Tagging with substructure

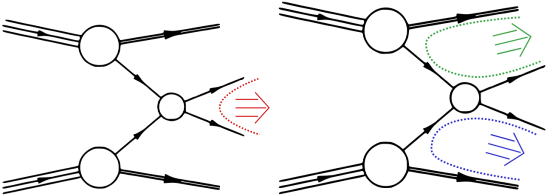
Designing robust taggers with high-level observables

L. Cavallini, A. Coccaro, O. Fedkevych, C. Khosa, G. Manco, S. Marzani, F. Parodi, D. Rebuzzi, A. Rescia, F. Sforza, G. Stagnitto

17 August 2022

Introduction

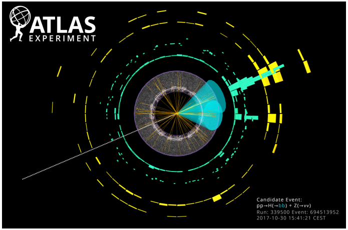
- Many interesting searches involve final states containing b-quarks stemming from colour singlets
  - VHbb, HH4b, $t\bar{t}H(b\bar{b})$
- Need to distinguish whether b-jets originate from Higgs or elsewhere
  - Gluon splitting, t-decay, etc.
- **Idea: Design taggers using jet substructure observables which exploit different colour configurations**
- Can additionally exploit this information along with radiation patterns and hadronisation features for b-tagging purposes



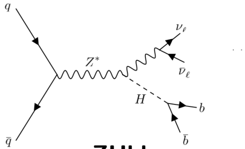
[arXiv:1001.5027]

Colour configuration tagger

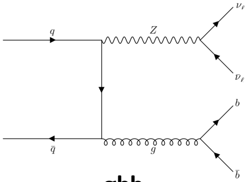
- Tagger designed with $H \to b\bar{b}$ decays in mind
- Topology similar to that of QCD-induced processes
- VH production can be used to improve experimental signature
- More information ultimately needed to effectively distinguish signal from background
  $\Rightarrow$ **colour flow**



ATLAS-PHOTO-2018-022-7



ZHbb



gbb

Based on
arXiv:2112.09650

Procedure

Processes

- Signal: $pp \to ZH$, $Z \to \nu_\ell \overline{\nu}_\ell$, $H \to b\bar{b}$ (ZHbb)
- Background: $pp \to b\bar{b}\nu_\ell \overline{\nu}_\ell$ (gbb)

Observables

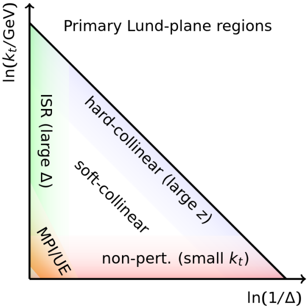
- Isolate high-level variables which distinguish decays from color singlets vs. color octects
- Analyse simulated data and extract the **9 variables**
  - Primary Lund Plane CNN (**lpcnn**)
  - Pull vector components and pull angle $t_{\|a}$, $t_{\perp a}$, θ_{pa} relative to Jet J_a
  - Pull vector components and pull angle $t_{\|b}$, $t_{\perp b}$, θ_{pb} relative to Jet J_b
  - Colour Ring ($\mathcal{O}$)
  - D_2

Machine Learning

- Use variables to train Boosted Decision Tree (BDT)

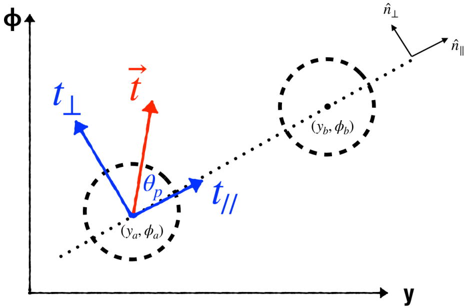
Lund Plane CNN

- "De-cluster" a C/A jet following the hardest branch in each splitting
- Plot coordinates of branch in a plane
  - $\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}$
  - $k_t = p_{Tb}\Delta$
- Feed Lund Plane for jet to a CNN
  - Train on different radiation patterns for signal/background
- Use output as discriminant



[arXiv:1807.04758]

