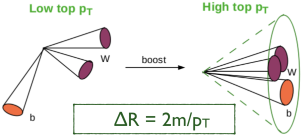
Jet sub-structures studies in FCC-hh

M. Aleksa, J. Faltova, C. Helsens, A. Henriques, C. Neubüser,
M. Selvaggi, V. Völkl, A. Zaborowska

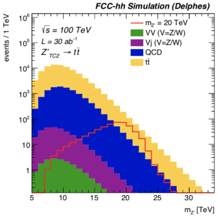
FCC Week 2019, Brussels

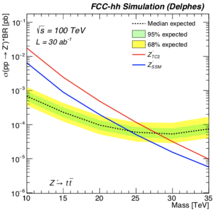
Why jet sub-structures?

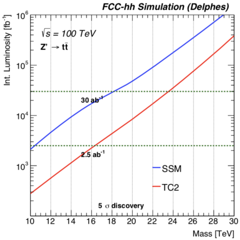
- physics case of FCC-hh: discovery machine for new heavy resonances!
- searches of heavy resonances in hadronic decay channels (`arXiv:1902.11217v2`)
  requirements:
  - id of boosted top quarks and W bosons essential!
  - collimated jets with angular sizes $< 0.01 - 0.02$
    $\rightarrow$ hard to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



$Z' \rightarrow t\bar{t}$







pre-selection:

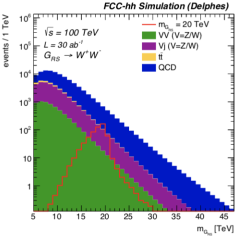
- p_T >3 TeV, $|\eta| < 3$
- m_{SD}>100 GeV
- $\tau_{21}, \tau_{32} > 0$
- 2 b-tagged jets
- di-top mass corrected for MET

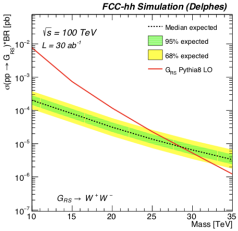
final selection based on top tagger:

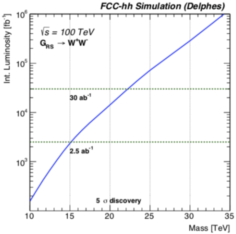
exploring jet-shapes to suppress QCD background, using multi variate analysis techniques (BDTs)

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

$Q^* \to W^+W^-$







pre-selection:

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

final selection based on W tagger:

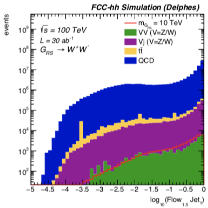
essential to find the few events in orders of magnitudes higher backgrounds

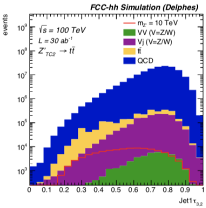
- exclusion up to 28 TeV
- discovery of 22 TeV.

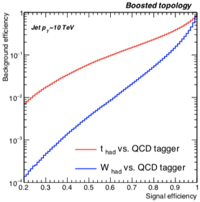
talk by C. Helsens

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` $\to$ used for t/QCD

$\to$ id of W bosons with 90 % signal efficiency, and $<$ 10 % QCD contamination

$\to$ id of top-jets with 90 % signal efficiency, and $<$ 30 % QCD contamination

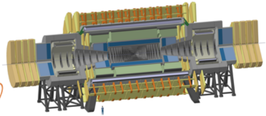
$\to$ based on..

- detector parametrisation.
- track based jets including pfa corrections

Interplay of full/fast simulation

in testing and developing a detector design

Detector model



Full simulation
in **FCCSW**

Fast simulation
Delphes

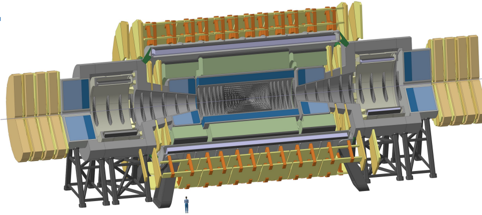
Realistic
Detector performance

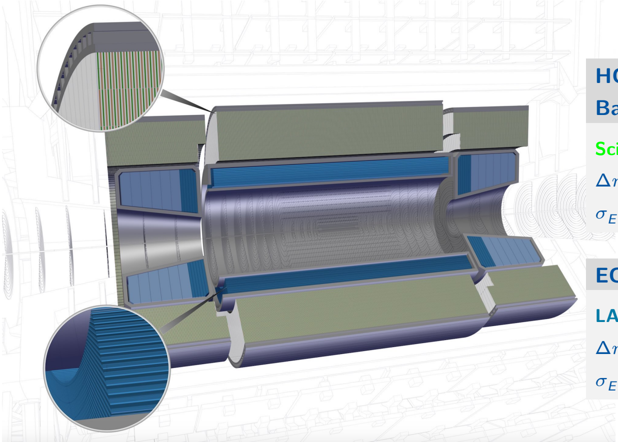
targeted Detector
performance

Full physics analysis chain
-> results (see CDR)

