Chapter 5

EVENT RECONSTRUCTION

The goal of particle physics experiments is to reconstruct and measure the outgoing particles produced in proton-proton collisions to describe the hard scatter process. After an event is accepted by the ATLAS trigger systems to be recorded to disk, the objects of interest such as electrons, muons, and jets must be reconstructed from the low-level detector signals. These complex objects, meant to be representative of the true Standard Model (SM) particle, are built from some of the low-level detector signals, such as muon spectrometer tracks or energy depositions in the electromagnetic or hadronic calorimeters. As the Large Hadron Collider (LHC) is a hadron collider, the LHC tends to produce colored final states through the collisions of gluons. Many Beyond the Standard Model (BSM) physics models contain these hadronic objects which are crucial to reconstruct accurately, amidst the initial and final state radiation and multiple simultaneous proton-proton collisions. Once reconstructed, the measured properties of these objects may be calibrated to a particular energy scale.

This chapter describes in detail the methods of standard ATLAS event reconstruction used for jets (section 5.1), b-jets (section 5.2), muons (section 5.3), electrons and photons (section 5.4), taus (section 5.5), and missing transverse momentum (section 5.6).

5.1 Jets

The first question with a particle physics detector is how to measure the hadronic final state. The difficult in reconstructing quark and gluons is because, due to the nature of QCD described in **??**, colored particles cannot be observed directly in the detector. Quarks and gluons get “smeared” by the showering process, and “obscured” through the hadronization process. The solution is to build objects called **jets**, the name for collimated sprays of particles produced by quarks and gluons as they shower and hadronize. The final step to connect theory in **??** to the calorimeters in **??** in order to measure jets are to cluster inputs from the calorimeter as shown in fig. 5.1. Proton-proton collisions produce partons which shower (parton jet) and then hadronize (particle jet) into colorless objects which deposit their energies in the hadronic calorimeter (HCal) and electromagnetic calorimeter (EMCal).

From these energy depositions, one can group them up to form jets, as shown in the candidate event from my Supersymmetry (SUSY) search (fig. 5.2). In the event display, the six jets are represented by cones around the groups calorimeter energy deposits, but this is where things start getting tricky. How do we deal with the underlying event that consists of initial and final state radiation, of pile-up, of the multiple, simulateneous proton-proton collisions that can obscure the physics of the hard scatter? How can we start to identify more than two or three jets cleanly even as we move to a higher luminosity with more proton-proton interactions per Bunch Crossing (BC)?

The rest of this section is dedicated to describing the goal of jet clustering algorithms (section 5.1.1), calibrating the energy of jets (section 5.1.2), characterizing the uncertainty of the jet calibrations (section 5.1.4), and some kinematic properties of jets at $\sqrt{s} = 13\,\mathrm{TeV}$ (section 5.1.5).

5.1.1 Jet Algorithms

The shower of particles in the calorimeter originating from the fragmentation and hadronization of quarks and gluons produces objects known as jets. However, jets are not unique, and are instead defined based on the clustering algorithm, its parameters, and its inputs. In the Snowmass Accords of 1990 [3], there was a first attempt to define a set of requirements jet

