(Domain-Adversarial) Neural Network classifiers of *b*-jets origins in top quark pair productions (*ttbar*) with additional *b*-jets in the final state

Master Thesis Defense by **Ning Ni**

Professor Supervisor: **Dmitri Tsybychev**
Postdoc Advisor: **Egor Antipov**

Outline

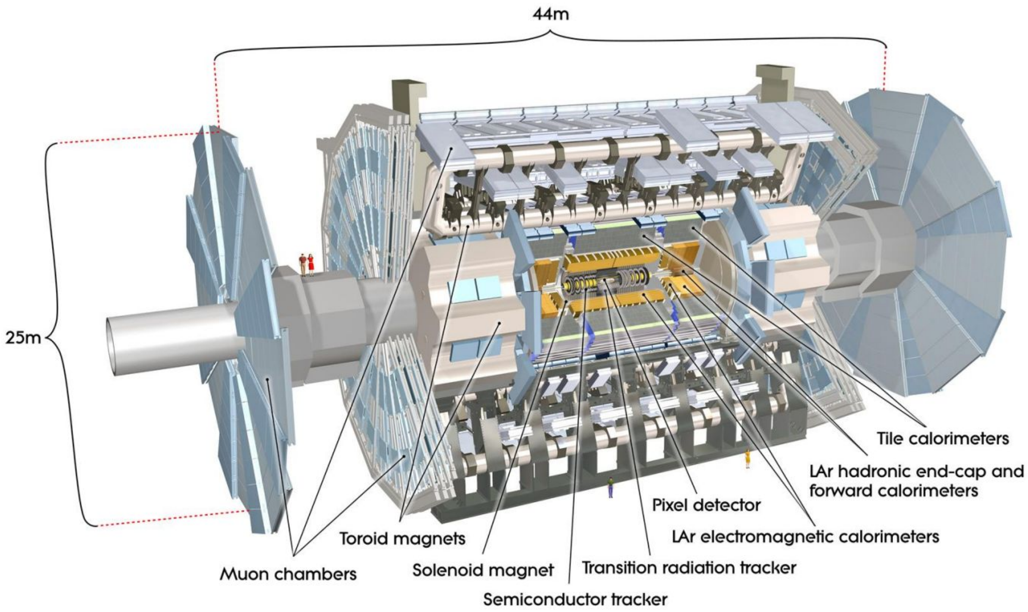
Background Introduction

- The ATLAS detector
- Motivation to analyze top quark pair production

Thesis Work

- Event Selection
- Reconstructed discriminating features
- Construction of neural network (NN)
- Analyzed p_T bias and Compared with Boosted Decision Tree (BDT)
- Exploration of Domain-Adversarial Neural Network (DANN)

The ATLAS Detector

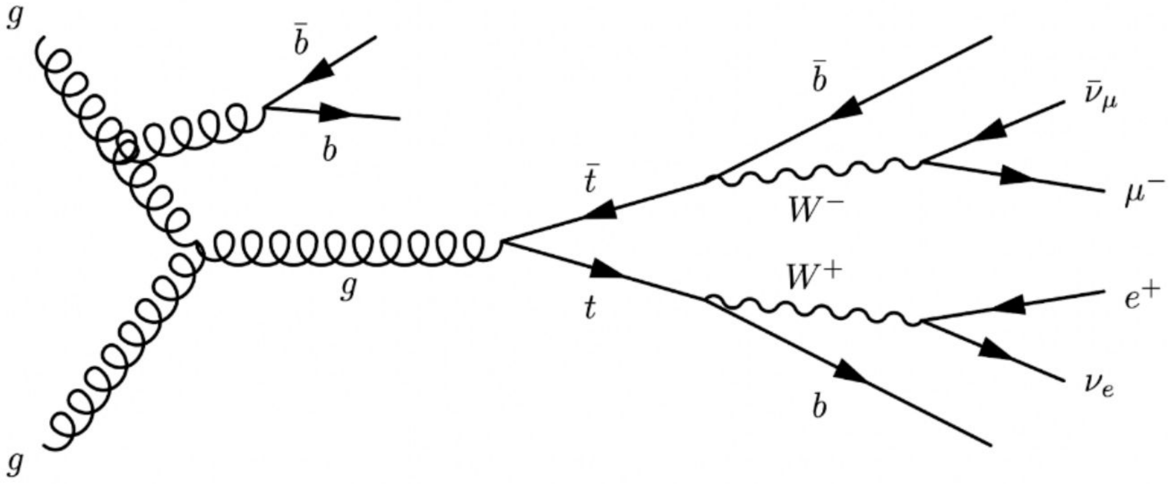


- *pp* collision in Run2 has $\sqrt{s}$ = 13 TeV.
- Integrated luminosity: ~140 fb^{-1}.
- Major systems:

1. Inner Detector
2. Calorimeters
3. Muon Spectrometer
4. Magnet system

Motivation

- *ttbar* production in association with additional *b*-jets is the major background for many BSM searches (SUSY) as well as to rare SM processes (*ttH*, 4-tops …).
- Measurement of QCD radiation jets produced in *ttbar* is crucial for tuning MC generators parameters.
- Theoretical prediction of ttbar + *b*-jets final state is highly non-trivial.



Event Selection

ATLAS data
Run2

MC simulation
(Powheg+Pythia8)

Apply cuts:

1. One electron and one muon with opposite charges
2. Lepton p_T at least 28 GeV
3. Jets p_T at least 25 GeV
4. At least 3 *b*-jets

Selected data

Z+jets

Other

Diboson

ttbar+H

ttbar+V

single top

ttbar

p_T : Transverse Momentum

Removal of Overlap

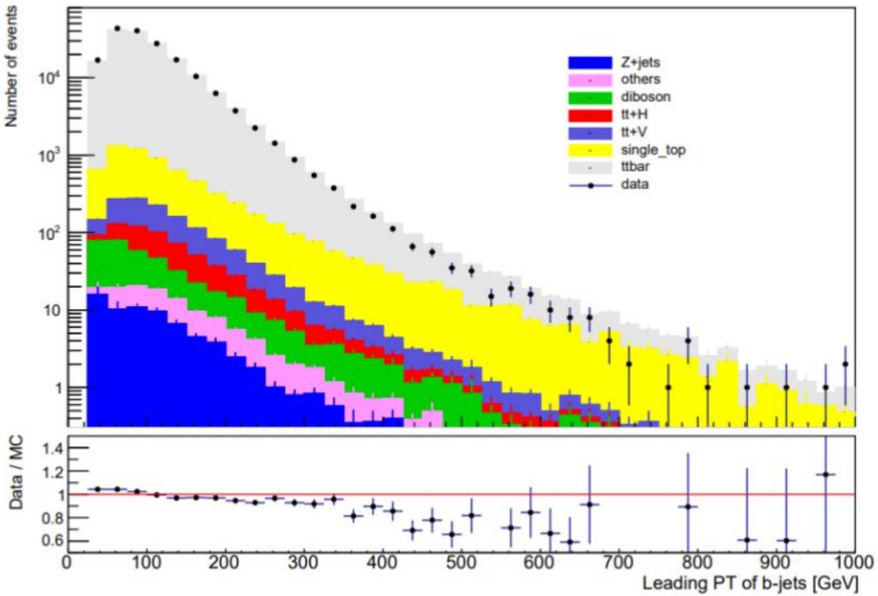
Use Heavy Flavor Filter Flag (HFFF) to avoid overlap of inclusive sample

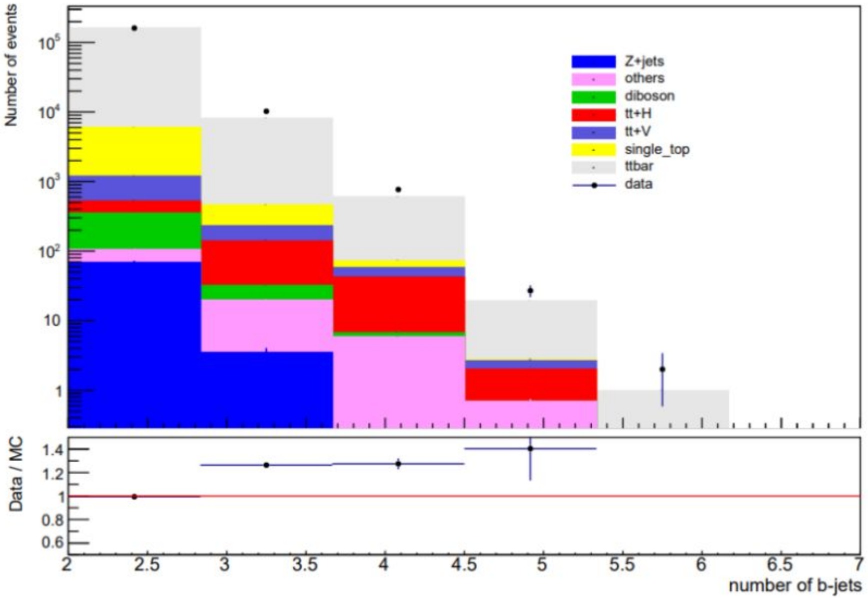
Sample	DID number	Only keep when HFFF is
$t\bar{t}$ inclusive	410472	0
$t\bar{t} + b\bar{b}$	411076	1
$t\bar{t} + b$	411077	2
$t\bar{t} + c$	411078	3

Reweighted event yield from each source after event selection.

Source	2015+2016	2017	2018
Data	44615	55004	72485
Total MC	44586	55789	70277
$t\bar{t}$	42994	53556	67601
single top	1261	1627	2135
$t\bar{t} + V$	201	253	324
$t\bar{t} + H$	81	99	131
diboson	17	212	30
Z+jets	17	24	31
others	16	19	25

Monte Carlo (MC) Modeling





Reconstructed Discriminating Variables

Input Features

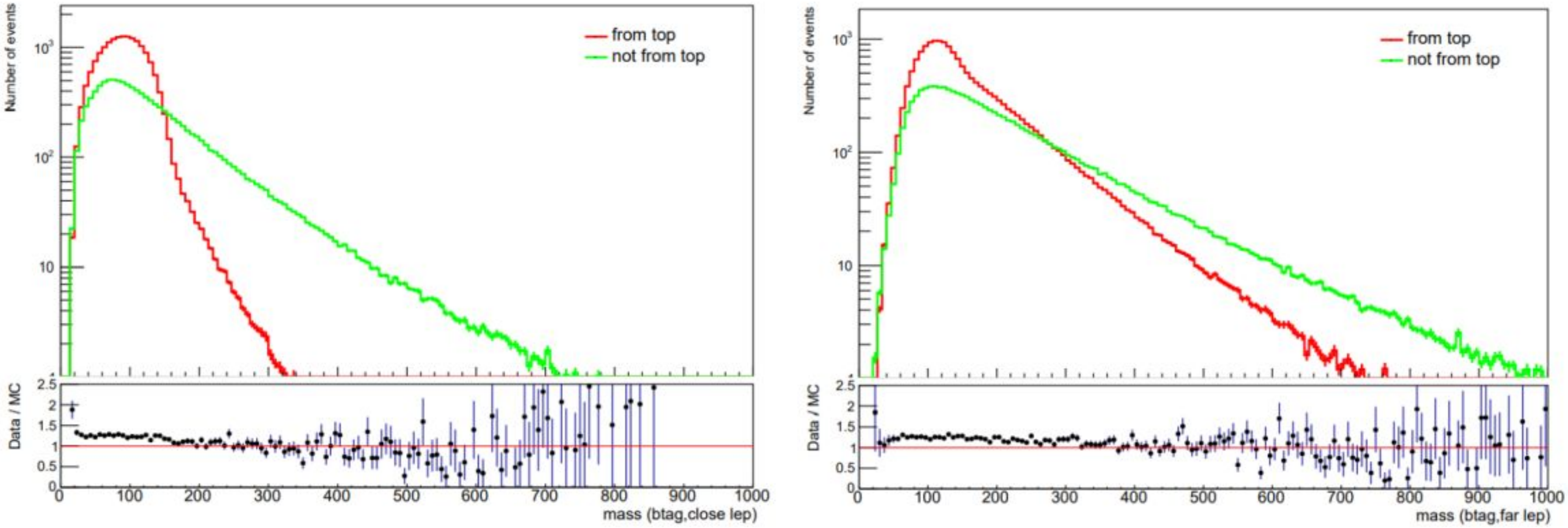
Mass_lep_close, Mass_lep_far
Mass_lep_el, Mass_lep_mu
Mass_lep_1pt, Mass_lep_2pt
Mass_bjet_large, Mass_bjet_small

