A boost to Higgs Physics: new regimes at high energy

Silvia Biondi
University & INFN of Bologna
silvia.biondi@cern.ch / silvia.biondi@bo.infn.it

Corso di Dottorato - 2019/2020

Course outline

- Theory reminder
- Higgs boson production and decay modes
- Higgs boson discovery by ATLAS and CMS
- Higgs boson mass measurement by ATLAS and CMS
- Overview of ATLAS and CMS analyses about Higgs
- **Signal/background discrimination techniques**
  - **boosted regimes**
    - **tagging, large-radius jets substructure, re-clustering**
  - multivariate analysis and deep neural network
- Signal extraction techniques
  - likelihood and test statistic
  - CLs method
- ttH analysis: an example

Signal/background discrimination techniques

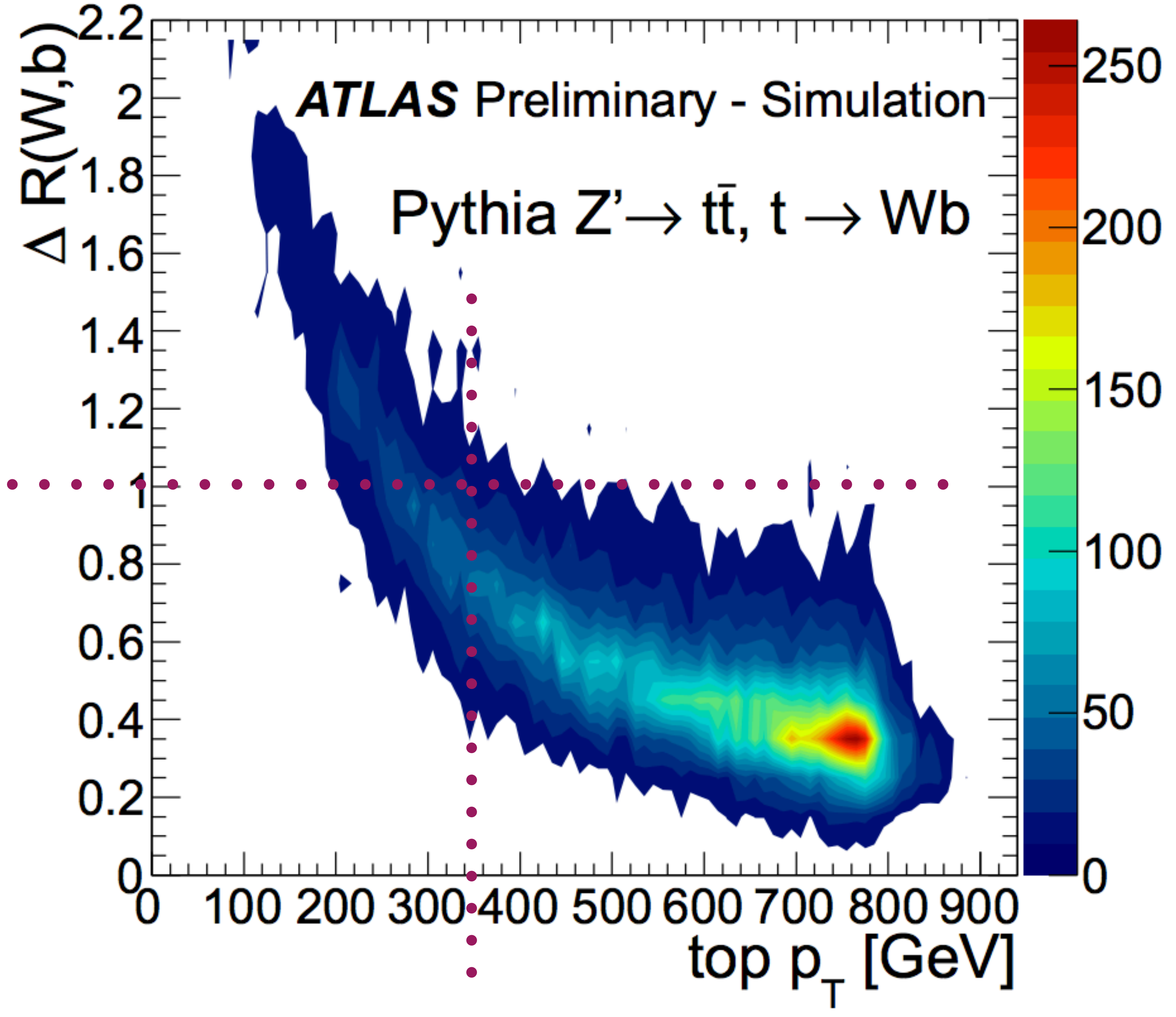
Boosted regimes

Importance of the topology

- during the Run 2, increasing more and more the energy, LHC is **exploring a completely new physics regime**;
- available center-of-mass energy far exceeds the masses of known SM particles;
- heavy particles are often produced with large transverse momentum (**boosted particles**) that implies large Lorentz boost for their decay products;
- they are **collimated to the decaying particle direction** in the detector rest frame.

Example: Z' → tt ($m_{Z'}$ = 1.6 TeV)

- true angular separation between W and b decay products as a function of top p_T
- for top p_T > 350 GeV, the decay products of the hadronic top quark tend to have a separation $\Delta R < 1.0$;
- **not possible anymore to resolve W from b quark!**



Boosted regimes

Consequences of higher energies

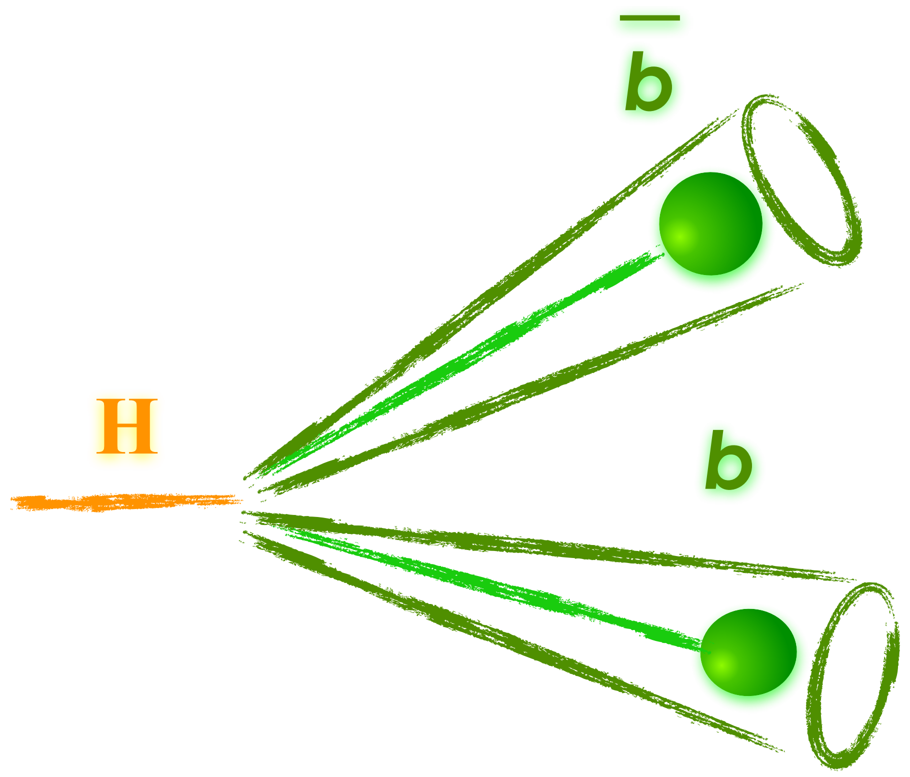
- **traditional reconstruction algorithms loose significantly the efficiency**, due to overlapping of jets coming from hadronic decay of a mother particle;
- the ability to resolve the individual hadronic decay products using standard narrow-cone jet algorithms begins to degrade.
- At high p_T, the decay products of a hadronically decaying object merge into a single, **energetic and large radius jet (large-R jet)**:

Angular separation of the decay products

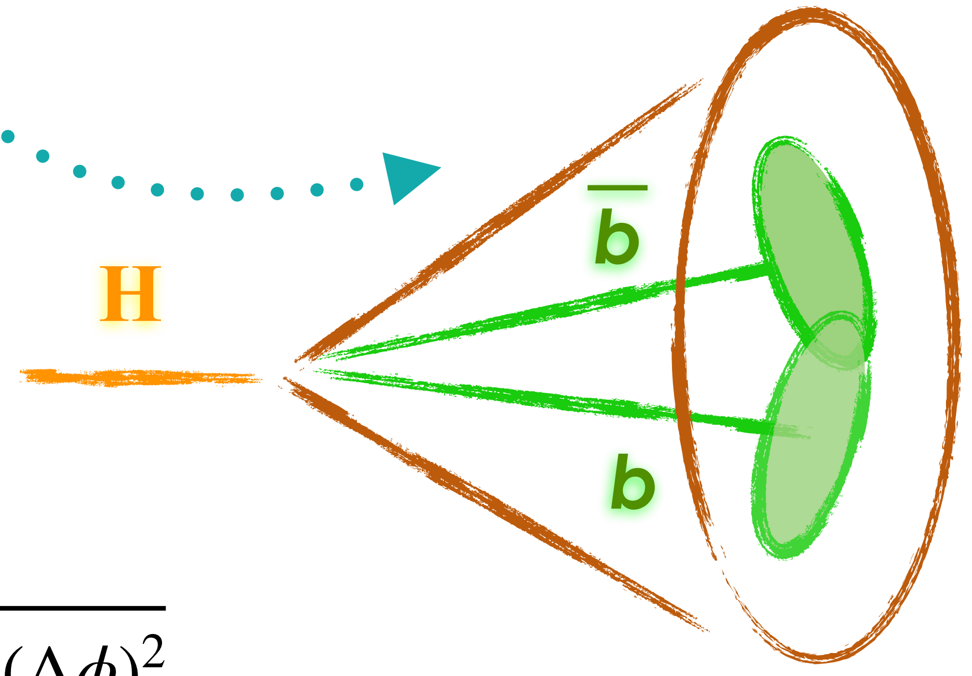
$$\Delta R_{decay} \simeq 2m/p_T$$

m and p_T of the **decaying particle**

Standard (small-R) jets



Boosted (large-R) jets



* $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$

Advantages of high p_T

- significantly **reduced combinatorial background** in busy final states;
- a single large-R jet containing all decay products of a massive particle will have significantly different properties than a single large-R jet originating from light-quark or gluon;
  - characteristic 2-body or 3-body decays of a vector boson or top quark result in a hard **substructure** that can be more resolved by removing soft radiation from jets ("grooming");
- **tagging algorithms** reliable to recognise the origin of a large-R jet
  - increase efficiency and purity in high energy analyses.

Large-R jets: substructure

Importance of the internal structure of a jet

- widest application of jet substructure tools is to **disentangle different kinds of jets**;
- **isolating boosted W/Z/H or top jets** (our signal) from the much more abundant QCD background of "standard" quark and gluon-initiated jets;
- aims to study the **internal kinematic properties of a high-p_T jet** in order to distinguish whether it is more likely to be a signal or background jet;
- large variety of methods have been proposed over the last ten years;
- can be grouped into **three wide categories**, according the physical observation that they mostly rely on:

Category I: prong finders

- when a boosted massive object decays into partons, all the partons typically carry a sizeable fraction of the initial jet transverse momentum, resulting in **multiple hard cores in the jet**;
- quark and gluon jets are dominated by the radiation of soft gluons, and are therefore mainly single-core jets.
- **look for multiple hard cores in a jet**, hence reducing the contamination from "standard" QCD jets.

Category II: radiation constraints

- second main difference between signal and background jets is their **colour structure**;
- **QCD radiation associated with an EW-boson jet, which is colourless, it is expected to be less than what we typically find in a QCD jet**.

Category III: groomers

- because of boosted jets large area, these jets are particularly sensitive to **soft backgrounds**;
- **removing the soft radiation far from the jet axis**, where it is the most likely to come from a soft contamination rather than from QCD radiation inside the jet.

Large-R jets: substructure

Looking inside jets

N-subjettiness

- jet shape that aims to discriminate jets according to the **number N of sub-jets *i*** (reconstructed from constituents *k* of the jet) they are made of

$$\tau_N = \frac{1}{d_0} \sum_k p_{Tk} \times \min(\delta R_{1k}, \delta R_{2k}, \ldots, \delta R_{Nk})$$

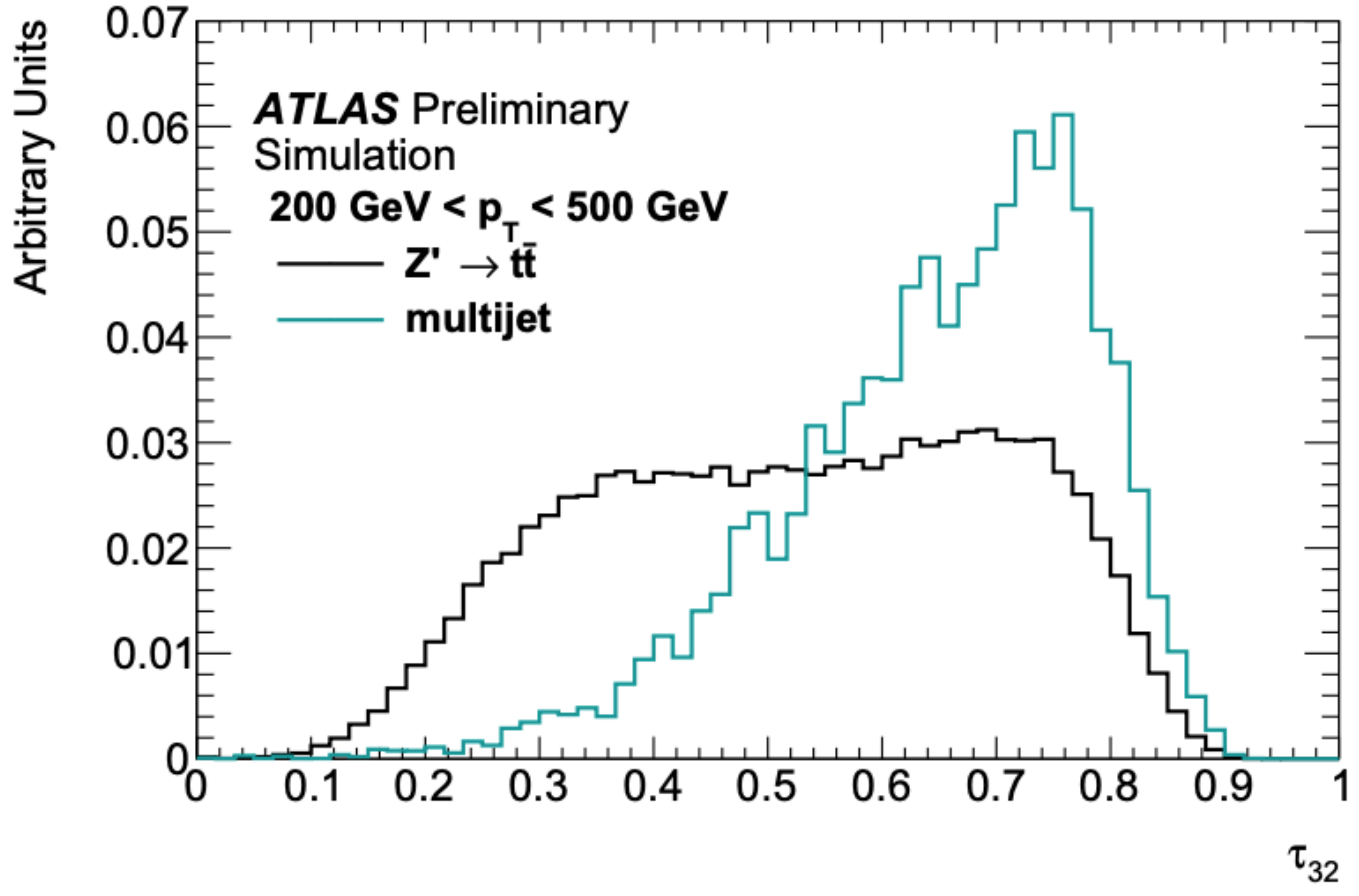
distance between sub-jet i and constituent k

$$d_0 = \sum_k p_{Tk} \times R$$

R = radius parameter of the jet

- For a jet with N prongs, one expects **$\tau_1, \ldots, \tau_{N-1}$ to be large and $\tau_{\geq N}$ to be small**;
- value of τ_N will also be larger when the prongs are gluons;
- **N-subjettiness ratio** is good discriminating variable for N-prong signal jets against the QCD background.

$$\tau_{N,N-1} = \frac{\tau_N}{\tau_{N-1}}$$



Examples:

- cut $\tau_{21} < \tau_{cut}$ to discriminate W/Z/H jets against QCD jets
- $\tau_{32} < \tau_{cut}$ to discriminate top jets against QCD jets

Large-R jets: substructure

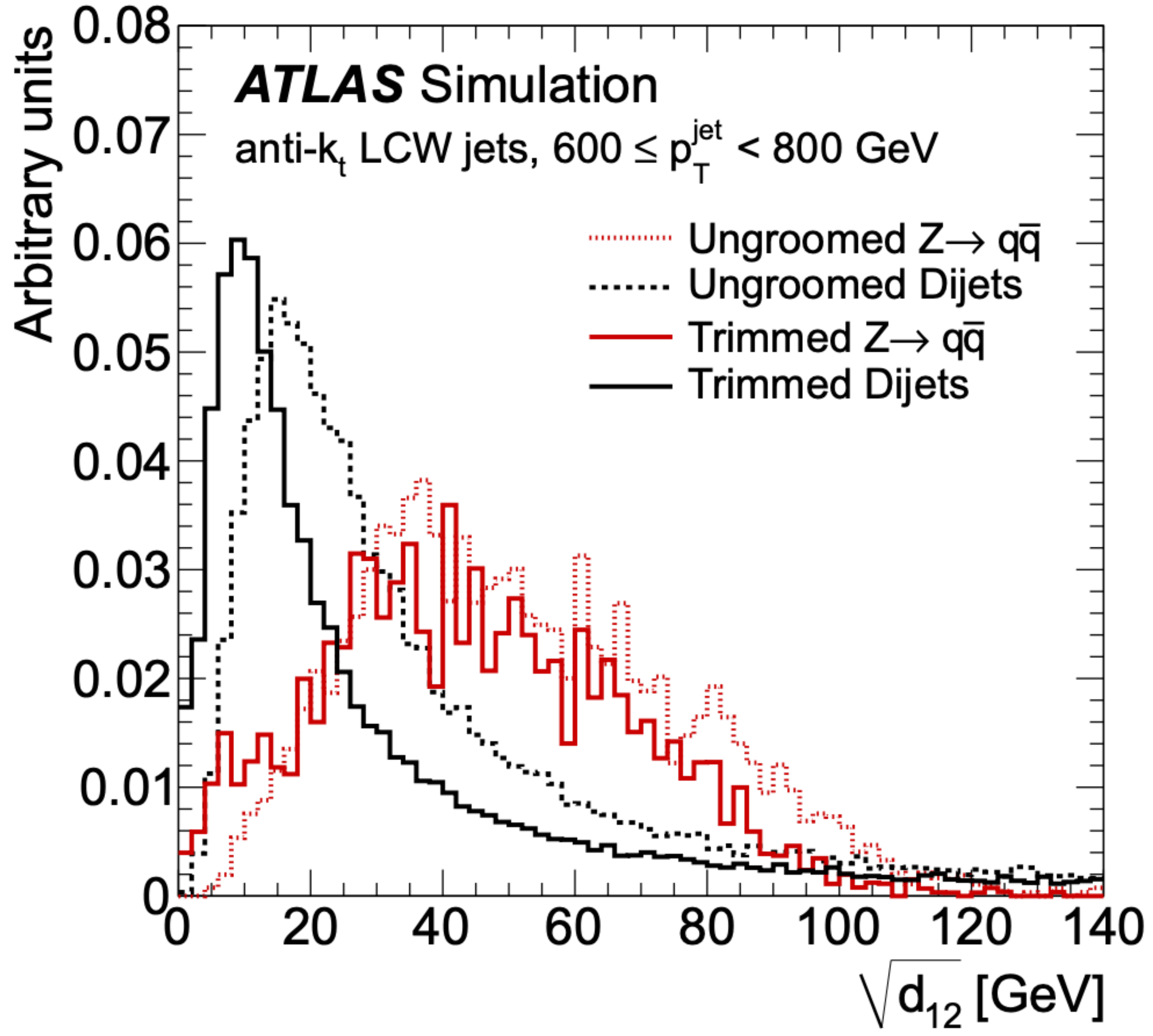
Looking inside jets

k_t splitting scales

- defined by **re-clustering the constituents of a jet with the k_t recombination algorithm**, which tends to combine the harder constituents last;
- at the final step of the jet recombination procedure, the k_t distance measure, **d_{ij}**, for the two remaining sub-jets can be used to define a splitting scale variable:

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

- the **sub-jets identified at the last step** of the re-clustering in the k_t algorithm provide the **$\sqrt{d_{12}}$** observable;
- **$\sqrt{d_{23}}$** characterises the **splitting scale in the second-to-last step** of the re-clustering;
- used to distinguish heavy-particle decays, which tend to be reasonably symmetric, from the largely asymmetric splittings that originate from QCD radiation in light-quark or gluon jets.



