DeXTer: Deep Sets based Neural Networks for Low-p_T X→ $b\bar{b}$ Identification in ATLAS

Yuan-Tang Chou (University of Massachusetts Amherst)

on behalf of ATLAS Collaboration

Aug 15-19 BOOST 2022@Hamburg

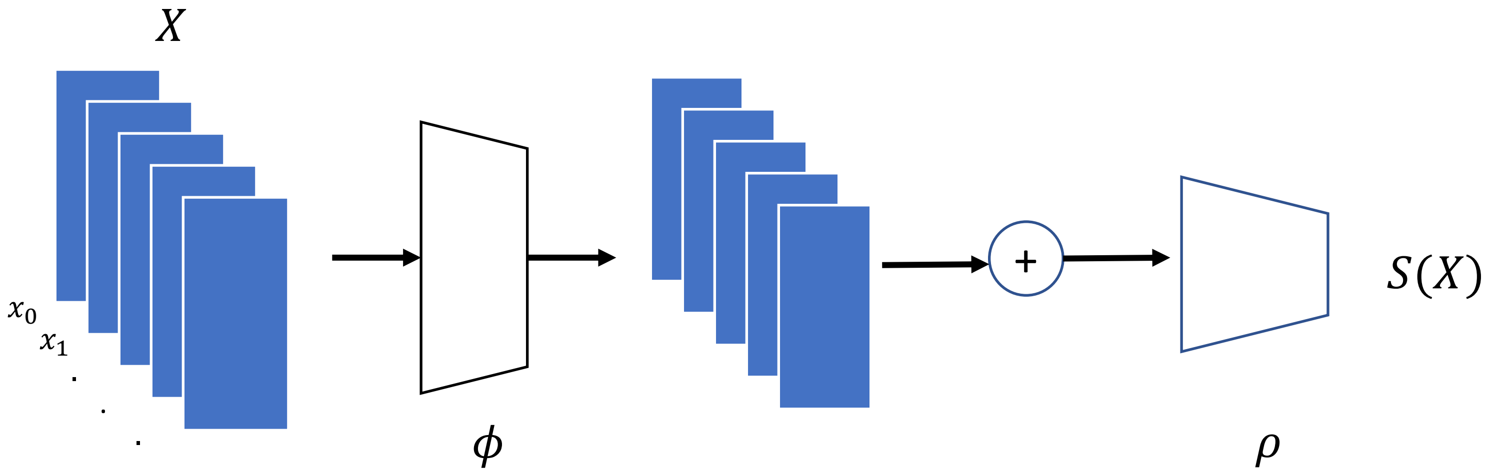
Introduction

- Several Standard Model and BSM processes produce collimated hadronic final states such as some Higgs portal and dark matter models
- Machine learning techniques offer significant improvements in many areas in HEP
  - Recurrent neural network (RNN)
  - Convolutional neural network (CNN)
- Dedicated flavor tagging algorithms are developed to tag collimated objects for heavy resonance
- Several BSM search for light resonance motivate the development of general-purpose low-mass double-b tagging
  - VH, $H \to aa \to b\bar{b}b\bar{b}$

We present the first general-purpose low-mass/p_T double-b tagging algorithm in ATLAS

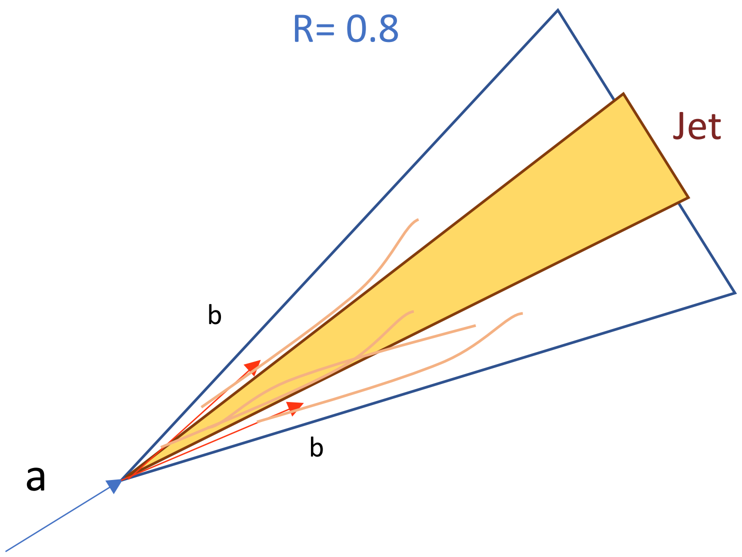
Deep Sets in jet flavor tagging

- Collection of objects treated as a set without empirical ordering
- Architecture suitable for jet flavor tagging with a variable number of tracks
- Prior development using this architecture has shown improvement on single b-tagging in ATLAS



DeXTer: Deep Sets $X \rightarrow b\bar{b}$ Tagger

A low-mass end-to-end double-b identification algorithm

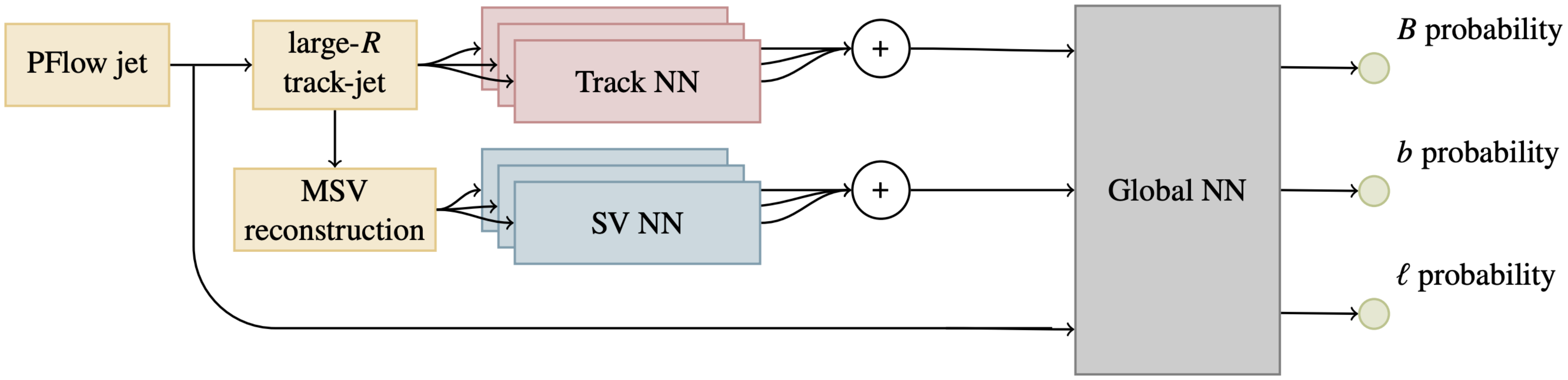


- First general-purpose low-mass $b\bar{b}$ tagger in ATLAS specialized for jets with $20 < p_T < 200$ GeV.
- DeXTer does multi-flavor tagging using jets and surrounding displaced tracks and SVs
- End-to-end training starting from displaced tracks and dedicated multiple secondary vertices (MSV) reconstruction

