

# Identification of Charged Leptons in Jets at Future Higgs Factories

Master Colloquium

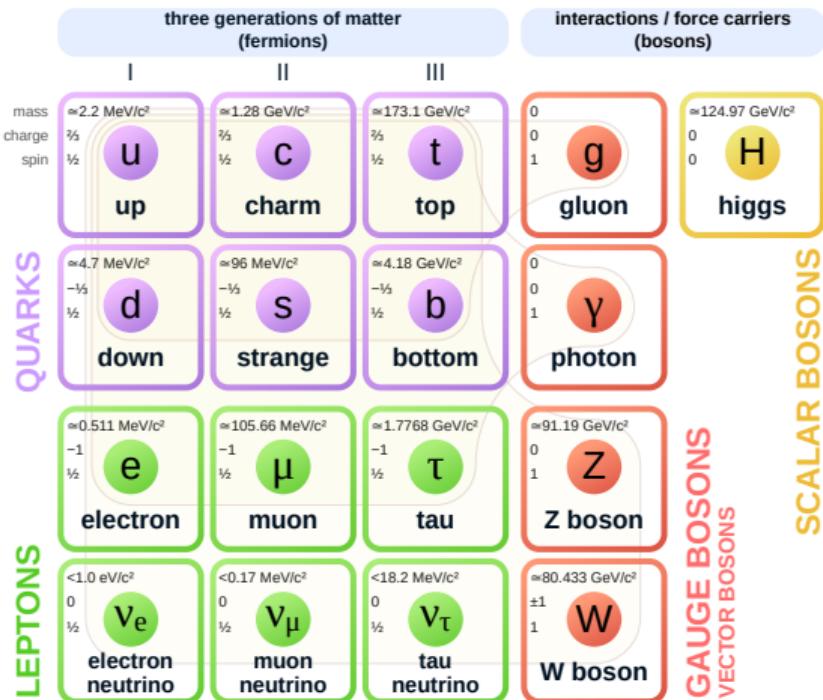
Leonhard Reichenbach  
Hamburg, 01.12.2022

HELMHOLTZ



# Introduction

## Standard Model of Elementary Particles

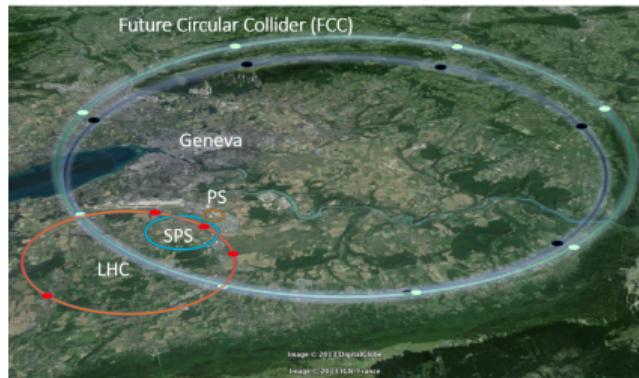
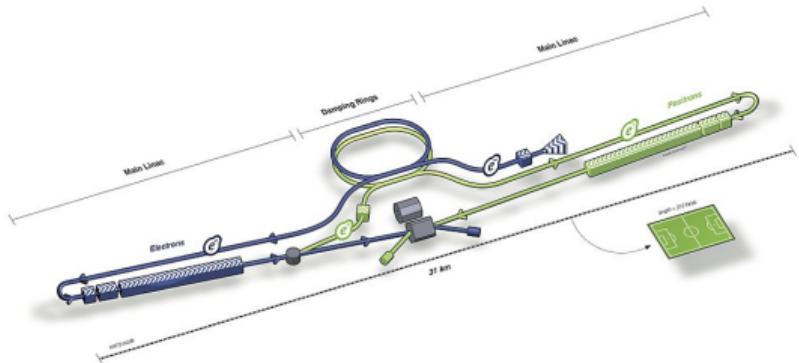


- Higgs discovered in 2012 at the LHC as the 'final' SM particle
- precision studies of its properties desirable to answer further questions
- we want to study as many Higgses as possible, ideally in a clean environment

# Future Higgs Factories

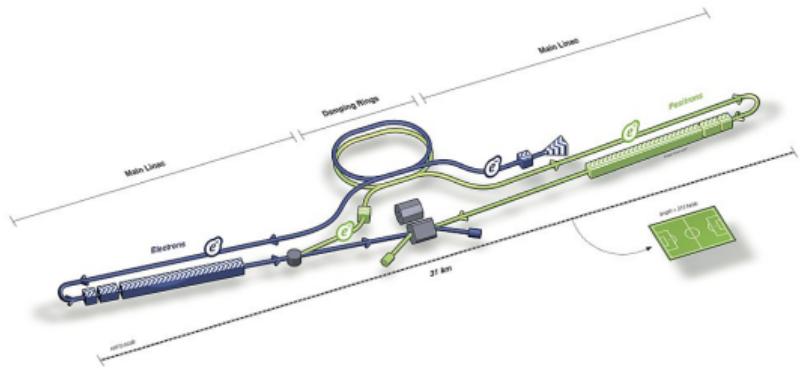
The next highest priority collider

- > electron-positron colliders → clean collision environment
- > linear or circular → extendability vs. luminosity
- > lots of concepts in varying stages of readiness e.g. ILC, FCC-ee, C3/HELEN, CEPC, (CLIC)
- > operation at the Higgsstrahlung peak around 240 GeV to 250 GeV but also plans for Z-pole (91 GeV), WW (160 GeV) and  $t\bar{t}$  (350 GeV) runs

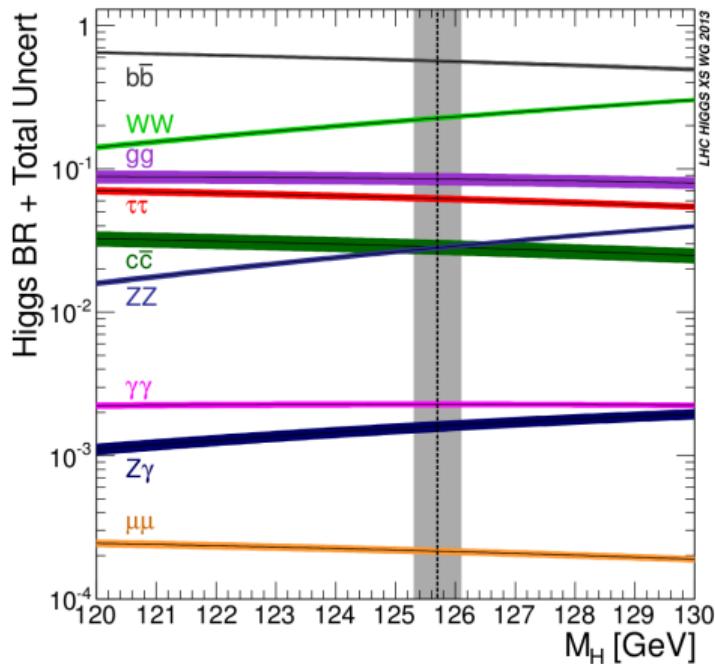
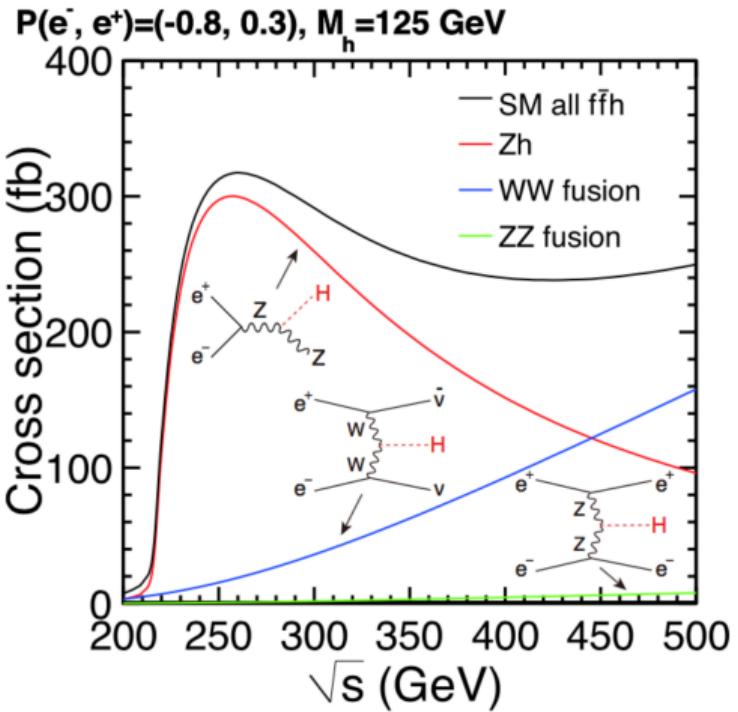


# International Linear Collider (ILC)

- > 31 km long linear collider
- > well tested accelerator technology (TESLA/XFEL)
- > initially 250 GeV, planned upgrades to 500 GeV, 1 TeV
- > 'shovel-ready', TDR published in 2013
- > currently under political consideration in Japan



# Higgs Physics

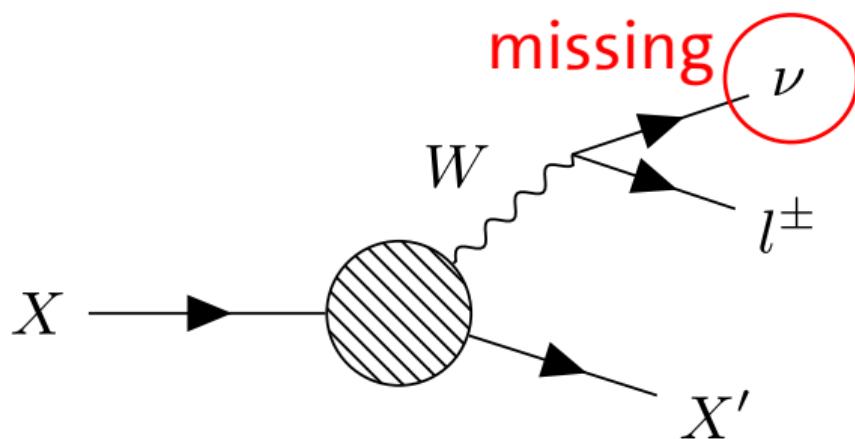


# The Problem

We are interested in Higgs physics, but:

- >  $H \rightarrow b\bar{b}$  most common decay
- >  $\sim 2/3$  of those decays have at least one subsequent semi-leptonic decay (SLD) of a  $B, D$  hadron
- > the neutrino energy goes missing
- > gets worse e.g.  $\sim 4/5$  for  $HH \rightarrow b\bar{b}b\bar{b}$  important for Higgs self-coupling measurements.

This can be fixed under certain conditions.

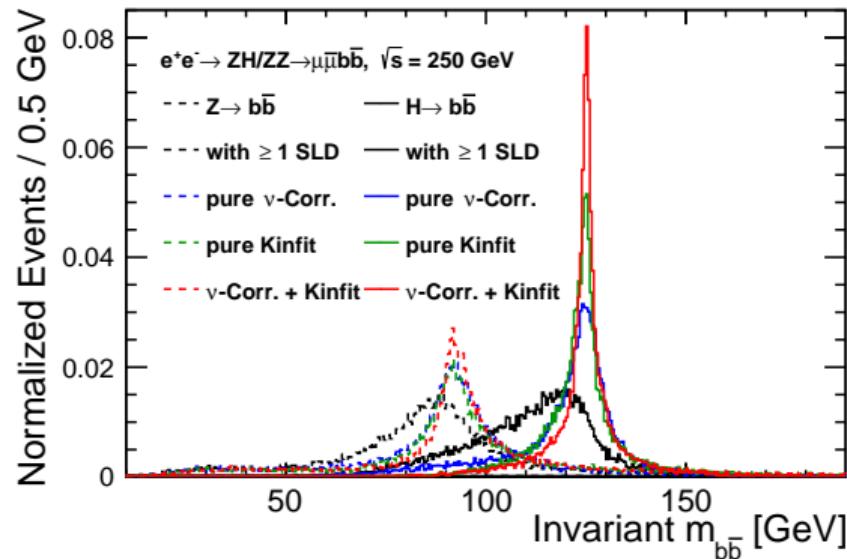


# Kinematic Fitting and $\nu$ -correction

Work done by Y. Radkhorrami

Reconstructing neutrinos:

- > if we find the  $e/\mu$  from the semi-leptonic decay and its production vertex we can reconstruct the  $\nu$  momentum up to a sign
- > do a kinematic fit to find the right sign afterwards



[arXiv: [2111.14775](https://arxiv.org/abs/2111.14775)]

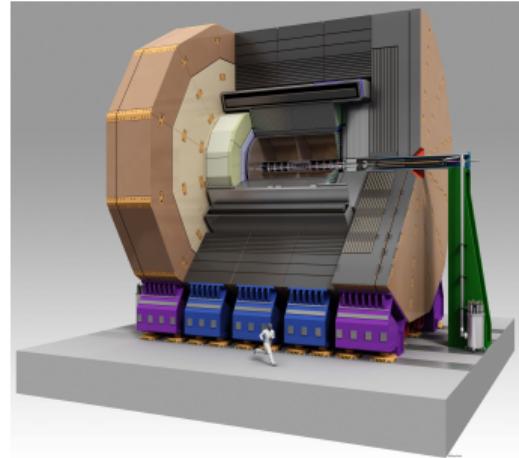
# Ingredients

What do we need?

The  $\nu$ -correction already works!  
But it needs to know the SLD lepton!

