Jet sub-structures studies at FCC-hh

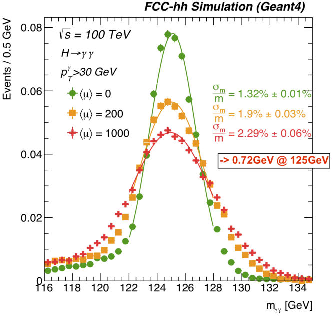
C. Neubüser on behalf of the FCC-hh Detector group

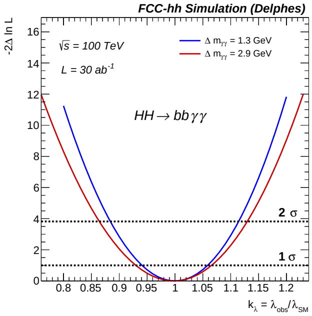
BOOST 2019, Boston

Interplay of fast and full-simulation

validation of physics reach reported in FCC *CDR*

example: di-photon invariant mass resolution, in Higgs self-coupling constant (with 1,000 pile-up)

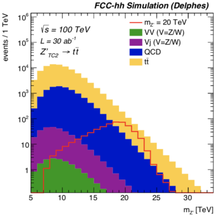


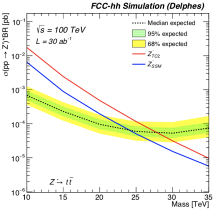


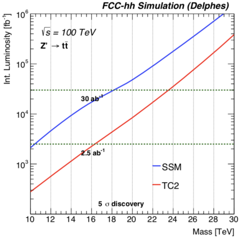
Here: tests with boosted objects

Physics analyses

$Z' \rightarrow t\bar{t}$ (full hadronic)







pre-selection:

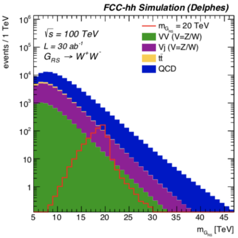
- p_T >3 TeV, $|\eta| < 3$
- m_{SD}>100 GeV
- τ_{21}, $\tau_{32} > 0$
- 2 b-tagged jets
- di-top mass corrected for MET

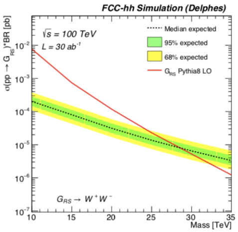
final selection based on top tagger:

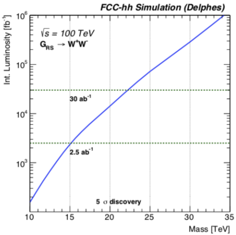
exploring jet-shapes to suppress QCD background, using multi variate analysis techniques (BDTs)

- exclusions between 25-28 TeV
- discoveries between 18-24 TeV.

$Q^* \to W^+W^-$ full hadronic)







pre-selection:

- p_T >3 TeV, $|\eta| < 3$
- m_{SD}>40 GeV

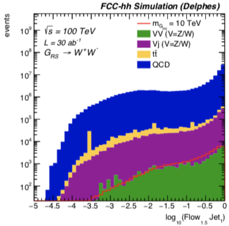
final selection based on W tagger:

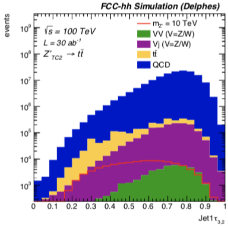
essential to find the few events in orders of magnitudes higher backgrounds

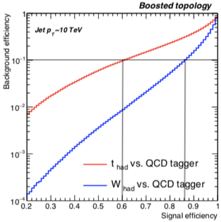
- exclusion up to 28 TeV
- discovery of 22 TeV.

Fast-simulation top/W tagger

BDTs in TMVA toolkit







distinguish between QCD and top/W jets:

- most-central energy fraction $Flow_{1,5}$
- $\tau_{3,2} = \tau_3/\tau_2$ sub-jettiness `arXiv:1011.2268`
  → used for t/QCD

→ W boson id:

90 % background rejection, and

85 % efficiency

→ top-jet id:

90 % background rejection, and

60 % efficiency

→ based on..

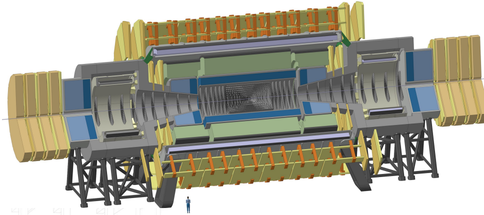
- detector parametrisation.
- track based jets including pfa corrections → **validation in full-sim**

