

MAR 12TH 2026

ML FOR NUCLEON ID IN THE CENTRAL DETECTOR

GRACE RAY

CONTEXT

Protons: Identified as hits in the CTOF with associated CVT tracks

Neutrons: Identified as hits in the CTOF and CND with no associated CVT tracks

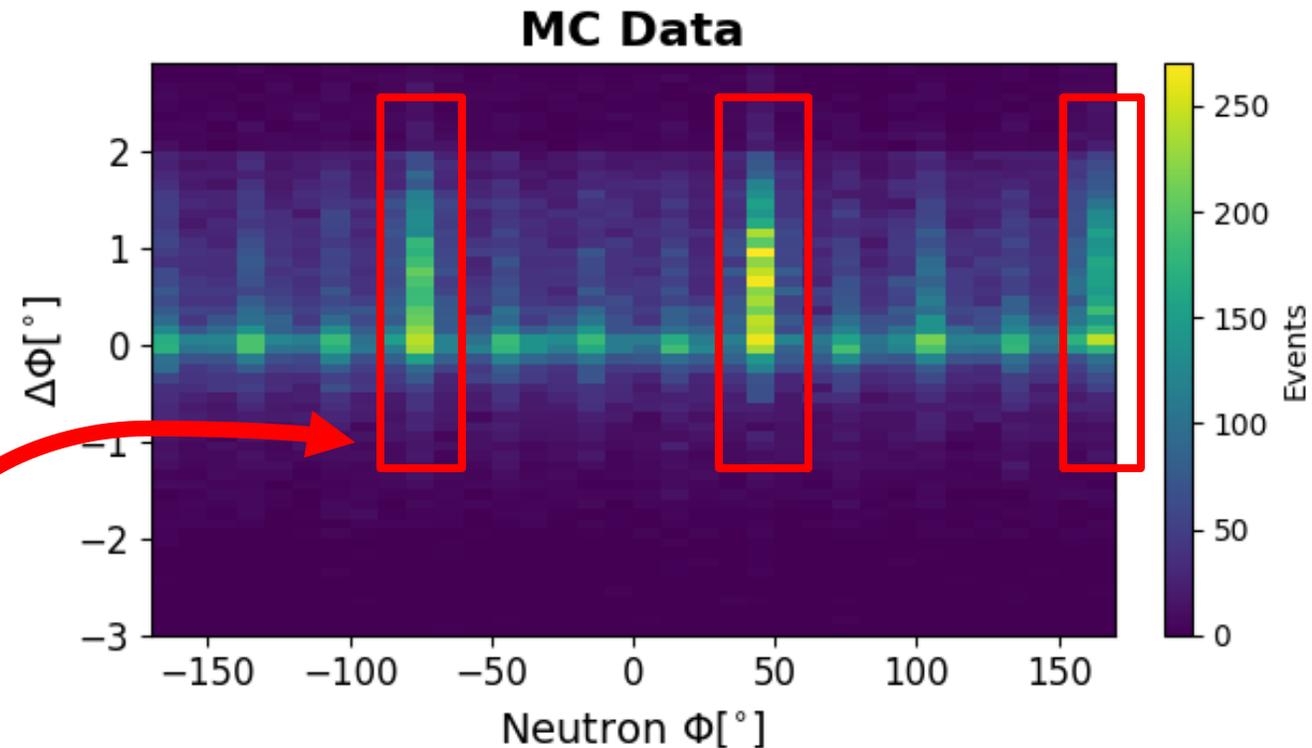
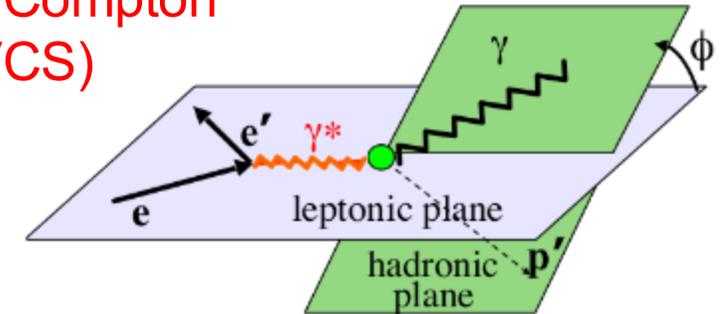
In low efficiency zones in CVT:

- proton tracks are missed
- protons are misidentified as neutrons and assigned a straight tracks

OBJECTIVE: use **detector-level information** to better classify neutrons (signal) and protons (background)!