For that we need:

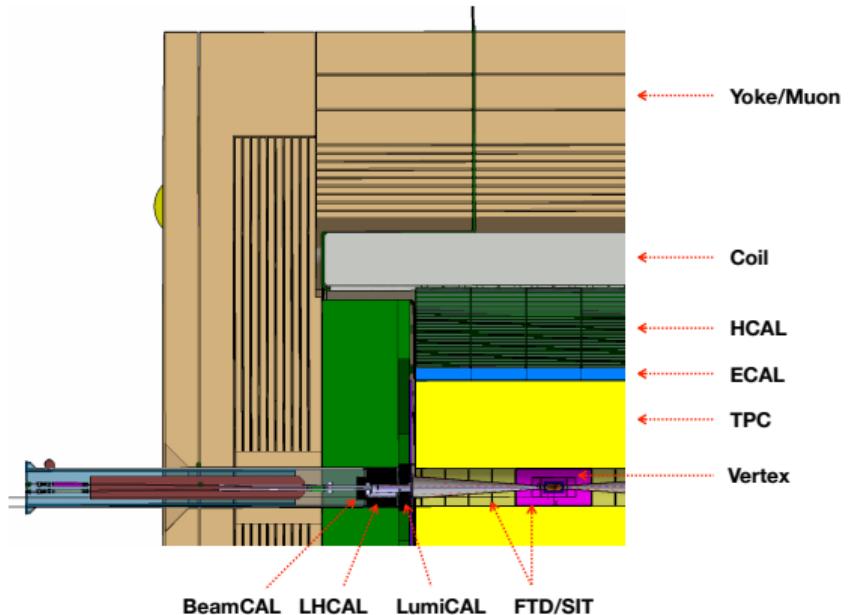
- > a highly granular detector able to resolve all charged jet particles (i.e. particle flow) → ILD
- > good lepton ID to identify jet leptons → this work
- > a way to determine if a jet lepton comes from a SLD  
→ also this work



# The International Large Detector (ILD)

Planned for the ILC

- > general-purpose detector
- > TPC providing continuous tracking and dE/dx measurements
- > highly-granular calorimeters (cell size:  
Ecal  $(0.5 \text{ cm})^2$ , Hcal  $(3 \text{ cm})^2$ )
- > PandoraPFA for particle flow  
used as an example here.

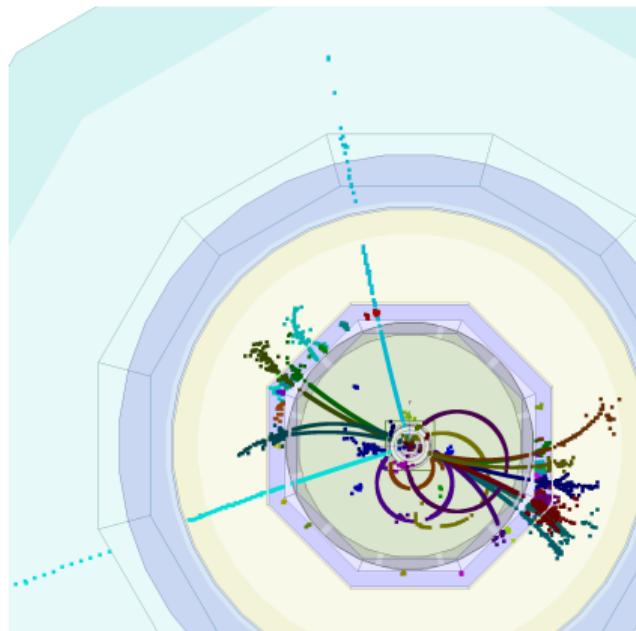


# Lepton ID

How does it work?

Directly done in the particle flow reconstruction

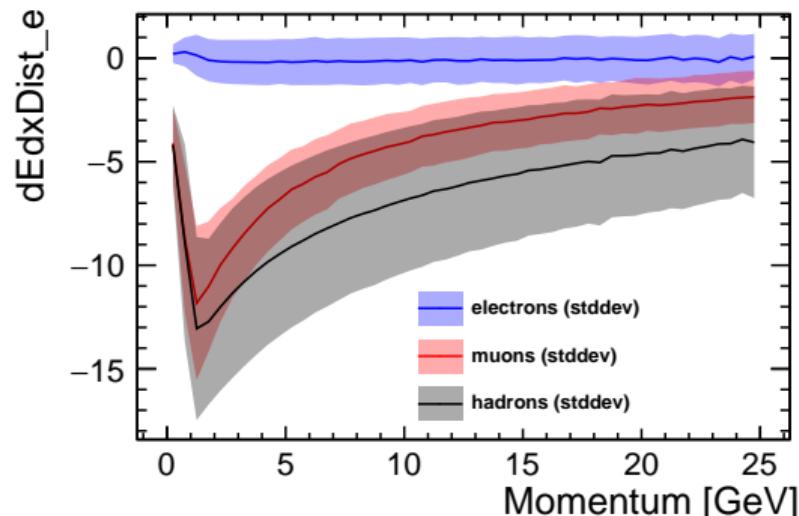
- > differentiates between  $e$ ,  $\mu$  and charged hadrons
  - > based on topological shower shape information
  - > full hit data available
  - > complicated and hard to modify
- try to build a more easily extendable method with improved  $e/\mu$  ID performance.



ILD  $ZH \rightarrow \mu^+\mu^- b\bar{b}$  @ 250 GeV

# Lepton ID Recipe

- > 3 classes  $e/\mu/\text{hadrons}(\pi + K + p)$
- > ROOT TMVA Multiclass BDTG
- > focused on  $ZH \rightarrow \mu^+ \mu^- b\bar{b}$
- > only used information already in DST files (no single detector hits)
- > 2 versions:
  - both using precomputed shower/cluster shape information
  - one additionally uses the  $dE/dx$  distance to the electron curve
  - 21 variables in total



## Definitions:

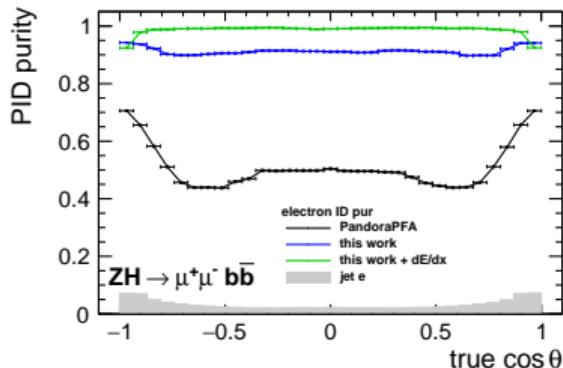
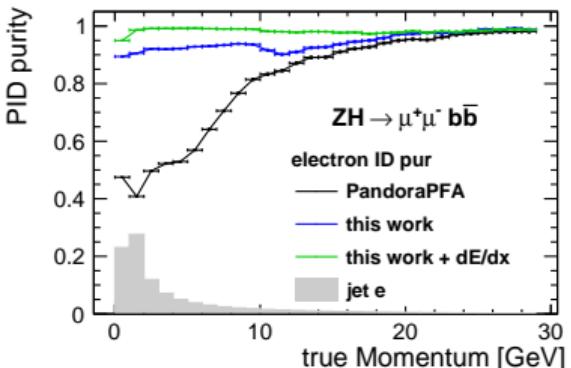
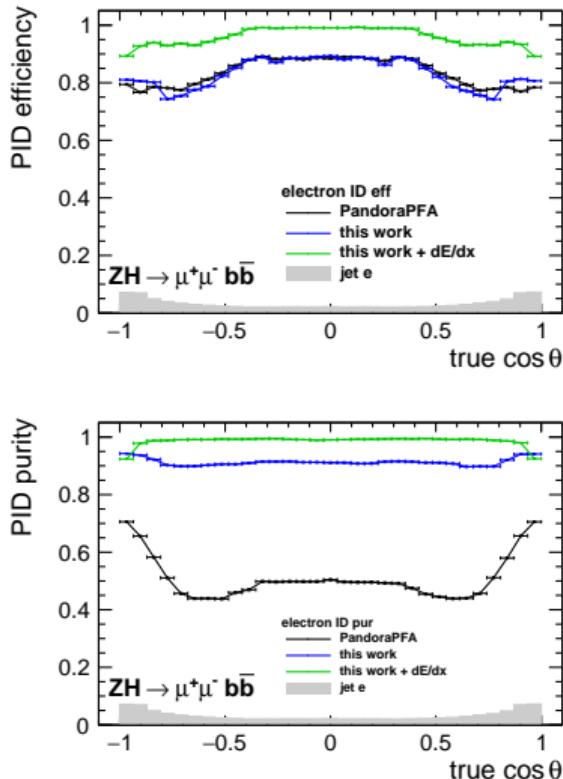
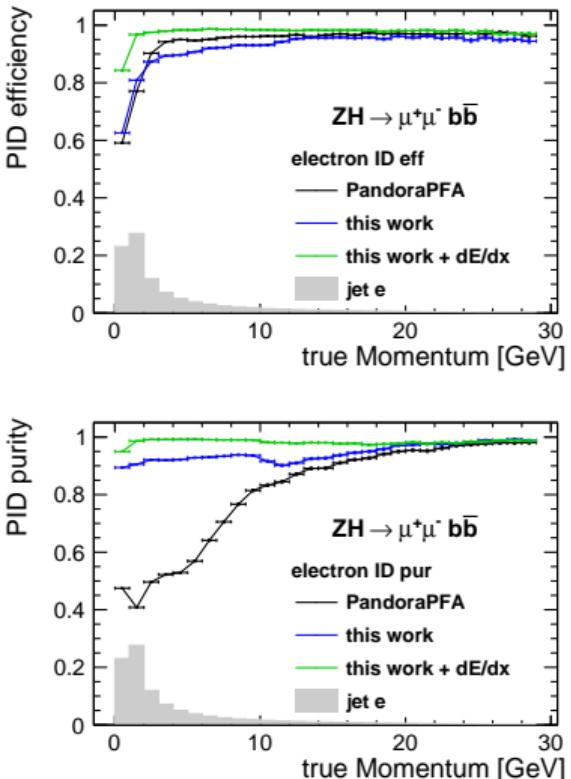
- > PID efficiency =  $\frac{\text{correctly identified}}{\text{identifiable}}$
- > PID purity =  $\frac{\text{correctly identified}}{\text{identified as}}$

Identifiable:

- > PFO has one track and one cluster
- > track/cluster were caused by the same MC particle to at least 50%
- > 50% of all hits from the MC particle are assigned to the track/cluster

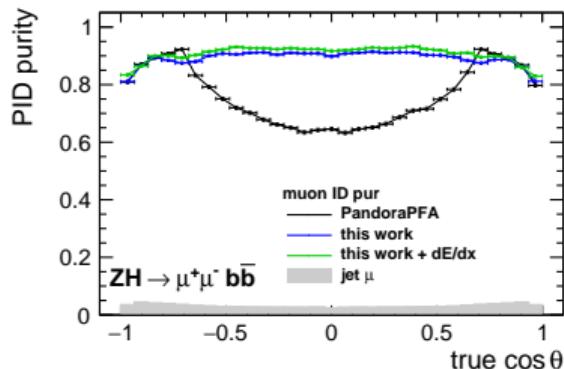
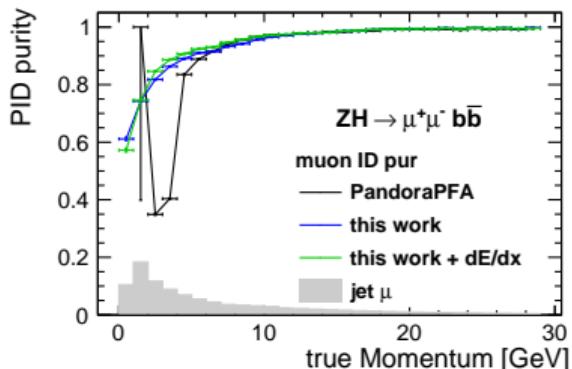
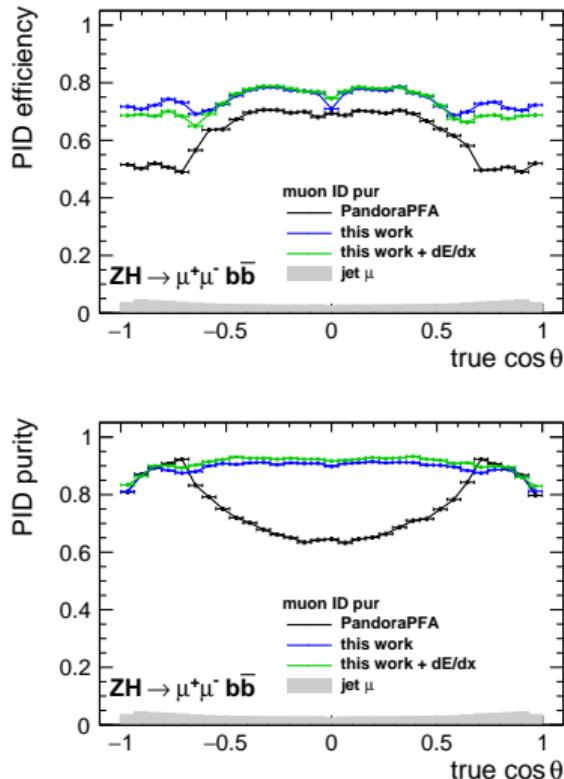
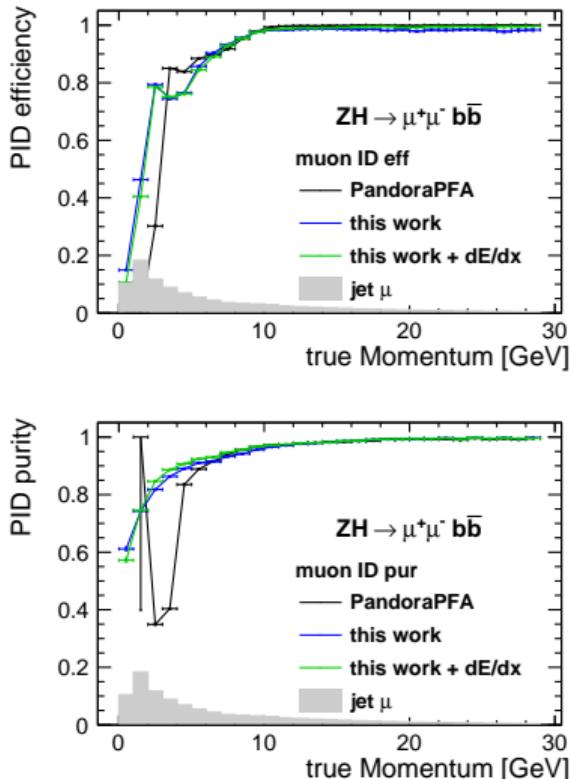
# Lepton ID Results: $b$ -Jet Electrons

- > efficiency similar to PandoraPFA
- > but much purer
- > highlight: dE/dx adds a significant improvement for electrons!