Examples:

- expected value for a two-body heavy-particle decay is approximately $\sqrt{d_{12}} \approx m_{particle}/2$
- jets from the parton shower of gluons and light quarks tend to have smaller values of the splitting scales and to exhibit a steeply falling spectrum for both $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$

Large-R jets: substructure

Looking inside jets
Jet substructure performances

Energy Correlation Functions (ECFs)

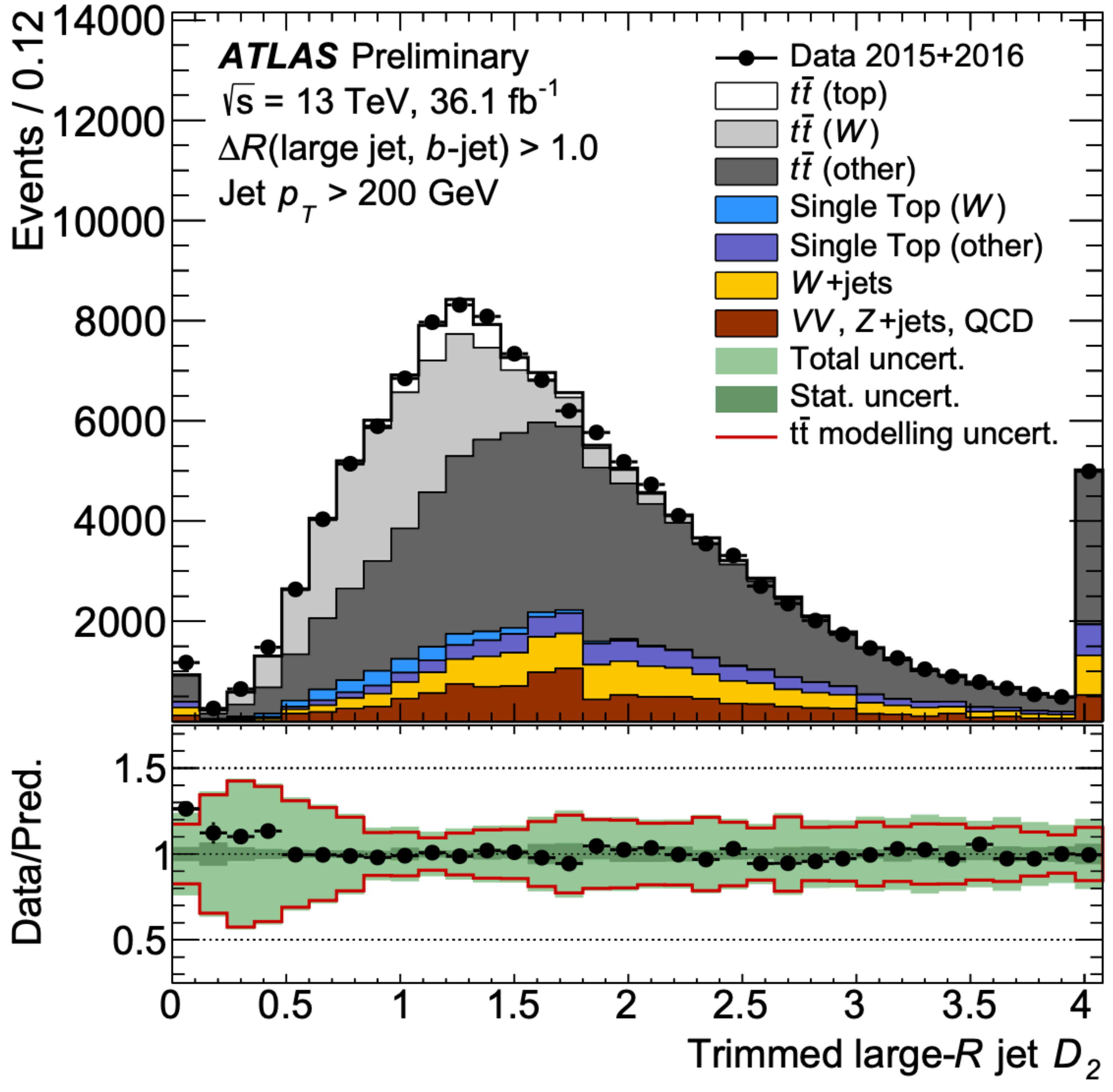
- achieve essentially the same objective than N-subjettiness without requiring the selection of N reference sub-jets:

$$e_2^{(\beta)} = \sum_{i<j\in \text{jet}} z_i z_j \, \Delta R_{ij}^{(\beta)} \;, \quad e_3^{(\beta)} = \sum_{i<j<k\in \text{jet}} z_i z_j z_k \, \Delta R_{ij}^{(\beta)} \, \Delta R_{jk}^{(\beta)} \, \Delta R_{ik}^{(\beta)} \quad \ldots \quad e_N^{(\beta)} = \sum_{i_1<\ldots<i_N\in \text{jet}} \Big(\prod_{j=1}^{N} z_i\Big)\Big(\prod_{k<l=1}^{N} \Delta R_{i_k j_l}^{(\beta)}\Big)$$

$$z_i = p_{t,i} / \sum_j p_{t,j}$$

- Compared to N-subjettiness, energy-correlation functions have the advantage of **not requiring a potentially delicate choice of reference axes**;
- Similarly to N-subjettiness, in order **to discriminate boosted massive particles from background QCD jets**, we again introduce ratios of ECFs:

$$\mathbf{C}_2^{(\beta)} = \frac{\mathbf{e}_3^{(\beta)}}{(\mathbf{e}_2^{(\beta)})^2} \qquad \mathbf{D}_2^{(\beta)} = \frac{\mathbf{e}_3^{(\beta)}}{(\mathbf{e}_2^{(\beta)})^3} \qquad \beta = 1$$



Examples:

- cut $C_2 < C_{cut}$ or $D_2 < D_{cut}$ to discriminate W/Z/H jets against QCD jets

Large-R jets: reconstruction and grooming

Reconstruction algorithms

both small-R jets and large-R jets

Jet are reconstructed with an iterative algorithm which combines calo deposits inside a given radius R = 1.0 (R = 0.4 for small-R jets).

Large-R jets: reconstruction and grooming

Reconstruction algorithms

both small-R jets and large-R jets

Jet are reconstructed with an iterative algorithm which combines calo deposits inside a given radius R = 1.0 (R = 0.4 for small-R jets).

only large-R jets

Jets are then cleaned, with "**grooming**" algorithms, from contamination due to the high particles concentration.

Large-R jets: reconstruction and grooming

Looking inside jets
Jet substructure performances

both small-R jets and large-R jets

Reconstruction algorithms

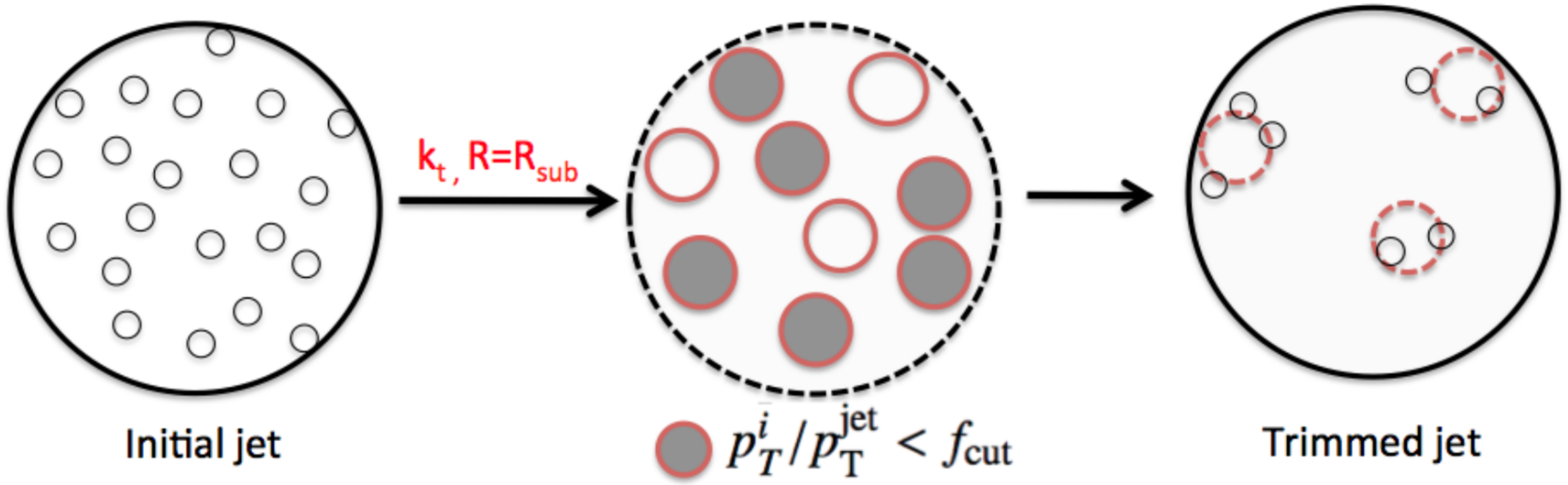
Jet are reconstructed with an iterative algorithm which combines calo deposits inside a given radius R = 1.0 (R = 0.4 for small-R jets).

Jets are then cleaned, with "**grooming**" algorithms, from contamination due to the high particles concentration.

only large-R jets

Trimming algorithm

Jet constituents are reconstructed again into jets with smaller radius $\mathbf{R_{sub}}$ (subjet). Subjets with lower p_T than a fraction $\mathbf{f_{cut}}$ of initial jet p_T are dropped off. The final jet is reconstructed using only the remaining subjets.



$\mathbf{R_{sub} = 0.2}$
$\mathbf{f_{cut} = 0.05}$

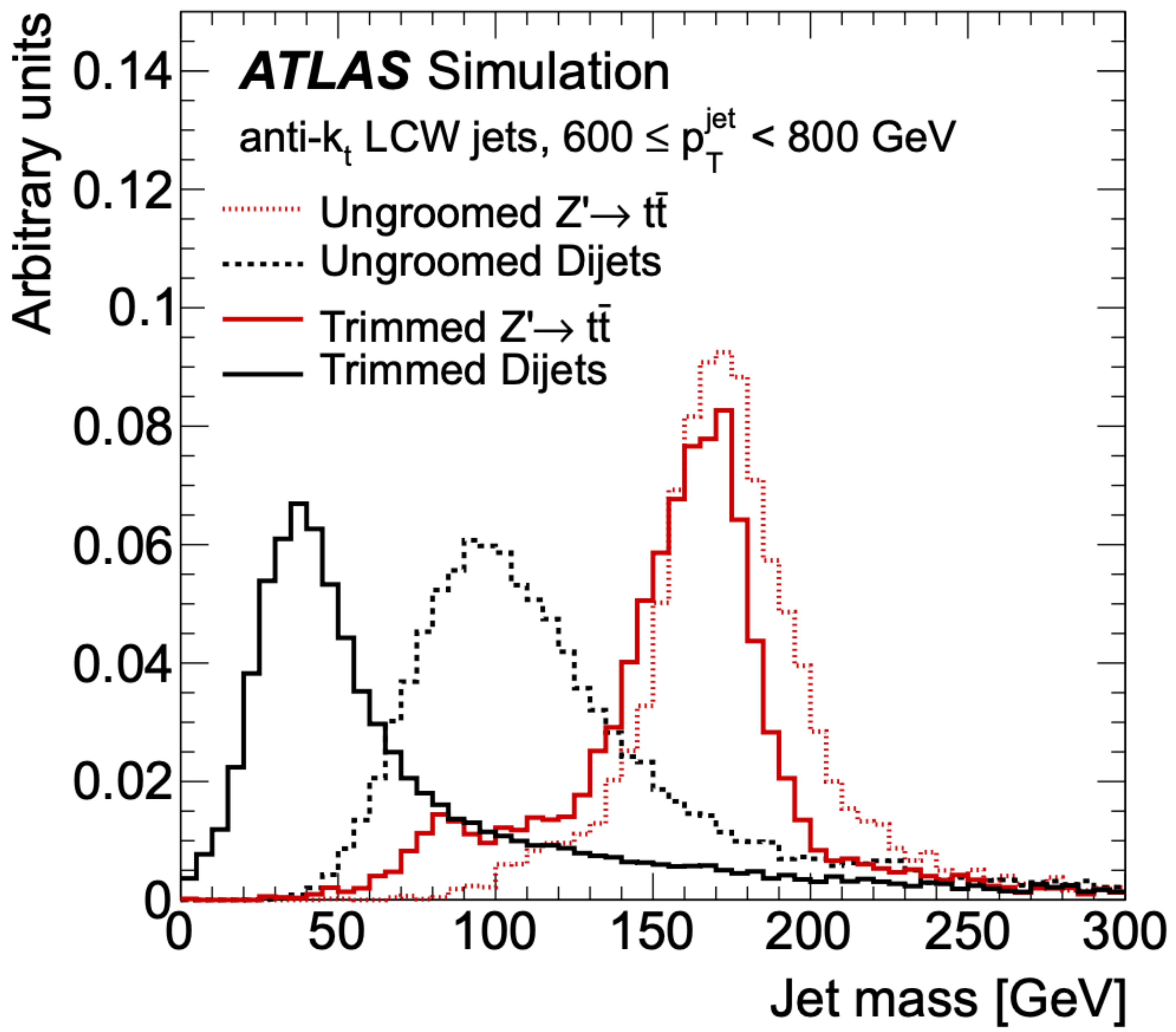
Large-R jets: reconstruction and grooming

Application of trimming algorithms

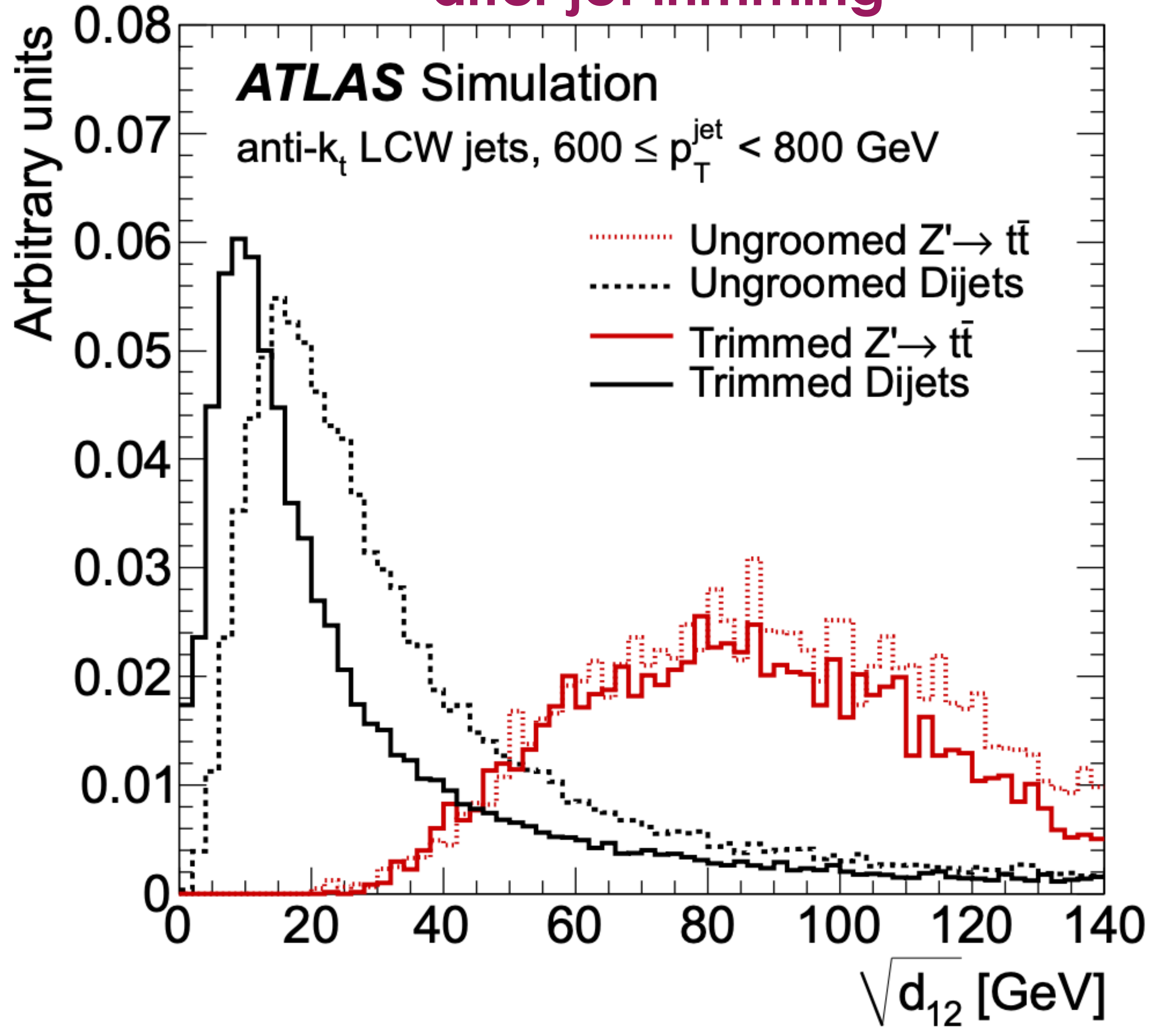
- example of Z'→tt and dijets events;
- anti-k_t reconstructed jets with $600 < p_T < 800$ GeV.

Grooming has a significant effect on substructure!

