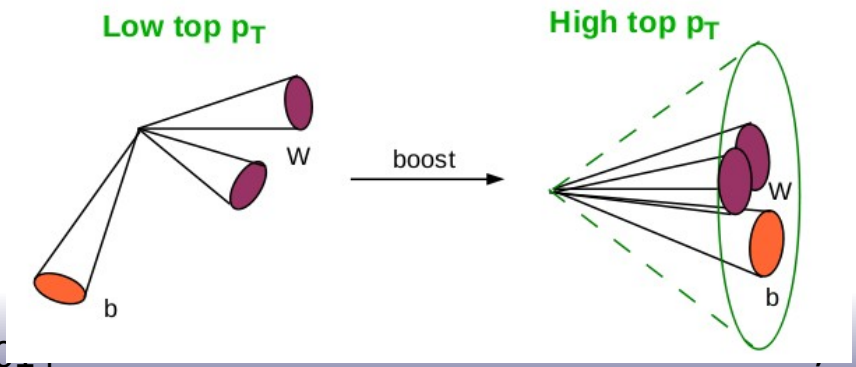
Jet substructure, top tagging & b-tagging at high pt

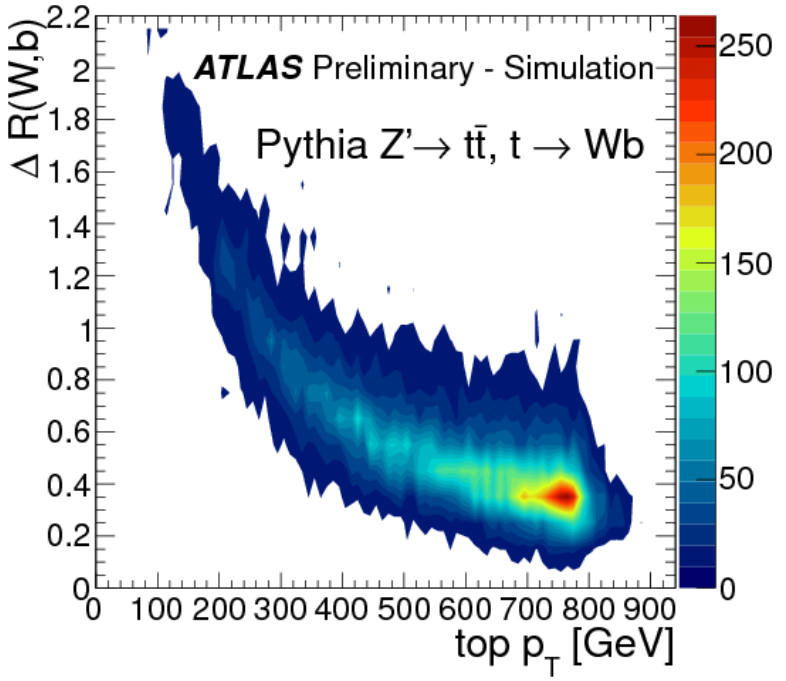
Pierre-Antoine Delsart
and Jeremy Andrea

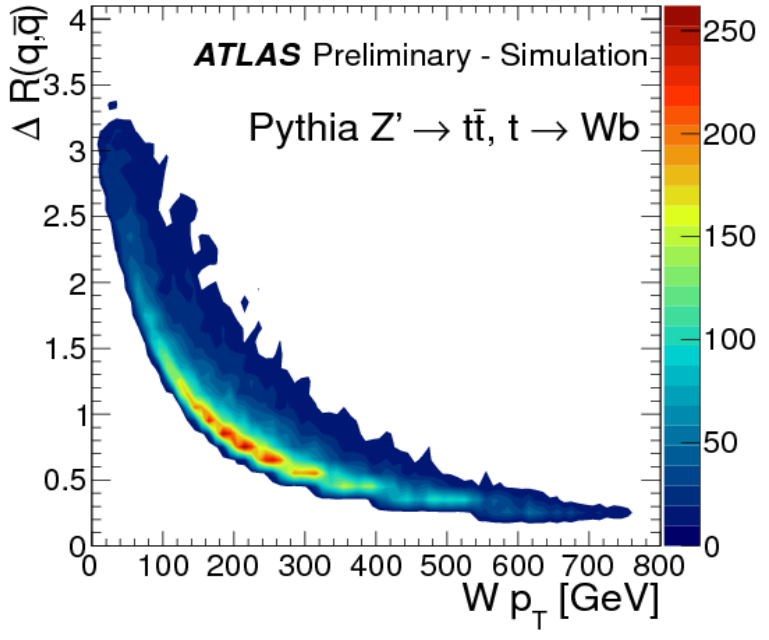
Laboratoire de Physique Subatomique et Corpusculaire
CNRS/IN2P3

Motivations

- Now exploring very high pt regions
- boosted objects
  → colimated decays
- Classic top reco techniques fail
  - Lepton isolation
  - **Hadronic jets merging**







Boosted top reconstruction

Must use large R jets because of collimation

- Pick-up more pile-up and underlying events
  - Fake large mass, decrease resolution, ...
  - Evaluate new procedures to filter/clean jet constituents

Boosted top reconstruction

Must use large R jets because of collimation

- Pick-up more pile-up and underlying events
  - Fake large mass, decrease resolution
  - Evaluate new procedures to filter/clean jet constituents

=> Jet Grooming

- A single large R jet : more QCD-induced background
  - Top decays through weak force
  - Top jets have different structures than QCD jets

Boosted top reconstruction

Must use large R jets because of collimation

- Pick-up more pile-up and underlying events
  - Fake large mass, decrease resolution
  - Evaluate new procedures to filter/clean jet constituents

=> Jet Grooming

- A single large R jet : more QCD-induced background
  - Top decays through weak force
  - Top jets have different structures t

=> Jet Substructures

- Combine substructure variables and kinematics of top's decay products

Boosted top reconstruction

Must use large R jets because of collimation

- Pick-up more pile-up and underlying events
  - Fake large mass, decrease resolution
  - Evaluate new procedures to filter/clean jet constituents

=> Jet Grooming

- A single large R jet : more QCD-induced background
  - Top decays through weak force
  - Top jets have different structures t

=> Jet Substructures

- Combine substructure variables and kinematics of top's decay products

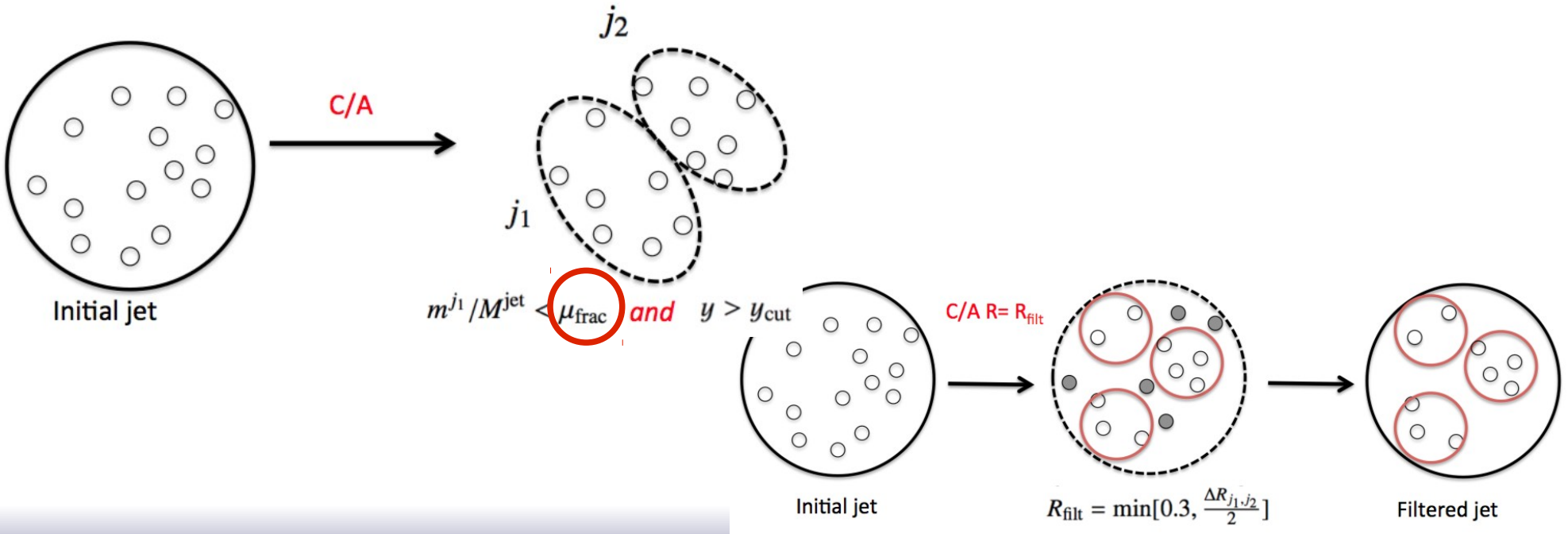
=> Top tagging

- Is B-tagging possible for highly boosted tops ?

