Signal vs Background Discrimination for LDMX using ParticleNet

Pradyun Solai
Carnegie Mellon University

This material is based upon work supported by the National Science Foundation under Grant No. 2244348.

Abstract

Dark matter is a form of matter that appears to only interact with standard matter through gravity. It is known that dark matter makes up around 85% of the universe's matter, but very little is known about its nature. Previous experiments looking for dark matter particles have not given promising results, but the Light Dark Matter Experiment (LDMX) proposes to look in a mass range that has not been extensively probed before. In the experiment, a beam of electrons is incident on a fixed target, in which it can hypothetically produce dark matter particles. Using the data from the LDMX detectors, it is possible to determine whether an event might have produced dark matter. In this project, I train a machine learning model called ParticleNet to discriminate between background and signal events in LDMX, and find it to be an effective and accurate discriminator to identify a hypothetical dark matter signal.

Introduction

The search for Dark Matter

Dark matter is an elusive form of matter that does not interact with light and electromagnetism. Astrophysics has seen its large-scale gravitational effects, but its small-scale nature is still unknown. The search for dark matter's particle nature has not yielded many results yet. Most of this search has been done in GeV-TeV mass range by searching for WIMPs, but not much has been done in the sub-GeV range.

Light Dark Matter eXperiment (LDMX)

- Experiment to detect theorized dark matter particles in the sub-GeV mass range
- 8 GeV electron beam incident on a thin target
- Electron interacts with the target, releasing a recoil electron and other particles
- Resulting particles detected by Electromagnetic Calorimeter (ECal) and other detectors, in which particles can deposit energy
- If the total energy/momentum of the resulting particles is much less than the original energy/momentum, then a dark matter particle may have been produced

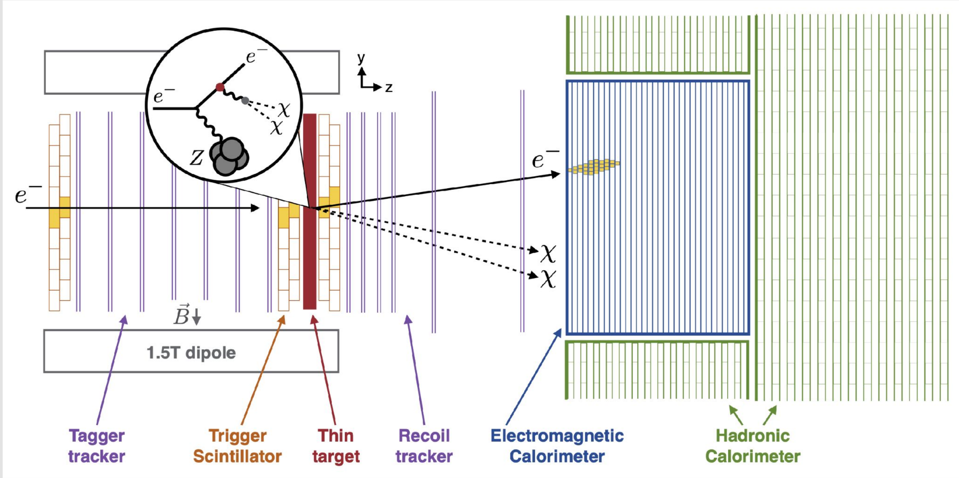
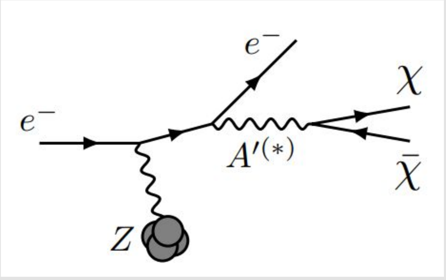


