Developing a BDT for finding Dark Photon Decays to Two Electrons within the ATLAS Detector

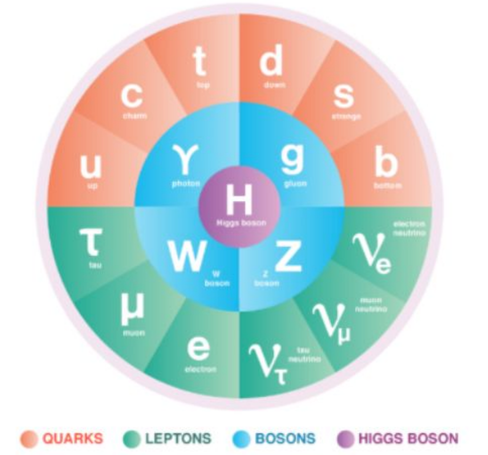
2024 REU Program at Columbia University - Nevis Labs
Amelia Stevens, Haverford College
August 2, 2024

Outline

The Standard Model of Particle Physics

The Standard Model (SM) is the current theory that describes the fundamental particles and the forces that govern their interactions

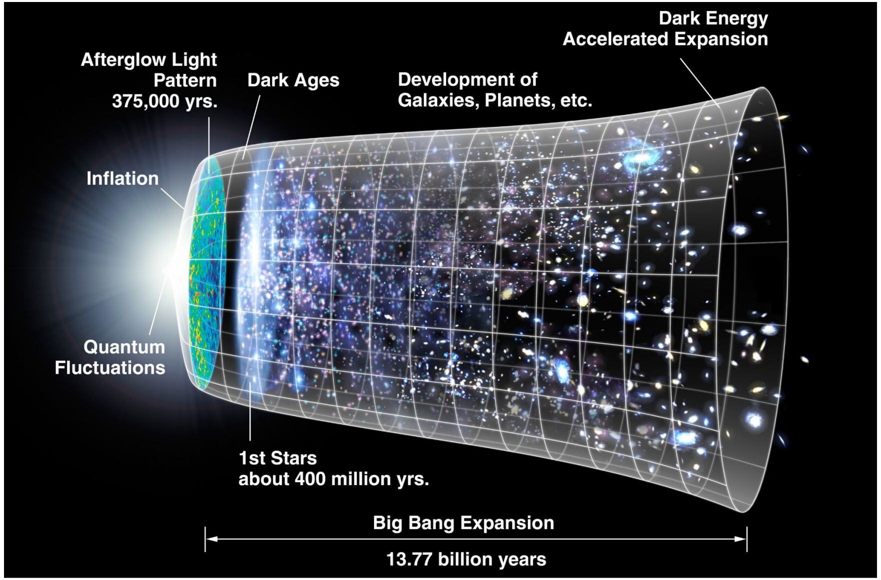
- Fermions - constituents of matter
    - Six quarks
        - Up, down, strange, charm, top, bottom
    - Six leptons
        - Electron, muon, and tau, three neutrinos
- Bosons - mediate the fundamental forces
    - Gluon - strong interaction inside atomic nuclei
    - Photon - electromagnetic interaction
    - W and Z boson - weak force
- Higgs Boson - excitation of Higgs field that gives particles mass



The Standard Model of Particle Physics

The Standard Model is inherently flawed…

- Gravity, the force that holds galaxies together, remains an outsider
  - Its influence at the subatomic level is negligible
- Dark Matter and Dark energy constitute most of the universe's energy and mass, yet the SM offers no explanation of their existence
- Big Bang should have created equal amounts of matter and antimatter, yet we see very little antimatter in the universe
  - all of these particles have anti-particle partners which have same quantum numbers except for EM charge



The Large Hadron Collider (LHC)

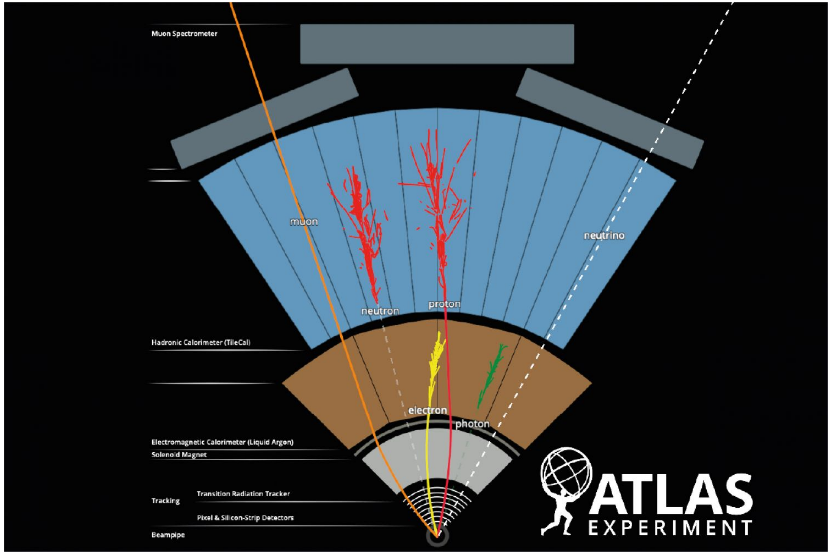
The LHC is the largest particle accelerator in the world at the European Organization for Nuclear Research (CERN) on the Swiss-French border