Invariant Mass Variables

dR_lep_close, dR_lep_far
dR_bjet_close, dR_bjet_far
dR_lep_el, dR_lep_mu
dR_lep_1pt, dR_lep_2pt

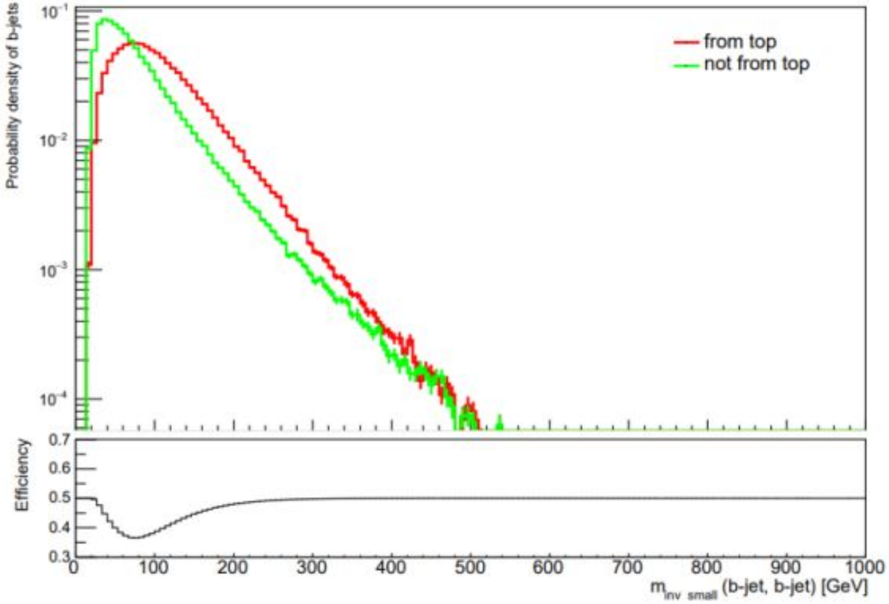
Radial Distance Variables

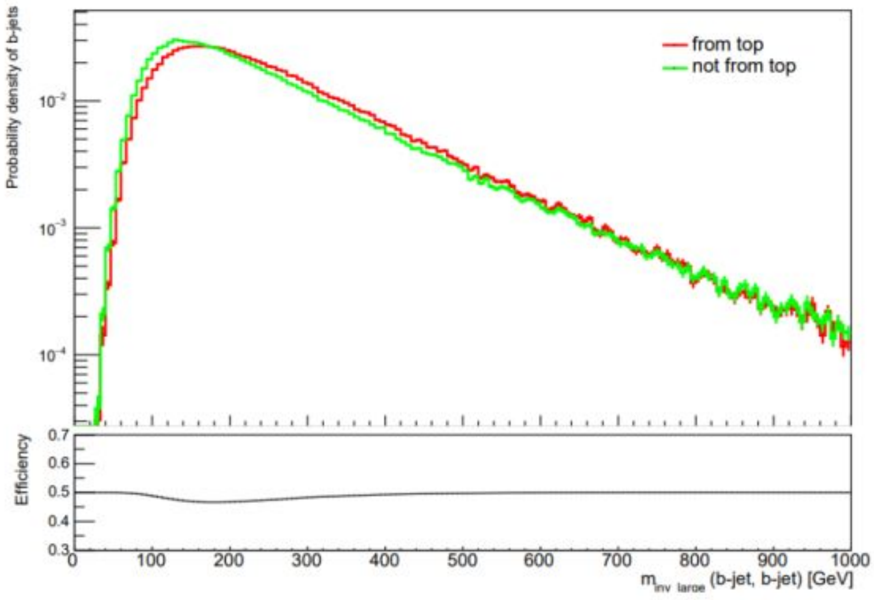
Discriminating Variables: Invariant Mass with lepton



b-jets from tops are expected to have invariant mass ~m_{top} (smaller) with either of the two leptons.

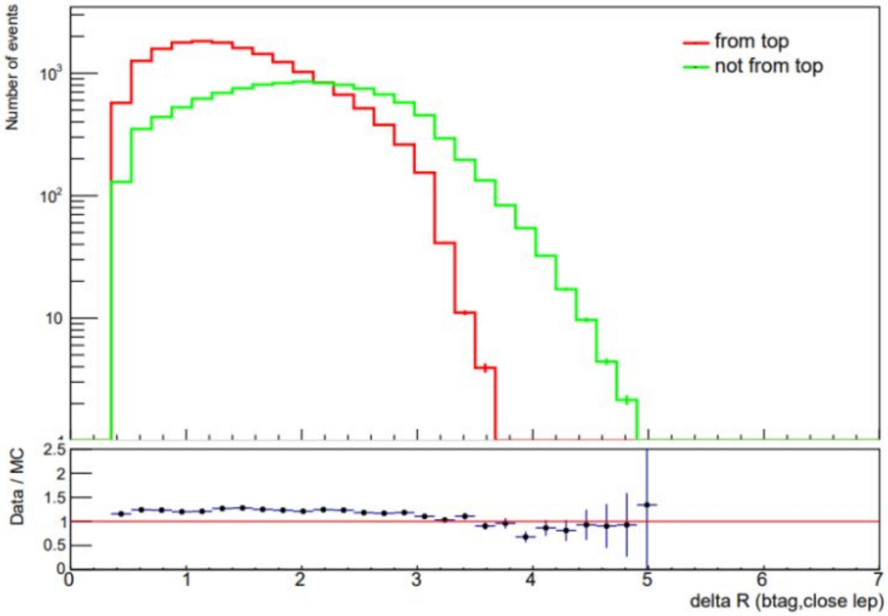
Discriminating Variables: Invariant Mass between *b*-jets

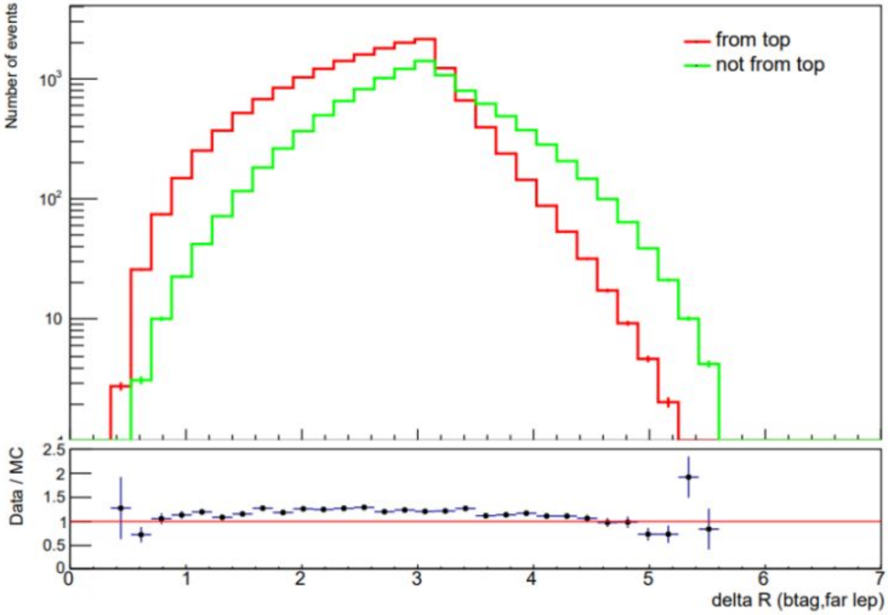




b-jets from tops are expected to have invariant mass ~m_{top} (larger) with each other.

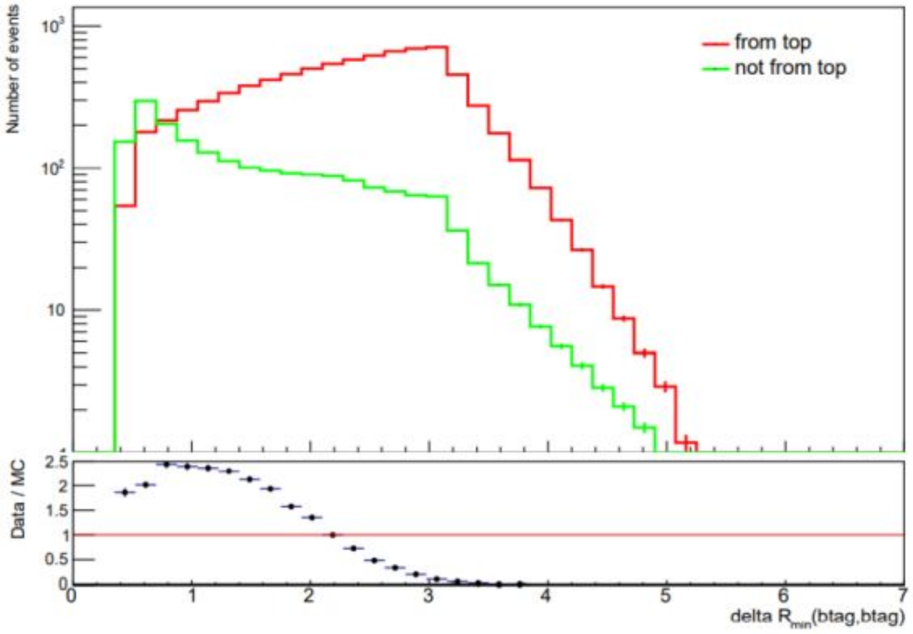
Discriminating Variables: Radial Distance with lepton

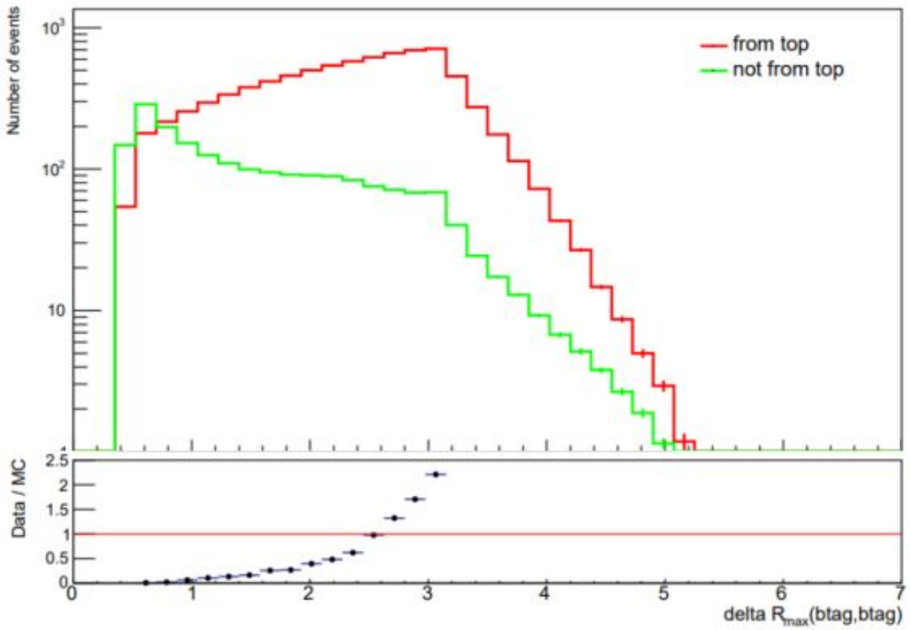




b-jets from tops are expected to be closer to leptons.