Jet Grooming techniques

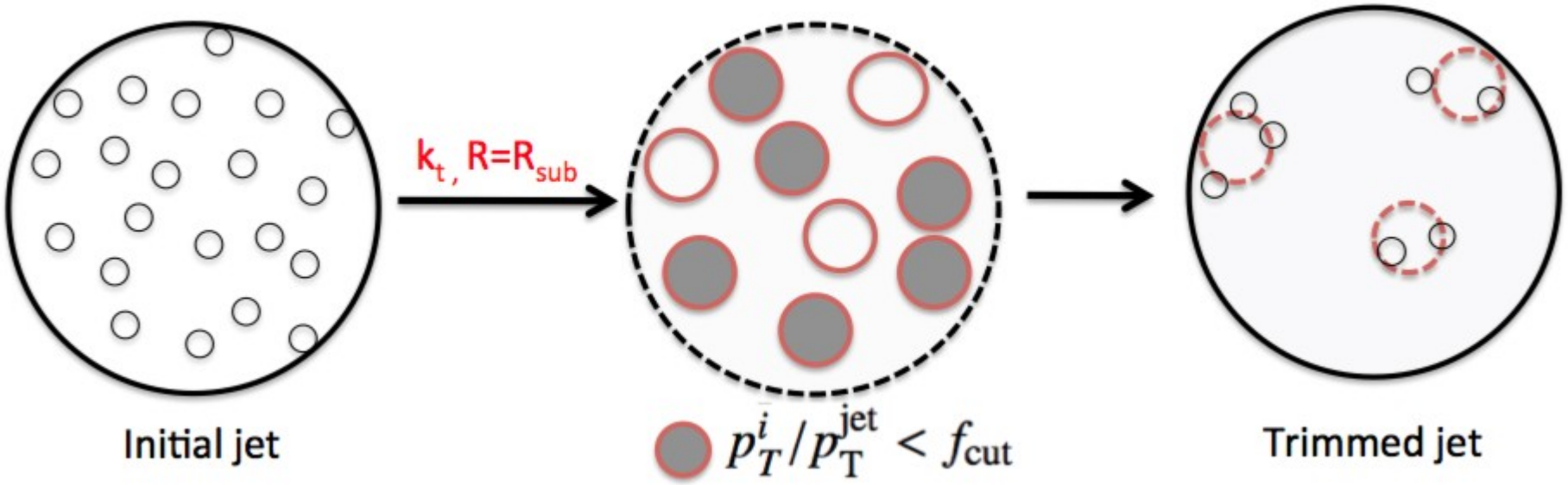
3 grooming techniques have been studied by Atlas

- **Mass-drop**
  - Find 2 symmetric subjets
  - If found, recluster their constituents with small R, keep 3 highest subjets



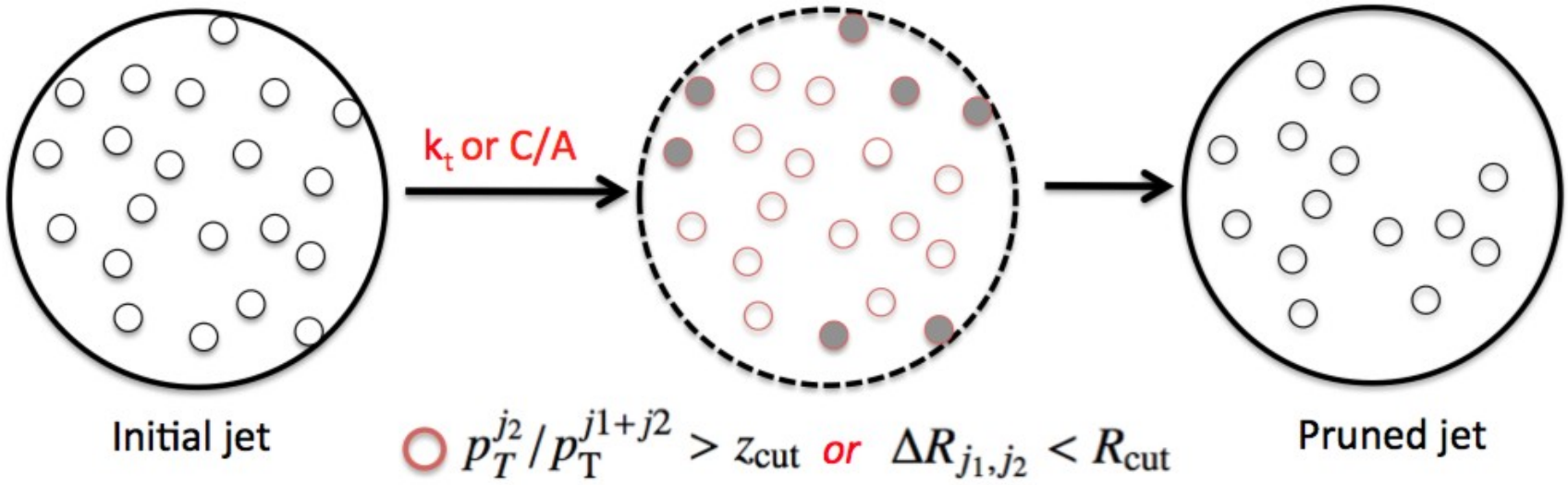
Jet Grooming techniques

- **Trimming**
  - Recluster jets with small R kt alg
  - Remove subjets with low pt fraction. Remaining constituents form the trimmed jet.



Jet Grooming techniques

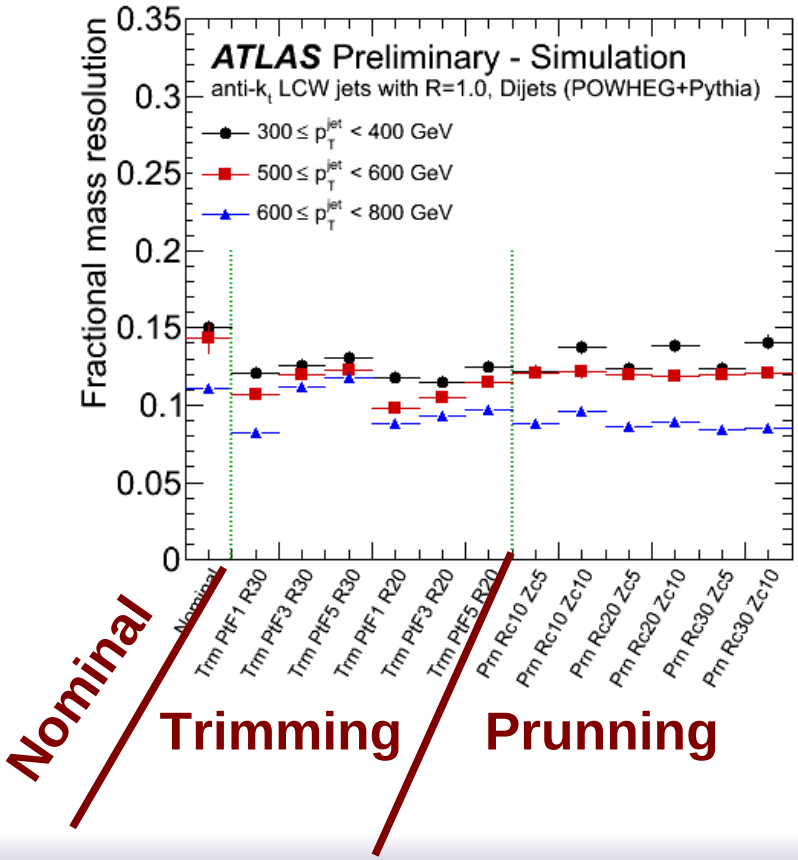
- **Pruning**
  - Recluster jet constituent with modified Kt or Cam alg
  - Modified alg requires balance in pt or not too large angle before merging 2 proto-jets. Else, discard low pt protojet.