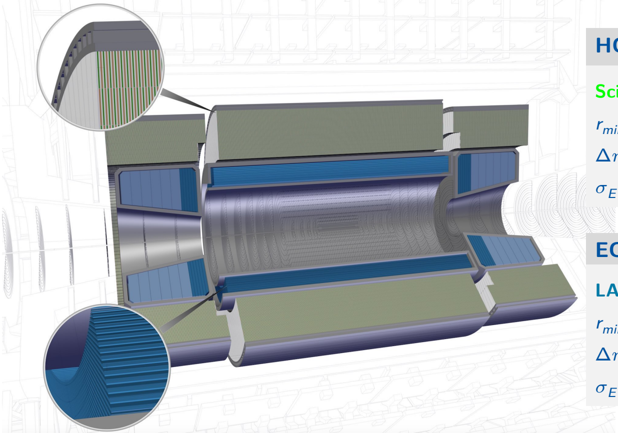
Full simulation

FCC-hh reference detector

100 TeV proton-proton collisions
up to on average 1,000 pileup events

- full simulations in **FCCSW**
- here emphasis on calorimetry in 4 T magnetic field





HCAL (Ext.) Barrel

Sci-Pb-Steel, 10/8 layers

$r_{min} = 2.85\,\mathrm{m}$

$\Delta\eta \times \Delta\phi = 0.025 \times 0.025$

$\sigma_E/E \sim 50\%/\sqrt{E} \oplus 3\%$

ECAL Barrel

LAr-Pb (1:3), 6-8 layers

$r_{min} = 1.75\,\mathrm{m}$

$\Delta\eta \times \Delta\phi = 0.01 \times 0.009$

$\sigma_E/E \sim 10\%/\sqrt{E} \oplus 0.7\%$

Simulation/reconstruction in FCCSW

Monte Carlo simulations:
QCD di-jets / W→qq / $t \rightarrow WWb$ in $|\eta| \leq 0.5$

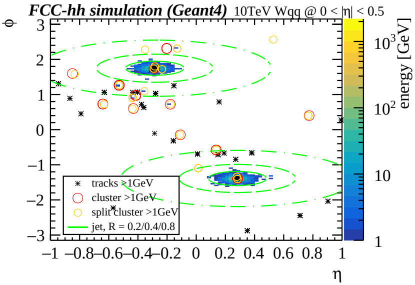
- Generator: Pythia8
- Particle propagation: Geant4 10.4.0

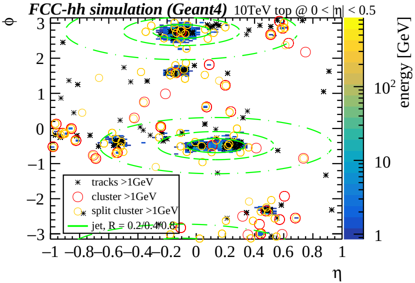
Reconstruction algorithms:

- calorimeter **cells**
- calorimeter **cluster** (sliding window, topological clustering) includes electronics/pile-up noise
- **tracks**, smeared charged generated particles according to tracker resolution

→ input for FastJet package 3.3.0 for jet clustering:

- matching gen/reco jets within $\Delta R = 0.3$
- select 2 highest p_T reco-jets / event
- requires jet 2.5 $GeV < p_T <$ 20 TeV

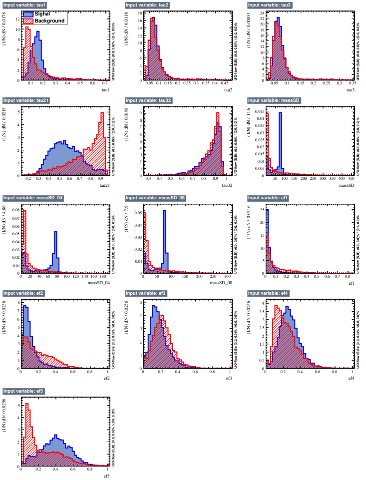




W tagger in full-simulation – calorimeter cells

13 BDT input variables

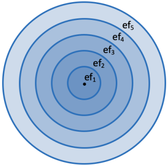
500 GeV Wqq_vs_QCD, R=1.0



- large jet cone necessary to ensure full energy containment
- τ_{21} and m_{SD} most powerful observables ($\beta = 0$)
- energy flow observables, transverse momentum fraction within 5 angular slices of size $\Delta R = 0.01$

$$ef_n = \sum_{\frac{n-1}{5}\alpha < \Delta R_i < \frac{n}{5}\alpha} \frac{p_{T,i}}{p_{T,sum}}$$

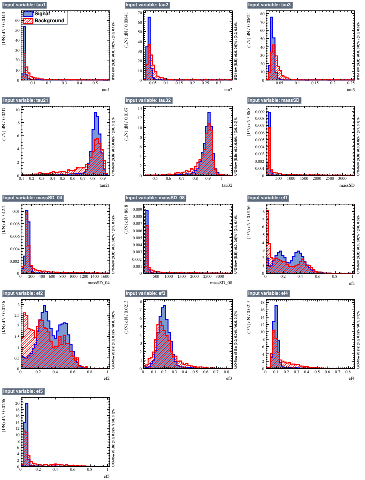
with $\alpha = 0.05$, $p_{T,i}$ of the ith cell with distance ΔR_i to jet axis.



