U.S. DEPARTMENT of ENERGY

Training BDTs for ALP Signal and Coherent Rho Production

Alyssa Wheeler, Alessandro Tricoli, Jae Nam, and Ganesh Parida

August 5th 2026

@BrookhavenLab

Introduction and BDT Plan

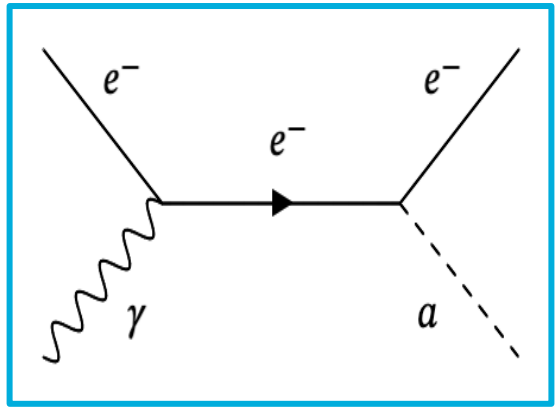
Introduction

- ALP Signal Production Campaign is complete

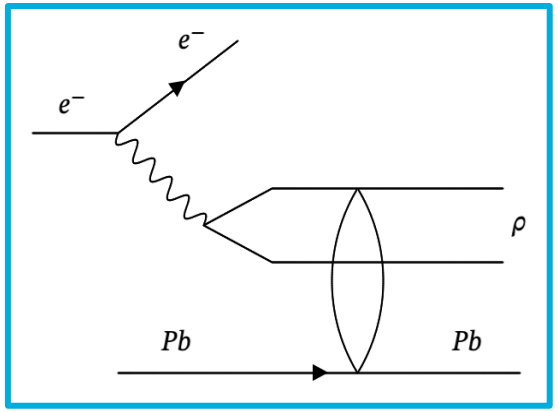
Backgrounds:

- **Coherent Rho Production: Official June production requested by Maci Kesler, eAu at 10x100 GeV2, $1 < Q^2$ (GeV^2) < 20 production**
- Inclusive DIS Production: eAu at 10x100 for $1 < Q^2 < 10$ GeV^2 and $10 < Q^2 < 100$ GeV^2 produced by Maci Kesler
- Recently finished producing Coherent Photon Production samples

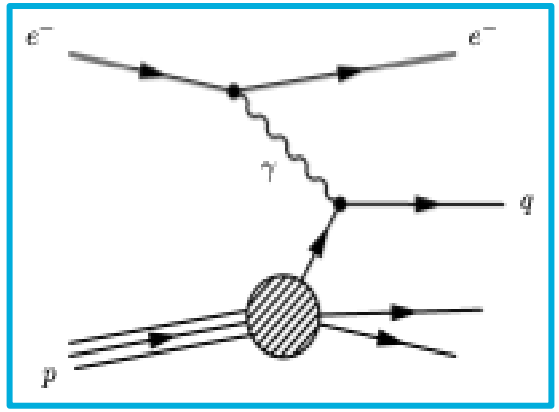
- This update focuses only on Coherent Rho Production



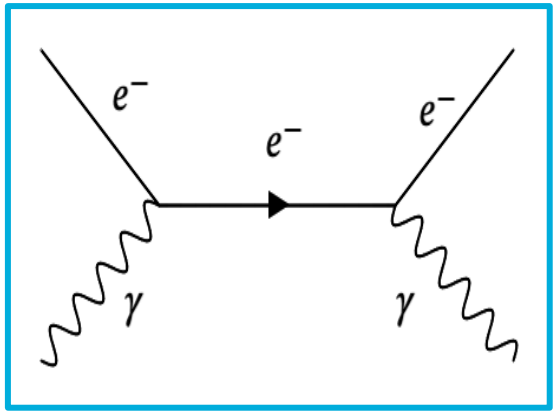
ALP Production



Coherent Rho Production



Inclusive DIS



Coherent Photon Production

BDT Training Plan:

1. Train BDTs for m_{ALP} = 0.1, 1.0 GeV signal and each background
    1. DIS and Coherent Rho in progress
2. Narrow down set of input variables and Adjust BDT Parameters
    1. Analyze Input variables
    2. Input Variable Importance
    3. BDT Response distributions
    4. Overtraining Checks
    5. ROC Curves and AUC

BDT Training Procedure:

Coherent Rho BDT Progress

Coherent Rho: Input Variable Distributions

Quality Cuts

	Cut:	Value:
1	Electron momentum	$p \geq 0.5$ GeV
2	Pseudorapidity	$\lvert\eta\rvert < 4$
3	Central Detector	Rejects tracks in the low-Q^2 tagger
4	Tracker hits	Nhits ≥ 5
5	Vertex cuts	$\lvert Z_{vtx}\rvert < 100\ mm$ $\lvert XY_{vtx}\rvert < 3\ mm$
6	Ecal Matching	$N_{Ecal,cls} \geq 1$
7	eID	Reject if there's no electron in the Central Detector

Using ROOT TMVA BDT Method

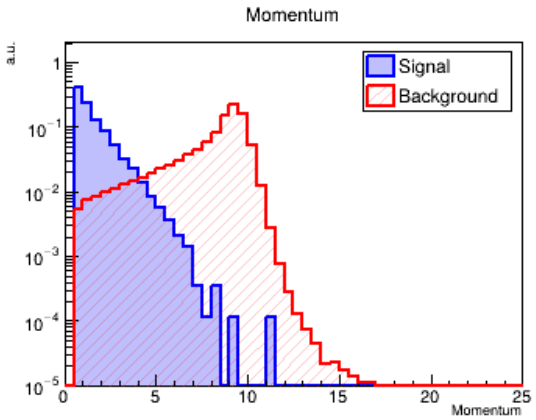
Model parameters:

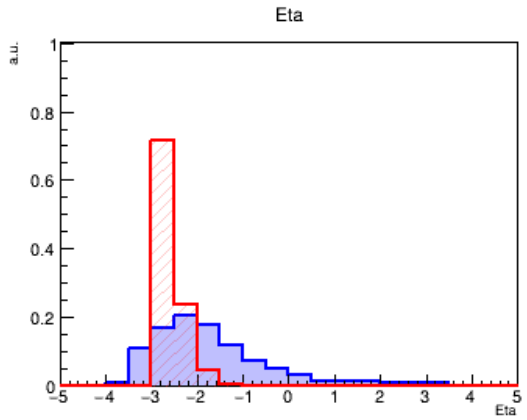
- Number of trees = 300
- Max Depth = 4
- Number of Cuts = 40