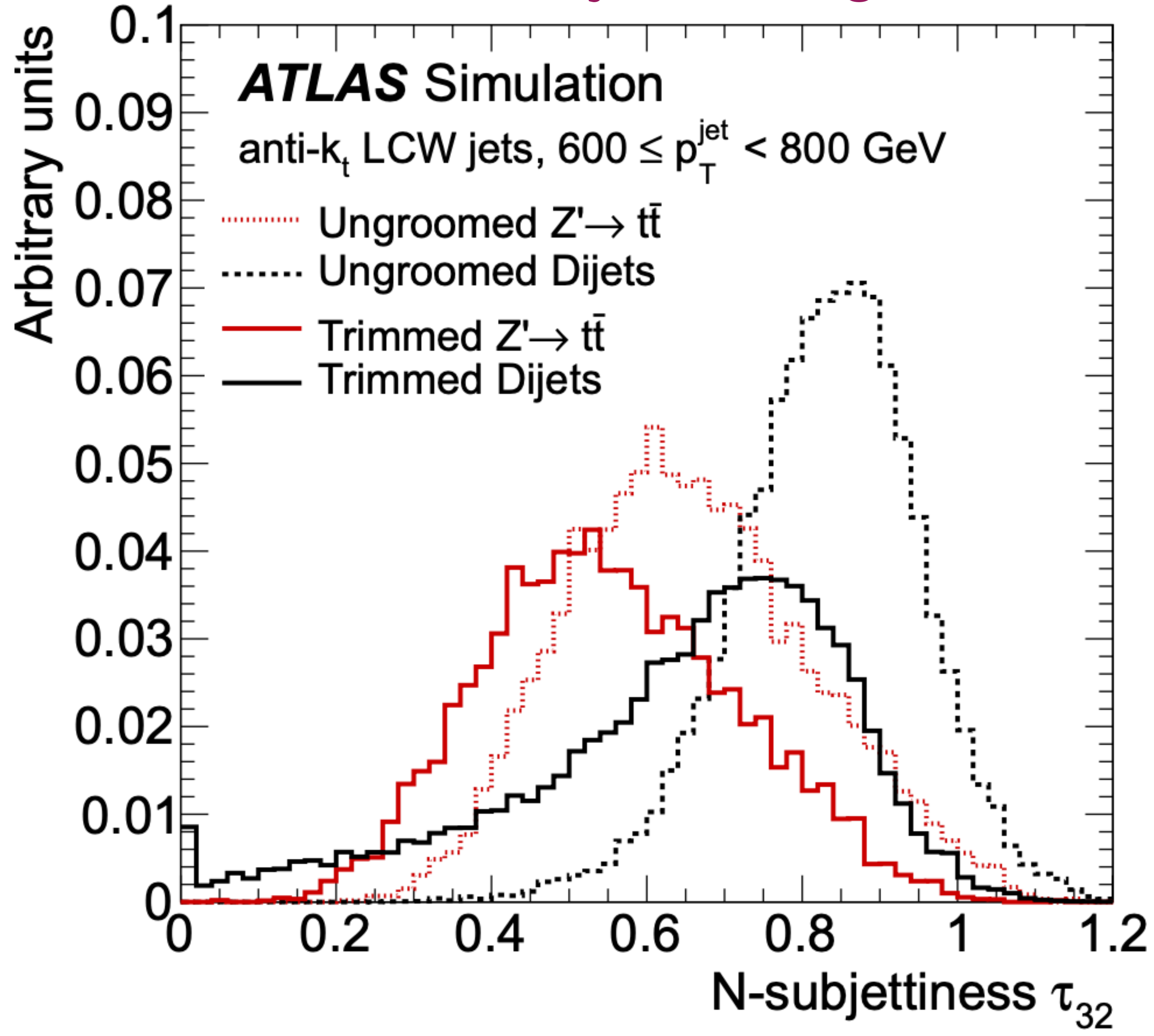
mass resolution for the signal improves after grooming



signal discrimination with the splitting scales is enhanced after jet trimming



trimming of anti-k_t jets results in similar discrimination between signal and background, but useful to reduce the jet mis-tag rate



Large-R jets: tagging technique

Combining all the info

- After reconstructing the jet as a collection of constituents, a number of methods can be used **to classify a jet as originating from a heavy particle (W boson or top quark) decay** as opposed to a light jet originating from gluons and quarks of all flavours other than top quarks;
- substructure observables do not commute and therefore, when considering combinations tools, the order in which we apply the different algorithms does matter;
- therefore important that the description of the tagging strategy clearly specify all the details of the combination including for example what jet, groomed or ungroomed, is used to compute jet shapes.
- Tagging techniques allow us to **exploit all the substructure characteristic of large-R jets**.

Example:

two-prong tagger

ATLAS

Strategy:

- trimming for **anti-k_t (R=1.0)** jets:
  - trimming radius $\mathbf{R_{sub} = 0.2}$;
  - energy cut $\mathbf{f_{cut} = 0.05}$;
- requirement on the **mass window**, depending on particle we are looking for;
- computation of $\mathbf{D_2}$ on the trimmed jet and impose a cut on this variable.

CMS

Strategy:

- mass-drop for **anti-k_t (R=0.8)** jets:
  - $\mathbf{z_{cut} = 0.1}$, keeping the hard structure and excluding soft emissions, starting from large angles;
- requirement on the **mass window**, depending on particle we are looking for;
- computation of $\boldsymbol{\tau_{21}}$ on the jet and impose a cut on this variable.
- Recently: $\mathbf{N_2}$ cut computed using the groomed jets.

Large-R jets: tagging technique with jet moments

EPJ C (2019) 79-375

Primary goal: provide a simple set of selections on jet moments that yield a constant signal efficiency as a function of the transverse momentum of the jet across a broad p_T range.

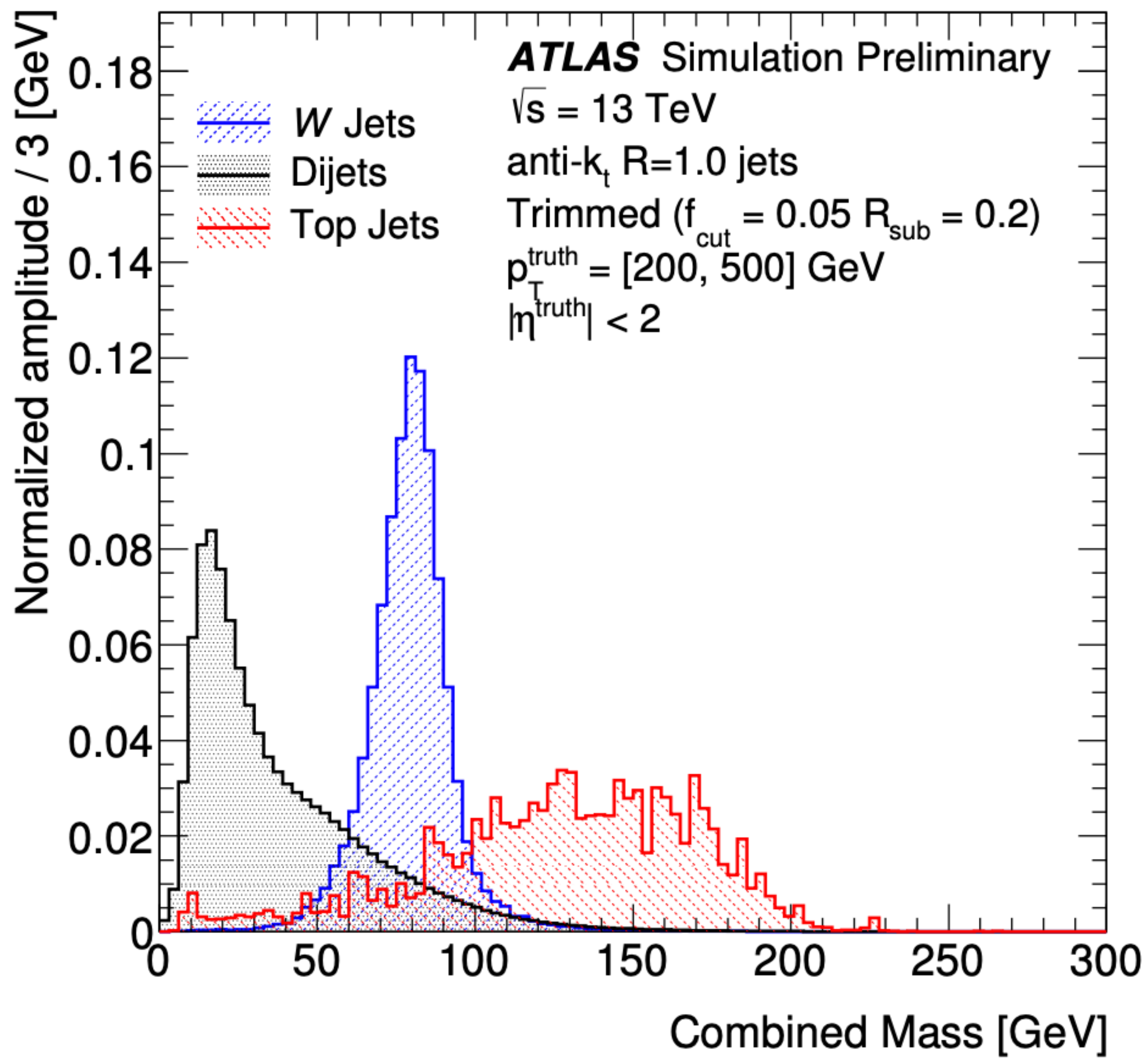
Based on the constituents of the trimmed jet and attempt to quantify a particular feature of the jet in an analytic way

- jet mass is the most powerful:
  - to achieve good performance across a broad range of jet transverse momenta;
  - **m_{calo}**: calculated as the invariant mass of the collection of topo-clusters of the trimmed jet;
  - **m_{TA}**: calculated as the invariant mass of the ghost-associated charged particle tracks scaled by the ratio of the transverse momenta of the trimmed jet and the associated tracks;
  - **combined mass (m_{comb})**: average of m_{calo} and m_{TA}, weighted by the inverse of their resolutions:

$$\mathbf{m}^{\mathrm{TA}} = \mathbf{m}^{\mathrm{track}} \times \frac{\mathbf{p}_{\mathrm{T}}^{\mathrm{calo}}}{\mathbf{p}_{\mathrm{T}}^{\mathrm{track}}}$$

$$\mathbf{w}^{\mathrm{TA}} = \frac{\sigma_{\mathrm{TA}}^{-2}}{\sigma_{\mathrm{calo}}^{-2} + \sigma_{\mathrm{TA}}^{-2}}$$

$$\mathbf{w}^{\mathrm{calo}} = 1 - \mathbf{w}^{\mathrm{TA}}$$



Large-R jets: tagging technique with jet moments

EPJ C (2019) 79-375

Primary goal: provide a simple set of selections on jet moments that yield a constant signal efficiency as a function of the transverse momentum of the jet across a broad p_T range.

Based on the constituents of the trimmed jet and attempt to quantify a particular feature of the jet in an analytic way

- jet mass is the most powerful:
  - to achieve good performance across a broad range of jet transverse momenta;
  - **m_{calo}**: calculated as the invariant mass of the collection of topo-clusters of the trimmed jet;
  - **m_{TA}**: calculated as the invariant mass of the ghost-associated charged particle tracks scaled by the ratio of the transverse momenta of the trimmed jet and the associated tracks;
  - **combined mass (m_{comb})**: average of m_{calo} and m_{TA}, weighted by the inverse of their resolutions:
- number of other observables quantify the extent to which the jet constituents are clustered or uniformly dispersed and can be used to **augment the discrimination power from the jet mass alone**
  - N-subjettiness, splitting scales, or using all jet constituents to quantify the dispersion of the jet constituents in an axis-independent way (ECFs).

Observable	Variable	Used for
Calibrated jet kinematics	p_T, m^{comb}	Top, W
Energy correlation ratios	e_3, C_2, D_2	Top, W
N-subjettiness	$\tau_1, \tau_2, \tau_{21}$	Top, W
	τ_3, τ_{32}	Top
Fox–Wolfram moment	R_2^{FW}	W
Splitting measures	z_{cut}	W
	$\sqrt{d_{12}}$	Top, W
	$\sqrt{d_{23}}$	Top
Planar flow	$\mathcal{P}$	W
Angularity	a_3	W
Aplanarity	A	W
KtDR	$KtDR$	W
Qw	Q_w	Top

Large-R jets: tagging technique with jet moments

EPJ C (2019) 79-375

Performances of a tagger

- each technique is explored and optimised;
- performances quantified in exclusive kinematic regimes based on p_T of the associated anti-k_t R=1.0 truth jet (p_T^{true}) to more closely resemble the kinematics of the parent particle;
- comparison between **signal efficiency** and **background rejection**:

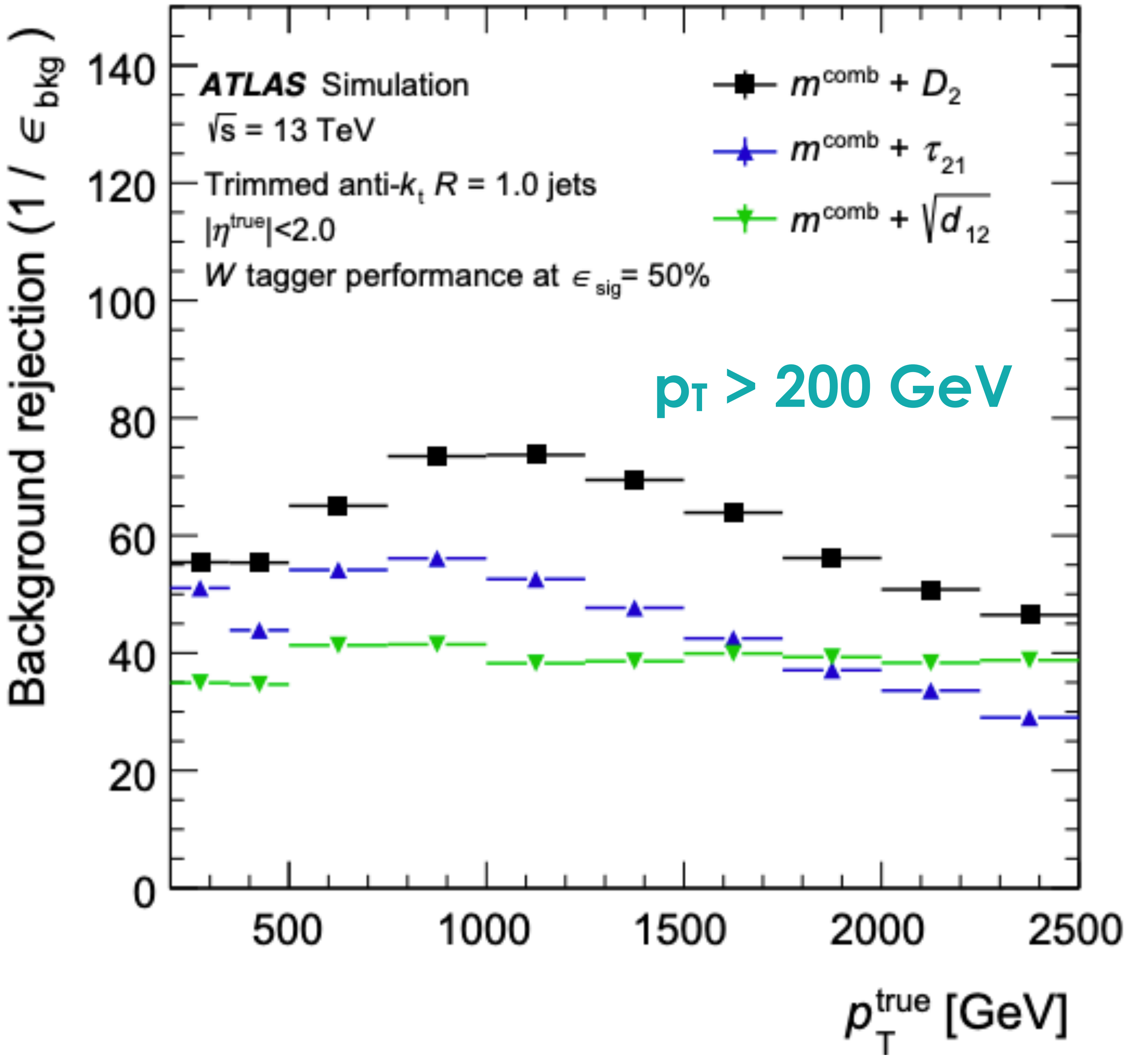
$$\epsilon_{signal} = \left(\frac{N_{tagged}}{N_{total}}\right)_{signal} \qquad r_{bkg} = \frac{1}{\epsilon_{bkg}} = \left(\frac{N_{total}}{N_{tagged}}\right)_{bkg}$$

Strategy:

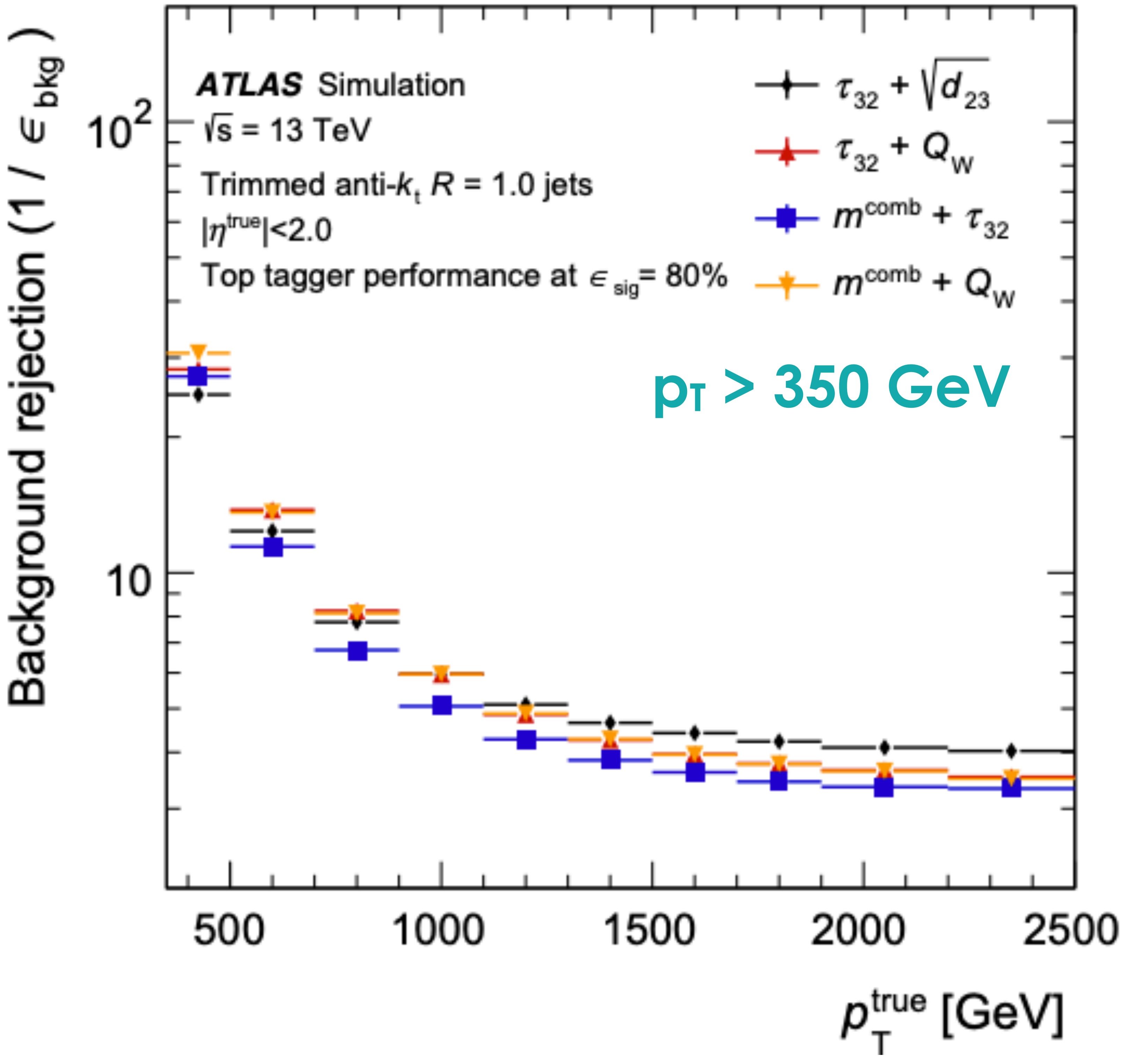
- reconstructed jets are matched to truth jets with a matching criterion of **$\Delta R(j_{true}, j_{reco}) < 0.75$**;
- only two reco jets matched to the two highest-p_T truth jets within $|\eta|<2$.