Jet Pull Components



[arXiv:1911.05090v2]

$$\vec{t} = \frac{1}{p_{t_a}} \sum_{i \in J_a} p_{t_i} |\vec{r}_i|^2 \hat{r}_i$$

$$\vec{r}_i = (y_i - y_a, \phi_i - \phi_a)$$

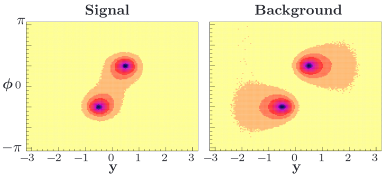
$$\hat{n}_\parallel = \frac{1}{\sqrt{\Delta y^2 + \Delta\phi^2}} (\Delta y, \Delta\phi)$$

$$\hat{n}_\perp = \frac{1}{\sqrt{\Delta y^2 + \Delta\phi^2}} (-\Delta\phi, \Delta y)$$

$$t_\parallel = \vec{t} \cdot \hat{n}_\parallel$$

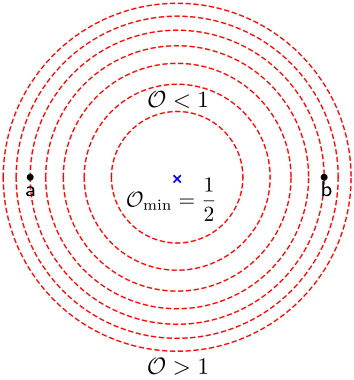
$$t_\perp = \vec{t} \cdot \hat{n}_\perp$$

$$\theta_p = \text{arccos}\left(\frac{t_\parallel}{|\vec{t}|}\right)$$



[arXiv:1001.5027v3]

Colour Ring



[arXiv:2006.10480v2]

- Use Neyman-Pearson lemma to design optimal observable to distinguish signal and background configurations

$$\mathcal{O} = \frac{|\mathcal{M}_{\mathcal{S}}|^2}{|\mathcal{M}_{\mathcal{B}}|^2} \approx \frac{\theta_{ak}^2 + \theta_{bk}^2}{\theta_{ab}^2}$$

- θ_{ak} (θ_{bk}) angle between hard parton a/b and gluon (k)
  $\Rightarrow$ Requires 3 objects!
- θ_{ab} angle between hard partons

D_2

- D_2 is defined as in [arXiv:1409.6298v1]

$$D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}$$

- $e_n^{(\beta)}$ is the normalised n-point Energy Correlator function
- For this case,

$$e_2^{(\beta)} = \frac{1}{p_{TJ}^2} \sum_{1 \leq i < j \leq n_J} p_{T_i} p_{T_j} R_{ij}^{\beta}$$

$$e_3^{(\beta)} = \frac{1}{p_{TJ}^3} \sum_{1 \leq i < j < k \leq n_J} p_{T_i} p_{T_j} p_{T_k} R_{ij}^{\beta} R_{ik}^{\beta} R_{jk}^{\beta}$$

- Sensitive to soft, large-angle radiation
  $\Rightarrow$ different between colour singlet and octect

Analysis

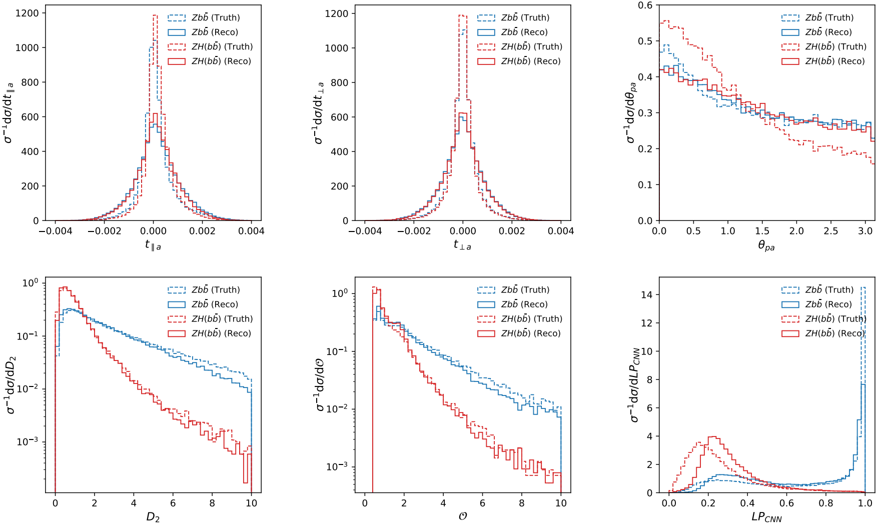
- Generate 300k signal & 4M background events in MG5 aMC@NLO v2.8.3.2
  - Require $p_T > 200$ GeV for ν pair
- Shower in Pythia v8.305 and simulate detector effects w/ Delphes v3.5.0 using modified ATLAS card
- Extract Monte Carlo truth and fast detector reconstruction
- **Truth:** Remove ν's and cluster stable particles w/ $p_T > 0.5$ GeV into jets
- **Reco:** Cluster into jets
  - EM Calo Towers
    - $E_{min} = 0.5$ GeV
    - $S_{min} = 2.0$ (Significance)
  - Hadron Calo Towers
    - $E_{min} = 1.0$ GeV
    - $S_{min} = 2.0$ (Significance)
  - Muon tracks w/ $p_T > 0.5$ GeV

