Alternative Higgs Tagger Group Approval

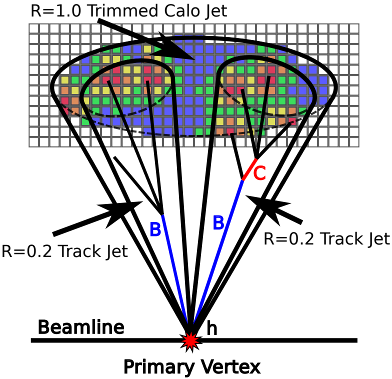
Nikola Whallon, Nurfikri Norjoharuddeen, Dan Guest, Qi Zeng, Jia Yu, Chunhui Chen, Shih-Chieh Hsu, Sam Meehan, Soeren Prell

University of Washington

April 11, 2017, Flavour Tagging Meeting

Introduction

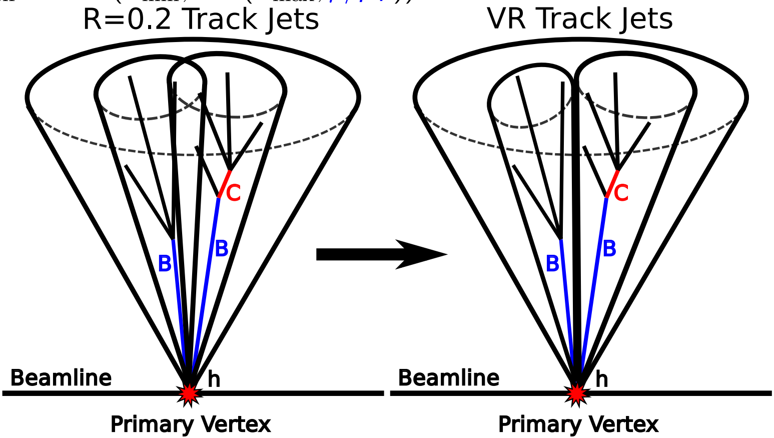
- part of $X \to b\bar{b}$ is working on advanced Higgs tagging at very high p_T
- nominal Higgs tagging is illustrated here - this method fails when the fixed radius (FR) track jets merge at high Higgs p_T
- the advanced taggers we developed fix this issue in different ways
- we are aiming to publish the new taggers in a pubnote



cartoon not for approval

Variable Radius (VR) Track Jets

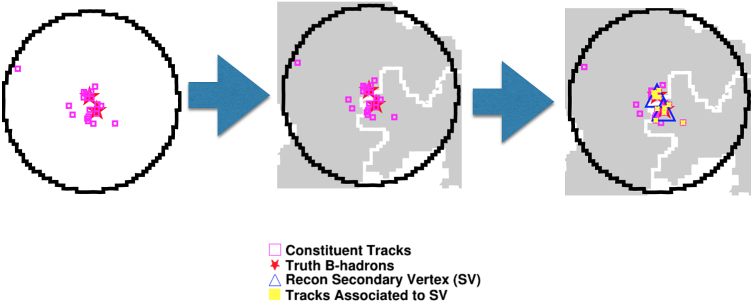
cluster anti-k_T track jets using
$R_{\mathrm{eff}} = \max(R_{\min}, \min(R_{\max}, \rho/p_T))$



cartoon not for approval

Exclusive k_T Subjets

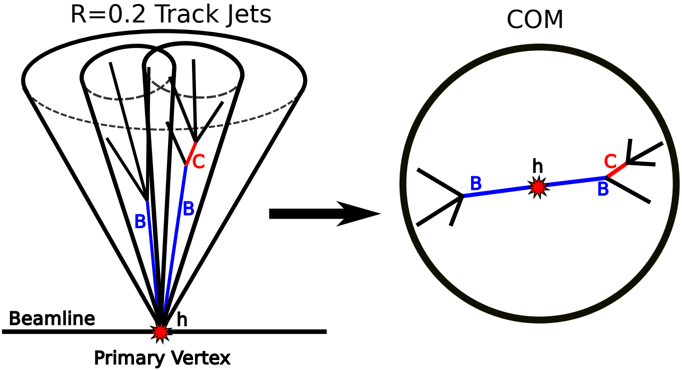
- use Higgs jet constituents to cluster k_T jet
- undo the last clustering step to form exactly 2 subjets
- we are trying three approaches:
  - use untrimmed large-R jet constituents (ExKt (Untrimmed))
  - use trimmed large-R jet constituents (ExKt (Trimmed))
  - use ghost associated tracks as constituents (ExKt Trackjets)



cartoon not for approval

COM Subjets

- boost to the Higgs jet COM
- use Higgs jet constituents to cluster 2 EECambridge subjets
- boost back to the lab frame and apply *b*-tagging
- if EECambridge y_{cut} parameter large enough, always get 2 subjets
- y_{cut} also controls track selection for *b*-tagging



Pubnote Structure

twiki link
cds link

- Introduction
- Monte Carlo Simulation Samples
- Event Reconstruction and Selections
- Subjet Tagging Algorithms
  - $R = 0.2$ trackjets - description
  - VR trackjets - description and ρ optimization plots
  - ExKt subjets - description and track vs calo comparisons
  - COM subjets - description and y_{cut} optimization plots
- Results
  - performance plots without b-tagging
  - performance plots with b-tagging
- Conclusion
- Auxiliary Materials

We are requesting approval for all following tables and figures.

