MACHINE-LEARNING BASED IDENTIFICATION OF HIGHLY COLLIMATED ELECTRON PAIRS FROM BOOSTED Z-BOSON DECAYS

NETZWERK TEILCHENWELT

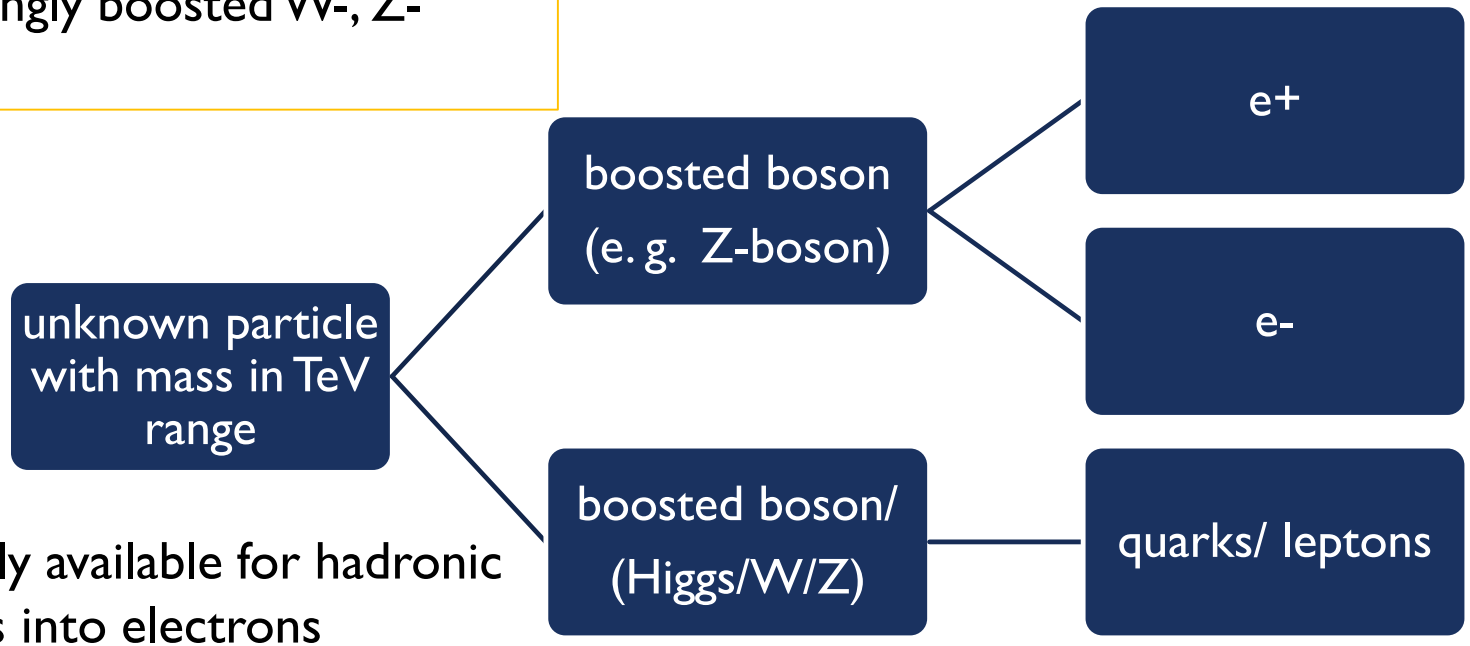
Max-Planck-Institut für Physik

ATLAS EXPERIMENT

CERN PROJECT WEEKS - FINAL PRESENTATION: 18.10.2019
SOPHIA VENERIS, RHABANUS-MAURUS-GYMNASIUM ST. OTTILIEN
SUPERVISOR: DR. DOMINIK DUDA, MPP MÜNCHEN

INTRODUCTION

Many bsm-theories (= **b**eyond the **s**tandard **m**odel – theories) predict the existence of high-mass particles which decay dominantely in strongly boosted W-, Z- and/or Higgs-bosons.



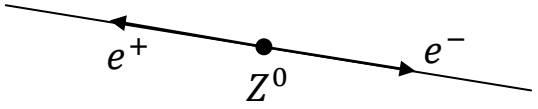
- standard ATLAS techniques only available for hadronic boson decays, but not for decays into electrons

BASICS
DECAY OF HIGH PT Z-BOSONS

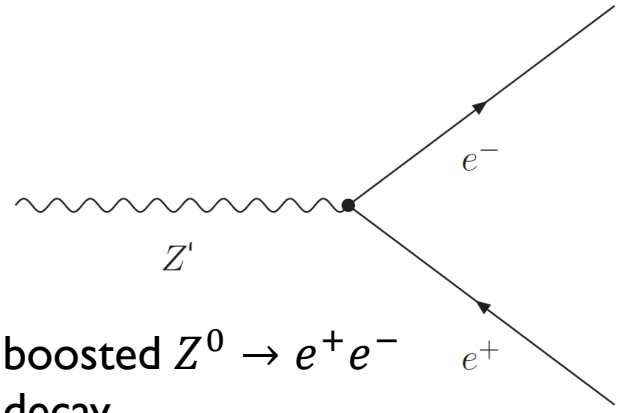
- pT = transverse momentum
- why are high pT Z-bosons so interesting to us?
  - Key to searches for new physics
  - Unique signature in the detector

energy and momentum conservation law:
energy and momentum cannot be lost

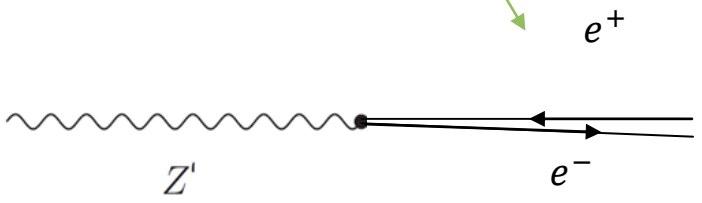
Highly collimated :
parallel and very close to each other



$Z^0 \rightarrow e^+e^-$ decay at rest (back-to-back)