- For each pair of observables, the **selection criteria which give the chosen signal efficiency and the largest background rejection** are considered optimal and taken as the selection criteria in that region of jet p_T;
- this sequence of selection criteria is parameterised by a **smooth function dependent on the jet p_T**.

W boson



Top quark



Large-R jets: tagging technique with jet moments

EPJ C (2019) 79-375

Performances of a tagger

- each technique is explored and optimised;
- performances quantified in exclusive kinematic regimes based n the p_T of the associated anti-k_t R=1.0 truth jet (p_T^{true}) to more closely resemble the kinematics of the parent particle;
- comparison between **signal efficiency** and **background rejection**:

$$\epsilon_{\text{signal}} = \left(\frac{N_{\text{tagged}}}{N_{\text{total}}}\right)_{\text{signal}} \qquad r_{\text{bkg}} = \frac{1}{\epsilon_{\text{bkg}}} = \left(\frac{N_{\text{total}}}{N_{\text{tagged}}}\right)_{\text{bkg}}$$

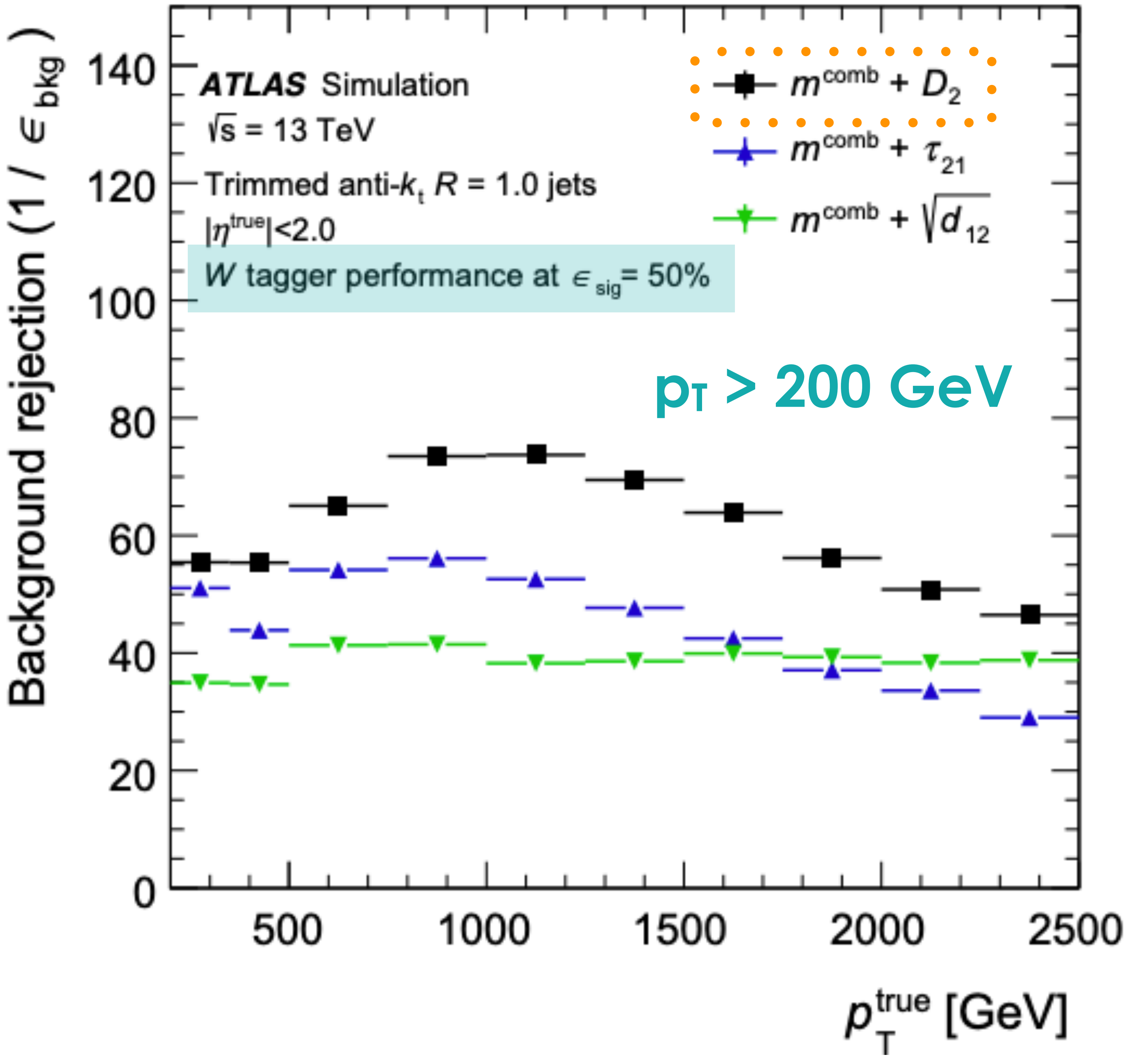
Results for W-tagging:

- **m_{comb} and D_2** is the most powerful in the kinematic range of interest;

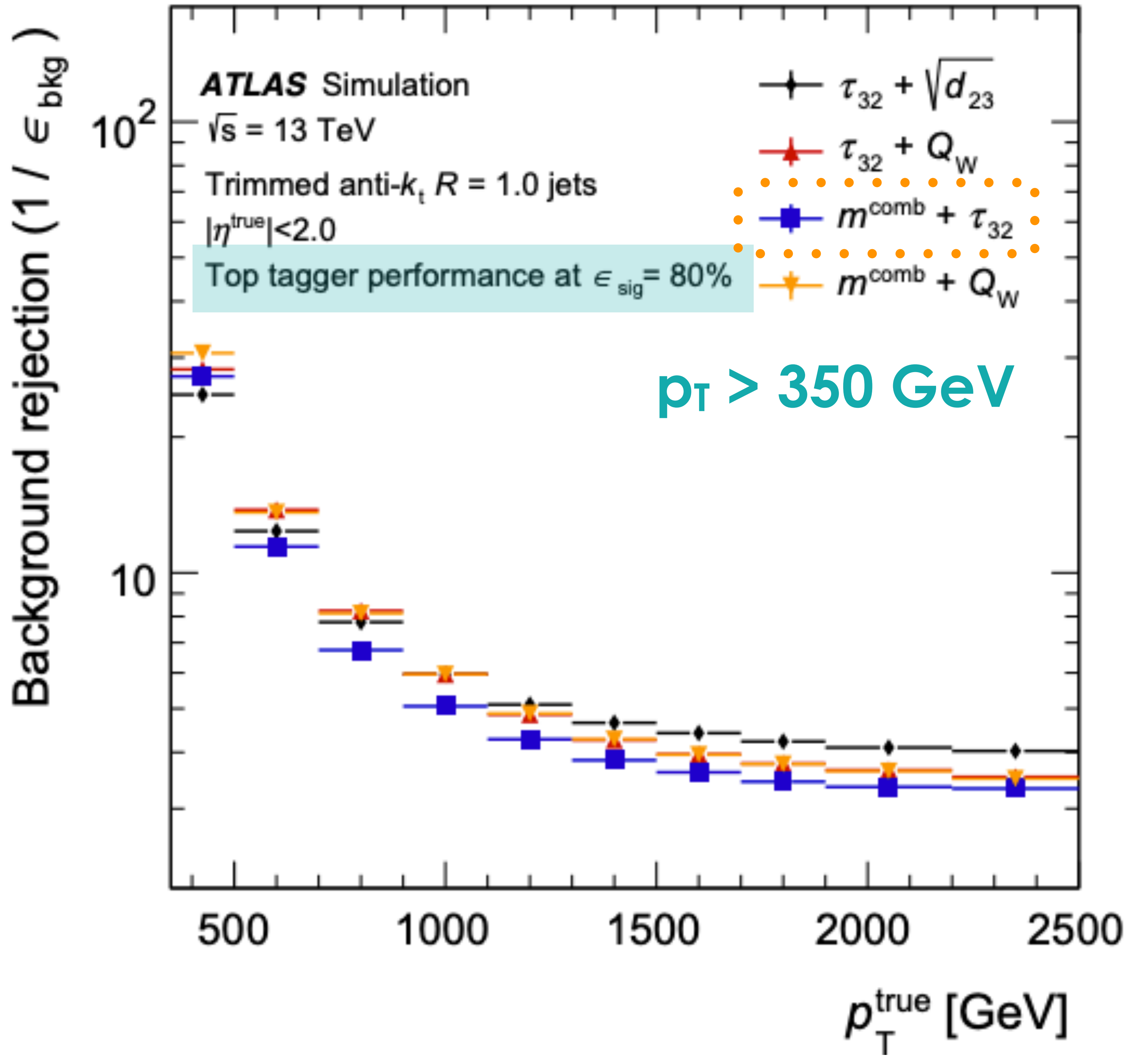
Results for top-tagging:

- plateauing at a lower value for high jet p_T^{true} mostly due to the migration of the light-jet mass distribution to higher values and a looser τ_{32} cut to maintain the constant signal efficiency;
- **m_{comb} and τ_{32}** chosen but specific top-quark jet tagger used in an analysis may depend on the context of the analysis and not on the performance alone.

W boson



Top quark



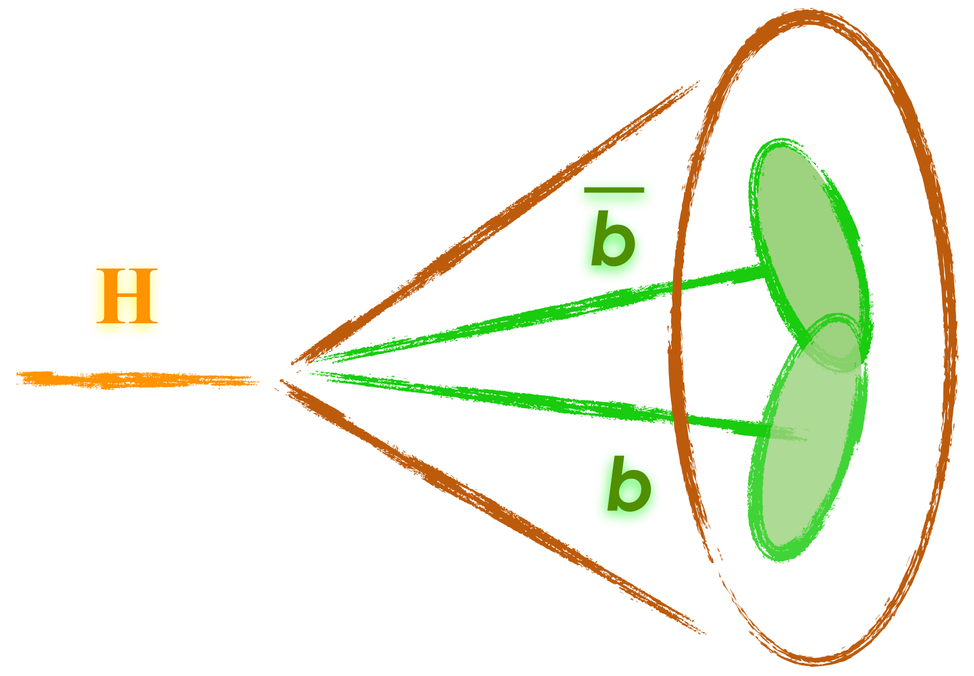
Large-R jets: tagging the Higgs boson

EPJ C (2019) 79-836

LHC centre-of-mass energy of 13 TeV greatly extends the sensitivity of the ATLAS experiment to heavy new particles which may have decay chains including a (boosted) Higgs boson

H→bb decay has the largest branching fraction within the SM, thus it is a major decay mode to use when searching for resonances involving high-momentum Higgs bosons

1. Higgs boson candidate is reconstructed as a large-R jet:
   ✓ $p_T > 250$ GeV, $|\eta| < 2.0$
2. b-tagging requirement is applied to track-jets associated with the large-R jet:
   ✓ $p_T > 5$ GeV, $|\eta| < 2.5$;
3. b-tagged large-R jet mass can be required to be around the SM Higgs boson mass of 125 GeV,
4. a requirement on other large-R jet substructure variables can be applied depending on the Higgs-jet tagger working point.



Large-R jets: tagging the Higgs boson

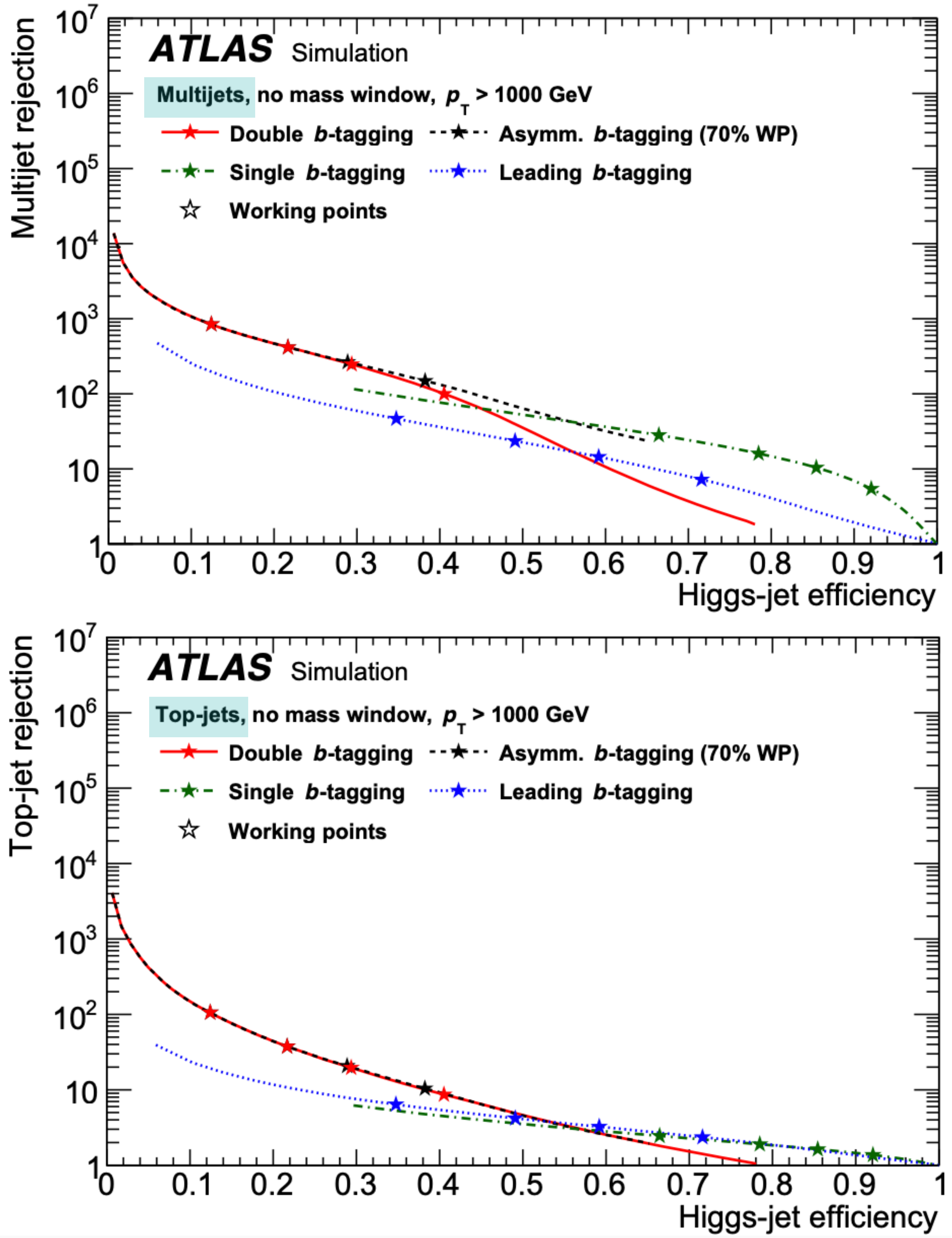
EPJ C (2019) 79-836

LHC centre-of-mass energy of 13 TeV greatly extends the sensitivity of the ATLAS experiment to heavy new particles which may have decay chains including a (boosted) Higgs boson

H→bb decay has the largest branching fraction within the SM, thus it is a major decay mode to use when searching for resonances involving high-momentum Higgs bosons

1. Higgs boson candidate is reconstructed as a large-R jet:
   ✓ p_T > 250 GeV, |η| < 2.0
2. **b-tagging requirement** is applied to track-jets associated with the large-R jet:
   ✓ p_T > 5 GeV, |η| < 2.5;
3. b-tagged large-R jet mass can be required to be around the SM Higgs boson mass of 125 GeV,
4. a requirement on other large-R jet substructure variables can be applied depending on the Higgs-jet tagger working point.

b-tagging benchmarks
- double: the two highest-p_T track-jets pass the same requirement;
- asymmetric: the two highest-p_T track-jets pass different requirements (WPs);
- single: at least one of the two highest-p_T track-jets pass the requirement;
- leading single: the highest-p_T track-jet pass the requirement.