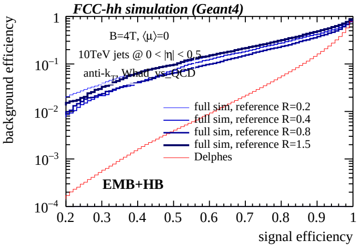
W tagger in full-simulation – calorimeter cells

13 BDT input variables

10 TeV Wqq_vs_QCD, R=0.8



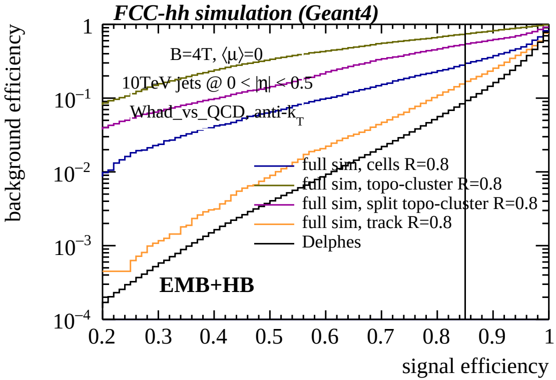
- extreme collimation of 10 TeV jets
  $\rightarrow \Delta R = 2m/p_T \approx 0.016$
  calo cell sizes 0.01/0.025 of E/HCal
- week separation power on the sub-jettiness variable



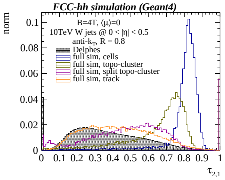
$\rightarrow$ optimisation of jet cone size $R = 0.8$

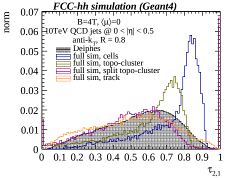
- 90 % signal efficiency with 20 % QCD mis-id

Performance of W tagger



example: $\tau_{2,1}$





- W tagger degrades with clustering/noise
  → need optimisation of cluster splitting
- remaining differences: definition of energy flow variable, and stats

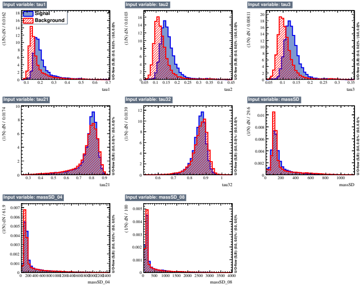
remaining contam.
at working point

Delphes	$\approx$ 10%
track	$\approx$ 18%
cells	$\approx$ 30%
split-cluster	$\approx$ 60%
cluster	$\approx$ 80%

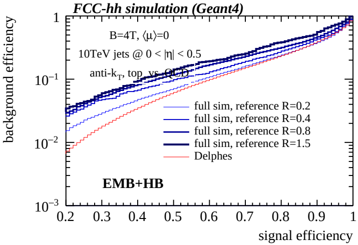
Top tagger in full-simulation – calorimeter cells

8 BDT input variables

10 TeV thad_vs_QCD, R=0.2



- extreme collimation of 10 TeV jets
  $\rightarrow \Delta R = 2m/p_T \approx 0.03$
  calo cell sizes 0.01/0.025 of E/HCal
- τ_2, τ_3 and the energy flow variables give discrimination power



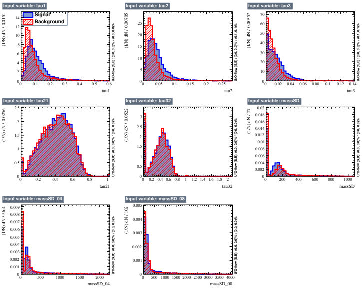
$\rightarrow R = 0.2$ best result

- 90 % signal efficiency with 20 % QCD mis-id

Top tagger in full-simulation – calorimeter clusters

8 BDT input variables

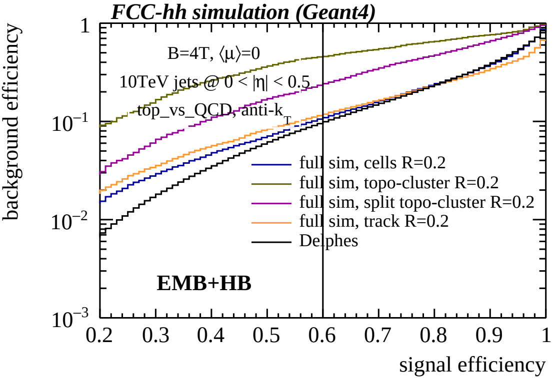
10 TeV thad_vs_QCD, R=0.2



- clustering of calorimeter cells effectively reduces granularity
- clusters are split, looking for local maxima, **but** if maxima too close to each other → no splitting
- low discrimination power

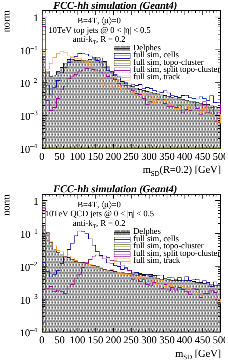
→ splitting needs p_T dependent optimisation