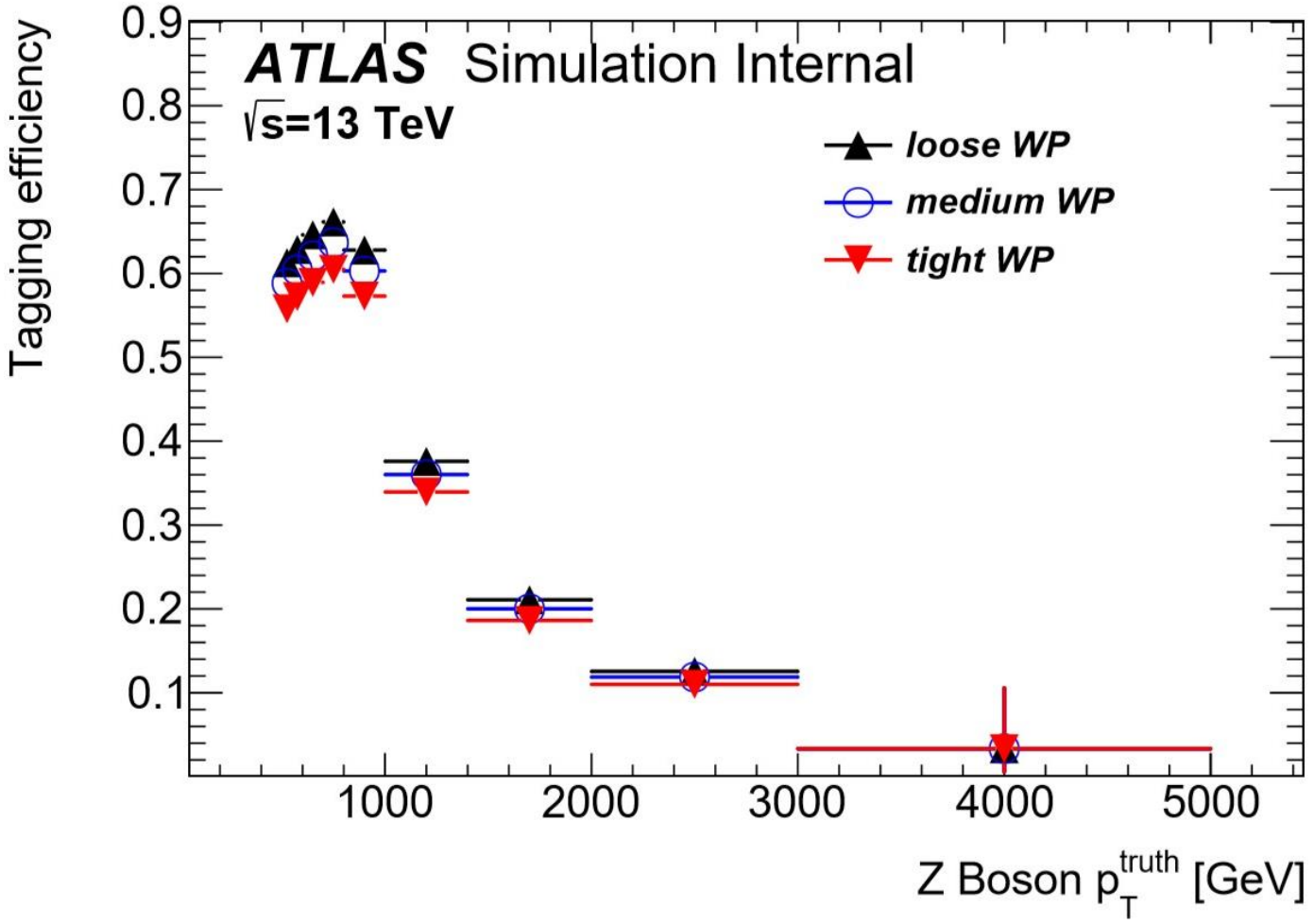
boosted $Z^0 \rightarrow e^+e^-$ decay



strongly boosted $Z^0 \rightarrow e^+e^-$ decay

BASICS
HIGHLY COLLIMATED ELECTRON PAIRS

- Efficiency to find both electrons with standard ATLAS reconstruction/identification techniques is strongly degraded for Z-bosons with a pT beyond 1 TeV



BASICS
THE ATLAS-EXPERIMENT

- multi-purpose particle detector
- for our analysis relevant:
  - the tracking system
    - Pixel detector
    - Silicon-strip detectors
    - TRT = transition radiation tracker
  - the electromagnetic calorimeter

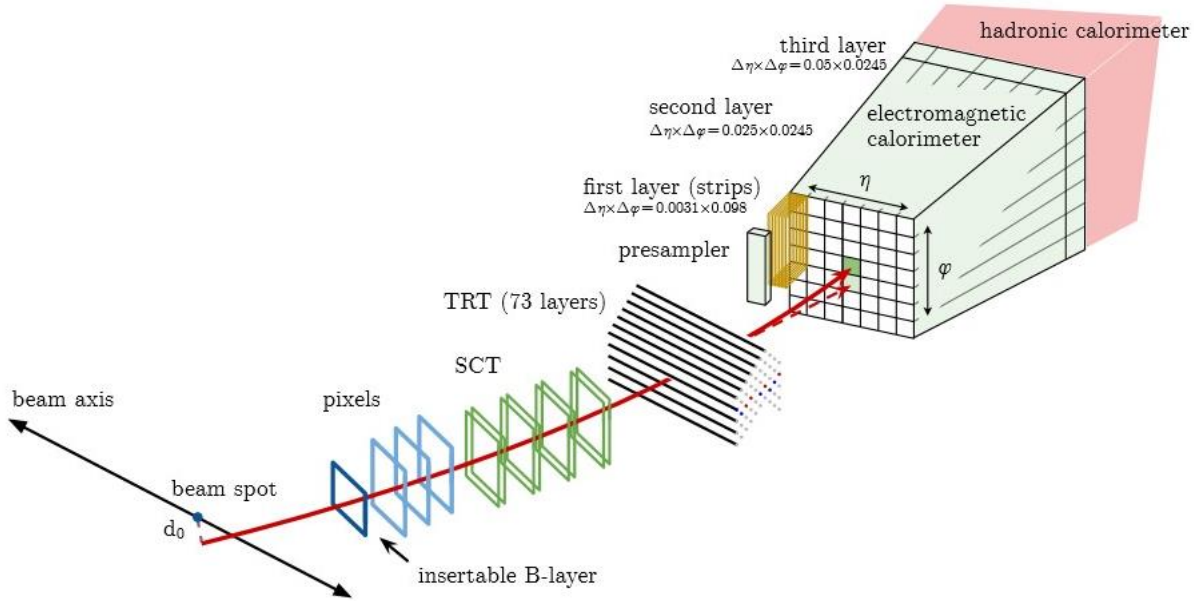
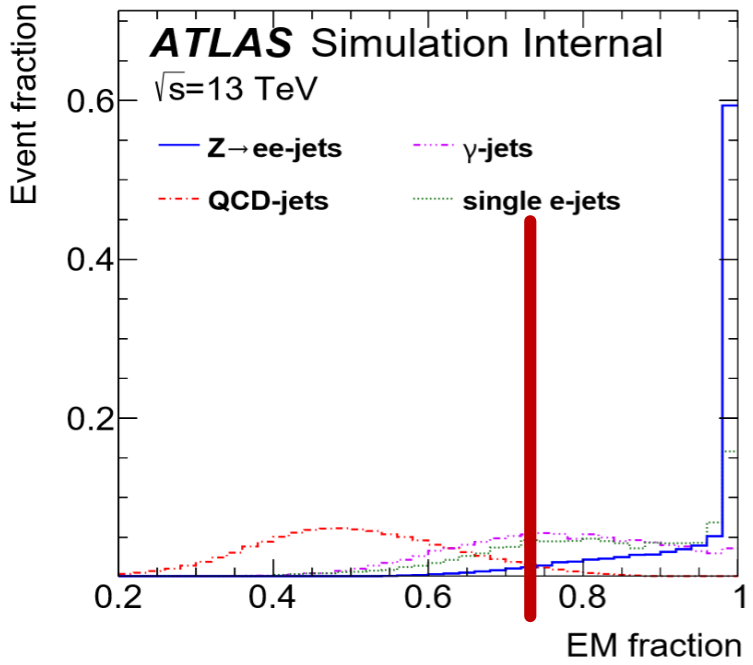


Figure 1: A schematic illustration of the path of an electron through the detector. The red trajectory shows the hypothetical path of an electron, which first traverses the tracking system (pixel detectors, then silicon-strip detectors and lastly the TRT) and then enters the electromagnetic calorimeter. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material in the tracking system.

BASICS MACHINE-LEARNING

- BDT = **B**oosted **D**ecision **T**ree
  - =>decision tree that weighs incorrect classifications in the next tree-modeling higher
    in order to improve the separation efficiency