Samples and Selections

Signal: $G \to hh \to b\bar{b}b\bar{b}$ DSIDs 301488-301507, 305776-305780
QCD: DSIDs 361023-361032
$t\bar{t}$: DSIDs 303722-303726

- fatjet collection: AntiKt10LCTopoTrimmedPtFrac5SmallR20
- fatjet selection: $p_T > 250$ GeV, $|\eta| < 2.0$, 76 GeV $< m <$ 146 GeV
- fatjet signal: $== 1$ h && $== 2$ b-hadron $\Delta R < 1.0$ to parent
- signal and top fatjets reweighted to QCD fatjet p_T spectrum
- trackjet selection: $p_T > 10$ GeV, $|\eta| < 2.5$, $\#tracks > 1$
- other subjet selection: $p_T > 5$ GeV, $|\eta| < 2.5$
- subjet truth matching: $\Delta R(\mathrm{subjet}, b{-}\mathrm{hadron}) < 0.3$ (excl)
- VR trackjets ghost-assoc. to untrimmed parent
- FR/VR/ExKt subjets use nominal track-to-jet associator
- COM subjets use special track-to-jet associator

Definitions

Truth double b-tagging efficiency: the efficiency to select a fatjet whose leading 2 subjets are truth matched to b-hadrons

Single b-tagging efficiency: the efficiency to select a fatjet with at least 1 of the leading 2 subjets passing an MV2c10 cut

Double b-tagging efficiency: the efficiency to select a fatjet whose leading 2 subjets pass an MV2c10 cut

Table 1: Signal Samples

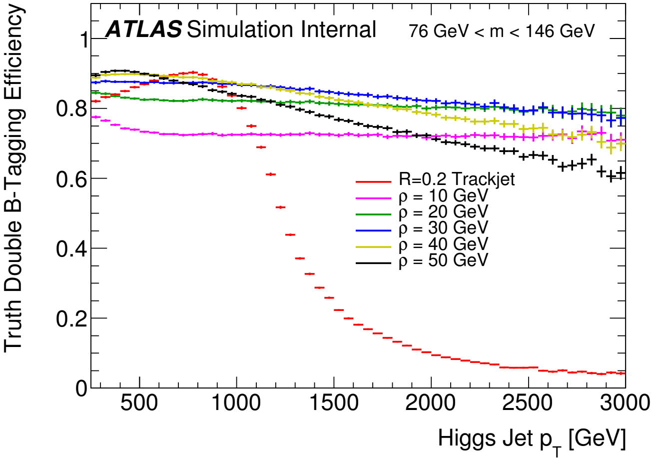
m_{G^*} [GeV]	N_{events}	m_{G^*} [GeV]	N_{events}	m_{G^*} [GeV]	N_{events}
300	79800	1100	99800	2250	99800
400	99800	1200	99800	2500	60000
500	94400	1300	19800	2750	59600
600	99800	1400	99600	3000	78000
700	54800	1500	99400	4000	100000
800	70000	1600	99800	4500	99000
900	83000	1800	15000	5000	99000
1000	10000	2000	89800	6000	99000

- it was requested to include explicitly which signal samples we use due to an issue with sample dependent performance

Table 1 Caption

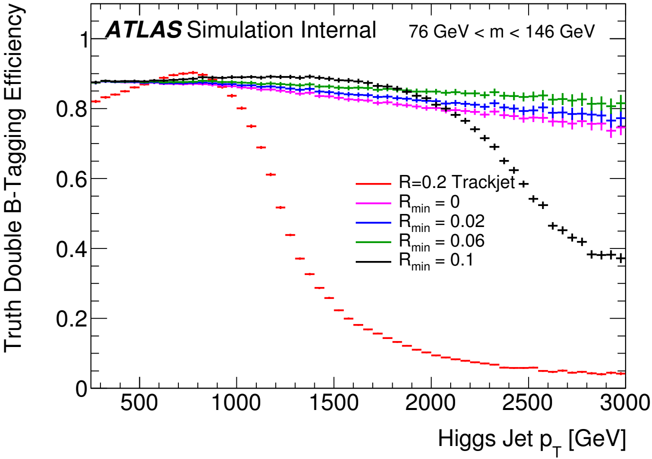
The graviton mass value (m_{G^*}) and the number of simulated events (N_{events}) for each MC signal graviton sample.

Fig. 1a: VR ρ Optimization



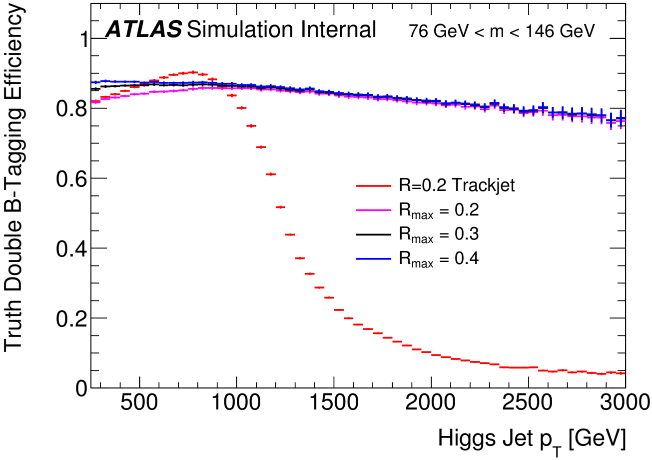
- VR trackjets maintain good performance at high p_T
- $\rho = 30$ GeV seems to be the best choice