Performance of top tagger



- W/Top tagger trained on tracks closest to fast-simulation (Delphes) tagger
- 10 k calo cells in R=0.2
- !ATTENTION! assumption of perfect track reconstruction

example: $m_{SD}(R=0.2)$



remaining contam.
at working point

Delphes	$\approx 8\%$
track	$\approx 10\%$
cells	$\approx 9\%$
split-cluster	$\approx 20\%$
cluster	$\approx 40\%$

Conclusions

- jet sub-structures are crucial for boosted object id, and background rejections
- in FCC-hh environment, strong boost at FCC-hh especially challenging
- first full simulation studies of calorimeters, and simplified tracks show promising results

next steps:

- evaluate possible reconstruction optimisation to improve performance on W tagger
- matching of tracks and clusters within a jet to profit from complementarities

Conclusions

- jet sub-structures are crucial for boosted object id, and background rejections
- in FCC-hh environment, strong boost at FCC-hh especially challenging
- first full simulation studies of calorimeters, and simplified tracks show promising results

next steps:

- evaluate possible reconstruction optimisation to improve performance on W tagger
- matching of tracks and clusters within a jet to profit from complementarities

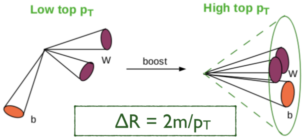
Thank you for your attention.

www.cern.ch

BACKUP

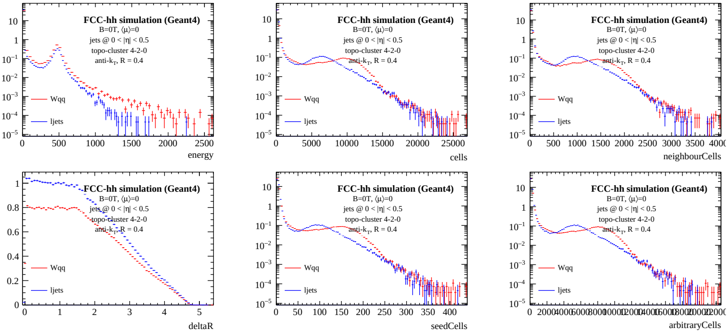
Why jet sub-structures?

- physics case of FCC-hh: discovery machine for new heavy resonances!
- searches of heavy resonances at frontier colliders in hadronic decay channels (`arXiv:1902.11217v2`)
  requirements:
  - id of boosted top quarks and W bosons essential!
  - collimated jets with angular sizes $< 0.01 - 0.02$
    $\rightarrow$ hard to distinguish e.g. 3-prong topology
    $\rightarrow$ set requirements on detectors granularity



Topo-clusters of 500 GeV jets

cut on cluster energy > 1 GeV, y-axis # of clusters/event, deltaR is the expansion of the cluster in R.



→ next step: implement cluster splitting, do distinguish jet components

Topo-cluster splitting

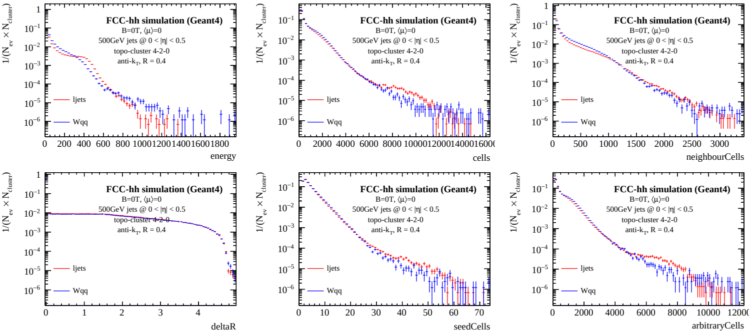
following ATLAS example [link]

for each topo-cluster:

1. identify local maxima:
   1.1 get seed cells above threshold t
   1.2 check if 4 neighbouring cells exist with energy $>$ 2nd topo-clustering threshold
   1.3 if more then one maximum found...
2. start splitting:
   2.1 use local maxima as new cluster seeds, starting with the highest energy one.
   2.2 collecting neighbouring cells for all clusters in iteration
   2.3 if cell has been identified for two clusters, distance from the cog of the clusters it determined, and it's assigned to the closest one
3. sanitary checks
   3.1 energy/cells preserved
   3.2 $\rightarrow$ write out new collection of clusters