- Protons are accelerated by superconducting magnets to an energy of 13.6 teraelectronvolts (TeV) and collide at 40 MHz
- Collisions give rise to creation of new particles
- Four main experiments
  - ATLAS
  - CMS
  - ALICE
  - LHCb

The ATLAS Detector

ATLAS is a gigantic particle detector situated at the Large Hadron Collider designed to investigate the particles and forces described beyond the Standard Model

- **Four Major Subsystems:**
    - **Inner Detector:** This innermost layer tracks the trajectories of charged particles produced in collisions
    - **Calorimeters:** These detectors measure the energy of different particles: electrons and photons (electromagnetic calorimeter), hadrons (hadronic calorimeter)
    - **Solenoid Magnet:** A powerful magnet that bends the trajectories of charged particles
    - **Muon Spectrometer:** The outermost layer detects and measures momenta of muons, which are heavy particles can pass through most other detector parts



Outline

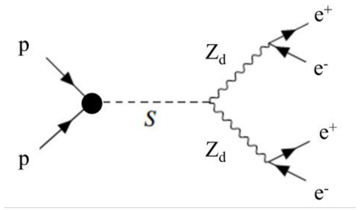
1. Introduction
    a. The Standard Model
    b. Instrumentation
        i. The LHC
        ii. The ATLAS Detector
2. Search for Displaced Electron from Long-Lived Dark Photon Decays
    a. Motivation
    b. Data and Simulated Samples
    c. Signal Variable Distributions
    d. BDT Introduction
3. Results
    a. BDT high mass vs. BDT low mass
    b. BDT with high delta M vs. BDT with low delta M
4. Conclusion and Next Steps

The Search for Displaced Electrons from Long-Lived Dark Photon Decays

The Standard Model of particle physics is incomplete. We are searching for new particles, including a hypothetical "dark photon." The Hidden Abelian Higgs Model introduces two new particles: the "dark Higgs" and the "dark photon." Focus on instances where the dark photon lives long enough to travel a measurable distance before decaying into two electrons.

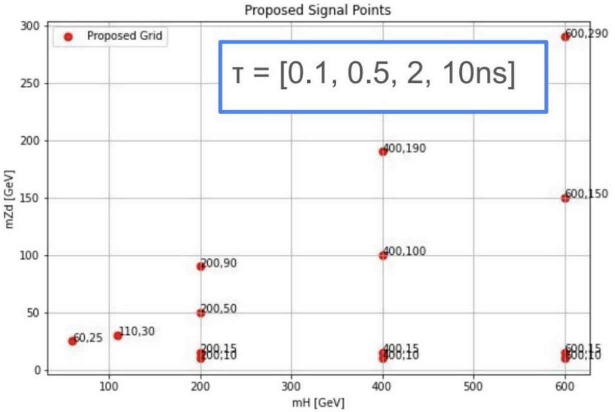
- Most previous searches for dark photons have focused on the case where the dark photon decays promptly
- Study the tracks of these electrons in a particle detector
- Analyze electron properties (momentum, time, distance from the interaction point, etc.) to identify potential dark photon decays
- Use these properties as input for a machine learning model (BDT) to discriminate electrons which come from dark photon decays from background

Feynman Diagram of electron pairs induced by Gauge Boson (Zd) in the Dark Higgs Boson (S) model



Data and Simulated Samples

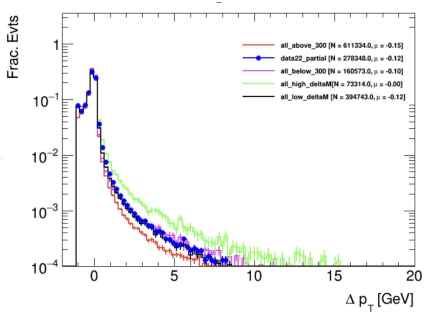
- Training is done on simulation of HAHM dark photons which decay to electrons (signal) and data taken with the ATLAS detector in 2022
- I select only events with at least 2 electrons
- Signal: refer to signal grid (mS, mZd, tZd)
  - a scenario with a Dark Higgs boson of 400 GeV and a Dark Photon of 100 GeV with a 2 nanosecond lifetime
  - Trained 309394 events
- Background: 3 million events of data 22