Fig. 1b: VR $R_{\min}$ Optimization



- $R_{min} = 0.06$ seems to be the best choice
- we went with $R_{min} = 0.02$ in order to be close to 0

Fig. 1c: VR R_{max} Optimization

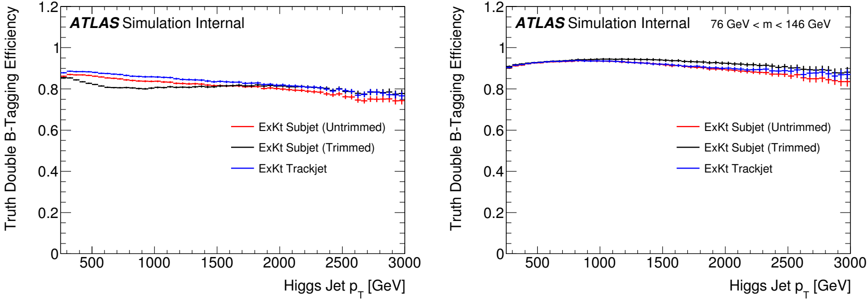


- $R_{max} = 0.4$ seems to be the best choice

Fig. 1 Caption

Truth double b-tagging efficiency of a Higgs jet as a function of the Higgs jet p_T. (a) The efficiency for VR trackjets with $R_{min} = 0.02$ and $R_{max} = 0.4$ for several ρ values. (b) The efficiency for VR track jets with $\rho = 30$ GeV and $R_{max} = 0.4$ for different values of R_{min}. (c) The efficiency for VR trackjets with $\rho = 30$ GeV and $R_{min} = 0.02$ for varying values of R_{max}. The efficiency for R $= 0.2$ trackjets are also included all the plots. Statistical errors are present.

Fig. 2: ExKt Truth Double B-Tagging

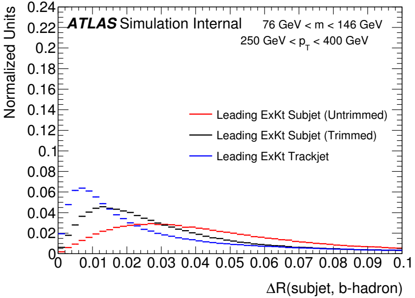


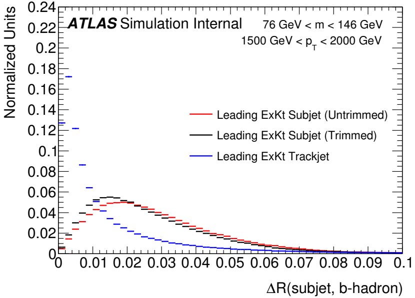
- before a mass window is applied, ExKt trackjets perform the best
- after a mass window is applied, ExKt (Trimmed) performs the best

Fig. 2 Caption

Probability that both leading and sub-leading subjets have exactly one b-hadron for different exclusive-k_T variations, as function of Higgs jet p_T, before (left) and after (right) Higgs boson mass cut $76 < m_H < 146\mathrm{GeV}$. Statistical errors are present.

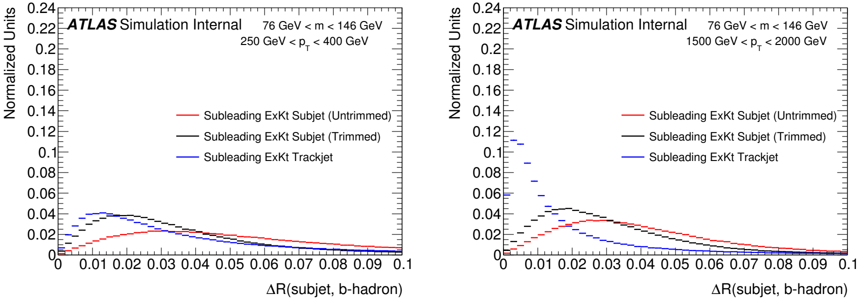
Fig. 3 (Top): ExKt ΔR(leading subjet, b−hadron)





- the trackjets have a better axis reconstruction than the calo jets
- the resolution differences are much smaller than the track gathering radius for b-tagging (0.24), so this does not result in significant b-tagging performance differences

Fig. 3 (Bot.): ExKt ΔR(subleading subjet, b−hadron)

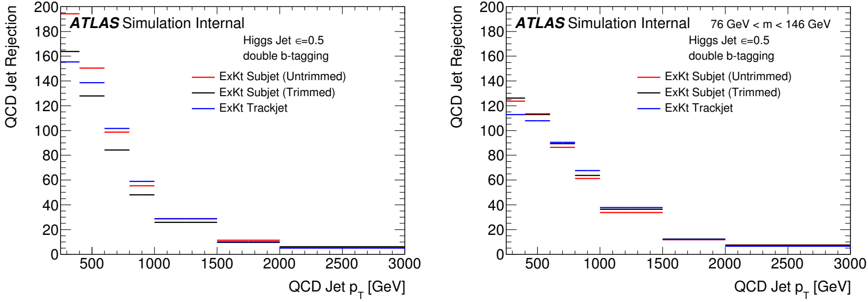


- similar conclusions as the previous slide

Fig. 3 Caption

ΔR distribution between reconstructed subjet axis and associated b-hadron flight direction for low and high p_T regions for leading and subleading exclusive-k_T subjets. Statistical errors are present.