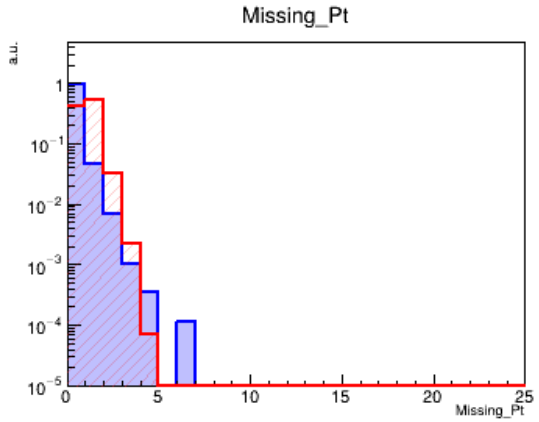
Input Variables

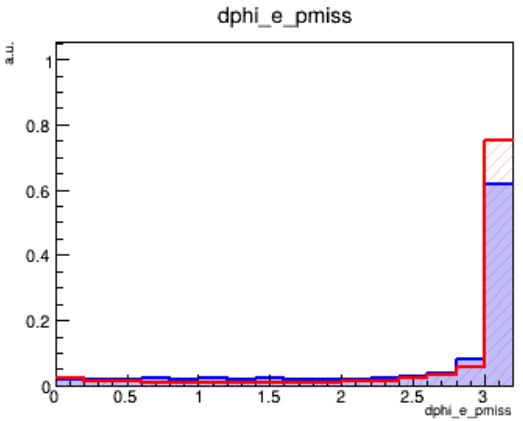
Variable Name	Definition
Momentum	Scattered electron momentum
Eta	Scattered electron η
Missing_Pt	$\left[-\left(\sum_{i=0}^{N_{tracks}} \vec{p_i} + \sum_{i=0}^{N_{unmatched\ clusters}} \vec{p_i}\right)\right]_T$
dphi_e_pmiss	Change in ϕ between $\vec{p}_{ele}$ and $\vec{p}_{miss}$
Multiplicity	Number of reconstructed particles
Empz	$\sum_{i=0}^{N_{tracks}} \lvert\vec{p_i}\rvert - pi_{,z} + \sum_{i=0}^{N_{unmatched\ clusters}} E_i - Ei\cos\theta$
pTET	$\frac{\vec{p}_{miss}}{\sum_i^{N_{tracks}} p_{i,T} + \sum_i^{N_{unmatched\ clusters}} p_{i,T}}$

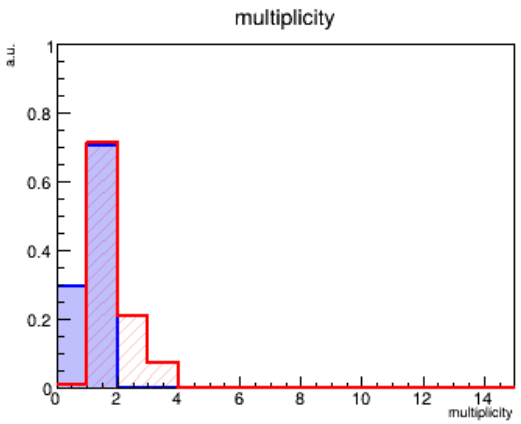
Coherent Rho: Input Variable Distributions