B: double-b tagged jets

DeXTer: Deep Sets $X \to b\bar{b}$ Tagger

A low-mass end-to-end double-b identification algorithm



- First general-purpose low-mass $b\bar{b}$ tagger in ATLAS specialized for jets with $20 < p_T < 200$ GeV.
- DeXTer does multi-flavor tagging using jets and surrounding displaced tracks and SVs
- End-to-end training starting from displaced tracks and dedicated multiple secondary vertices (MSV) reconstruction

B: double-b tagged jets

Input features for DeXTer

PFlow Jet
- Features to global NN

Track
- Inner detector hits
- Impact parameters
- Angular separation to track subjets

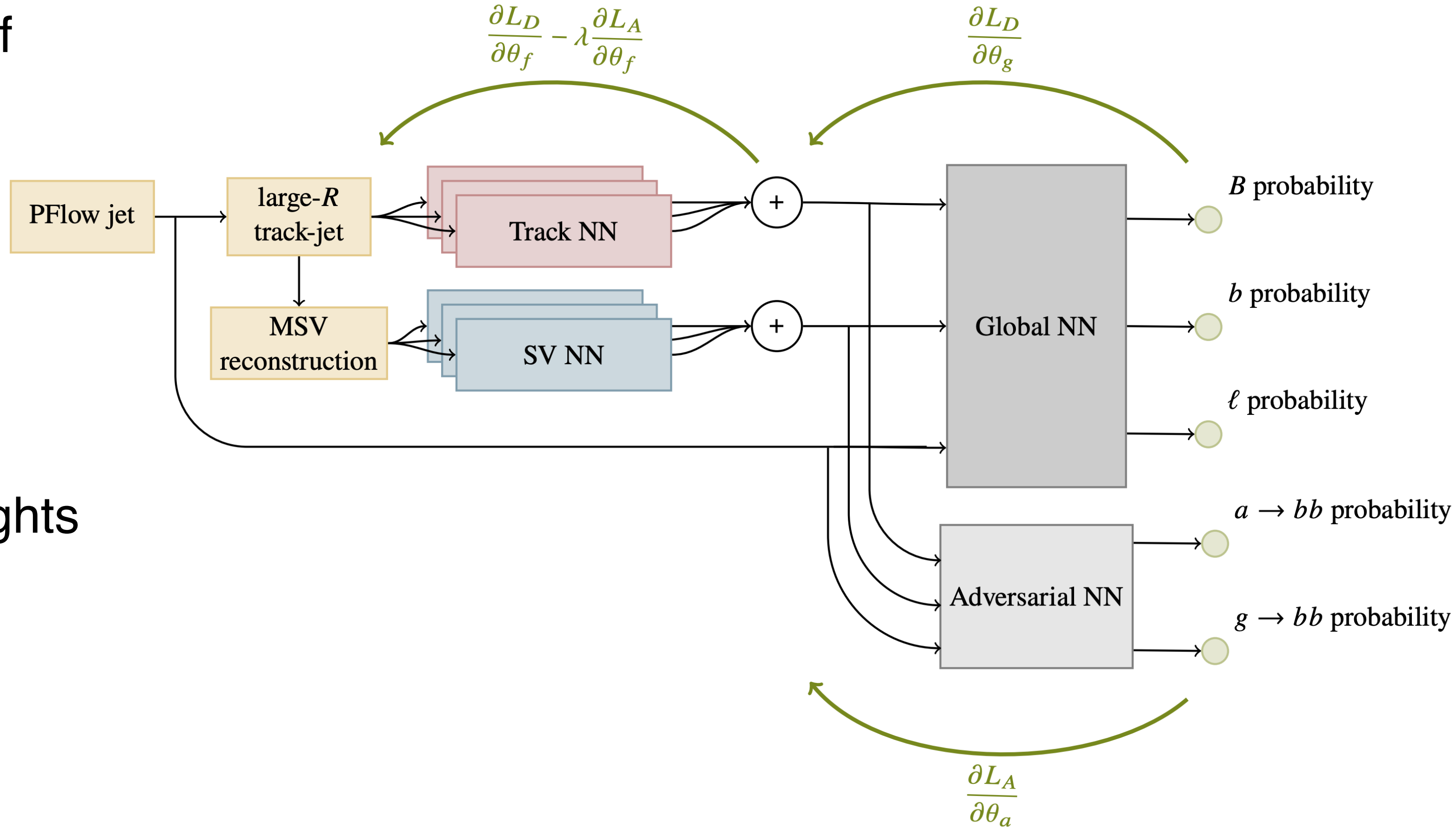
Secondary Vertex
- Track mass
- Angular separation to track subjets
- Decay length and significance

Feature	Description
	Jet
p_{T}	Jet transverse momentum
η	Jet pseudorapidity
	Track
$\log\, p_{\mathrm{T}}^{\mathrm{frac}}$(track, PFlow jet)	$\log\, p_{\mathrm{T}}^{\mathrm{track}}/p_{\mathrm{T}}^{\mathrm{PFlow\ jet}}$ the p_{T} fraction between track and PFlow jet p_{T}
$\Delta\eta$ (track, Ex$k_t^{(2)}$ track jet)	Pseudorapidity difference between track and Ex$k_t^{(2)}$ jet
$\Delta\phi$ (track, Ex$k_t^{(2)}$ track jet)	Angular difference between between track and Ex$k_t^{(2)}$ jet
d_0	Transverse impact parameter
$z_0 \sin\theta$	Longitudinal impact parameter
S_{d_0}	d_0/σ_0 : transverse IP significance
$S_{z_0 \sin\theta}$	$z_0 \sin\theta/\sigma_{z_0 \sin\theta}$: longitudinal IP significance
PIX1 hits	Number of hits in the first pixel layer
IBL hits	Number of hits in the IBL
Shared IBL Hits	Number of shared hits in the IBL
Split IBL Hits	Number of split hits in the IBL
Shared pixel hits	Number of shared hits in the pixel layers
Split pixel hits	Number of split hits in the pixel layers
Shared SCT hits	Number of shared hits in the SCT
nPixHits	Number of hits in the pixel layers
nSCTHits	Number of hits in the SCT layers
	Secondary Vertex
$\log(m)$	Track mass of the secondary vertex
$\log\, p_{\mathrm{T}}^{\mathrm{frac}}$(vertex, PFlow jet)	$\log\, p_{\mathrm{T}}^{\mathrm{SV}}/p_{\mathrm{T}}^{\mathrm{PFlow\ jet}}$ the p_{T} fraction between the secondary vertex and PFlow p_{T} jet
$\Delta\eta$ (vertex, Ex$k_t^{(2)}$ track jet)	Pseudorapidity difference between the secondary vertex and the Ex$k_t^{(2)}$ jet
$\Delta\phi$ (vertex, Ex$k_t^{(2)}$ track jet)	Angular difference between between the secondary vertex and the Ex$k_t^{(2)}$ jet
L_{xy}	Transverse decay length relative to primary vertex
L_z	Longitudinal decay length relative to primary vertex
$S_{L_{xy}}$	Transverse decay length significance
S_{L_z}	Longitudinal decay length significance