Fig. 4: ExKt QCD Double B-Tagging Rej. vs p_T

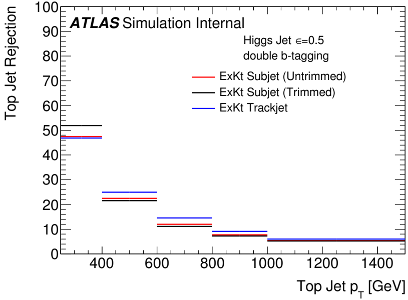


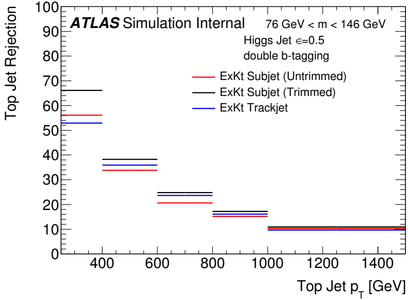
- after the mass window is applied, ExKt (Trimmed) performs the best - it was chosen to be the optimal configuration

Fig. 4 Caption

Rejection against QCD jet background as function of Higgs jet p_T for a fixed Higgs jet efficiency of 50% both with and without a masscut. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 5: ExKt Top Double B-Tagging Rej. vs p_T



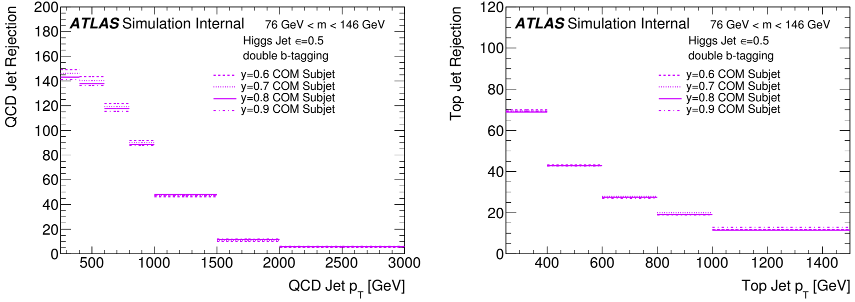


- after the mass window is applied, ExKt (Trimmed) performs the best - it was chosen to be the optimal configuration

Fig. 5 Caption

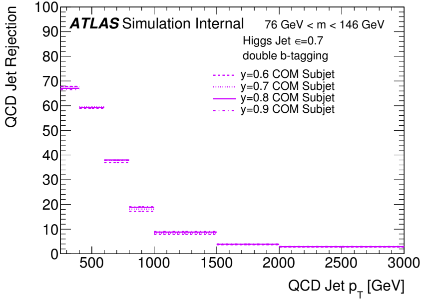
Rejection against top jet background as function of Higgs jet p_T for a fixed Higgs jet efficiency of 50% both with and without a masscut. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 6 (Top, $\epsilon = 0.5$): COM Double B-Tagging Rej. vs p_T



- all of our choices for $y_{\rm cut}$ perform similarly
- the small differences in certain bins are pointed out in the note

Fig. 6 (Bot., $\epsilon = 0.7$): COM Double B-Tagging Rej. vs p_T

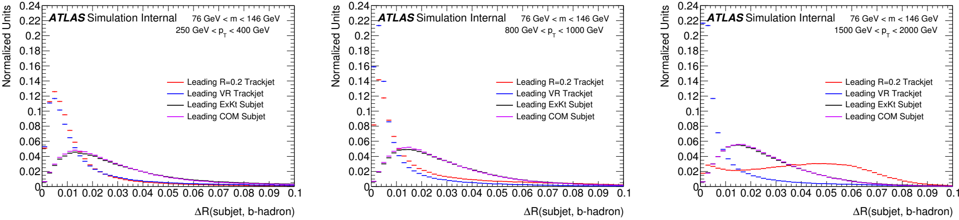


- all of our choices for y_{cut} perform similarly
- the small differences in certain bins are pointed out in the note

Fig. 6 Caption

Double b-tagging background rejection at fixed signal efficiency of 50%(top) and 70%(bottom) respectively against the background of QCD jets (left) and large-R calorimeter jet initiated from a top-quark decay products (right). The track-to-subjet association cone size parameter, y_{cut}, in CoM method is studied. Values of y_{cut} are varied from 0.6 – 0.9 with a gap of 0.1. At the signal efficiency of 50% $y_{cut} = 0.6$ performs better in terms of background rejection at $p_T < 800\mathrm{GeV}$. At higher efficiency of 70%, $y_{cut} = 0.9$ is better at high p_T. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 7: ΔR(leading subjet, $b-$hadron)

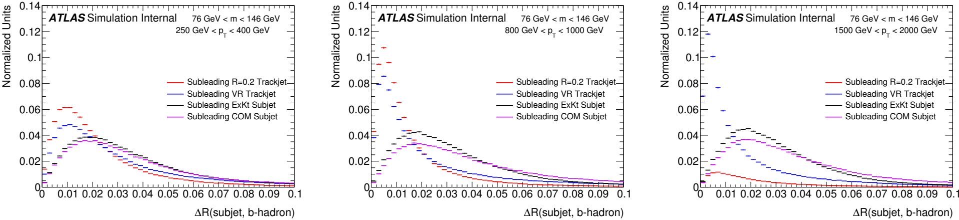


- at high p_T, the large tail for $R = 0.2$ trackjets indicates the b-hadron is poorly reconstructed
- trackjets in general have a better resolution here than calo jets

Fig. 7 Caption

Distributions of the ΔR between leading subjets and matched truth b-hadrons for three different Higgs jet p_T bins. Statistical errors are present.