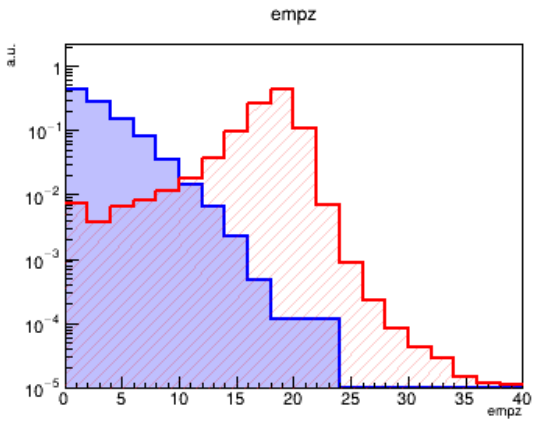


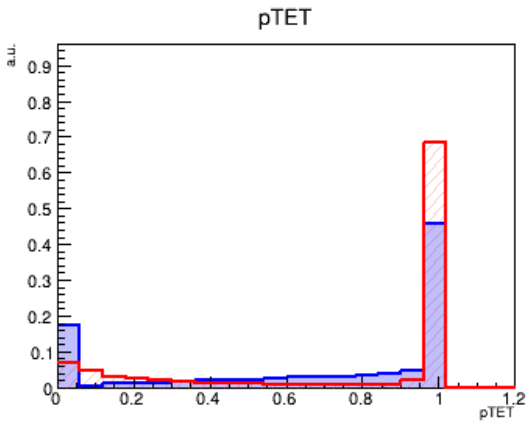










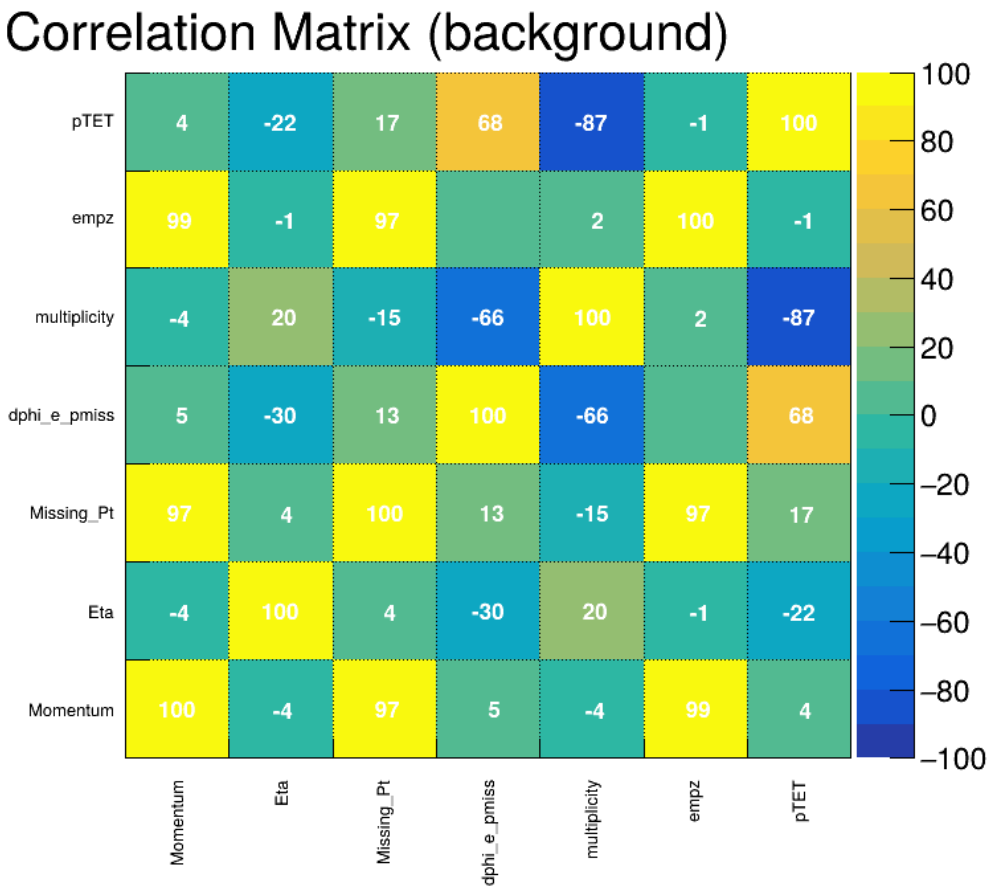


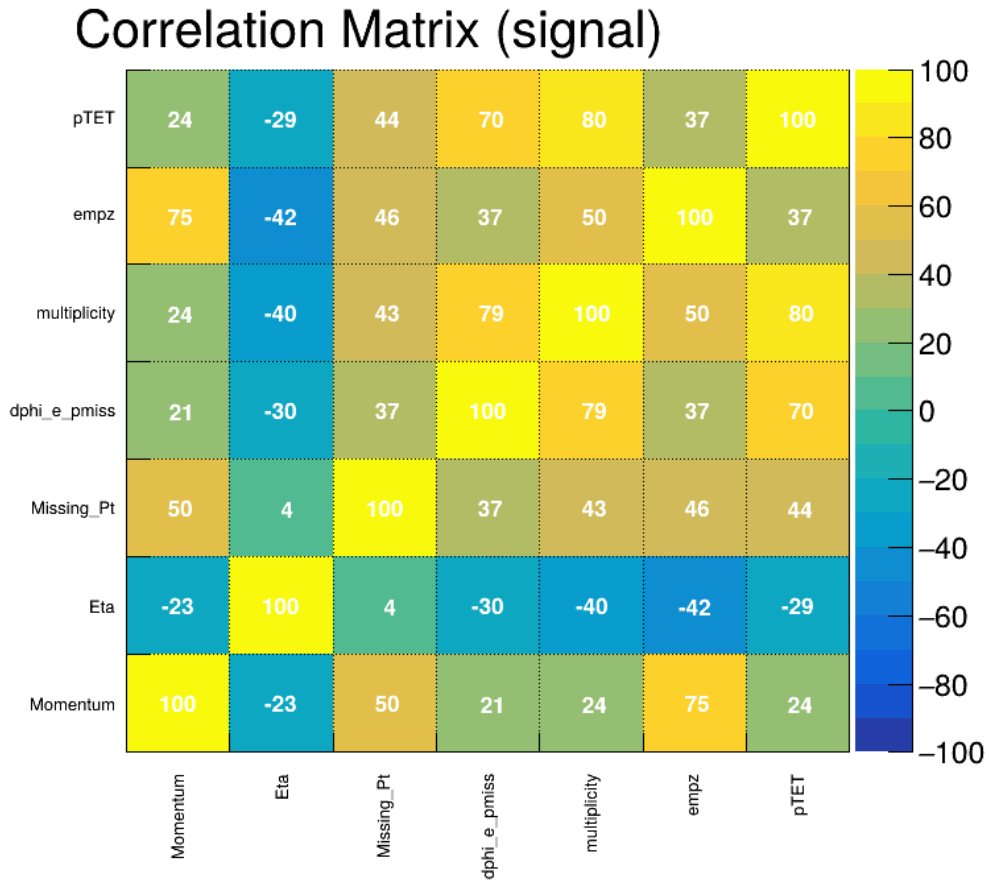
m_{ALP} = 1.0 GeV, $beam_e$ = 10 GeV

Coherent Rho: Variable Importance

Number	Variable	Variable Importance
1	Missing_Pt	1.931e-01
2	Momentum	1.724e-01
3	Eta	1.547e-01
4	pTET	1.424e-01
5	dphi_e_pmiss	1.177e-01
6	empz	1.173e-01
7	multiplicity	1.024e-01