DeXTer with domain adaptation

- An adversarial classifier is added to regulate the feature extractor part of DeXTer
- Adversarial classifier tried to categorizes B-jets originated from a → bb signals and g → bb
- The gradients of the total loss with respect to the feature extractor weights (θ_f)

$$\frac{\partial L}{\partial \theta_f} = \frac{\partial L_D}{\partial \theta_f} - \lambda \frac{\partial L_A}{\partial \theta_f}$$

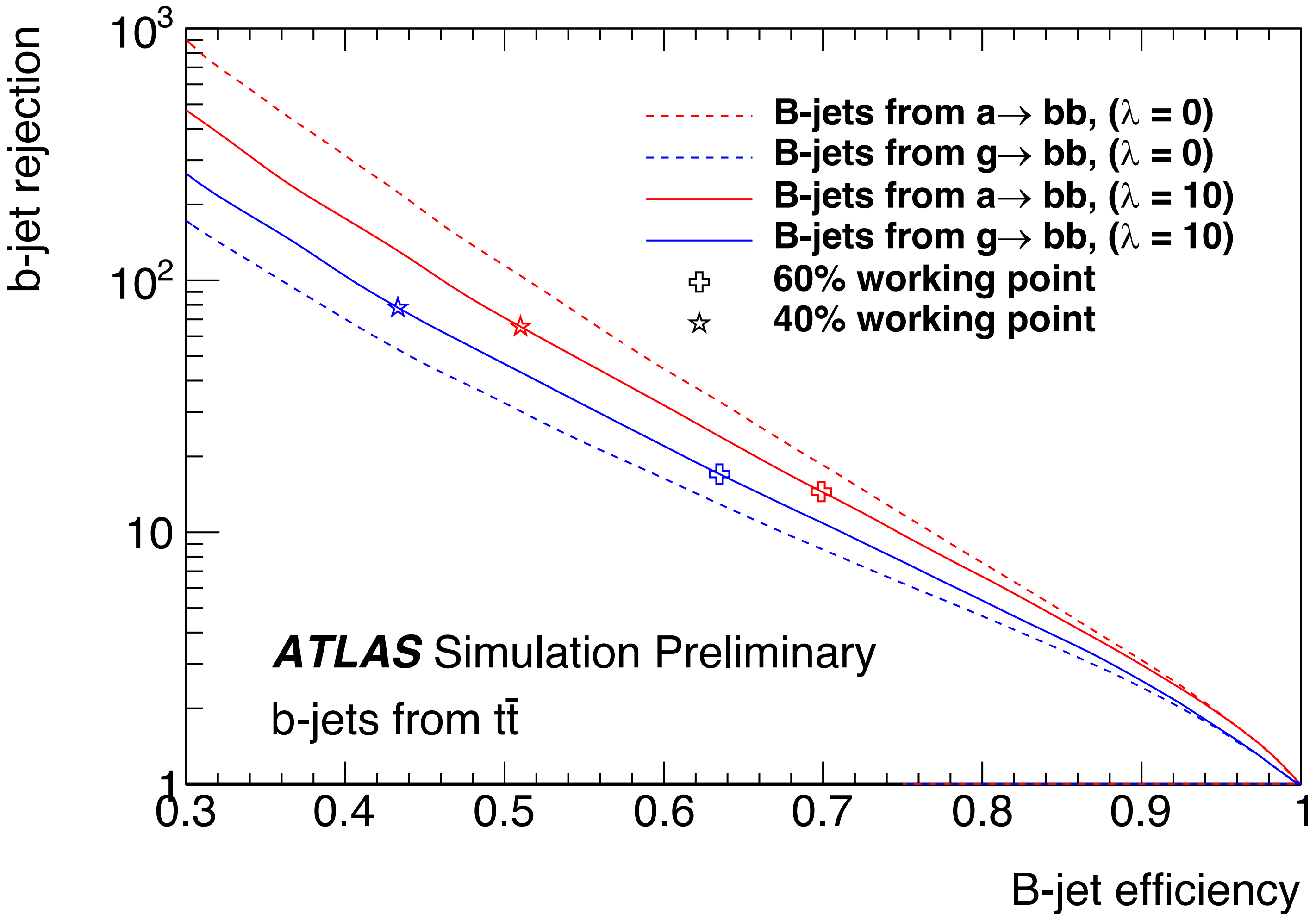


Calibrateability

Tuning λ parameter

- Extra parameter (λ) to tuned the strength of constraint from adversarial classifier

- λ = 10 was found to be the optimum value which has the smallest difference between a → bb and g → bb