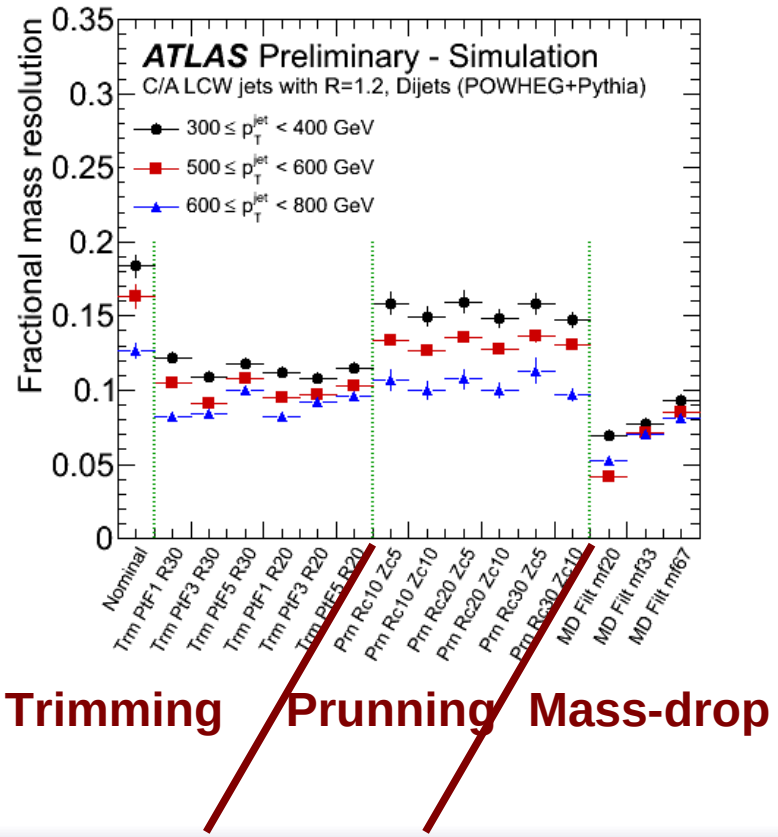


Grooming performances

Studied in Atlas on inclusive jet samples (Data and MC)

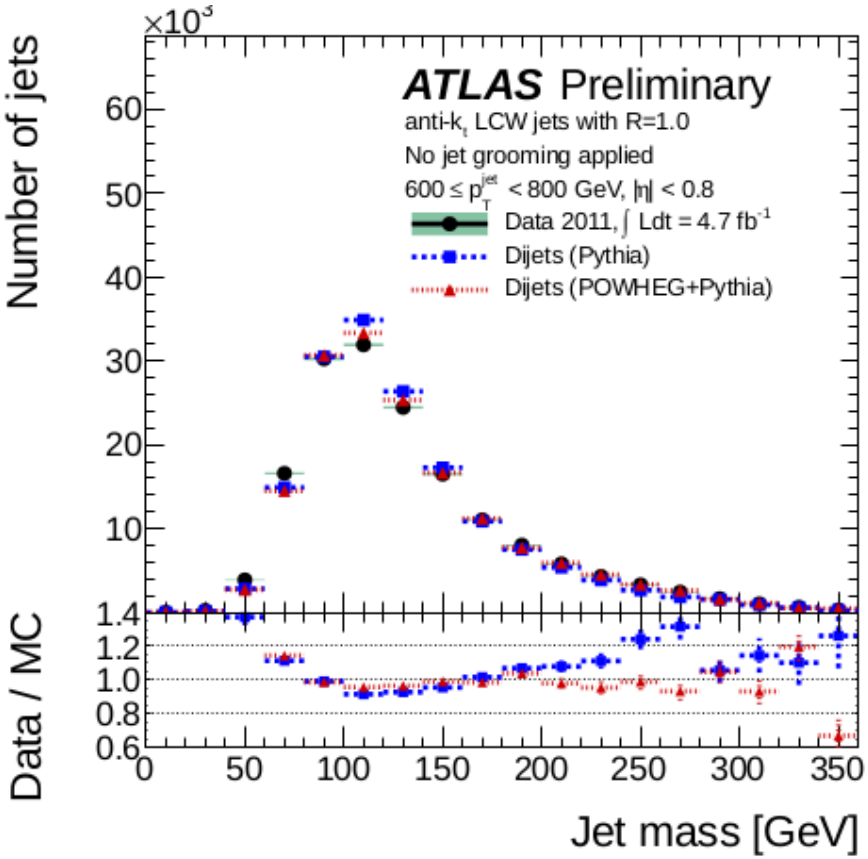
- **Jet mass resolution** $\sigma\left(\frac{Mass_{truth} - Mass_{reco}}{Mass_{truth}}\right)$



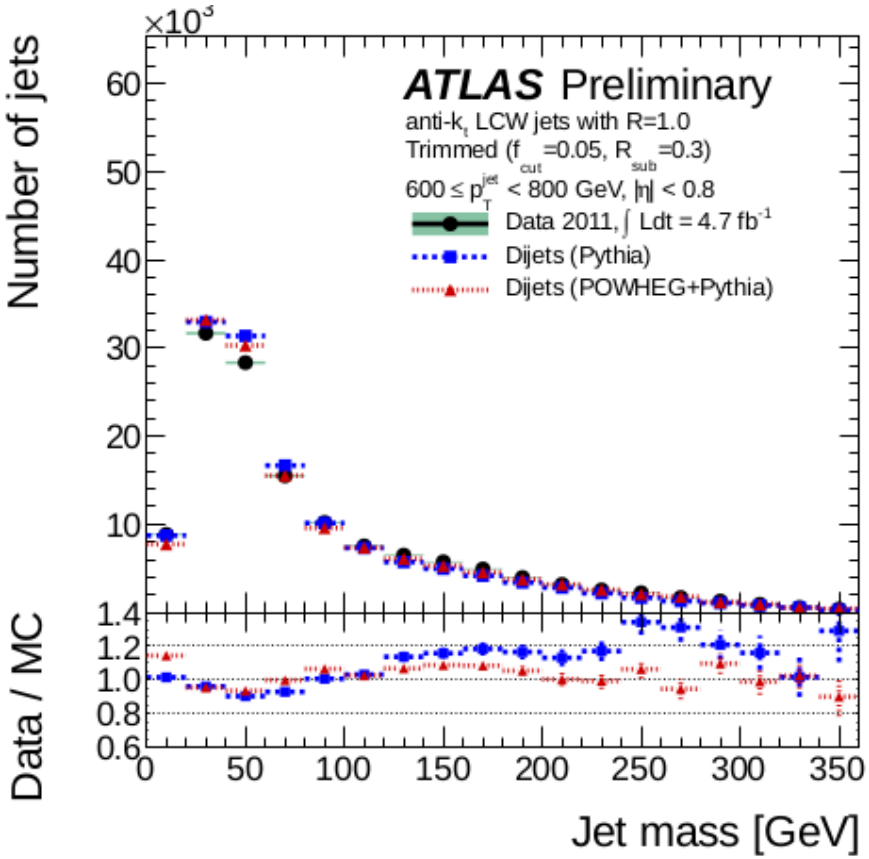


Grooming performances

Data/MC, effect on mass distribution



No grooming



Trimmed

Jet Grooming usage at LHC

- Atlas prefers Trimming and Mass-drop
  - better mass resolution
  - less sensitive to PU
- Atlas uses trimmed jets in physics searches (ex : ttbar resonnances)
- CMS uses pruning as part of "partial tagging"
- Pruning also uses in some substructure calculations (ex : Q-jets)

Substructure variables

Many substructure variables invented.

Atlas and CMS studied several of them

- Jet Mass
- Kt split scale
- N-subjettiness
- Q_w
- Shape variables
- Shower deconstruction
- Q-jets
- ....

