Semi-leptonic $t\bar{t}$ Calibration for $X \rightarrow b\bar{b}$ tagger

Qibin Liu [1][2], Changqiao Li [1], Shu Li [1], Shih-Chieh Hsu [2]

[1] Shanghai JiaoTong Univ. & TDLI.

[2] Univ. of Washington, Seattle

Abstract

The identification of massive particles decaying into bottom quark pairs is important for the physics program of the ATLAS experiment at the Large Hadron Collider. A neural network (NN) based double b-tagging algorithm named $X \rightarrow b\bar{b}$ tagger is developed and calibrations of the tagger are performed using proton-proton collision data corresponding to $139fb^{-1}$ collected at a centre-of-mass energy of $\sqrt{s} = 13TeV$.
The technique of $X \rightarrow b\bar{b}$ mis-tag rate calibration [1] is developed based on semi-leptonic decay $t\bar{t}$ events which provide typical non-$b\bar{b}$ flavor combination and high statistics. The mis-tag efficiency is measured and the scale factor, which is defined as the ratio of the mis-tag rate measured in the data over the one in simulation, is found to be in a range of 1~1.1 with uncertainty less than 16%.

$X \rightarrow b\bar{b}$ tagger

The neutral network (NN) based $X \rightarrow b\bar{b}$ tagger[2] developed in ATLAS extends the flavor tagging to the large-*R* jet level and combines the strong b-tagging discriminants of variable-radius (VR) subjets and the kinematics of large-*R* jet. It shows good ability to reject both boosted top quark jets and jets arising from multijet process.