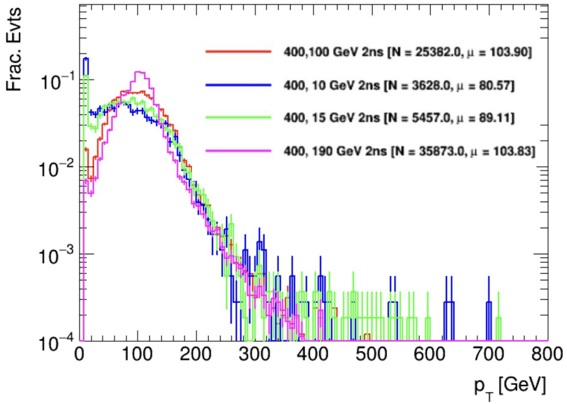


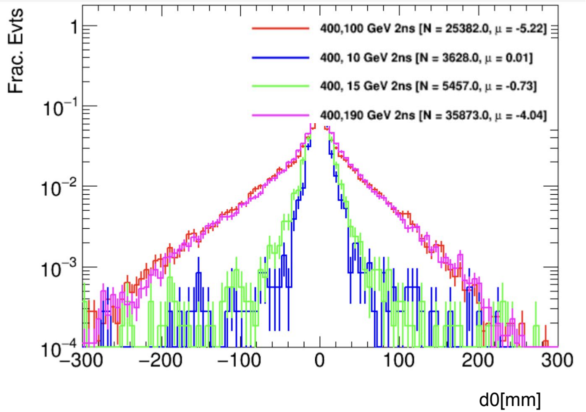
Signal Variable Distributions for Leading Electron

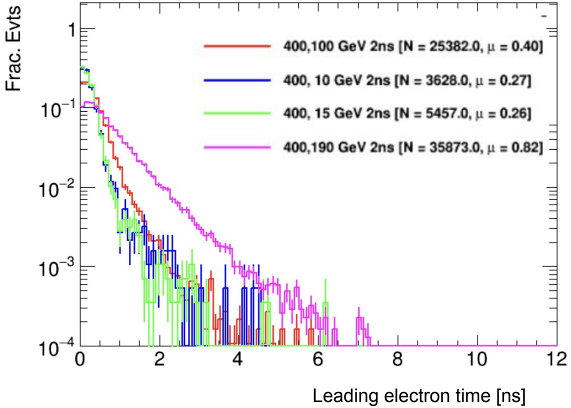
Three crucial properties of the leading electron in a particle collision

- Pt (transverse momentum) - measures the momentum of the electron perpendicular to the beam axis
- d0 (transverse impact parameter) - measures the distance of closest approach in the transverse plane of the track from the primary vertex (assumed collision point)
- Time - calibrated time of detection of the electron as measured by the LAr calorimeter



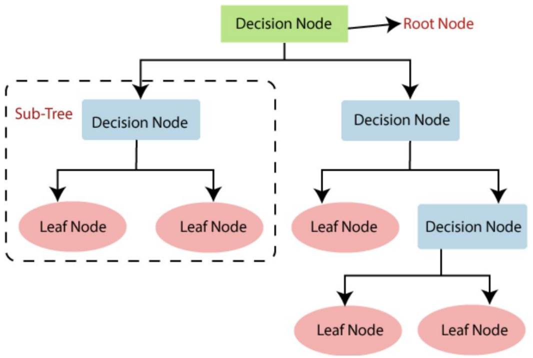






Boosted Decision Tree (BDT) Introduction

- Sets up a training process for a machine learning algorithm (TMVA) used to distinguish between signal (presence of Dark Photon) and background events
- decision tree in the ensemble makes a simple yes/no decision based on a cut of the data (e.g., electron_time < 0)



Boosting

- refers to how the individual decision trees are trained sequentially
- focuses on learning from the mistakes of the previous ones, leading to a more robust final model

How to use a BDT (TMVA)

1. Training - code defines cut to set features, BDT learns by splitting data based on features
2. Testing - assess BDT performance on designated subset of signal and background
3. Application - BDT used to analyze real detector data

BDT Preselection

BDT preselection is the process of applying a set of cuts to a dataset before training a Boosted Decision Tree (BDT). The goal of preselection is to remove events that are unlikely to be signal events and reduce dimensionality of data, which can improve the BDT's training speed and performance.