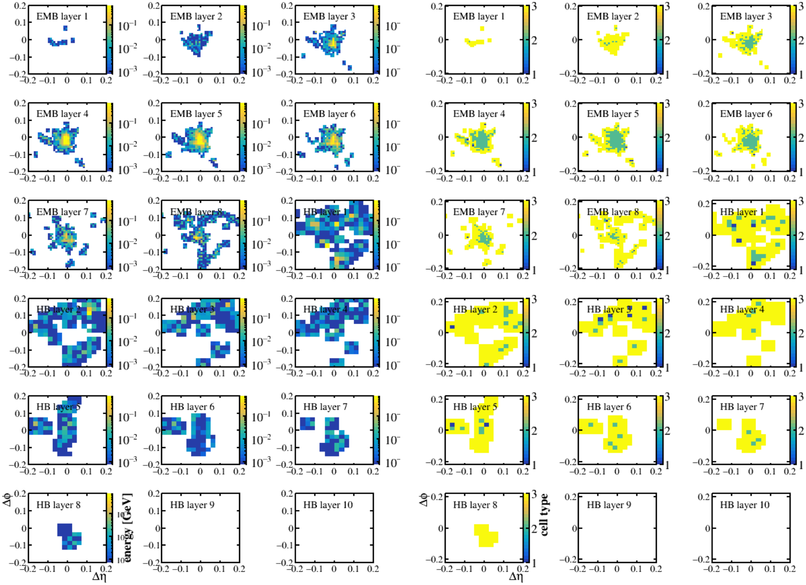
Split topo-clusters of 500 GeV jets

cut on cluster energy > 1 GeV, y-axis # of clusters/event, deltaR is the expansion of the cluster in R.



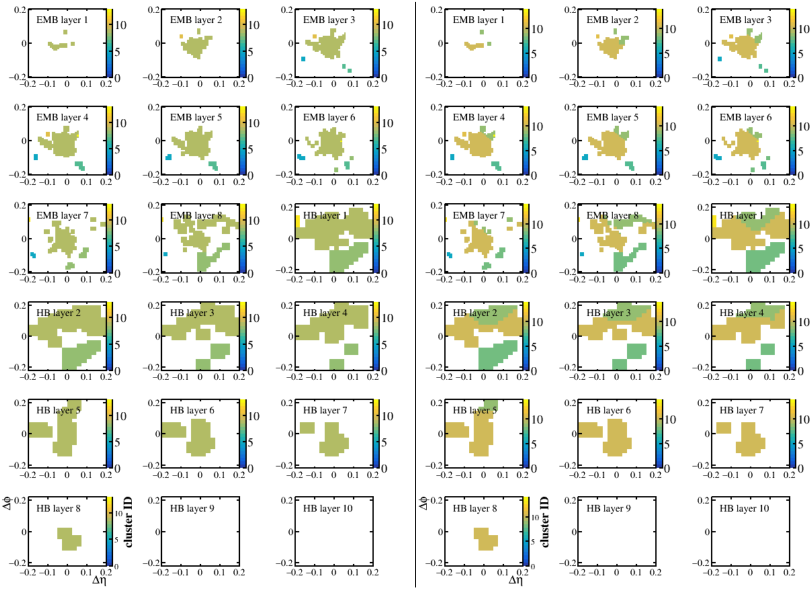
Topo-cluster splitting

example 100 GeV π^-



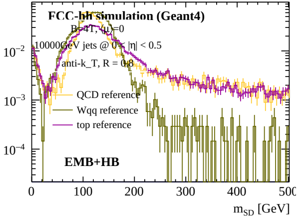
Before/after topo-cluster splitting

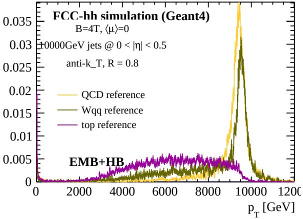
1 cluster split in 2:



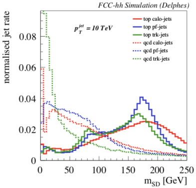
Jet mass and momentum of 10 TeV jets

calorimeter cells

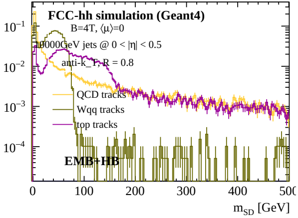


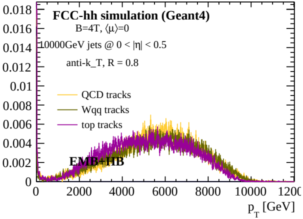


→ no efficient p_T, no m_{SD} reconstruction for $t\bar{t}$ jets



tracks (smeared charged particles)



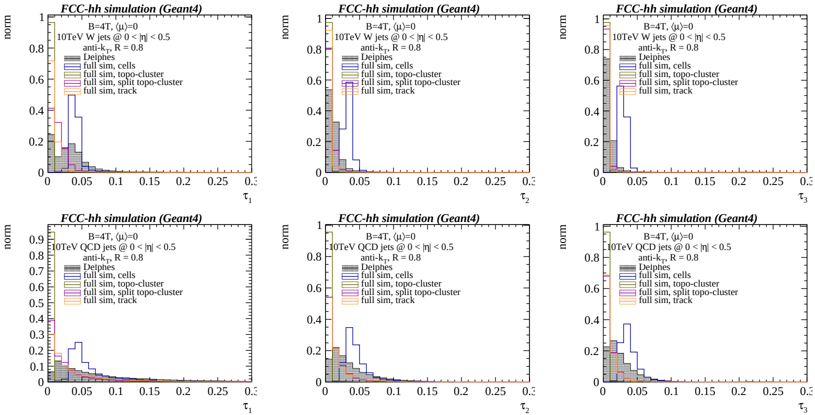


→ m_{SD} well reconstructed, strong expected error on the momentum.