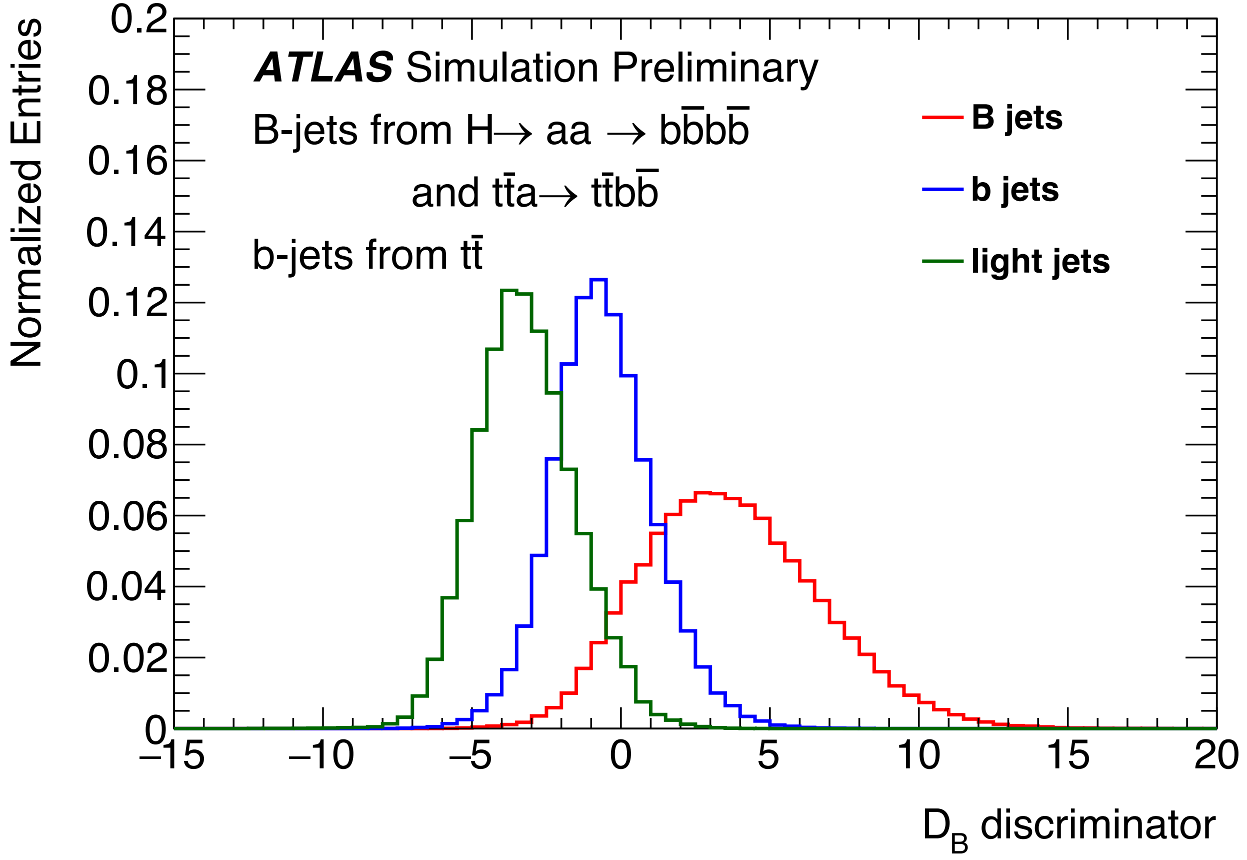


DeXTer discriminant and performance

The class probabilities predicted by the model outputs (p_B , p_b, and p_l), are combined into a B-tagging discriminant:

$$D_B = \ln \frac{p_B}{(1 - f_b)p_l + f_b p_b}$$

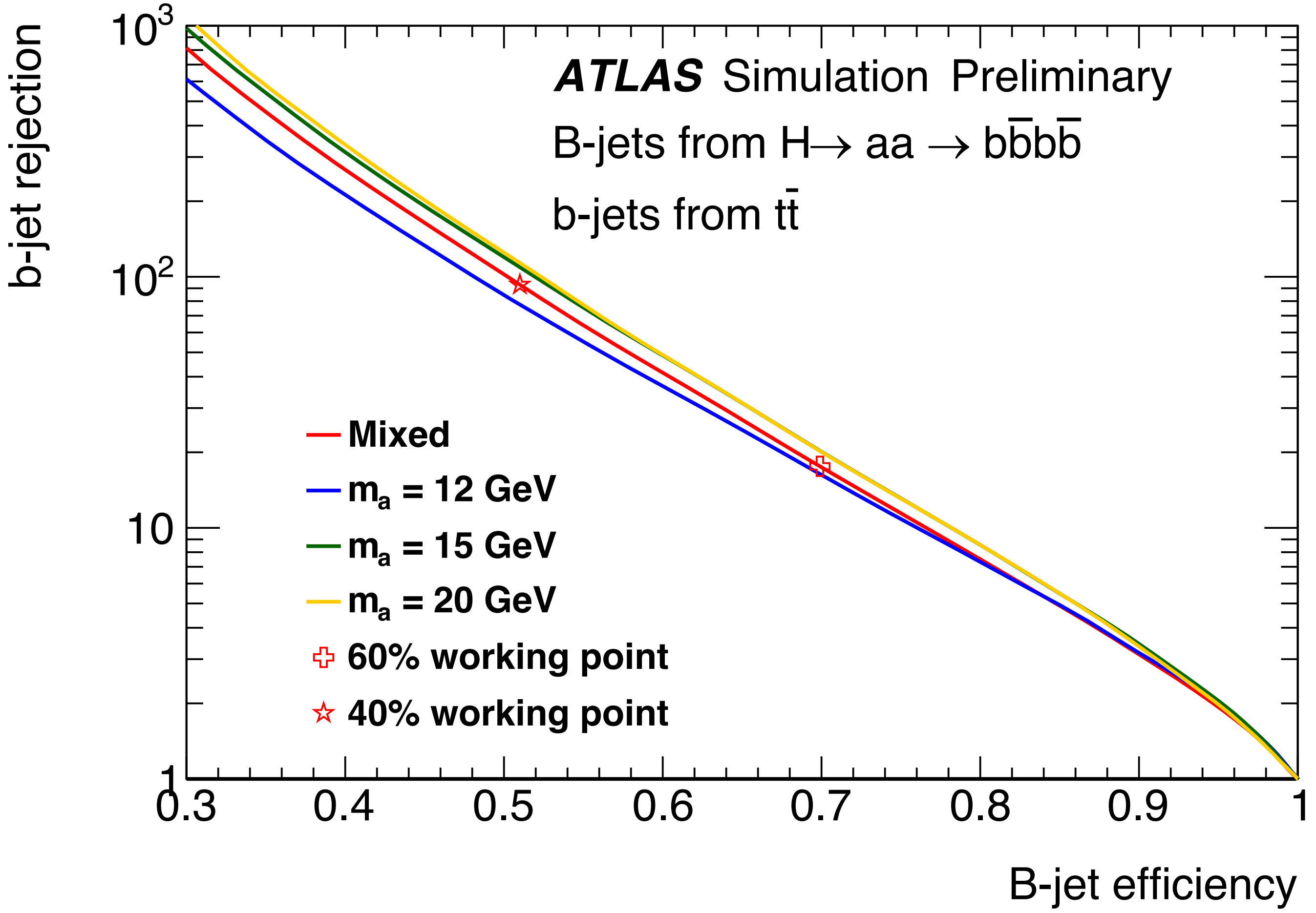
where f_b is a free parameter that balances between the rejection of light-flavor vs b-jets for a given efficiency of selecting b-jets. f_b=0.4 was used in the results