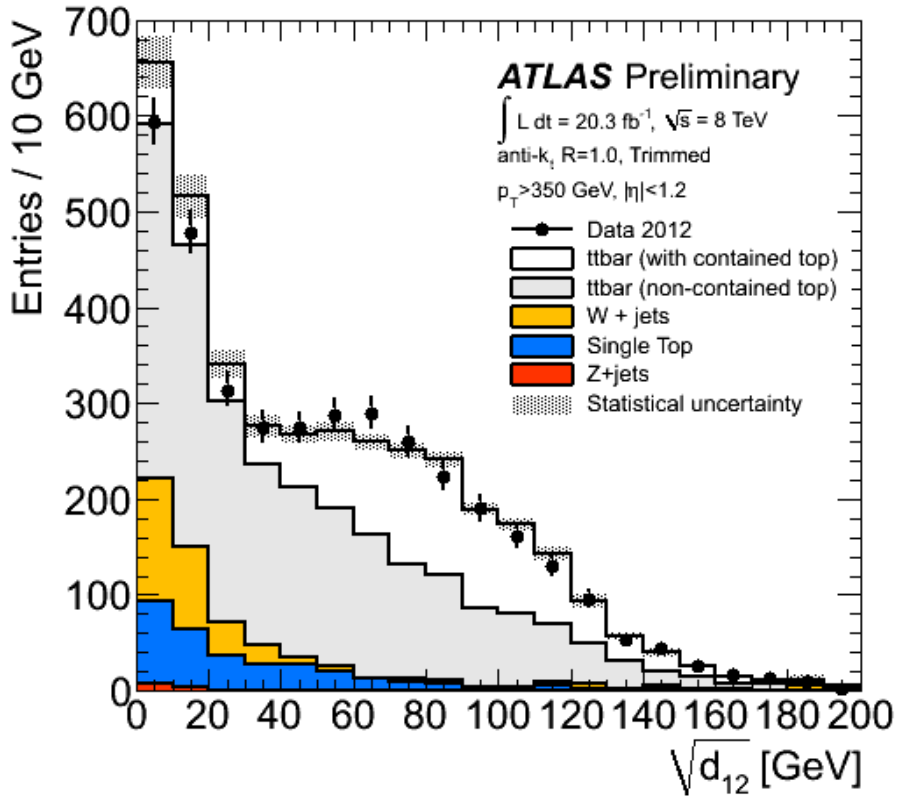
- Show only a selection based on recent public results

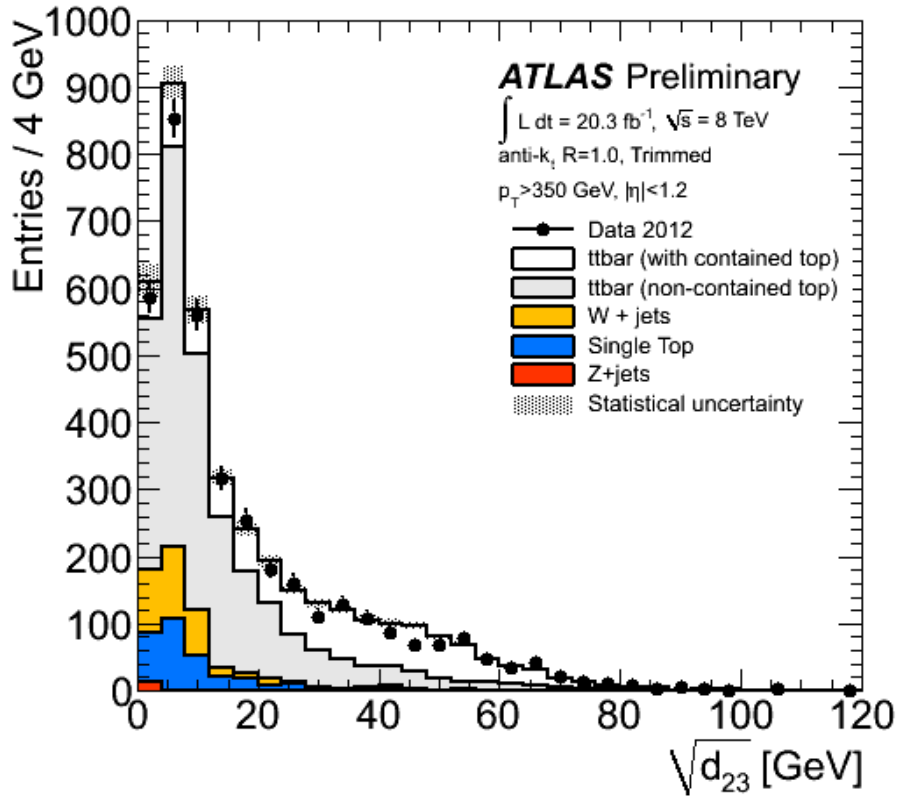
Substructure : d_{12} (split scale)

- Last recombination distances used in Kt alg :

$$\sqrt{d_{ij}} = \min(p_{Ti}, p_{Tj}) \times \Delta R_{ij}$$

Example : in a boosted ttbar selection :





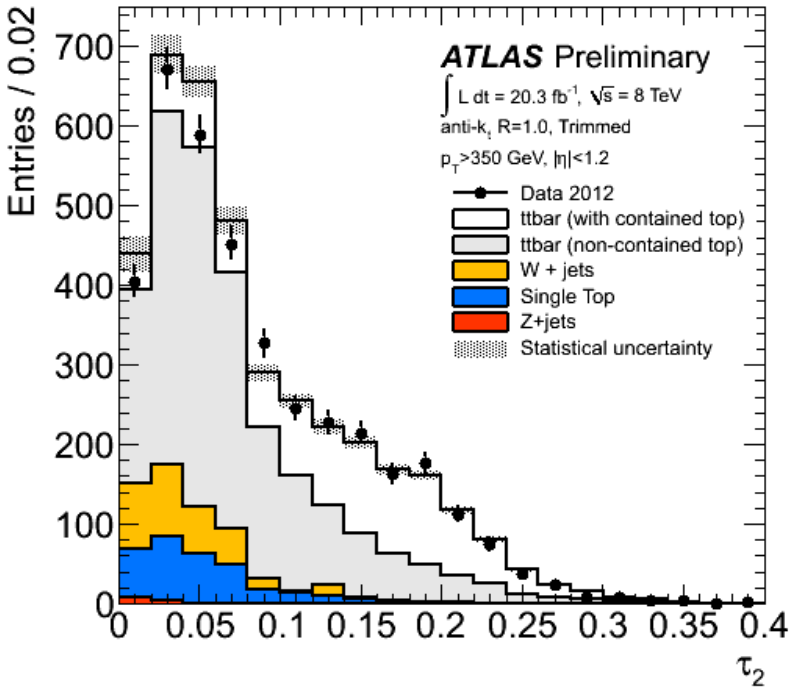
Substructure : n-subjettiness

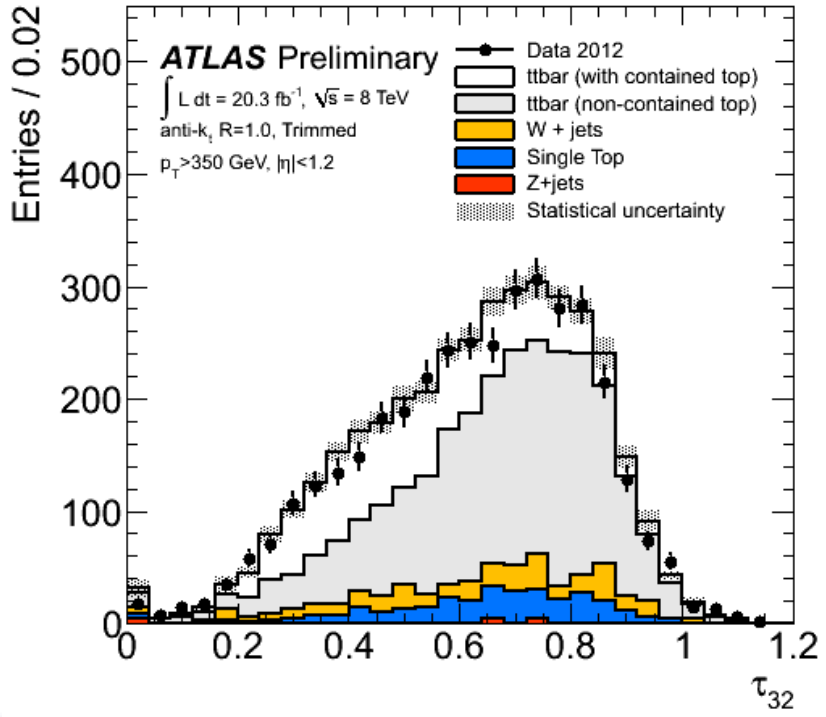
- Recluster a jet in N kt subjets, then compute :

$$\tau_N = \frac{1}{d_0} \sum_k p_{Tk} \times \min(\delta R_{1k}, \delta R_{2k}, ..., \delta R_{Nk}) \,, \text{ with } d_0 \equiv \sum_k p_{Tk} \times R$$

  - Consider ratios of these quantities (ex : $\tau_{21} = \tau_2/\tau_1$)

Example : in a boosted ttbar selection :