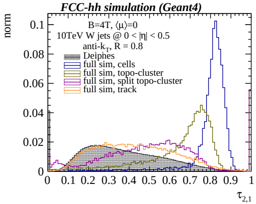
W tagger variables

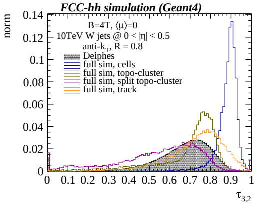
W/QCD, $R = 0.8$

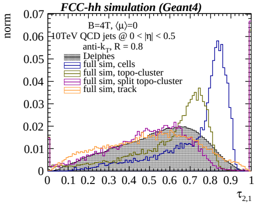
$\tau_{1,2,3}$

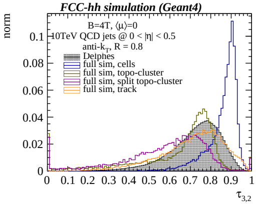

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.8
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
τ$_1$
τ$_2$
τ$_3$
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.8
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track

$\tau_{21,32}$


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.8
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{2,1}$


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.8
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{3,2}$

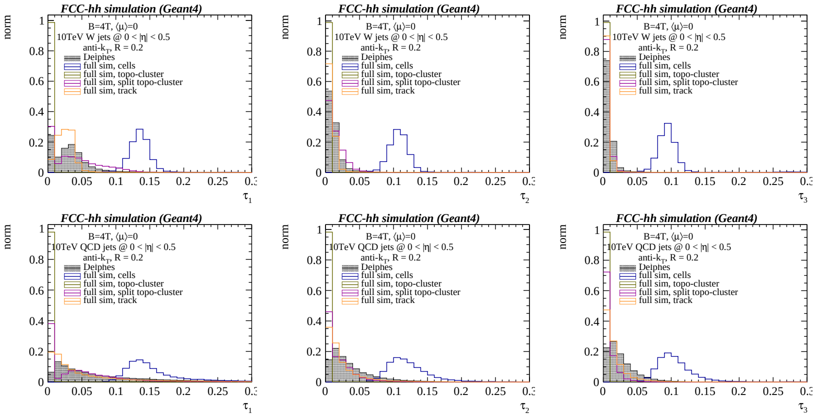

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.8
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{2,1}$


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.8
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{3,2}$