- Identify R = 1.0 jets and select the hardest
  - $p_T > 250$ GeV
  - $|y| < 1.5$
- Identify R = 0.2 subjets
  - $p_T > 10$ GeV
  - ΔR < 0.8 from large-R jet
- Angular b-labelling and selection of events w/ exactly 2 b-subjets
  - b-parton $p_T > 5$ GeV
  - ΔR $= 0.2$
  - $|\eta| < 2.5$
- Choose hardest non-b subjet as third jet for colour ring
  - If not present, set $\mathcal{O} =$ -1

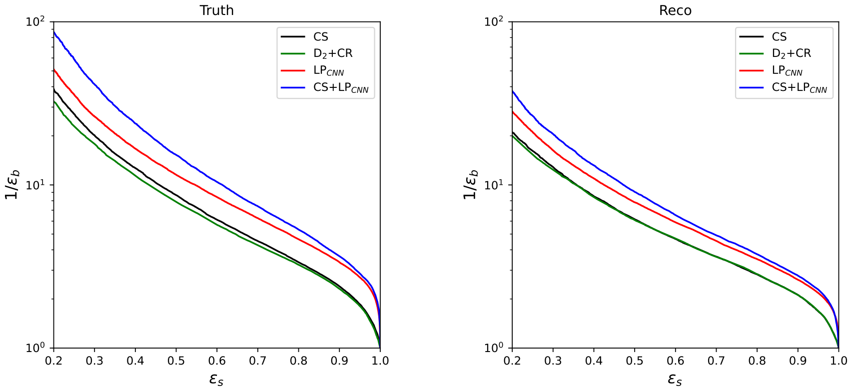
Events Passed

	Truth	Reco
Signal	20%	17%
Background	1.6%	1.3%

Distributions

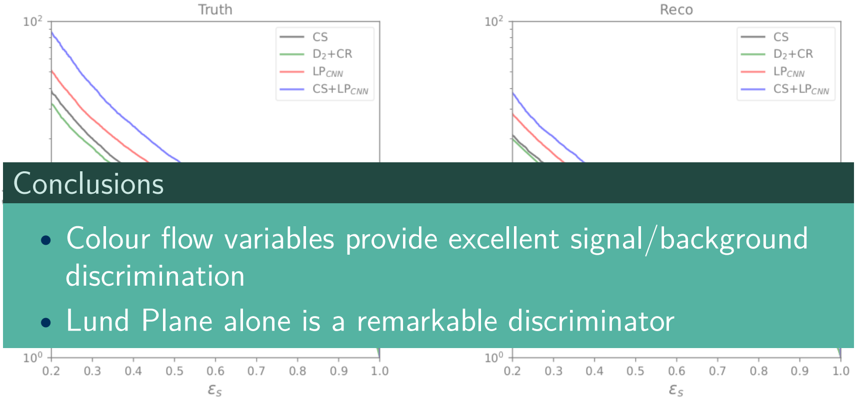


Results



	AUC - Test Sample	
	Truth	Reco
CS + LP	0.893	0.846
LP	0.876	0.828
D_2+CR	0.817	0.787
CS observables	0.826	0.788

Results



Conclusions

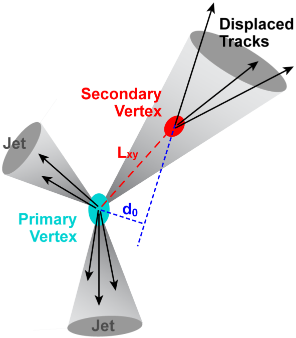
- Colour flow variables provide excellent signal/background discrimination
- Lund Plane alone is a remarkable discriminator