# Lepton ID Results: $b$ -Jet Muons

- > efficiency similar, better in forward region
- > purity better in barrel region



# Lepton ID Conclusion

- > overall performance for  $H \rightarrow b\bar{b}$  jet leptons:
  - $e$ : 95% efficiency and 98% purity with dE/dx
  - $e$ : 82% efficiency and 92% purity without dE/dx
  - $\mu$ : 74% efficiency and 89% purity
- > already works well while a lot of obvious improvements are not implemented yet
- > having dE/dx information available improves electron identification significantly
- > using full reconstruction data instead of just DST will offer further improvement

# Identification of Semi-Leptonic Decays

Still need to decide if an  $e/\mu$  is from SLD

- > not a completely new task: also used in  $b$ -tagging
- > usually done via cut-based selection on track/vertex parameters
- > the number of so found 'soft' leptons is then used as flavor tag input

Here:

- > do this with another set of classifiers (BDTs)
- > 2x2 categories:  
 $\{e, \mu\} \times \{\text{vtxState1}, \text{vtxState2}\}$
- > vtxState:  
$$\begin{cases} 1: & \text{pfo has assigned LCFIPlus vertex} \\ 2: & \text{interpolated vertex} \end{cases}$$

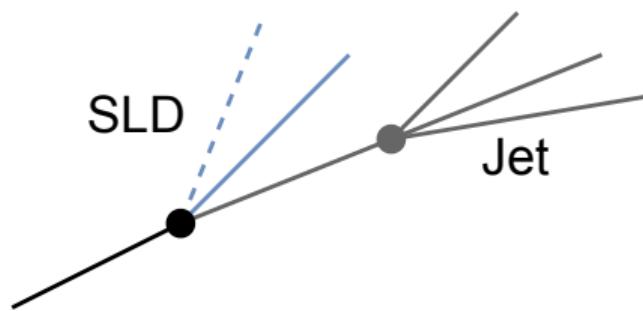
# SLD ID Inputs Example

$b$  lifetime long enough to travel a few mm from the interaction point (IP)

- > look at impact parameters ( $z_0, d_0$ ) and vertex position

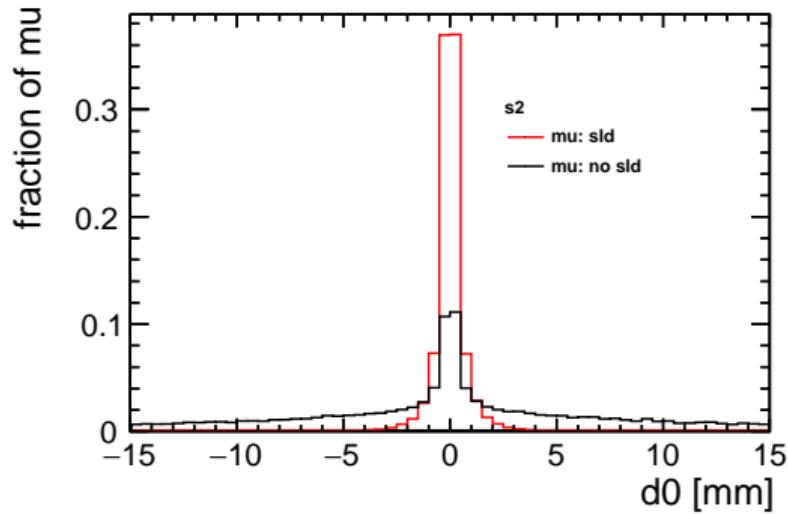
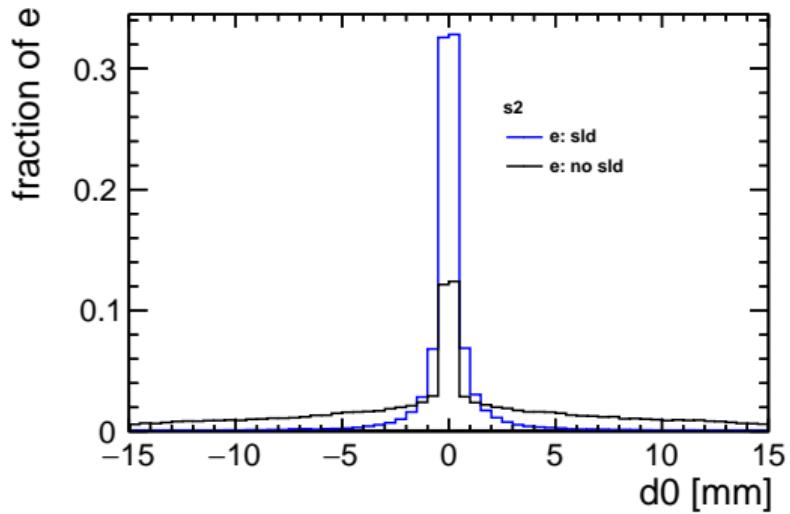
SLD first decay in  $b$ -jet

- > higher momentum fraction, higher transverse momentum relative to the jet axis
- > different angle relative to the jet axis

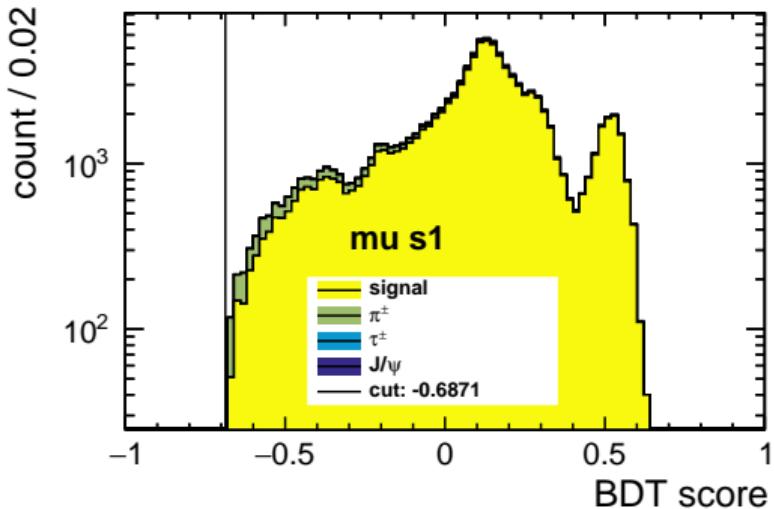
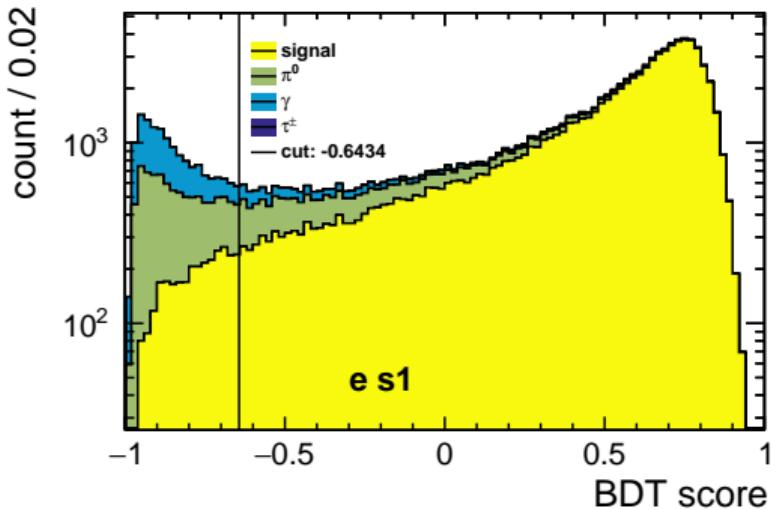


# SLD ID Inputs

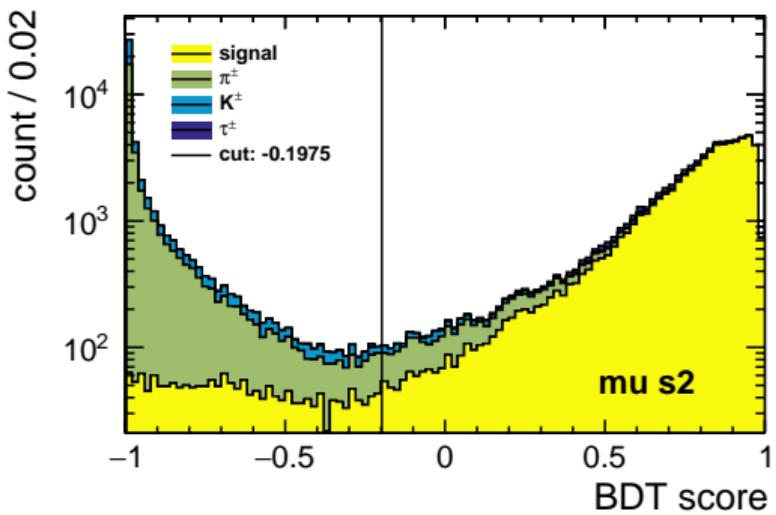
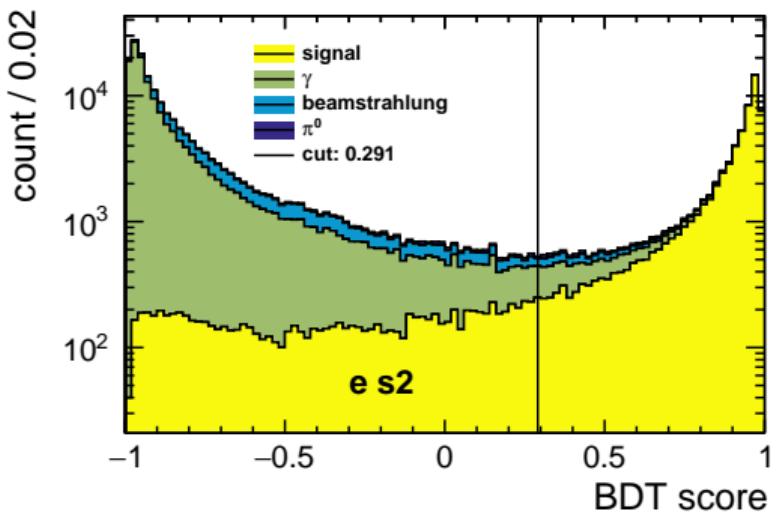
vtxState 2



# SLD ID Results: BDT-score Distributions



# SLD ID Results: BDT-score Distributions



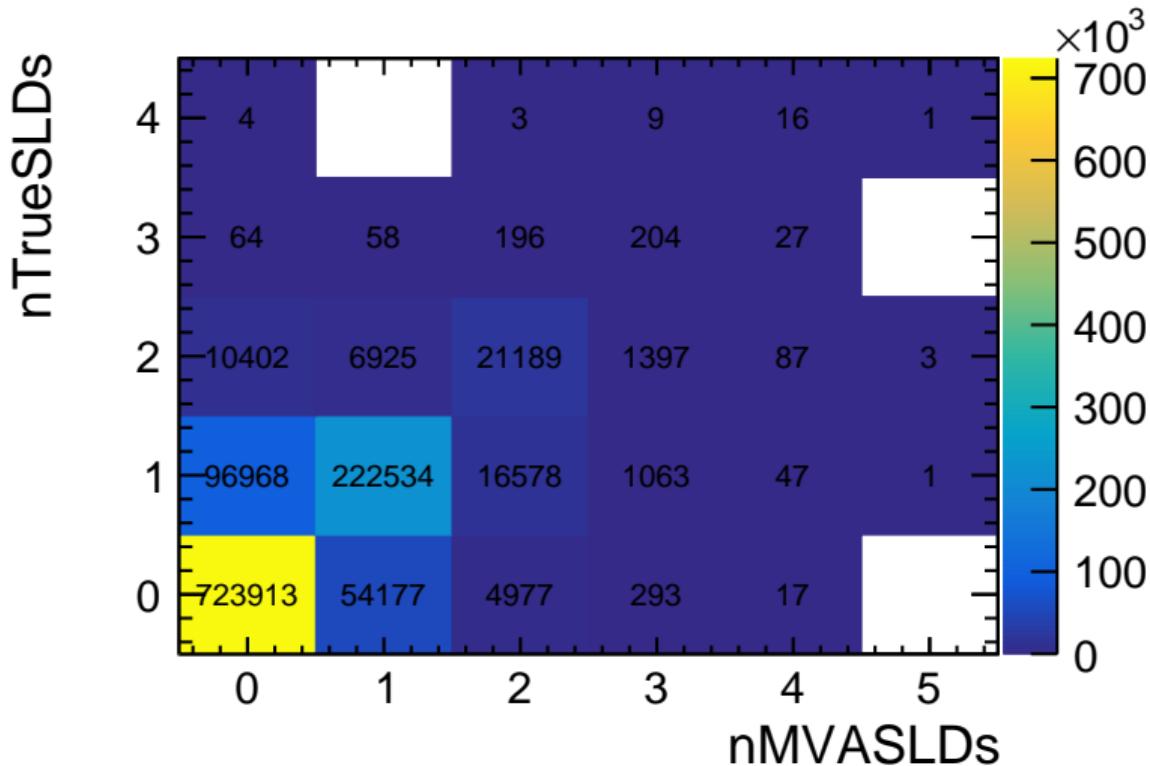
# SLD ID: Particle-by-Particle Results

	efficiency	purity	pre-purity
electrons state 1	97%	86%	77%
electrons state 2	86%	86%	27%
muons state 1	100%	93%	93%
muons state 2	97%	90%	57%