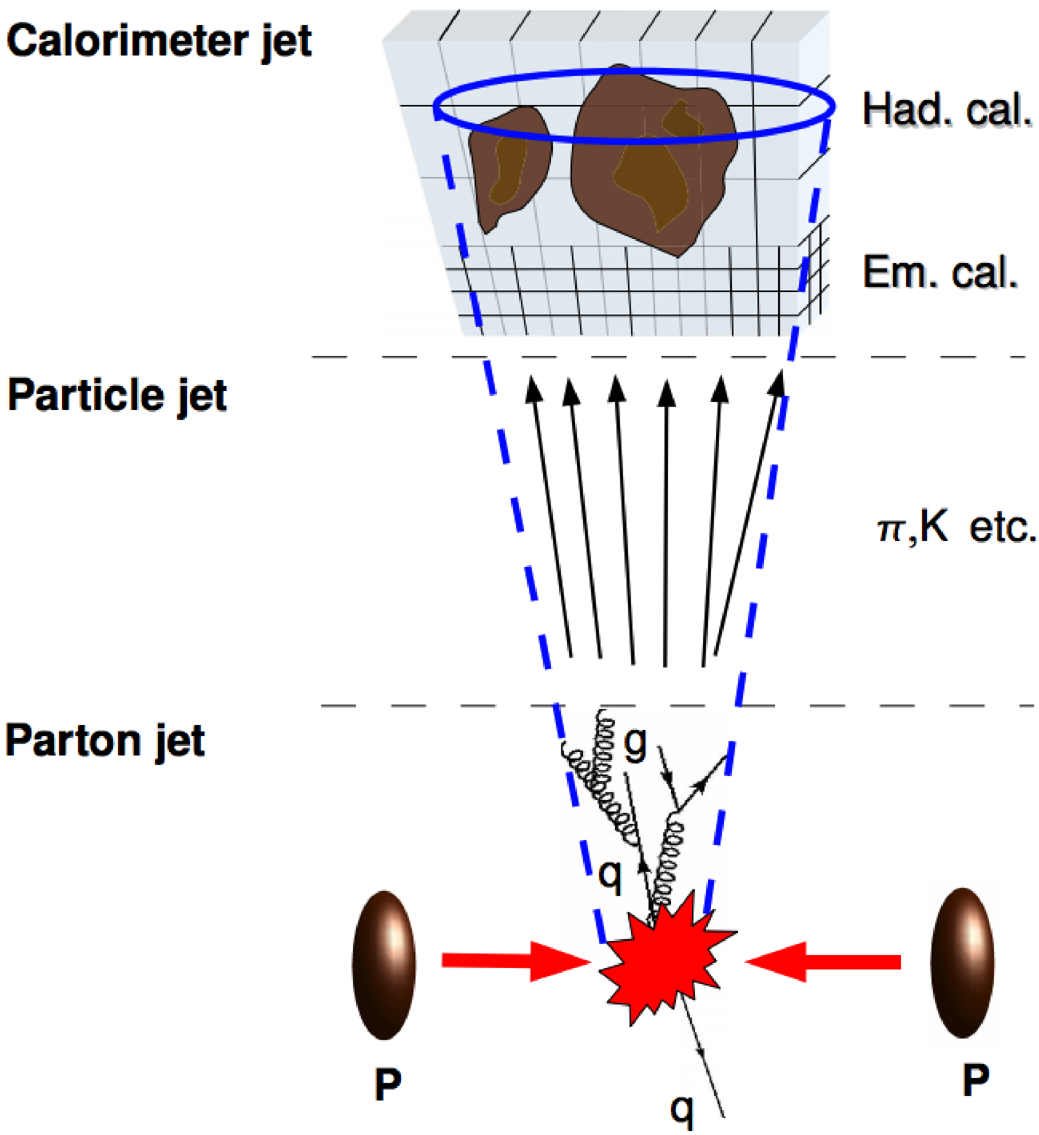


Figure 5.1: [1] The evolution of the partonic system which demonstrates how jets are formed and measured by the calorimeters in ATLAS.