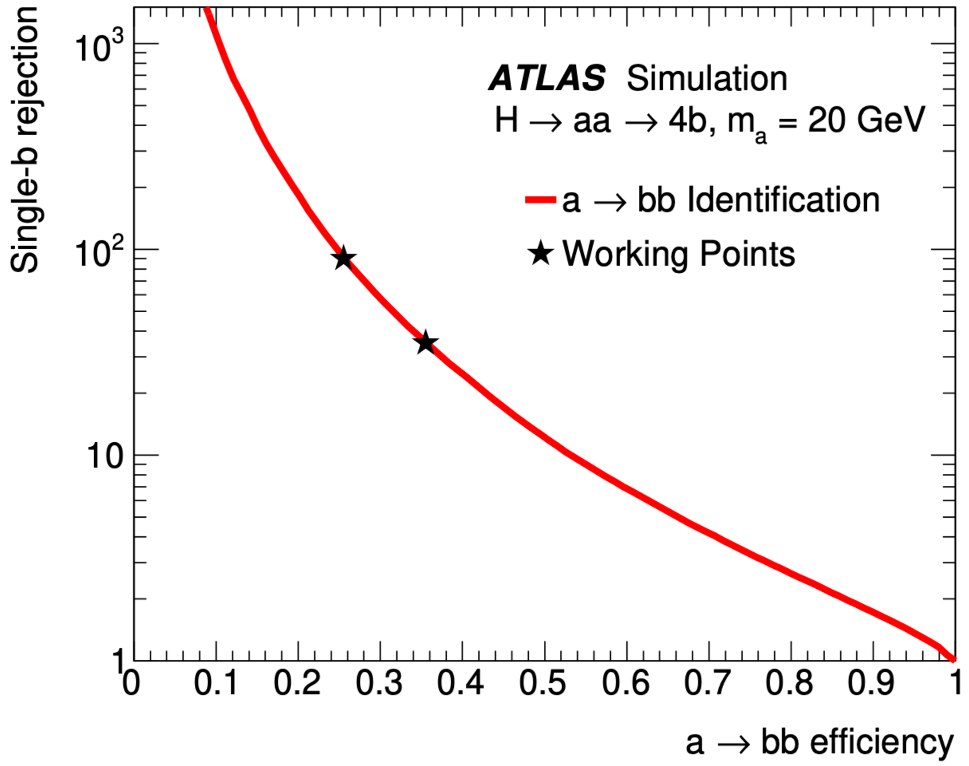
Dependency on Resonance Mass

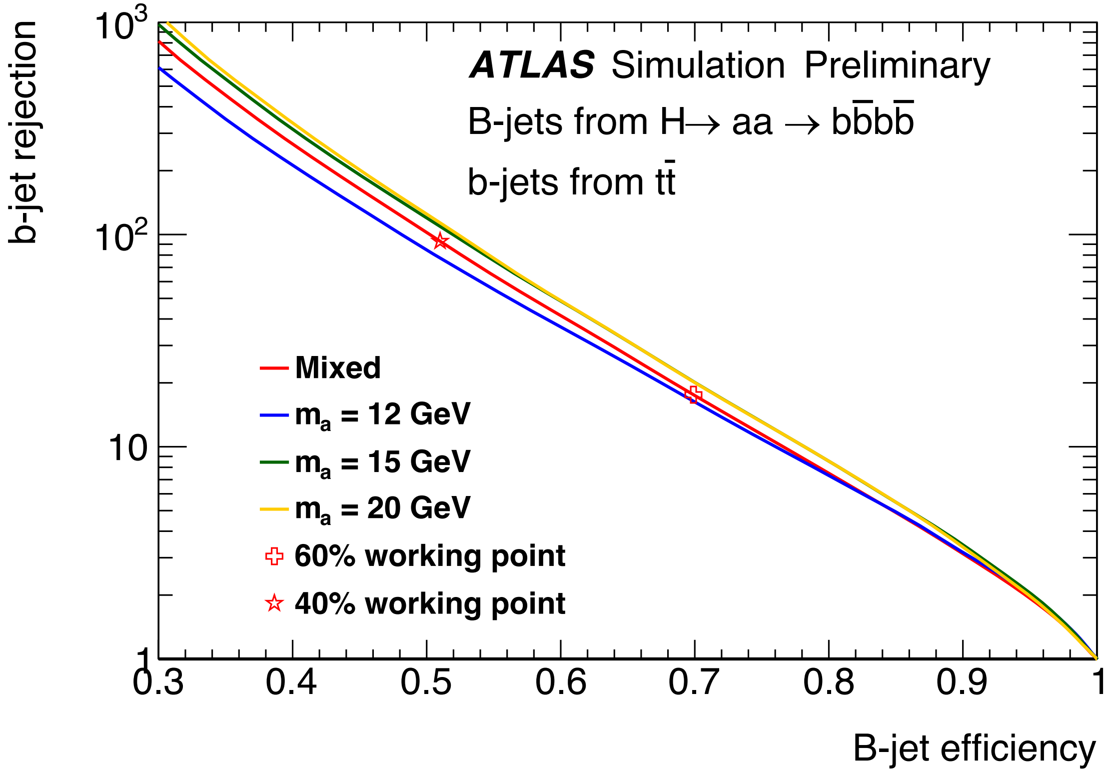
- An ensemble mix of H→aa → b$\bar{b}$b$\bar{b}$ and t$\bar{t}$a, a → b$\bar{b}$ sample with different a boson mass are used as B-labeled jet training sample

- Features are redefinition or drop if it's large different in ROCs was observed between different m_a



Compare with previous BDT Tagger





We achieve much better signal efficiency with the same background rejection!

[Phys. Rev. D 102 (2020) 112006]
[ATL-PHYS-PUB-2022-042]

Efficiency measuremnt for DeXTer
Tagging Interval definition

Efficiency estimated using the Z+jets and $t\bar{t}$ MC as a function of transverse momentum (p_T)