Caution: this is the efficiency of just the SLD ID classifier with cheated PID

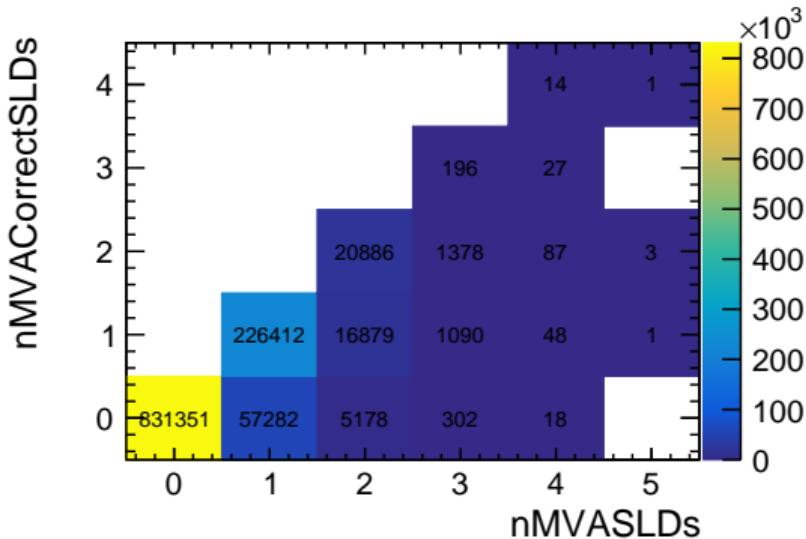
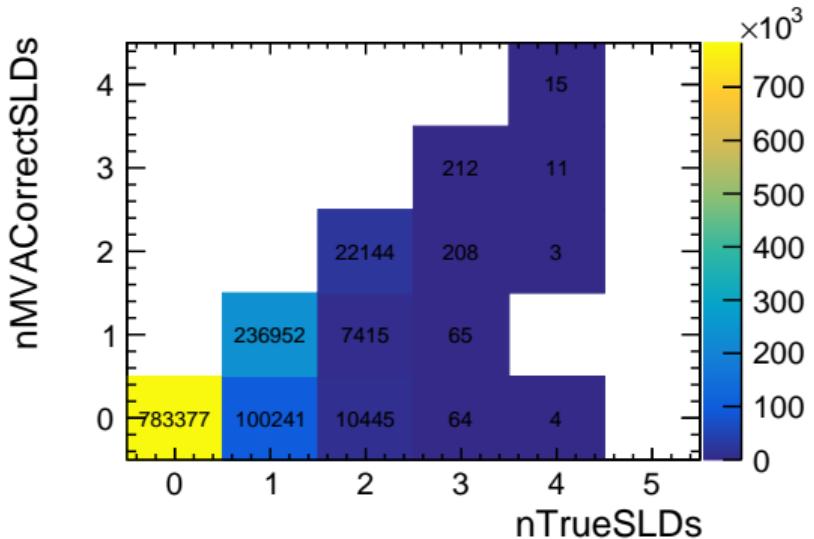
# From Particles to Jets

with new lepton ID



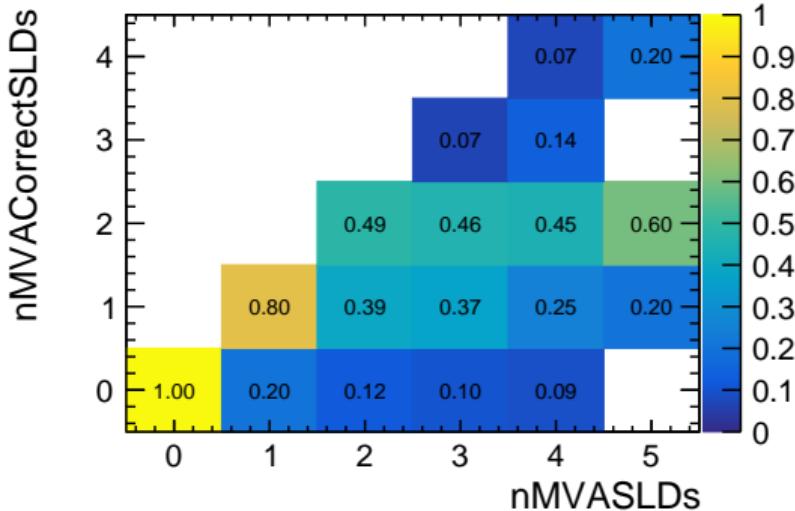
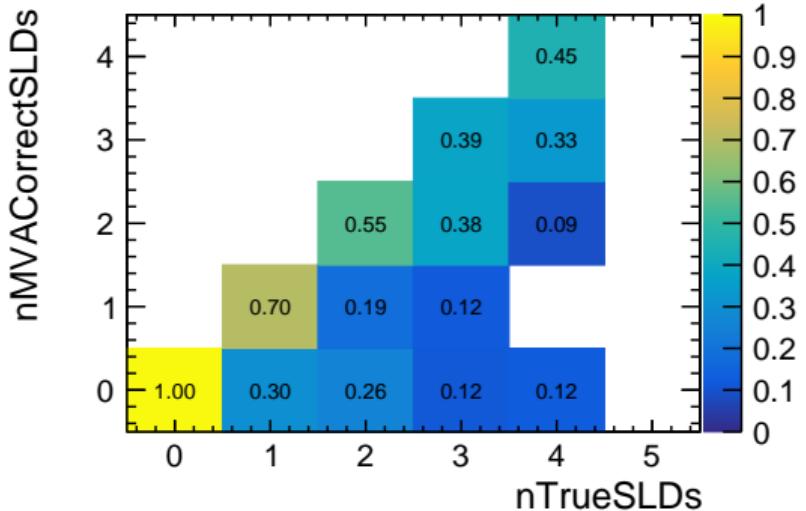
# From Particles to Jets

with new lepton ID



# From Particles to Jets

with new lepton ID

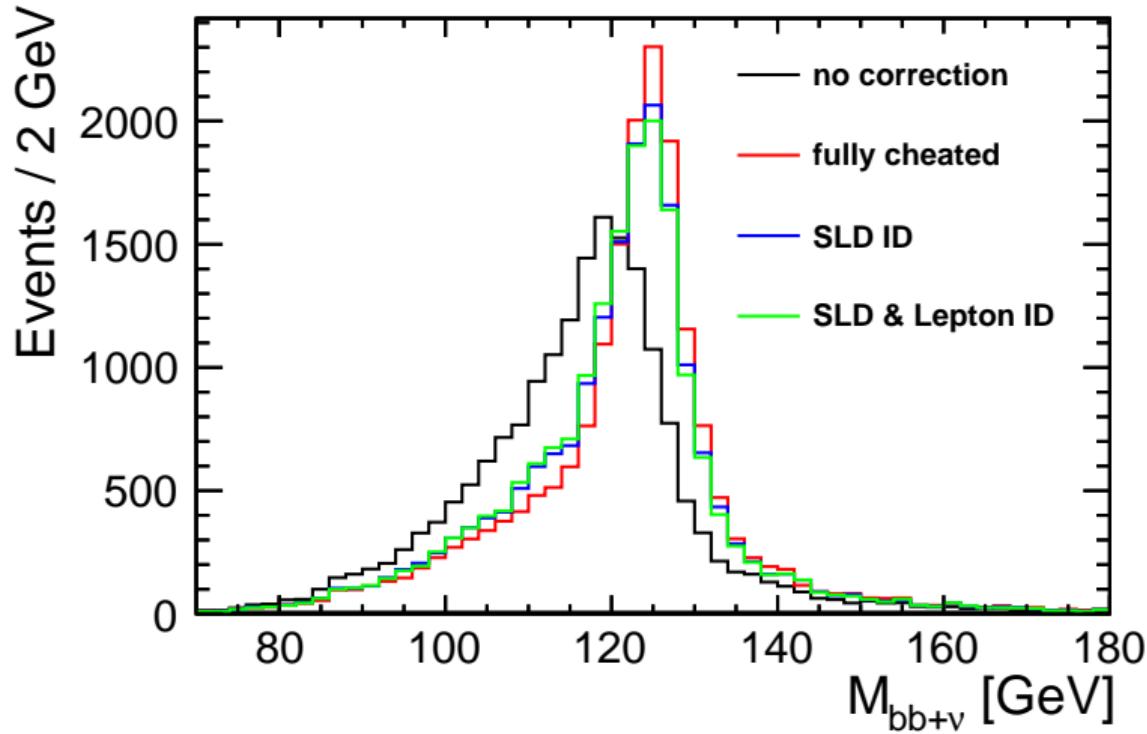


## From Particles to Jets

- > neutrino correction can only do one neutrino per jet at the moment
- > SLD ID classifier can find multiple (fake) candidates per jet
- > need a decision mechanism on what to do if there is more than one SLD candidate for a jet.
- > simplest case: only use events where one SLD candidate per jet is found

PID	efficiency	purity
cheated	73%	83%
new	65%	77%

# Teaser: Cheated Neutrino Correction with SLD ID



# Possible Improvements

- > more sophisticated methods: binned classifiers, more advanced ML, etc.
- > for lepton ID:
  - fundamental improvements to PandoraPFA muon track/cluster association
  - search for better suited shape variables, specifically for lower momenta
- > for SLD ID:
  - piece-wise background removal
  - more holistic approaches integrated with vertex finding, jet clustering and flavor tagging (LCFIPlus)

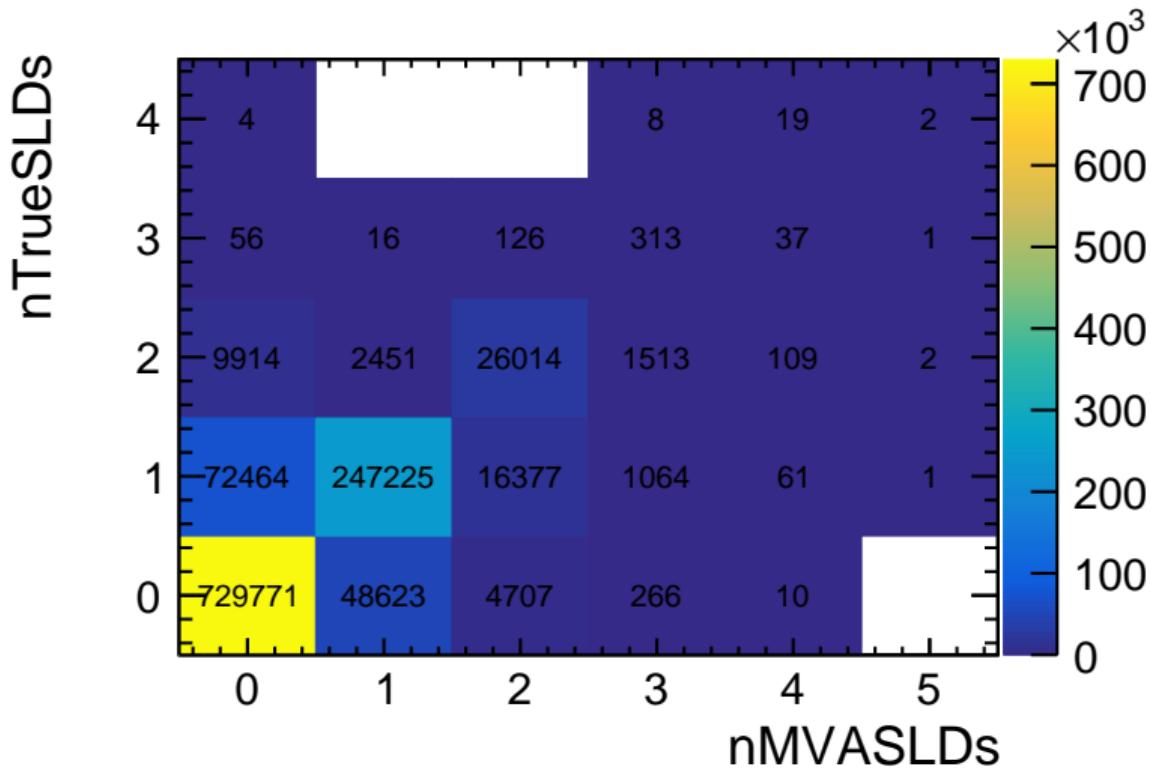
# Conclusion

- > developed an easily extendable BDT-based lepton ID method identifying  $b$ -jet leptons with
  - $e$ : 95% efficiency and 98% purity with  $dE/dx$
  - $e$ : 82% efficiency and 92% purity without  $dE/dx$
  - $\mu$ : 74% efficiency and 89% purity
- > demonstrated the feasibility of SLD lepton identification with
  - 73% efficiency and 83% purity with cheated lepton ID
  - 65% efficiency and 77% purity with the new lepton ID with  $dE/dx$

# Backup

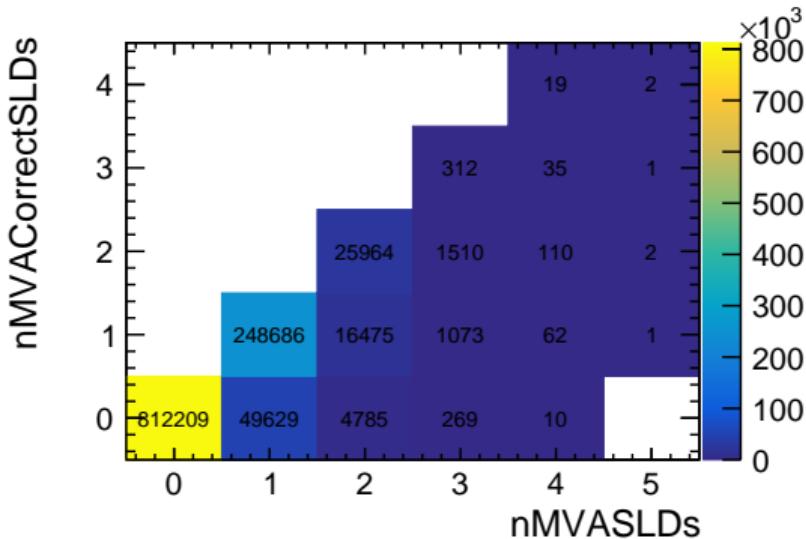
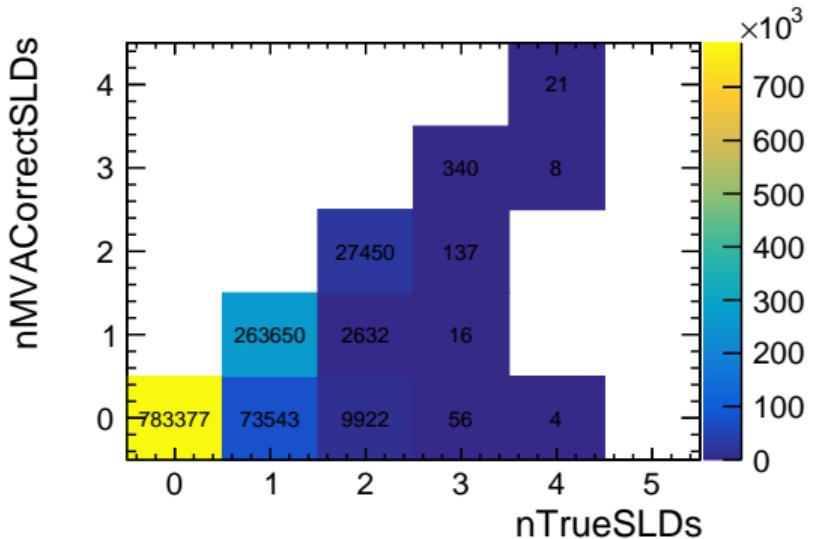
# From Particles to Jets

