches have been explored. Missing mass searches such as Babar and PADME use kinematics to attempt to find the mass of the $A'$, but rely on reconstructing the trajectories of every particle involved in a signal event. Dark matter re-scattering searches such as LSND, E137, and MiniBoone involve looking for the annihilation of $\chi$ particles into a virtual $A'$ that then decays into standard model fermions. LDMX can look for the same annihilation but does not need the extra process that produces the fermions, as the missing momentum approach of LDMX is sensitive to the $A'$ itself: this makes it more sensitive than a dark matter re-scattering search. [1]
Chapter 3

Radius of Containment

The goal of my work on LDMX has been to produce novel features to implement in the ECAL BDT to improve its background rejecting performance. When an electron or photon hits the ECAL, it produces a shower of charged particles, and the shower’s size and shape depend on the trajectory and energy of the initial particle. We characterized the size of these particle showers as they moved through the ECAL to produce a set of numbers called the radius of containment. The ECAL is made of flat, discrete layers oriented perpendicular to the axis of propagation of the electron beam: the radii of containment are just the average radii of circles in the plane of each ECAL layer such that they would encompass a fixed amount of the total energy deposited by the shower, in our case either 68 percent or 90 percent of the total energy in that layer. The BDT was then designed to draw these Radii of Containment around the trajectory of the electron using tracking information, and around the inferred trajectory of whatever particle was produced in the bremsstrahlung process, and add up the energy contained inside and outside the radii. If there is very little energy contained in a region where one would expect to see energy deposits from a photon shower, this is a strong indication of a dark matter signal event. The radii of containment are far more sophisticated than any other variable used in the BDT, and ultimately have been shown to produce significant improvements in the BDT’s performance.

The radius of containment idea was actually inspired by similar work done for the Compact Muon Solenoid experiment (CMS) at CERN. Their original figure (Fig. 3.4) from a technical proposal for the Phase-II upgrade served as a benchmark for the design of LDMX’s radius of containment variables.

3.1 Boosted Decision Trees (BDTs)

As BDTs are at the heart of the radius of containment studies, it is worth understanding them in a bit of detail. Boosted Decision Trees are a relatively simple form of machine learning used in the case of LDMX as classification trees, classifying whether events are signal or background: "decision trees" are basically what they sound like, branching sets
CHAPTER 3. RADIUS OF CONTAINMENT

Figure 3.1: A simple example of the ideas behind a decision tree. The datasets being used get broken into consecutively smaller subsets while an associated decision tree is simultaneously incrementally developed. Decision nodes, like Temperature for example, have branches (Hot, Mild, Cool). Leaf nodes, the only one of which here is Play Golf, represent a classification (Yes or No). The topmost decision node in the tree corresponds to the best predictor, in this case Outlook [13].

of checks that increase or decrease the confidence of the algorithm in classifying an event as signal or background, while the "boosting" refers to assigning weights to the relevance of different decision branches as they are shown to be more or less significant over time. The simplest way to understand the use of decision trees is through Fig. 3.1. It should be noted that this figure overgeneralizes the process used for LDMX: in our BDT, we actually employ a tree ensemble model which has a set of classification and regression trees (CART), a more nuanced version of a decision tree. When a CART makes a choice between branches, it does not arrive at an absolute yes or no answer. Rather, the tree develops a discriminant value which varies between 0 to 1 and determines its confidence in a yes or no in a given leaf. Each decision node applies a score that raises or lowers the discriminant based on the branch chosen. The trick in using CARTs is to choose the best minimum cutoff value on the discriminant such that the BDT ends up with the fewest false positives and the greatest true positives in the leaves. In the case of LDMX, this means the fewest background events being passed off as signal and the most signal events being retained and not thrown out with the background.

LDMX implements XGBoost, a gradient boosting library in Python. In the following, we’ll go into a little depth as to how XGBoost and supervised learning actually works, as it is central to the radius of containment studies.

The use case for XGBoost is in "supervised learning", where training data with "features" \( x_i \), just the different CARTs mentioned earlier, are used to predict a target variable \( y_i \). In
supervised learning, the "model" is the framework for making a prediction $\hat{y}_i$. For example, one could have a simple linear model where $\hat{y}_i = \Sigma_j w_j x_{ij}$, where the $w_j$ are weights.