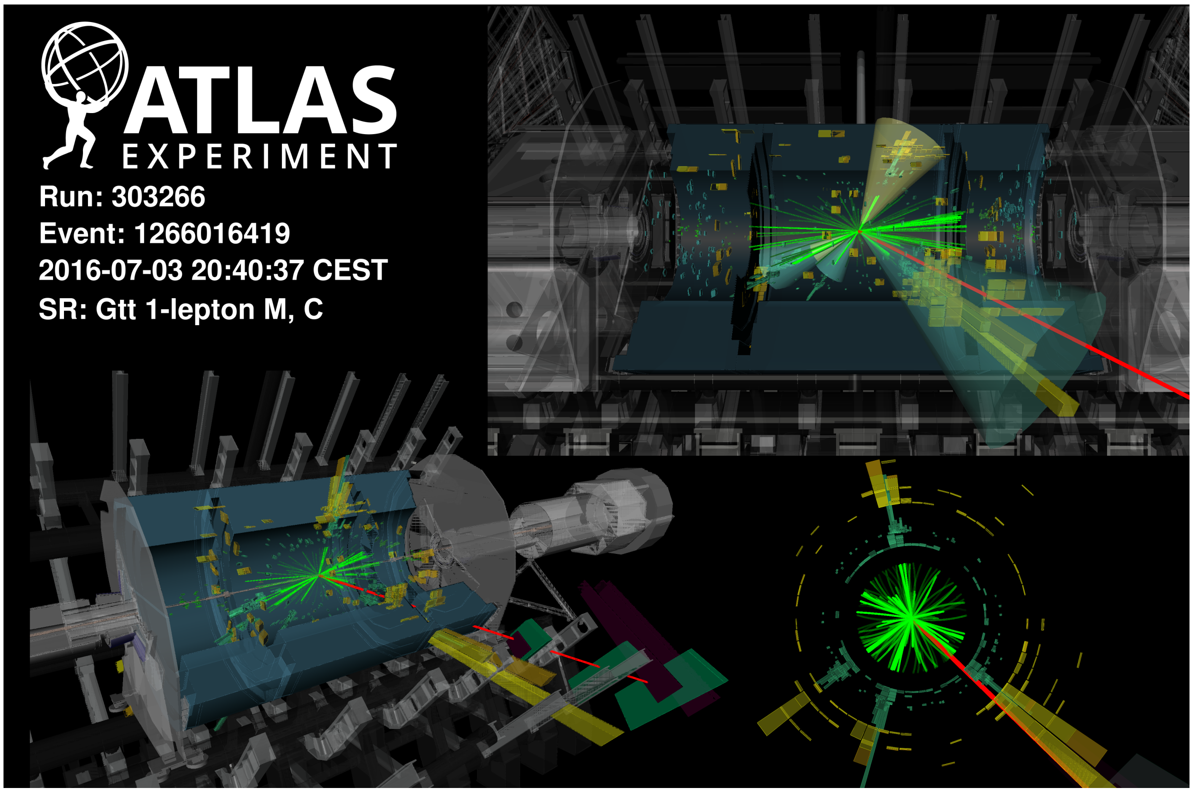


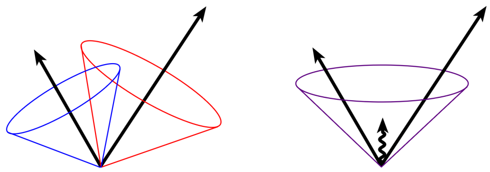
Figure 5.2: [2] A candidate supersymmetry event with 6 jets (shown as cones) and a muon track which is colored red. The b-quark tagged jets are colored blue, while the non-b-quark tagged jets are colored yellow. The size of the jet cones are proportional to the measured and calibrated p_{T} of the jet. Inner Detector (ID) tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ are colored green and their brightness is proportional to their p_{T}. The LAr and Tile calorimeter (Tile) are colored teal and yellow, respectively, and their length is proportional to the measured transverse energy deposit.

algorithms needed to fulfill including:

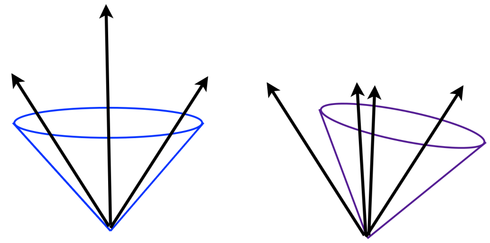
- simple to implement in an experimental analysis;
- simple to implement in the theoretical calculation;
- defined at any order of perturbation theory;
- yields finite cross section at any order of perturbation theory;
- yield s cross section that is insensitive to hadronization.

Over the past 25 years, many algorithms have come and gone. The so-called cone algorithms [4], have become the most popular option by experimental physicists [5]. Typically, these algorithms collect all particules within some stable, seeded cone defined by a size parameter R, the cross-sectional area of the cone in the η-ϕ plane. These "simple" cone algorithms were used with mixed success as they did manage to satisfy many of the requirements for jet algorithms, however they were found to not be **infrared-safe** (IR-safe) and **collinear-safe** as shown in fig. 5.3. IR safety is a theoretical guideline that adding or removing soft particles should not change the result of jet clustering. In practice, the underlying event and pile-up activity should not affect the jet final state in hard scattering. Collinear safety states that the splitting of a large p_{T} particle into two collinear particles should not affect the jet finding, or the number of jets identified.

For example, a seeded cone algorithm fails to be collinear-safe, as the choice of seed is often the one with the most energy, and accidental collinear fragmentation can suppress the seed. Similarly, cone algorithms that varied the center of the jet as it clustered more inputs together would not be IR-safe as the addition of soft gluon radiation could shift the center enough to overlap with an existing jet. Both IR-safety and collinear-safety have dictated the guidelines for the next generation of cone algorithms. By changing focus from seeds and the cones



(a) infrared-safety



(b) collinear-safety

Figure 5.3: [6] Illustration of the (a) infrared sensitivity and (b) collinear sensitivty of a cursory designed jet algorithm.

around seeds, to relationships between close-by inputs, one can satisfy both of the safety requirements. Instead of identifying inputs based on their distance to a seed, identify pairs of inputs based on their distance parameter. If two inputs have a distance parameter below some threshold, they are merged and the process continues until no merges are possible [4]. This was able to solve the safety requirements by merging together both collinear particles as well as soft particles. This next generation, known as sequential recombination algorithms, was defined in six steps depending on two parameters, R (size parameter) and P (algorithm choice, explained further below):

1. For each protojet i, define the **jet-beam distance measure**,

$$d_i = p_{\mathrm{T},i}^{2P}. \tag{5.1}$$

2. For each pair of protojets $i, j \forall i \neq j$, define the **jet-jet distance**,

$$d_{i,j} = \min(d_i, d_j) \left(\frac{\Delta R_{i,j}^2}{R^2} \right). \tag{5.2}$$

3. Minimize the set of d_i and $d_{i,j}$ so far and call this $d_{\min}$.

4. If $d_{\min}$ belongs to a protojet i (the set of d_i), then the protojet is not mergeable, removed as an input, and is defined as a jet.

5. If d_{min} belongs to the distance between protojets i, j (the set of $d_{i,j}$), then the two protojets are removed, merged into a single protojet k, and added as an input.
6. This repeats until there are no more remaining protojets i.

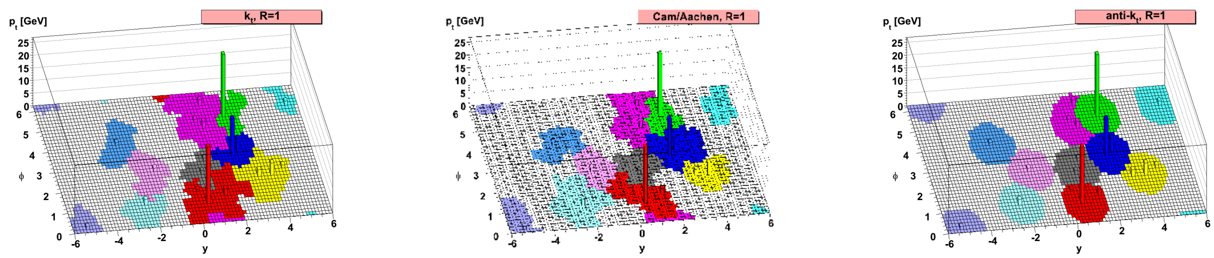
The choice of the parameter P corresponds with the choice of particular algorithm which is applied.