- **Final State Selection**
  - I am selecting only events with >= 2 electrons
- **Baseline Cuts**
  - Require |electron time of flight| < 12.5 ns
  - collisions occur every 25 ns
  - **Transverse momentum (pT)**
    - Selects high momentum electrons
  - **Electron identification (ID)**
    - Ensures good quality electrons and reduces background from jets

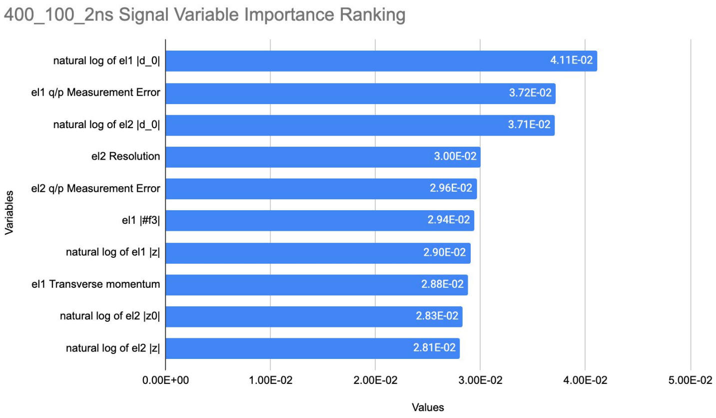
Variable	Requirement
pT	> 10 GeV
eta	< 2.47
ID	$<$ VeryLooseNoPix
Crack Veto	True
passOR	True
nTrackParticles	> 0
track z0	< 500

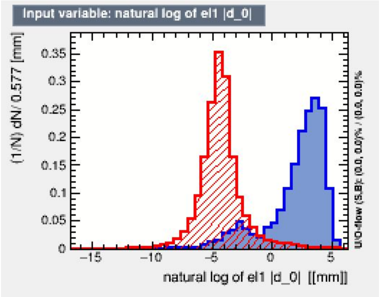
Table 1: *Electron Selection Criteria*

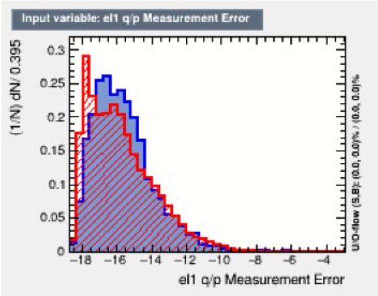
Input Variable Selection

BDT ranks variables by their variable importance, trained over variables related to the leading and subleading electron

- Started with around 80 variables, and 46 were in the final version
- Feature Importance - analyzing the feature importance scores, less important variables can then be removed
- Physical relevance - seeing physical difference of the signal and background process
- To optimize the BDT's performance in different signal regions, we trained 4 models

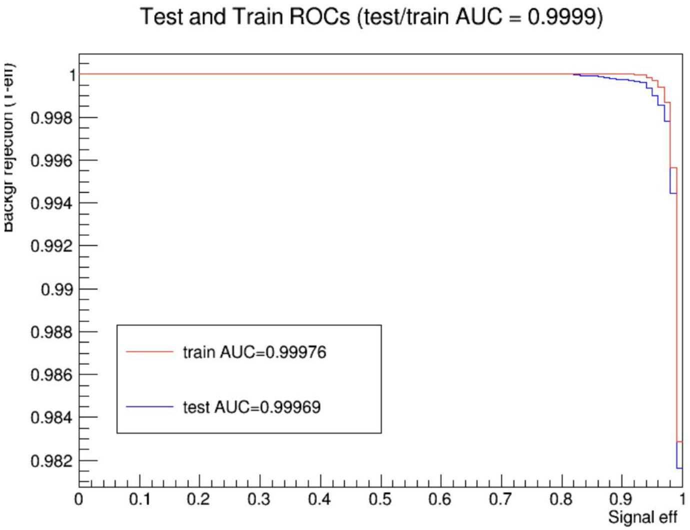


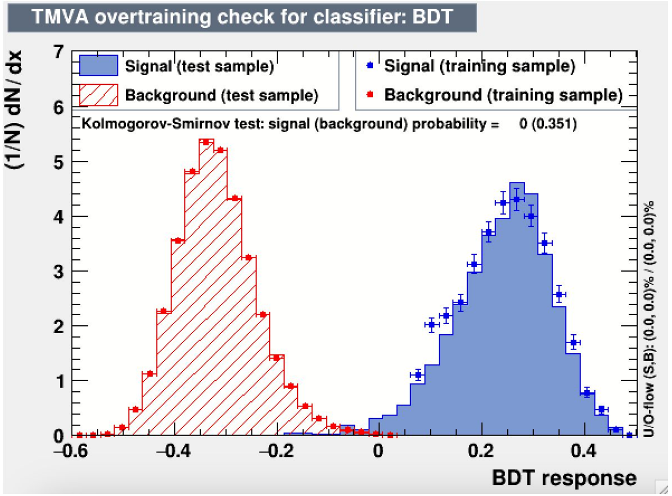




Validation Procedure

- Receiving operating characteristic (ROC) curves show the false positive vs true positive rate for a machine learning (ML) model
- Area Under the Curve (AUC) shows the percentage chance that the ML model successfully identifies signal from background
- doing a test for overtraining (machine learning model becomes too specialized in the training data and performs poorly)
  - Some concern on overtraining that may need further study