In order to build the BDT, it needs to be "trained". This involves finding the best "parameters" $\Theta$, or the things we need to learn from the data. In a linear model, $\Theta = \{w_j | j = 1, \ldots, n\}$. We define an objective function that measures how well our model fits the training sample we give it, which we will attempt to optimize through training:

$$obj(\Theta) = L(\Theta) + \Omega(\Theta)$$

$L(\Theta)$ is the training loss, $L = \Sigma_{i=1}^{n} l(y_i, \hat{y}_i)$, where $l$ is the function describing the deviation of the model from the training data. The simplest example would be square loss:

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$$

which is basically just variance. $\Omega(\Theta)$ is regularization, a measure of how complex the model is. A simple example would be $\Omega(\Theta) = \|w\|^2$. Model complexity matters, since low complexity models tend to avoid overfitting, analogous to a lower order polynomial regression. As we mentioned earlier, XGBoost employs CARTs which produce a score function that ultimately translates to a discriminant value based on the branches chosen. We will call the score function that the CART produces for a given event $f_k(x_i)$. It becomes useful to reformulate our parameters with this in mind, so $\Theta = \{f_i | i = 1, \ldots, K\}$.

Instead of learning weights, the BDT is learning the function scores of the different trees. For example, the actual regularization used by XGBoost is $\Omega(f) = \gamma T + 0.5 \lambda \Sigma_{j=1}^{T} w_j^2$, where $T$ is the number of leaves.

With the mathematical machinery in place, it’s not much harder to understand the basic operation of XGBoost. Our prediction $\hat{y}_i$ is just the sum of the scores in all the trees:

$$\hat{y}_i = \Sigma_{k=1}^{K} f_k(x_i)$$

where $K$ is equal to the total numbers of CARTs in the ensemble. This sum is built up step by step in the training stage of the BDT: one score is added to the prediction at a time, such that at the $t^{th}$ step the prediction is $\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$. This is done simply because it would be far too complex to learn all of the tree scores at one time. If we take a look at the objective, where the loss function is just the mean squared error, we see that the key to optimization actually lies in the way in which we employ this additive training.

We have the objective function at step $t$:

$$obj^{(t)} = \Sigma_{i=1}^{n} [2(\hat{y}_i^{(t)} - y_i)f_t(x_i) + f_t(x_i)^2] + \Omega(f_t) + \text{constant}$$

where the constant contains all terms not of the current step $t$. It is extremely useful to Taylor expand the loss function in terms of $f_t(x_i)$:

$$obj^{(t)} \simeq \Sigma_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) + \text{constant}$$

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) ; \ h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$$
Now the above is what we're aiming to optimize with each addition of a new tree at a given step \( t \). The optimization process has been greatly simplified by introducing the Taylor expansion, as the objective function ultimately only depends on \( g_i \) and \( h_i \), and it should be noted that this allows for flexibility in terms of what loss functions can be used in XGBoost.

With a little more work (that we will skim over) using the regularization we mentioned earlier that is implemented by XGBoost, one can arrive at the following equation for the objective function:

\[
obj^{(t)} = -\frac{1}{2} \sum_{i=1}^{T} \frac{G_j^2}{H_j} + \gamma T
\]  

This equation is basically rearranging the objective as terms by individual leaves, where \( T \) is the number of trees and we define an instance of a particular leaf as \( I_j \), and the terms \( G_j \) and \( H_j \) sum the Taylor expanded terms of the loss over each instance of a given leaf: 
\( G_j = \sum_{i \in I_j} g_i \) and \( H_j = \sum_{i \in I_j} h_i \). The tree is then grown "greedily", where the summed term of the objective is split at each leaf node into different terms for different parts of the tree in hopes of finding an optimization of the objective [5].

One final thing to mention is the receiver operating characteristic (ROC) curve. As shown in Fig. 3.2, the ROC curve is a key visualization tool for analyzing the performance of a BDT and comparing the differences in performance between various features.

With a slightly more in-depth understanding of the operation of a BDT in hand, we will now show the progress made with the LDMX ECAL BDT by introducing a new group of features, all based on the radius of containment.

### Previous Background Rejection Methods

As the radius of containment studies were designed to improve upon previous background rejection efforts, it is worth mentioning previous efforts to fight backgrounds in the ECAL. These were largely restricted to manual cuts on specific variables, the most impactful of which were the number of total hits and the total energy deposited in the ECAL, and a basic implementation of a BDT. The features implemented in the BDT before the radii of containment included things such as the maximum depth of the shower in the ECAL, the standard deviation of hits in the x and y directions, and recording which ECAL layers had the most hits.