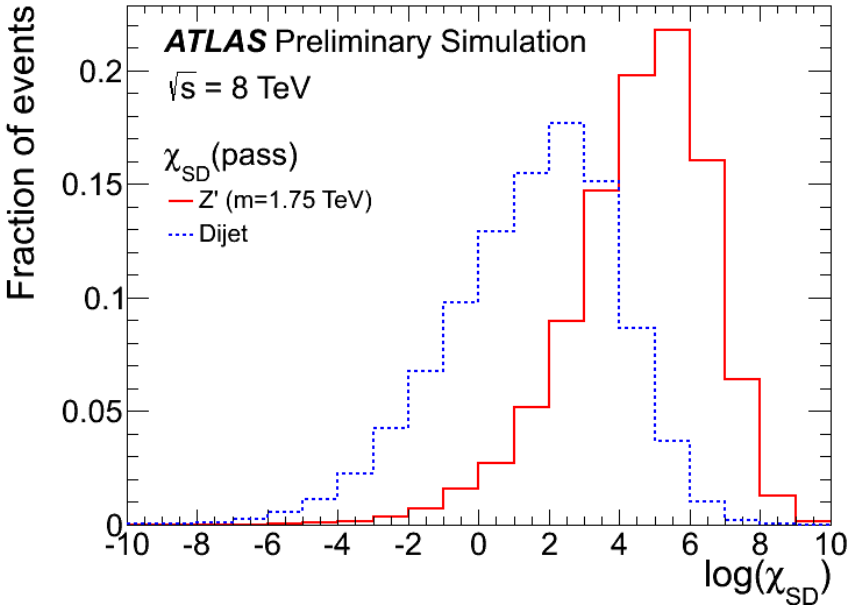


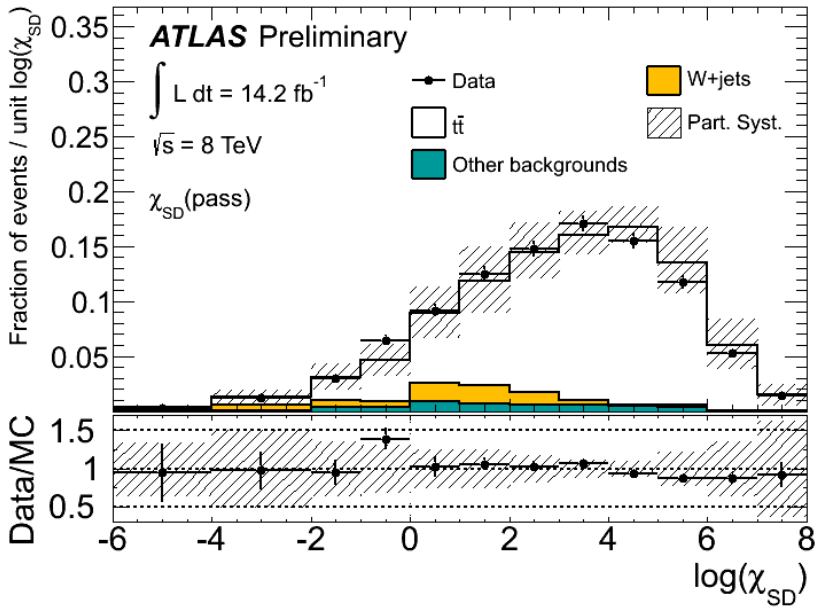


Substructure : shower deconstruction

- Cluster small subjets from a large jet => configuration of N subjets
- Calculate probabilities for this subjet configuration assuming signal (**top**) or assuming background (**QCD**)
  - Probabilities calculated as in parton shower generator, using Sudakov factors
- Build a likelihood ratio from these proba χ_{SD}

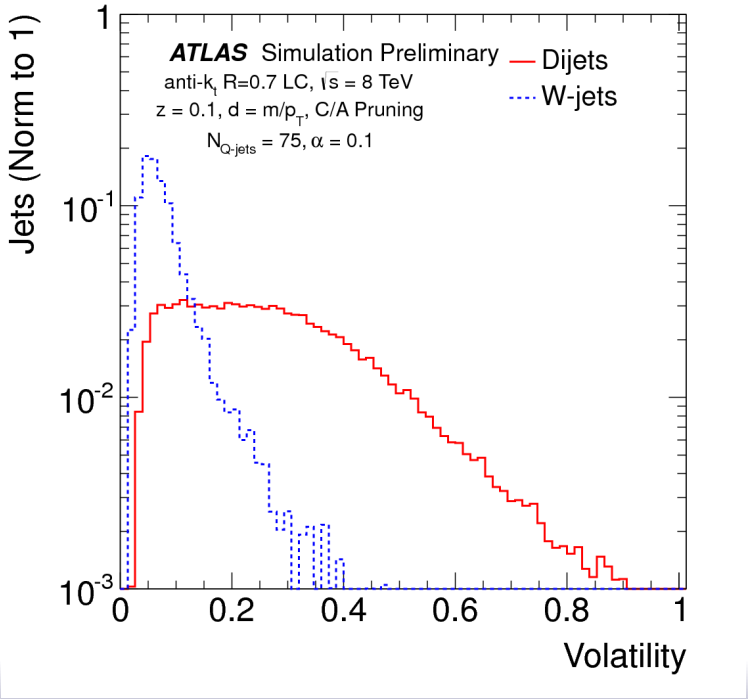
D. E. Soper and M. Spannowsky





Substructure : Q-jets

- Recluster jet constituents several times introducing randomness in merging criteria
  - Select pair for merging randomly as in $\exp\left\{-\alpha \frac{d_{ij} - d^{\min}}{d^{\min}}\right\}$
  - This simulates different possible showering history
- Obtain a mass distribution for a given jet : take the RMS



Ellis et al.

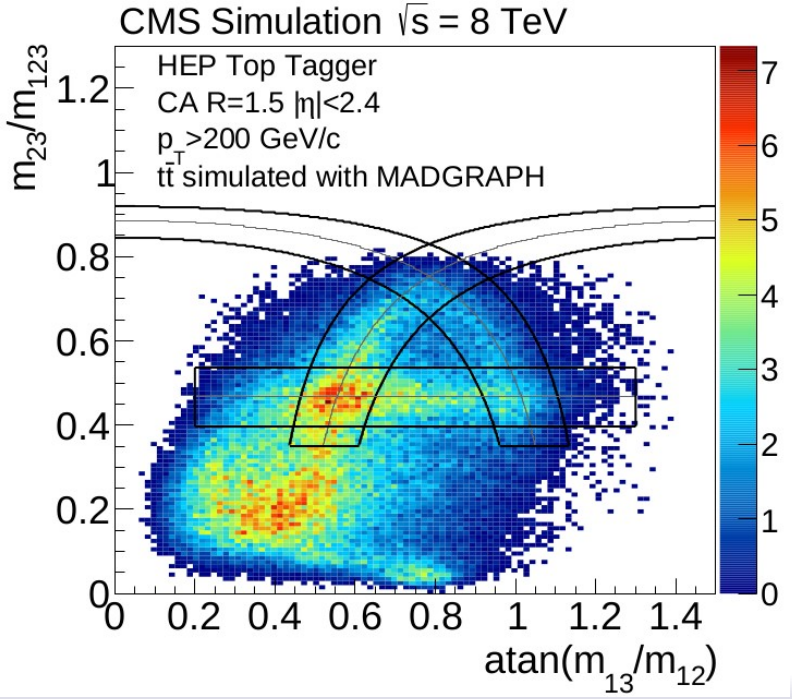
So far used only in boosted W studies

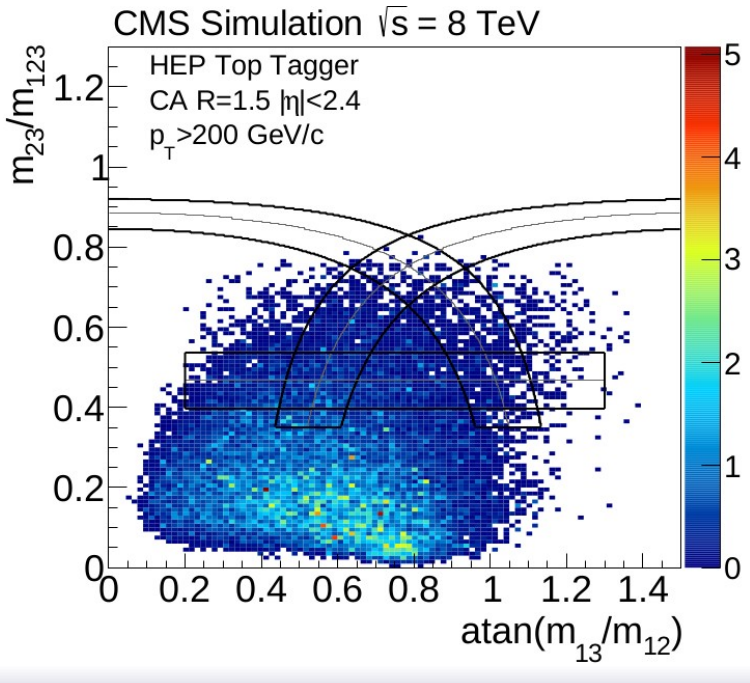
Top tagging

- Substructure variables are used as discriminant to select boosted tops
  - Ex : d_{12} part of event selection in Atlas ttbar resonances search
  - Combination of variables are being studied
- More complex procedure have been proposed : Top taggers
  - HepTopTagger
  - CMS TopTagger
- CMS also investigated « partial top tagging »
  - i.e btagging + tagging a merged W jet