$P = 1$ This defines the k_t algorithm [7]. Softest protojets are clusters first into harder ones. As soft radiation is prioritized, this algorithm can be susceptible to energy fluctuations from pile-up and detector noise. This typically forms irregularly-shaped jets.

$P = 0$ This defines the Cambridge-Aachen (C/A) algorithm [8]. This omits all trace of p_{T}-dependence in the clustering and depends only on angular properties. This was still susceptible to some of the same problems at the k_t algorithm, being sensitive to soft gluon radiation. Also like k_t algorithm, this typically forms irregularly-shaped jets as well.

$P = 1$ This defines the anti-k_t algorithm [9, 10]. This algorithm prefers the hardest protojets and clusters them first. This is similar to the cone algorithms because it prefers hardest protojets which are seed-like.

The `FastJet` software package [10] contains the technical execution of the jet clustering algorithms and was able to reduce the complexity of jet finding to $\mathcal{O}(n \ln n)$ for n inputs [11]. Figure 5.4 depicts the three different algorithm choices for $P = 0, \pm 1$. For this thesis, the anti-k_t algorithm is used with an R parameter $R = 0.4$ which is a typical parameter for small-radius jets in ATLAS. Larger jets are typically $R = 1.0$ or $R = 1.2$, are more sensitive to pile-up contributions, but are not used in this analysis.



(a) k_t (b) C/A (c) anti-k_t

Figure 5.4: [9] A sample parton-level monte-carlo simulated event illustrating the shapes and areas of the resulting hard jets for $R = 1.0$ and (a) $P = 1$, (b) $P = 0$, and (c) $P = -1$.

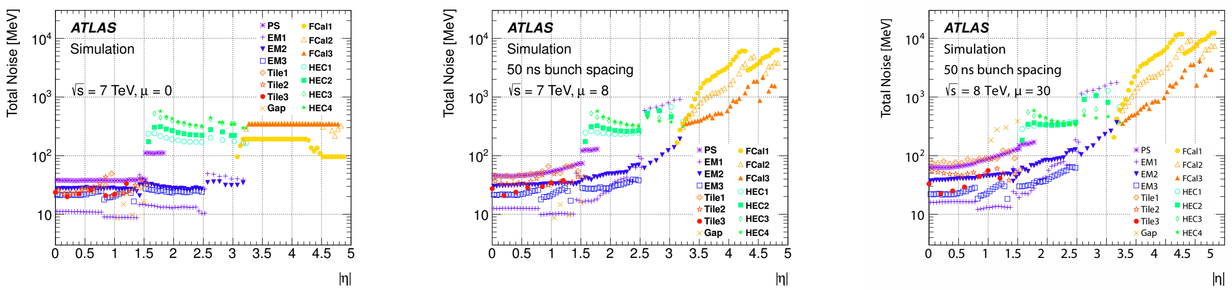
The last consideration is about the inputs to these jet algorithms which form different kinds of jet collections for a given algorithm. For example, to produce **truth jets**, inputs are simulated particles by a Monte-Carlo generator. These are generally used to study the performance of algorithms without detector effects and to calibrate/study the resolution of other reconstructed jets. Jets can be formed from tracks, **track jets**, inputs are the hits in the ID which correspond to trajectories of charged particles. Another set of jets are **calorimeter jets**[1], inputs are energy deposits left in the calorimeter for both neutral and charged particles.

Given the computing budget for ATLAS and the ability to do offline reconstruction quickly for analyses, using the energy measurements in the calorimeter at the cell level, readouts, is computationally intensive. Calorimeter cells are also very sensitive to pile-up and a single quark could shower over many cells. In pre-2011 ATLAS operations, the cell noise was almost entirely electronic noise. Now that pile-up is significant, the noise contribution from pile-up is also dominating as shown in fig. 5.5 and summarized in eq. (5.3) [12].

$$\sigma_{\text{noise}} = \begin{cases} \sigma_{\text{noise}}^{\text{electronic}} & \text{(2010 and prior)} \\ \sigma_{\text{noise}}^{\text{electronic}} \oplus \sigma_{\text{noise}}^{\text{pile-up}} & \text{(2011+)} \end{cases} \tag{5.3}$$

[1]Colloquially known as "reco jets", short for reconstructed jets.

In eq. (5.3), $\sigma_{\text{noise}}^{\text{electronic}}$ is the electronic noise and $\sigma_{\text{noise}}^{\text{pile-up}}$ is the noise from pile-up determined with Monte-Carlo simulations for $\langle\mu\rangle \geq 8$.