Large-R jets: tagging the Higgs boson

EPJ C (2019) 79-836

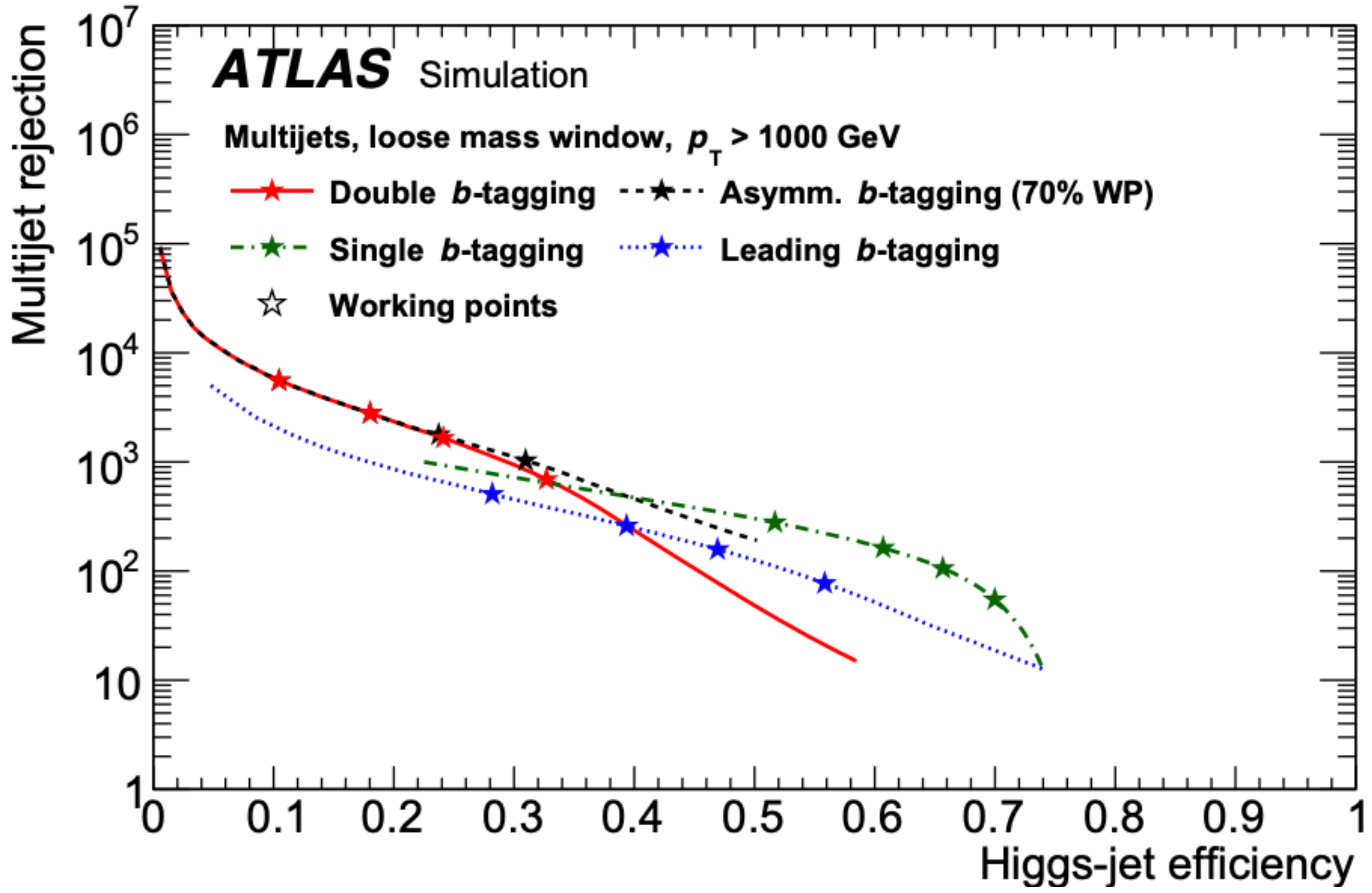
LHC centre-of-mass energy of 13 TeV greatly extends the sensitivity of the ATLAS experiment to heavy new particles which may have decay chains including a (boosted) Higgs boson

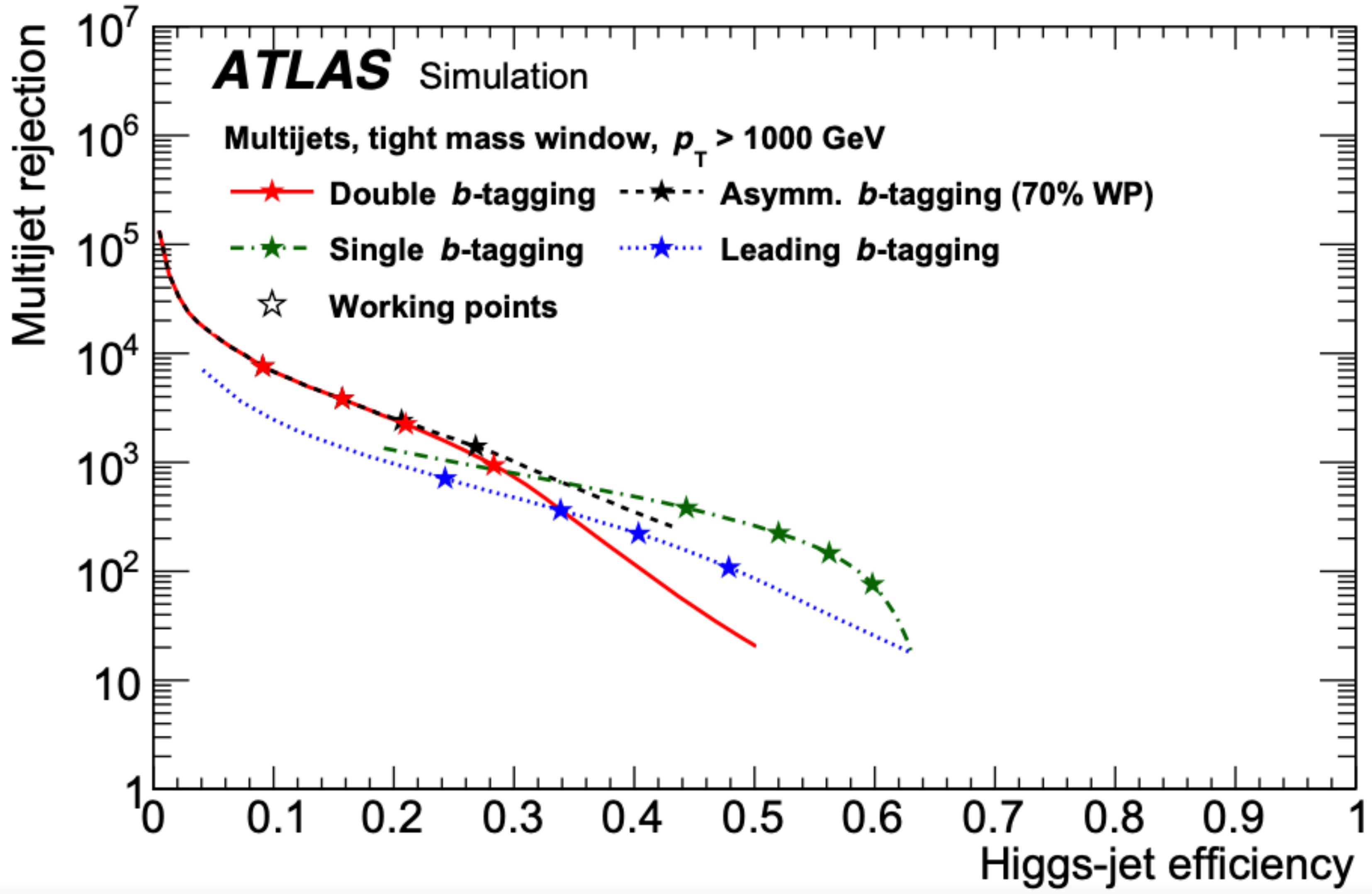
H→bb decay has the largest branching fraction within the SM, thus it is a major decay mode to use when searching for resonances involving high-momentum Higgs bosons

1. Higgs boson candidate is reconstructed as a large-R jet:
   ✓ p_T > 250 GeV, |η| < 2.0
2. b-tagging requirement is applied to track-jets associated with the large-R jet:
   ✓ p_T > 5 GeV, |η| < 2.5;
3. **b-tagged large-R jet mass** can be required to be around the SM Higgs boson mass of 125 GeV,
4. a requirement on other large-R jet substructure variables can be applied depending on the Higgs-jet tagger working point.

Mass window

- parametrised as a function of Higgs p_T;
- defined as the smallest window containing the given fraction of Higgs-jets:
  - tight mass window, containing 68% of Higgs-jet;
  - loose mass window, containing 80% of Higgs-jets;





Large-R jets: tagging technique with ML algorithms

Motivation

- recent simulation-based studies have found that the more direct use of the jet constituents as **inputs to a machine-learning algorithm** can lead to significant improvements in discriminating power as compared to more traditional, jet-moment-based discriminants.

Still missing some important concepts, we'll come back here after the MVA lecture!

Large-R jets: re-clustering technique

Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC

Motivation

- **radius parameter R of jet clustering algorithms** aimed at targeting the hadronic decays of a particle should be **process dependent and scale with the momentum** under consideration.

$$\Delta R_{decay} \simeq 2m/p_T$$

- Only a **few choices of R** are used for all analyses:
  - every jet configuration, which includes the algorithm, radius, and grooming parameters, must be calibrated to account for unmeasured energy deposits and other experimental effects

- **Jet calibration** includes also jet energy and mass scale corrections, which provide a full calibration by also correcting particles that were missed, merged, or below noise thresholds, energy loss in un-instrumented regions of the calorimeter, and additionally takes into account correlations between particles.

nice to have calibration also for "non-standard" R jets!

Large-R jets: re-clustering technique

Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC

Motivation

- **radius parameter R of jet clustering algorithms** aimed at targeting the hadronic decays of a particle should be **process dependent and scale with the momentum** under consideration.

$$\Delta R_{decay} \simeq 2m/p_T$$

- Only a **few choices of R** are used for all analyses:
  - every jet configuration, which includes the algorithm, radius, and grooming parameters, must be calibrated to account for unmeasured energy deposits and other experimental effects

- **Jet calibration** includes also jet energy and mass scale corrections, which provide a full calibration by also correcting particles that were missed, merged, or below noise thresholds, energy loss in un-instrumented regions of the calorimeter, and additionally takes into account correlations between particles.

Solution

introduction of a new angular scale $r < R$, such that jets of radius r can be the inputs to the clustering algorithm of large radius R jets

Advantages

- automatic calibration of the re-clustered large radius jets;
- with no additional calibration needed, any large radius R, any clustering algorithm, and many grooming strategies can be used;
- uncertainties on the re-clustered p_T and mass are also automatic consequences of propagating the corresponding uncertainties computed for small radius jets.

Large-R jets: re-clustering technique

[Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC]

Sequential recombination: algorithms recursively combine sub-jets until there are none left.

- d_{ij} = distance between particle 4-vectors;
- d_{iB} = distance between particle and beam 4-vectors;

- list of **sub-jets** is initialised by the **set of jet inputs**;
- at every level of recursion, combination of particles based on

$$\mathbf{d_{ij} = \min_{i',j'}(d_{i',j'}, d_{i',B})}$$

$\mathbf{j \neq B}$

sub-jets i and j are combined into a new sub-jet with a 4-vector that is the sum of the 4-vectors of i and j

$\mathbf{i = B}$

the sub-jet i is declared a jet and removed from the list of sub-jets

Large-R jets: re-clustering technique

[Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC]

Sequential recombination: algorithms recursively combine sub-jets until there are none left.

- d_{ij} = distance between particle 4-vectors;
- d_{iB} = distance between particle and beam 4-vectors;

- list of **sub-jets** is initialised by the **set of jet inputs**;
- at every level of recursion, combination of particles based on

$$\mathbf{d_{ij}} = \min_{i',j'}(\mathbf{d_{i',j'}}, \mathbf{d_{i',B}})$$

$$\mathbf{j \neq B}$$

sub-jets i and j are combined into a new sub-jet with a 4-vector that is the sum of the 4-vectors of i and j

$$\mathbf{i = B}$$

the sub-jet i is declared a jet and removed from the list of sub-jets

Most widely used metrics:

$$\mathbf{d_{ij}} = \min(\mathbf{p_{Ti}^{2n}}, \mathbf{p_{Tj}^{2n}})\ \mathbf{R_{ij}^2/R^2}$$

- n = 1 — k_t
- n = 0 — C/A
- n = -1 — anti-k_t

In all three of these algorithms, **R is roughly the size of the jet in (y, φ) space**, though C/A and k_t jets can have very irregular jet areas

anti-k_t is the most used one

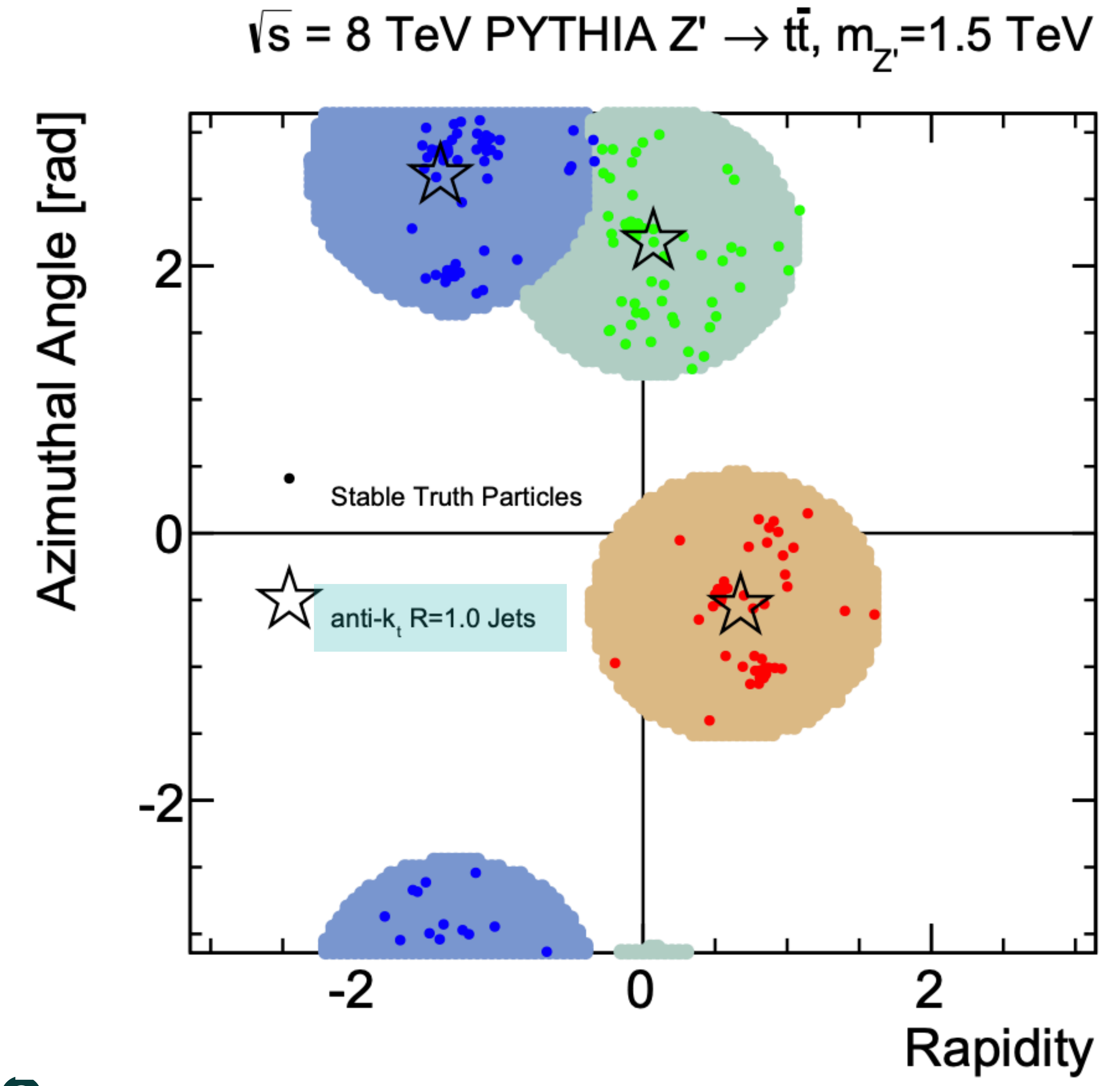
$$R_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}$$

Large-R jets: re-clustering technique

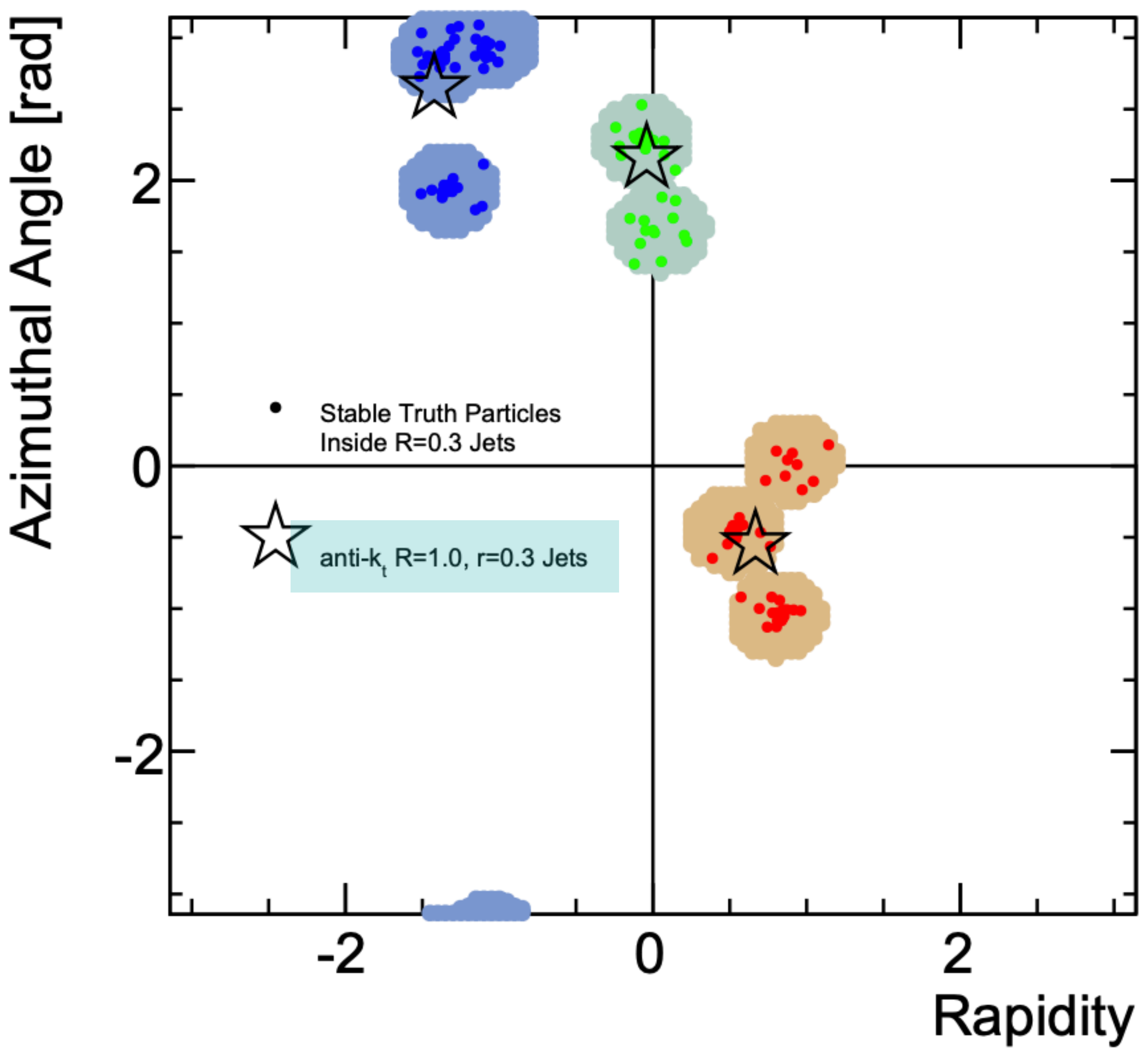
Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC

- Inputs of jet in a clustering algorithm are typically **stable particles** (Monte Carlo truth studies), **topological clusters** (ATLAS), or **particle flow objects** (CMS);
- Re-clustered large radius R jets take as input the **output of the small radius r jet clustering**.
- In general, the algorithm used to cluster the small radius jets can be different than the algorithm used for re-clustering the entire event.

standard large-R jet (R = 1.0)



re-clustered jet (R = 1.0, r = 0.3)



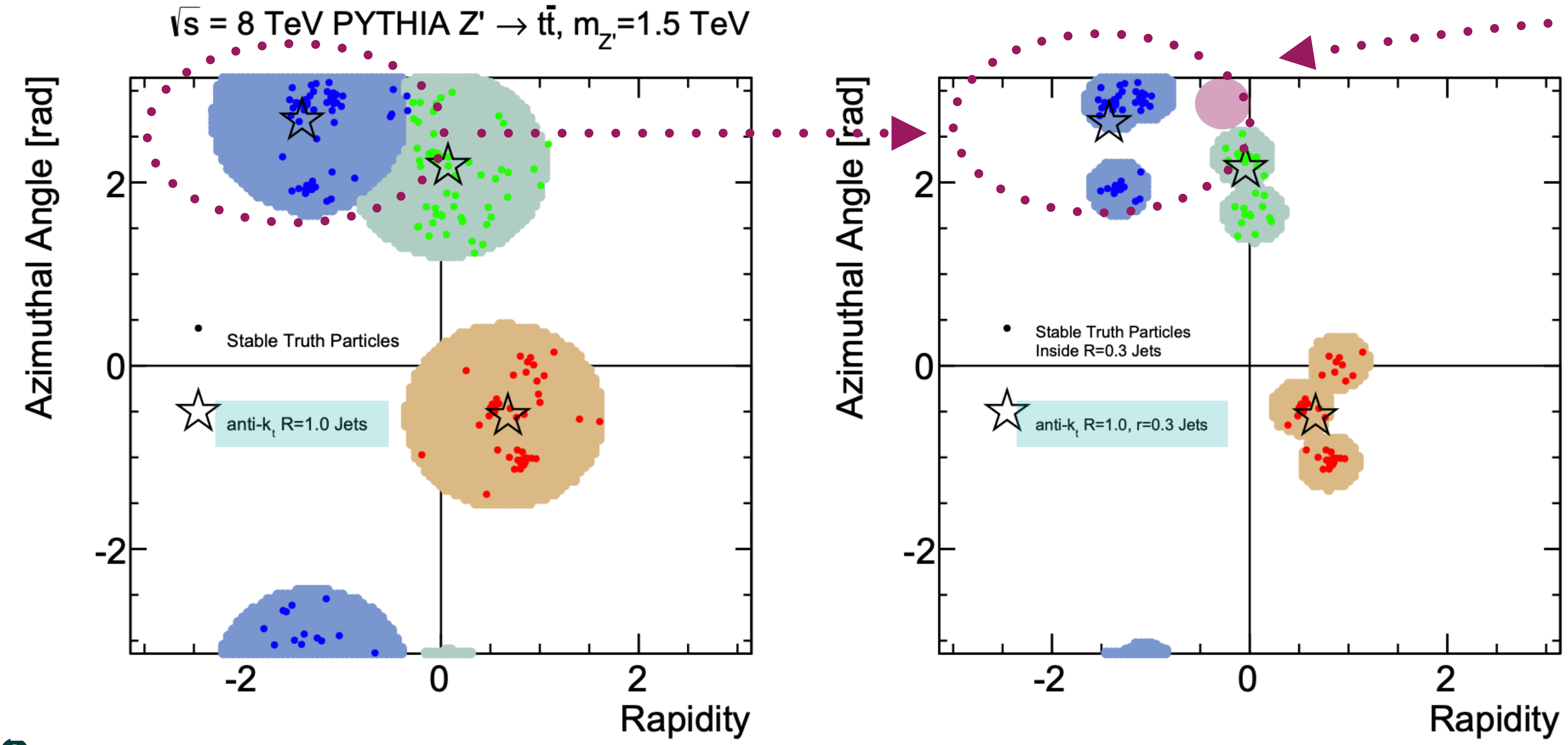
Large-R jets: re-clustering technique

Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC

- Inputs of jet in a clustering algorithm are typically **stable particles** (Monte Carlo truth studies), **topological clusters** (ATLAS), or **particle flow objects** (CMS);
- Re-clustered large radius R jets take as input the **output of the small radius r jet clustering**.
- In general, the algorithm used to cluster the small radius jets can be different than the algorithm used for re-clustering the entire event.

standard large-R jet (R = 1.0) **re-clustered jet (R = 1.0, r = 0.3)**

can be seen here!