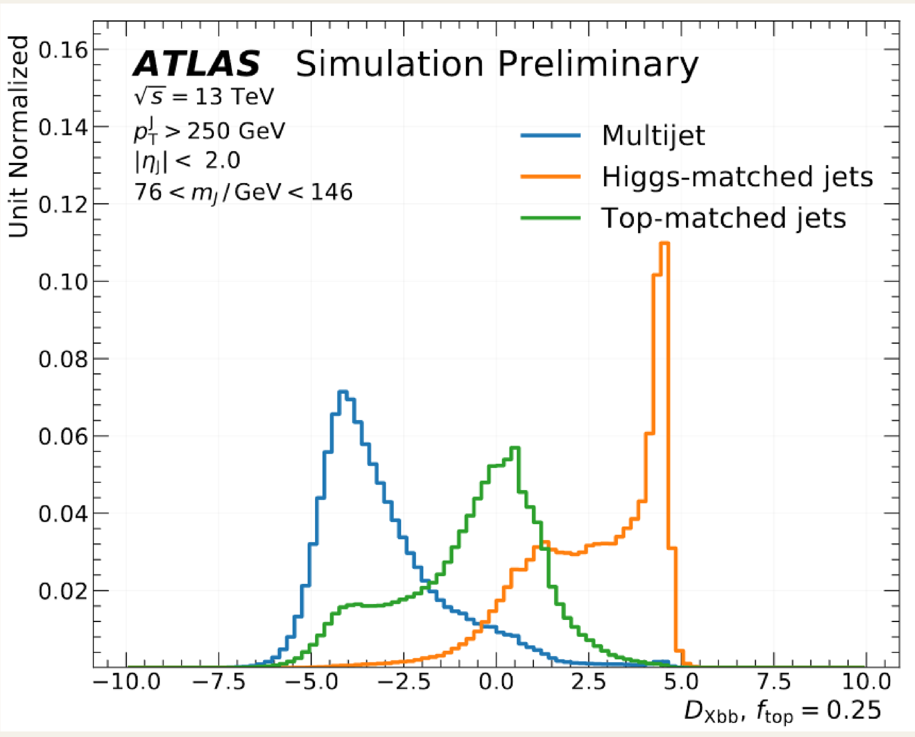

Figure 1: Distribution of $X \rightarrow b\bar{b}$ tagger score

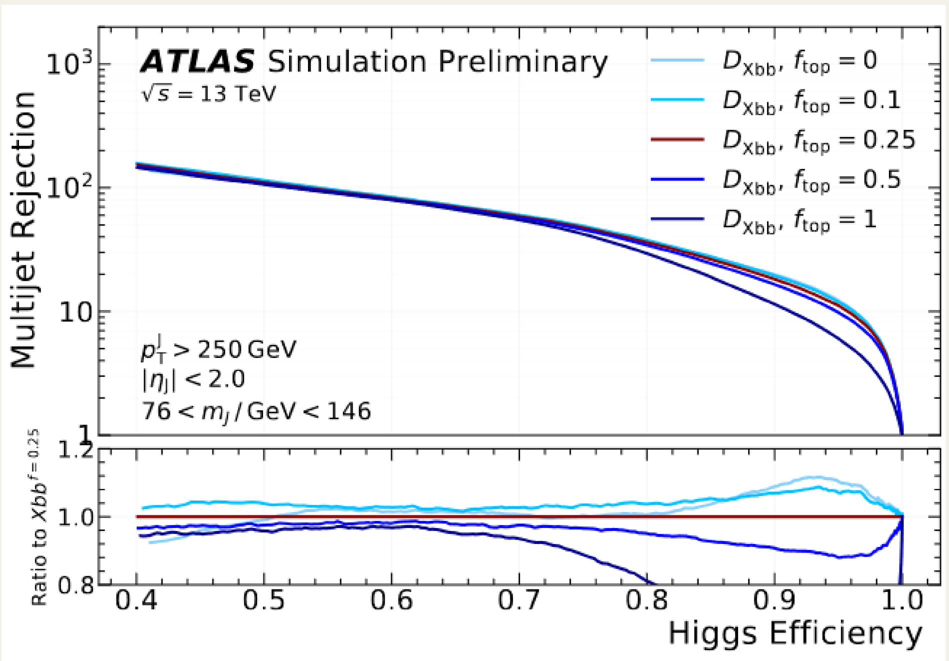

Figure 2: ROC curve of Multijet rejection

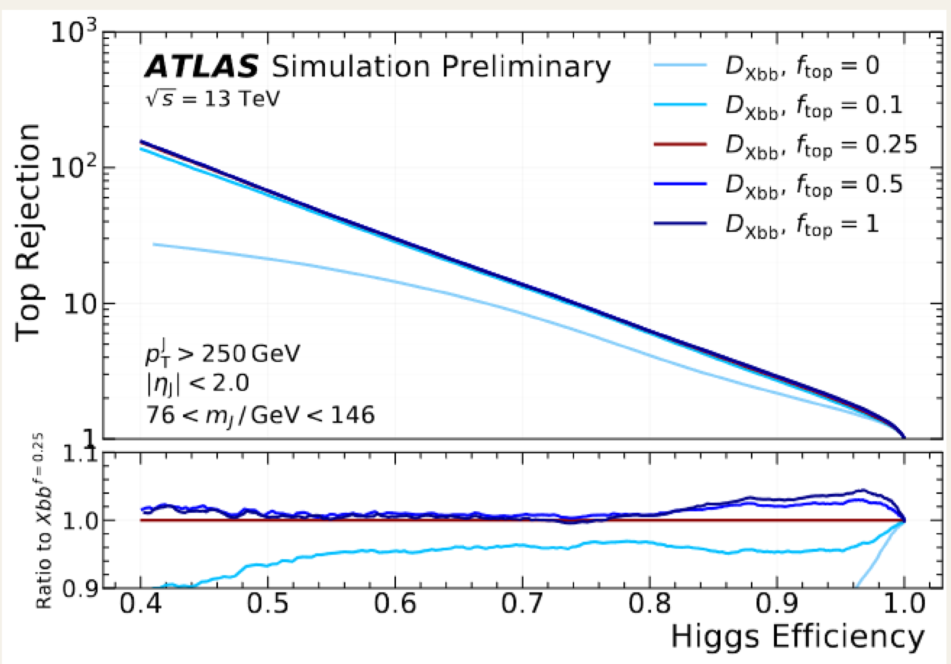

Figure 3: ROC curve of Top jet rejection

Event Selection

Pure semi-leptonic events are selected with the help of the leptonic decay of top quark, where the single lepton is used as trigger and the high transverse momentum requirement reduces the fake. Additional requirements of one-b-tagging and mass of lepton-jet system suppress the *W+jet* background. The requirement of the angle between the probing large-*R* jet and the leptonic side, increases the probability to capture the hadronic decay where the tagging efficiency is measured.