(a) 2010 (b) 2011 (c) 2012

Figure 5.5: [13] The energy-equivalent cell noise in the ATLAS calorimeters on the electromagnetic scale as a function of $|\eta|$ in the detector for the (a) 2010 configuration with $\langle\mu\rangle = 0$, (b) 2011 configuration with $\langle\mu\rangle = 8$, and (c) 2012 configuration with $\langle\mu\rangle = 30$. The various colors indicate the noise in the first layer of the forward calorimeter (FCal) and the three layers of the LAr EMCal, the three layers of the Tile, the four layers of the HEC, and the three modules of the FCal. The higher levels in 2011 and 2012 compared to 2010 indicate the contribution from increased pile-up.

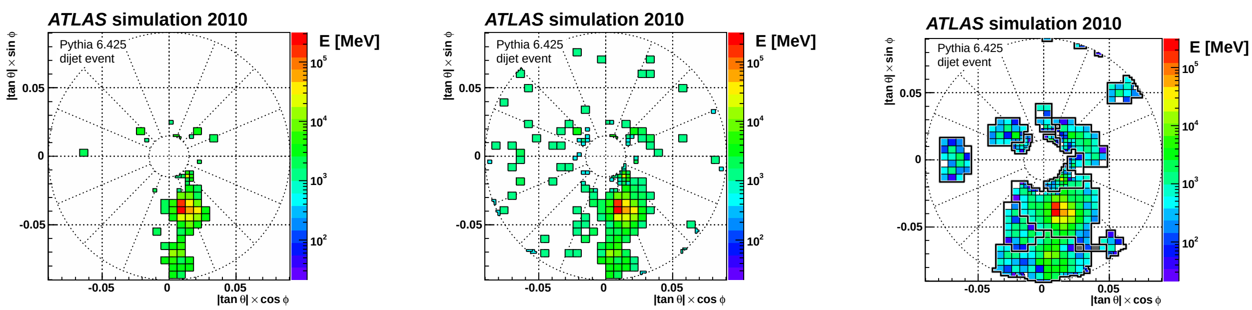
Given this noise contribution, calorimeter cells were clustered together to form **Three-dimensional topological clusters** (or topoclusters, for short) using an algorithm designed to maintain a certain amount of cell signal to the average expected noise in the electromagnetic calorimeters [13]. These topoclusters are seeded by cells with a large signal-to-noise ratio[2], S, and then these seeds are summed with the scalar energy measured in adjacent cells, N, up to a boundary condition, P. This algorithm is shown in eq. (5.4) and the three particular parameters $S > 4, N > 2, P > 0$ were chosen and optimized using studies with ATLAS test-beam data [12].

[2]Here, the noise is the expected noise σ_{noise} in eq. (5.3).

$$\left|E_{\text{cell}}^{\text{EM}}\right| > S\sigma_{\text{noise, cell}}^{\text{EM}} \qquad \text{(primary seed threshold);} \tag{5.4a}$$

$$\left|E_{\text{cell}}^{\text{EM}}\right| > N\sigma_{\text{noise, cell}}^{\text{EM}} \qquad \text{(threshold for growth control);} \tag{5.4b}$$

$$\left|E_{\text{cell}}^{\text{EM}}\right| > P\sigma_{\text{noise, cell}}^{\text{EM}} \qquad \text{(boundary cell filter).} \tag{5.4c}$$



(a) $S > 4$ (b) $N > 2$ (c) $P > 0$

Figure 5.6: [13] Stages of topocluster formation in the first module of the FCal for a simulated dijet event with at least one jet entering the calorimeter is shown for $\sqrt{s} = 8\,\text{TeV}$. Shown are (a) seed cells for topoclustering, (b) growth cells, and (c) boundary cells. Cells which are not colored, but contained inside a topocluster boundary are cells with negative energy.

Figure 5.6 shows the three stages of topoclustering described by eq. (5.4). Since negative energies are allowed to join topoclusters, primarily caused by the pulse shaping of the LAr calorimeter and caused by pile-up, are expected to partially cancel the positive fluctuations by pile-up. Finally, topoclustering also improves the performance of the calorimeter by suppressing isolated fluctuations due to pile-up and electronic noise.

5.1.2 Jet Calibrations

The jets formed from topoclusters are reconstructed at the electromagnetic scale as described in section 5.1.2 are still not ready for use by analyses. A series of further corrections are derived from both monte-carlo simulation and data[3] to account for the non-compensating[4] nature of the ATLAS calorimeters [14, 15]. Calibrations derived using both monte-carlo and data are applied in sequence as described in section 5.1.3 in order to get the energy **scale** of the jet as close to the scale of the particle. Section 5.1.4 describes the systematics and uncertainties associated with these corrections and calibrations that need to be considered by analyses using the calibrated jets. Note that MC simulation which is built on test beam data, prior collision data, and theoretical considerations is not a perfect description of the detector response, hence the need for in-situ calibrations after all other considerations.