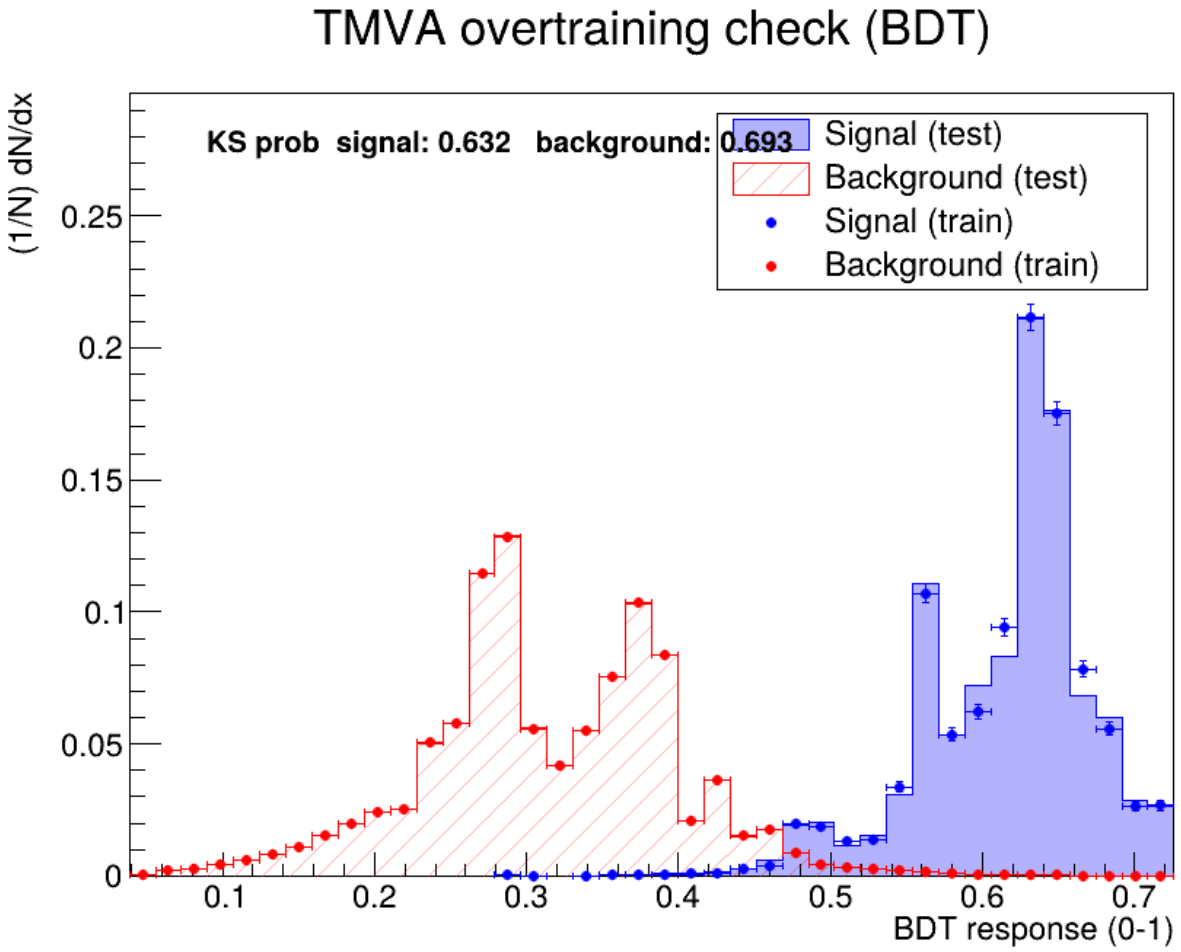
Coherent Rho Background and Signal Correlations