HEPTopTagger

- Multi steps algorithm on large R(1.5) Cam jet
  - Mass-drop technique to identify hard subjets
  - Filtering of soft components
  - Kinematic constraints on final subjets (W, top mass...)
- If succeed, results in 3 identified subjets





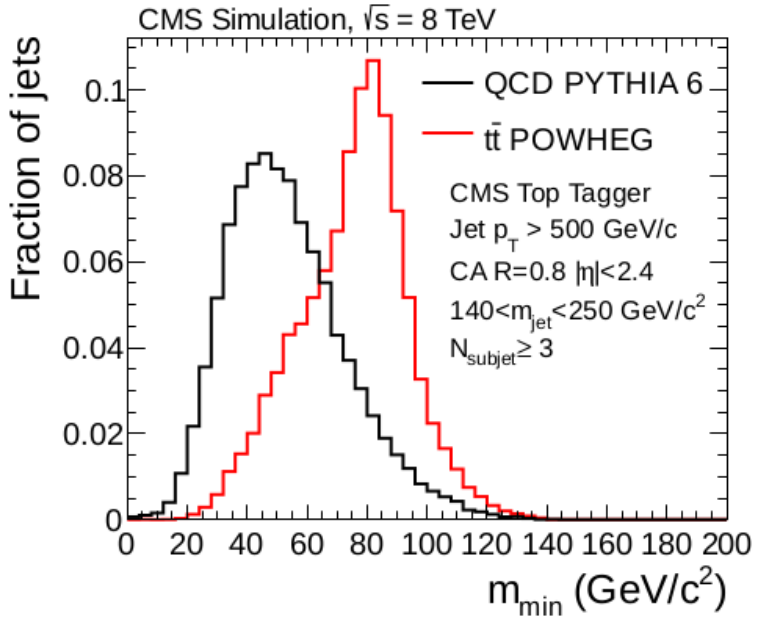
Plehn et al.

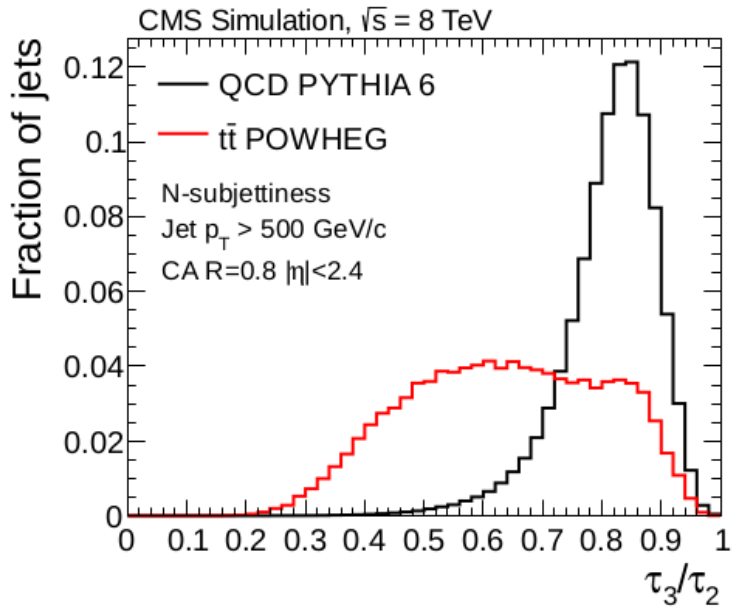
CMS Top Tagger

Starts from a lar R Cam algorithms

- Iteratively undo the merging until 2 balanced & separated subjets are found (else fails)
- Re-decompose these subjets
- If succeed, results in 3 or 4 subjets
- Variables : all subjet mass, minimal mass of pair of subjets

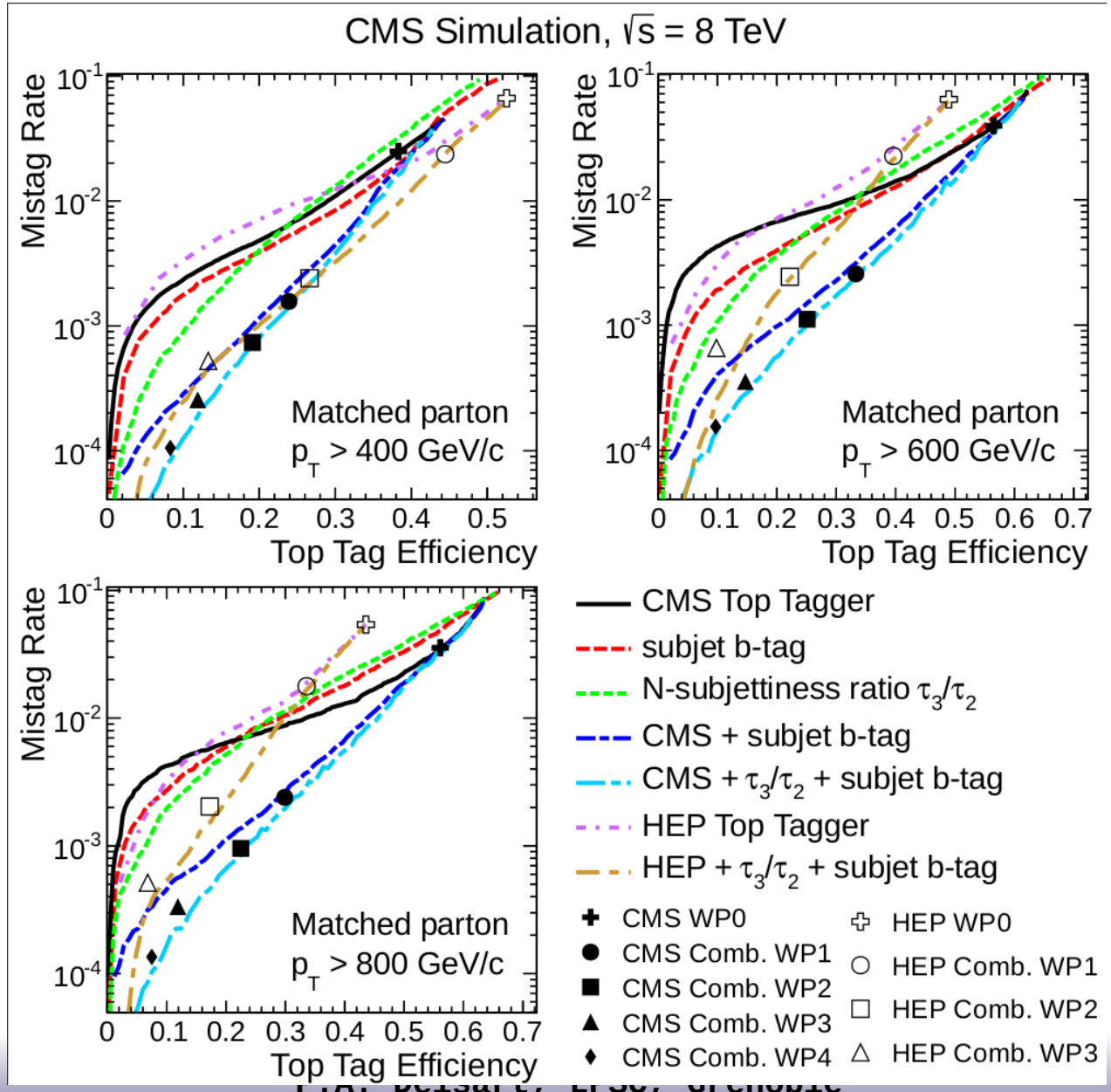
Kaplan et al.