FCC-hh reference detector

full simulations in FCCSW
here emphasis on calorimetry





HCAL Barrel / Ext. Barrel
Sci-Pb-Steel, 10/8 layers
$\Delta\eta \times \Delta\phi = 0.025 \times 0.025$
$\sigma_E/E \sim 50\%/\sqrt{E} \oplus 3\%$

ECAL Barrel
LAr-Pb (1:3), 6-8 layers
$\Delta\eta \times \Delta\phi = 0.01 \times 0.009$
$\sigma_E/E \sim 10\%/\sqrt{E} \oplus 0.7\%$

Simulation/reconstruction in FCCSW

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

- Generator: Pythia8
- Particle propagation: Geant4 10.4.0

Reconstruction algorithms, see **talk** by V.Völkl:

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

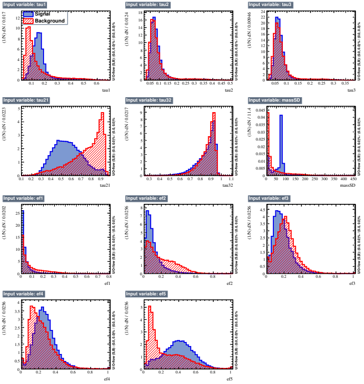
$\rightarrow$ input for FastJet package 3.3.0 for jet clustering:

- 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

11 BDT input variables

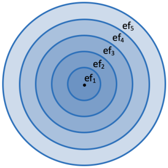
500 GeV Wqq_vs_QCD, R=1.0



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

$$ef_n = \sum_{\frac{n-1}{5}\alpha < \Delta R_i < \frac{n}{5}\alpha} \frac{e_i}{e_{sum}}$$

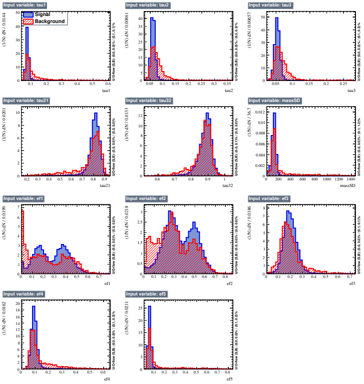
with $\alpha = 0.05$, e_i energy of the ith cell with distance ΔR_i to jet axis.



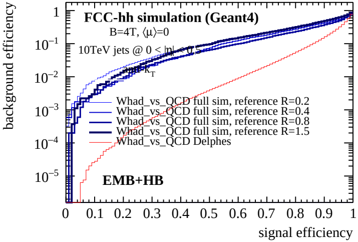
W tagger in full-simulation – calorimeter cells

11 BDT input variables

10 TeV Wqq_vs_QCD, R=0.4



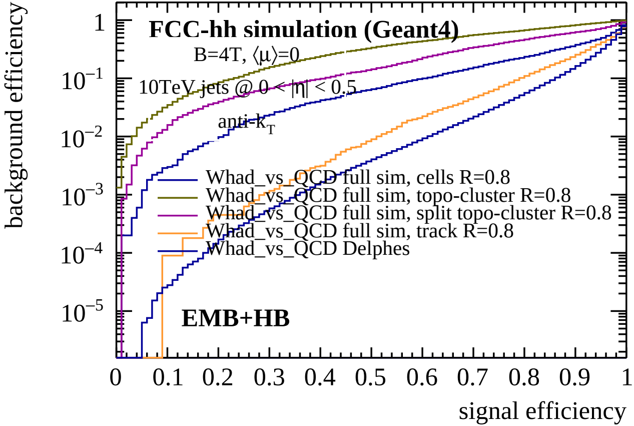
- extreme collimation of 10 TeV jets $\rightarrow \Delta R = 2m/p_T \approx 0.016$ calo cell sizes 0.01/0.025 of E/HCal
- week separation power on the sub-jettiness variable