Discriminating Variables: Radial Distance between *b*-jets

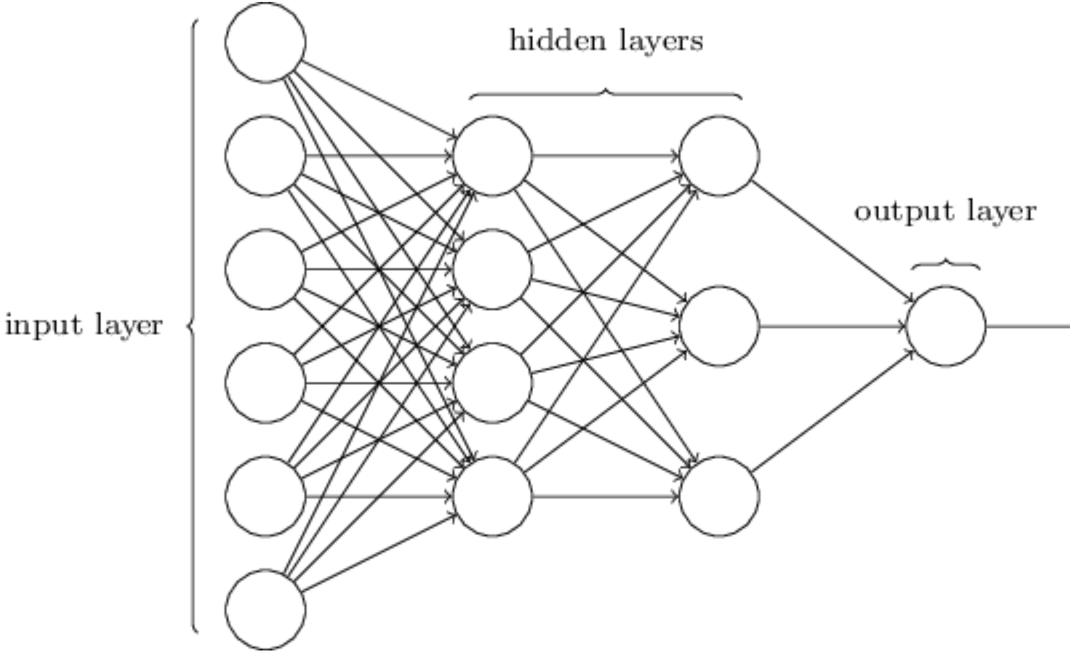




b-jets from gluons are expected to be closer to each other.

Construction of Neural Network (NN)

- Machine learning system made up of connected nodes/neurons
- Train a model by a set of discriminating features to learn a way that minimizes loss function
- In this project, used to solve a binary classification problem
- Built with the Tensorflow and Keras deep learning API



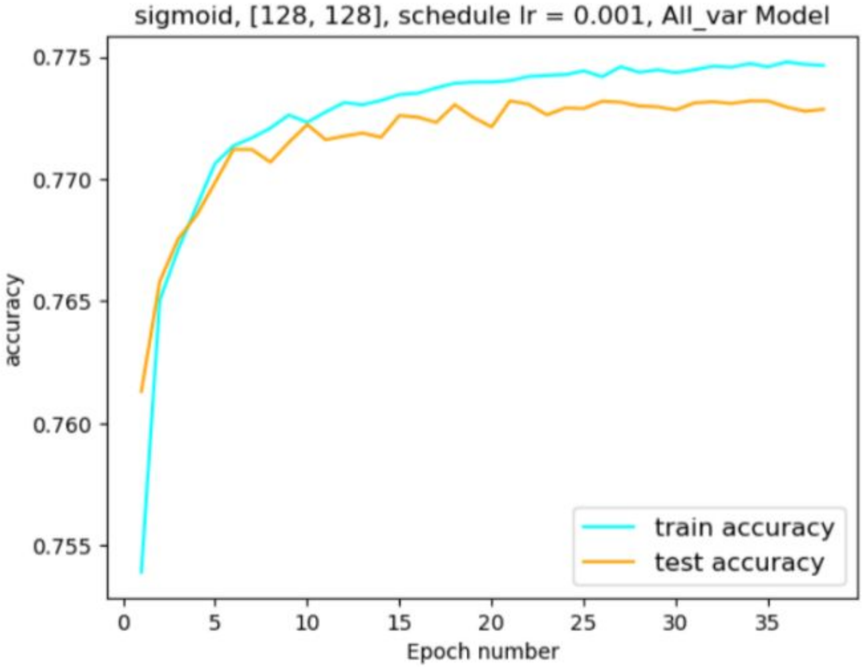
Hyper-Parameter Setup

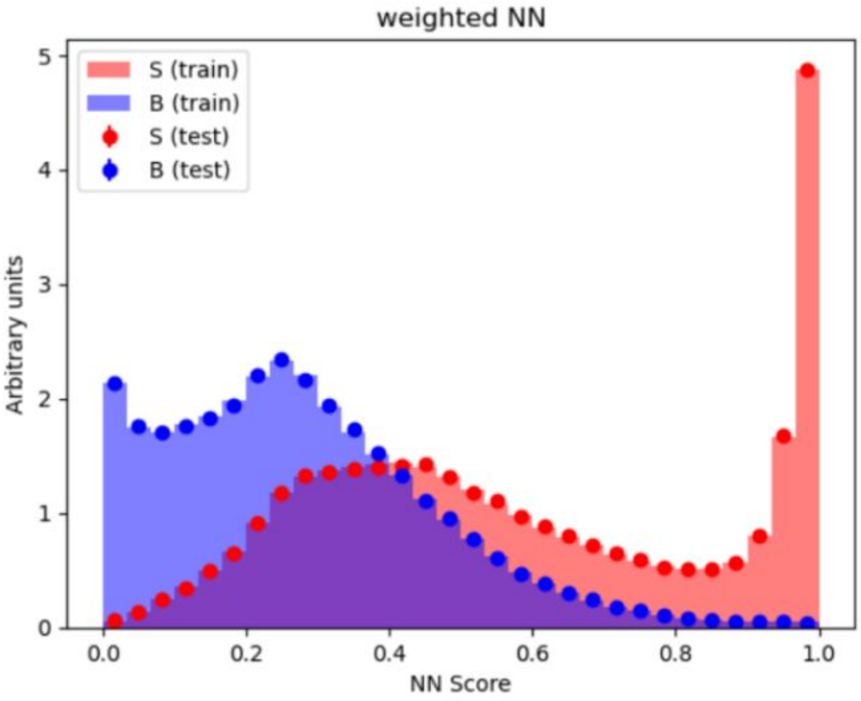
Activation Function	Sigmoid
Loss Function	binary crossentropy
Hidden Layers	128, 128
Output Layer	1
Optimizer	Keras Adam
initial learning rate L_0	0.003
decay rate k	0.1

$$L = L_0 * e^{-k*n}$$

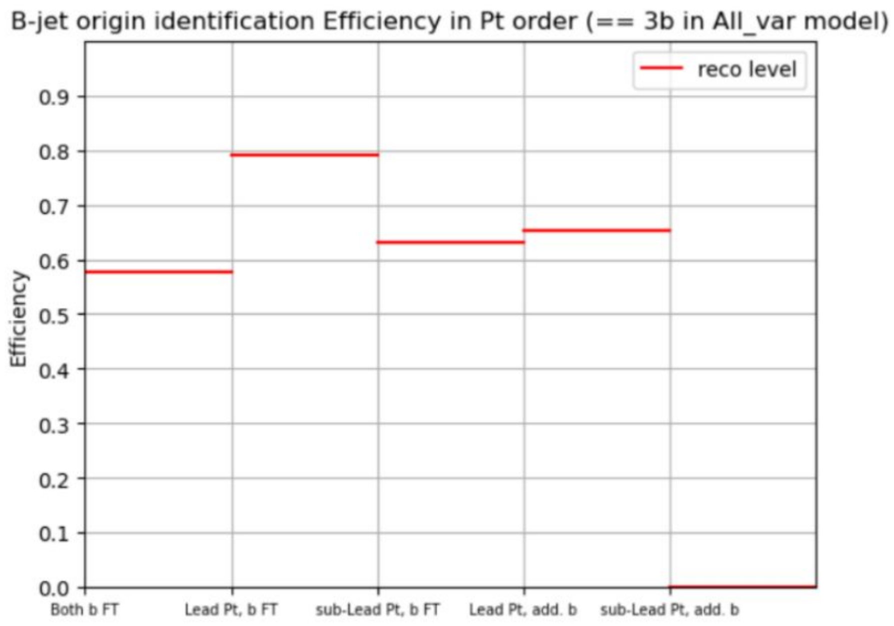
L: current learning rate
L_0 : initial learning rate
k : decay rate
n : current epoch number

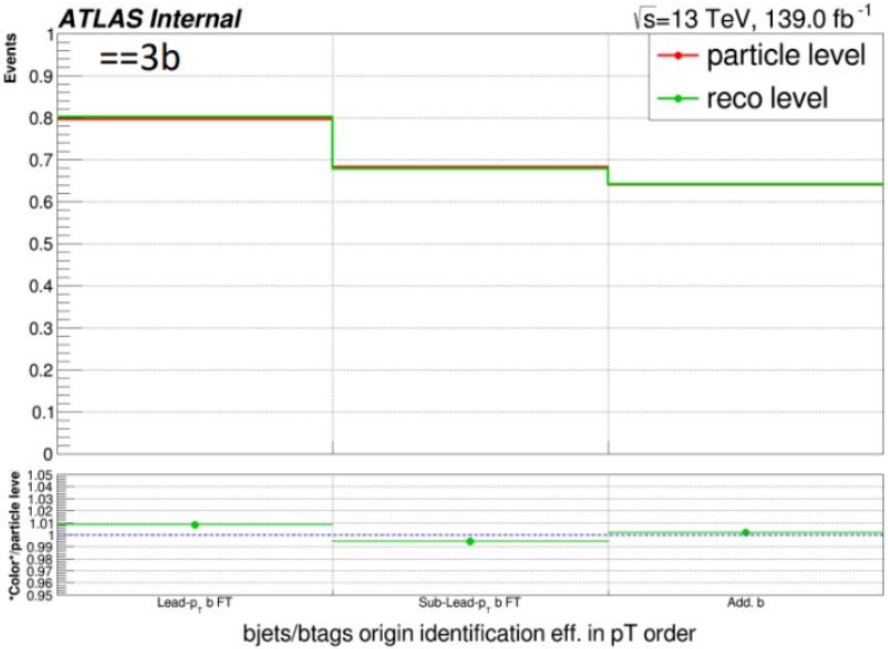
NN Performance





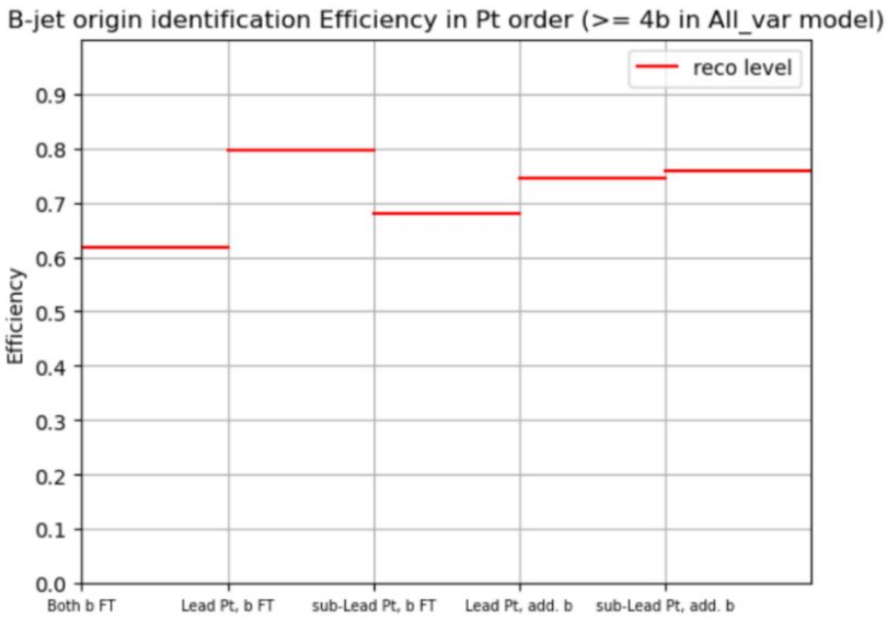
Efficiencies Comparison between NN and BDT