The calibration are provided in 3 jet p_T interval bins

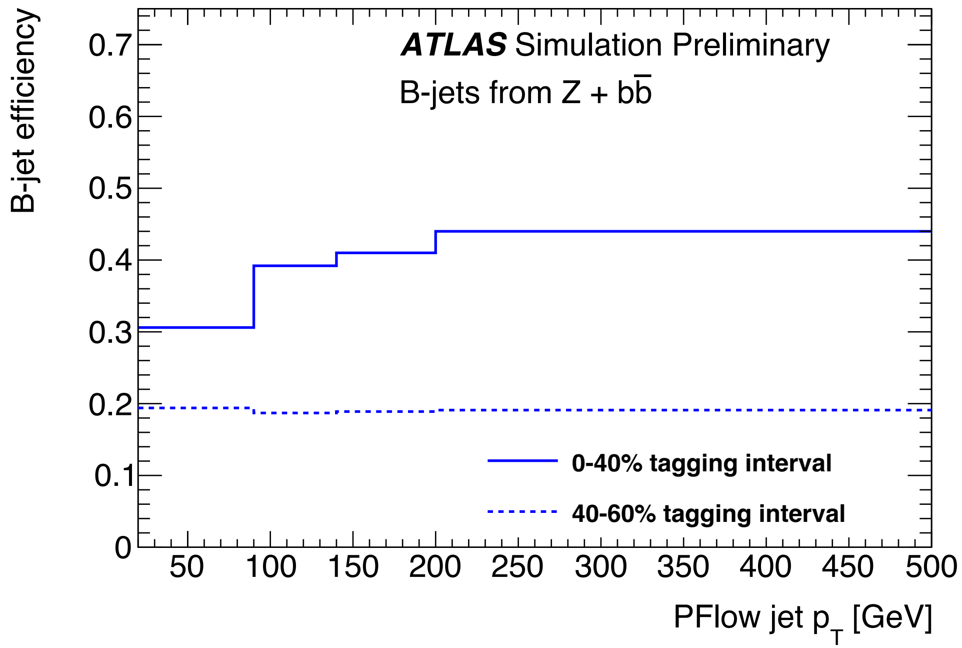
[20, 90], [90, 140], [140, 200] GeV

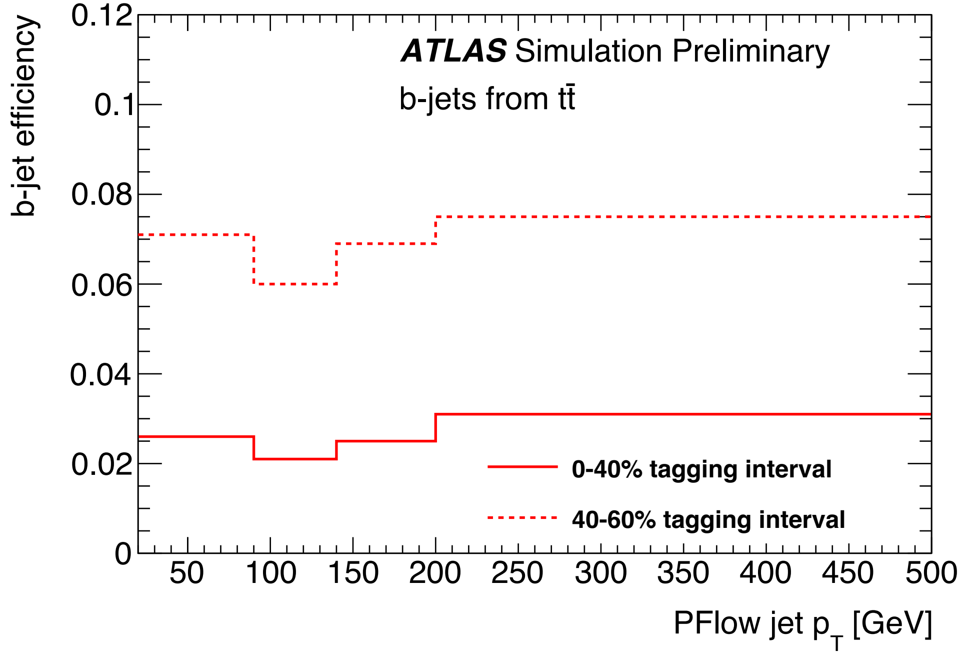
and further split into three tagging intervals are defined

[0 - 40], [40, 60], [60, 100]% tagging interval

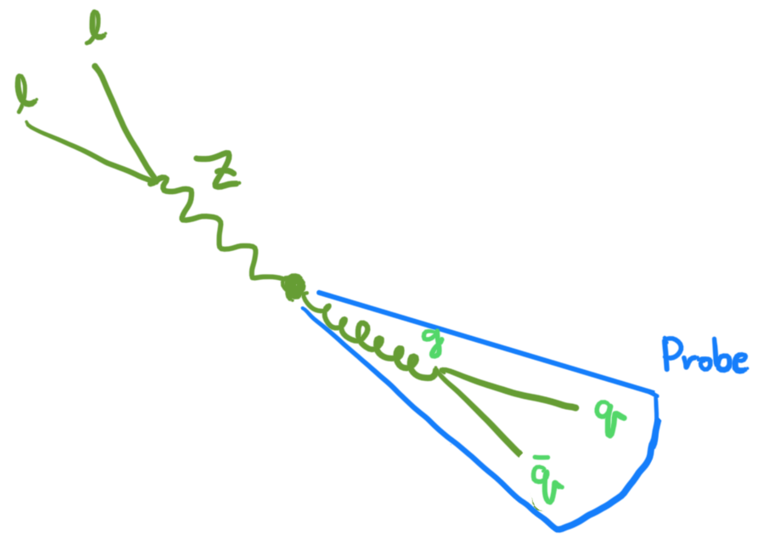
The Scale factor (SF) are defined as

$$SF = \frac{\varepsilon_{Data}}{\varepsilon_{MC}}$$





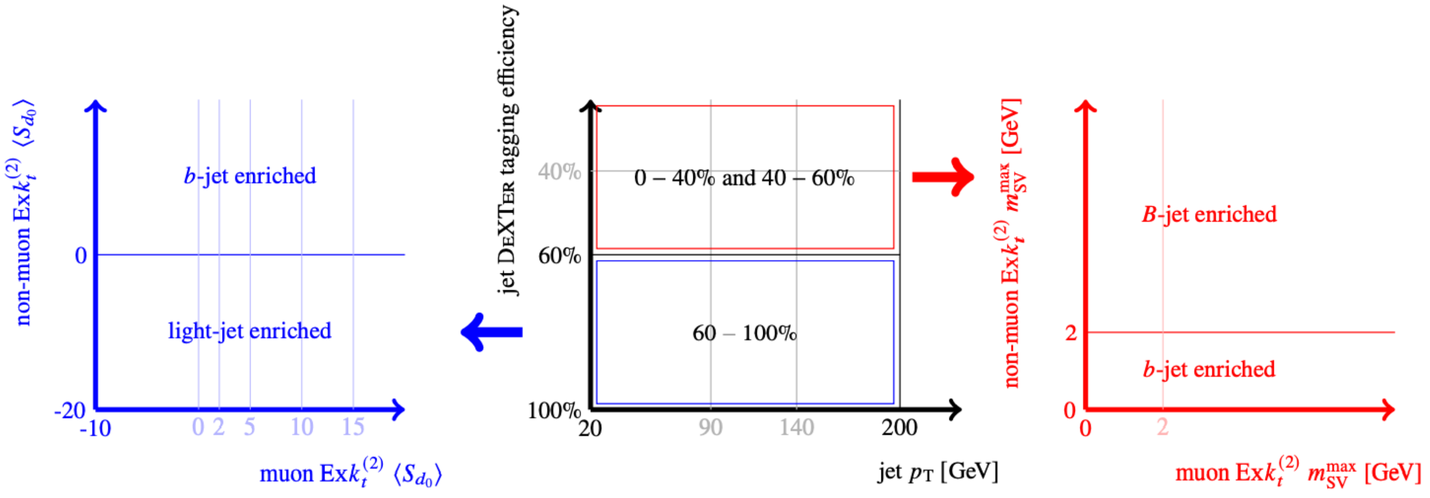
Calibration of B-tagged SFs with Z+jets events