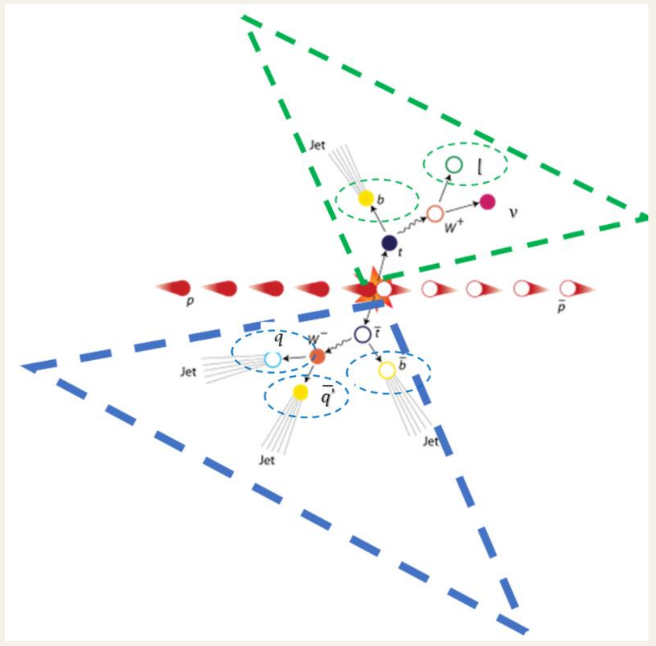

Figure 4: Topology of typical semi-leptonic $t\bar{t}$ events

Calibration Strategy

Normalization correction of the $t\bar{t}$ modelling and data efficiency (or scale factor) on top are extracted simultaneously with fitting on the large-*R* jet mass, in 4 regions covering large-*R* jet p_T ranging from 300GeV-1000GeV.

$$\begin{cases} N_{pass}^{data} = SF * f^{t\bar{t}} * N_{pass}^{t\bar{t},MC} + N_{pass}^{non-t\bar{t},MC} \\ N_{fail}^{data} = SF' * f^{t\bar{t}} * N_{fail}^{t\bar{t},MC} + N_{fail}^{non-t\bar{t},MC} \end{cases}$$

$$\begin{cases} SF' \equiv \frac{1-\epsilon^{data}}{1-\epsilon^{MC}} = \frac{1-\epsilon^{MC} * SF}{1-\epsilon^{MC}} \\ f^{t\bar{t}} \equiv \frac{N^{t\bar{t},data}}{N^{t\bar{t},MC}}, \epsilon^{MC} \equiv \frac{N_{pass}^{t\bar{t},MC}}{N^{t\bar{t},MC}} \end{cases}$$

Equation 1: Formulas used in the fitting

Systematic Uncertainties

Dominant systematic uncertainties from modelling of $t\bar{t}$, especially the parton showering and initial/final state radiation and up to 16% in high p_T region for medium working point of $X \rightarrow b\bar{b}$ tagger

p_T [GeV]	300 – 400	400 – 500	500 – 600	600 – 1000
SF	1.06	1.08	1.14	0.99
Total unc.	0.045	0.10	0.11	0.16
Statistic unc.	0.018	0.029	0.046	0.06
Systematic unc.	0.041	0.095	0.095	0.15
$t\bar{t}$ modelling	0.039	0.094	0.088	0.14
$t\bar{t}$ PS	<0.001	0.002	0.003	0.002
$t\bar{t}$ FSR	0.022	0.075	0.036	0.093
$t\bar{t}$ ISR	0.031	0.055	0.078	0.11
$t\bar{t}$ generator	<0.001	<0.001	<0.001	<0.001
$t\bar{t}$ PDF	0.01	0.015	0.019	0.022
$t\bar{t}$ cross-section	–	<0.001	<0.001	<0.001
Single-top modelling	0.007	0.009	0.020	0.023
Single-top Wt DR vs DS	0.005	0.007	0.014	0.015
Single-top PS	<0.001	0.002	0.007	0.015
Single-top generator	0.004	–	0.011	0.002
Single-top cross-section	0.003	0.002	0.003	0.003
W + jets (scale, cross-section)	0.004	0.003	0.004	0.005
Small-*R* jet energy	0.008	0.011	0.022	0.016
Large-*R* jet energy and mass	0.004	0.008	0.014	0.008
Small-*R* jet Flavour tagging related	0.001	0.001	0.001	0.002
Others	0.003	0.004	0.004	0.006