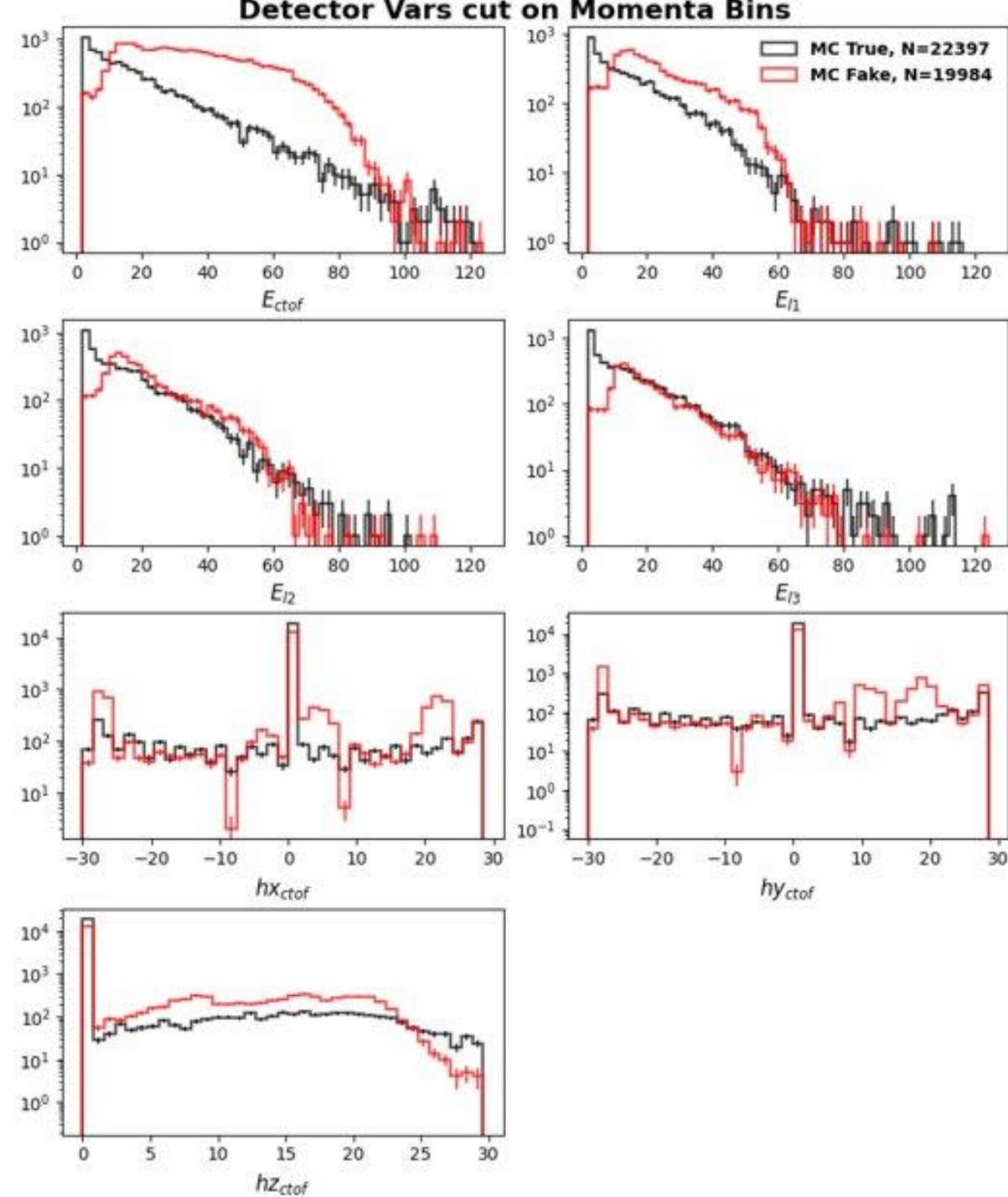
Regions where momentum is not conserved in $en \rightarrow en\gamma$ (low eff zones)