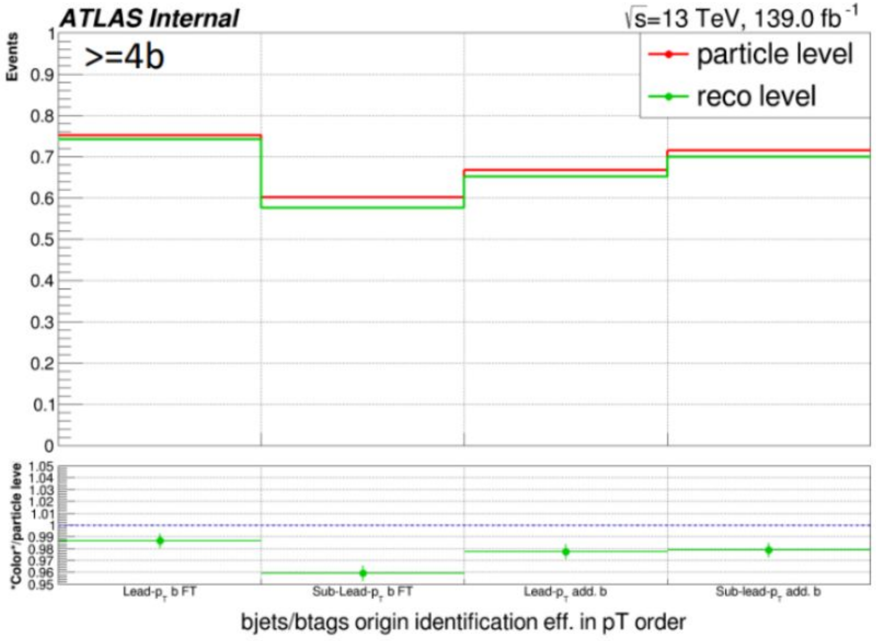




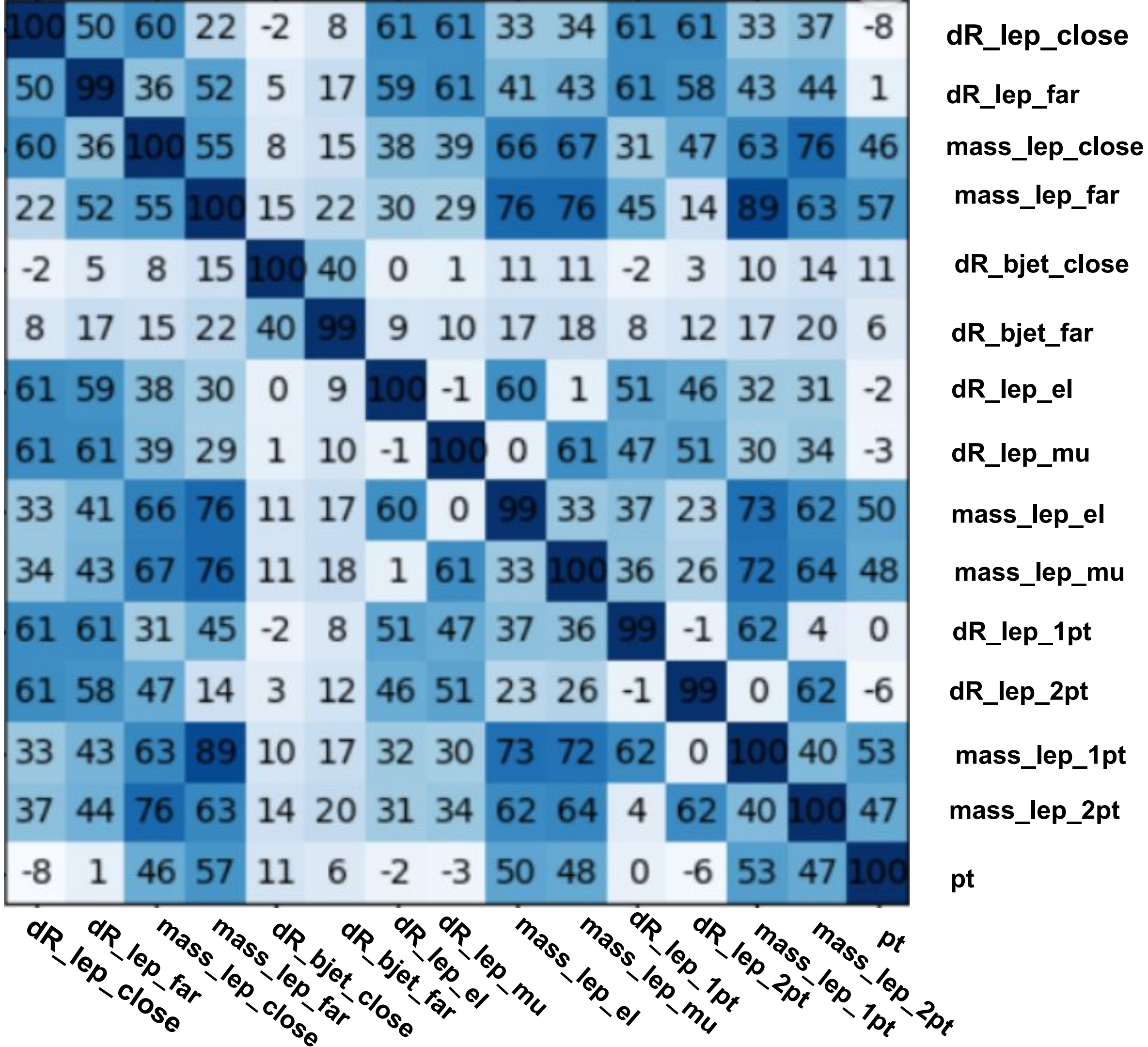
Efficiency: ratio of the number of correctly assigned *b*-jets by the classifier to the truth number of *b*-jets in one category

Efficiencies Comparison between NN and BDT





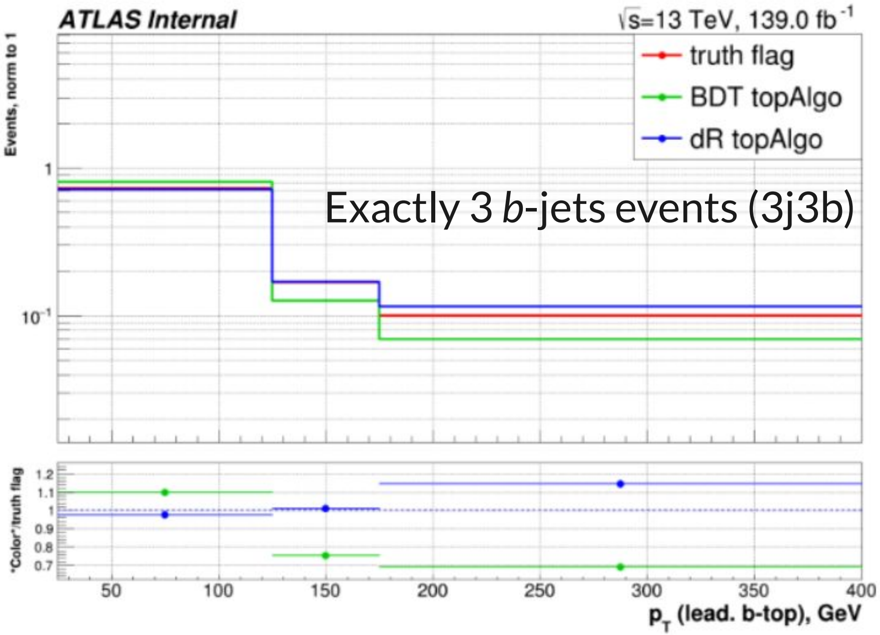
Correlation matrix

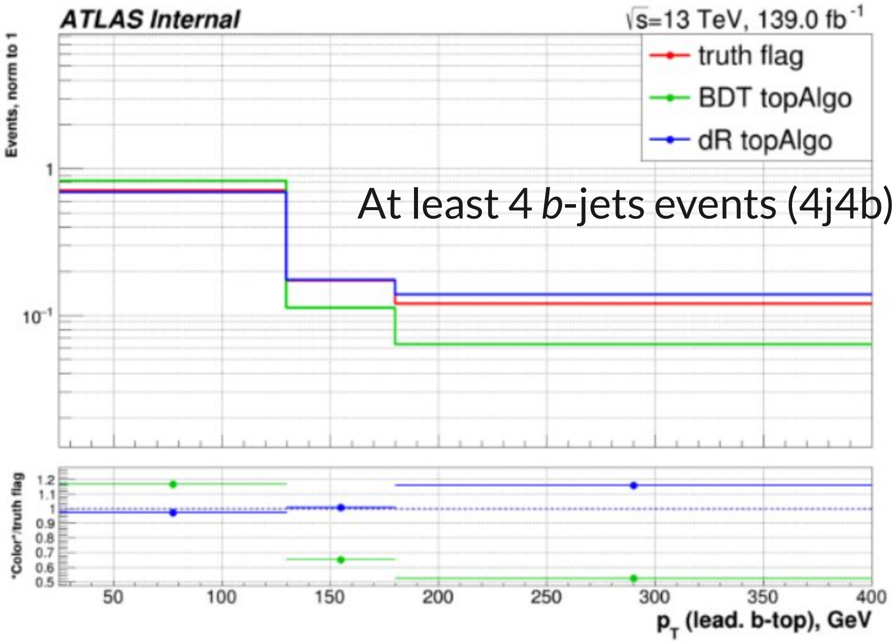


$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$

As the classifier will be used for unfolding of leading, sub-leading from tops and additional *b*-jets p_T (4 variables in total), we aim that prediction of classifier is independent of transverse momentum p_T

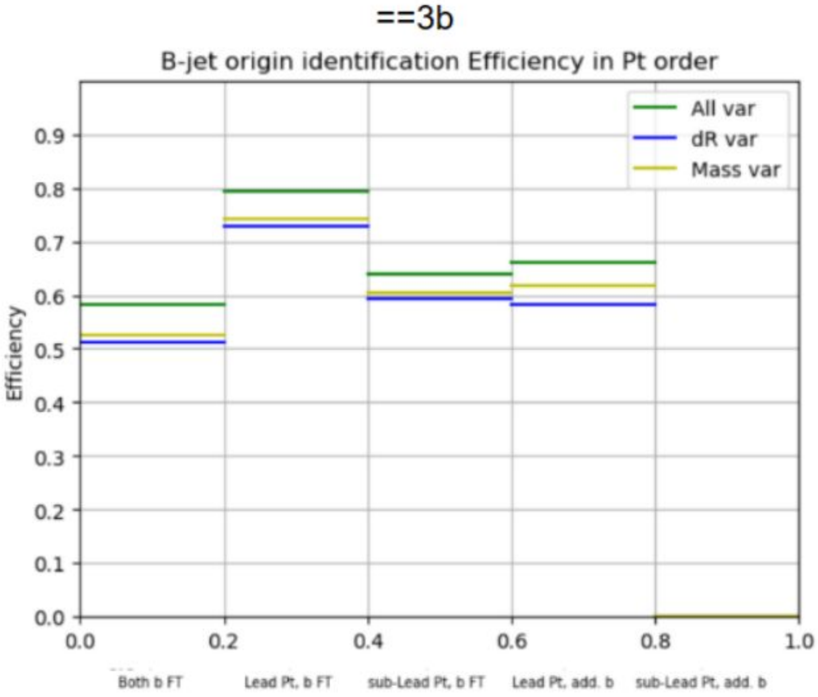
pT bias in Boosted Decision Tree (BDT) Model