Table 1: Summary of Systematic Uncertainties

Results

The mis-tag efficiency of $X \rightarrow b\bar{b}$ tagger in data is measured with semi-leptonic $t\bar{t}$ events and the data-to-Monte-Carlo scale factors are derived as a function of large-*R* jet p_T and found to be in the range between 1.1 ± 0.12 to 1.0 ± 0.16.

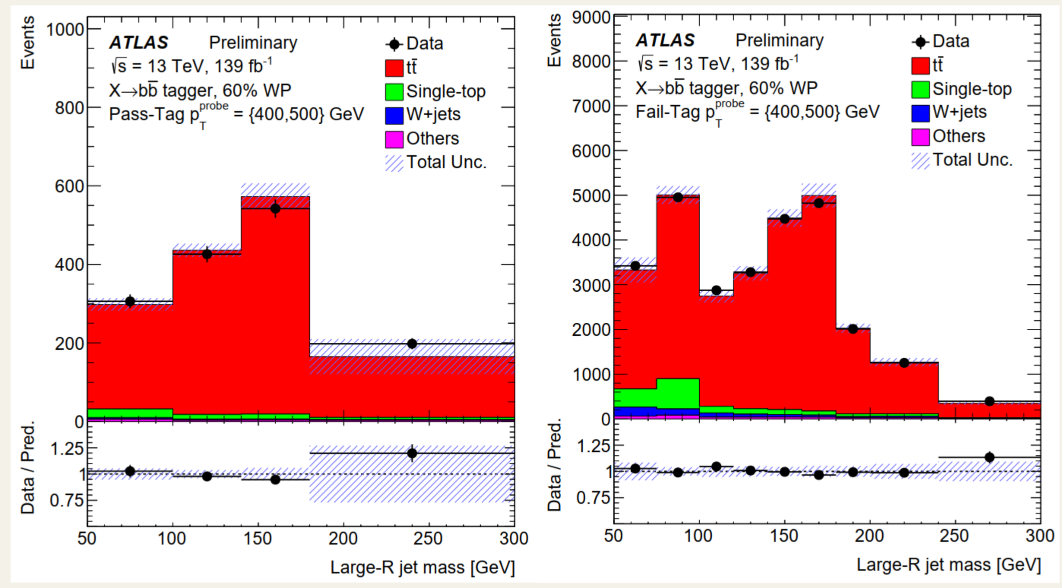

Figure 5: Post-fit for pass- and fail-tag in one p_T region

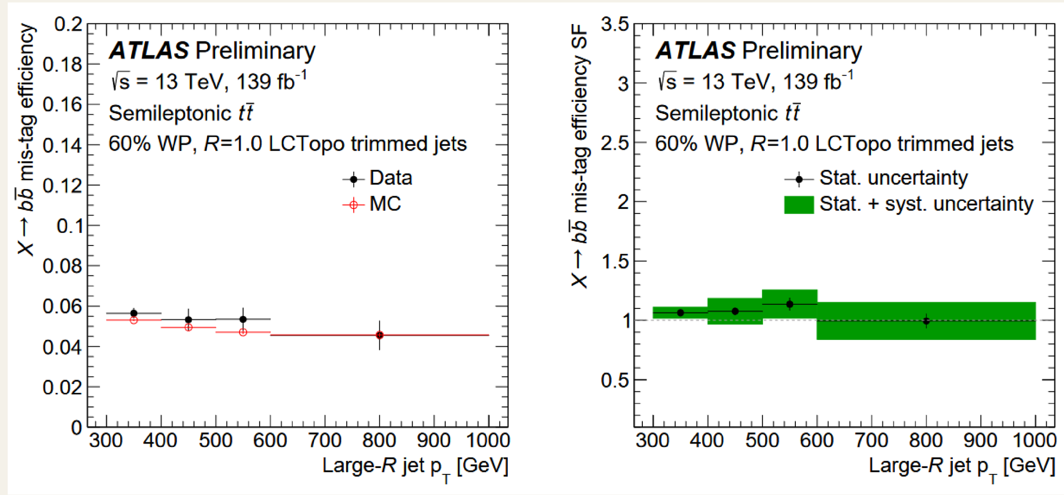

Figure 6: $X \rightarrow b\bar{b}$ mis-tag eff. in data/MC and scale factor

Other Calibrations

The signal efficiency of $X \rightarrow b\bar{b}$ tagger is calibrated with another dedicated $Z(\rightarrow b\bar{b}) + jets$ and $Z(\rightarrow b\bar{b})\gamma$ analysis. And the modelling of large-*R* jet kinematics in Monte Carlo simulation, after