cheated PID



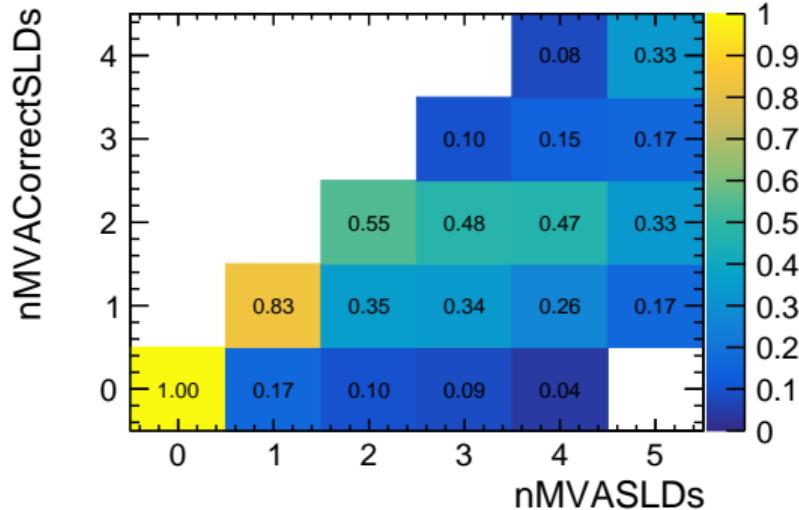
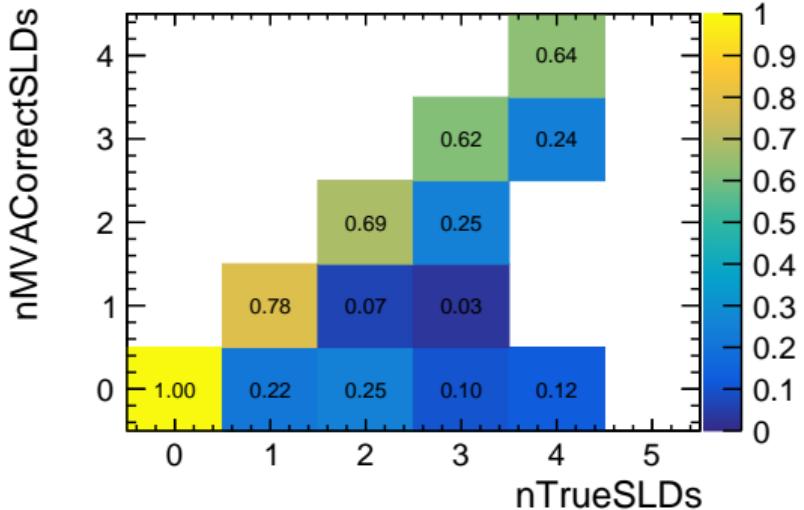
# From Particles to Jets

cheated PID

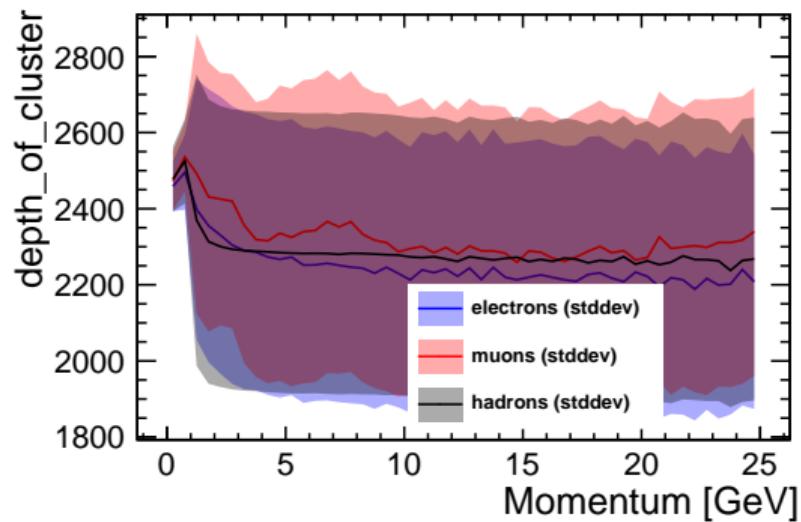
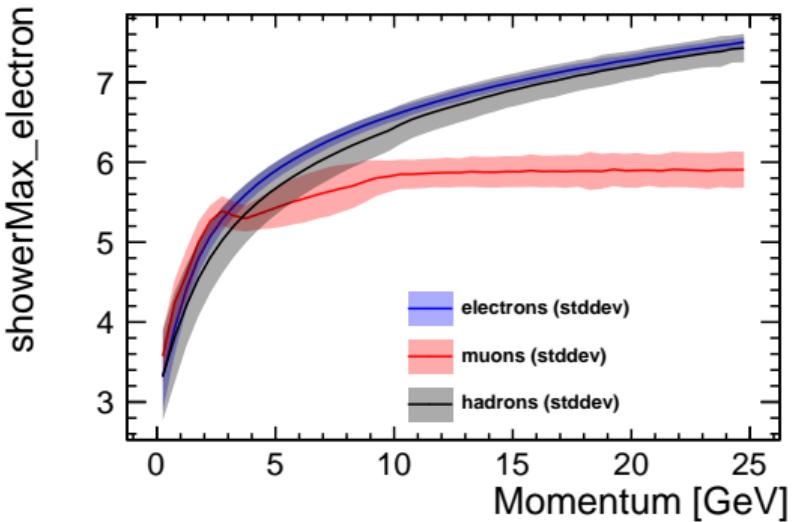


# From Particles to Jets

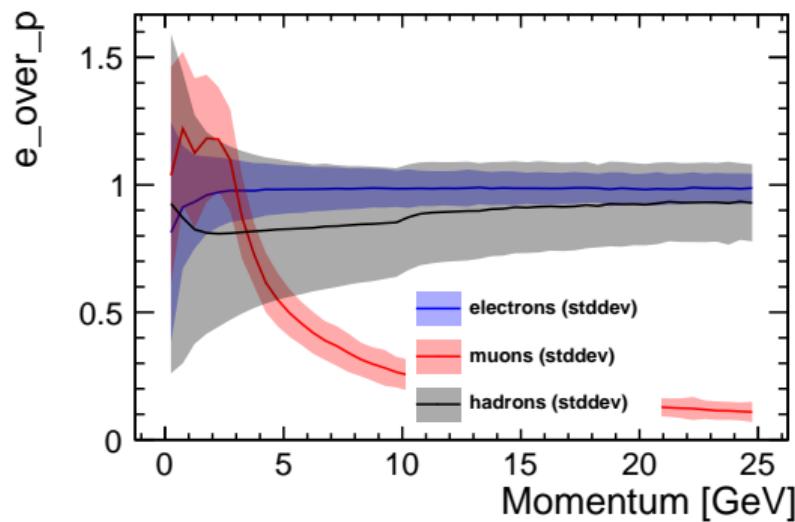
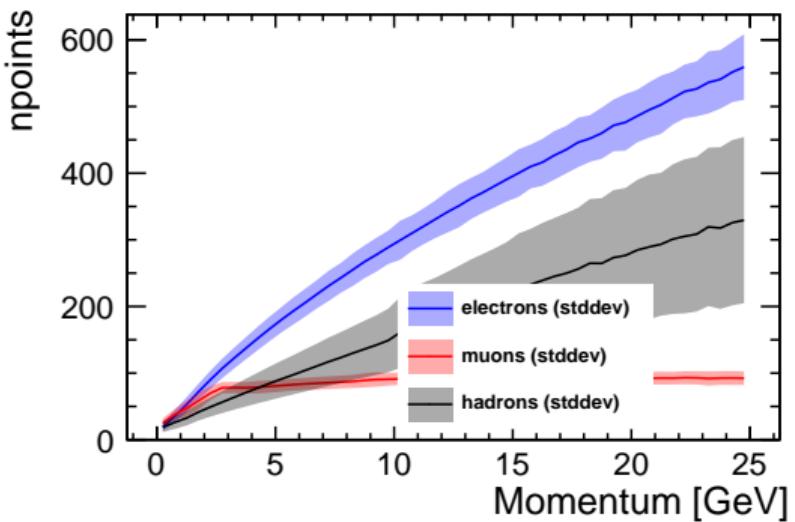
cheated PID



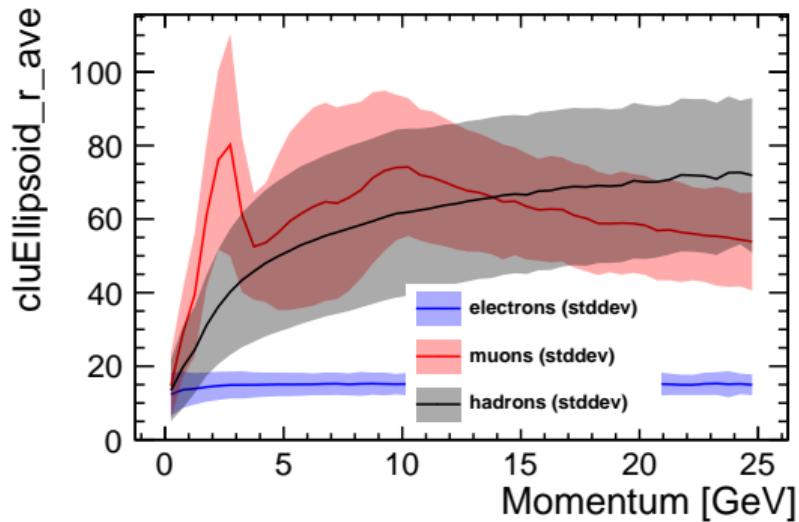
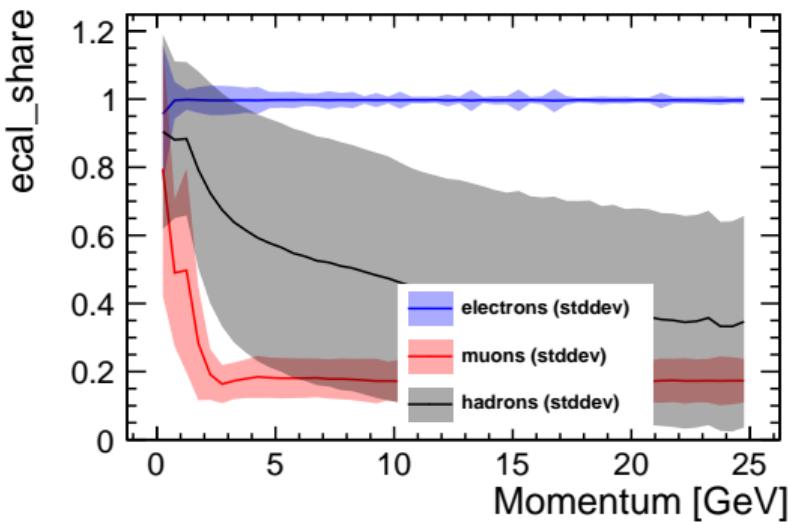
# Lepton ID Inputs



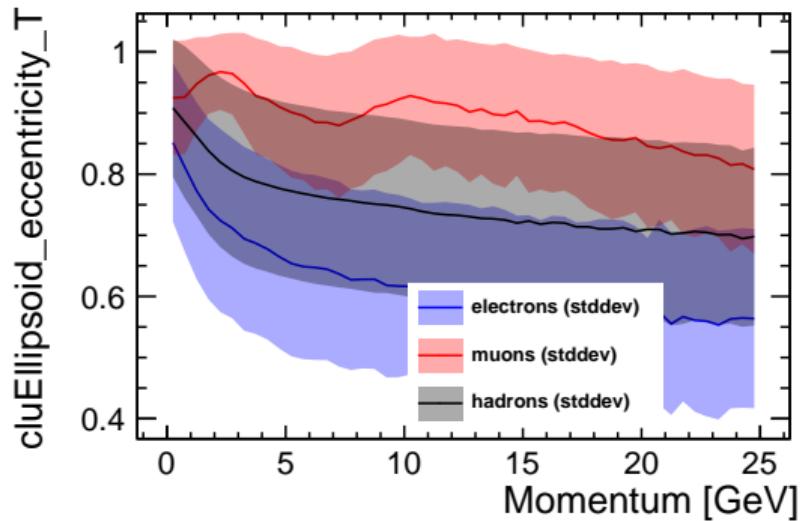
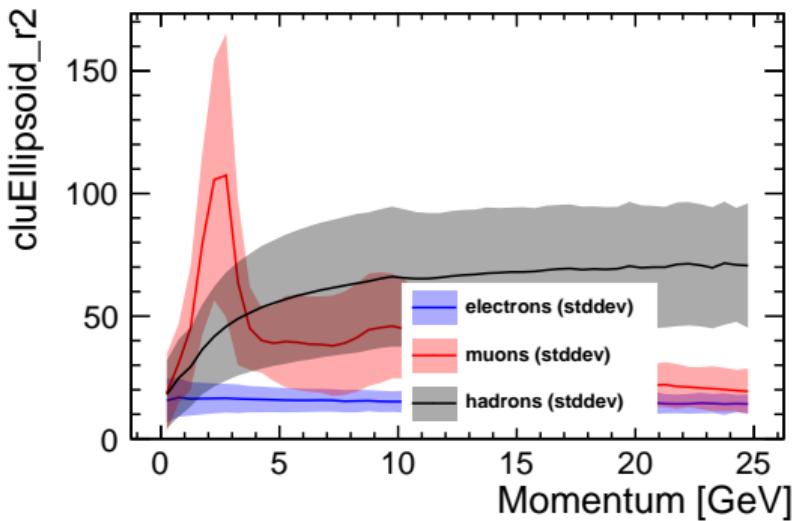
# Lepton ID Inputs