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

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

Performance of W tagger

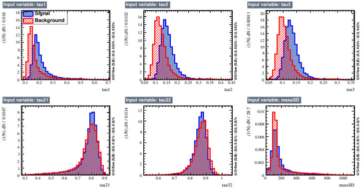


- W tagger degrades with clustering (incl. electronics noise) → need optimisation of cluster splitting

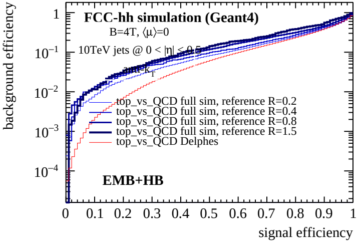
Top tagger in full-simulation – calorimeter cells

6 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



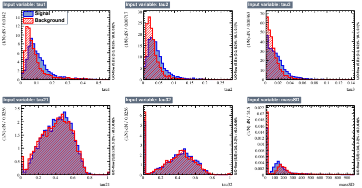
$\rightarrow R = 0.2$ best result

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

Top tagger in full-simulation – calorimeter clusters

6 BDT input variables

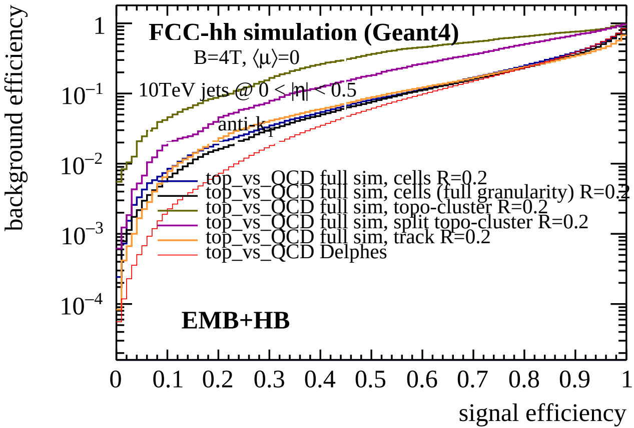
10 TeV thad_vs_QCD, R=0.2



- 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

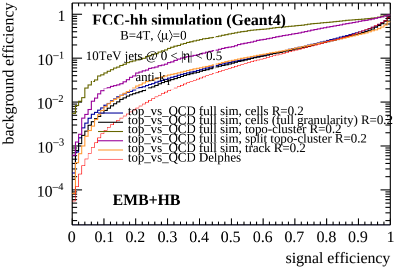
→ splitting needs p_T dependent optimisation

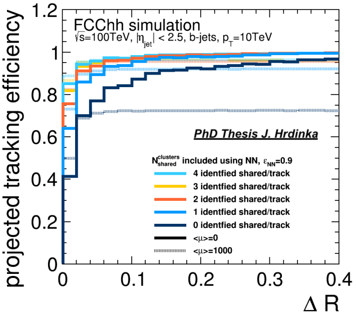
Performance of top tagger



- W/Top tagger trained on tracks closest to fast-simulation (Delphes) tagger
- !ATTENTION! assumption of perfect track reconstruction

Performance of top tagger





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

Conclusions

- jet sub-structures are crucial for jet 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 (cluster splitting)
- matching of tracks and clusters within a jet to profit from the individual strengths

Conclusions

- jet sub-structures are crucial for jet 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 (cluster splitting)
- matching of tracks and clusters within a jet to profit from the individual strengths

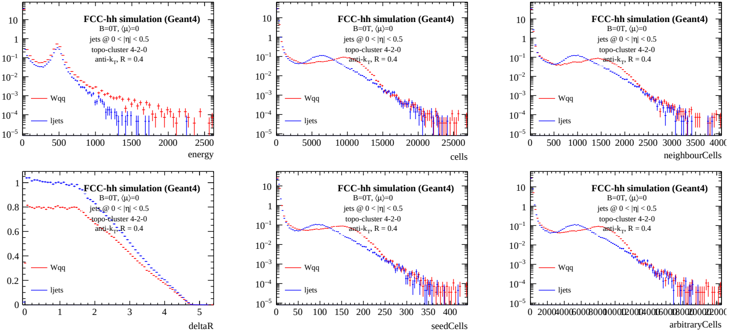
Thank you for your attention.