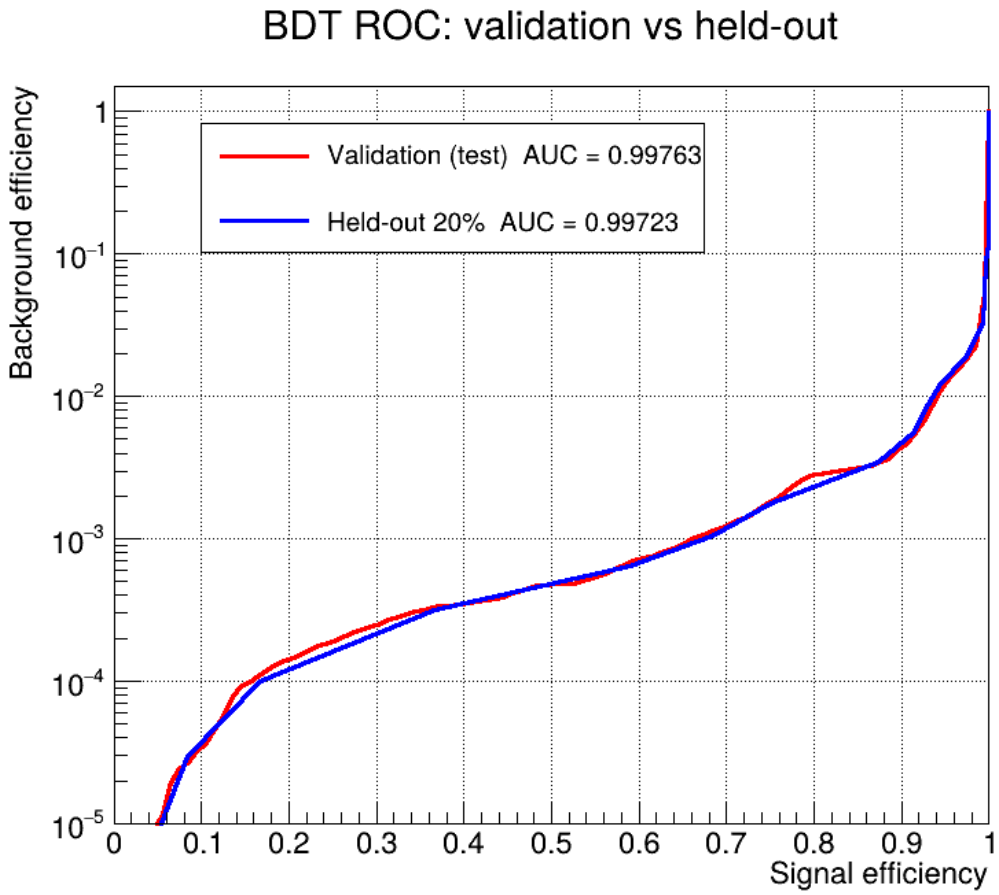
m_{ALP} = 1.0 GeV, $beam_e$ = 10 GeV

Coherent Rho: Overtraining test



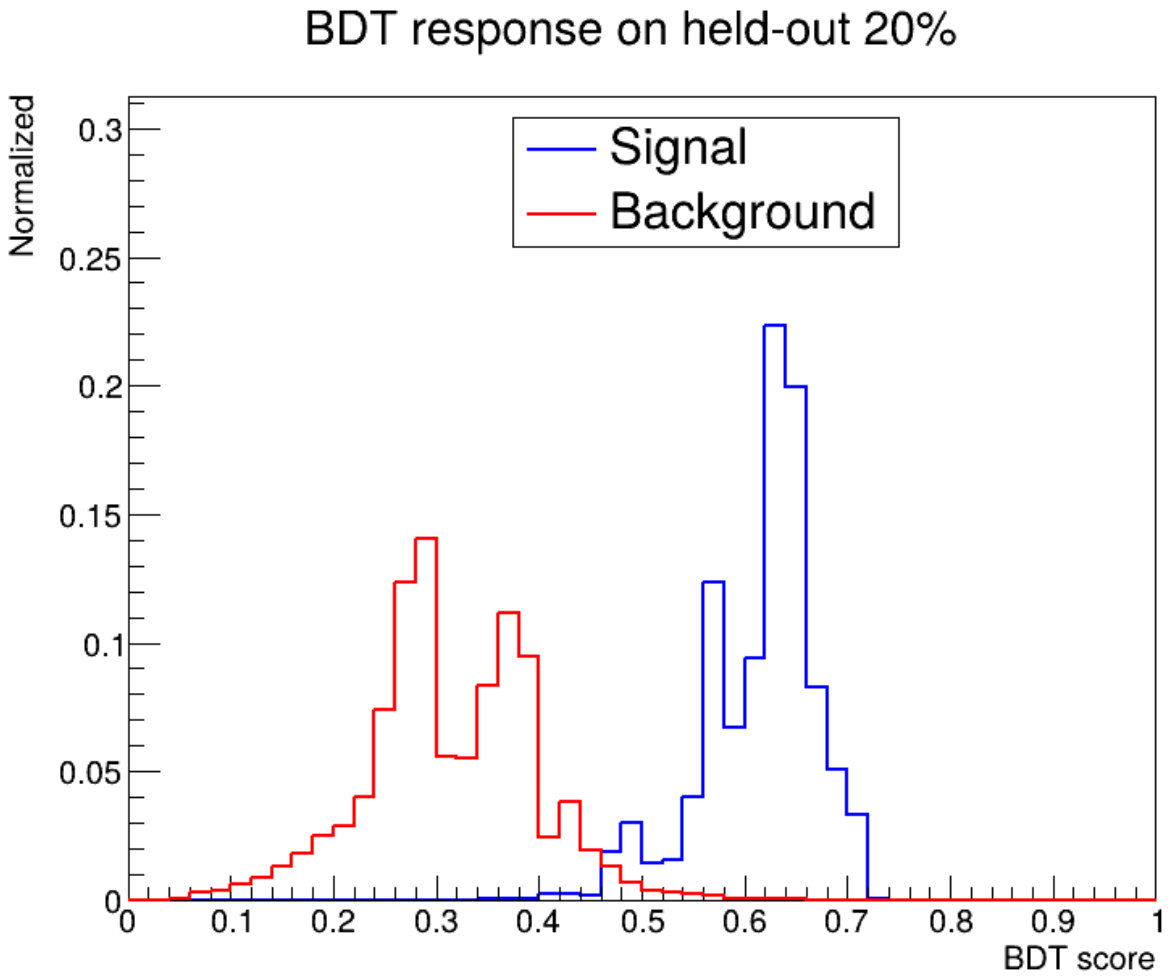
m_{ALP} = 1.0 GeV, $beam_e$ = 10 GeV

Coherent Rho: ROC Curve



m_{ALP} = 1.0 GeV, $beam_e$ = 10 GeV

Coherent Rho: BDT Training Results



Overall good discrimination

Persistent double peak in the background BDT response

Plan: try eliminating variables one by one in BDT training to see if the double peak remains

m_{ALP} = 1.0 GeV, $beam_e$ = 10 GeV

Summary

- Overall, the BDT has good discrimination between ALP signals and coherent rho backgrounds.
- More adjustments need to be made to settle on a final set of variables and quality cuts

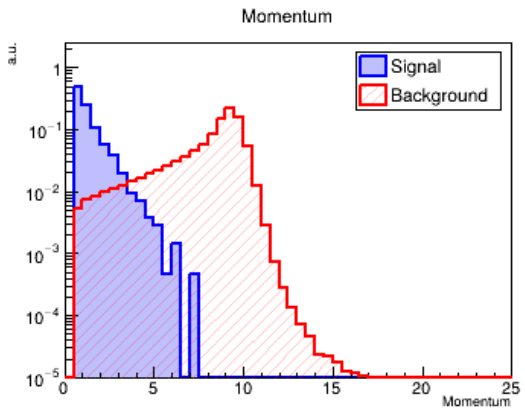
Outlook

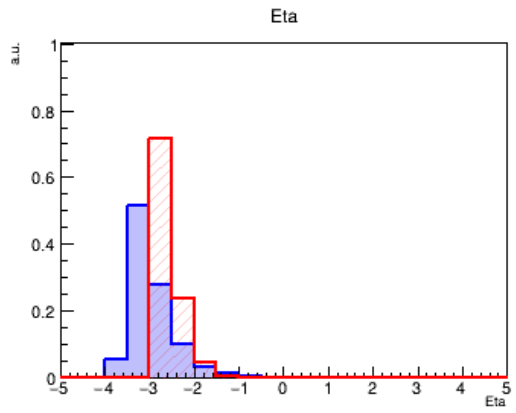
- Train BDTs for coherent photon production samples
  - narrow down variable and cut selections
  - Analyze input variable distributions, variable importance, BDT responses, overtraining checks, etc.
- Analyze signal efficiency and selection acceptance as a function of ALP mass to optimize

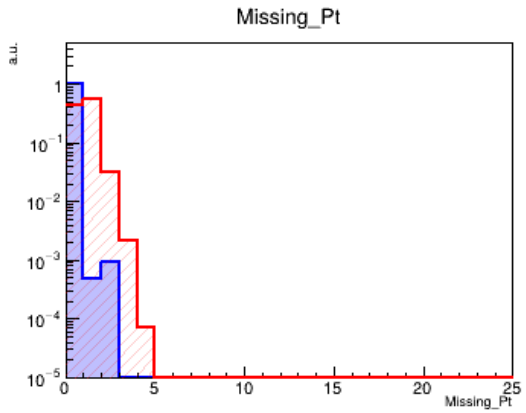
Thank you!