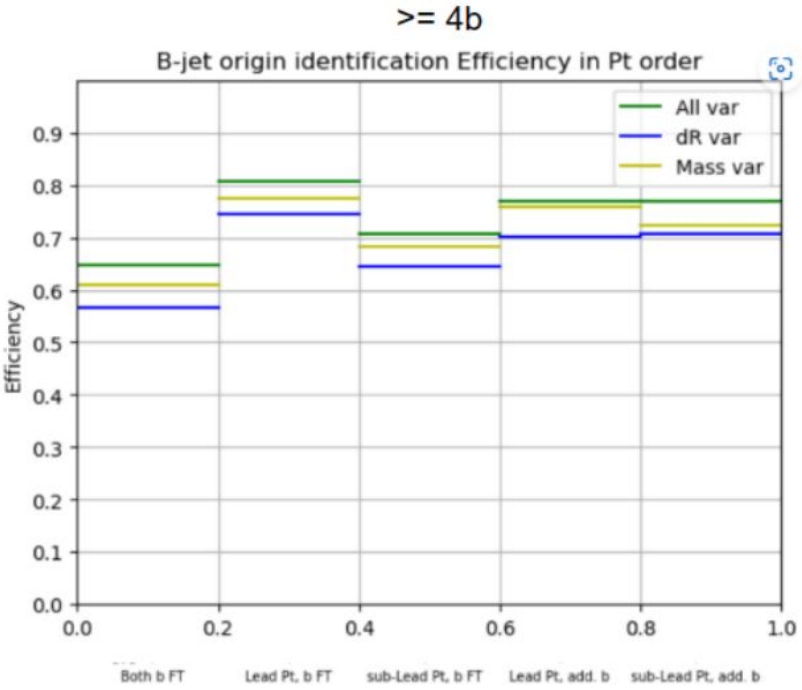




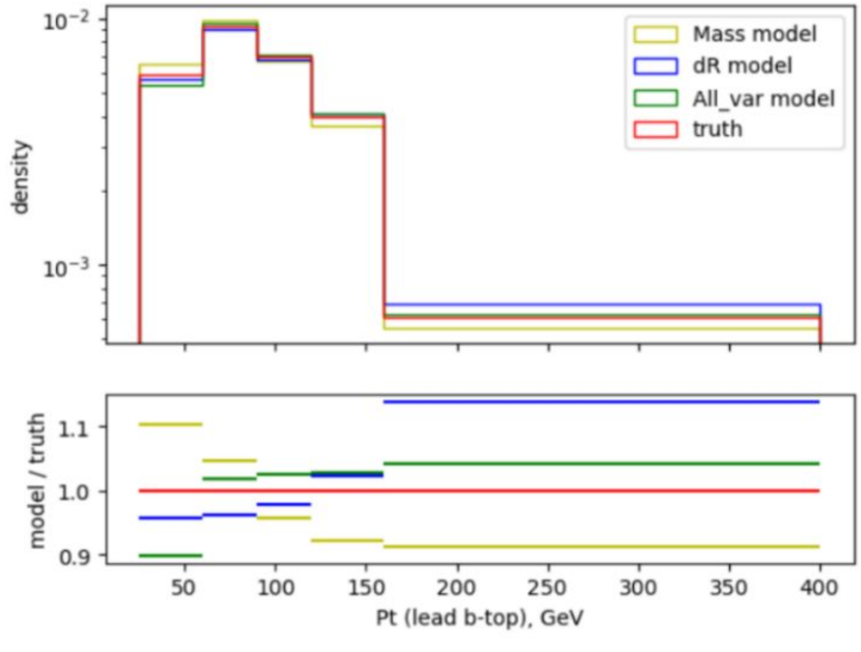
p_T of the leading *b*-jet from tops assigned by NN algorithms and compared to the truth from leading p_T tops spectrum.

Efficiencies of Three NN Classifiers



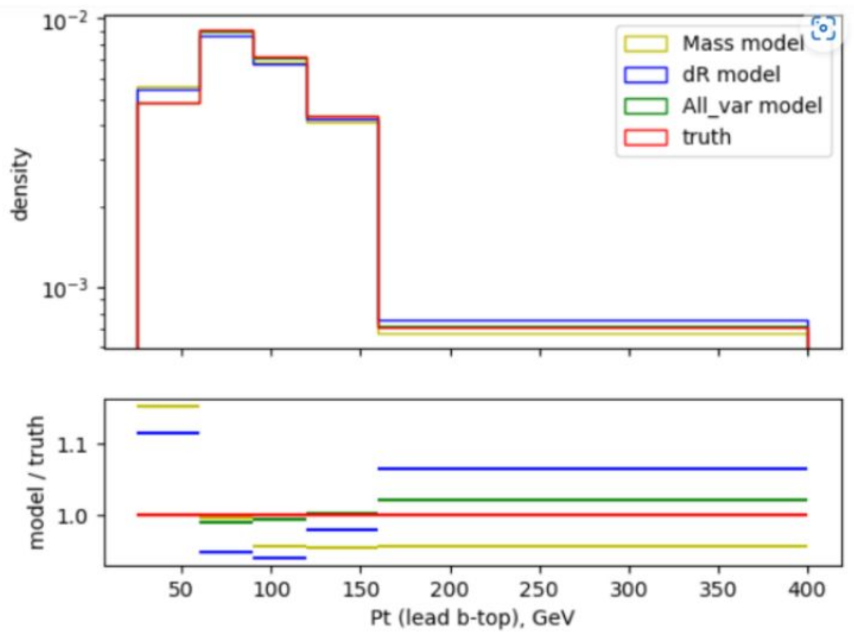


pT bias in NN Model




```
deviation of All_var model: 0.11638680870001683
deviation of dR model: 0.1530109837116903
deviation of Mass model: 0.16917171089315283
weighted deviation of All_var model: 1.905610432638008
weighted deviation of dR model: 1.4705694526182662
weighted deviation of Mass model: 2.590659454649556
```

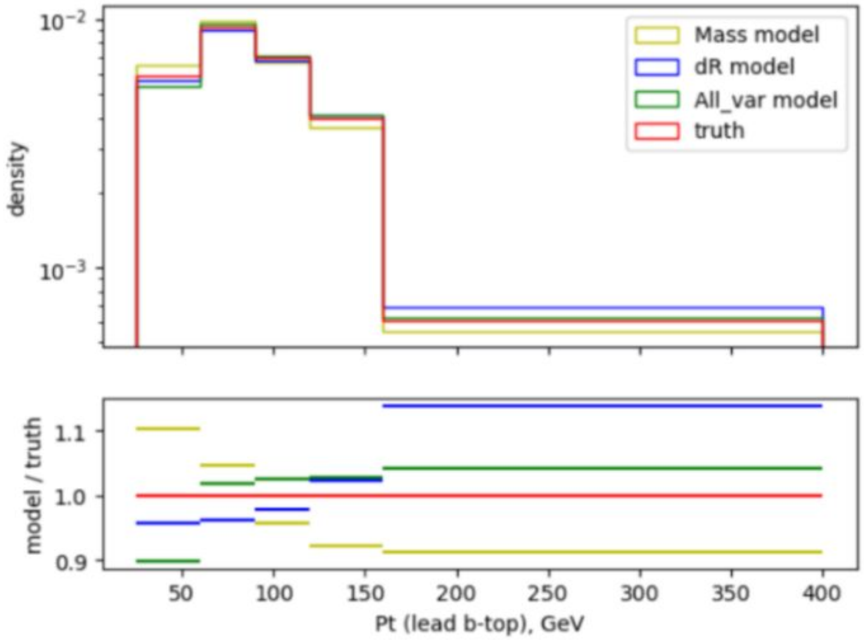
3j3b

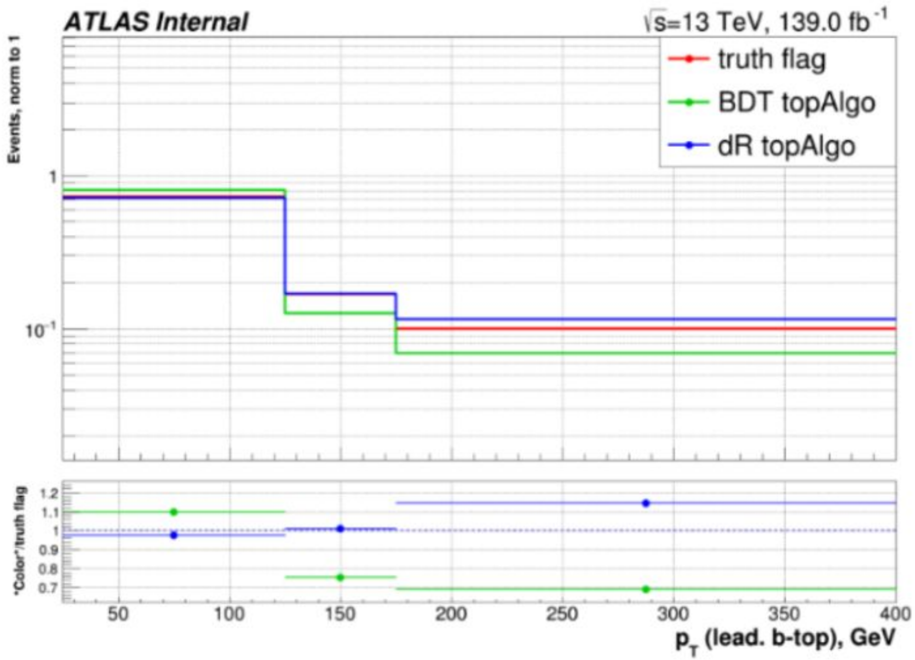



```
deviation of All_var model: 0.023575461021797226
deviation of dR model: 0.15537811271183247
deviation of Mass model: 0.17183578715599848
weighted deviation of All_var model: 0.3243938529932606
weighted deviation of dR model: 2.6663535106861938
weighted deviation of Mass model: 2.434090543234854
```