	AUC - Test Sample	
	Truth	Reco
CS + LP	0.893	0.846
LP	0.876	0.828
D_2+CR	0.817	0.787
CS observables	0.826	0.788

QCD b-tagging

Using QCD variables to discriminate b-jets vs. light jets

- B-tagging (in ATLAS) currently based on techniques which exploit properties of b-hadrons
  - Long lifetime of b-hadron leads to secondary vertex
  - Use tracking information to measure signed impact parameter and to measure mass (~ 5 GeV)
  - Feed information to low-level ML algorithms and ultimately to high-level DL1 NN
- QCD variables can be used to design a simpler ML-based tagger using jet substructure observables
  - Primary Lund Plane CNN
  - Jet angularities



[Source: D0 Collaboration]

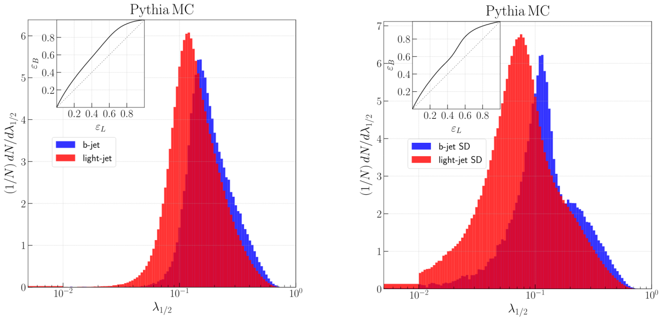
Based on
arXiv:2202.05082

Jet angularities

- Jet angularities defined as in [arXiv:2112.09545]:

$$\lambda_\alpha^\kappa = \sum_{i \in jet} \left(\frac{p_{T,i}}{\sum_{j \in jet} p_{T,j}} \right)^\kappa \left(\frac{\Delta_i}{R_0} \right)^\alpha$$

- IRC safety obtained for $\kappa = 1$ and $\alpha > 0$
  - $\lambda_{1/2}^1, \lambda_1^1, \lambda_2^1$ chosen for this work
- Observables provide some discrimination between b/light-jets
- Jet grooming (SoftDrop) changes shape of distribution but not discrimination potential

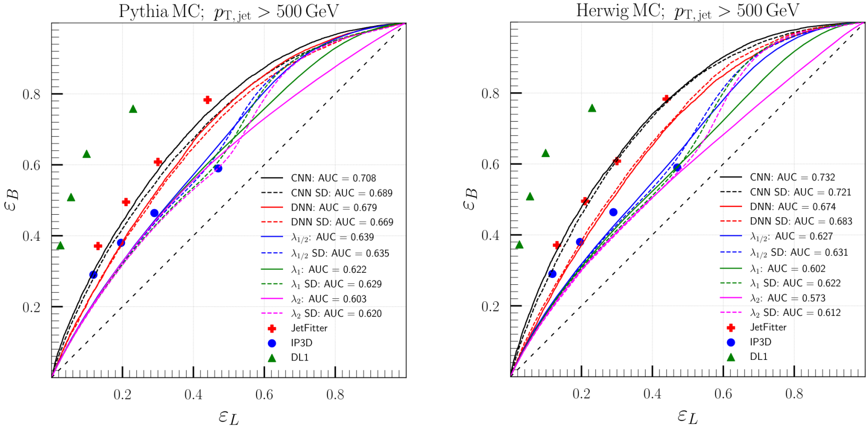


Analysis

- Simulate $Z(\mu^+\mu^-)$+hadronic jet events at the LHC
- ME simulation:
  - **Train:** PYTHIA v8.303 w/ NNPDF2.3 LO
  - **Validation:** HERWIG v7.2.1 w/ CT14 LO PDF set
- Shower with default tunes
- Muon requirements:
  - $p_{T,\mu} > 26$ GeV
  - $|m_{\mu^+\mu^-} - m_Z| < 20$ GeV
  - $|\eta_\mu| < 2.4$
- Require at least one R $= 0.4$ anti-k_t jet:
  - $|y_{jet}| < 2.5$
  - $p_{T,jet} > 500$ GeV
  - $|\frac{p_{T,jet} - p_{T,\mu^+\mu^-}}{p_{T,jet} + p_{T,\mu^+\mu^-}}| < 0.3$
  - $|\phi_Z - \phi_{jet}| > 2$