Deeply Virtual Compton Scattering (DVCS)
 $ep \rightarrow e'p'\gamma$



DATASET & TRAINING VARIABLES

- **MC – ND3 target** (any neutrons simulated from both pdvcs and ndvcs events)
 - Used for training
 - **True**: neutrons from generated neutron (**en** -> **eny**)
 - **Fake**: neutrons from generated proton (**ep** -> **eny**)
- **Training Vars:**
 - Energy deposition variables
 - Cluster distance from straight track



MODELS

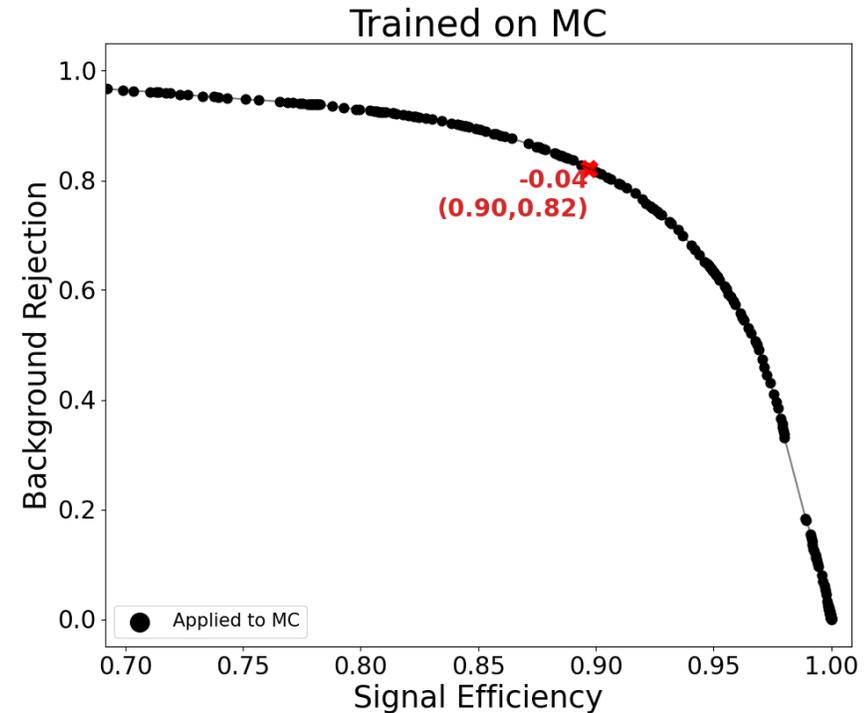
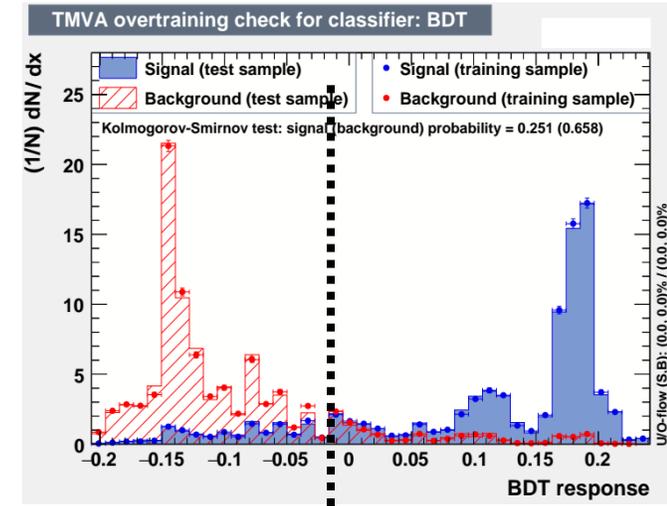
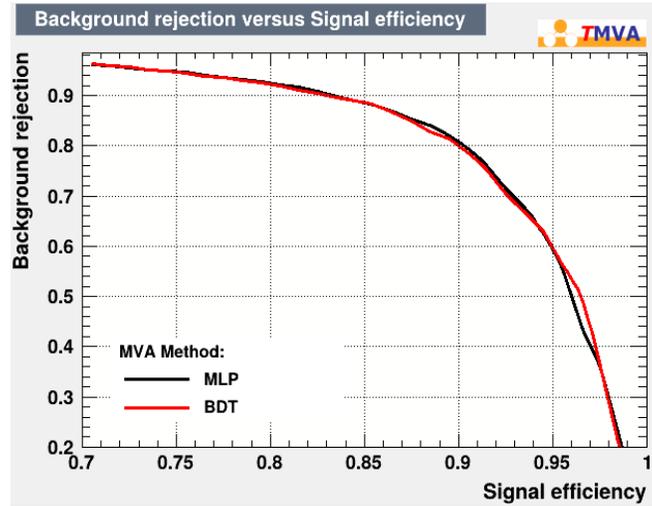
Software: ROOT TMVA