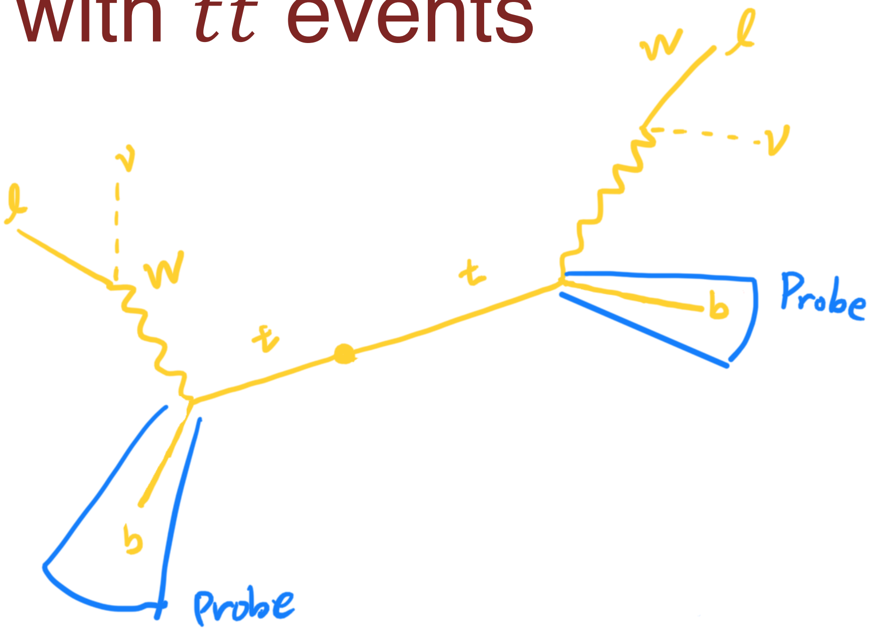
Events are selected with
- Exactly two same flavor and oppsite charge leptons
- Exactly one probe jet with $\Delta R(jet, \mathrm{Z\ or\ lepton}) > 1$

Muon-in-jet tagging is used in the track-subjet to enrich HF fraction from Z+jets events

Flavor-sensitive variables are used to define regions
- 60-100 % tagging interval : $\mathrm{Ex}K_t^{(2)}\ \langle S_{d_0}\rangle$
- 0-40 and 40-60 % tagging intervals : $\mathrm{Ex}K_t^{(2)}\ m_{SV}^{max}$



Calibration of bottom mis-tag rate with $t\bar{t}$ events



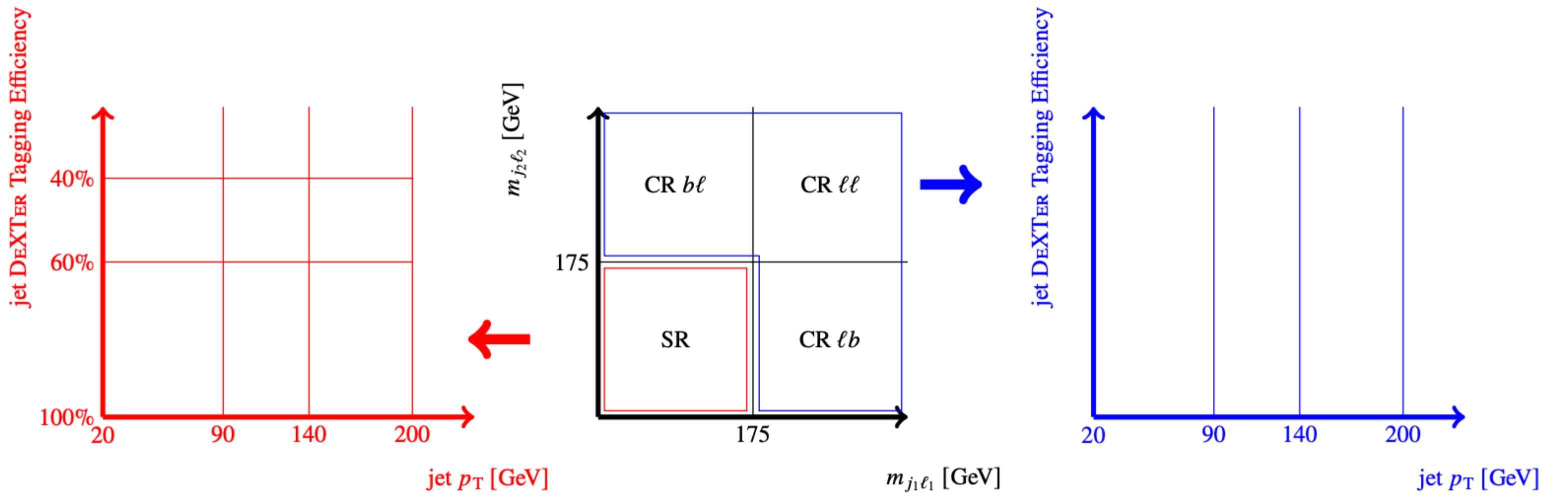
Events are selected with

- Exactly one electron and one muon with opposite sign with $m_{e\mu}$ > 50 GeV
- Exactly two jets with $\Delta R(jet,\ \text{lepton})$ > 0.8