www.cern.ch

BACKUP

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

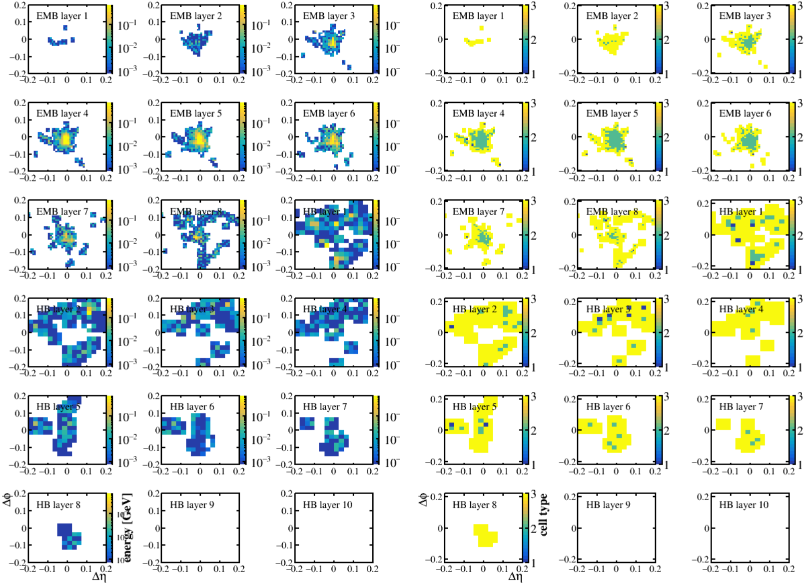
following ATLAS example [link]

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 → write out new collection of clusters

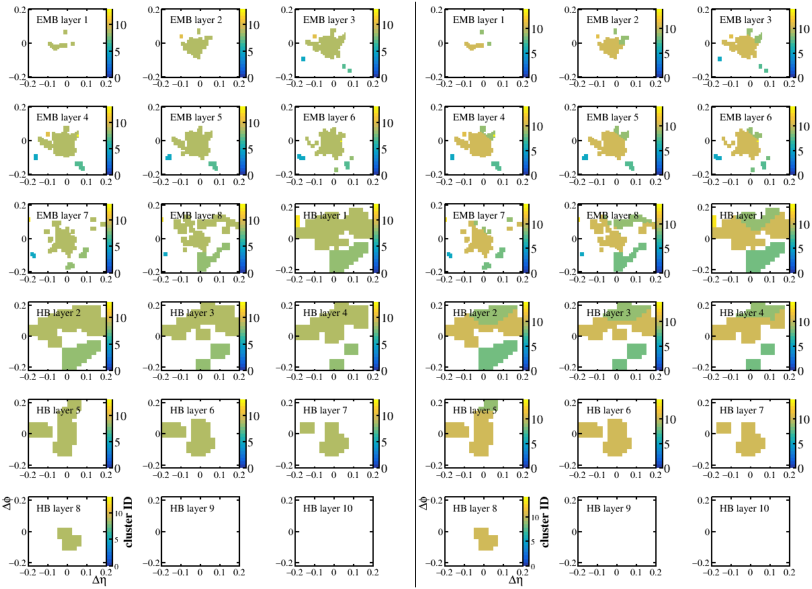
Topo-cluster splitting

example 100 GeV π^-



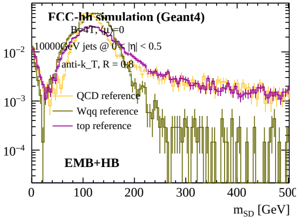
Before/after topo-cluster splitting

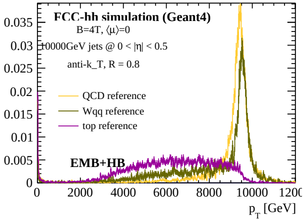
1 cluster split in 2:



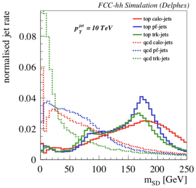
Jet mass and momentum of 10 TeV jets

calorimeter cells

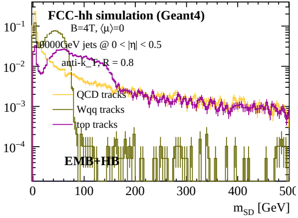


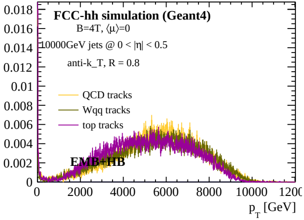


→ no efficient p_T, no m_{SD} reconstruction for $t\bar{t}$ jets



tracks (smeared charged particles)





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