Fig. 8: ΔR(subleading subjet, $b-$hadron)

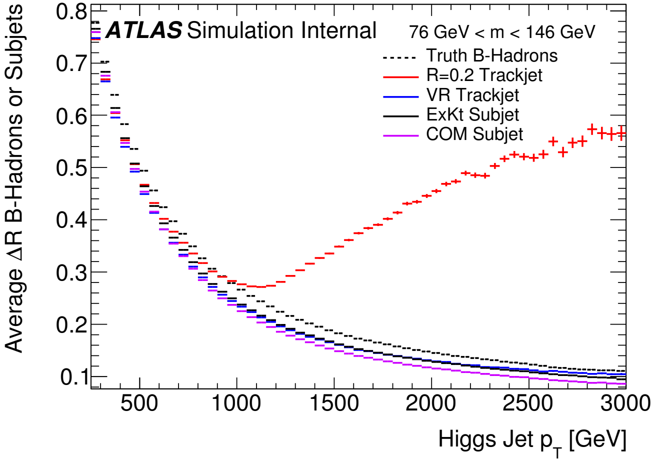


- trackjets in general have a better resolution here than calo jets, but this does not show up in *b*-tagging performance plots because the track collection radius is large (0.24)
- $R = 0.2$ trackjets behave fine here, but have a very small normalization

Fig. 8 Caption

Distributions of the ΔR between subleading subjets and matched truth b-hadrons for three different Higgs jet p_T bins. Statistical errors are present.

Fig. 9: $\Delta R(\text{leading subjet}, \text{subleading subjet})$

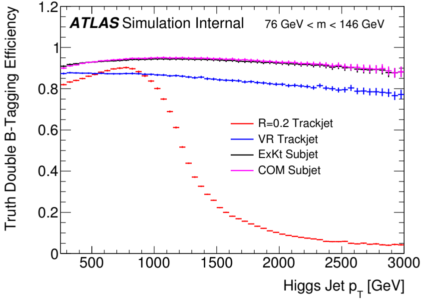


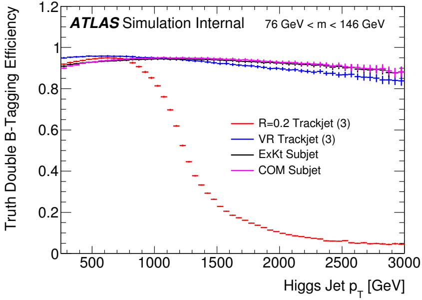
- $R = 0.2$ trackjets deviate for $p_T > 1000$ GeV - the other techniques do not

Fig. 9 Caption

The ΔR between the two leading truth b-hadrons or subjets associated to Higgs jets as a function of Higgs jet p_T. Statistical errors are present.

Figs. 10 and 11: Truth Double B-Tagging





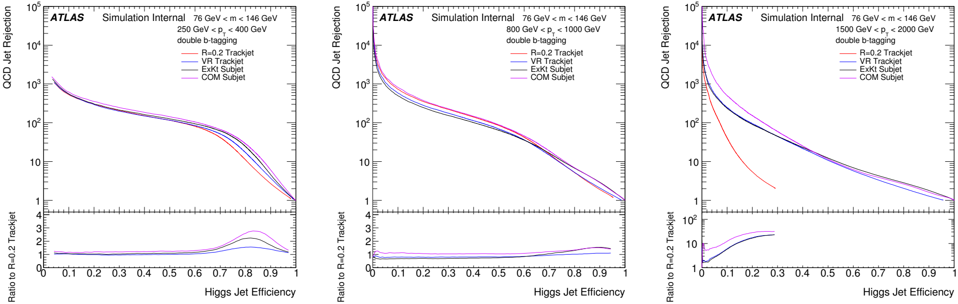
- for the nominal truth double *b*-tagging definition (left), the alternative techniques perform better than $R = 0.2$ trackjets for almost all p_T
- for $600\ \mathrm{GeV} < p_T < 900\ \mathrm{GeV}$, VR trackjets perform worse than $R = 0.2$ trackjets unless you consider the subsubleading subjet (right)

Fig. 10 Caption and Figure 11 Caption

Figure 10 Caption:
The efficiency for a Higgs jet whose two leading associated subjets are matched to truth b-hadrons vs Higgs jet p_T. Statistical errors are present.

Figure 11 Caption:
The efficiency for a Higgs jet to have two of the leading three associated subjets matched to truth b-hadrons vs Higgs jet p_T. Statistical errors are present.