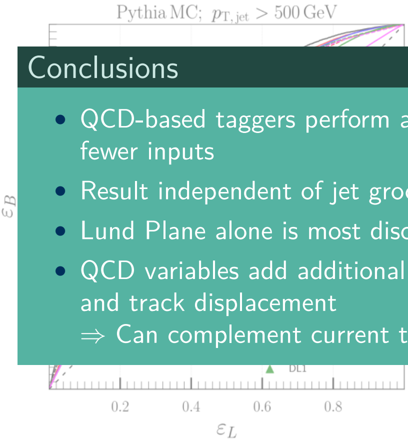
- Jet tagged as b (c) if matched to decaying b (c) hadron within $\Delta R = 0.3$
  - Remaining jets tagged as light
  - c-jets discarded (b vs. light tagging)
- Apply SOFTDROP grooming from fjcontrib library
  - $\beta = 0$
  - $z_{cut} = 0.1$
- Measure jet angularities and Lund Plane
  - Train DNN with jet angularities
  - Train CNN with Lund Plane
- Compare to ATLAS taggers

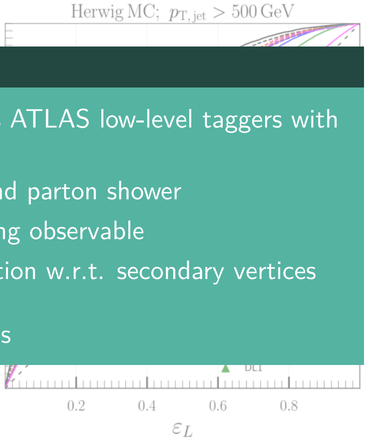
Results



Pythia MC; $p_{\mathrm{T,jet}} > 500$ GeV
ε_B
ε_L
CNN: AUC = 0.708
CNN SD: AUC = 0.689
DNN: AUC = 0.679
DNN SD: AUC = 0.669
$\lambda_{1/2}$: AUC = 0.639
$\lambda_{1/2}$ SD: AUC = 0.635
λ_1: AUC = 0.622
λ_1 SD: AUC = 0.629
λ_2: AUC = 0.603
λ_2 SD: AUC = 0.620
JetFitter
IP3D
DL1
Herwig MC; $p_{\mathrm{T,jet}} > 500$ GeV
ε_B
ε_L
CNN: AUC = 0.732
CNN SD: AUC = 0.721
DNN: AUC = 0.674
DNN SD: AUC = 0.683
$\lambda_{1/2}$: AUC = 0.627
$\lambda_{1/2}$ SD: AUC = 0.631
λ_1: AUC = 0.602
λ_1 SD: AUC = 0.622
λ_2: AUC = 0.573
λ_2 SD: AUC = 0.612
JetFitter
IP3D
DL1

Results





Conclusions

- QCD-based taggers perform as well as ATLAS low-level taggers with fewer inputs
- Result independent of jet grooming and parton shower
- Lund Plane alone is most discriminating observable
- QCD variables add additional information w.r.t. secondary vertices and track displacement
  $\Rightarrow$ Can complement current techniques

Conclusions

- It is possible exploit high-level jet substructure observables to design QCD-based taggers which complement (b-tagging) or augment (colour configuration tagging) those currently in use
- Taggers are simple: effective discrimination with relatively few inputs
- Lund Plane is particularly effective discriminator for both b-tagging and colour configuration tagging
- Ongoing effort underway to include colour configuration tagger in ATLAS Xbb tagger

Conclusions

- It is possible exploit high-level jet substructure observables to design QCD-based taggers which complement (b-tagging) or augment (colour configuration tagging) those currently in use
- Taggers are simple: effective discrimination with relatively few inputs
- Lund Plane is particularly effective discriminator for both b-tagging and colour configuration tagging
- Ongoing effort underway to include colour configuration tagger in ATLAS Xbb tagger

Thank you for your attention!