- Small radius jets can only be **reliably fully calibrated for p_T >15 GeV**;
- this minimum p_T threshold acts as an **effective grooming** for the re-clustered jets;
- different grooming algorithm can be used with tighter threshold:
  - a **dedicated trimming (RT)** has been developed.

$$f_{cut} = \frac{p_T^{sub-jet}}{p_T^{large-R\,jet}} = 0.1$$

Large-R jets: re-clustering technique

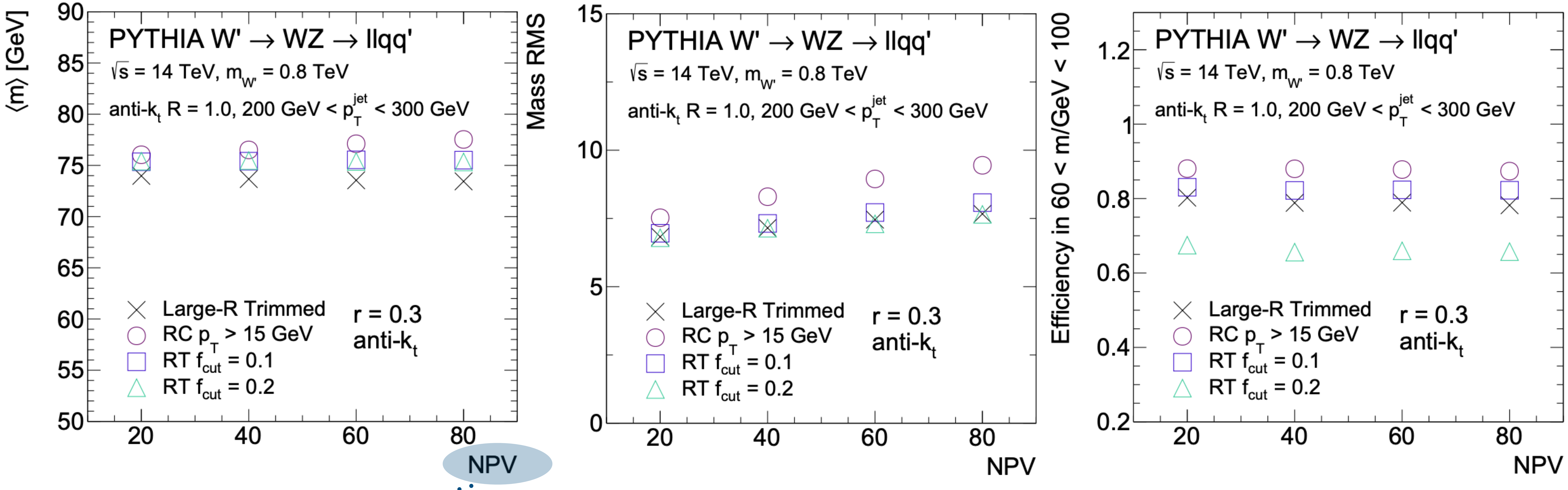
Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC

Performance of jet mass

- The most widely used large radius jet observable;
- to compare the performance of large radius jets with re-clustered jets, study of **performance of the jet mass for the various re-clustering schemes**;
- averages and deviations computed over a fixed mass range: **60-100 GeV**.

$$m_{RCjet} = \left(\sum_{i\in jet} E_i\right)^2 - \left(\sum_{i\in jet} \vec{p}_i\right)^2$$

- Jet mass performance is quantified by the average jet mass $\langle m \rangle$, a mass resolution, **σ**, and the dependance of these quantities with the amount of **pileup**.



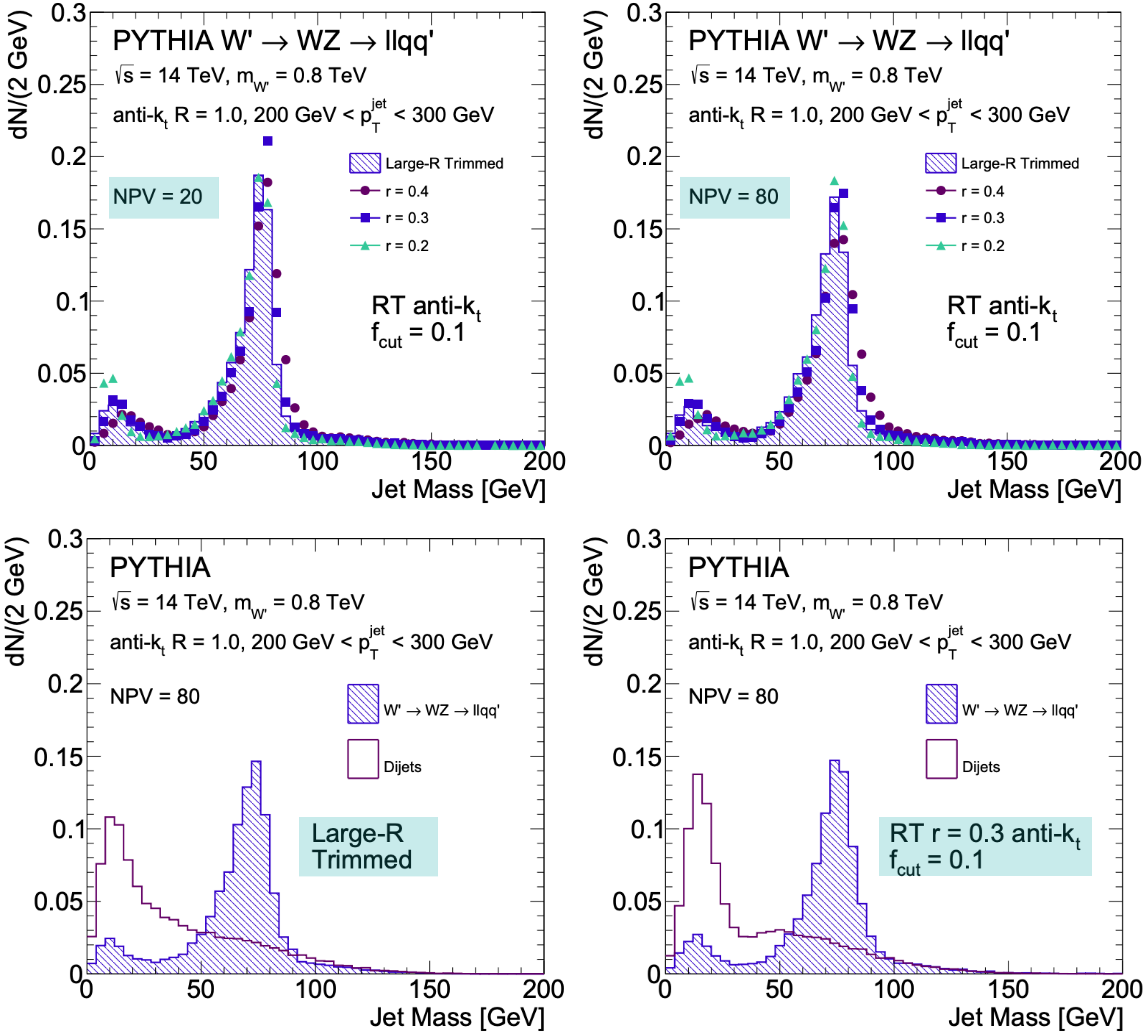
number of additional interaction vertices

- $\langle m \rangle$ for RT is very stable, whereas there is a slight slope for RC;
- **σ** for RC is slightly worse than for RT;
- **efficiency** of RC is better because it avoids the peak at low masses well below the W boson mass.

for low pileup levels, essentially no differences between standard and RC jets.

Large-R jets: re-clustering technique

[Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC]



- mass distribution for several small radius jet sizes;
- same RT applied;
- 2 different NPV;
- minimum p_T cut is 15 GeV;
- r = 0.2 is the most peaked, both in W-peak region and in low mass region.

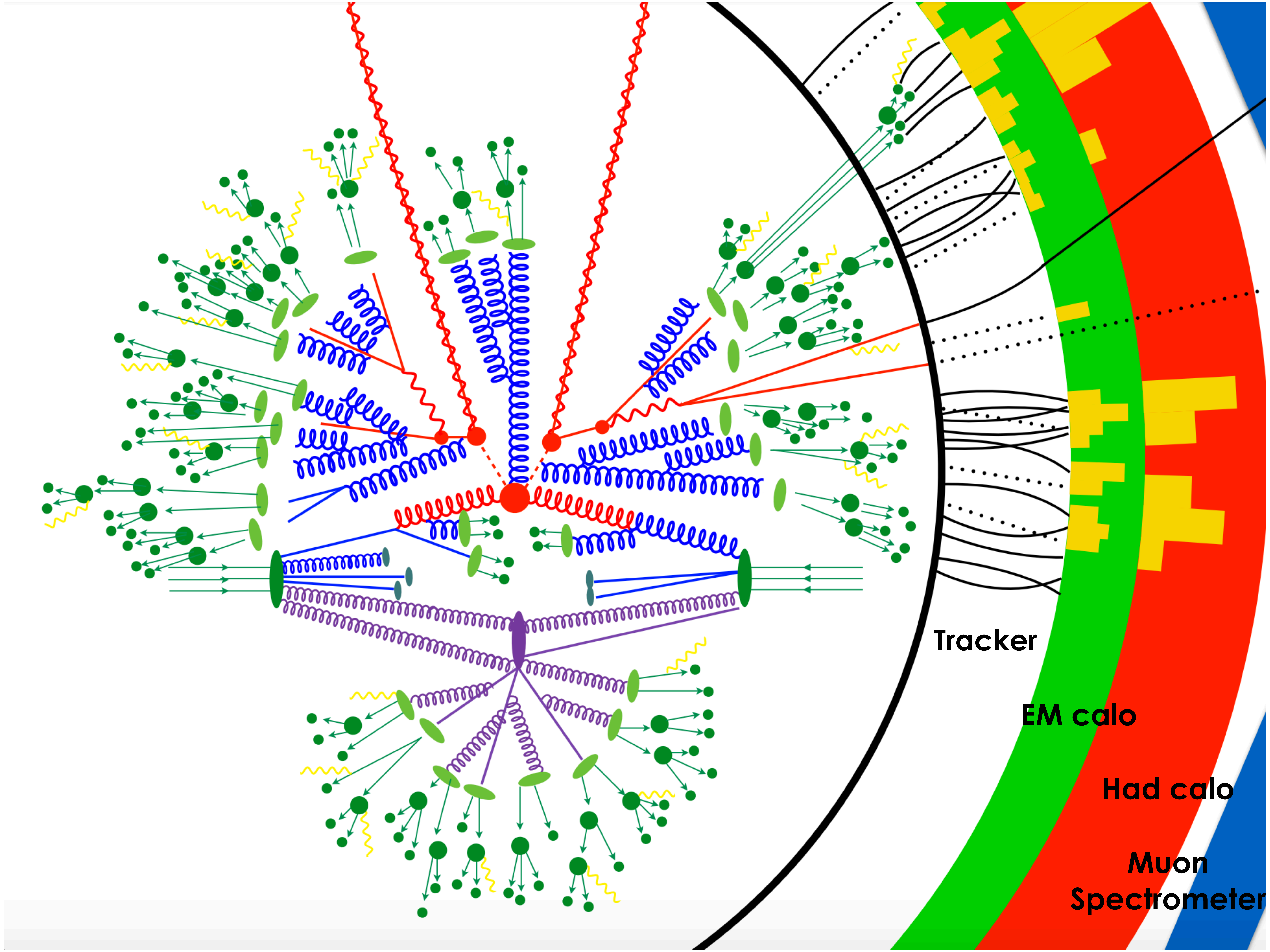
best combination* of parameters
r = 0.3 for radius
f_{cut} = 0.1 for RT

* but also depends on specific analyses or studies

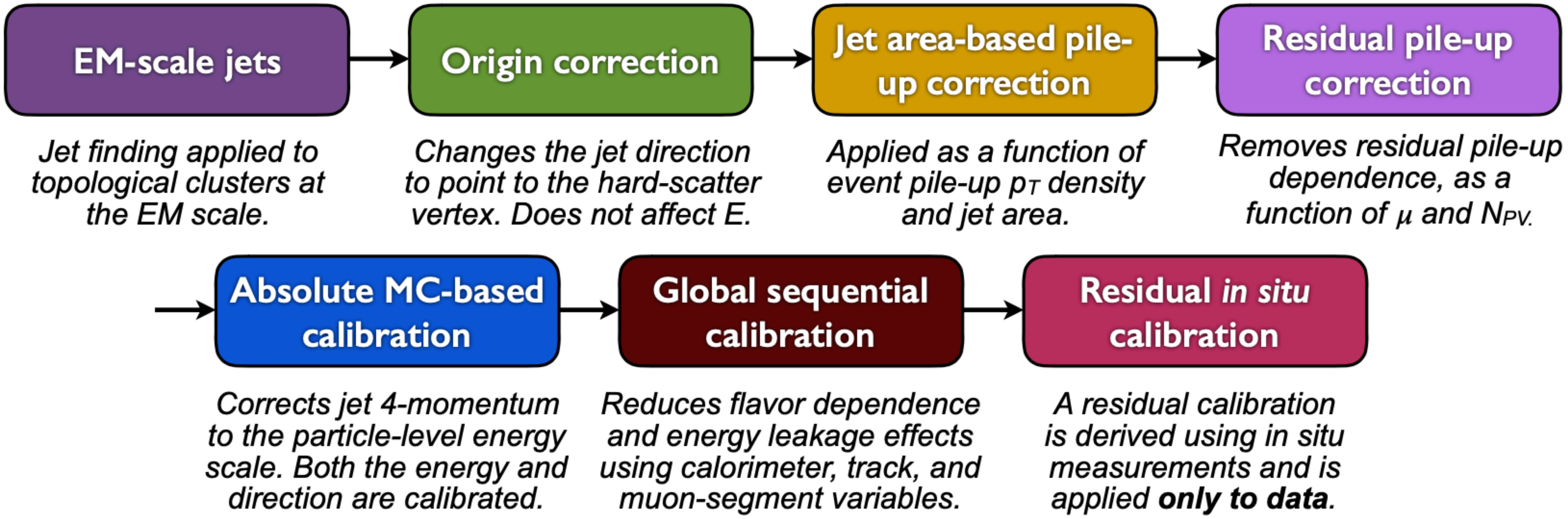
- important to study the **effect of re-clustering also on the most likely background-jets** originating from QCD multi-jet processes;
- various jet grooming techniques increase the separation in jet mass between signal and QCD jets;
- **both trimming and re-clustering allow for the successful discrimination of signal and background using the jet mass**
  - details of the optimisation to improve S/B left to the experimental analyses

Supporting material

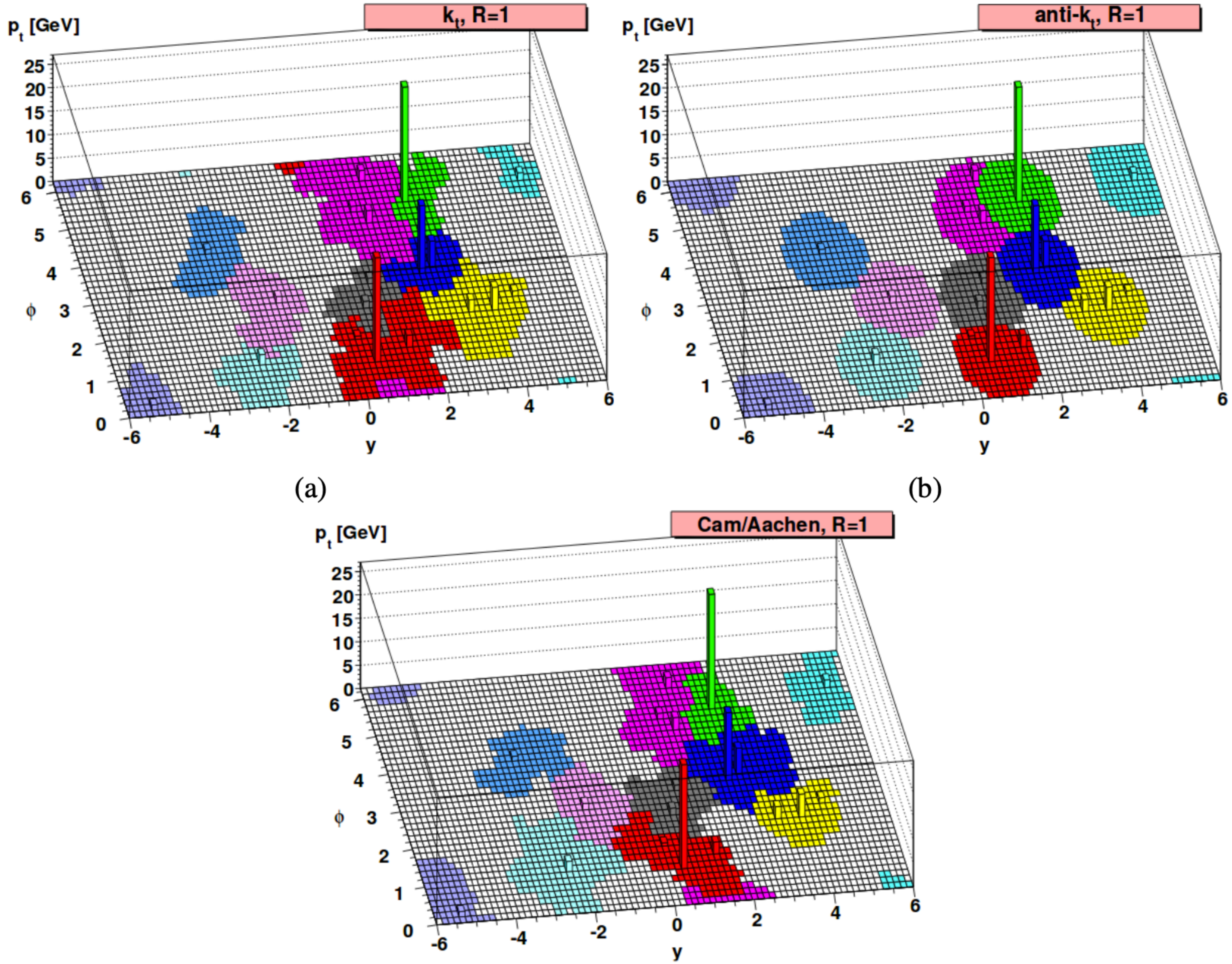
Reconstruction



Large-R jets: calibration in ATLAS



Large-R jets: reconstruction and grooming



Anti-k_t and k_t algorithms

- The iterative recombination procedure works by first cleaning a list of all objects (either hadrons, topo-clusters or tracks) in an event.
- The ordering of the list is irrelevant and proto-jets are built from these objects.
- Two distance measures in y-φ-space are associated to each member of the list, between the proto-jet and its closest neighbour:

$$\rho_{ij} = \min\left(p_{Ti}^{2p}, p_{Tj}^{2p}\right)\frac{(\Delta R_{ij})^2}{R^2}$$

measure of the opening angle between the two constituents

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$$

- and between the proto-jet and the beam:

$$\rho_{iB} = p_{Ti}^{2p}$$

- If $\rho_{iB} < \rho_{ij}$: the proto-jet is closer to the beam than it is to any other proto-jet in the event, so it is defined as a jet and removed from the list.
- If $\rho_{iB} > \rho_{ij}$: the two proto-jets i and j are combined into one, thereby forming a new proto-jet.
- This procedure continues through all proto-jets in the event.