Diagram of LDMX setup and detectors



Possible mechanism for DM production via A' mediator

Images from https://arxiv.org/abs/2203.08192

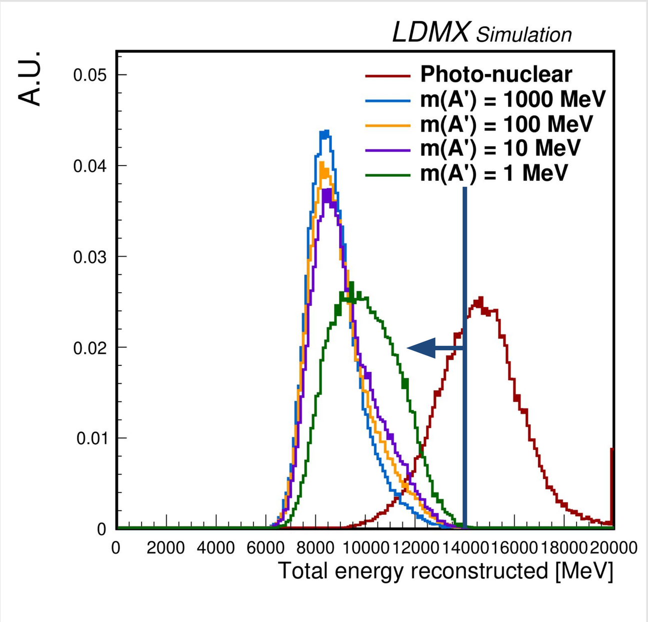
Methodology

Input data

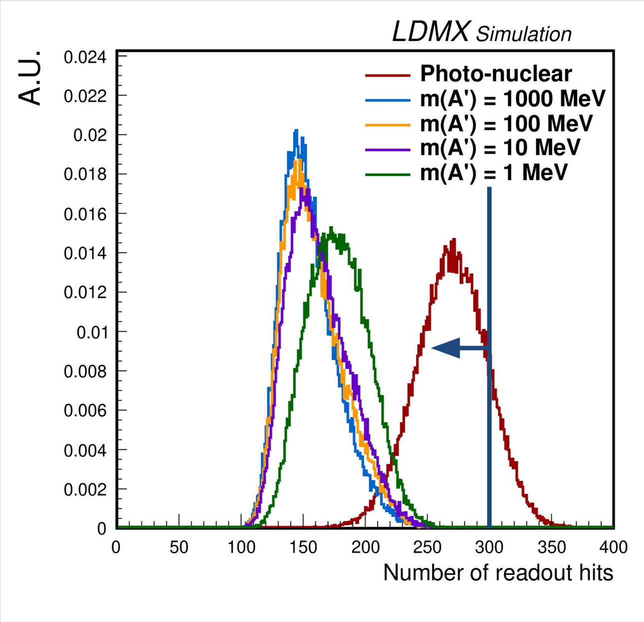
- Data generated using ldmx-sw simulation framework
- Each event is two 8 GeV electrons hitting the target at the same time
- 5 samples
  - 1 photonuclear background sample (DM not produced)
    - 130,000 events
  - 4 signal samples (DM produced)
    - A' (DM photon) masses of 1, 10, 100, and 1000 MeV
    - 200,000 events each
- Each ECal hit represented as a tuple - (x, y, z, energy)
  - (x, y, z) is the position of the hit in the ECal
  - energy is the energy deposited in the cell.

Preselection

- Filters out events based on total reconstructed energy and number of readout hits
- Total reconstructed energy < 14 GeV
- Number of Readout Hits < 300
- Increases computational efficiency



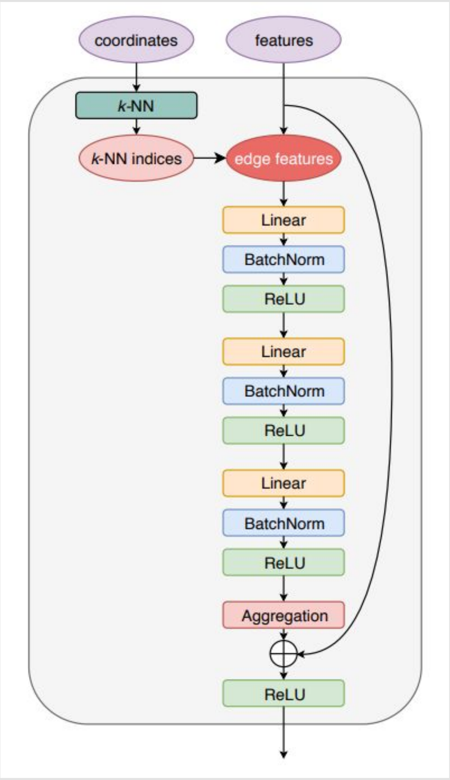
Total reconstructed energy distribution for each sample, with preselection threshold



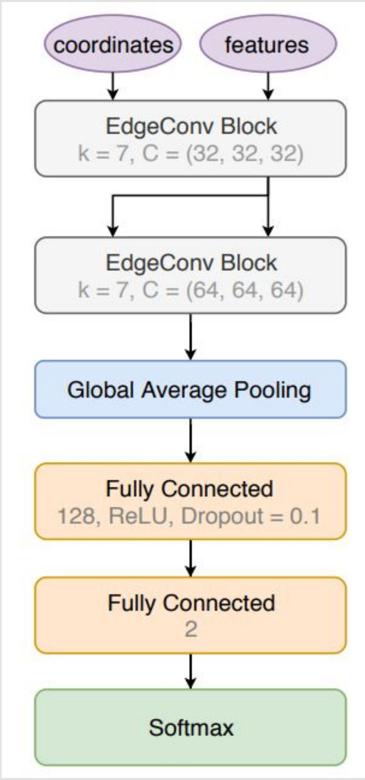
Number of readout hits distribution for each sample, with preselection threshold