- Multilayer perceptron (MLP) and adaptive BDT were trained
- **MLP** → AUC = **0.927**
- **BDT** → AUC = **0.926**
- The models perform similarly! **BDT was chosen because it was 20x faster than MLP on my desktop and has a simpler structure to visualize.**

NEXT: a cut was chosen on BDT response

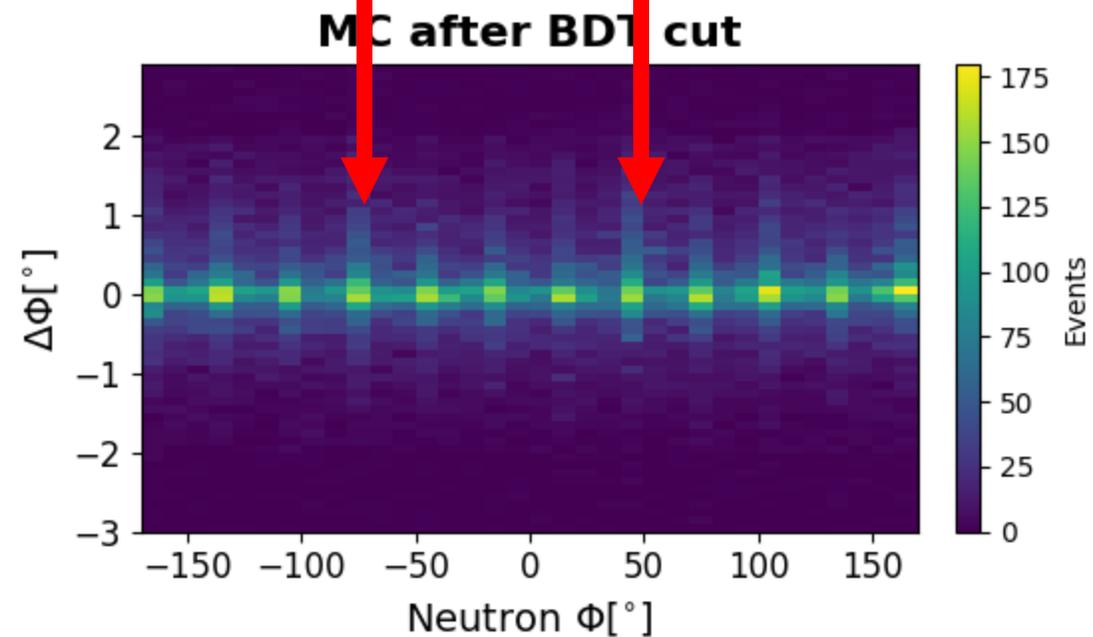
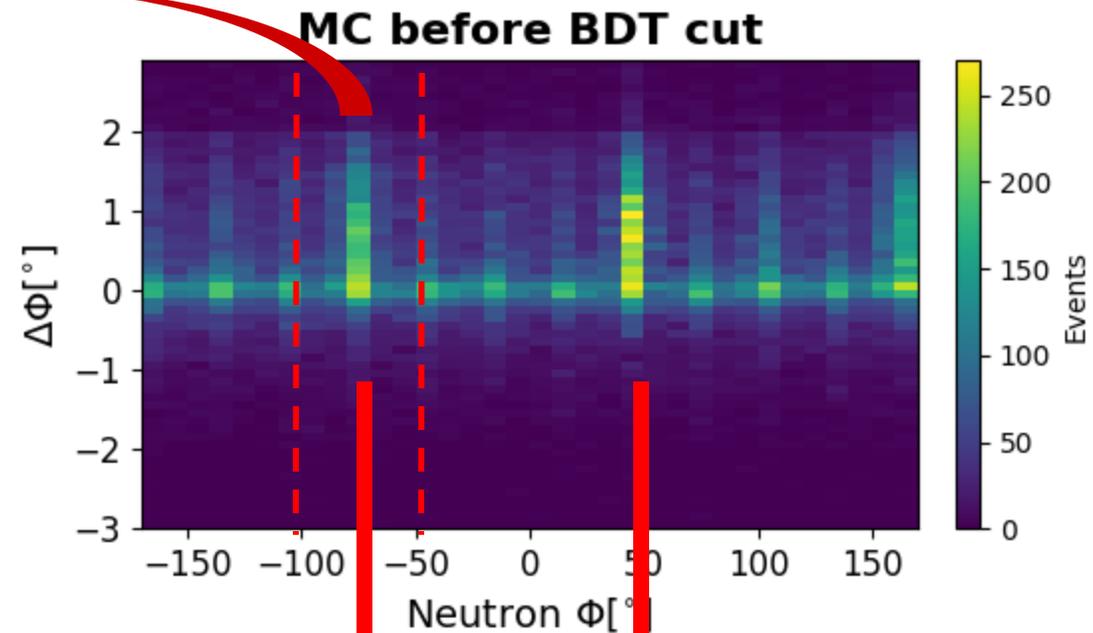
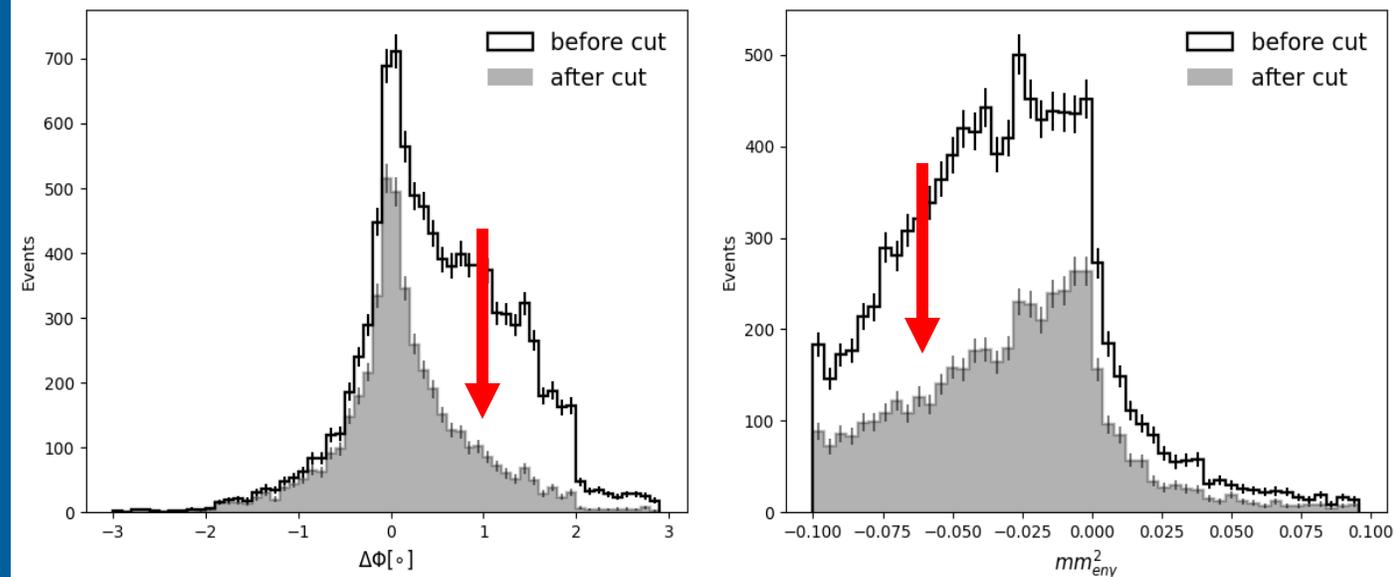
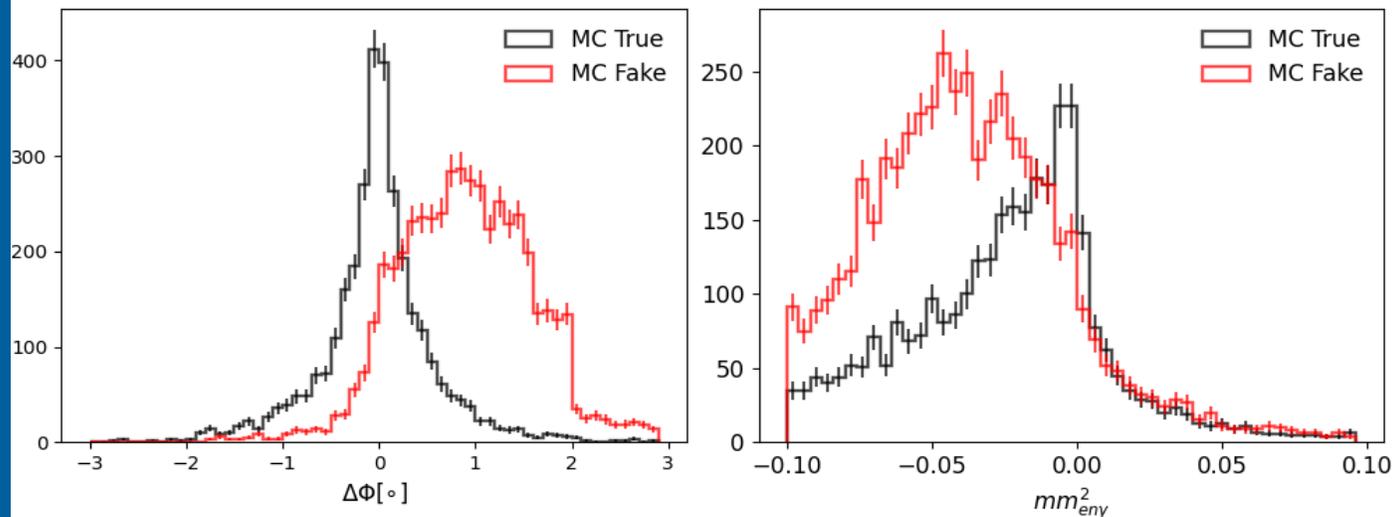
What is the cut criteria? In the context of **DVCS** studies which have low statistics and methods of removing remaining background, **high signal efficiency is preferred.**