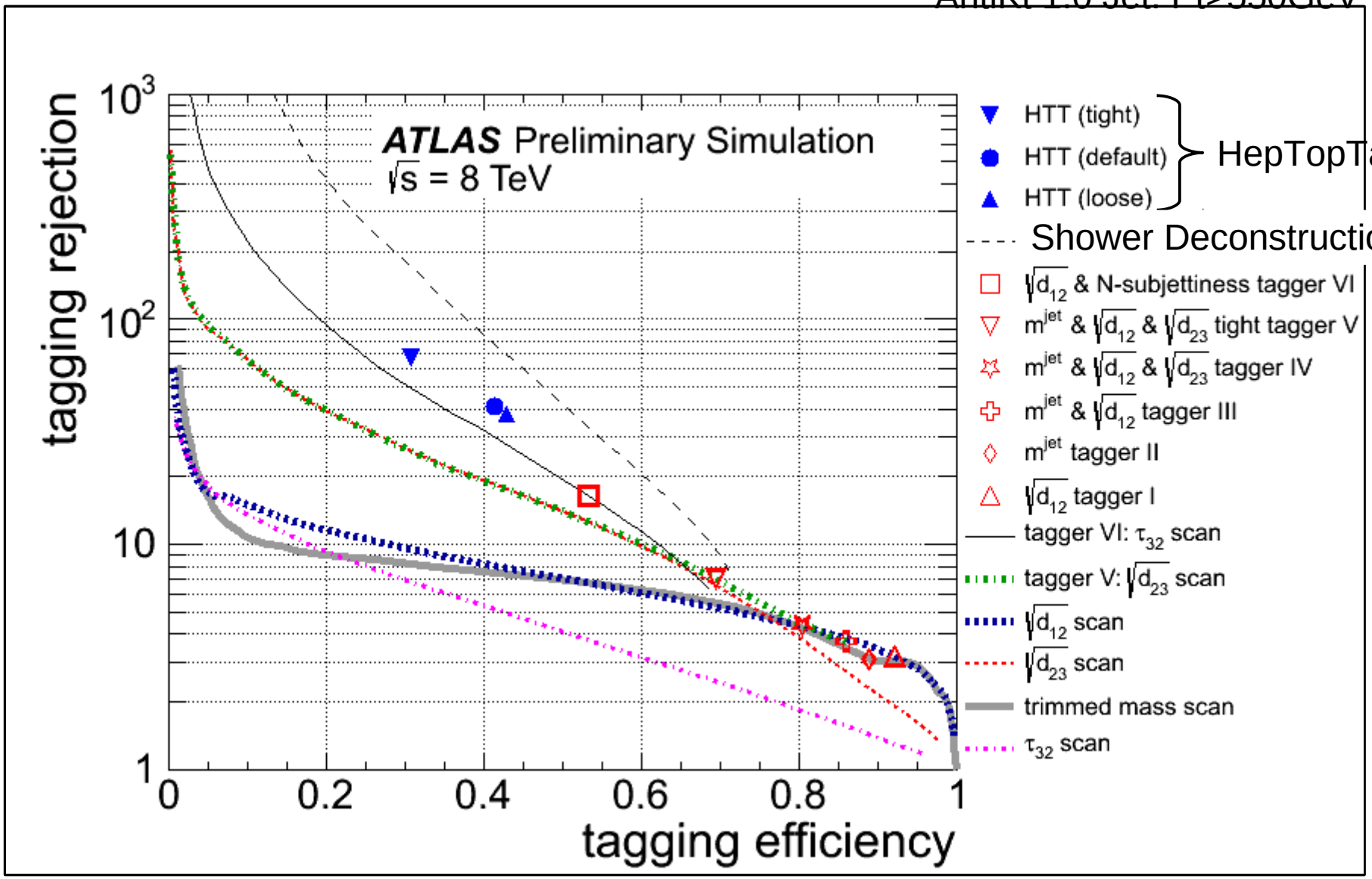


Top Tagging performances (CMS)

- CMS also combines tagger with n-subjetiness and/or b tagging



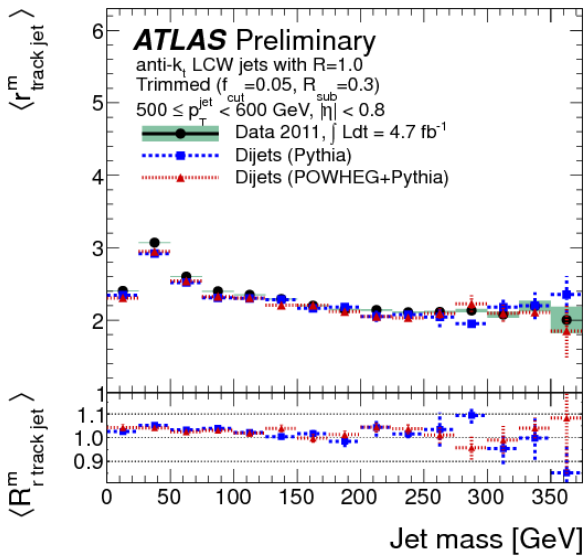
Top Tagging performances (Atlas)

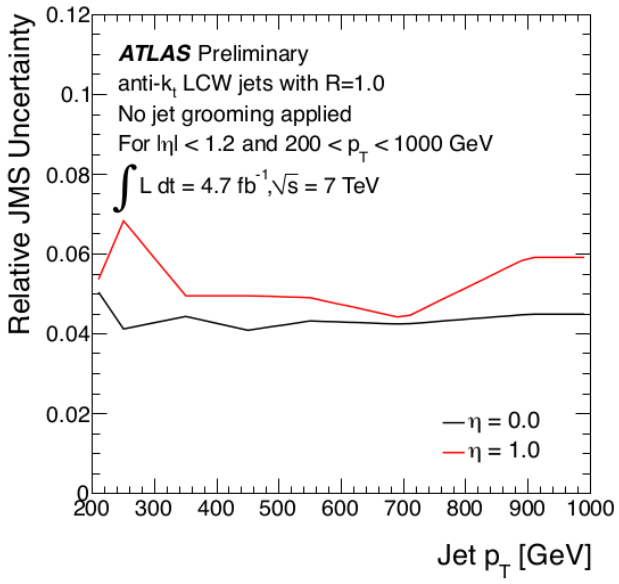


Systematics & substructures

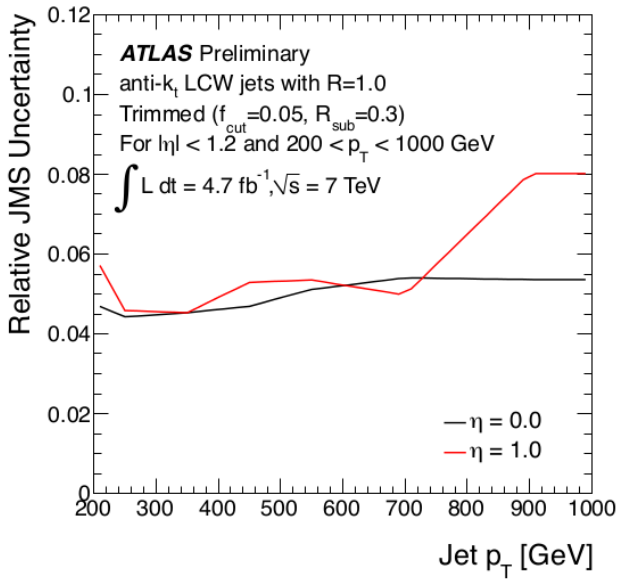
- Atlas evaluates uncertainties on substructure vars with "double ratios" = $(V_{calo}/V_{track})_{data}$ / $(V_{calo}/V_{track})_{MC}$

- Example : Mass





(a) anti-k_t, R = 1.0 (no jet grooming)



(b) anti-k_t, R = 1.0 (trimming)

- Have not seen public discussions on
  - Systematics for taggers, combination of vars & correlations
  - How final sensitivity in physics analysis is affected

Conclusions

- Understanding boosted topologies becomes mandatory
- Lots of ideas have been tested at the LHC
  - Many perf studies
  - New techniques start to reach physics analysis
- Still room for improvements and work
  - Multivariate combinations ?
  - Understanding of correlations and systematics...