4j4b

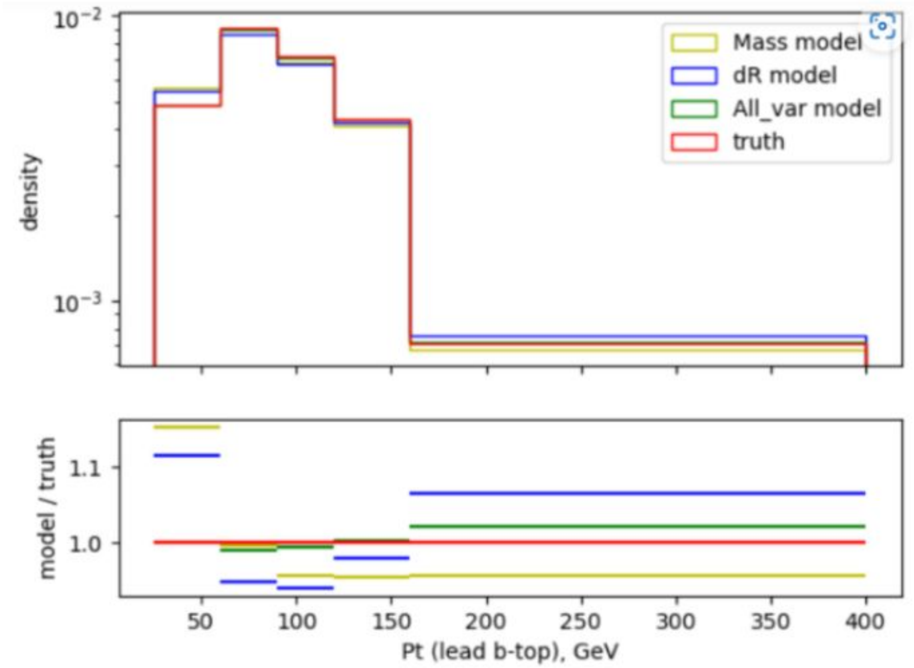
pT bias Comparison between NN and BDT

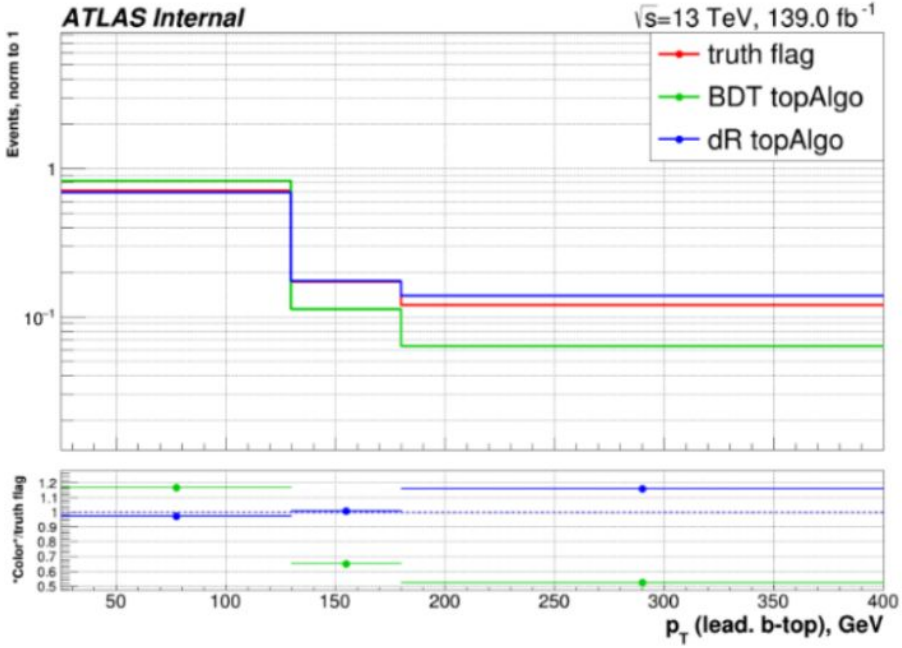




Leading *b*-jet from tops in 3j3b events

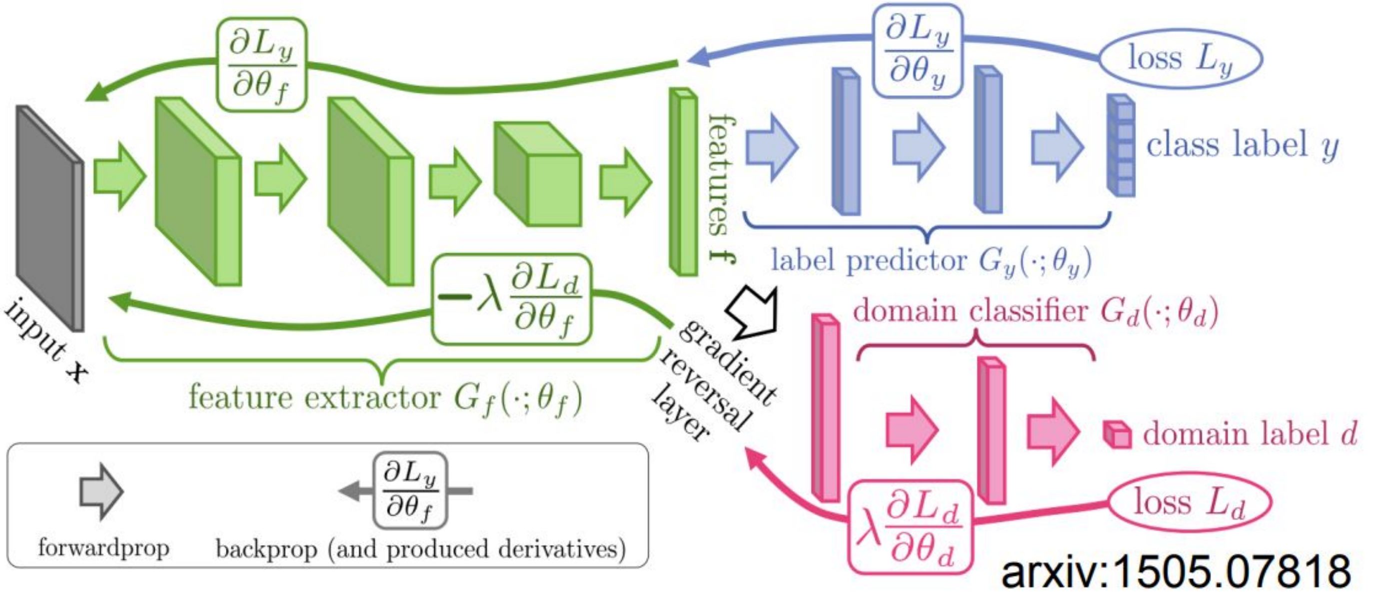
pT bias Comparison between NN and BDT





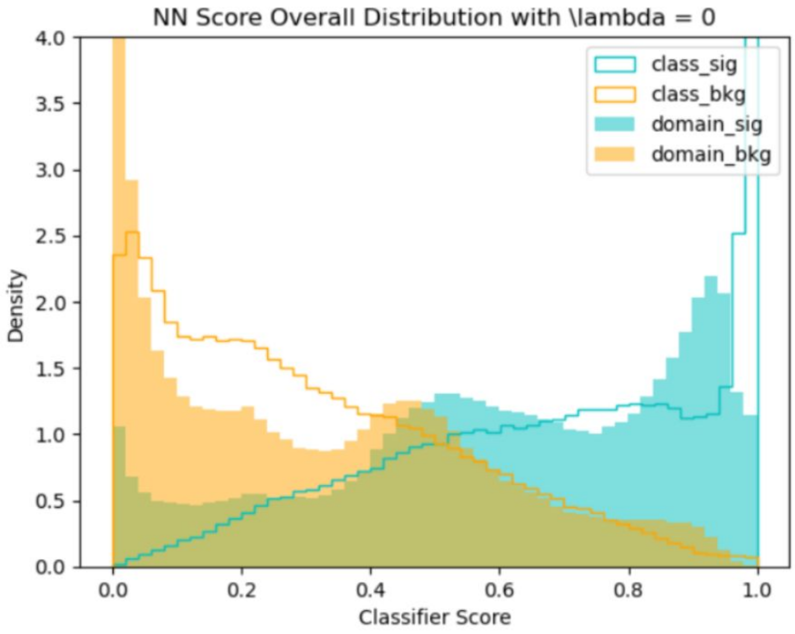
Leading *b*-jet from tops in 4j4b events

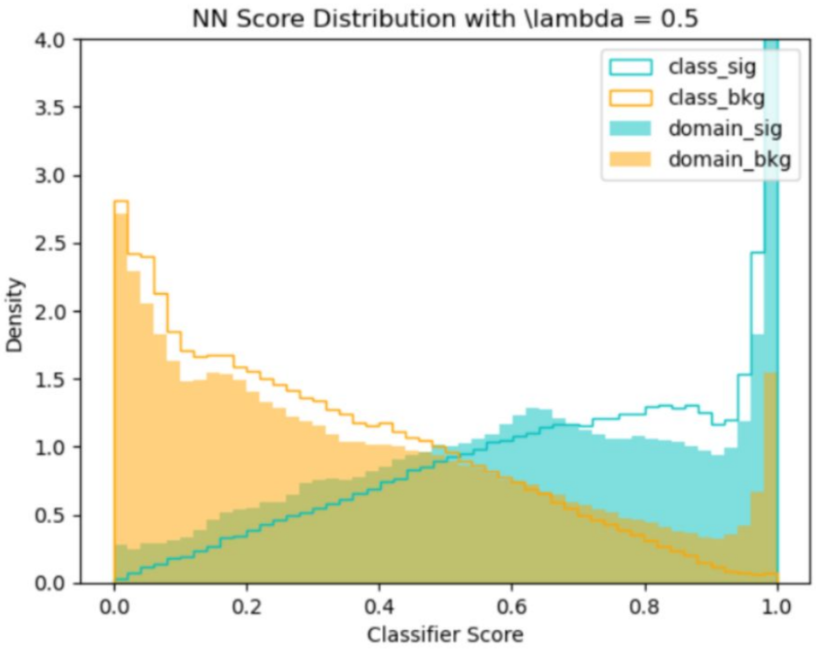
Domain-Adversarial Neural Network (DANN)



- p_T bias suppression setups
- Build nominal vs. alternative MCs-independent classifier

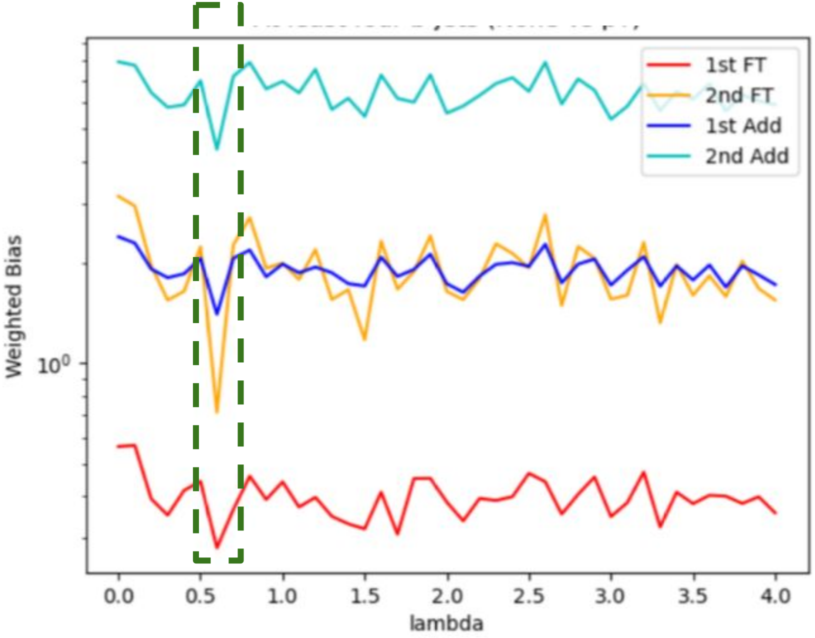
DANN: Suppress pT Bias

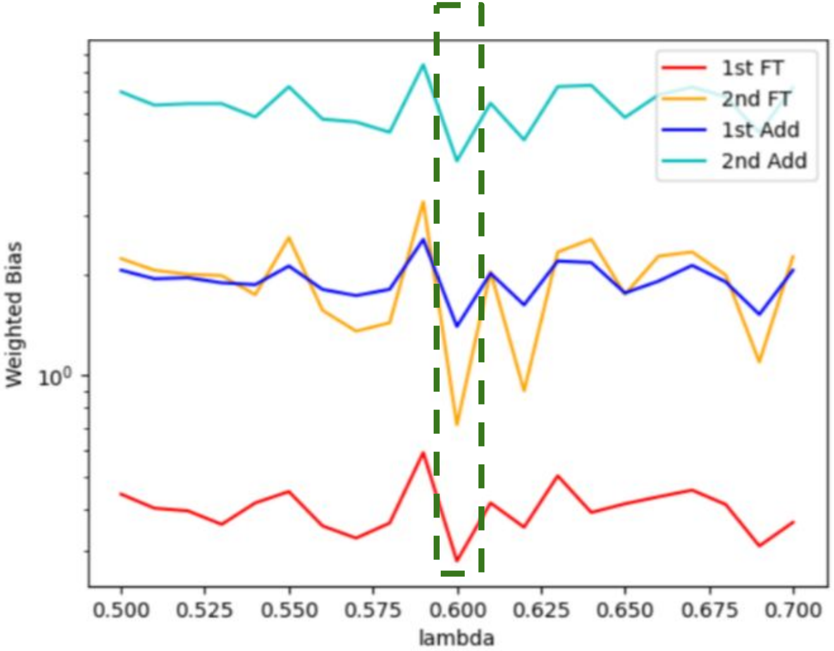




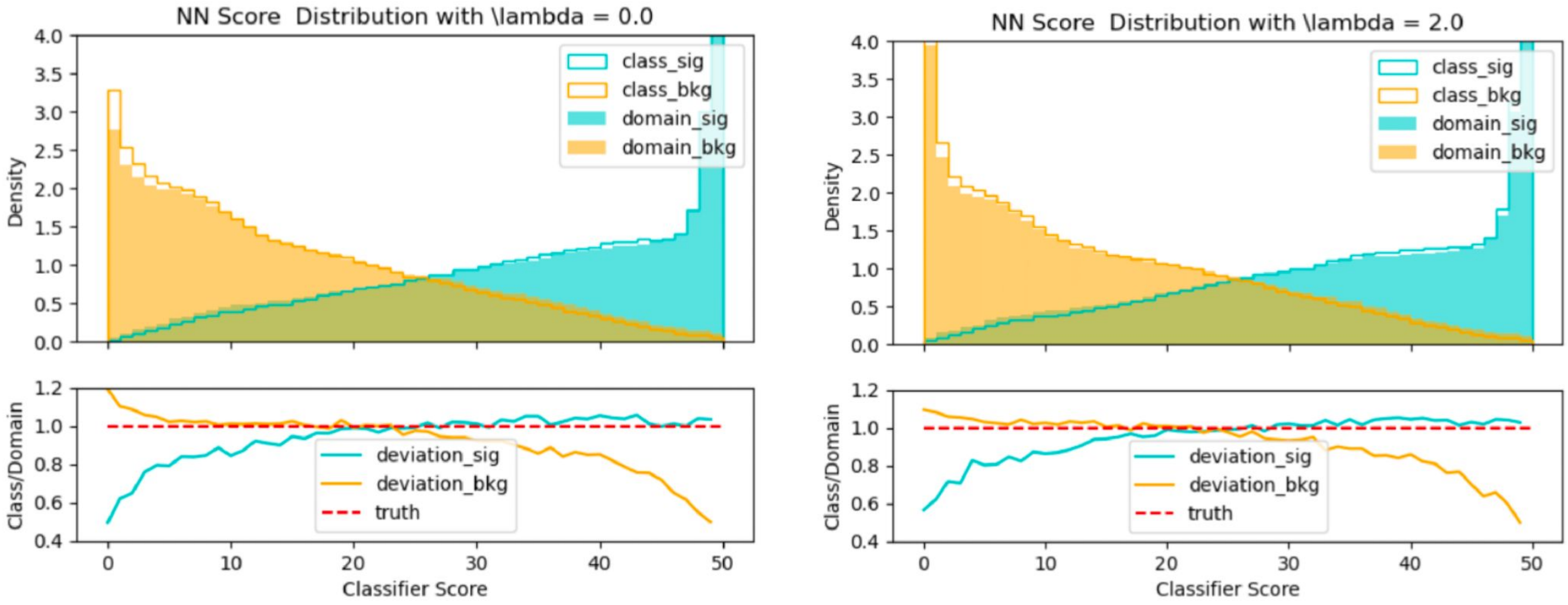
- Classifier label: the original set of variables for *b*-jets, no *b*-jet p_T (set to 0 for all).
- Domain label: the original set of variables for *b*-jets plus real *b*-jet p_T.