### 3.2 New Method: Implementing Radii of Containment

The power of the LDMX ECAL lies in its high granularity (i.e. the large number of small pads for energy deposition in each hexagonal module, 432 channels to be exact), and the fact that it is a sampling calorimeter with many layers to segment electromagnetic showers longitudinally. We aimed to make use of these two main aspects of the ECAL in a set of
Figure 3.2: The basics of ROC curves: the y axis is the true positive rate or TPR for classifications (for LDMX, how much the BDT correctly identifies a signal event), and the x axis is the false positive rate or FPR (for LDMX, how much the BDT mistakenly identifies backgrounds as signal events). The curve is made by selecting a continuous range of values of the discriminant produced by the BDT and plotting their performance. The optimal BDT has a normalized area under the curve (AUC) of 1, so the closer the curve gets to the top-left corner of the plot, the more efficient the BDT is. A diagonal line would represent a 50/50 chance of correct classification, i.e. the BDT is just guessing.
new features for the BDT, and what resulted were the radii of containment, a set of circles in the plane of each layer of the ECAL that characterize the width of the shower, measured to a high level of precision thanks to the granularity of the modules.

The basic concept of the radii of containment is illustrated in Fig. 3.3. First, the trajectory of an electron obtained from tracking information is projected through the ECAL. Next, the 68 percent radii of containment are drawn centered on the intersection of the particle’s trajectory with each layer of the ECAL, and concentric annuli that are integer multiples of the radius of containment are drawn as well, with the largest multiple being 5 times the radius of containment. The annuli/circles are then used to separate the hits in the ECAL: in each circle/annulus as well as outside the largest annulus we sum up the number of hits, the total energy, and the standard deviation in x and y. These numbers obtained in each layer for each annulus/circle and the outside region are then summed through all the layers (for example, all the energy in the central circle in each layer is added together, all the energy in the first annulus in each layer is added together, the second annulus, and so on). The result is ultimately a significant performance boost in the rejection capabilities of the ECAL BDT.

Defining the Radii

We took inspiration from the CMS experiment’s work on characterizing their ECAL showers (Fig. 3.4): they plot the radii of circles in the plane of each layer of the ECAL
which contain on average 68 percent and 90 percent of the total energy deposited by an electromagnetic shower in that given layer. The CMS showers are much higher in energy than the showers in LDMX are planned to be, so the first step was to reproduce the radii of containment plots for LDMX.

As discussed in Section 2.1, electromagnetic showers have shapes which are largely governed by the momentum of the particles creating them and the angle at which the particles enter the ECAL. With this in mind, we decided to plot the momentum and angle of particles entering the ECAL (Fig. 3.5) to see what general trends could affect the shape of the showers. We decided to utilize this plot to segment the radii of containment in different binnings of the momentum-theta phase space (Fig. 3.6). Before running the BDT, the angle and momentum of the particles would be obtained via tracking info (although currently tracking is not implemented in the LDMX simulations, so the angle and momentum is obtained right in front of the ECAL), and would thus dictate which binning of the radii to use for the event.

We then set about creating the radii of containment, shown in Figs. 3.7 and 3.8. For reasons addressed in Section 4.2, we opted to focus on the 68 percent radius of containment for implementation in the BDT, and thus the 90 percent radii of containment will be ignored for the moment. Figs. 3.7 and 3.8 are generated by averaging the containment radii of $O(10^5)$ events per binning in Fig. 3.6.
Figure 3.5: Momentum vs angle with respect to the normal of the ECAL for electrons entering the detector in single electron signal events. The lightest yellow curve in the plot indicates the most frequent values.

Figure 3.6: The 4 different binnings used for the radii of containment, as follows: 1) $p > 500, \theta < 10$  2) $p < 500, \theta < 10$  3) $10 < \theta < 20$  4) $\theta > 20$
Figure 3.7: The various 68 percent binned radii of containment generated for LDMX. Note that in the back layers there is a large deviation in radii, the radii are all very large, and there is very little shower energy deposited in the back layers. Most of this is because the ECAL is designed to be longer than the shower, and thus the hits in the back layers at the end of the ECAL are very sporadic.

Figure 3.8: Because of the low statistics in the back of the ECAL, we opted for a linear fitting in the back layers. This should help the BDT perform optimally.
Figure 3.9: This ROC curve demonstrates the improvement of the BDT containing the radius of containment features (New) over the previous features (Nominal) for a range of possible dark photon masses. This ROC curve is the result of running on a sample of the order $O(10^{14})$ electrons incident on the target.

**BDT Performance with Radii of Containment**

With the radii of containment calculated, the next step was to run the BDT with the new features. The result was an immediate marked improvement in the performance of the BDT as shown in Fig. 3.9. It is important to note that the order of the curves from most to least efficient has changed in the new BDT, in congruence with the transverse momentum distributions in Fig. 2.8 that indicate that it is easiest to identify high mass signal events and hardest to identify low mass signal events because of their kinematics being respectively least and most similar to that of background. The old BDT did not order the events in efficiency by mass, and we address possible reasons for this in Section 4.2. As Fig. 3.9 is the most recent ROC curve made from running the BDT over the full sample of approximately $10^{14}$ available simulation events, where an individual event constitutes a single electron entering the ECAL and producing either a dark or standard bremsstrahlung, this is the most reliable current performance of the radius of containment features in the BDT.
Figure 3.10: This ROC curve is the result of a BDT with features that are linear combinations of the outside and annulus variables: in addition to summing the energy, hits, etc. outside the largest annulus, we also include everything outside the fourth largest annulus, outside the third largest, etc. There is clear improvement, however this ROC curve was produced over a sample on the order of $O(10^6)$ events, and thus still requires more testing.