A simple top-quark pair reconstruction is adopted

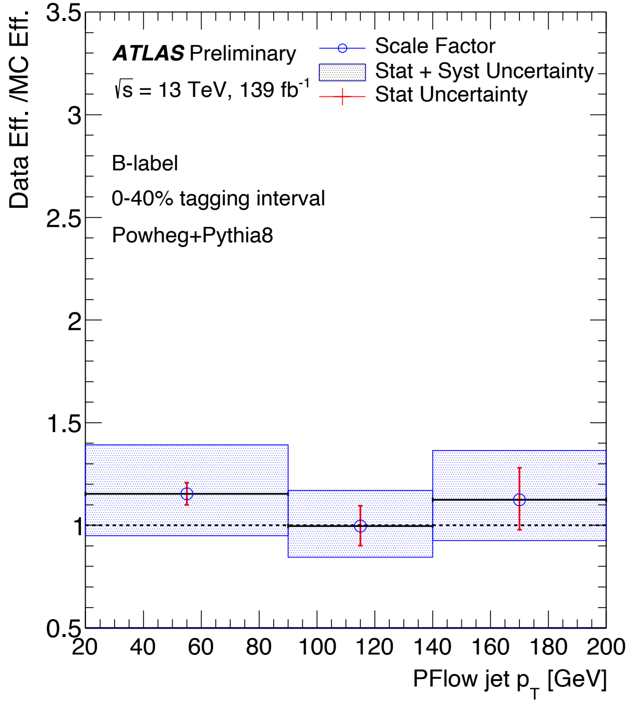
$$\underset{i,j\in\{e,\mu\}}{\text{argmin}}\,(m^2_{j_1,\ell(i)} + m^2_{j_2,\ell(j)}),$$

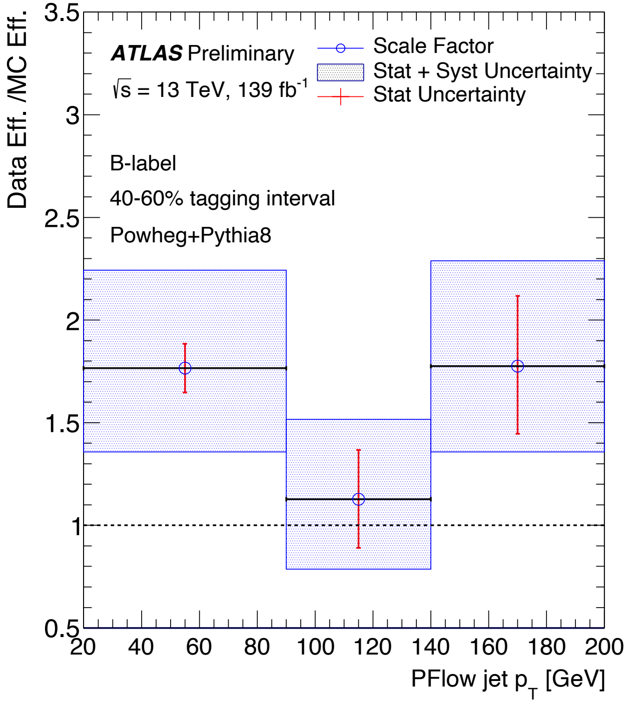
- The signal and control regions are defined using $\boldsymbol{m_{j_1 l}}, \boldsymbol{m_{j_2 l}}$ to better constraint the flavor composition corrections.

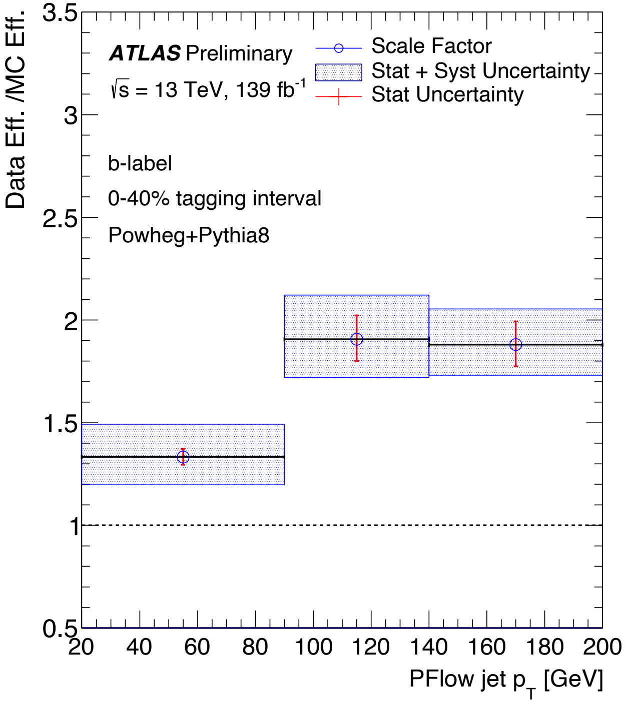


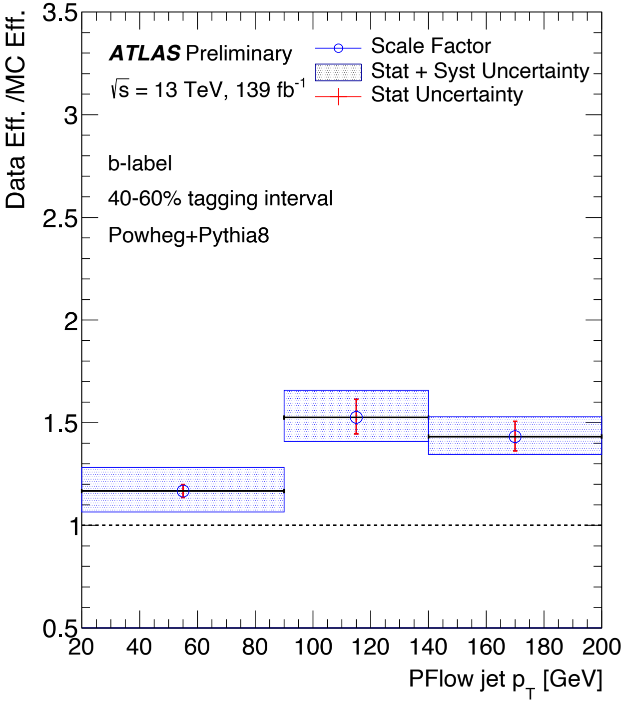
DeXTer Data/MC Scale Factor for analysis

- Z-regions and top-regions are used simultaneously to measure both B-tagging and b mis-tag efficiency in data
- The measurements are mostly systematically dominated which are from MC modeling or extrapolation.









Ready for analysis to use!

Summary

- A novel low-mass double-b tagger, DeXTer, is developed to search for new light resonances.
  - Deep Sets architecture allows us to develop an end-to-end tagger utilizing low-level features which significantly boosts tagger performance
- The simultaneous measurement of the tagging and mis-tag efficiency with data allows to account for the full correlation model in the propagation of uncertainties
  - Easier for analysis to apply the systematic uncertainties for different SFs
- DeXTer opens many new possibilities to probe the phase space that was known to be difficult in the past
  - ZH, H→aa → $b\bar{b}b\bar{b}$

  and more!