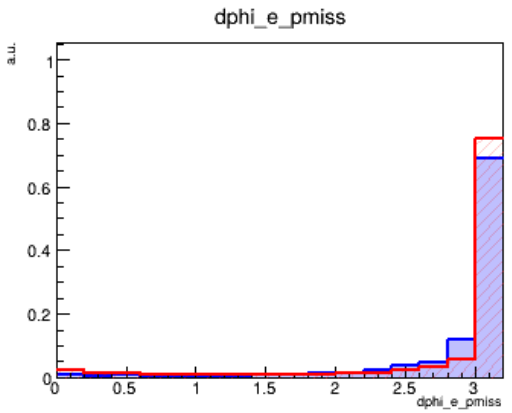
Backup Slides

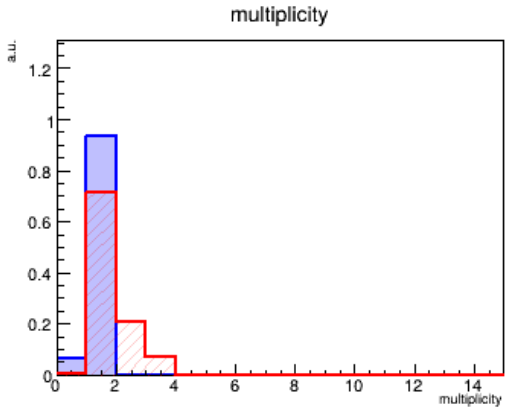
Coherent Rho: Input Variable Distributions