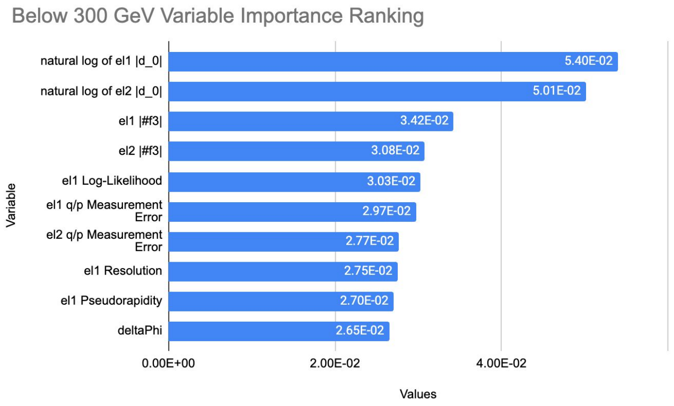


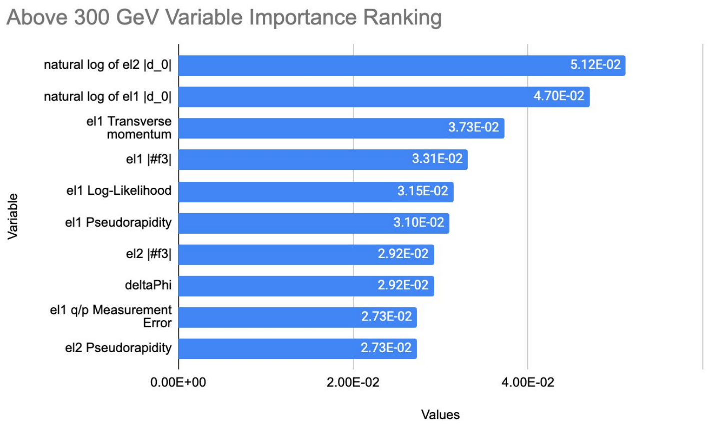
Outline

1. Introduction
    a. The Standard Model
    b. Instrumentation
        i. The LHC
        ii. The ATLAS Detector
2. Search for Displaced Electron from Long-Lived Dark Photon Decays
    a. Motivation
    b. Data and Simulated Samples
    c. Signal Variable Distributions
    d. BDT Introduction
3. Results
    a. BDT high mass vs. BDT low mass
    b. BDT with high delta M vs. BDT with low delta M
4. Conclusion and Next Steps

Low Mass Signals vs. High Mass Signals

- **Plot Comparison:**
  - Two plots display variable importance rankings: one for signals below 300 GeV, the other for signals above 300 GeV
  - Variable importance scores are represented by bar lengths
- **d0 Variable:**
  - "d0" variable (transverse impact parameter) is prominent in both plots
  - Significantly higher importance in the "Below 300 GeV" plot
  - Lower importance in the "Above 300 GeV" plot



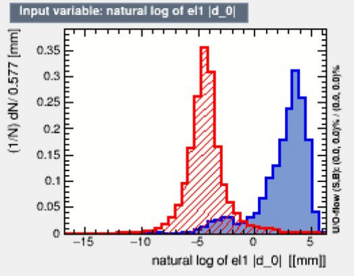


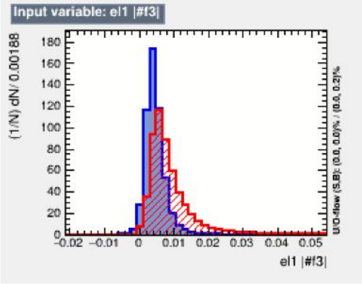
Low Mass Signals vs. High Mass Signals

Interpretation:

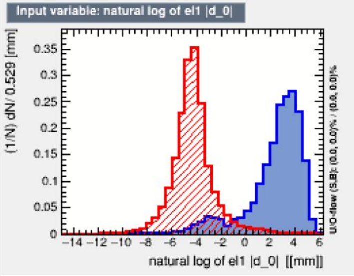
- Transverse impact parameter (d0) is a strong discriminator for low mass signals
- Its importance decreases in higher energy regions (high mass signals)
- f3 - fraction of the electron energy deposited in the third layer of the calorimeter
- pt - transverse momentum of electron (Transverse = in plane orthogonal to direction of the proton-proton beams)