DANN: Suppress pT Bias

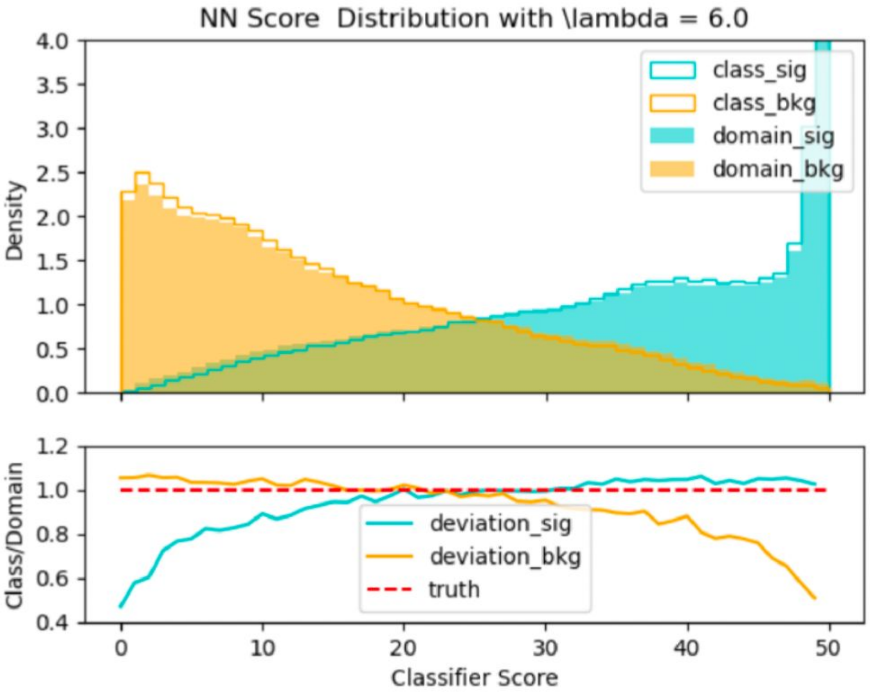


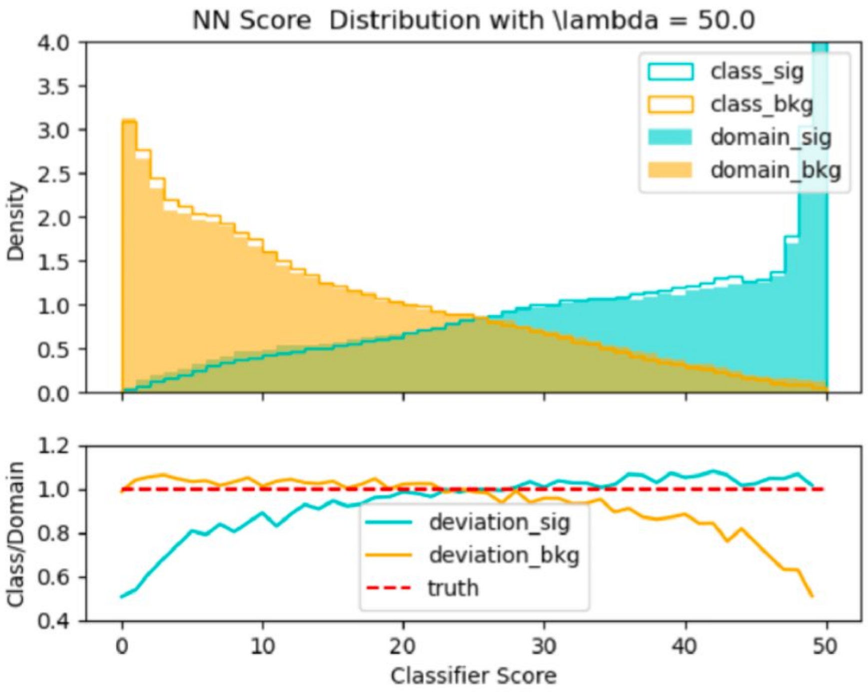


DANN: PowPy8 vs. Sherpa 2.2.10

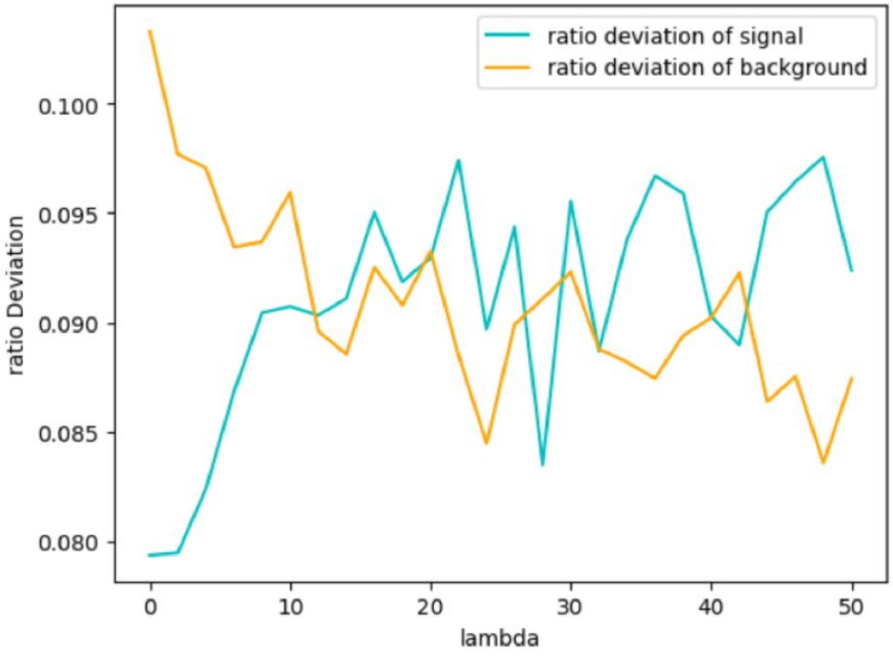


DANN: PowPy8 vs. Sherpa 2.2.10





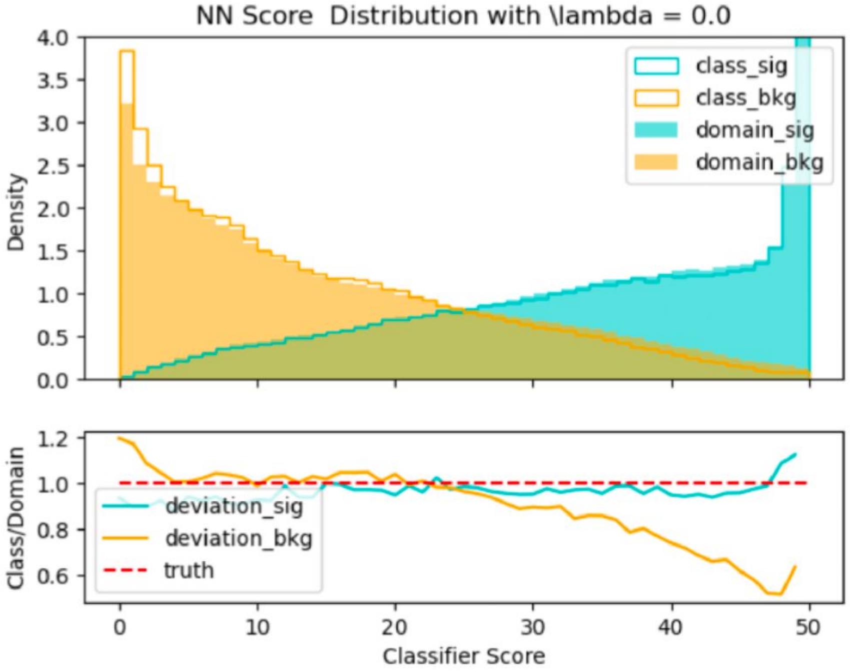
DANN: PowPy8 vs. Sherpa 2.2.10

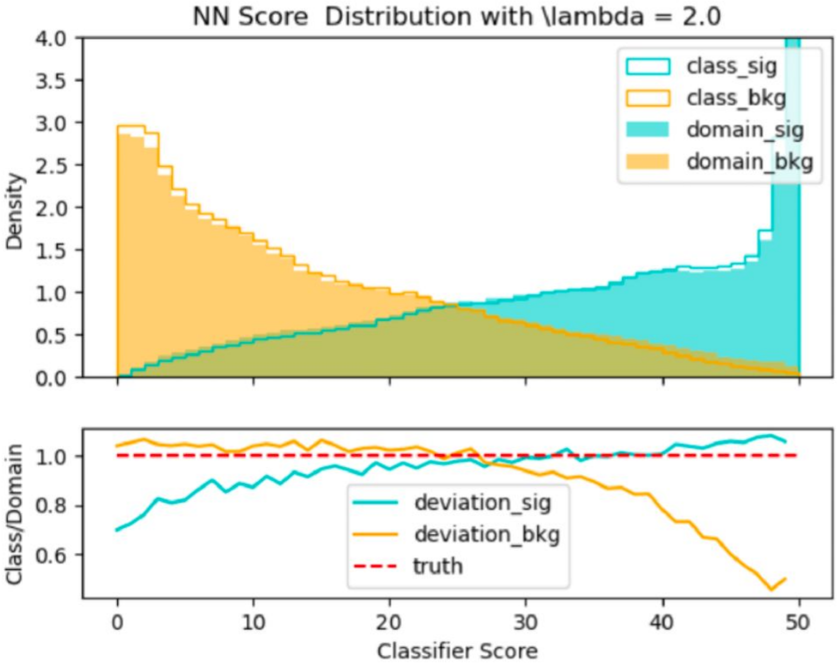


- Increasing deviation of *b*-jets from gluons
- Decreasing deviation of *b*-jets from tops

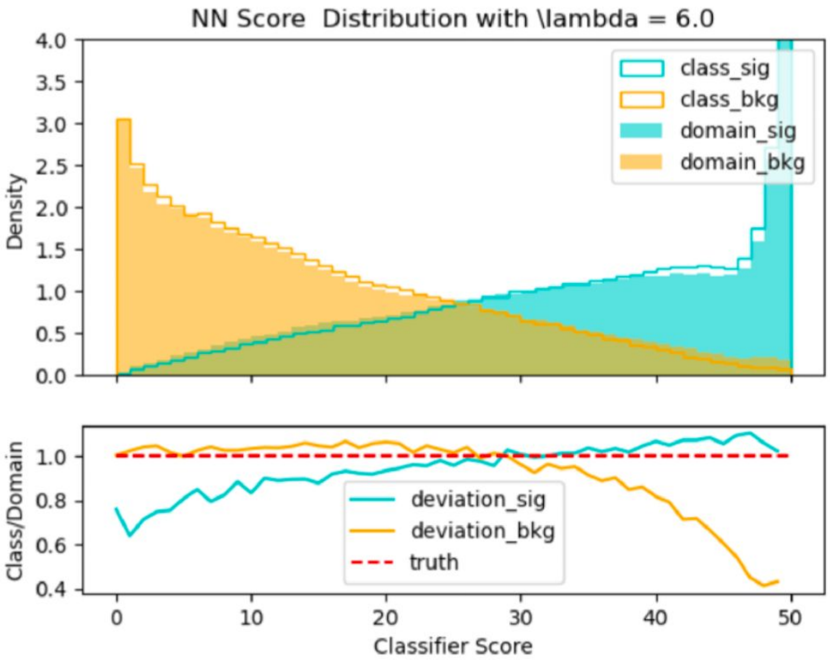
The absolute difference value between class/domain ratio and 1, in dependence of the scale factor.