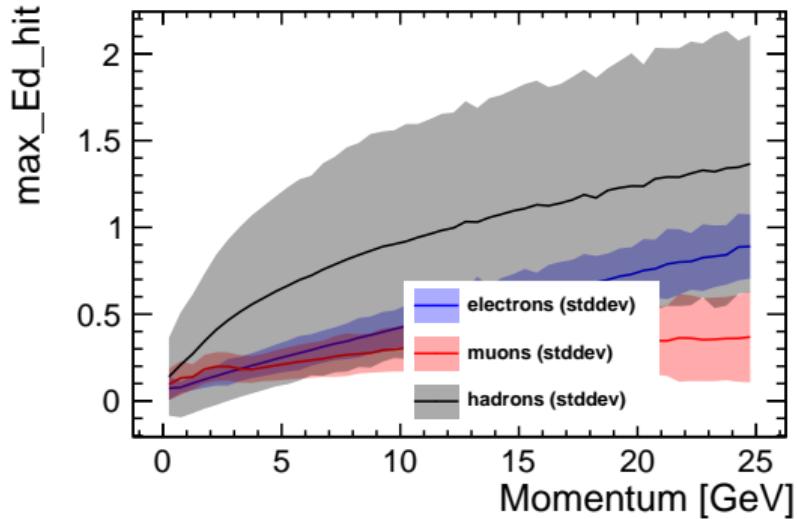
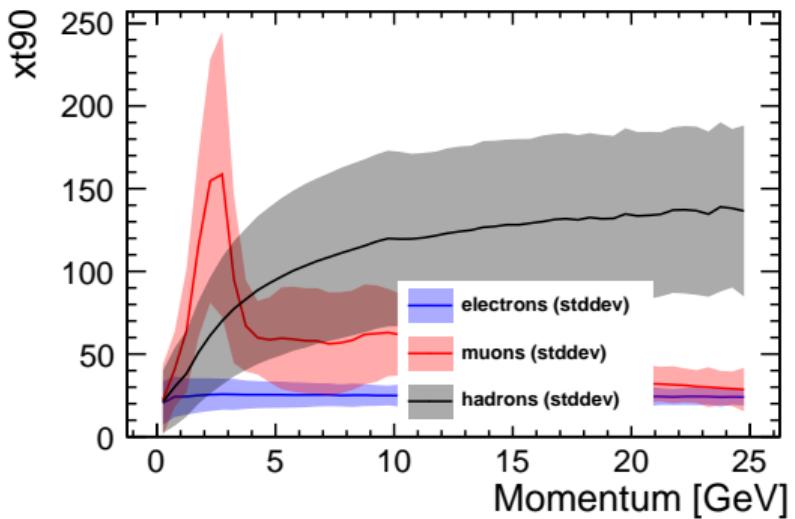
# Lepton ID Inputs



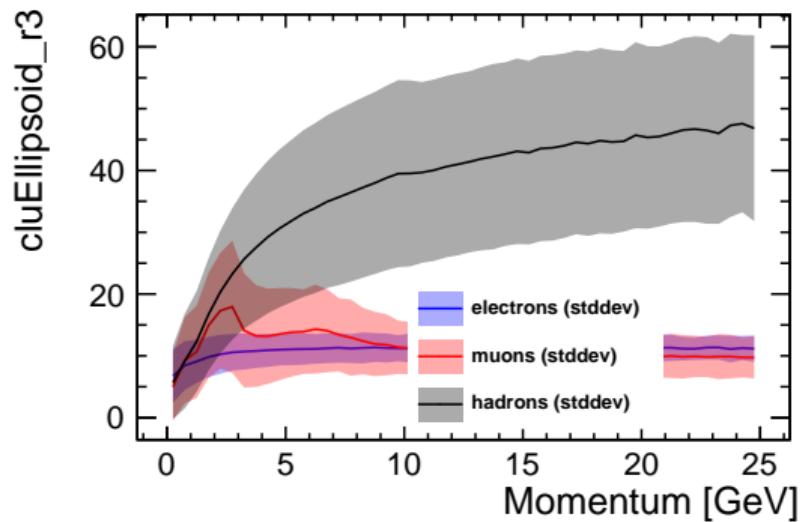
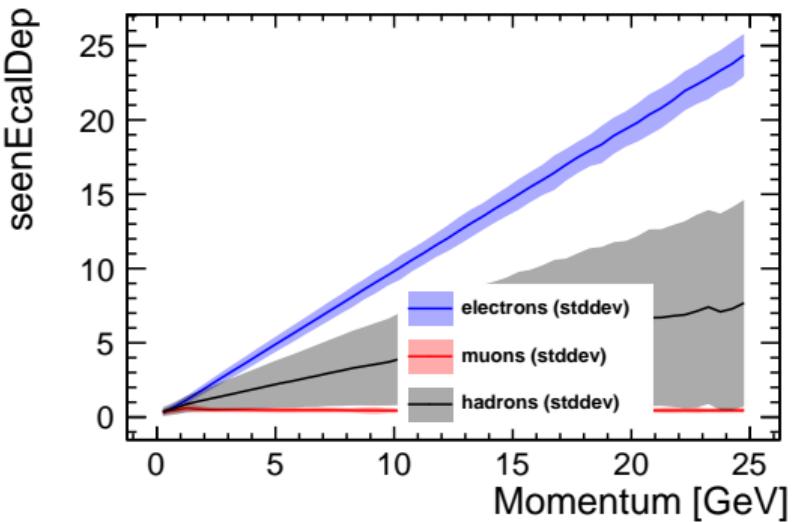
# Lepton ID Inputs



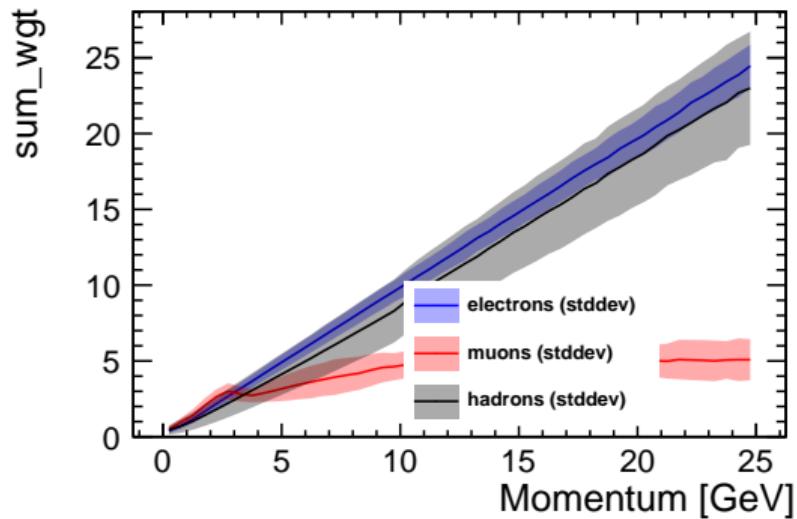
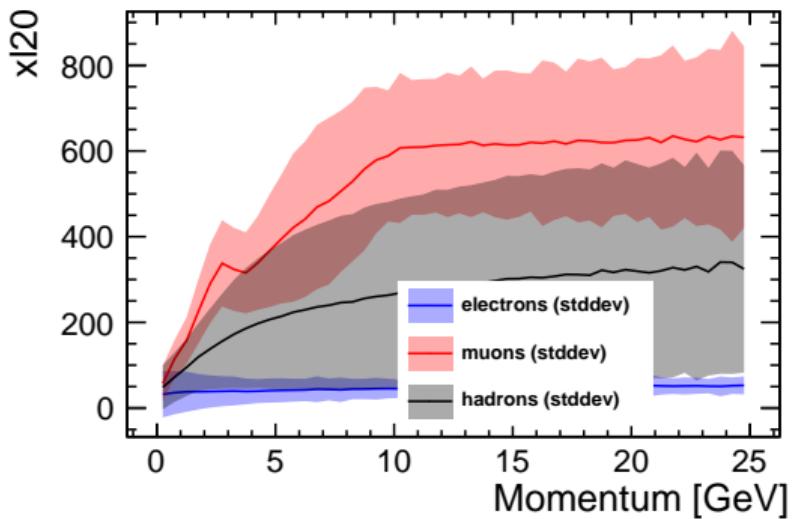
# Lepton ID Inputs



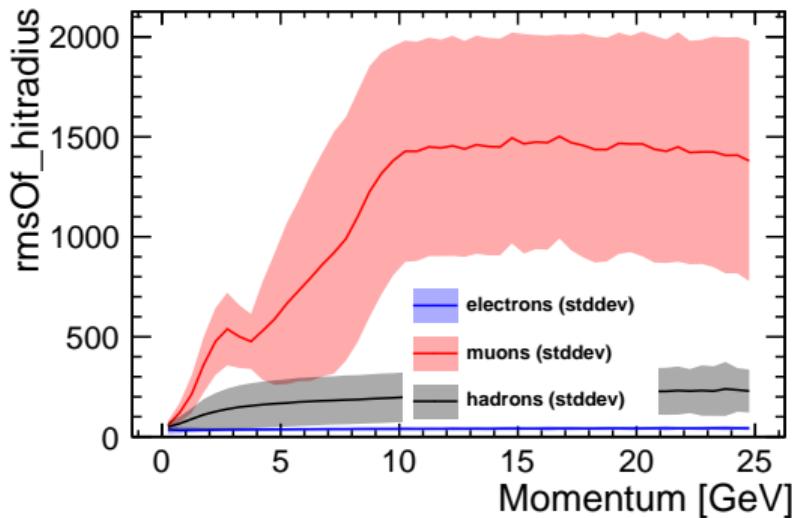
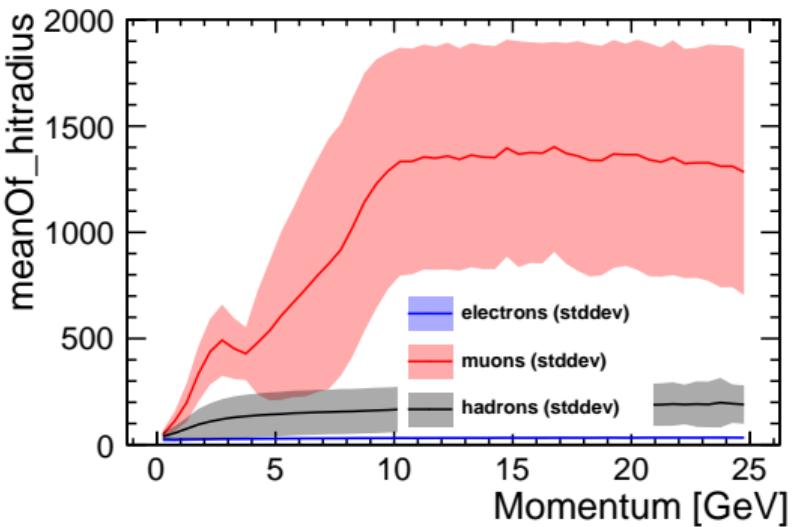
# Lepton ID Inputs



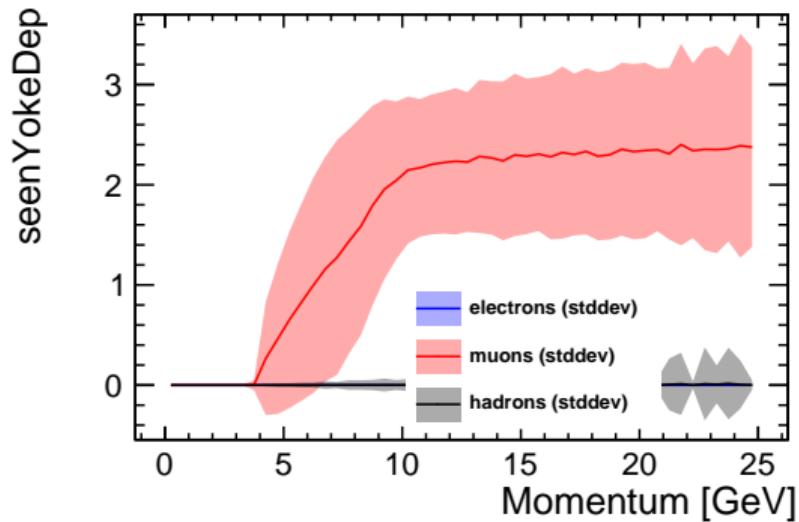
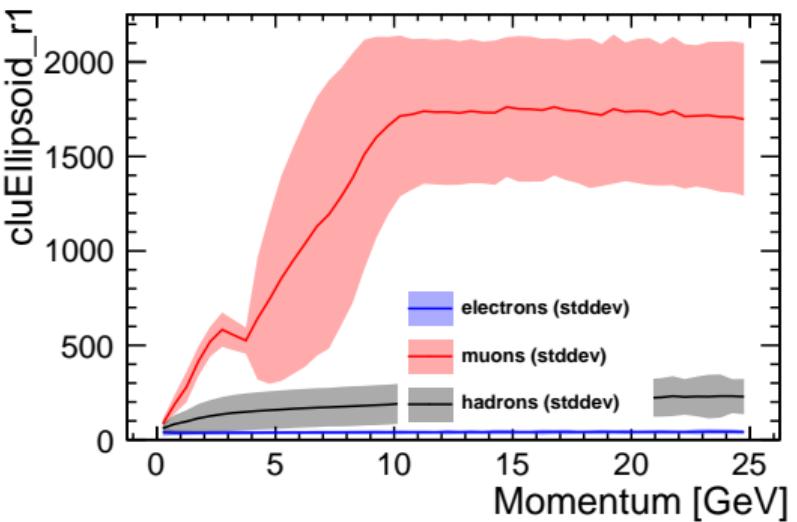
# Lepton ID Inputs