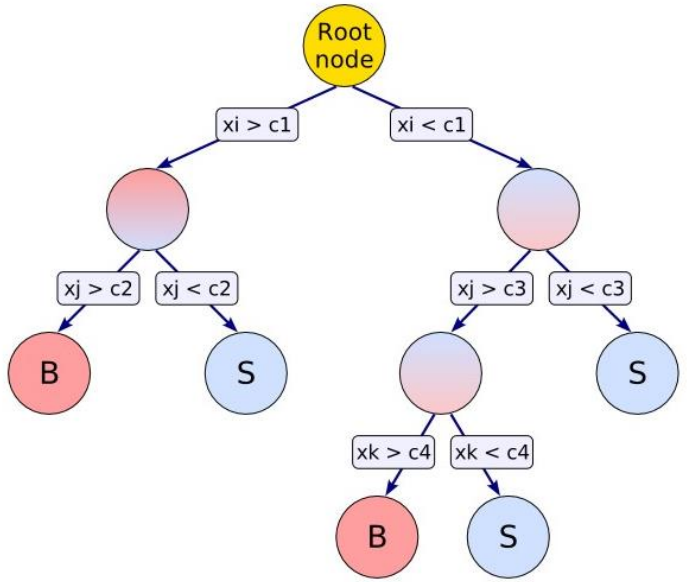
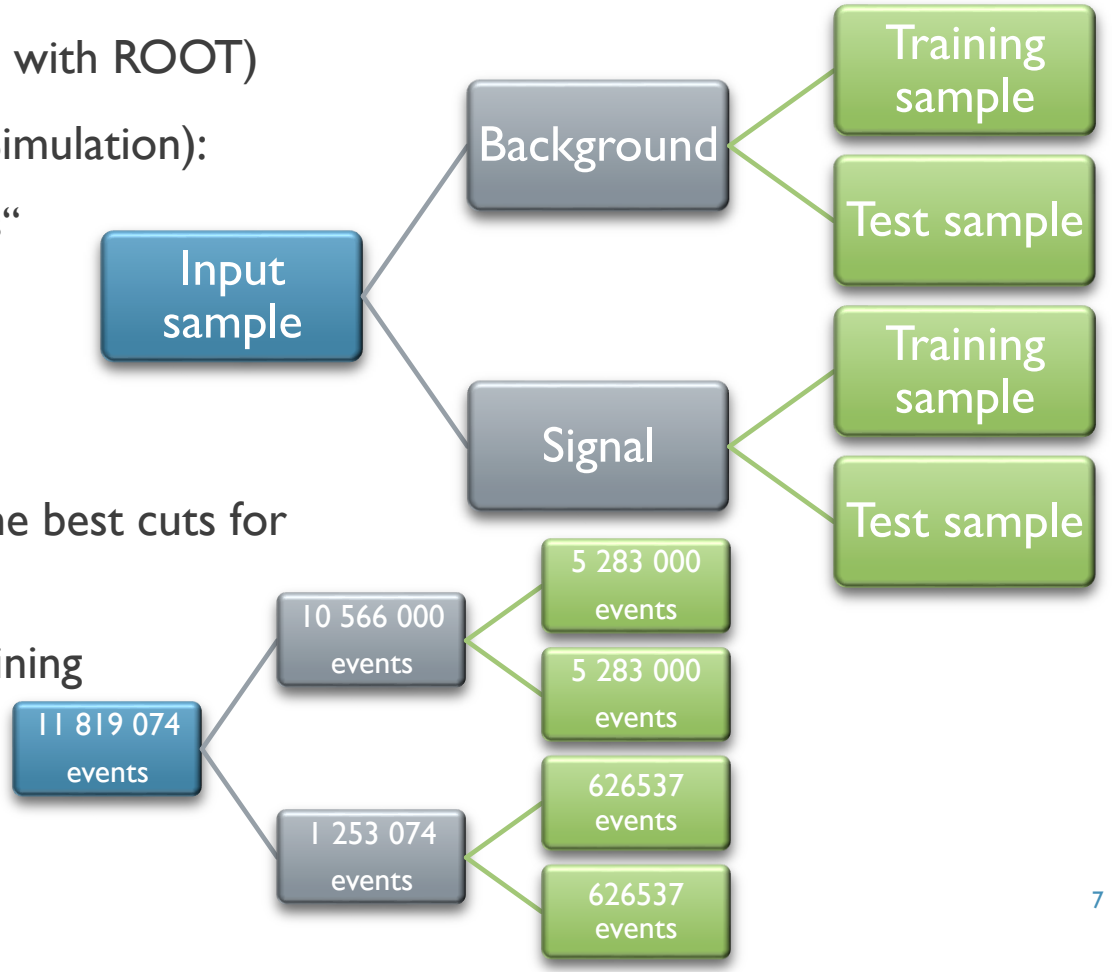


Figure 18: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is applied to the data. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled "S" for signal and "B" for background depending on the majority of events that end up in the respective nodes. For regression trees, the node splitting is performed on the variable that gives the maximum decrease in the average squared error when attributing a constant value of the target variable as output of the node, given by the average of the training events in the corresponding (leaf) node (see Sec. 8.13.3).

STRATEGY
TRAINING WITH TMVA

- TMVA (short for: **T**oolkit for **M**ulti**v**ariate Data **A**nalysis with ROOT)
- Training with two simulated data samples (Monte Carlo Simulation):
  - Monte Carlo simulation: based on the „law of large numbers“
  - Same amount of each, signal or background, in training and test sample



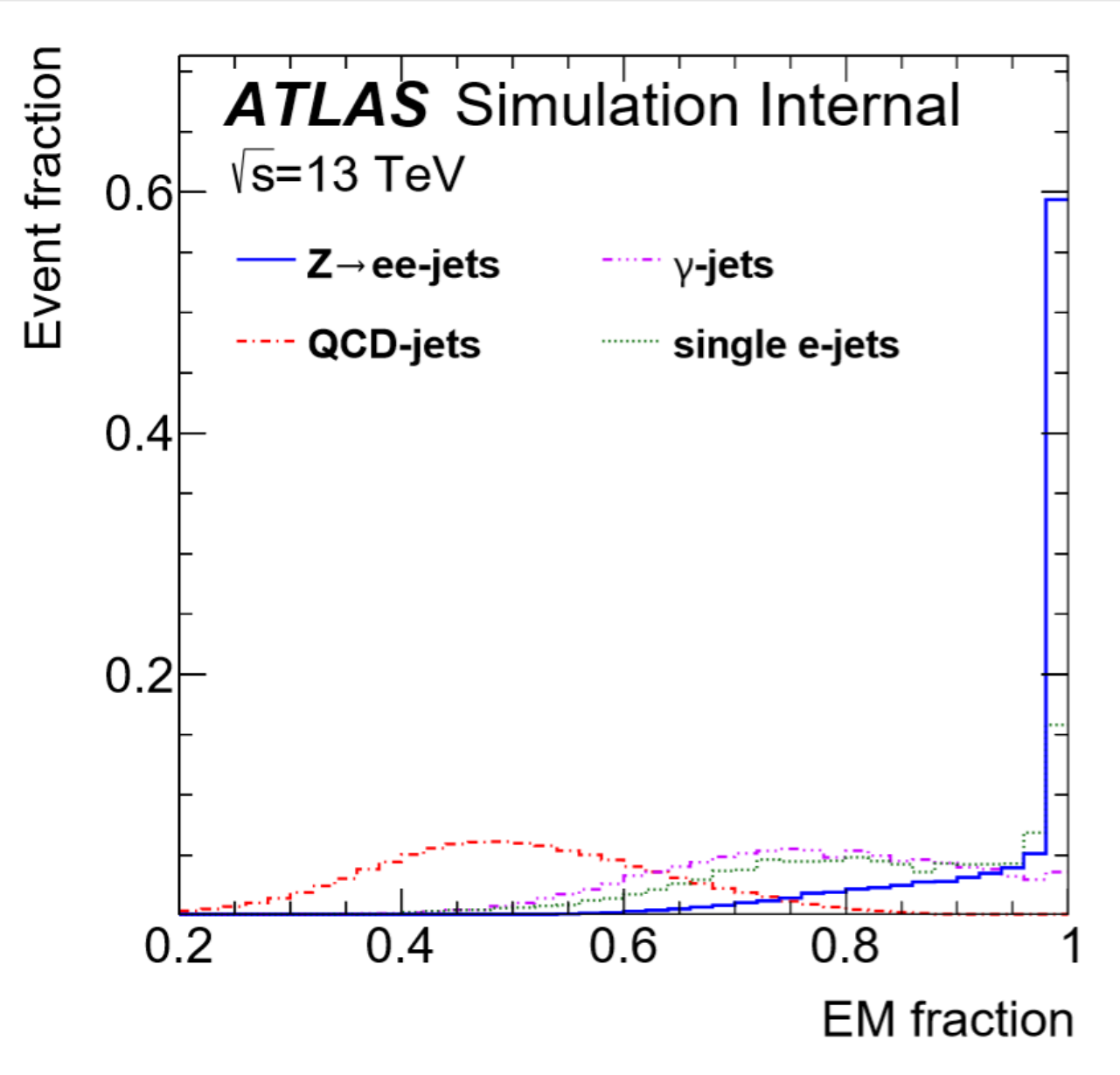
- (1.) the program trains with the training sample to find the best cuts for separating signal from background
- (2.) the test sample is used to probe for possible overtraining

```
                       : ---------------------------------------------------------------------------------------
                       :
                       : Testing efficiency compared to training efficiency (overtraining check)
                       : ---------------------------------------------------------------------------------------
                       : DataSet              MVA              Signal efficiency: from test sample (from training sample)
                       : Name:                Method:          @B=0.01             @B=0.10            @B=0.30
                       : ---------------------------------------------------------------------------------------
                       : dataset              SelectBTracks  : 0.988 (0.988)       1.000 (1.000)      1.000 (1.000)
                       : ---------------------------------------------------------------------------------------
                       :
Dataset:dataset        : Created tree 'TestTree' with 816190 events
                       :
Dataset:dataset        : Created tree 'TrainTree' with 816190 events
                       :
Factory                : Thank you for using TMVA!
                       : For citation information, please visit: http://tmva.sf.net/citeTMVA.html
Finished Training !!!
```

= ?