ParticleNet

- Dynamic Graph Convolutional Neural Network originally made for jet tagging at the Large Hadron Collider
- Uses EdgeConv blocks to take advantage of the structure of ECal
- Trained on around 36,000 background events and 12,000 events from each signal sample



Structure of an EdgeConv block

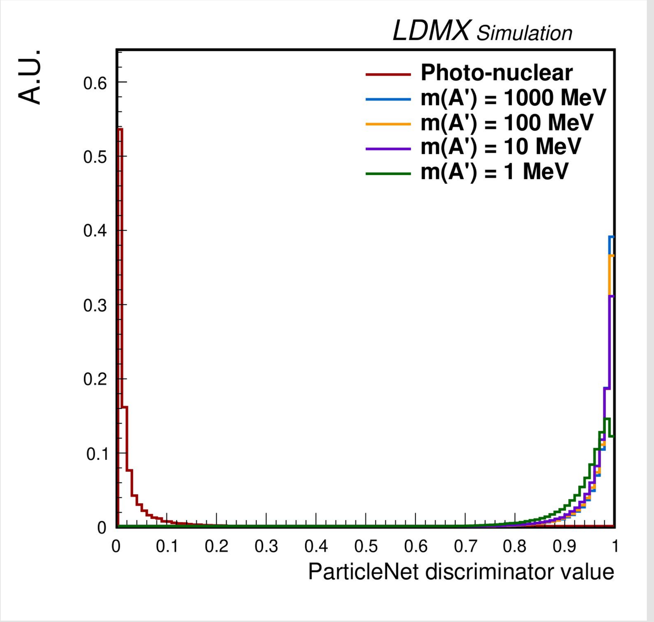


ParticleNetLite architecture

ParticleNet images from https://arxiv.org/abs/1902.08570

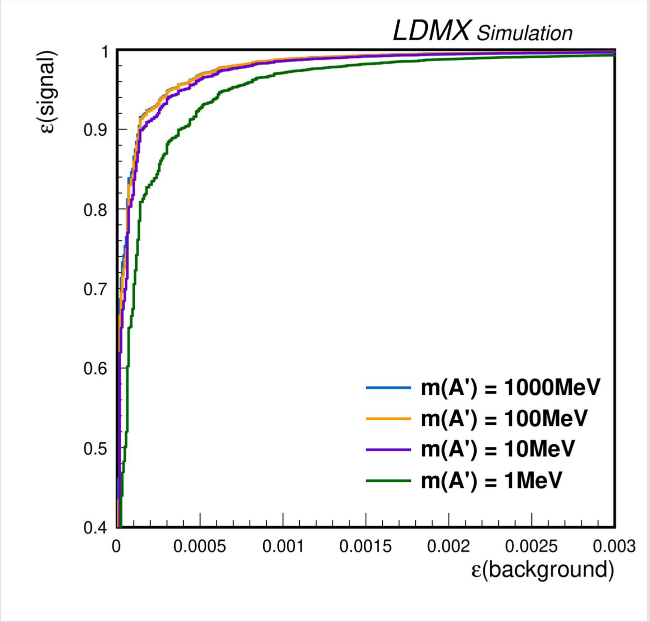
Results

Discriminator Distribution



ParticleNet discriminator distribution for background and signal. The discriminator is the output of ParticleNet, which should be close to 1 for signal and close to 0 for background.

ROC Curves



ROC curves adjusted by preselection efficiencies. An ROC curve plots the true positive (y-axis) and false positive (x-axis) rates at different classification thresholds.

Signal efficiency table

Signal efficiencies		
A' mass	10^{-3} background efficiency	10^{-4} background efficiency
1 MeV	97.10%	70.52%
10 MeV	98.58%	83.67%
100 MeV	98.78%	85.96%
1000 MeV	98.78%	86.57%

Conclusion

ParticleNet is an effective and accurate machine learning model to classify signal vs background events in LDMX when two electrons hit the target at the same time.

Acknowledgements

I would like to express my thanks to Dr. Valentina Dutta for giving me the opportunity to complete this project and for helping me throughout its duration.
Secondly, I would like to thank Dr. Pritam Palit for helping with this project as well.