Fig. 12: QCD Double B-Tagging ROC Curves

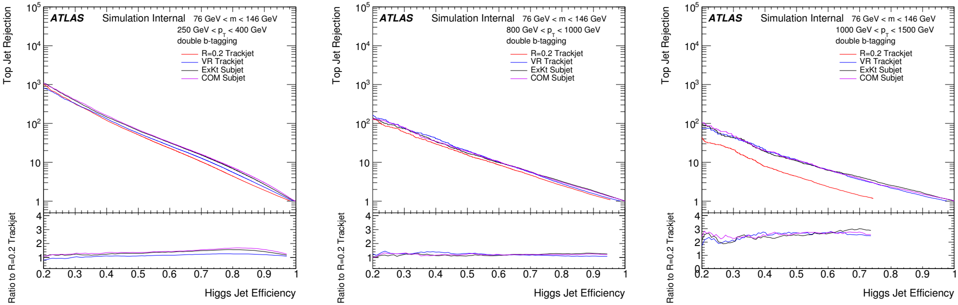


- COM performs the best
- $R = 0.2$ trackjets beat ExKt and VR in the middle p_T bin here
- all alternative techniques beat $R = 0.2$ trackjets in the high p_T bin here

Fig. 12 Caption

QCD jet rejection as function of $h \to b\bar{b}$ jet efficiency when applying double b-tagging on subjets found by the $R = 0.2$ trackjet, VR trackjet, exclusive-k_T subjet, and CoM subjet algorithms in different p_T regions. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 13: Top Double B-Tagging ROC Curves

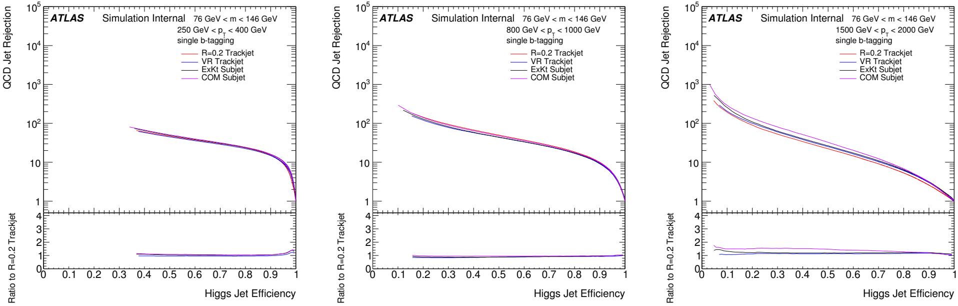


- all alternative techniques beat $R = 0.2$ trackjets, with the effect being most pronounced at high p_T

Fig. 13 Caption

Top jet rejection as function of $h \to b\bar{b}$ jet efficiency when applying double b-tagging on subjets found by the $R = 0.2$ trackjet, VR trackjet, exclusive-k_T subjet, and CoM subjet algorithms in different p_T regions. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 14: QCD Single B-Tagging ROC Curves

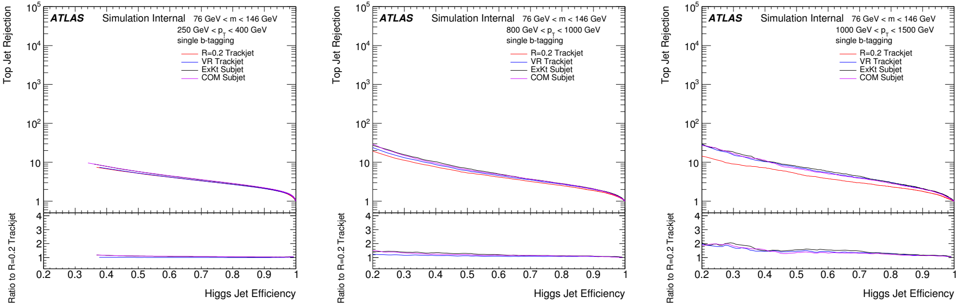


- all techniques perform similarly
- the alternative techniques perform somewhat better at high p_T, with COM performing the best

Fig. 14 Caption

QCD jet rejection as function of $h \to b\bar{b}$ jet efficiency when applying single b-tagging on subjets found by the $R = 0.2$ trackjet, VR trackjet, exclusive-k_T subjet, and CoM subjet algorithms in different p_T regions. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 15: Top Single B-Tagging ROC Curves

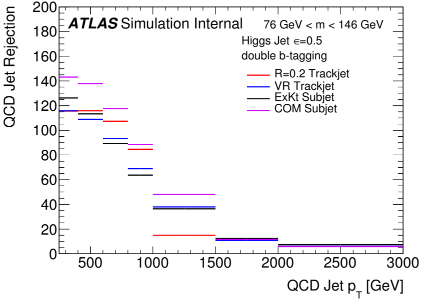


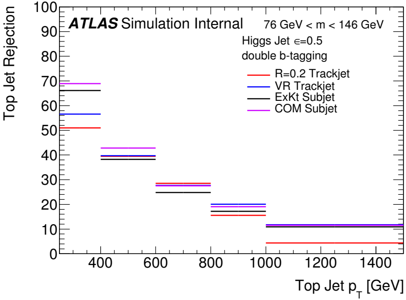
- all techniques perform similarly
- the alternative techniques perform somewhat better at high p_T, with COM performing the best

Fig. 15 Caption

Top jet rejection as function of $h \to b\bar{b}$ jet efficiency when applying single b-tagging on subjets found by the $R = 0.2$ trackjet, VR trackjet, exclusive-k_T subjet, and CoM subjet algorithms in different p_T regions. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 16: Double B-Tagging Rej. vs p_T



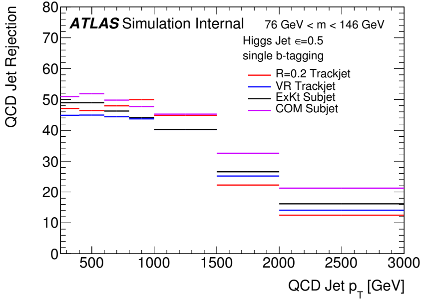


- same conclusions as with the ROC curves

Fig. 16 Caption

QCD and top jet double b-tagging rejection as a function of p_T for a fixed Higgs jet efficiency of 50%. Statistical errors are not shown, but are comparable in size to the line width.