eff. test sample = (eff. training sample)
⇒ (nearly) no overtraining

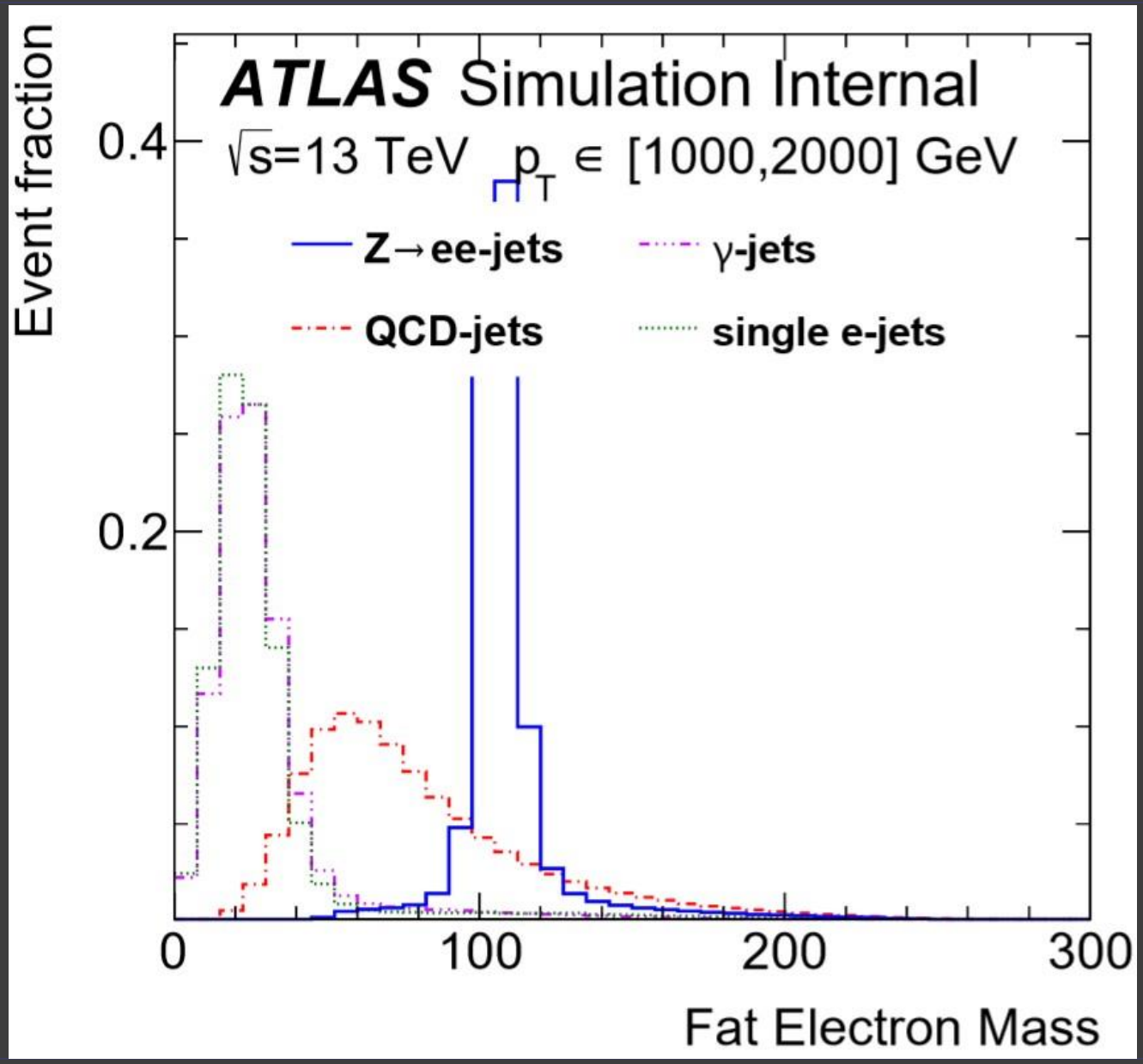
eff. test sample ≠ (eff. training sample)
⇒ indicates statistical fluctuations
⇒ overtraining
⇒ more input needed



EM Fraction Fat Electrons

- Electron jets:
- how much of the jets total energy was located in the electromagnetic calorimeter?
- very good discrimination betweeen signal and background possible

Fat Electron = *dielectron candidate jet*



Fat Electron Mass:

as measured by the electromgnetic calorimeter system

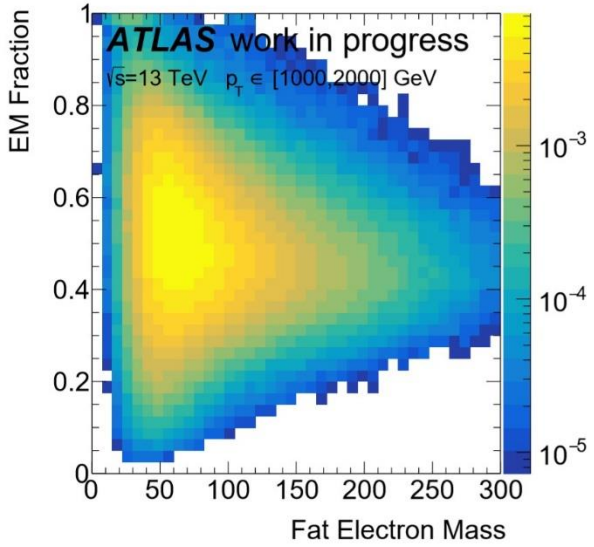
Fat Electron = *dielectron candidate jet*

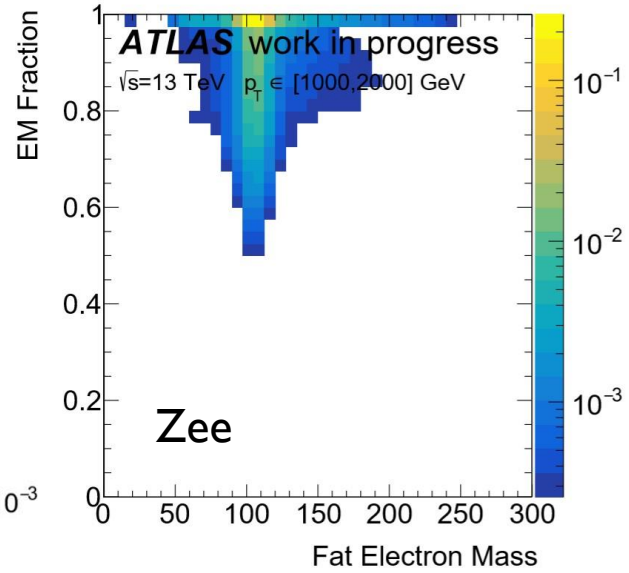
variable	definition
FatElectron_M	Mass of a Fat Electron
FatElectron_ntrks	Number of tracks
FatElectron_EMFrac	Fraction of the jet´s total energy which is located in the electromagnetic calorimeter
FatElectron_balance	$\frac{(\text{trk1 pT} - \text{trk2 pT})}{(\text{trk1 pT} + \text{trk2 pT})}$
Trk1_dRToJet	radial distance between the electron candidate track and the jet axis (highest pT track)
Trk2_dRToJet	radial distance between the eletron candidate track and the jet axis (second highest pT track)
dR_tt	radial distance between the highest and second highest pT tracks

variable	definition
Trk1_EOverP	$\frac{energy\ in\ electromagnetic\ calorimeter\ cluster}{momentum\ of\ the\ track}$ (highest pT track)
Trk2_EOverP	$\frac{energy\ in\ electromagnetic\ calorimeter\ cluster}{momentum\ of\ the\ track}$ (second highest pT track)
Trk1_dPhi	dPhi (distance in Phi-direction) between track and cluster (highest pT track)
Trk2_dPhi	dPhi (distance in Phi-direction) between track and cluster (second highest pT track)
Trk1_dEta	dEta (distance in Eta-direction) between track and cluster (highest pT track)
Trk2_dEta	dEta (distance in Eta-direction) between track and cluster (second highest pT track)

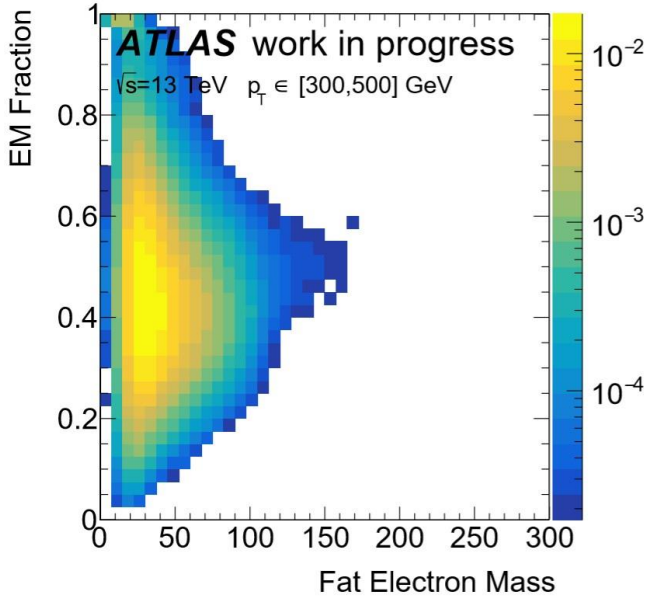
2DCOMPARISON FATELECTRON_M VS. FATELECTRON_EMFRAC