Backup

ML details - Lund Plane CNN (Colour Tagger)

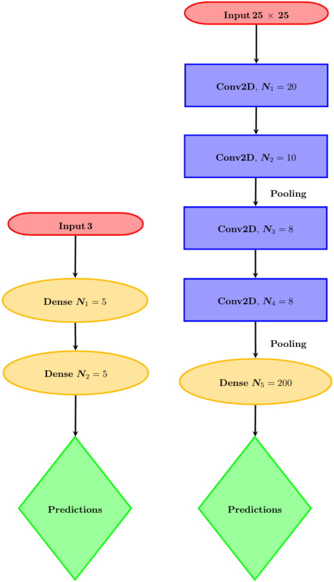
- 70/15/15 train/test/validation
- ReLu activation function used for intermediate layer
- Softmax activation used for output layer
- MaxPooling used after second and fourth layer

Parameter	Value
N_1 Conv2D	30
N_2 Conv2D	30
Dropouts	- (0.3)
N_3 Conv2D	30
Dropouts	- (0.3)
N_4 Conv2D	10
Dropouts	- (0.1)
Flat Layer	150
Epochs	30
Batch Size	800
Filter size	3×3
Optimiser	Adam

ML details - BDT (Colour Tagger)

Parameters	Value
No. of Trees	100
Max Depth	3
MinNodeSize	2.5%
Boost Type	AdaBoost
Train/Test	50/50
No. of Cuts	200
Downsampling	No
Optimisation	Gini index

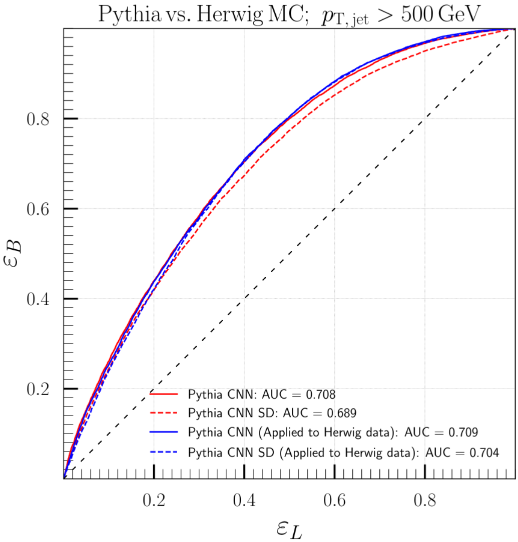
ML details (b-tagger)



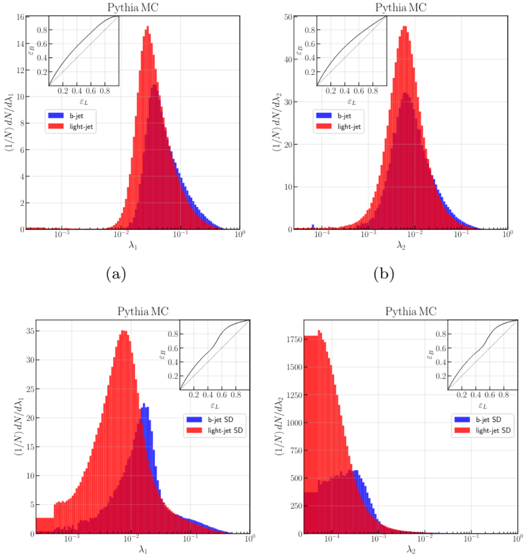
- **DNN** (left)
  - $\lambda_{1/2}, \lambda_1, \lambda_2$ used as inputs
  - ReLu activation function used for intermediate layer
  - Softmax activation function used for output layer
  - Adam optimiser
  - 60/20/20 train/test/validation
- **CNN** (right)
  - 4 convolutional layers followed by flat layer
  - $N_i =$ no. of filters in i-th layer
  - Same activation functions and optimisers as DNN

ROC curves (b-tagger)

Pythia train and Herwig test



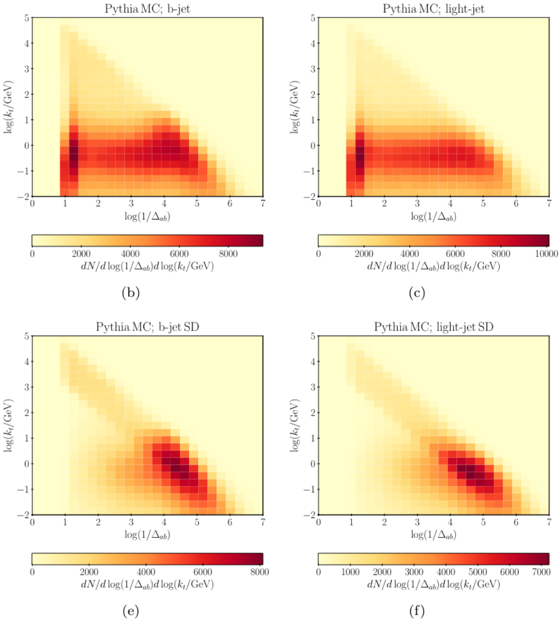
λ_1 & λ_2 distributions (b-tagger)