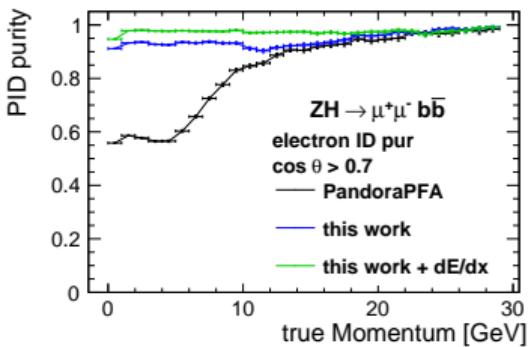
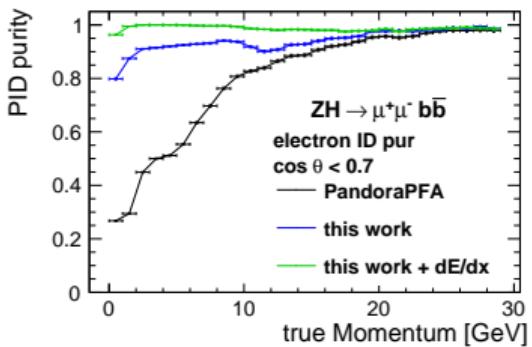
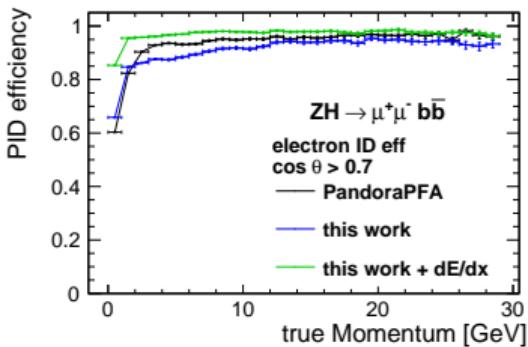
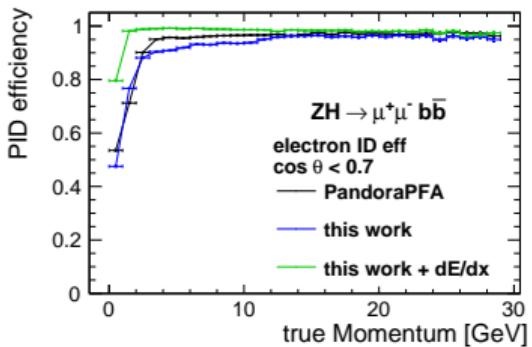
# Lepton ID Inputs



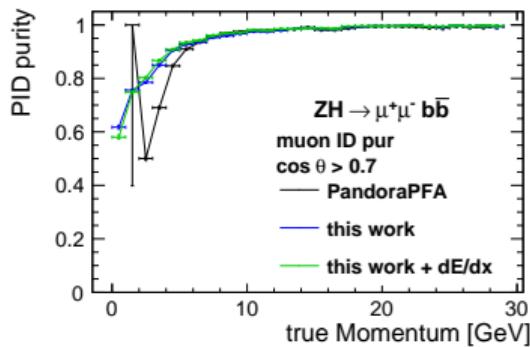
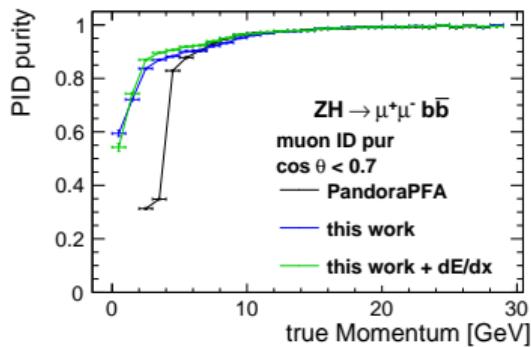
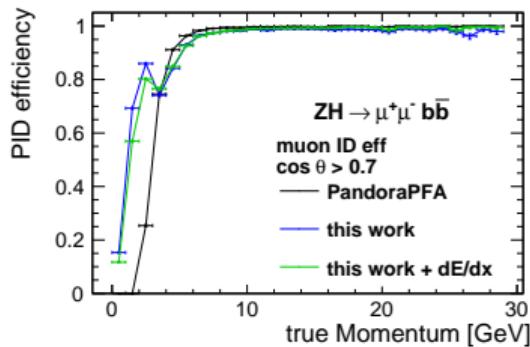
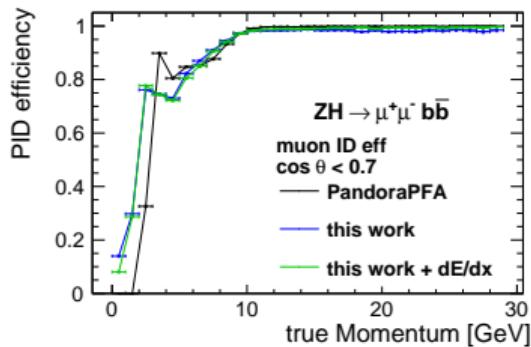
# Lepton ID Inputs



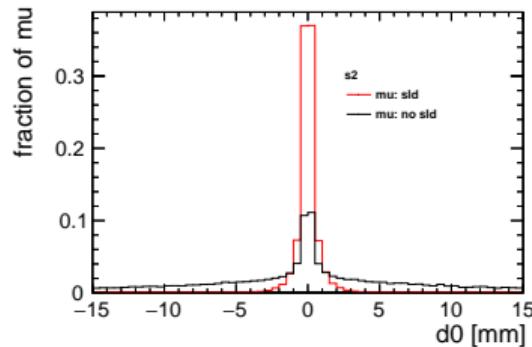
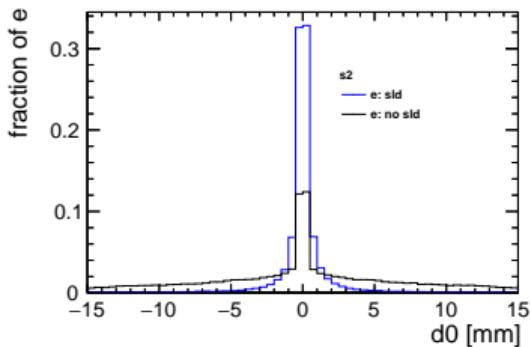
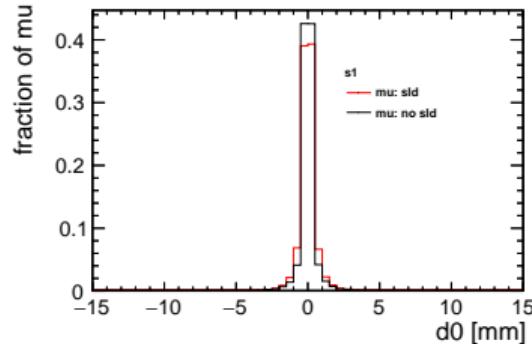
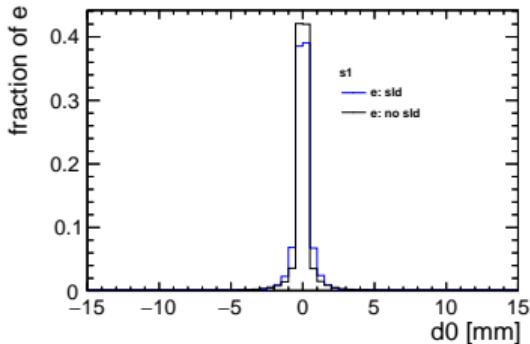
# Lepton ID Results: $b$ -Jet Electrons



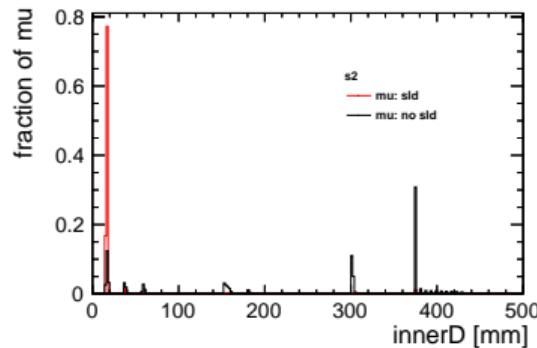
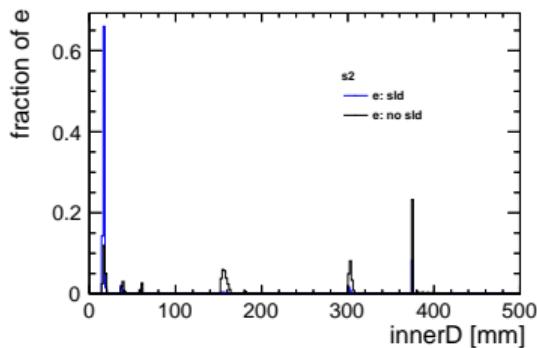
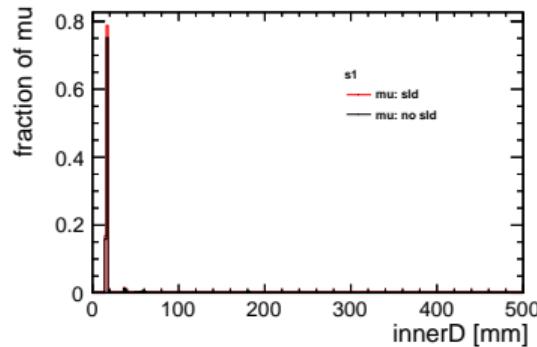
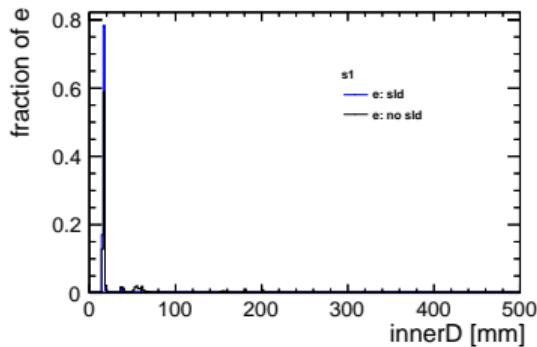
# Lepton ID Results: $b$ -Jet Muons



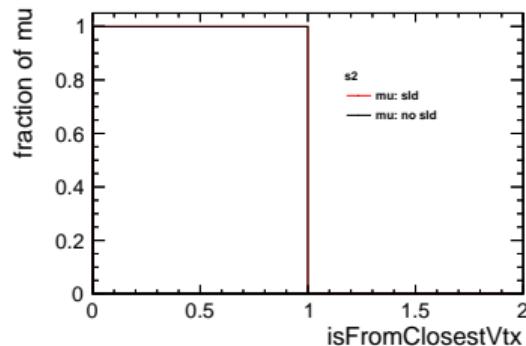
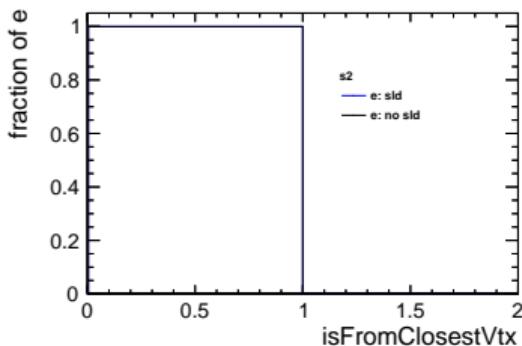
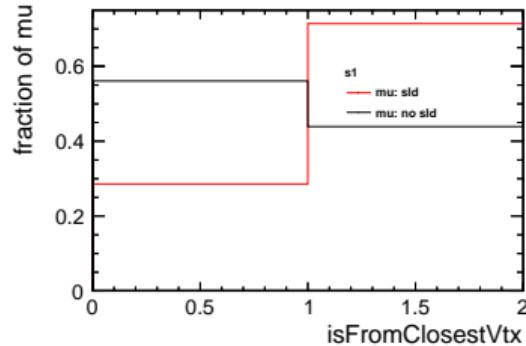
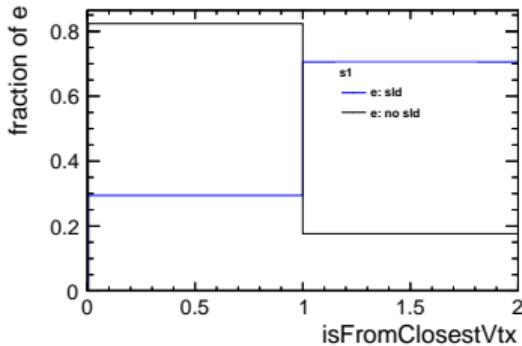
# SLD ID Inputs



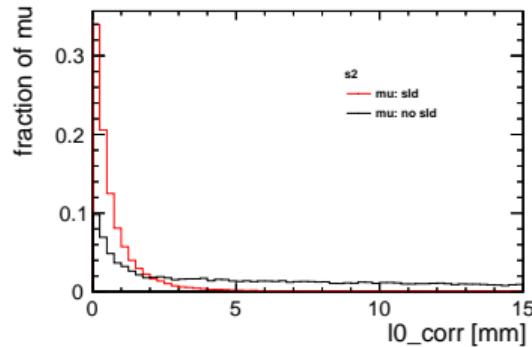
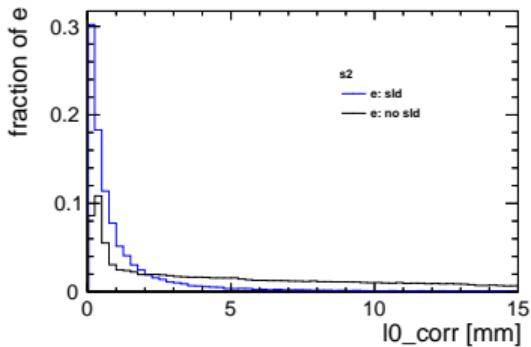
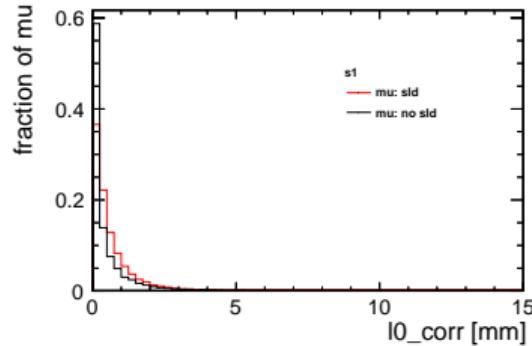
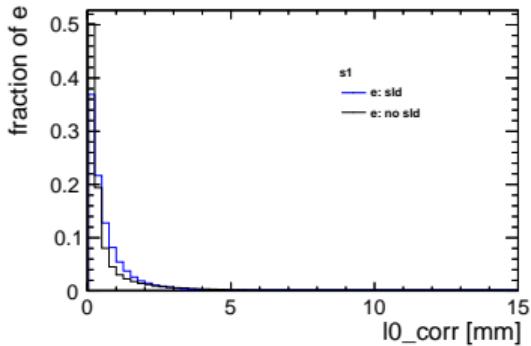
# SLD ID Inputs