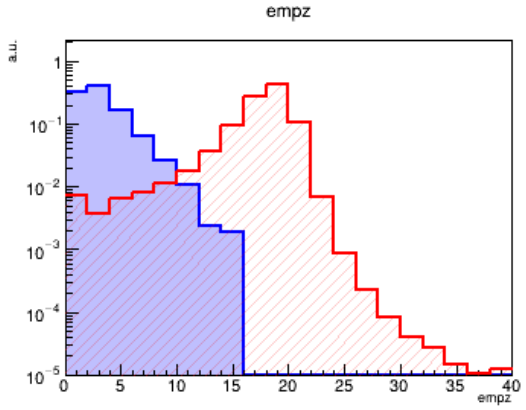


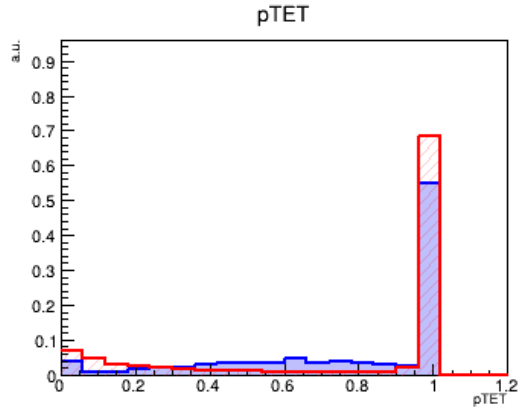










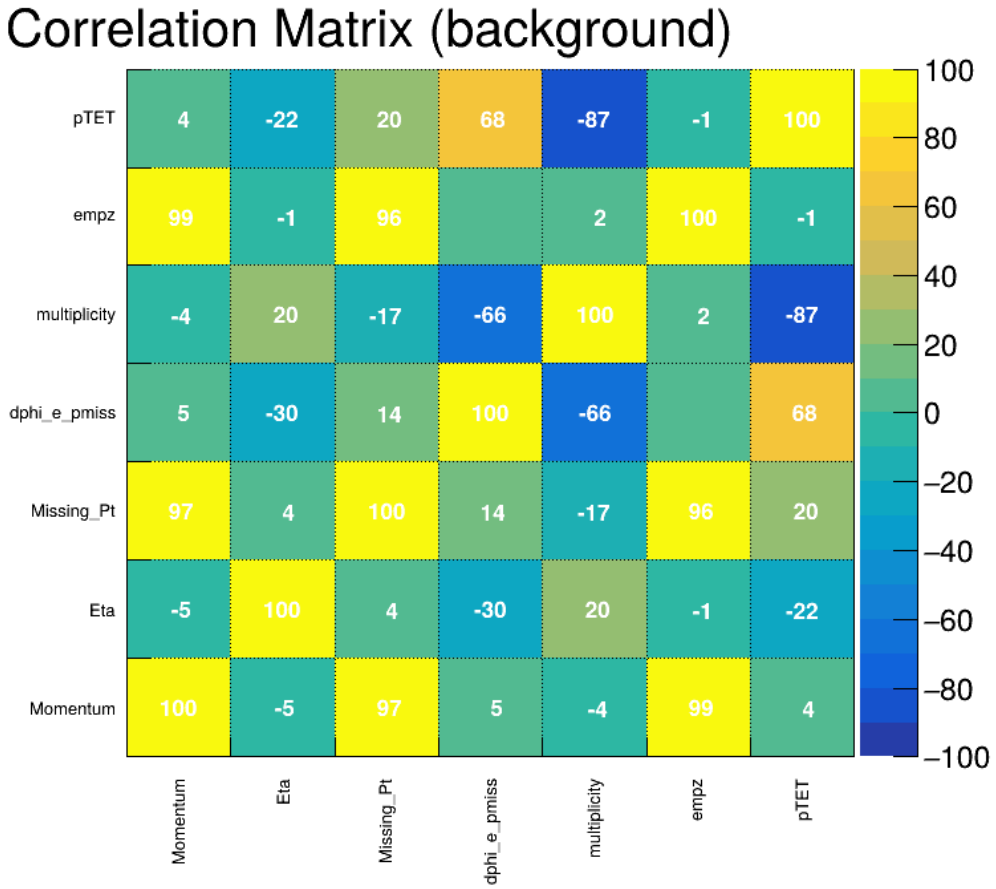


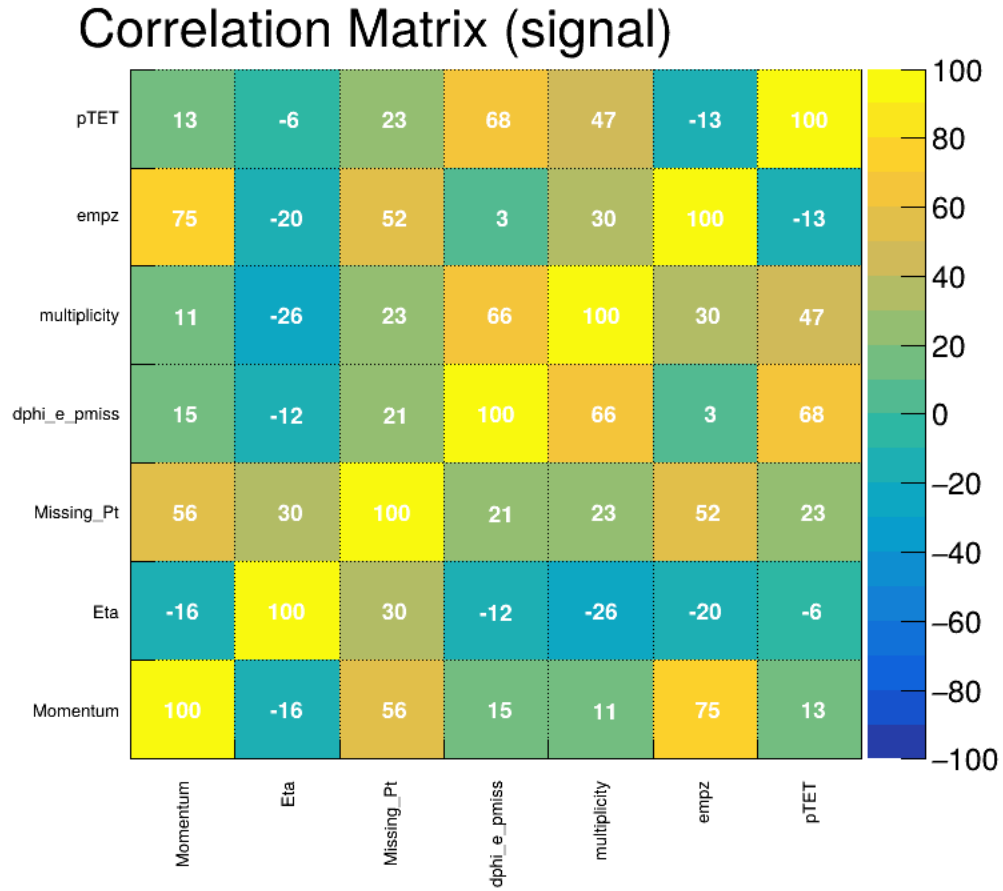
m_{ALP} = 0.1 GeV, $beam_e$ = 10 GeV

Coherent Rho: Variable Importance

```
: ----------------------------------------
                    : Rank : Variable     : Variable Importance
                    : ----------------------------------------
                    :    1 : Eta          : 1.842e-01
                    :    2 : Missing_Pt   : 1.736e-01
                    :    3 : dphi_e_pmiss : 1.565e-01
                    :    4 : pTET         : 1.382e-01
                    :    5 : Momentum     : 1.352e-01
                    :    6 : empz         : 1.214e-01
                    :    7 : multiplicity : 9.085e-02
                    : ----------------------------------------
```