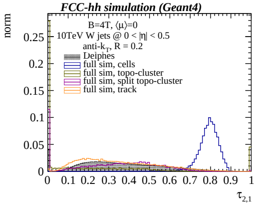
W tagger variables

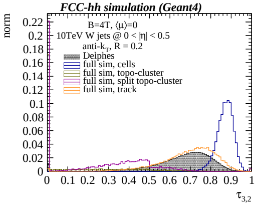
W/QCD, $R = 0.2$

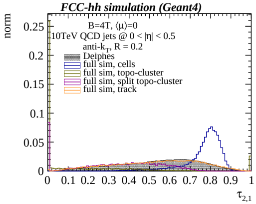
$\tau_{1,2,3}$

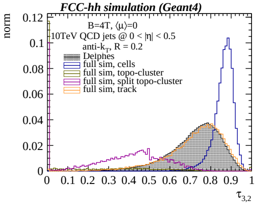

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
τ_1
τ_2
τ_3

$\tau_{21,32}$

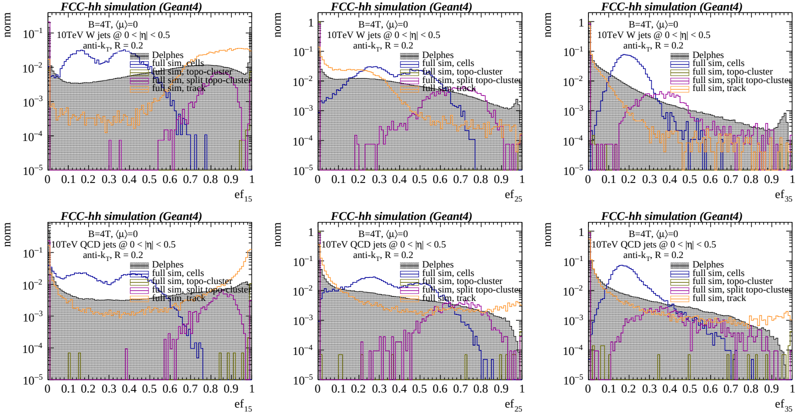

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{2,1}$


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{3,2}$

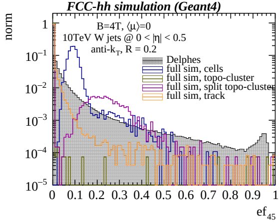

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{2,1}$

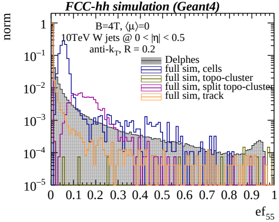

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
$\tau_{3,2}$

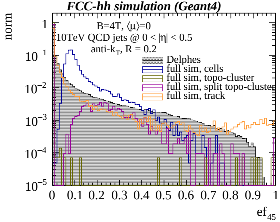
$ef_{15,25,35}$

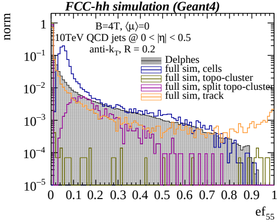

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
ef$_{15}$
ef$_{25}$
ef$_{35}$

$ef_{45,55}$

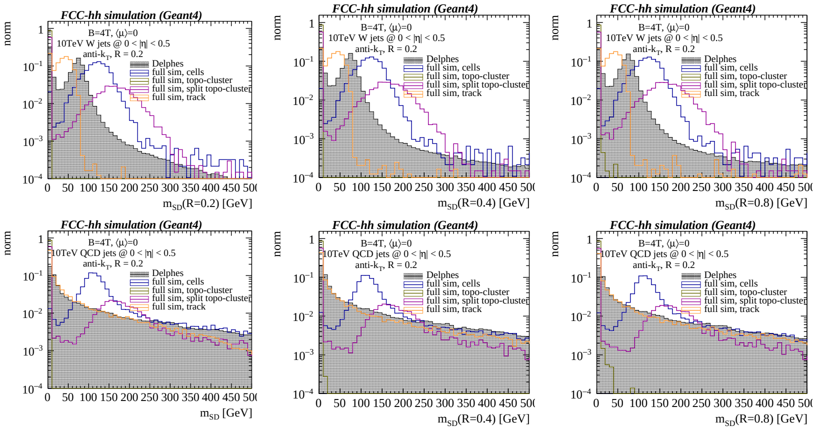

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
ef_{45}


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
ef_{55}


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
ef_{45}


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
ef_{55}

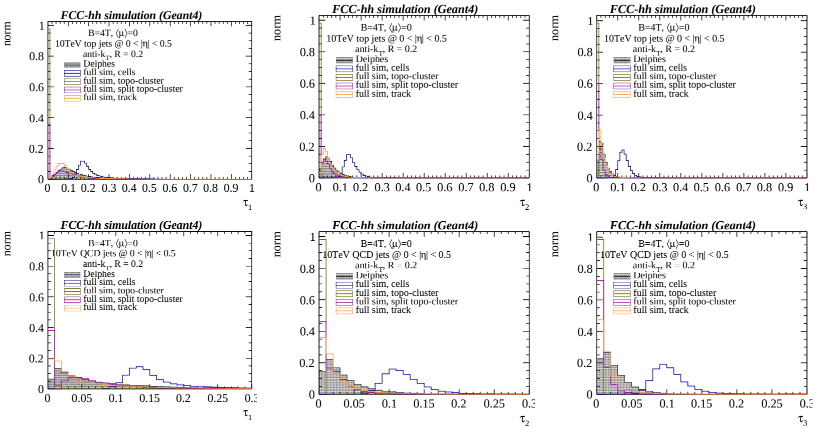
m_{SD}


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
m$_{SD}$(R=0.2) [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
norm
m$_{SD}$(R=0.4) [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV W jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
norm
m$_{SD}$(R=0.8) [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
norm
m$_{SD}$ [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
norm
m$_{SD}$(R=0.4) [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
norm
m$_{SD}$(R=0.8) [GeV]