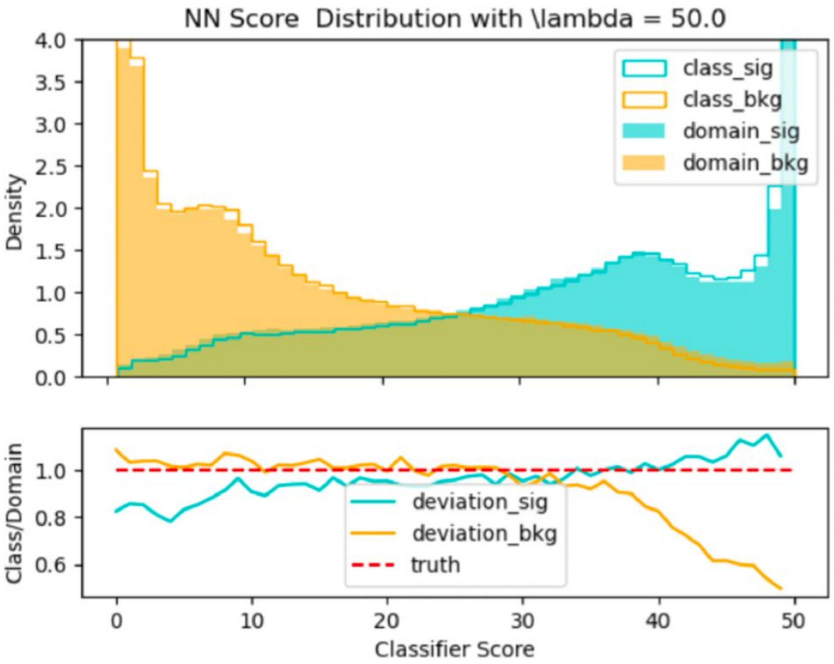
DANN: PowPy8 vs. MG_aMC@NLO+Py8



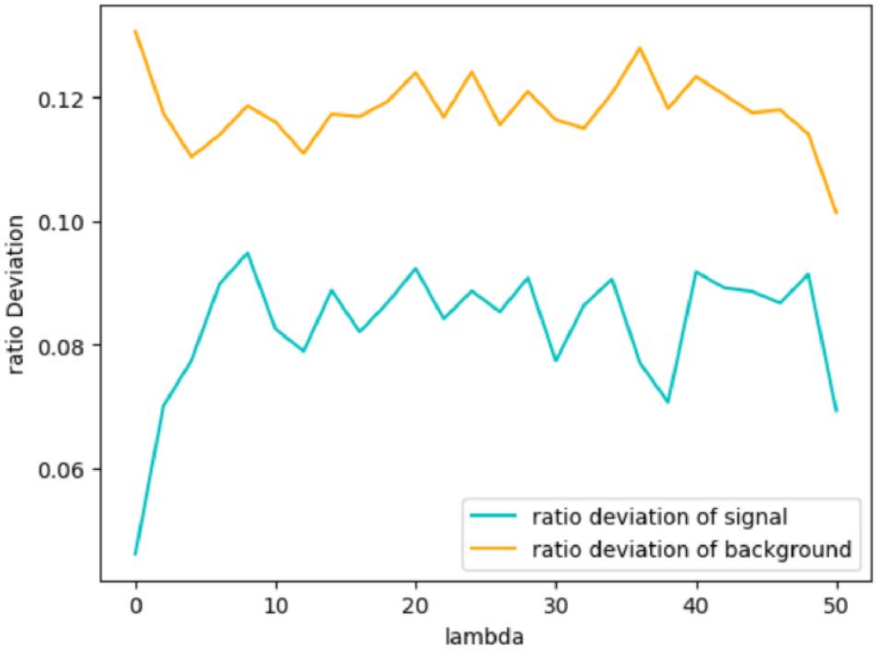


DANN: PowPy8 vs. MG_aMC@NLO+Py8





DANN: PowPy8 vs. MG_aMC@NLO+Py8



- Increasing deviation of *b*-jets from gluons
- Decreasing deviation of *b*-jets from tops
- Similar pattern but flat after 9

Conclusion

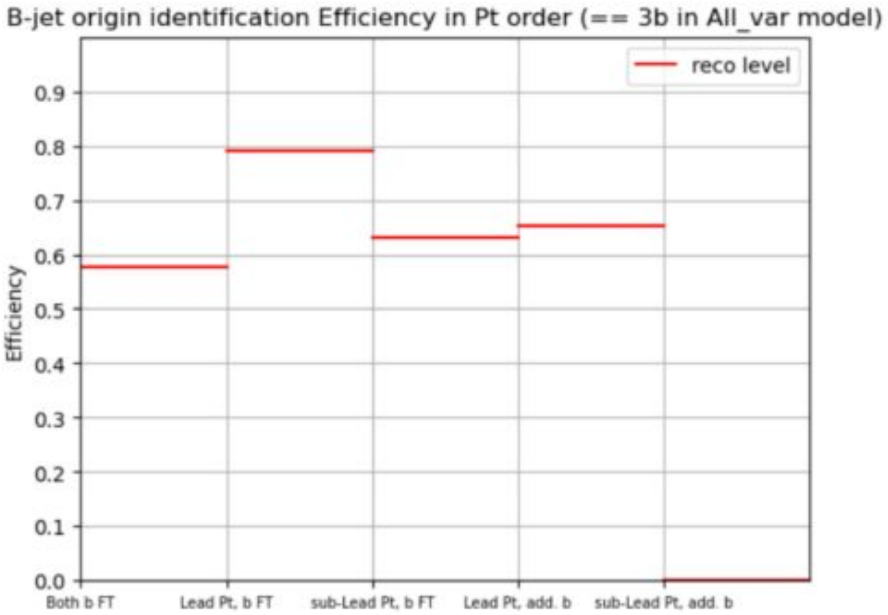
DNN:

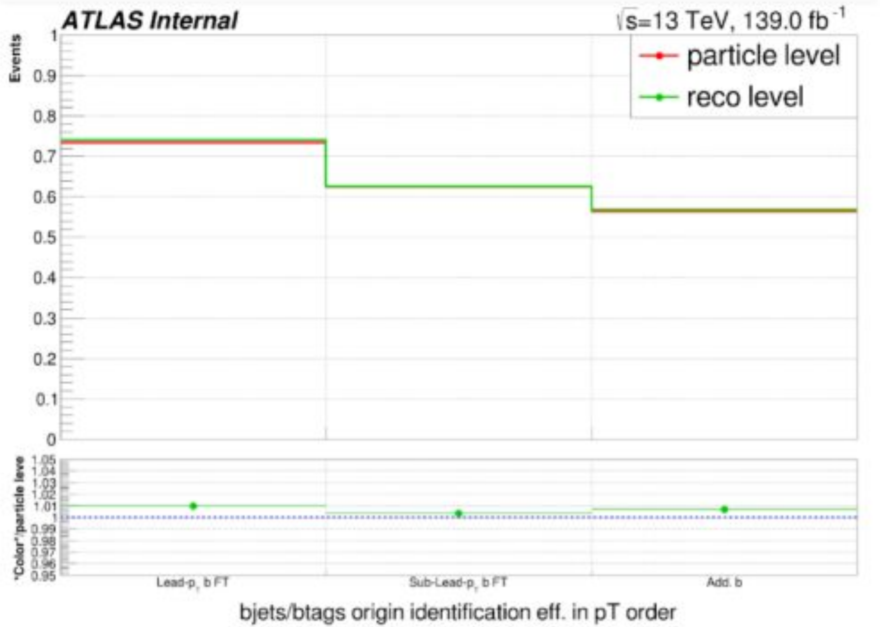
- Higher efficiency than the previous BDT classifier in 4j4b events (~5%).
- Lower p_T bias than the BDT, making DNN better
  (~6% in 3j3b events and ~10% in 4j4b events).

DANN:

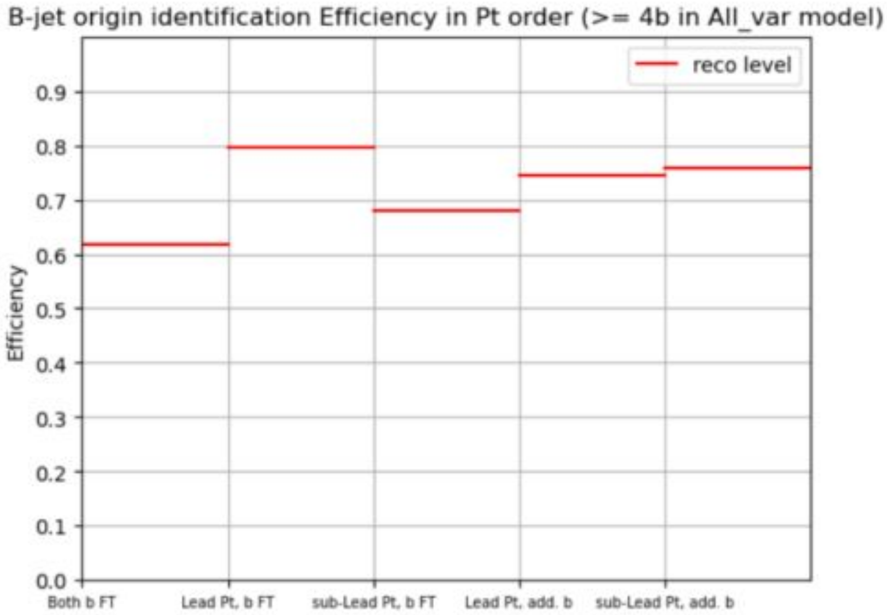
- Decease p_T bias even more by showing an optimal $\lambda = 0.6$.
- Potential of MC-independent classifier.

Back up

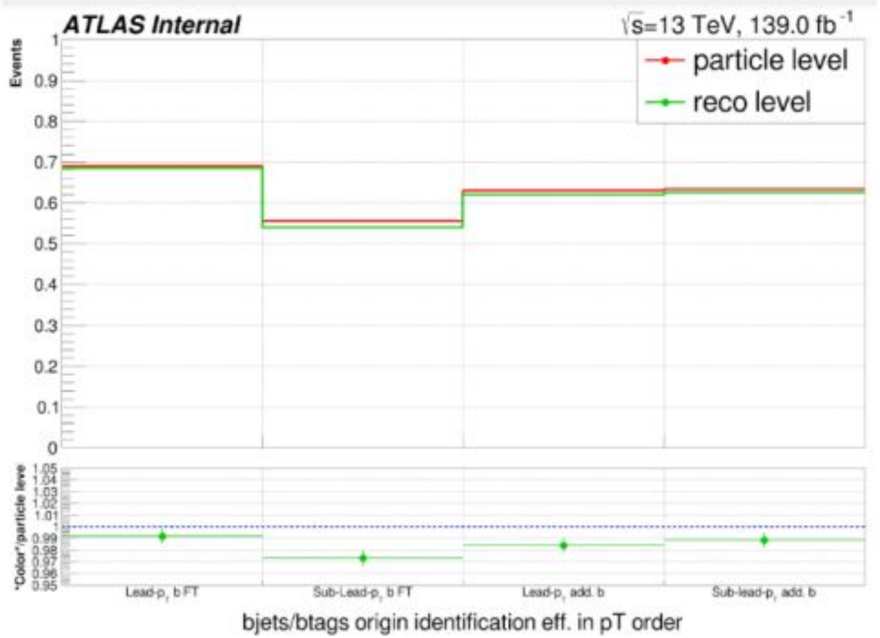

B-jet origin identification Efficiency in Pt order (== 3b in All_var model)
reco level
Efficiency
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Both b FT
Lead Pt, b FT
sub-Lead Pt, b FT
Lead Pt, add. b
sub-Lead Pt, add. b


ATLAS Internal
√s=13 TeV, 139.0 fb⁻¹
particle level
reco level
Events
Lead-p, b FT
Sub-Lead-p, b FT
Add. b
bjets/btags origin identification eff. in pT order

Back up



B-jet origin identification Efficiency in Pt order (>= 4b in All_var model)
reco level
Efficiency
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Both b FT
Lead Pt, b FT
sub-Lead Pt, b FT
Lead Pt, add. b
sub-Lead Pt, add. b



ATLAS Internal
√s=13 TeV, 139.0 fb⁻¹
particle level
reco level
Events
Lead-p, b FT
Sub-Lead-p, b FT
Lead-p, add. b
Sub-lead-p, add. b
bjets/btags origin identification eff. in pT order