Fig. 17: Single B-Tagging Rej. vs p_T



- same conclusions as with the ROC curves

Fig. 17 Caption

QCD and top jet single *b*-tagging rejection as a function of p_T for a fixed Higgs jet efficiency of 50%. Statistical errors are not shown, but are comparable in size to the line width.

CDS Comments

cds comments link

There are a few sets of CDS comments posted since last week that we need to address. Most of the comments are textual, asking for clarifications or fixing typos. These will be addressed ASAP. In the following few slides, I have picked out some comments which we should review here.

Pubnote Title

There has been some discussion about the title of the note. We have moved away from saying “advanced” taggers because the new taggers are not necessarily always better than the current nominal tagger. However, the current title is also not the best. (Andrea, Richard, Dan, and Sam have also made comments about this)

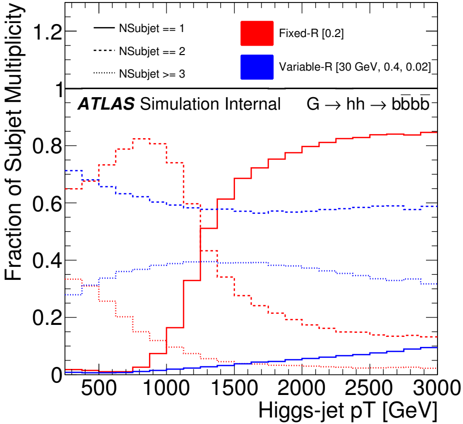
One suggestion from Dan is quite appealing:

“Center of Mass, Variable Radius, and Exclusive-k_T Subjet Reconstruction for Higgs Boson Tagging at ATLAS”

Appendices

We have some insightful material in the appendices, and they have brought up some CDS comments/questions about whether or not we want to include them in the main text. I present the material here and ask for suggestions.

Figure 18: VR Multiplicity Study

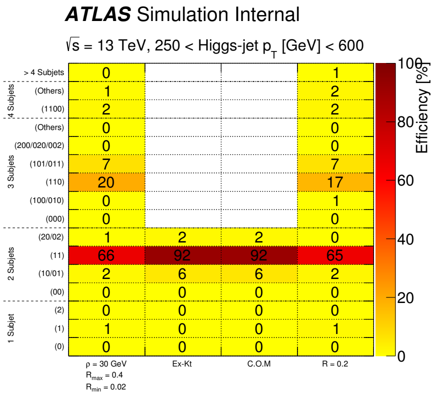


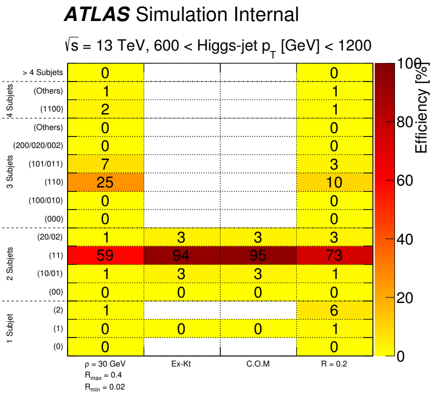
- we carried out a study showing how VR can recover efficiency when considering the subsubleading subjet
- this is one figure resulting from that study which is particularly informative
- should we expand and include this appendix, move it to the main text, or leave it out?

Figure 18 Caption

The subjet multiplicity efficiency as a function of the Higgs jet p_T for Higgs jets with exactly 1 subjet (solid line), 2 subjets (dashed line) and at least 3 subjets (dotted line). Red line refers to fixed radius track jets (anti-k_T R $= 0.2$) while the blue line refers to variable radius track jets ($\rho = 30$ GeV and $R_{max} = 0.4$, the $R_{min} = 0.02$).

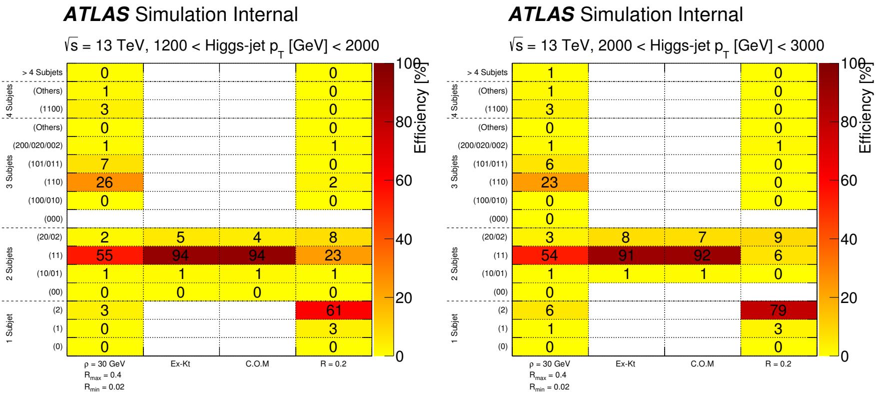
Figure 19 (Top): Efficiency Matrices





- these plots are very informative - they catagorize all of our subjet techniques
- they require some explanation, and we are wondering again whether we should expland and include this appendix, move it to the main text, or leave it out