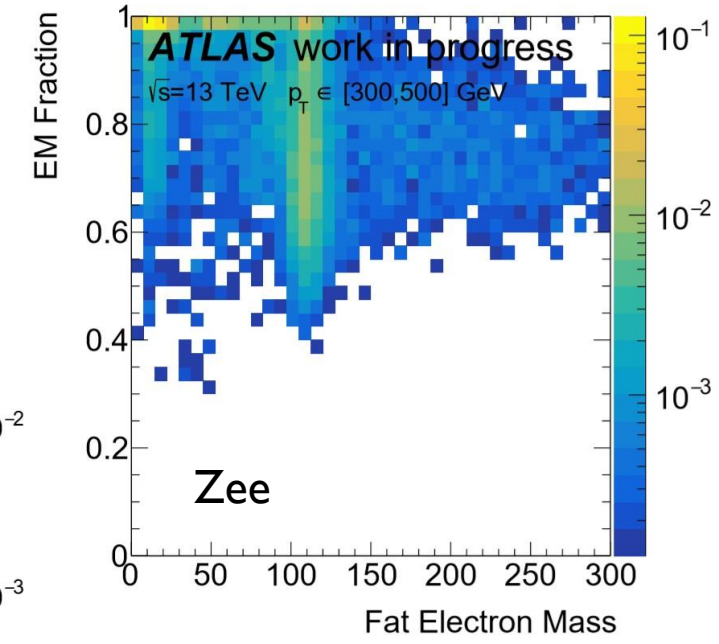
QCDjets





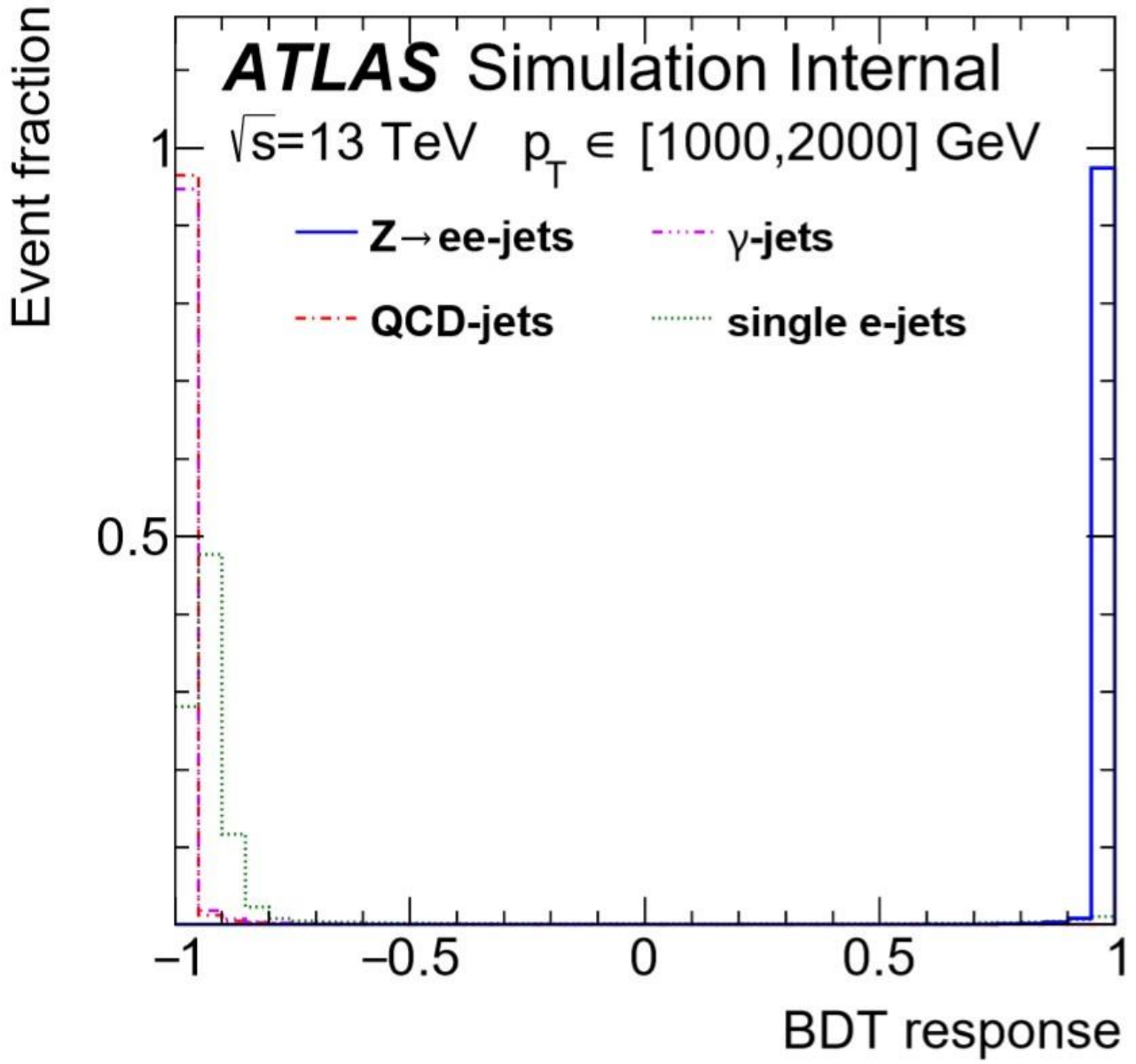
QCDjets





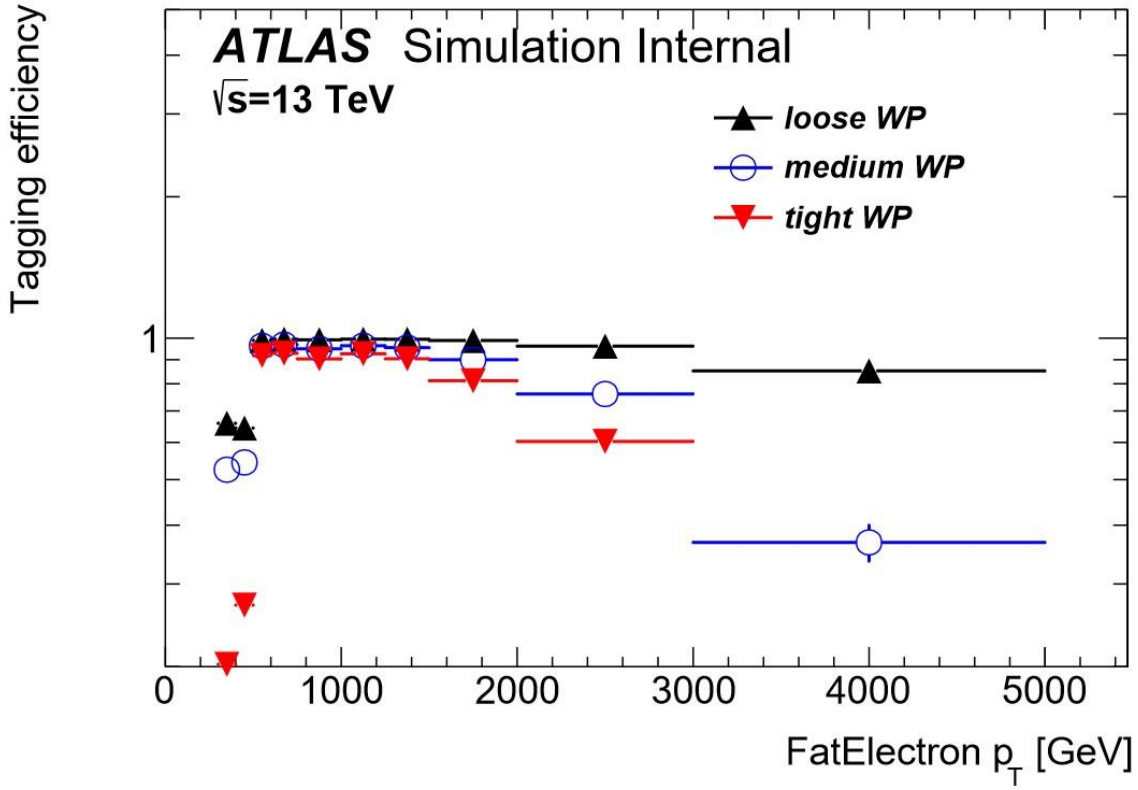
BDT RESPONSE

- Working points:
  - Loose: 99% efficient
  - Medium: 95% efficient
  - Tight: 90% efficient