Selected Cut: **BDT Response = -0.04**
signal efficiency = 90%
background rejection = 82%

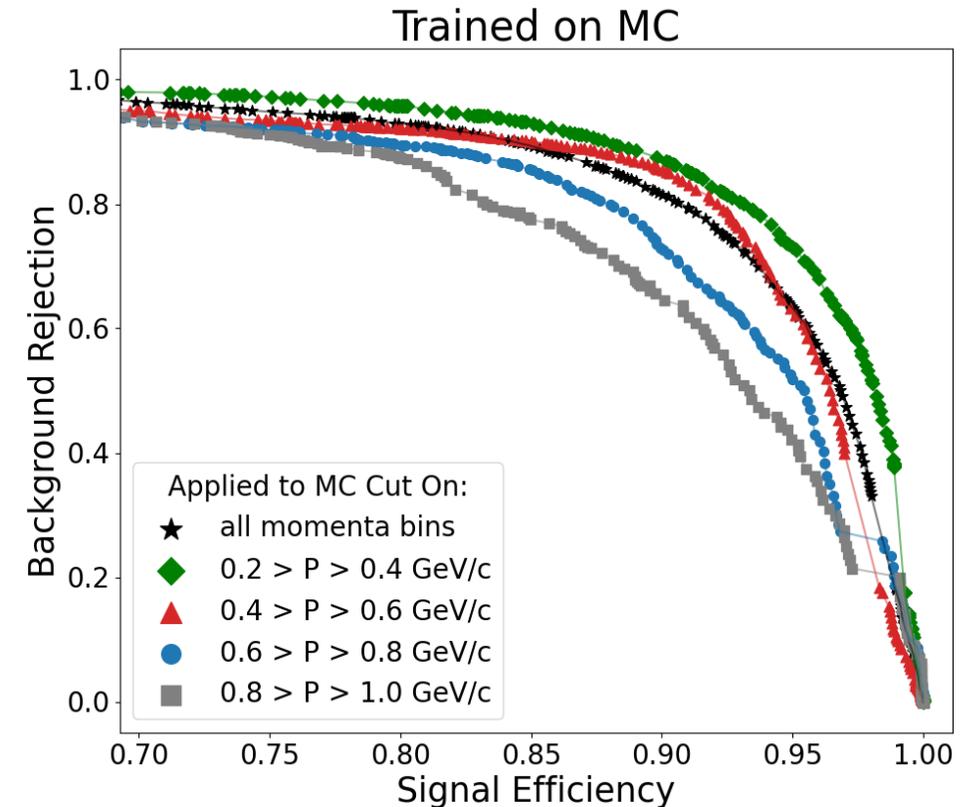
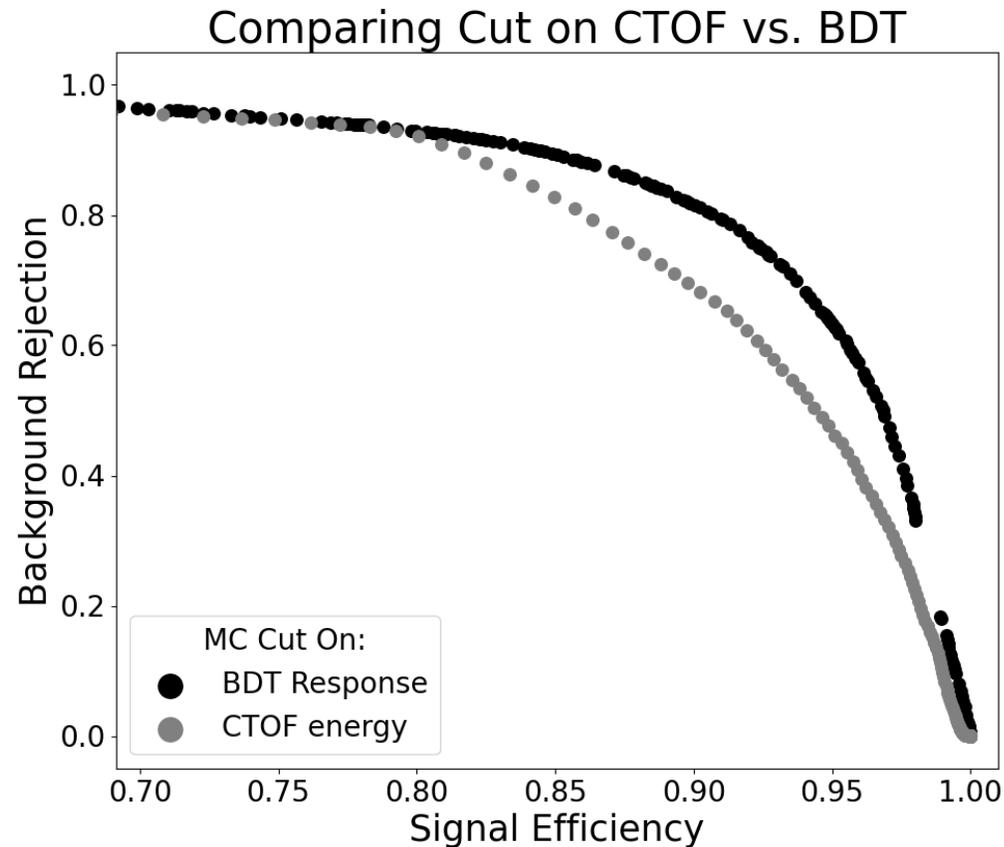


CUT APPLICATION

Low Efficiency Zones



PERFORMANCE & APPLICATION TO MOMENTA BINS



- **Momenta Binning** - At high momentum, protons move quickly through the CVT and have straight tracks which make them hard to distinguish from neutrons in the detectors
 - Applying training to high and low momenta bins can show us how we can adjust training in future tests

SUMMARY

- **Generic Problem:** Protons get misidentified as neutrons in the CD
 - Low efficiency zones in CVT
- **Specific Example:** Deeply Virtual Compton Scattering (ndvcs and pdvcs)
 - Trained BDT on MC with ND3 target for the purpose of cleaning up neutron data contaminated by misidentified protons
 - BDT response cut was made preserving high signal efficiency (~90%)
- Applied BDT trained on MC to MC cut on low and high momenta bins
 - Shows an area of interest for adjustments to BDT training (can train on momenta bins & see performance)

THANK YOU!