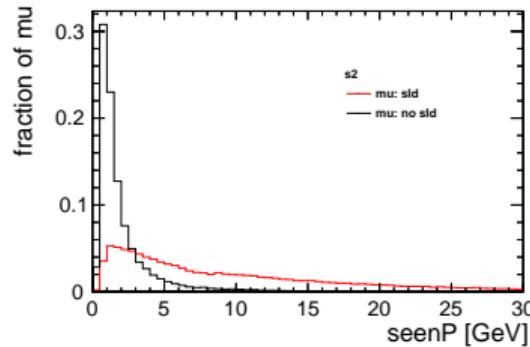
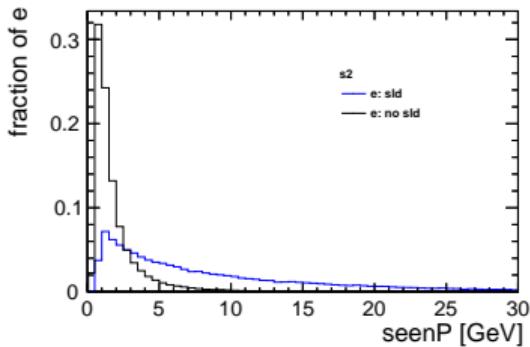
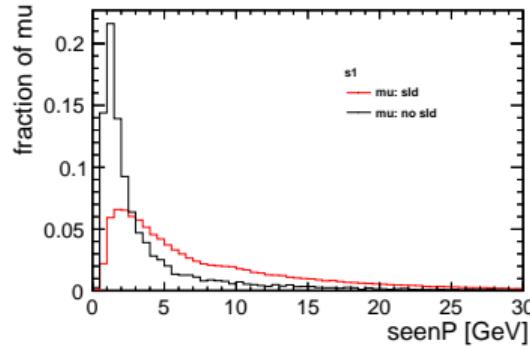
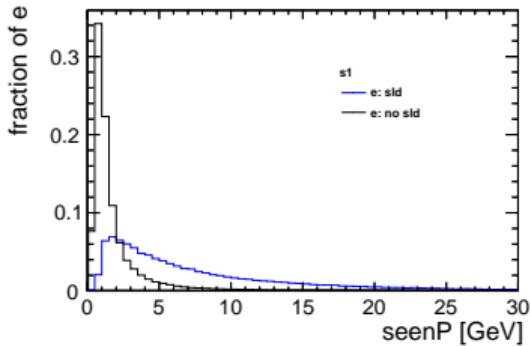
# SLD ID Inputs



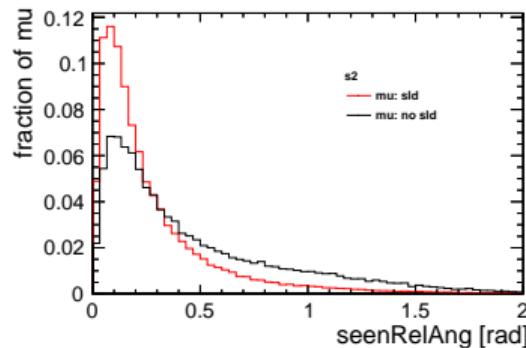
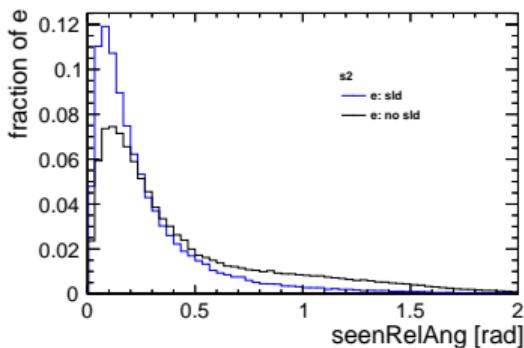
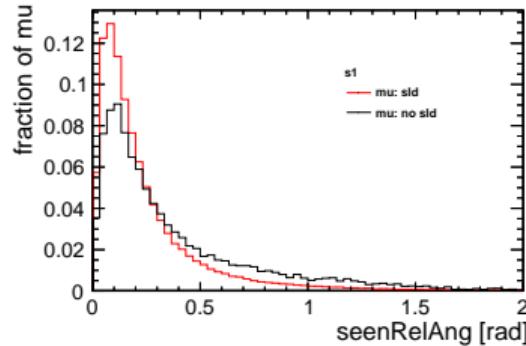
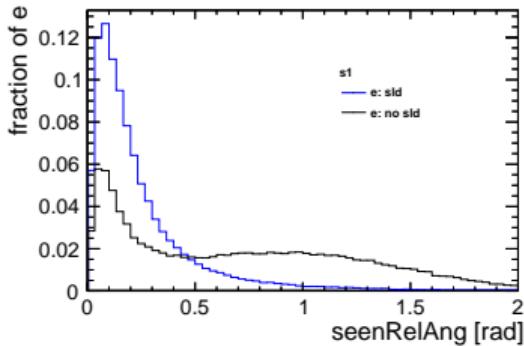
# SLD ID Inputs



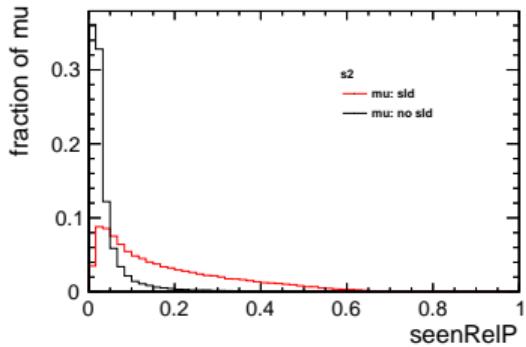
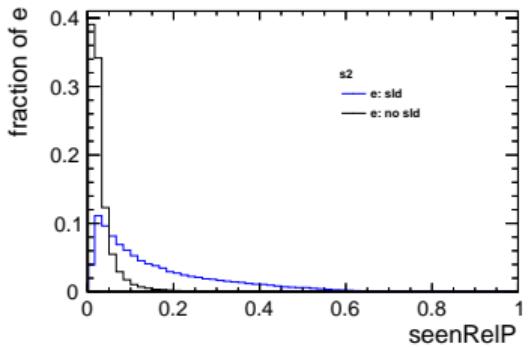
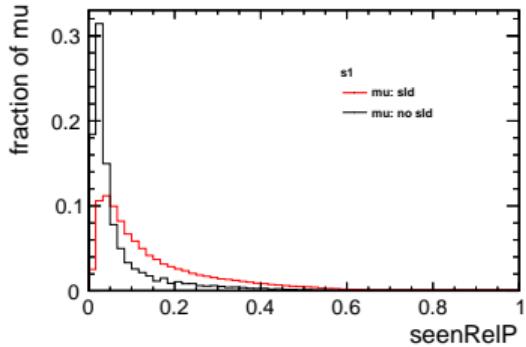
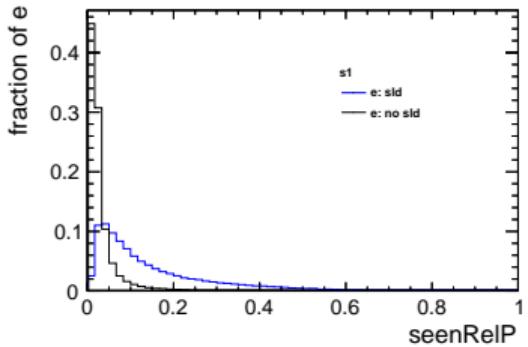
# SLD ID Inputs



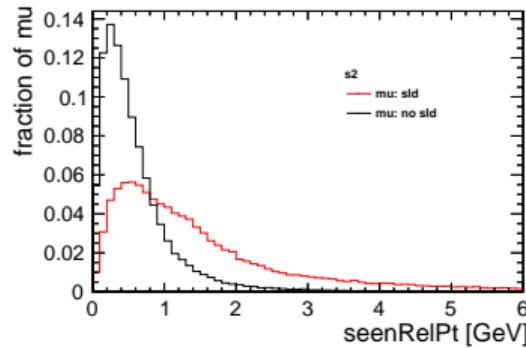
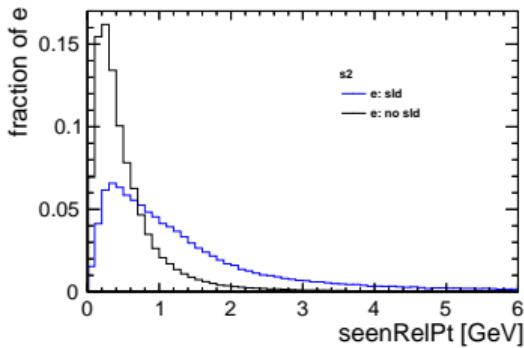
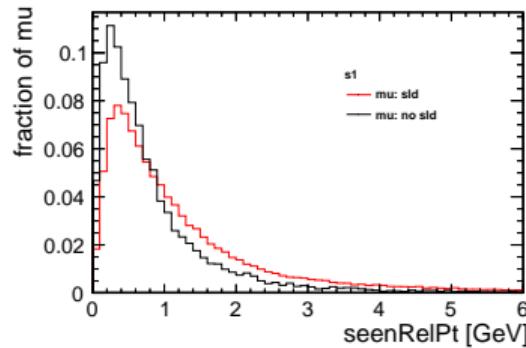
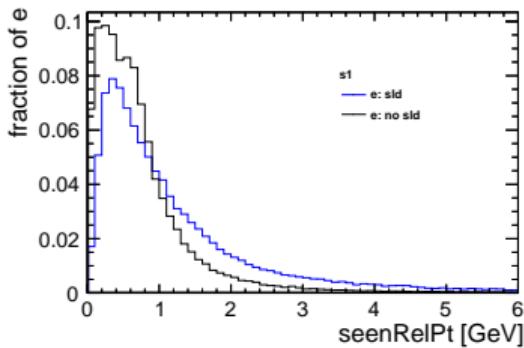
# SLD ID Inputs



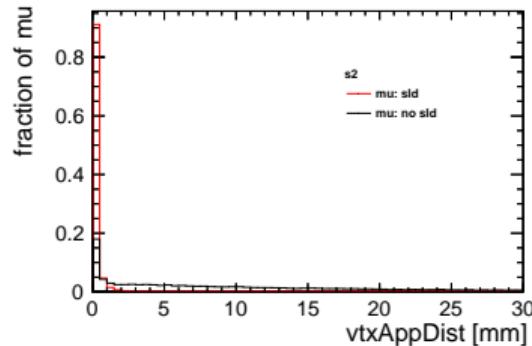
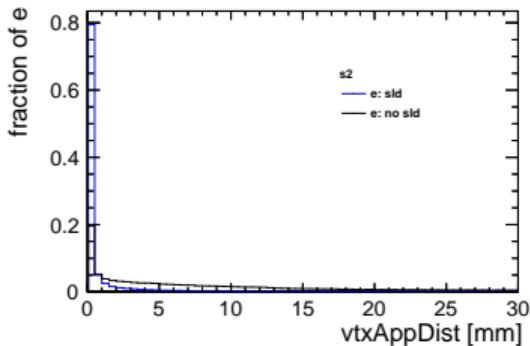
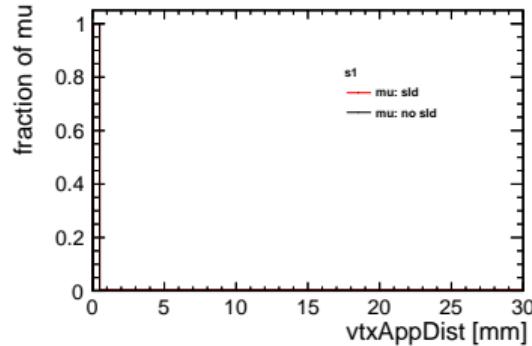
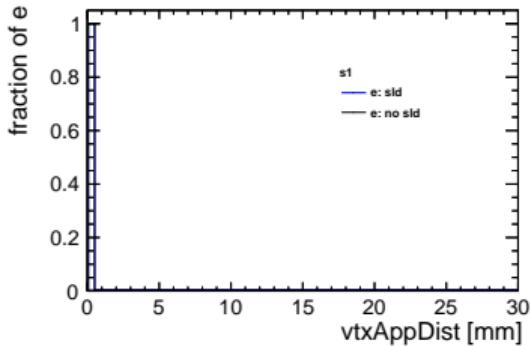
# SLD ID Inputs



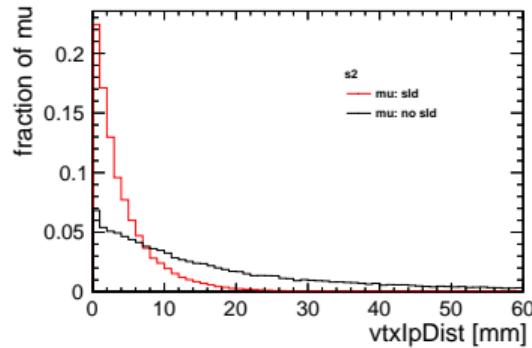
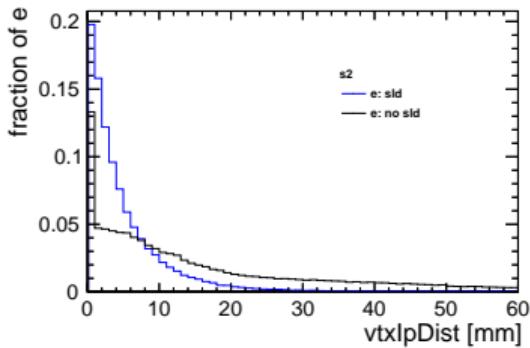
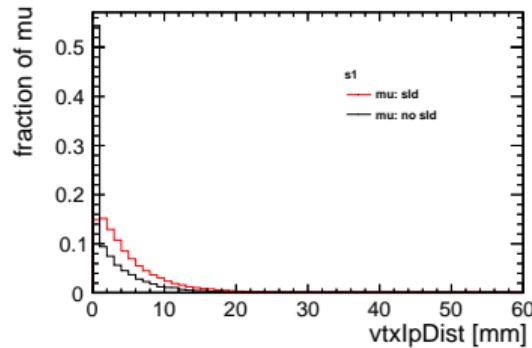
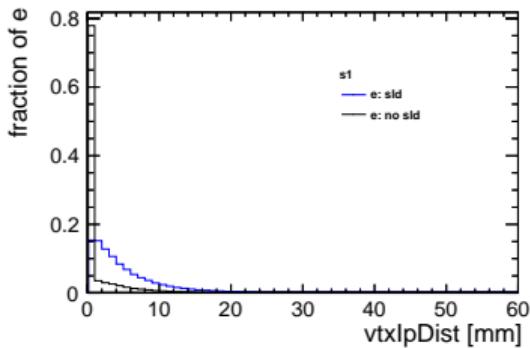
# SLD ID Inputs



# SLD ID Inputs



# SLD ID Inputs



# SLD ID Inputs