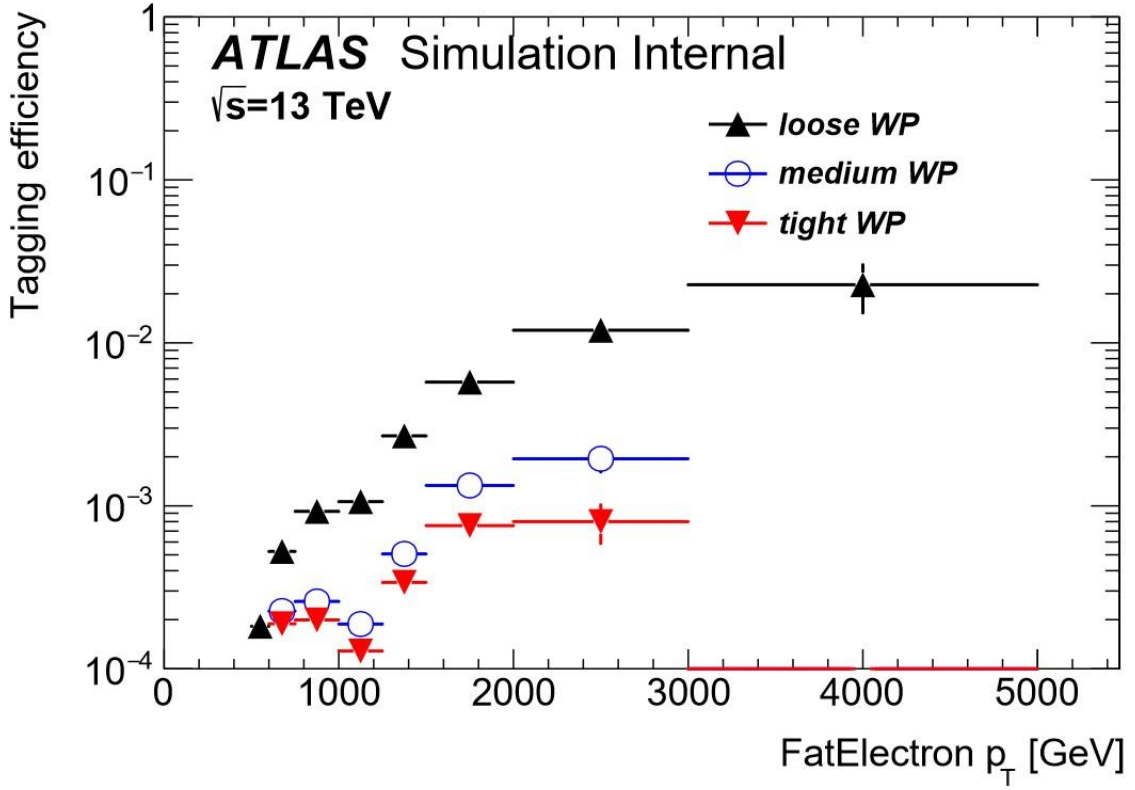


RESULTS

Zee



QCDjets



$$\text{Efficiency} = \frac{number\ of\ objects\ with\ required\ properties}{number\ of\ total\ objects}$$

3. CONCLUSION AND OUTLOOK

- Studied several quantities for their ability to separate signal from background
- the results improve the identification of Z -> ee substantially
- Outstanding tasks:
  - Test, if decorrelation with mass is needed
  - pT- and Eta- dependence must be taken into account during the training

=> Aiming to establish this technique eventually as a common tool in ATLAS

THANK YOU!

to

- my supervisor: Dr. Dominik Duda (Max-Planck-Institut für Physik, München)
- the organizer: Sebastian Fabianski (Netzwerk Teilchenwelt)
- Prof. Dr. Siegfried Bethke (Max-Planck-Institut für Physik, München)
- Prof. Dr. Andrzej Buras (TU München)
- Dr. Sandra Kortner (Max-Planck-Institut für Physik, München)
- Barbara Wankerl (Netzwerk Teilchenwelt / Max-Planck-Institut für Physik, München)
- Dr. Pater Timotheus Bosch OSB (Rhabanus-Maurus-Gymnasium St. Ottilien)
- Markus Schnell (Rhabanus-Maurus-Gymnasium St. Ottilien)
- my school´s headmasters (Michael Häußinger, Werner Hörmann : Rhabanus-Maurus-Gymnasium St. Ottilien)
- … and to all other scientists, teachers, classmates and friends who supported me in and during this project!

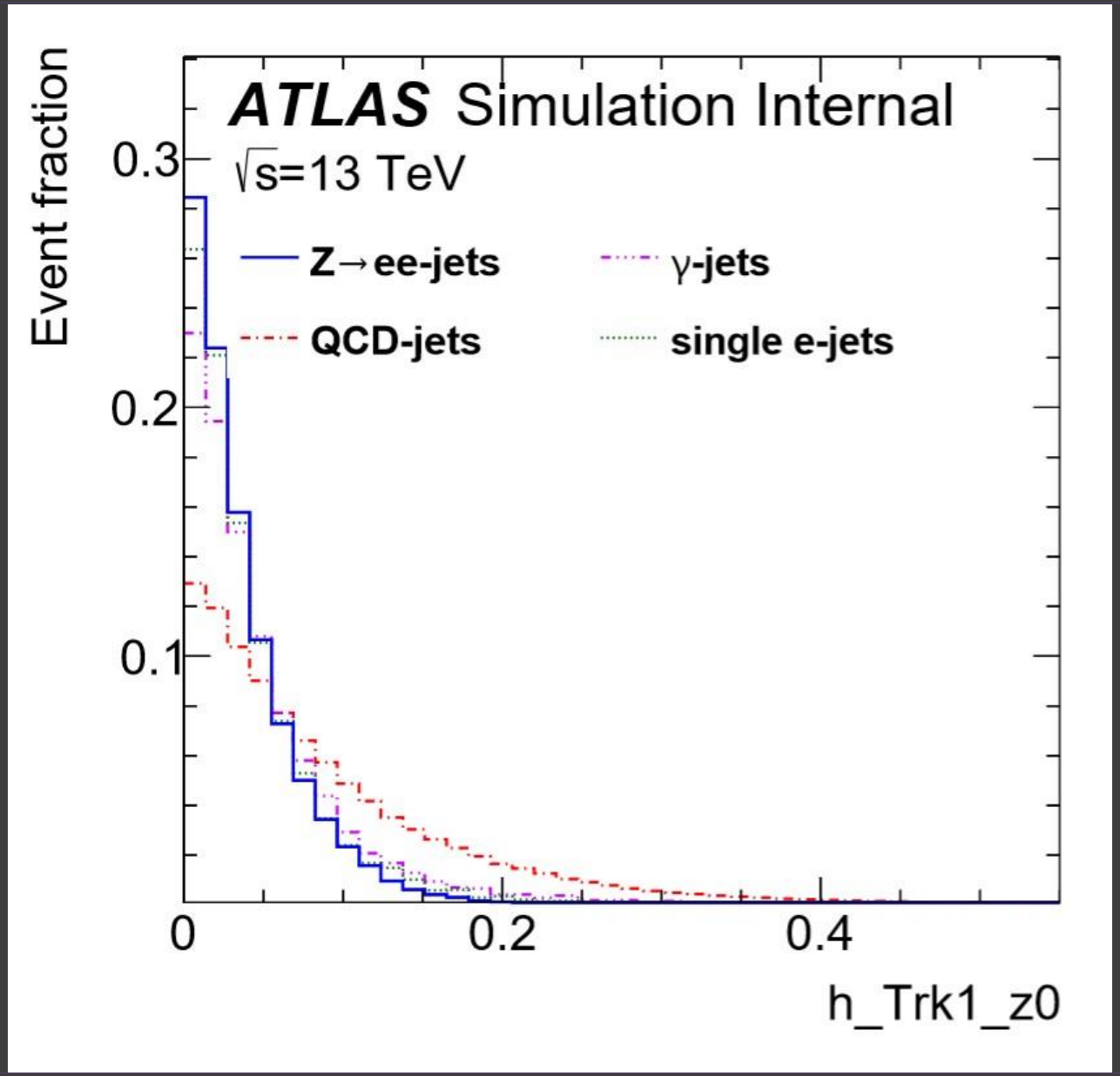
4. REFERENCES

- Graphics:
  - https://home.cern/sites/home.web.cern.ch/files/logo/cern-logo.png
  - http://atlasexperiment.org/photos/logos-2015-png.html
  - https://www.mpp.mpg.de/fileadmin/config/themes/mpp/images/mpp-mpg-logo.png
  - http://www.physik.uni-regensburg.de/service/schulen/nwt_logo.jpg
  - https://atlas.physicsmasterclasses.org/zpath_files/img/highslide/feynman/Zprime_ElectronPositron.png
- Content:
  - http://tmva.sourceforge.net/
  - https://arxiv.org/pdf/1902.04655.pdf
  - https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf

BACKUP SLIDES

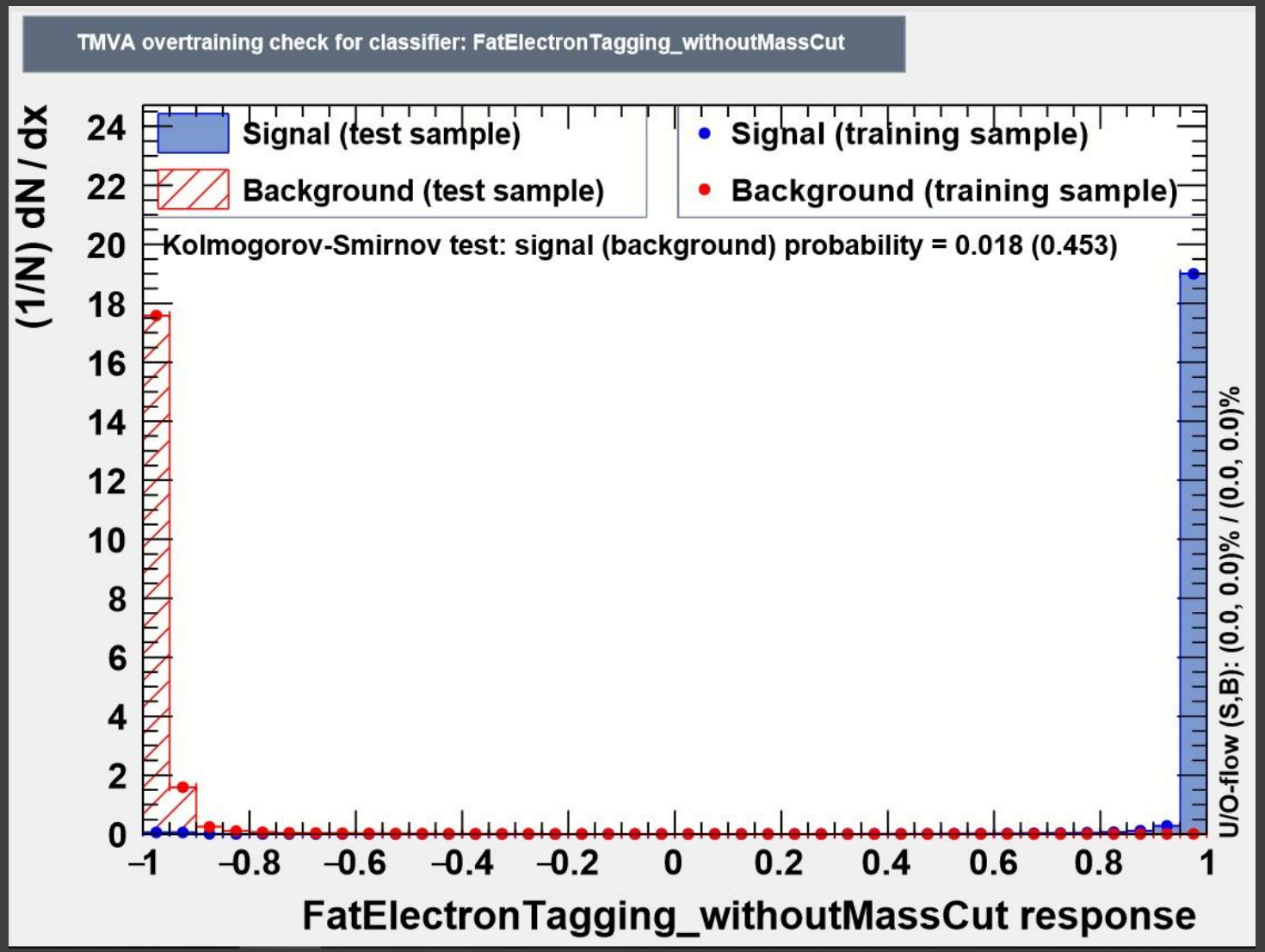
HISTOGRAMS

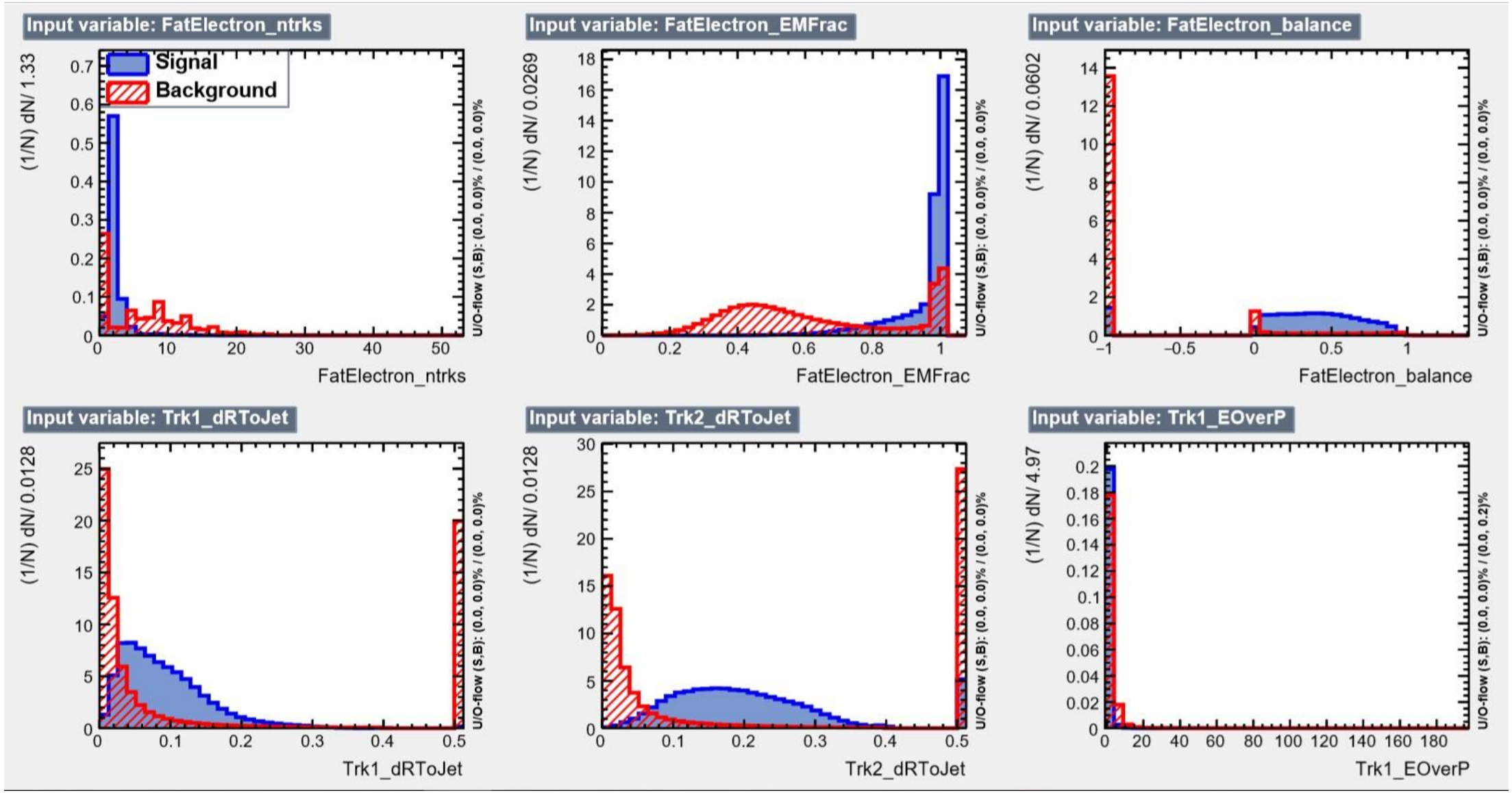
- plot making with python
- 1-component histograms:
  - A variable along the x-axis
  - the probability distribution is displayed on the y-axis
- 2-component histograms:
  - A variable along the x-axis
  - Another variable along the y-axis
  - The probability distribution is shown by different colours (colour range)