Figure 19 (Bot.): Efficiency Matrices

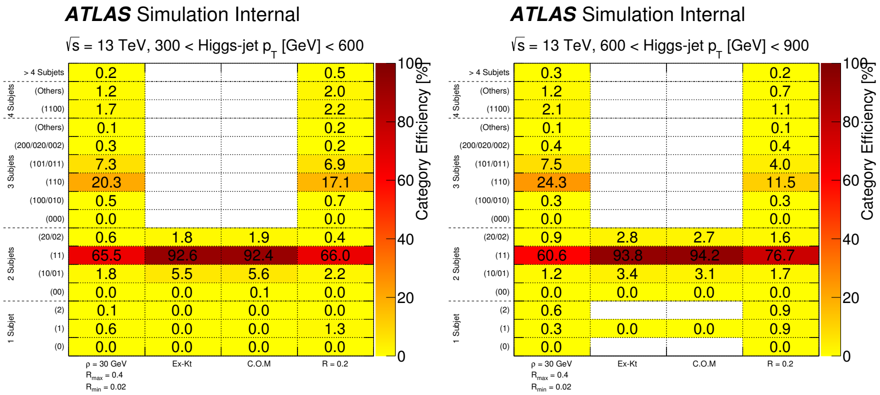


- these plots are very informative - they catagorize all of our subjet techniques
- they require some explanation, and we are wondering again whether we should expland and include this appendix, move it to the main text, or leave it out

Figure 19 Caption

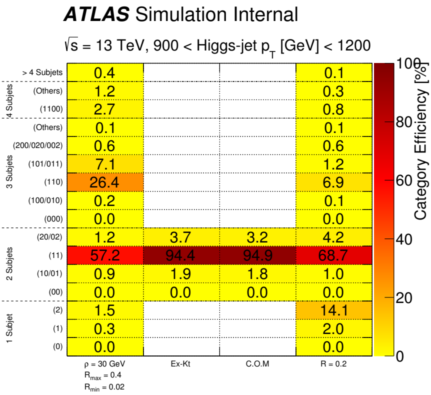
Category efficiency matrix for different subjet collections in various Higgs jet pT regimes.

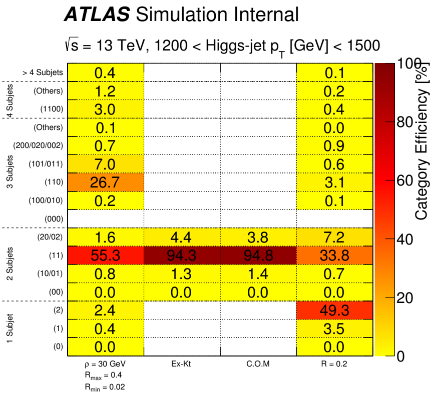
Figure 20 (Top): Fine Bin Efficiency Matrices



- these plots are very informative - they catagorize all of our subjet techniques
- they require some explanation, and we are wondering again whether we should expland and include this appendix, move it to the main text, or leave it out

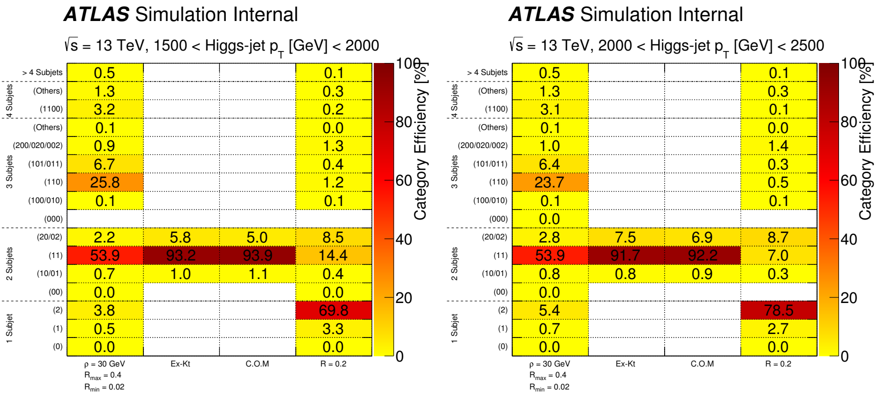
Figure 20 (Middle): Fine Bin Efficiency Matrices





- these plots are very informative - they catagorize all of our subjet techniques
- they require some explanation, and we are wondering again whether we should expland and include this appendix, move it to the main text, or leave it out

Figure 20 (Bot.): Fine Bin Efficiency Matrices



- these plots are very informative - they catagorize all of our subjet techniques
- they require some explanation, and we are wondering again whether we should expland and include this appendix, move it to the main text, or leave it out

Figure 20 Caption

Category efficiency matrix for different subjet collections in various fine Higgs jet pT bins.

Summary

Our pubnote hopes to document and demonstrate the merits of 3 new Higgs tagging techniques.

The note has undergone several revisions, and we are always open to new CDS comments.

The plots in the main text of the pubnote seem to have converged nicely, and most of the remaining CDS comments are contextual and will be addressed ASAP.