Topocluster Calibration

As topoclusters are used as inputs for the anti-k_t clustering algorithm in this analysis, the first step is to calibrate the topoclusters. The intention of the calibration scheme for topoclusters is to provide a calorimeter signal for physics object reconstruction that is agnostic to the kind of objeect being reconstructed. In addition, the signal from hadronically-interacting particles will always be smaller than the signal from electromagnetically-interacting particles depositing the same amount of energy due to the non-compensating nature of the hadronic calorimeters. Finally, one needs to account for energy losses due to dead material and inefficiencies of the clustering procedure itself. Therefore, a weight is assigned to each calorimeter cell based on the probability $P^{\text{EM}}_{\text{clus}}$ of the topocluster to have been generated by an electromagnetic shower, using the kinematics of the topocluster:

$$w^{\text{cal}}_{\text{cell}} = \underbrace{P^{\text{EM}}_{\text{clus}} \cdot 1}_{w^{\text{EM}}_{\text{cell}}=1} + (1 - P^{\text{EM}}_{\text{clus}}) \cdot \underbrace{\frac{E^{\text{dep}}_{\text{cell}}}{E^{\text{EM}}_{\text{cell}}}}_{w^{\text{Had}}_{\text{cell}}} \tag{5.5}$$

[3]Data-derived corrections are called **in-situ** corrections.

[4]The energy measured by the detectors is not the full energy of the particle that is being detected/measured.

By default, all topoclusters are reconstructed at the electromagnetic scale[5], so $w_{\text{cell}}^{\text{EM}} = 1$. $w_{\text{cell}}^{\text{Had}}$ is the ratio of the energy deposited per-cell to the same energy reconstructed at the electromagnetic scale. Neutral pion showers, $\pi^0 \to \gamma\gamma$, are used to calibrate the electromagnetic likelihood; postively charged pions, $\pi^+ \to \pi^0 + e^+ + \nu_e$, are used to train the hadronic likelihood.

For Run I in ATLAS, $0 < P_{\text{clus}}^{\text{EM}} < 1$ and jets built from topoclusters were known as LCW jets (local cell weighting)[6]. In Run 2, the default jets were EM jets[7] built from topoclusters at the electromagnetic scale with $P_{\text{clus}}^{\text{EM}} = 1$.

5.1.3 Jet Energy Calibration

The jet calibration procedure summarized in fig. 5.7 is intended to correct for several detector effects that affect the jet energy measurement:

- dead material: energy lost in the dead material of the detector, e.g. inactive absorbers and instrumentation
- non-compensation: difference in detector response between hadrons, leptons, and photons; specifically the response is lower for hadrons
- punch-through: energy leakage where the hadron showers deposit energy outside of the calorimetry system
- pile-up: energy originating from additional proton-proton collisions in the detector (part of the underlying event which includes initial and final state radiation)

[5]What this means is that the measured signal from the electromagnetic cells and hadronic cells are used with no other cell-level corrections.

[6]Colloquially called “LCTopo”.

[7]Colloquially called “EMTopo”.

- minimum threshold: hardware limits on energy deposits

- out-of-cone: inefficiencies in reconstruction due to not capturing the full particle shower in the jet

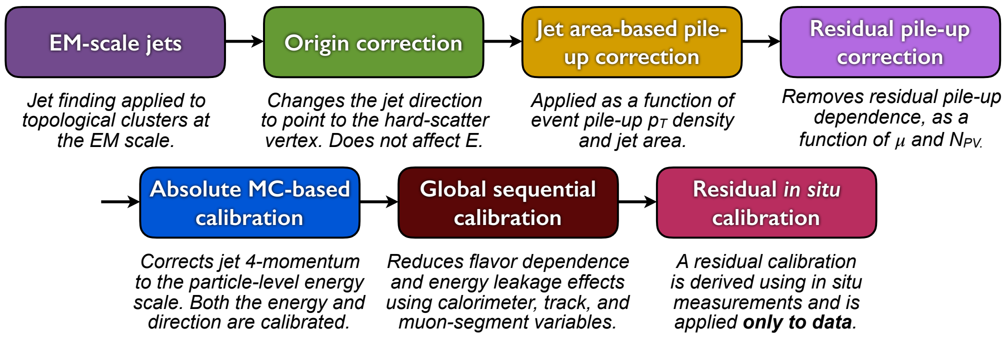


Figure 5.7: [16] Calibration stages for EM-scale jets. Other than the origin correction, each stage of the calibration is applied to the four-momentum of the jet.

The following subsections briefly describe each stage, more detailed information is found in [16].

Jet Origin Correction

In this stage, the jet direction is corrected for the difference between the default ATLAS detector origin, the geometric center of the detector, and the actual position of the primary proton-proton interaction. In reality, particles point back to vertices at the main interaction point. Therefore the primary vertex of the event, especially in the case of multiple proton-proton interactions, is identified by the vertex consisting of the largest $\sum \vec{p}_{\mathrm{T}}^{2}$ of tracks and the origin of the jet is corrected to point back towards this primary vertex. This correction improves the angular resolution of jets with a small effect on jet p_{T}. After this correction, the jets are said to be at the **origin corrected scale**.

Pile-up Correction

In this stage, the average additional energy due to multiple proton-proton interactions is subtracted from the jet using an area-based subtraction scheme [17]. The average energy density is calcuted using k_t $R = 0.4$ jets described in eq. (5.6) as the median energy density calculated from the area A^i and transverse momentum p_{T}^i of each jet i.