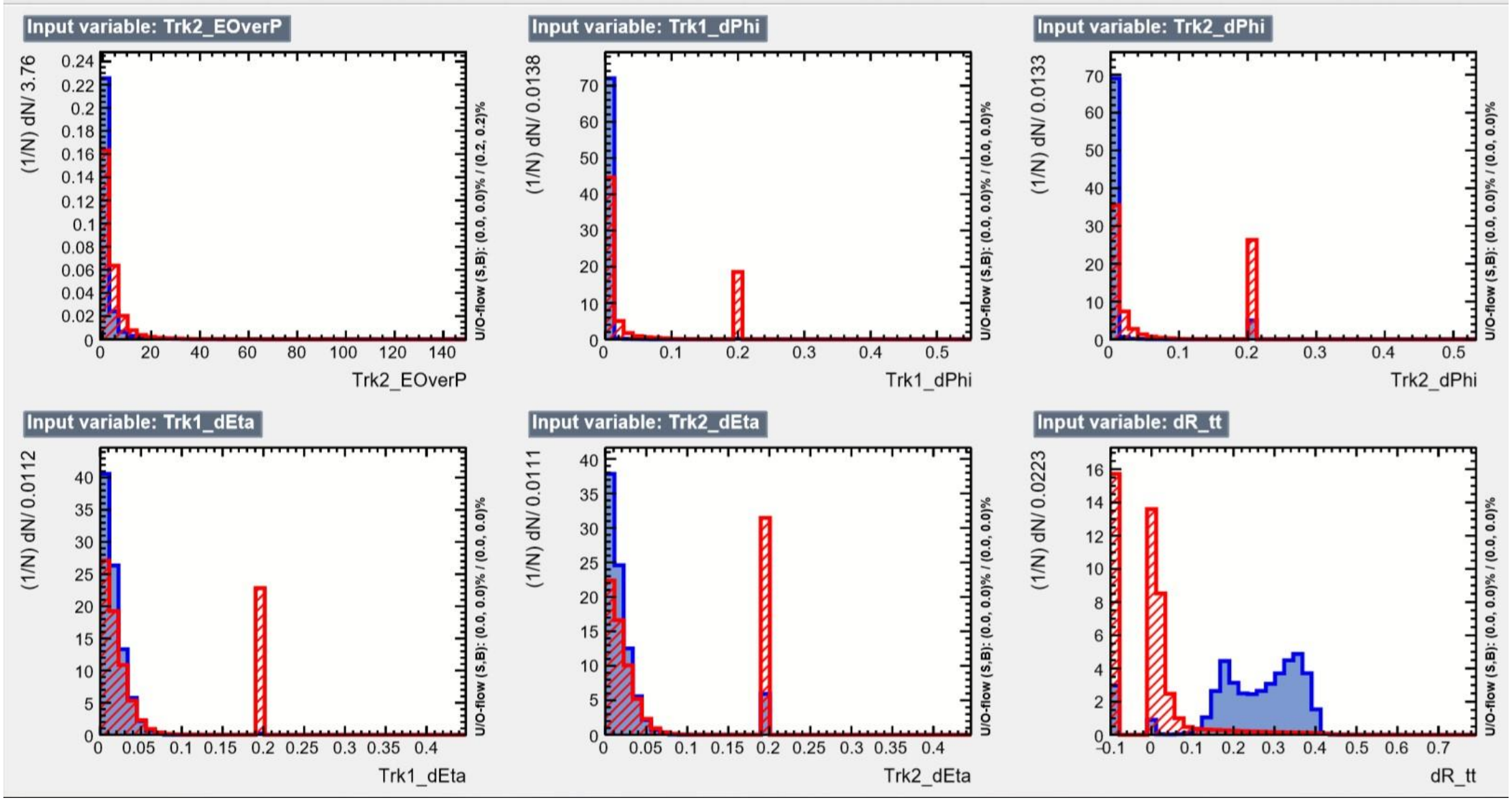

ATLAS Simulation Internal
√s=13 TeV
Z→ee-jets
γ-jets
QCD-jets
single e-jets
Event fraction
0.3
0.2
0.1
0
0.2
0.4
h_Trk1_z0

h_Trk1_z0:

Overtraining Check

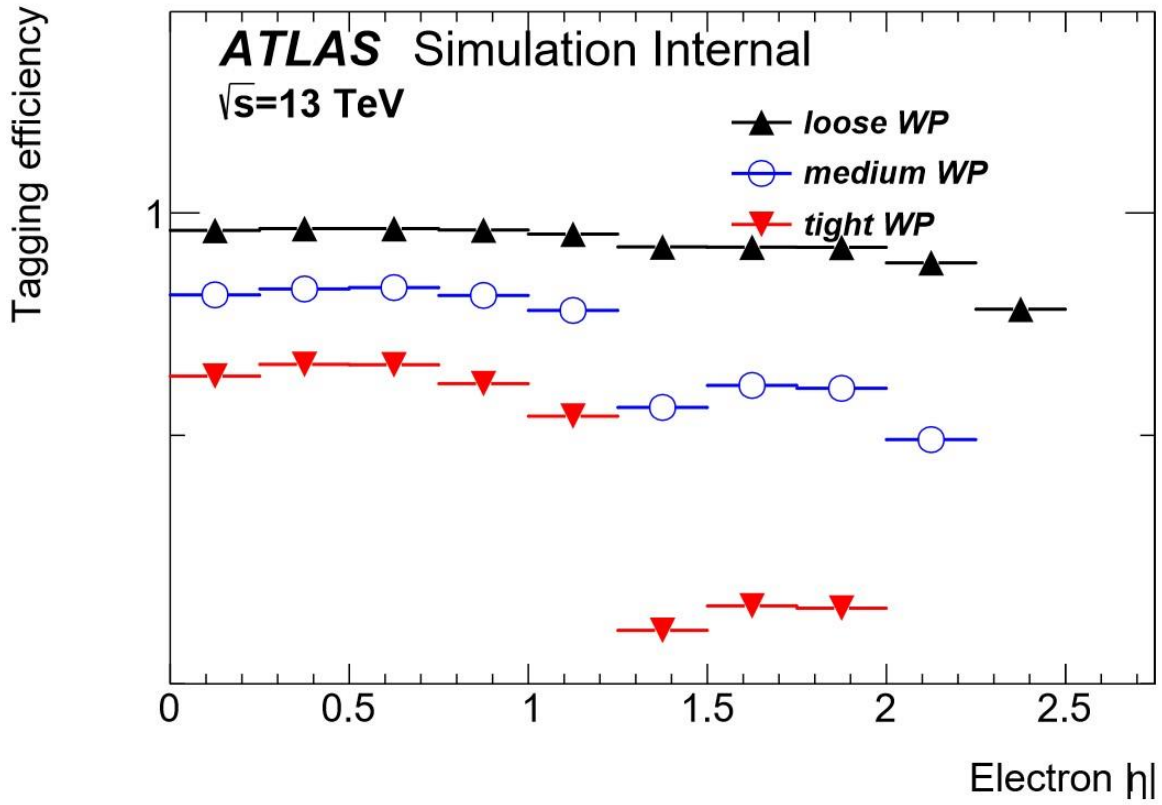



Input variable: FatElectron_ntrks
Signal
Background
(1/N) dN/ 1.33
FatElectron_ntrks
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: FatElectron_EMFrac
(1/N) dN/ 0.0269
FatElectron_EMFrac
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: FatElectron_balance
(1/N) dN/ 0.0602
FatElectron_balance
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: Trk1_dRToJet
(1/N) dN/ 0.0128
Trk1_dRToJet
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: Trk2_dRToJet
(1/N) dN/ 0.0128
Trk2_dRToJet
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: Trk1_EOverP
(1/N) dN/ 4.97
Trk1_EOverP
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.2)%


Input variable: Trk2_EOverP
(1/N) dN/ 3.76
Trk2_EOverP
U/O-flow (S,B): (0.0, 0.0)% / (0.2, 0.2)%
Input variable: Trk1_dPhi
(1/N) dN/ 0.0138
Trk1_dPhi
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: Trk2_dPhi
(1/N) dN/ 0.0133
Trk2_dPhi
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: Trk1_dEta
(1/N) dN/ 0.0112
Trk1_dEta
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: Trk2_dEta
(1/N) dN/ 0.0111
Trk2_dEta
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: dR_tt
(1/N) dN/ 0.0223
dR_tt
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

T-EFFICIENCY PLOTS: ETA

Zee



single ejets