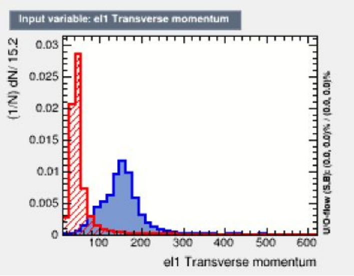
Low Mass Signals





High Mass Signals



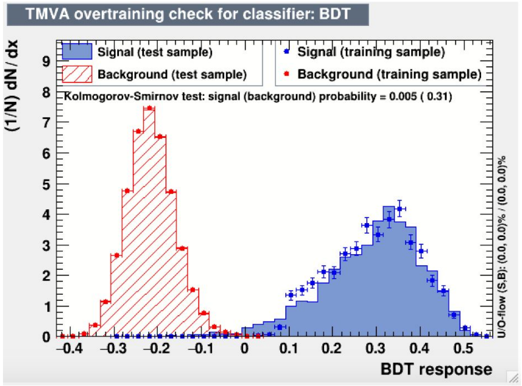


Low Mass Signals vs. High Mass Signals

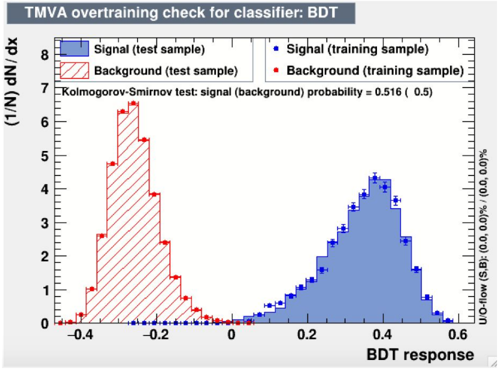
Observations:

- Background concentrates towards lower BDT response values
- TMVA classifier effective in separating signal and background below 300 GeV
- Challenges in discrimination above 300 GeV due to overlap

Low Mass Signal



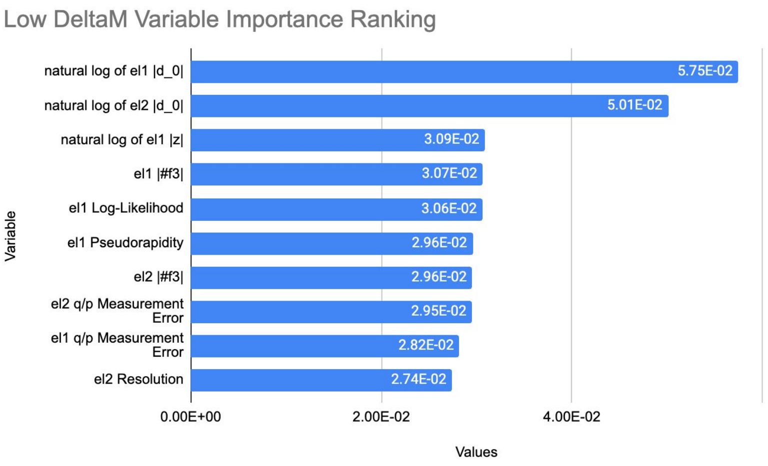
High Mass Signal

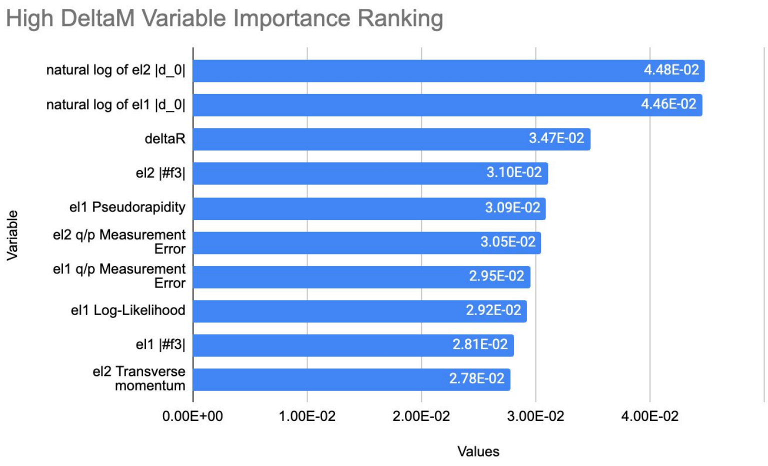


Low DeltaM vs. High DeltaM Signals

Observations

- Different variables are important for low and high DeltaM signals
- For low DeltaM, variables related to transverse impact parameter and "pointing" of electron are prominent
- For high DeltaM, variables associated with transverse impact parameter and deltaR take precedence