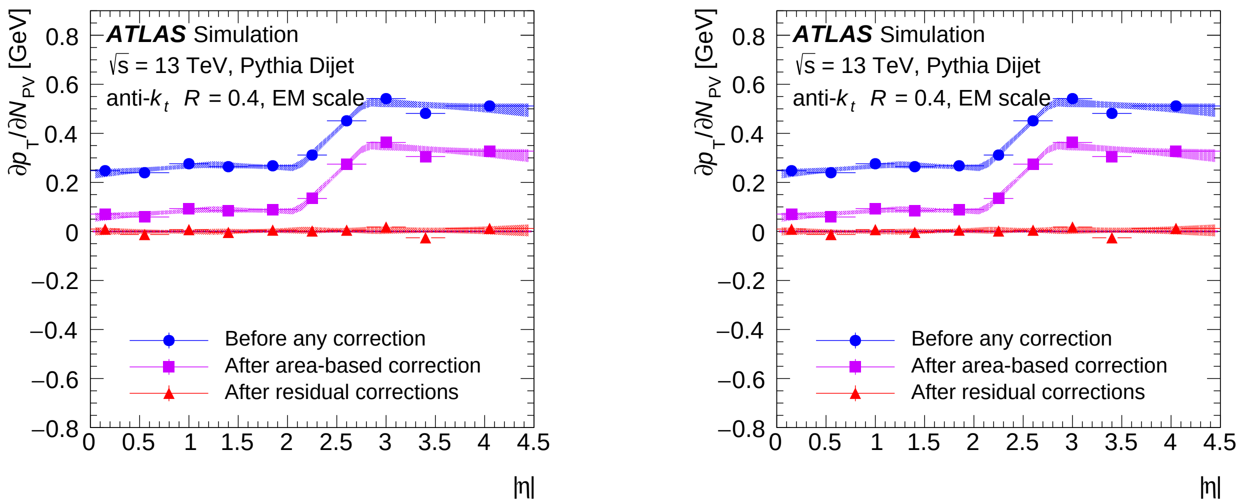
$$\rho = \mathrm{median}\left(\frac{p_{\mathrm{T}}^i}{A^i}\right)\Bigg|_{k_t\ \mathrm{jets}} \tag{5.6}$$

ρ represents the pile-up energy density of the calorimeter. Note an interesting feature is that while topoclustering mitigates pile-up correction partially by allowing for negative-energy cells to cancel out positive-energy cells, the k_t clustering only permits positive-energy topoclusters in the calculation of pile-up energy density. The inclusion of negative energies in the calculation of pile-up is an interesting topic that will be explored more in **??**.

The p_{T} of each jet is corrected by a variety of factors shown in eq. (5.7). One factor is to estimate the amount of pile-up in a jet of area A using $\rho \times A$; one factor is a simulation-based residual correction based on the position of the jet in pseudorapidity $\alpha(\eta)$ and number of primary vertices N_{PV}; and one factor is based on the average number of interactions per bunch crossing μ for the position of the jet in pseudorapidity $\beta(\eta)$. The different functions $\alpha(\eta)$ and $\beta(\eta)$ are determined from simulation and allows for both in-time and out-of-time pile-up effects to be accounted for as shown in fig. 5.8. The residual correction factors α and β are taken from a fit to the purple bands.

$$p_{\mathrm{T}}^{\mathrm{corr}} = p_{\mathrm{T}}^{\mathrm{EM}} - \rho \times A - \alpha(\eta) \times (N_{\mathrm{PV}} - 1) - \beta(\eta) \times \mu \tag{5.7}$$

Notice that there are four sources of uncertainty introduced by this calculation by N_{PV}, μ, ρ, and an uncertainty due to the p_{T}-dependence of the correction itself. These uncertainties are included as part of the jet energy scale (JES) uncertainties discussed in section 5.1.4. After the correction is applied to jets, the jets are said to be at the **pile-up corrected scale**.



(a) in-time pile-up

(b) out-of-time pile-up

Figure 5.8: [16] Dependence of the EM-scale anti-k_t jet p_{T} on (a) in-time pile-up and (b) out-of-time pile-up as a function of $|\eta|$ for $p_{\mathrm{T}}^{\mathrm{truth}} = 25\,\mathrm{GeV}$. The dependence is shown before pile-up corrections (blue), after area-based pile-up correction (purple), and after residual corrections (red) using N_{PV} and μ.