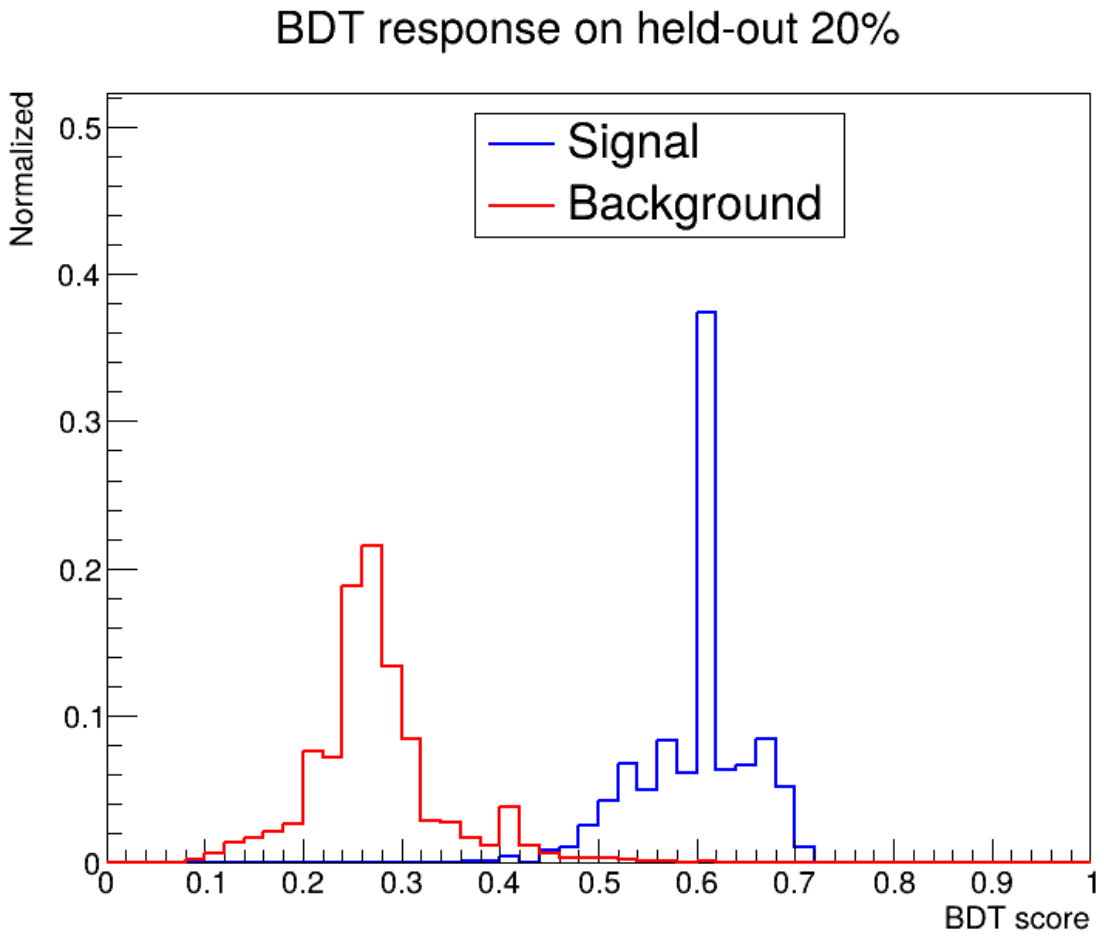
Coherent Rho: Variable Correlations





m_{ALP} = 0.1 GeV, $beam_e$ = 10 GeV

Coherent Rho: BDT Training Results

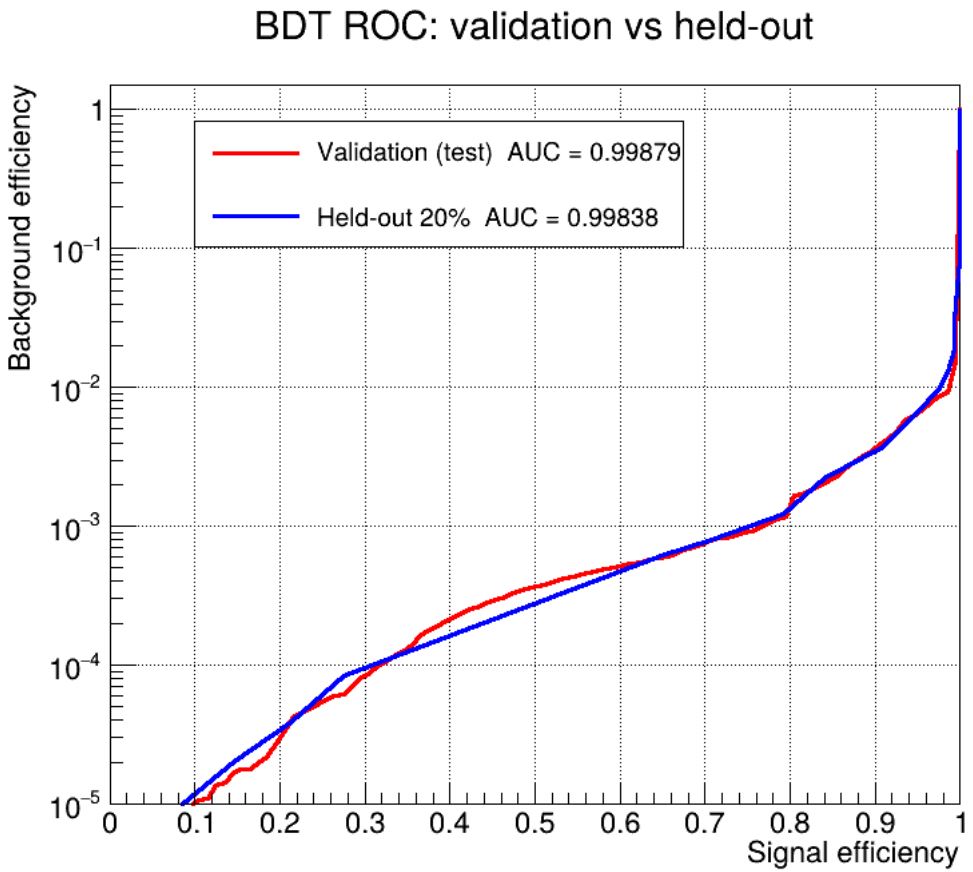


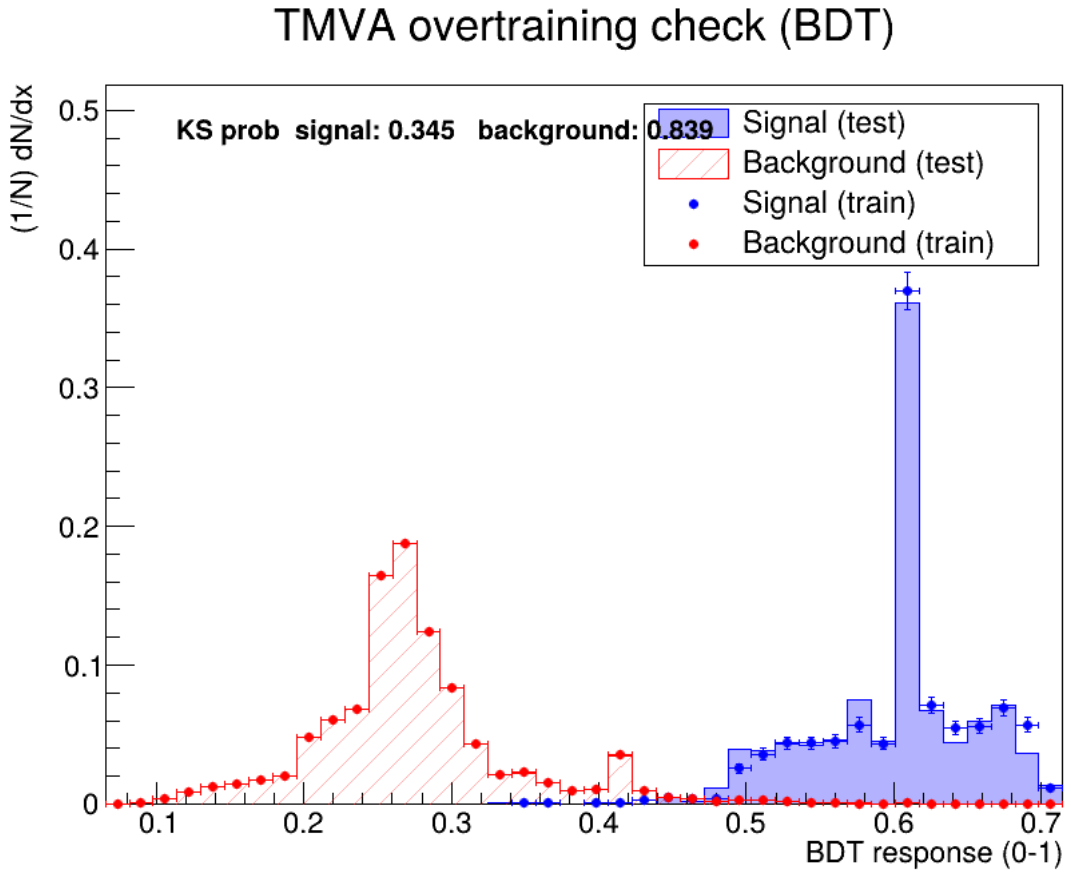
Overall good discrimination between the ALP signal and coherent ρ background

Single peak in BDT response to coherent rho production

m_{ALP} = 0.1 GeV, $beam_e$ = 10 GeV

Coherent Rho: ROC and Overtraining test





m_{ALP} = 0.1 GeV, $beam_e$ = 10 GeV

Backup slides: Double Peak in coherent Rho background at higher m_{ALP}