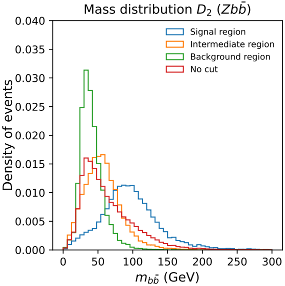
(a) (b)

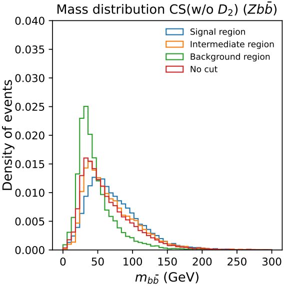
(c) (d)

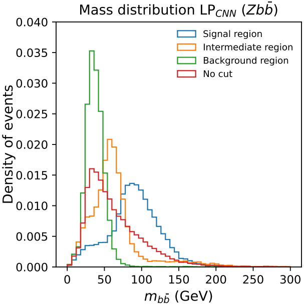
Lund Plane (b-tagger)



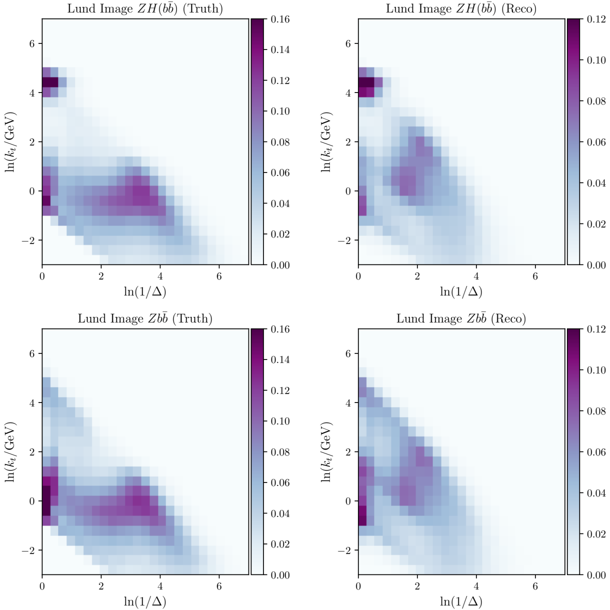
Mass bias (Colour Tagger)



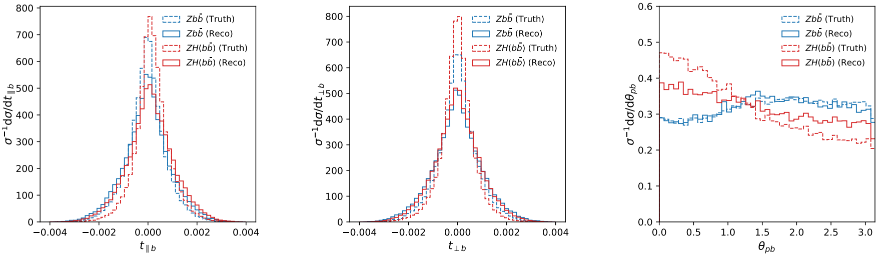




Lund Plane (Colour Tagger)



Jet 2 pull variable distributions (Colour Tagger)



BDT Variable Rankings (Colour Tagger)

Observable Ranking				
	Truth		Reco	
Rank	Obs.	Importance	Obs.	Importance
1	LP	$6.6\ \times 10^{-1}$	LP	4.8×10^{-1}
2	D_2	$1.4\ \times 10^{-1}$	$\mathcal{O}$	1.0×10^{-1}
3	$\mathcal{O}$	5.7×10^{-2}	D_2	9.3×10^{-2}
4	θ_{pb}	3.0×10^{-2}	θ_{pb}	7.0×10^{-2}
5	θ_{pa}	2.9×10^{-2}	θ_{pa}	6.5×10^{-2}
6	$t_{\|b}$	2.6×10^{-2}	$t_{\perp b}$	6.0×10^{-2}
7	$t_{\|a}$	2.4×10^{-2}	$t_{\|a}$	4.5×10^{-2}
8	$t_{\perp b}$	1.9×10^{-2}	$t_{\perp a}$	4.3×10^{-2}
9	$t_{\perp a}$	1.0×10^{-3}	$t_{\|a}$	3.3×10^{-2}