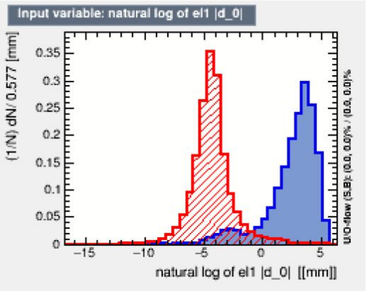


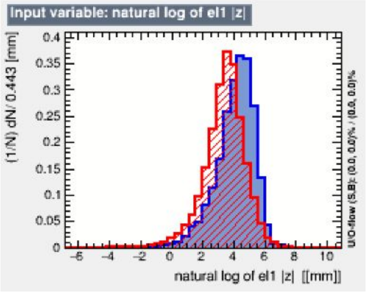
Low DeltaM vs. High DeltaM Signals

Interpretation:

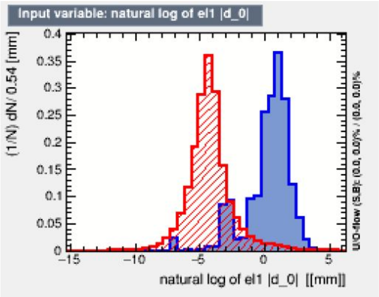
- Distributions of the top variables exhibit distinct shapes between the two DeltaM categories
- these variables can effectively differentiate between low and high DeltaM signals
- z - "pointing" of electron as measured using calorimeter deposits only

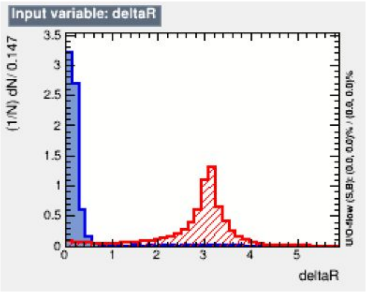
Low DeltaM Signal





High DeltaM Signal



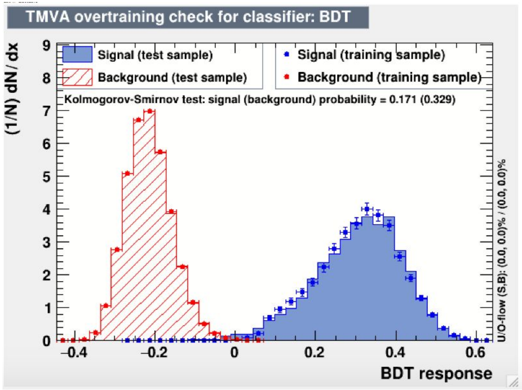


Low DeltaM vs. High DeltaM Signals

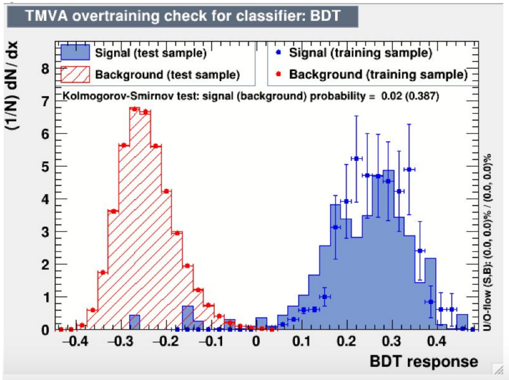
Interpretation

- BDT effectively discriminates signal from background in Low DeltaM region
- BDT performance less effective in High DeltaM region
- Both plots show good separation between signal and background
- Kolmogorov-Smirnov test results indicate significant differences between signal and background distributions for both low and high DeltaM

Low DeltaM Signal



High DeltaM Signal



Outline

1. Introduction
    a. The Standard Model
    b. Instrumentation
        i. The LHC
        ii. The ATLAS Detector
2. Search for Displaced Electron from Long-Lived Dark Photon Decays
    a. Motivation
    b. Data and Simulated Samples
    c. Signal Variable Distributions
    d. BDT Introduction
3. Results
    a. BDT high mass vs. BDT low mass
    b. BDT with high delta M vs. BDT with low delta M
4. Conclusion and Next Steps

Conclusions and Next Steps

Conclusion: By ranking electron variables by importance, we find that the high deltaM signal models look different compared to the other models. Thus, the high deltaM model may need to be treated separately. I have also narrowed down the list of input variables for the BDT to consider. From the ranking, electron d0 always ranks the highest for all signal models.