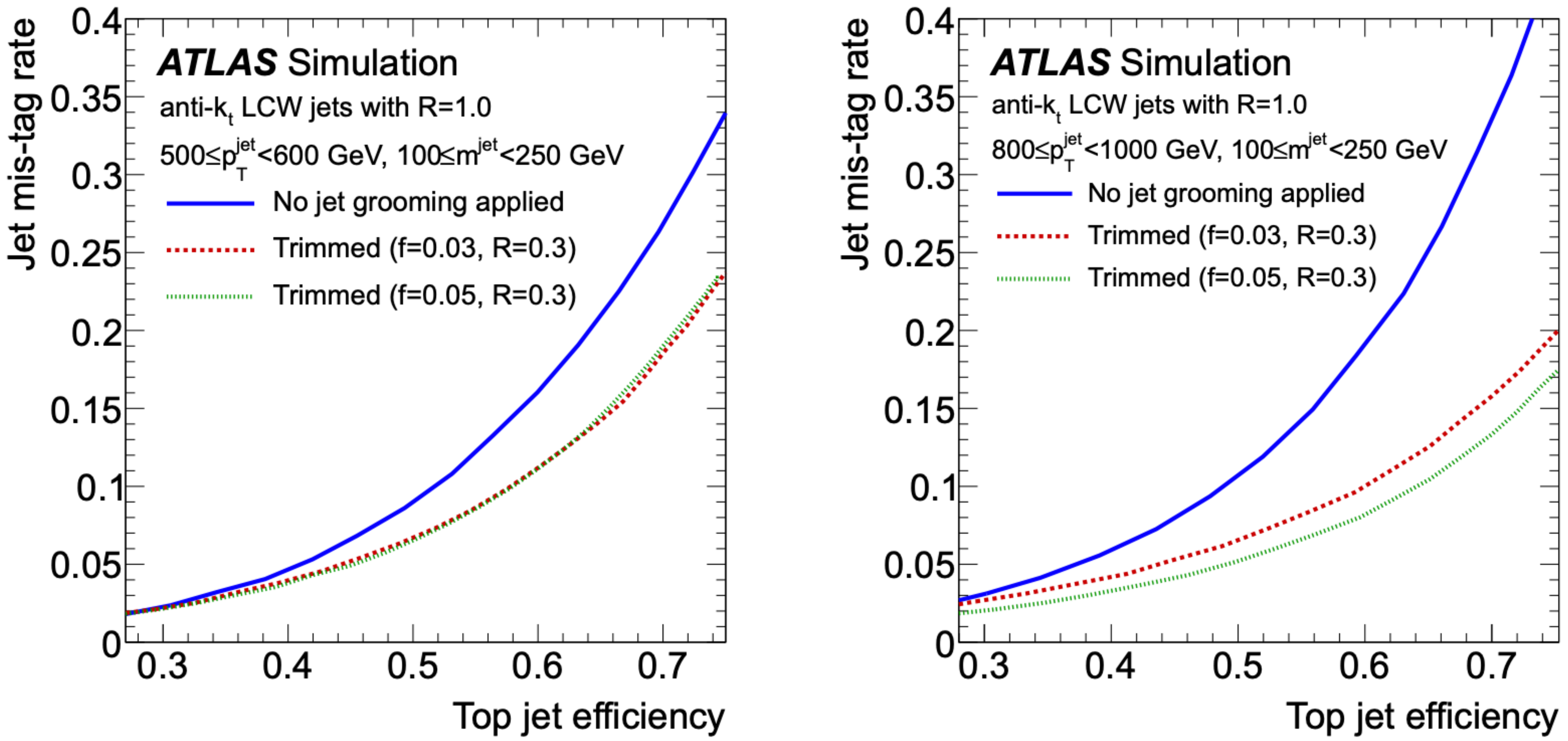
- If p = +1 k_t algorithm: proto-jets with the smallest p_T tend to be clustered first, so that the highest p_T proto-jets are clustered last.
- If p = -1 anti-k_t algorithm: proto-jets with the largest p_T are clustered first. A consequence of this is that isolated anti-k_t jets tend to be very close to circular in η-φ space, because the axis of the jet is relatively fixed after the first few steps of recombination. This stability makes anti-k_t jets more robust than k_t jets in high multiplicity environments.

Large-R jets: reconstruction and grooming

Jet substructure performances

Application of trimming algorithms

- example of Z' tt and diets events
- anti-k_t reconstructed jets with $600 < p_T < 800$ GeV
- utility of τ_{32} for reducing the jet mis-tag rate



(a) 600 GeV $\leq p_T^{jet} <$ 800 GeV (b) 800 GeV $\leq p_T^{jet} <$ 1000 GeV

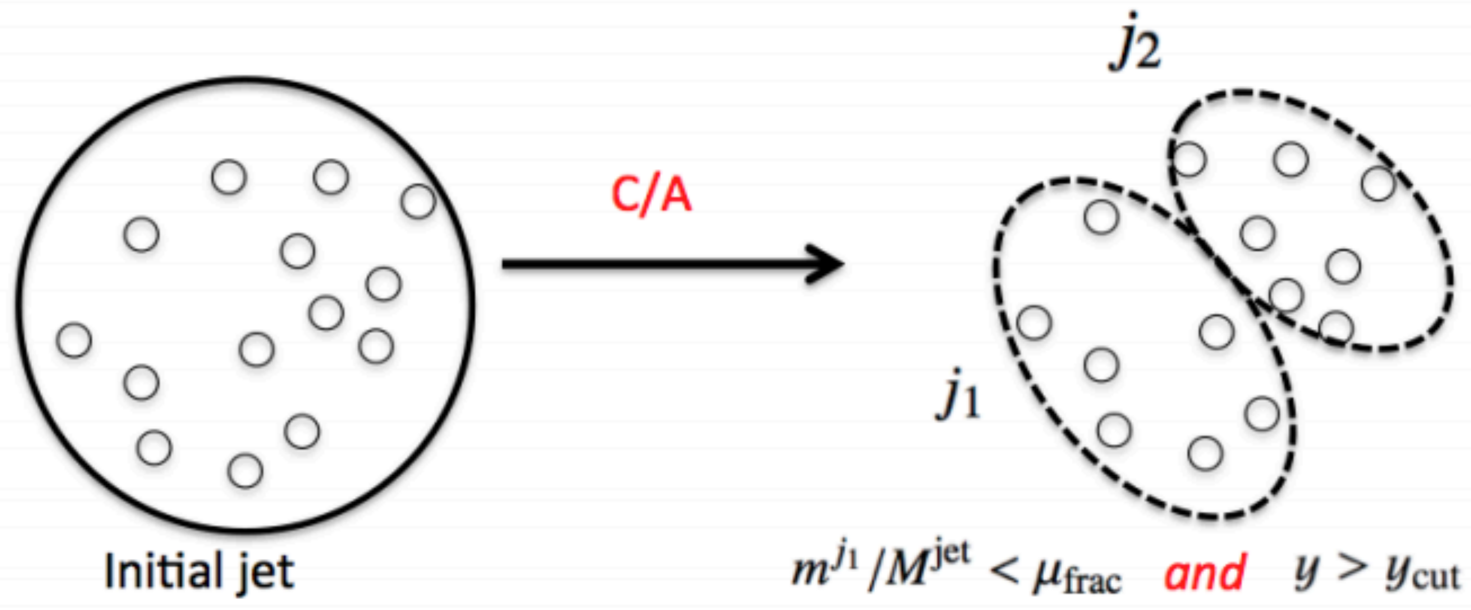
Figure 36. Jet mis-tag rate for dijet events vs. top-jet efficiency curves using τ_{32} as a top tagger for (a) 600 GeV $\leq p_T^{jet} <$ 800 GeV and (b) 800 GeV $\leq p_T^{jet} <$ 1000 GeV with masses in the range 100 GeV $\leq m^{jet} <$ 250 GeV.

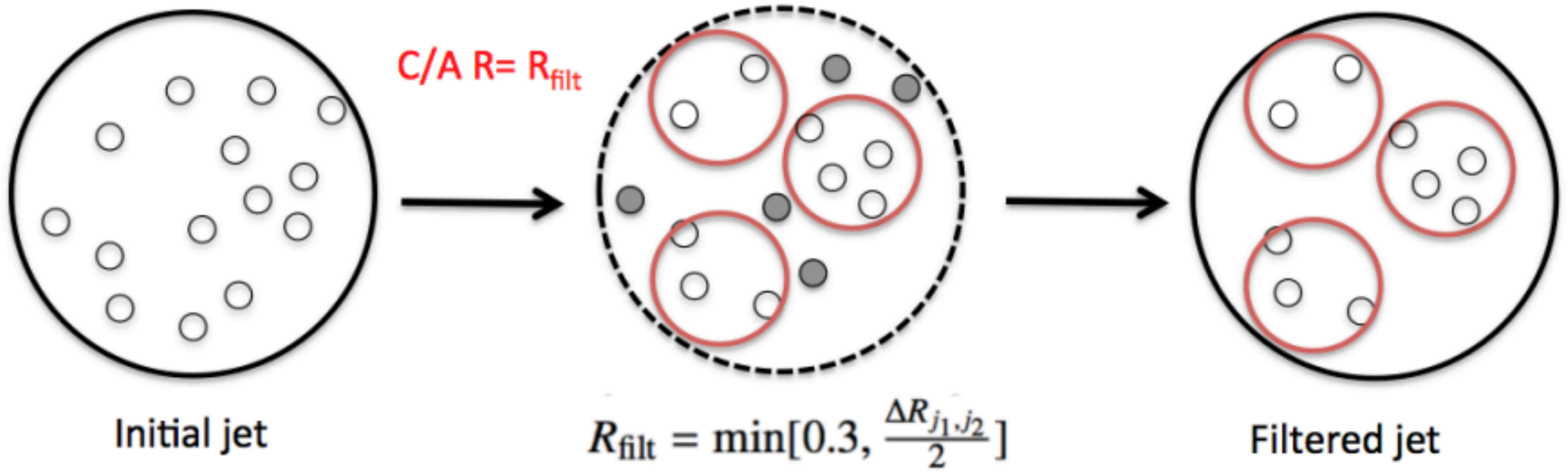
Large-R jets: reconstruction and grooming

Jet substructure performances

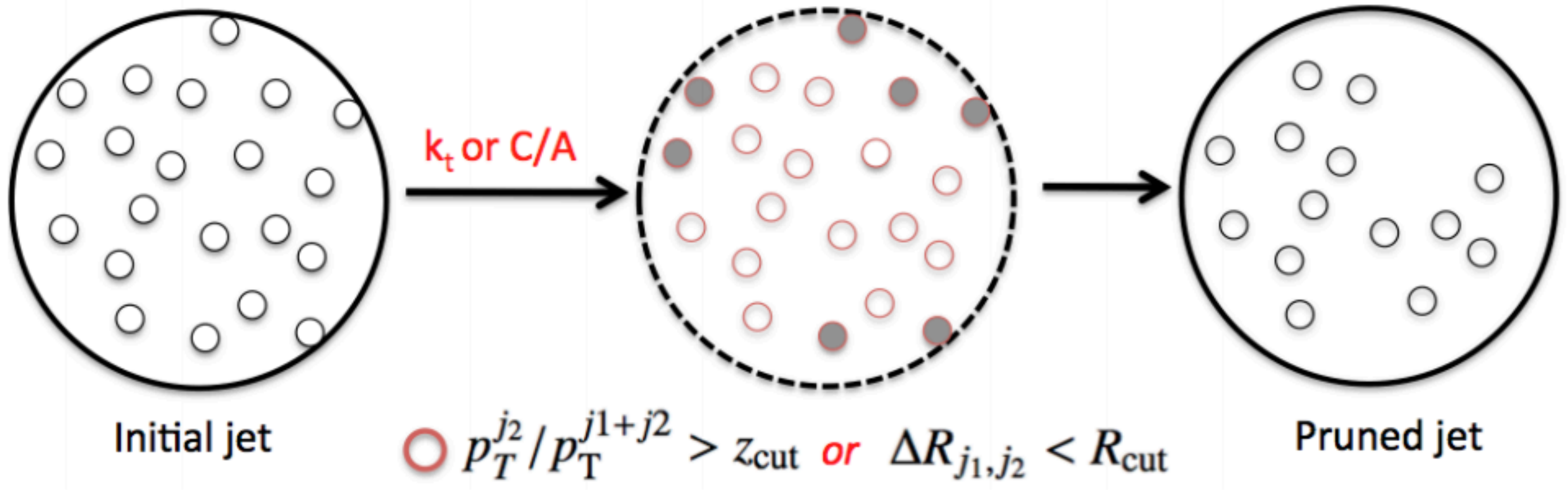
Other grooming algorithms

- **mass-drop filtering**





- **pruning**



- **trimming** (in main slides)

Large-R jets: tagging technique with jet moments

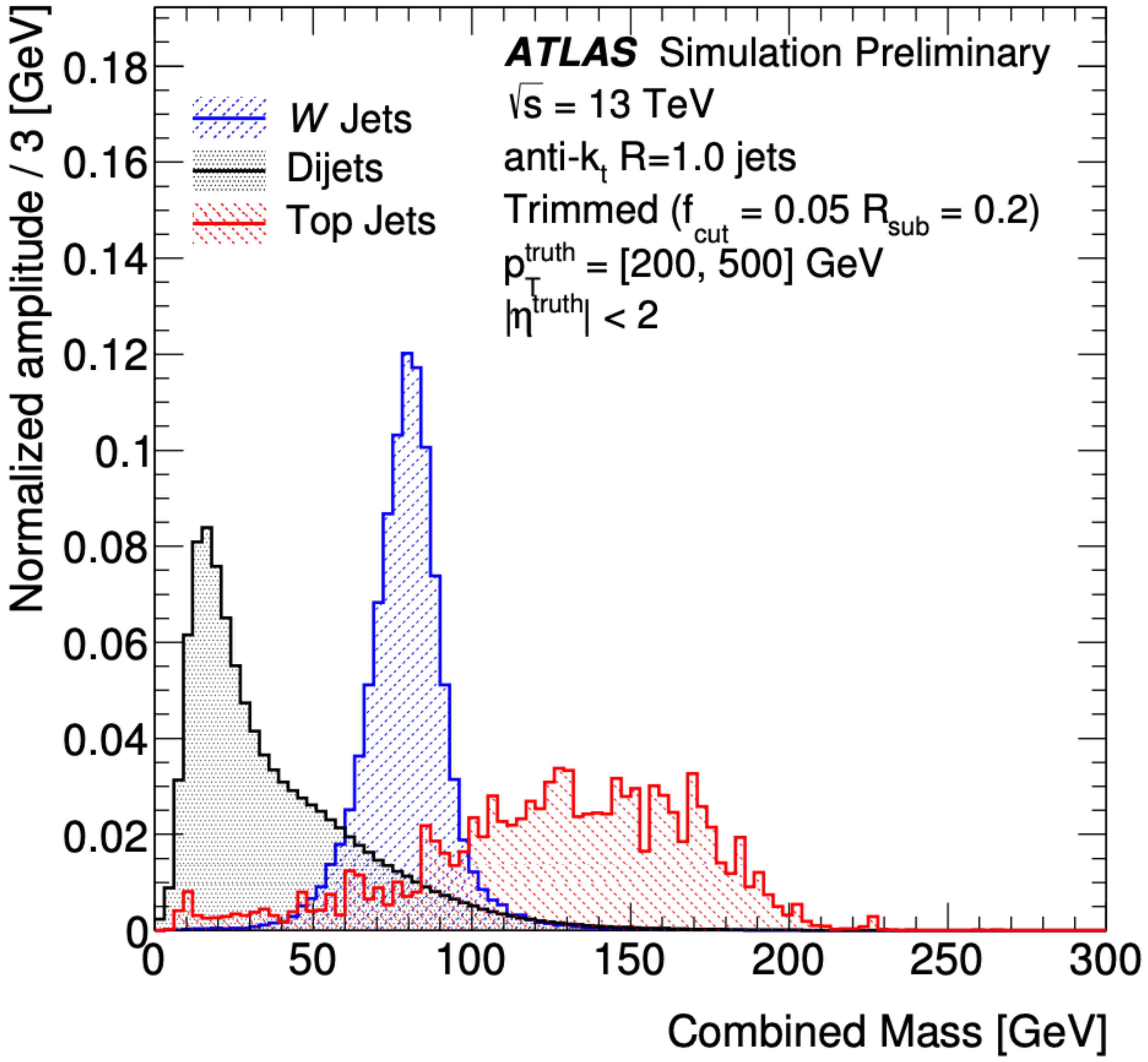
EPJ C (2019) 79-375

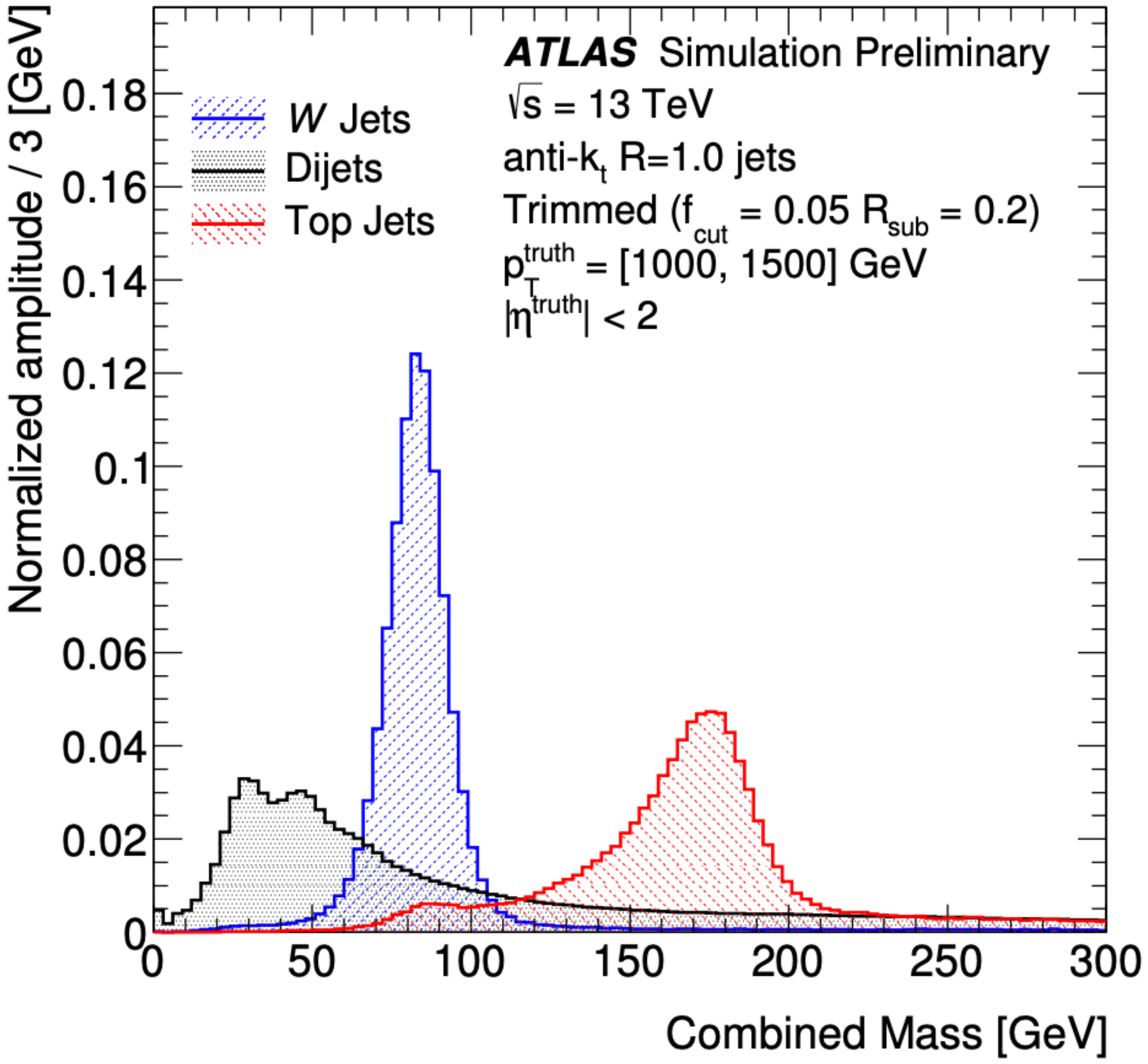
$$m^{\mathrm{TA}} = m^{\mathrm{track}} \times \frac{p_T^{\mathrm{calo}}}{p_T^{\mathrm{track}}}$$