- Topic even more important at 14TeV and increased PU
  - Detector resolution limits ?
    - Atlas can make more use of tracks (and eflow)

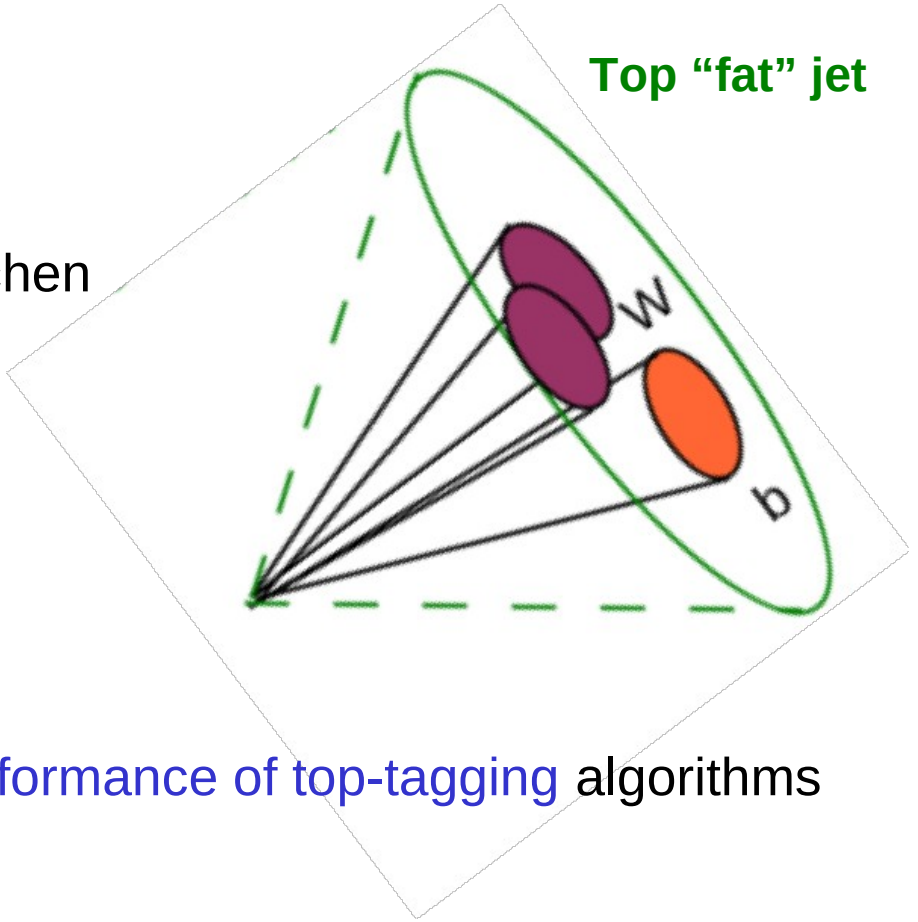
- Atlas references :
  - W-tagging: https://cds.cern.ch/record/1690048
  - top-tagging: https://cds.cern.ch/record/1571040
  - shower deconstruction: https://cds.cern.ch/record/1648661
  - Q-jets: https://cds.cern.ch/record/1572981
- CMS references
  - Boosted Top Jet Tagging
  - Jet Substructure Algorithms

B-tagging in boosted object @CMS

BTV-13-001

- At high p_T, jets from top decays start to be collimated.

- => Top-tagging such as the Cambridge/Aachen (CA) algorithm :
  - Search for 3 sub-jets (W->qq', b) in a fat jet (cone size 1.5).
  - Allows moderately boosted top quark to be reconstructed, down to 200 GeV.

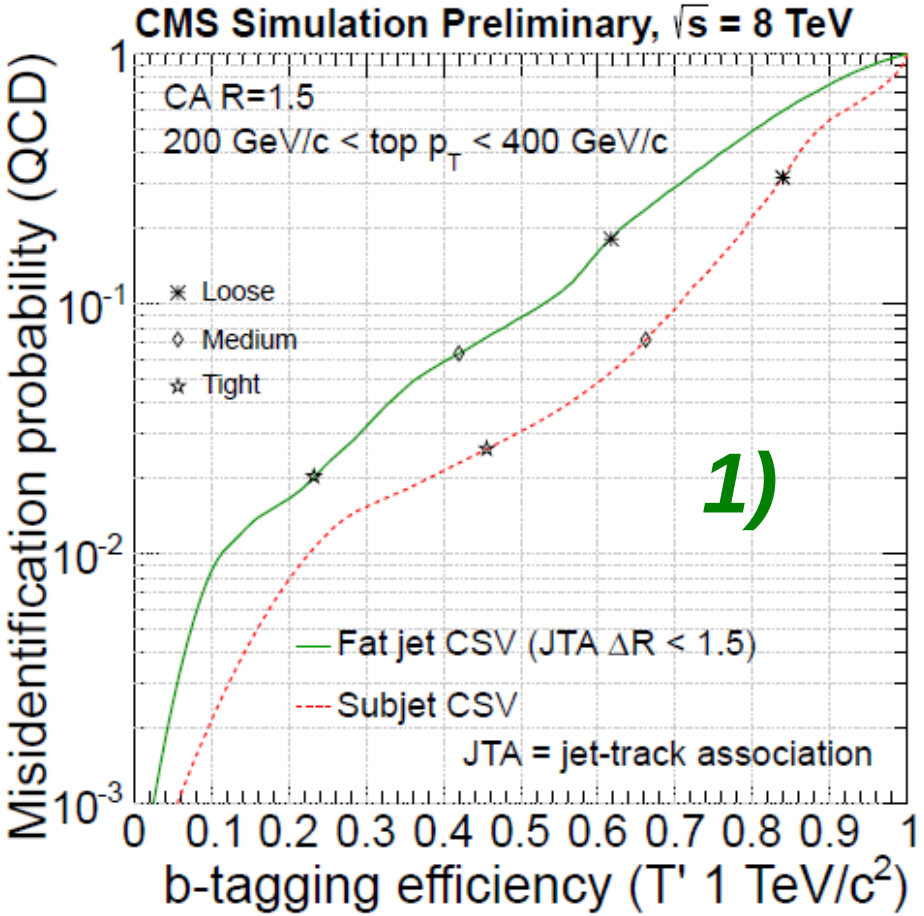


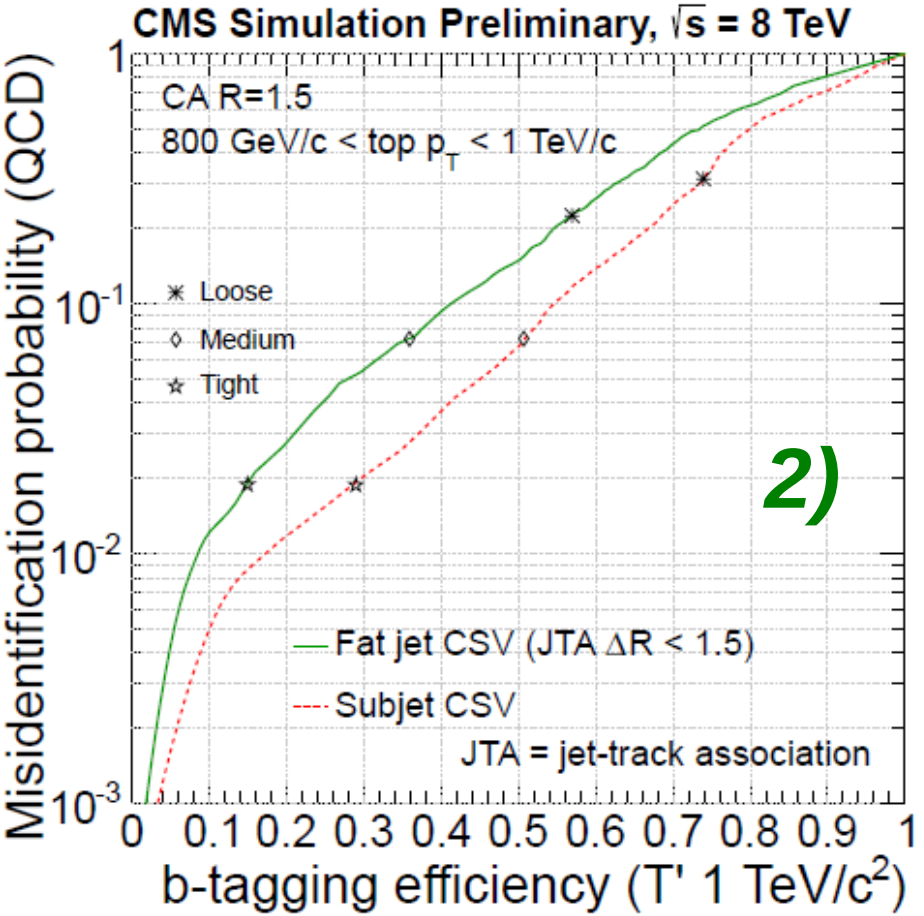
- Can b-tagging be used to improve the performance of top-tagging algorithms for boosted topologies ?

- Approach of the CMS collaboration :
  - Use standard b-tagging algorithm from all the tracks associated to the fat-jets (jets of cone ΔR=1.5) : poor performance.
  - Use standard b-tagging algorithm from tracks associated to sub-jets (ΔR=0.3) : better performance.

B-tagging in boosted object @CMS (2)

- B-tagging performance from fat-jets (green) and from sub-jets in fat-jets (red) for moderate top boost ***1)*** and high boost ***2).***





B-tagging in boosted object @CMS (3)

- Top-tagging efficiency ***1)*** and top-mistagging rate ***2)*** as a function of the reconstructed fat-jets p_T, asking for at least one b-tagged sub-jet.

- Large diminution of the mistagging rate for a moderate loss of efficiency.