However, there are many ways to optimize the performance of the BDT through feature engineering: things as simple as making linear combinations of the features, splitting the sums of the features along the ECAL longitudinally, etc. can produce noticeable improvements. Take Fig. 3.10: the inclusion of linear combinations of the outside and annulus features produces a clear performance boost. More work is still needed to verify these results.

A matter that has yet to be resolved is the extent to which the new BDT makes use of transverse momentum, which is something we are trying to avoid as discussed in Section 2.2. Fig. 3.11 indicates that some minor transverse momentum bias may be present. There are many ways to approach reducing transverse momentum, as we discuss in Section 4.2.
Figure 3.11: These figures demonstrate transverse momentum bias in the BDT with respect to signal efficiency - a distribution that is not flat or that rises with increasing transverse momentum ($p_T$) hints at the BDT making use of transverse momentum, which is undesirable. The orange line is for the New BDT, the blue line is for the Nominal BDT.
Chapter 4

Discussion

4.1 Conclusion

"It ain’t over ’til it’s over."

- Yogi Berra

LDMX is a uniquely powerful dark matter search currently in development that will look for light dark matter through fixed target electron beam interactions that generate a dark bremsstrahlung event. Its power comes in large part from its state of the art ECAL, which it borrows from the HL-LHC CMS upgrades.

Background rejection is crucial for the experiment’s success, and a BDT is implemented for rejection of background events in the ECAL. Here we have shown that a new set of features based on the radius of containment, which is the radius of a circle that on average would contain 68 percent of the total energy deposited by a shower in a given layer of the ECAL, produces a marked improvement in the performance of the BDT.

4.2 Further Work

There is plenty of room for further improvements to the BDT, as can already be seen in Fig. 3.10. There are also numerous issues that need to be resolved before significant progress can be made.

A major bug that was recently discovered to be plaguing our analysis is duplicate events in our simulation samples, produced by passing identical seeds for generating the events to GEANT4, the simulation framework we use. As a result of this duplication, which seems to be more prevalent in the 0.01 and 0.1 GeV signal mass ranges, training the BDT is not totally realistic. This duplication problem could also potentially explain why the nominal BDT was most efficient for the 0.01 and 0.1 GeV masses - the test is to see if without the duplication all of the signal masses for the nominal BDT have equivalent ROC curves. If
this is the case, it does still indicate that our new BDT is using fundamentally different and more powerful aspects of the showers for signal identification, as we do expect to have the easiest time identifying the heaviest A' signals.

It is likely that including the 90 percent containment radii will produce improvements. They were not originally implemented in the BDT due to a series of unfortunate events: originally the radii of containment were generated with events that had 2 electrons producing showers in the ECAL simultaneously instead of 1. These 2 electron files had one recoil electron and one 4 GeV electron from the beam entering the ECAL together. The goal was to make separate radii of containment for the beam electron and recoil electron, however it was not apparent at the time that the simulation information that linked the showers to each of the two separate electrons was not properly saved, and attempting to tease out the information for one shower and not the other was impossible. These preliminary radii of containment had very strange 90 percent containment curves, which we chose to ignore. Eventually we realized that there was a bug and went back to 1 electron events where there was no ambiguity as to the origin of the showers: the 90 percent curves then looked similar to the 68 percent curves and the curves present in Fig. 3.4. We had already begun training the BDT with only the 68 percent radii and their concentric annulus multiples, and chose to stick with that method because it had begun producing positive results.

Other feature engineering possibilities include splitting the radii of containment sums into multiple longitudinal regions, i.e. sum all the data points from layers 1 through 4, 5 through 15, and 16 through 34. Different linear combinations of the existing features beyond those implemented in Fig. 3.10 are also possible. However, reducing the number of features implemented currently to ascertain which are the most impactful is an essential next step before we can add even more features to the BDT. Having too many features is undesirable as it will simply overtrain the BDT.

To address concerns regarding $p_T$ biasing, it is worth re-running the BDT with radii of containment that do not have binning in momentum and angle: while the binning does likely introduce additional power, it may utilize more $p_T$ information because of its direct relation to recoil electron kinematics. The BDT training could also be re-run with weightings for different $p_T$ values to remove the biasing.
Bibliography