Top tagger variables

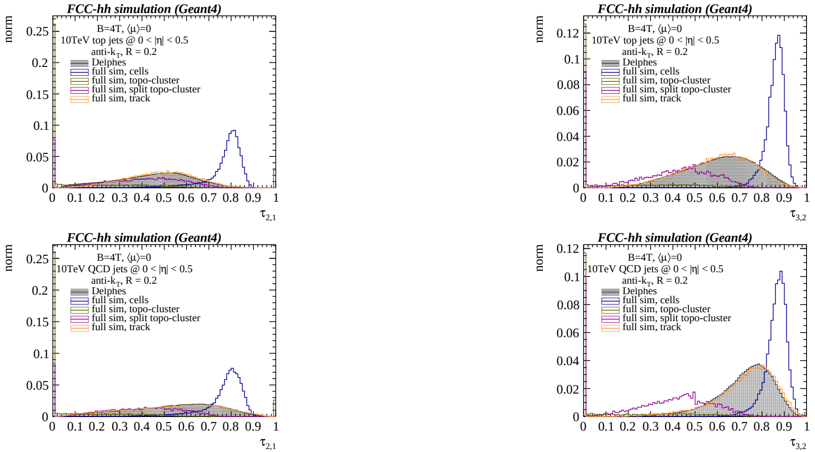
top/QCD, $R = 0.2$

$\tau_{1,2,3}$

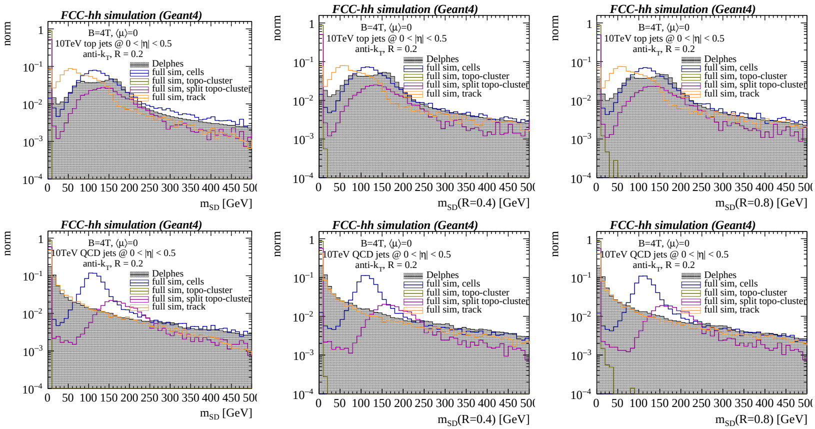


FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV top jets @ 0 < |η| < 0.5
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
τ_1
τ_2
τ_3

$\tau_{21,32}$


FCC-hh simulation (Geant4)
norm
B=4T, ⟨μ⟩=0
10TeV top jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
$\tau_{2,1}$
FCC-hh simulation (Geant4)
norm
B=4T, ⟨μ⟩=0
10TeV top jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
$\tau_{3,2}$
FCC-hh simulation (Geant4)
norm
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
$\tau_{2,1}$
FCC-hh simulation (Geant4)
norm
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k_T, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
$\tau_{3,2}$

m_{SD}

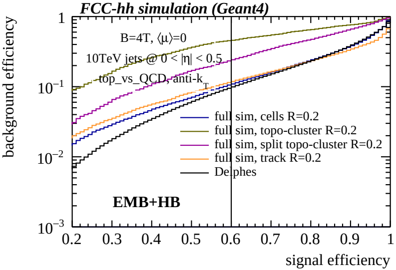

FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV top jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
Delphes
full sim, cells
full sim, topo-cluster
full sim, split topo-cluster
full sim, track
norm
m$_{SD}$ [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV top jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
m$_{SD}$(R=0.4) [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV top jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
m$_{SD}$(R=0.8) [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
m$_{SD}$ [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
m$_{SD}$(R=0.4) [GeV]
FCC-hh simulation (Geant4)
B=4T, ⟨μ⟩=0
10TeV QCD jets @ 0 < |η| < 0.5
anti-k$_T$, R = 0.2
m$_{SD}$(R=0.8) [GeV]

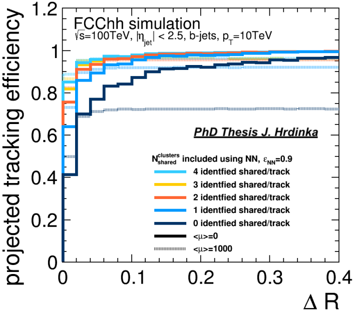
Delphes BDT overview

W tagger		top tagger	
variable	weight	variable	weight
τ_3 (track jet, R=0.2)	0.12	τ_1 (track jet, R=0.2)	0.21
m_{SD} (track jet, R=0.2)	0.11	m_{SD} (track jet, R=0.2)	0.17
τ_{31} (track jet, R=0.2)	0.10	τ_{31} (track jet, R=0.2)	0.11
$E_F(n = 5, \alpha = 0.05)$	0.09	τ_2 (track jet, R=0.2)	0.10
$E_F(n = 4, \alpha = 0.05)$	0.09	τ_3 (track jet, R=0.2)	0.09
$E_F(n = 1, \alpha = 0.05)$	0.08	m_{SD} (track jet, R=0.8)	0.09
$E_F(n = 2, \alpha = 0.05)$	0.07	m_{SD} (track jet, R=0.4)	0.09
$E_F(n = 3, \alpha = 0.05)$	0.06	τ_{32} (track jet, R=0.2)	0.08
τ_{21} (track jet, R=0.2)	0.06	τ_{21} (track jet, R=0.2)	0.06
m_{SD} (track jet, R=0.8)	0.06		
m_{SD} (track jet, R=0.4)	0.06		
τ_1 (track jet, R=0.2)	0.05		
τ_2 (track jet, R=0.2)	0.04		
τ_{32} (track jet, R=0.2)	0.02		

Table: Summary of the input variables to the BDT and their relative weight for both *W* and top taggers.

Performance of top tagger





- W/Top tagger trained on tracks closest to fast-simulation (Delphes) tagger