$$w^{\mathrm{TA}} = \frac{\sigma_{\mathrm{TA}}^{-2}}{\sigma_{\mathrm{calo}}^{-2} + \sigma_{\mathrm{TA}}^{-2}}$$

$$w^{\mathrm{calo}} = 1 - w^{\mathrm{TA}}$$

- weights used for the m_{comb} calculation (linear combination of m_{TA} and m_{calo})





Large-R jets: tagging the Higgs boson

EPJ C (2019) 79-836

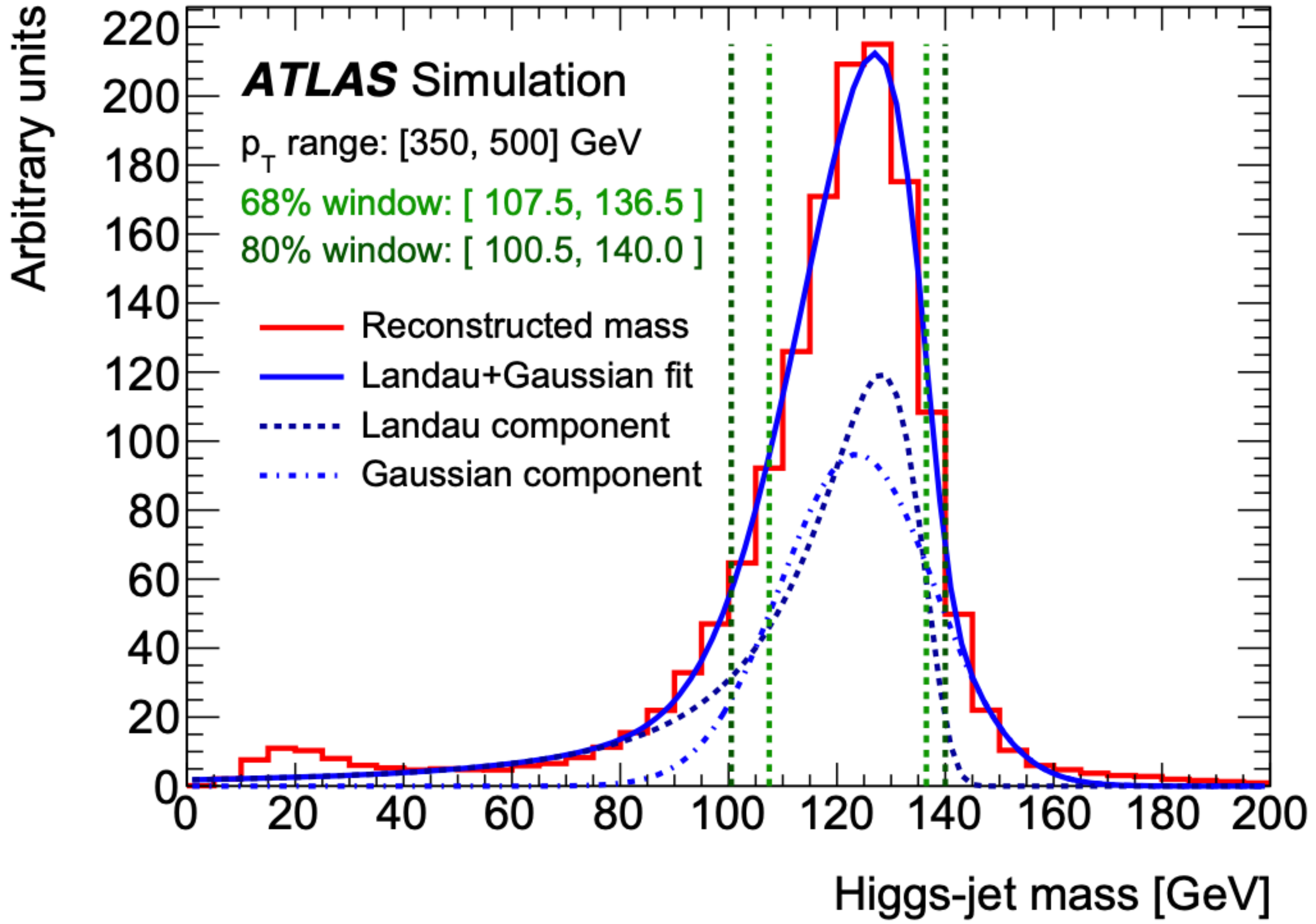


Fig. 4 The Higgs-jet mass distribution for jet transverse momenta in the range 350 to 500 GeV after reweighting the p_{T} spectrum. The dotted and dash-dotted blue curves correspond to the two components of the fit function, while the solid blue curve shows the combination thereof. The vertical lines indicate the boundaries of the mass ranges for 68% (light green) and 80% (dark green) containment

Mass window

- parametrised as a function of Higgs pT;
- defined as the smallest window containing the given fraction of Higgs-jets:
  - tight mass window, containing 68% of Higgs-jet;
  - loose mass window, containing 80% of Higgs-jets;

Large-R jets: tagging the Higgs boson

EPJ C (2019) 79-836

Large-R jets: tagging the Higgs boson

[EPJ C (2019) 79-836](#)

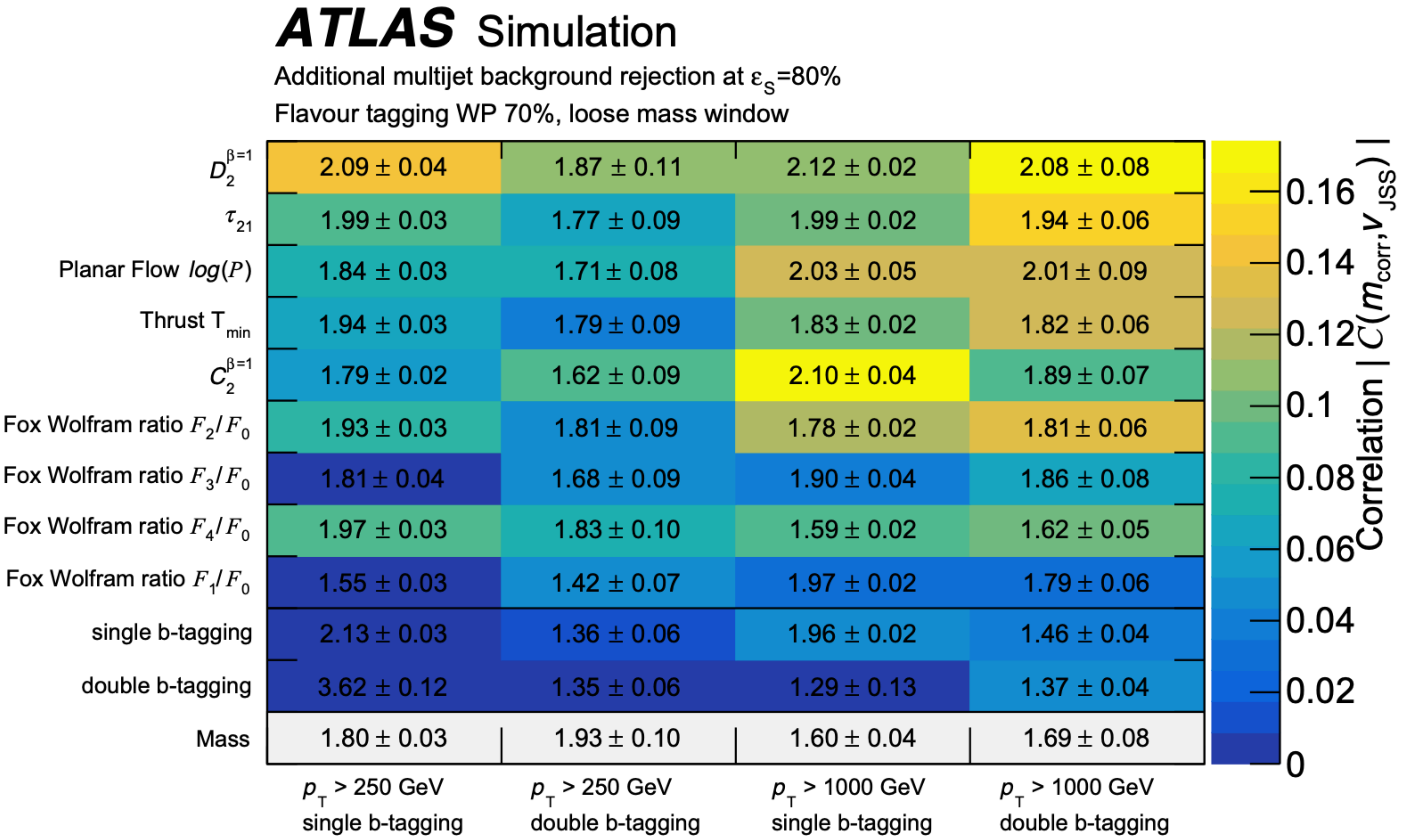


Fig. 11 Multijet background rejection at 80% signal efficiency (ε_S = 80%) for a variety of substructure variables using different benchmarks in terms of b-tagging strategy and transverse momentum range. The z-axis colour scale represents the absolute value of the linear-correlation coefficient, $|C(m_{\text{corr}}, v_{\text{JSS}})|$, between the jet mass and the jet substructure variables. The selection efficiency is determined relative to the mass window and b-tagging benchmark working points defined in Sects. 6.3 and 6.2 respectively

Large-R jets: tagging the Higgs boson

EPJ C (2019) 79-836

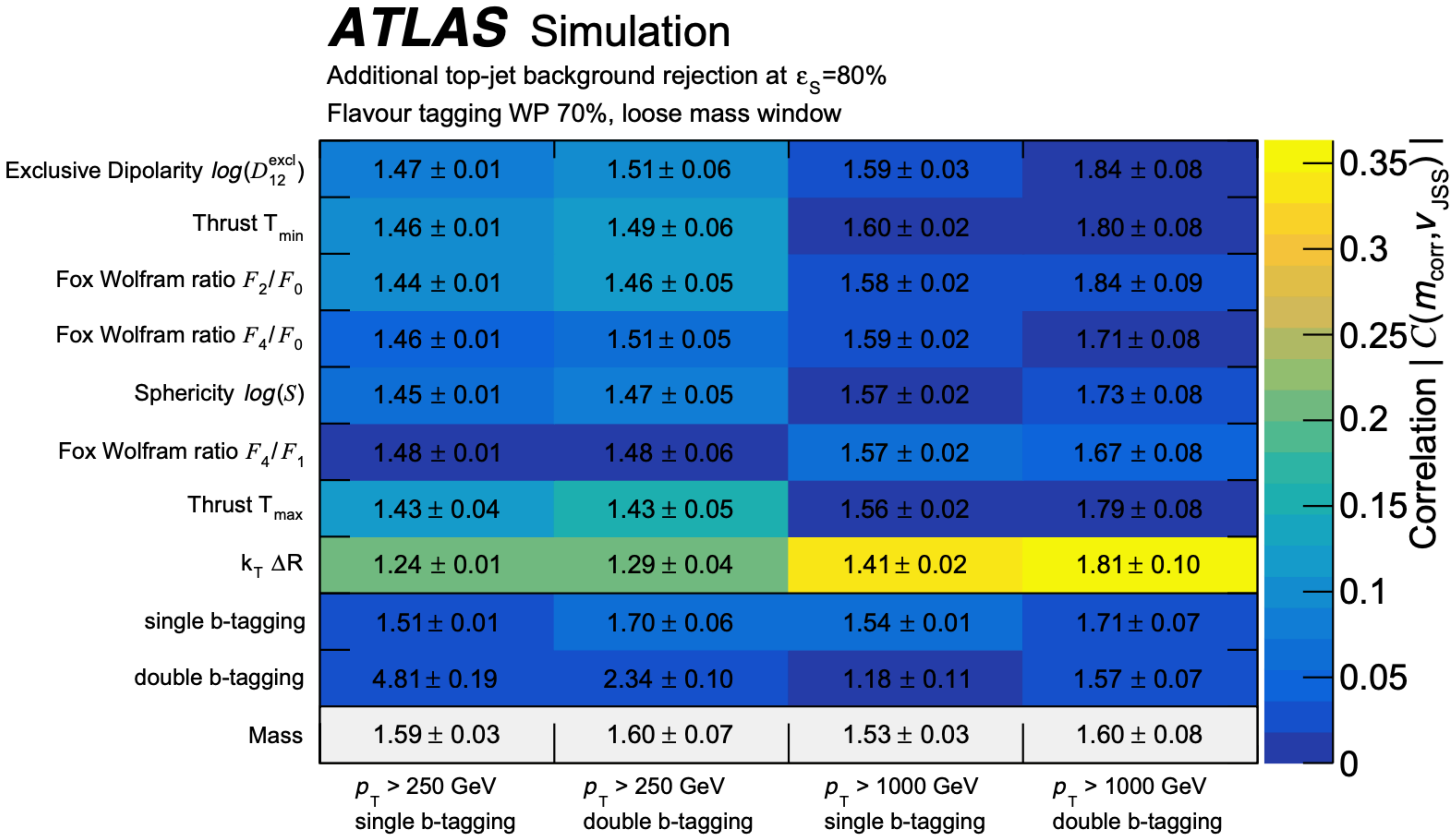


Fig. 12 Hadronic top-quark background rejection at 80% signal efficiency ($\varepsilon_S = 80\%$) for a variety of substructure variables using different benchmarks in terms of b-tagging strategy and transverse momentum range. The z-axis colour scale represents the absolute value of the linear-correlation coefficient, $|C(m_{corr}, v_{JSS})|$, between the jet mass and the jet substructure variables. The selection efficiency is determined relative to the mass window and b-tagging benchmark working points defined in Sects. 6.3 and 6.2 respectively

Large-R jets: re-clustering technique

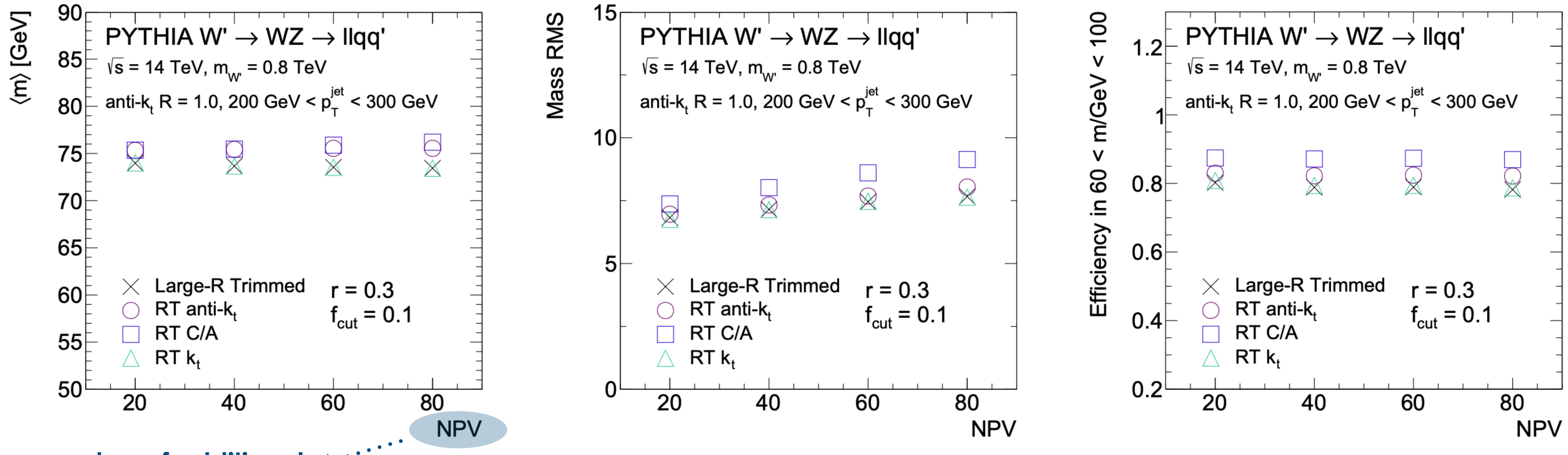
Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC

Performance of jet mass

- The most widely used large radius jet observable;
- to compare the performance of large radius jets with re-clustered jets, study of performance of the jet mass for the various re-clustering schemes;
- averages and deviations computed over a fixed mass range: 60-100 GeV.

$$m_{\mathrm{RCjet}} = \left(\sum_{i\in \mathrm{jet}} E_i\right)^2 - \left(\sum_{i\in \mathrm{jet}} \vec{p}_i\right)^2$$

- Jet mass performance is quantified by the average jet mass $\langle \mathbf{m} \rangle$, a mass resolution, **σ**, and the dependance of these quantities with the amount of **pileup**.



number of additional interaction vertices

for low pileup levels, essentially no differences between standard and RC jets.

Large-R jets: re-clustering technique

Jets from jets: reclustering as a tool for large radius jet reconstruction and grooming at the LHC

- performances for several small radius jet sizes;
- same RT applied.