the application of $X \rightarrow b\bar{b}$ tagger, is checked in another calibration using multijet events enriched in $g \rightarrow b\bar{b}$ splitting. The details of the two calibrations are documented in reference [1].

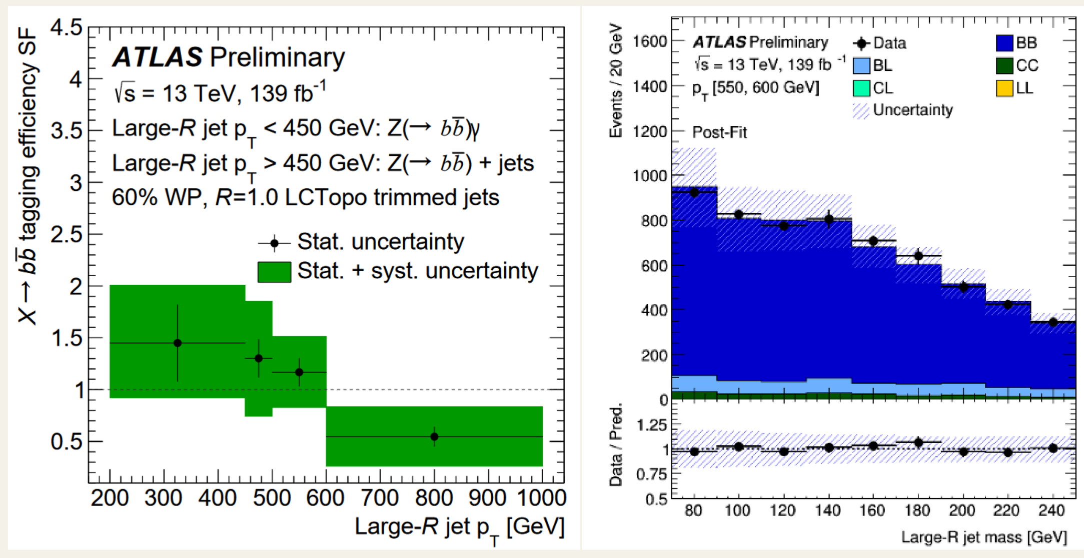

Figure 7: $X \rightarrow b\bar{b}$ signal eff. scale factor in $Z \rightarrow b\bar{b}$ calibration(left) and large-*R* jet mass distribution after tagging in $g \rightarrow b\bar{b}$ calibration(right)

Related Publications

[1] ATLAS Collaboration, Efficiency corrections for a tagger for boosted $H \rightarrow b\bar{b}$ decays in pp collisions at $\sqrt{s} = 13TeV$ with the ATLAS detector, ATL-PHYS-PUB-2021-035, 2021

[2] ATLAS Collaboration, Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS, ATL-PHYS-PUB-2020-019, 2020

qibin.liu@cern.ch