Next Steps:

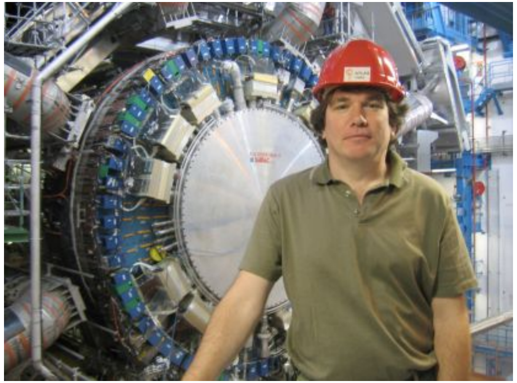
- Compute expected signal and background ratios from the BDT
- Improve training statistics for the high deltaM signal
- Decide whether we can have a common BDT for all signals or specialized one for high deltaM signal

References

1. Detector & Technology (atlas.cern)
2. The Large Hadron Collider | CERN (home.cern)
3. The Standard Model | CERN (home.cern)
4. Search for Higgs-like scalar decaying into new spin-1 bosons in four-lepton final states at the ATLAS detector with 139 fb^{-1} at $\sqrt{s}=13$ TeV - CERN Document Server
5. Search for a new scalar decays to beyond-the-Standard-Model light bosons in four-lepton final states with 140 fb^{-1} at $\sqrt{s}=13$ TeV - CERN Document Server
6. DOE Explains...the Standard Model of Particle Physics | Department of Energy
7. 1412.0018 (arxiv.org)
8. ATLAS detector slice (and particle visualisations) - CERN Document Server
9. TMVA - Users Guide (cern.ch)
10. README.md · master · Eleanor Luise Woodward / Dark Photons BDT · GitLab (cern.ch)

Acknowledgements

Faculty Mentor: John Parsons



Research Staff:

Jonathan Long

Graduate Students:

Eleanor Woodward and Maria Bressan

- ATLAS Collaboration, especially the Columbia ATLAS Group
- My fellow REU participants who I got to know in the past 10 weeks
- Entire team at Nevis Laboratories, particularly John, Georgia, Reshmi, and Amy
- The National Science Foundation (NSF) for funding this research under Grant No. PHY/2349438

Questions?

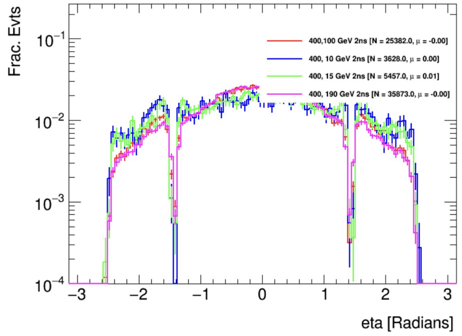
BACKUP

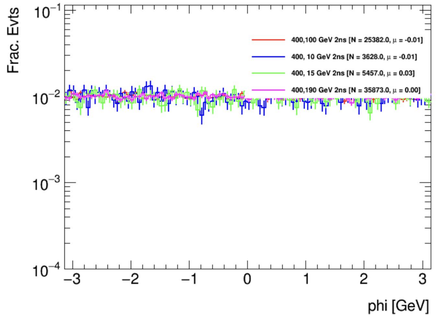
NEVIS Website

[2 Electron Final State BDT Training (columbia.edu)]()

BDT Inputs

- f3 - fraction of the electron energy deposited in the third layer of the calorimeter
- pt - transverse momentum of electron (Transverse = in plane orthogonal to direction of the proton-proton beams)
- phi - phi coordinate (azimuthal angle) of the electron trajectory
- eta - pseudorapidity coordinate of electron
- dz - resolution of calorimeter pointing measurement
- ntracks - number of tracks associated to this electron
- LHValue - continuous version of electron_ID variable, corresponding to the log-likelihood of electron corresponding to a "true" electron
- dpt - percent difference between pt of electron track and electron object
- d0 - transverse impact parameter of the electron track
- z - "pointing" of electron as measured using calorimeter deposits only
- z0 - longitudinal impact parameter of electron track
- qop_err - error of electron charge over momentum measurement
- chi2 - goodness of fit parameter for electron track from fit to the hits that form the track
- nPIX - number of pixel layers crossed by the electron track
- time - calibrated time of production of the electron as measured by the LAr calorimeter
- charge - electron electric charge (+/- 1)
- nSCT - number of layers crossed by the electron track in the Semiconductor Tracker
- topcone20 - measurement of electron isolation in calorimeter: sum of transverse energy of topological calorimeter clusters within a cone of radius dR < 0.2 of electron calorimeter cluster
- E - calibrated electron energy measurement





BDT Code

v1.0 of ntuples for full signal grid/data22 currently available on lxplus and xenia

How BDT is Used in the Code:

1. Define electron variables and signal sample (e.g., all_above_300ns)
2. Run root -q -b -l myBDT.C
3. Run root -l -e 'TMVA::TMVAGui("output_full.root")'

Files · master · Amelia Mei Stevens / Dark Photons BDT · GitLab (cern.ch)