MC-based Correction

This correction is also known as the JES which is meant to correct the response of the jet energy and jet η in the detector back to the truth level. A correction is derived from truth information in Monte-Carlo simulations in both p_{T} and η, due to both the non-compensating nature of the ATLAS calorimeters and the changing geometry as a function of η. Reconstructed jets are first required to be isolated which requires that the minimum ΔR between pairs of jets is no less than $2.5 \times R$. In events with isolated jets, the reconstructed jets are matched to the corresponding truth jet using a ΔR association requiring that the truth jet and reconstructed jet are less than $0.75 \times R$. If a reconstructed jet has no matching truth jet, it is discarded.

$$\mathcal{R}^{\text{jet}} = \frac{E^{\text{jet}}_{\text{reco}}}{E^{\text{jet}}_{\text{truth}}} \tag{5.8}$$

The jet energy response (see eq. (5.8)) of isolated, reconstructed jets in QCD multijet simulation is binned in energy of the matched truth-jet $E^{\text{jet}}_{\text{truth}}$ [8] (see fig. 5.9) and pseudorapidity η_{det} in the detector. Each bin produces a Gaussian distribution which is fit and the mean value is extracted, $\langle\mathcal{R}^{\text{jet}}\rangle$. This peak value is used to transform $\langle E^{\text{jet}}_{\text{truth}}\rangle \mapsto \langle E^{\text{jet}}_{\text{reco}}\rangle$ for each η_{det} bin, known as the "numerical inversion" procedure to derive corrections in reconstructed jets from truth jets.

$$E^{\text{jet}}_{\text{EM + JES}} = \frac{1}{\mathcal{F}(E^{\text{jet}}_{\text{reco}})} E^{\text{jet}}_{\text{EM}} \tag{5.9}$$

Finally, each entire η bin is fit to $\langle\mathcal{R}^{\text{jet}}\rangle, \langle E^{\text{jet}}_{\text{reco}}\rangle$ to produce the jet calibration function $\mathcal{F}_{\text{calib}}(E^{\text{jet}}_{\text{reco}})$ which is inverted to obtain the correction (see eq. (5.9)) and bring the jets to the **EM+JES scale**. Figure 5.10 shows the corrections done on the jet η derived in a similar fashion, but with the response defined as $\mathcal{R}^{\text{jet}} = \eta^{\text{jet}}_{\text{reco}} - \eta^{\text{jet}}_{\text{truth}}$.

Global Sequential Calibration

Following the previous calibrations and corrections on the origin, pile-up, and jet energy scale; there is still an observed dependence on the longitudinal and transverse features of

[8]We bin in truth-jet energy, rather than reco-jet energy to remove a dependence of the calibration on the reco-jet p_{T} spectrum which includes detector-level effects that almost certainly introduce a bias.

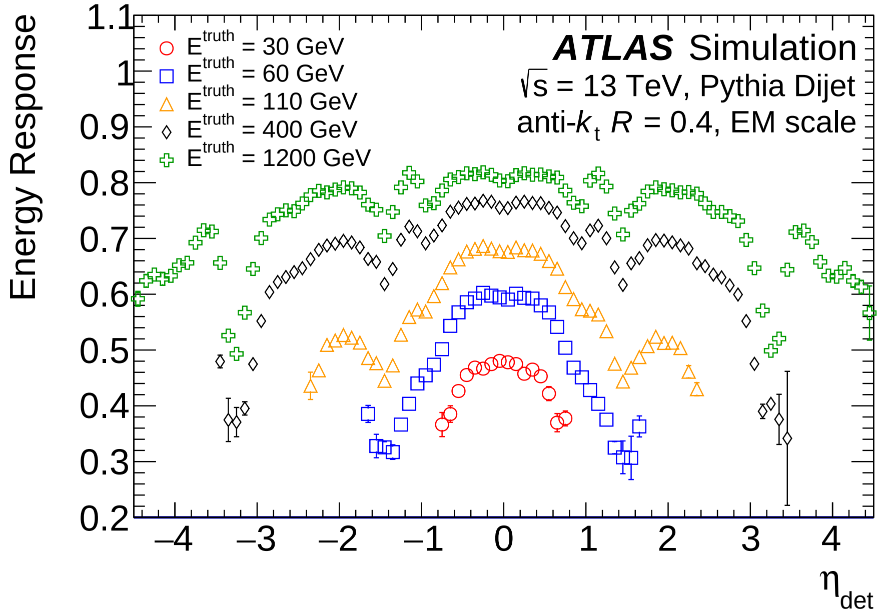


Figure 5.9: [16] The average jet energy response as a function of η_{det} for jets of a truth energy of 30, 60, 110, 400, and 1200 GeV. The energy response is shown after origin and pile-up corrections are applied. This shows the size of the calibration constants for jets built from topoclusters at the EM scale.

the jet. In particular, quarks and gluons shower and hadronize differently which means that quark and gluon jets will have a different response in the detector. As gluons split into pairs of quarks, gluon-initiated jets tend to have a high multiplicity of soft signals. Alternatively, quark-initiated jets will often include hadrons with a higher fraction of the jet p_{T} that penetrate deeper into the calorimeter. As the goal of these jet calibrations is to be independent of the "type" of jet, removing these dependencies is important. This particular procedure, known as Global Sequential Calibration (GSC), was explored during the Run I calibration effort [18] which identified five variables that are used to improve the resolution of the JES. Each of these variables exist in a subset of the detector and corrections are applied sequentially:

1. f_{Tile0}, $|\eta_{\text{det}}| < 1.7$: the fraction of jet energy in the first layer of Tile

2. f_{LAr3}, $|\eta_{\text{det}}| < 3.5$: the fraction of jet energy in the third layer of the EMCal

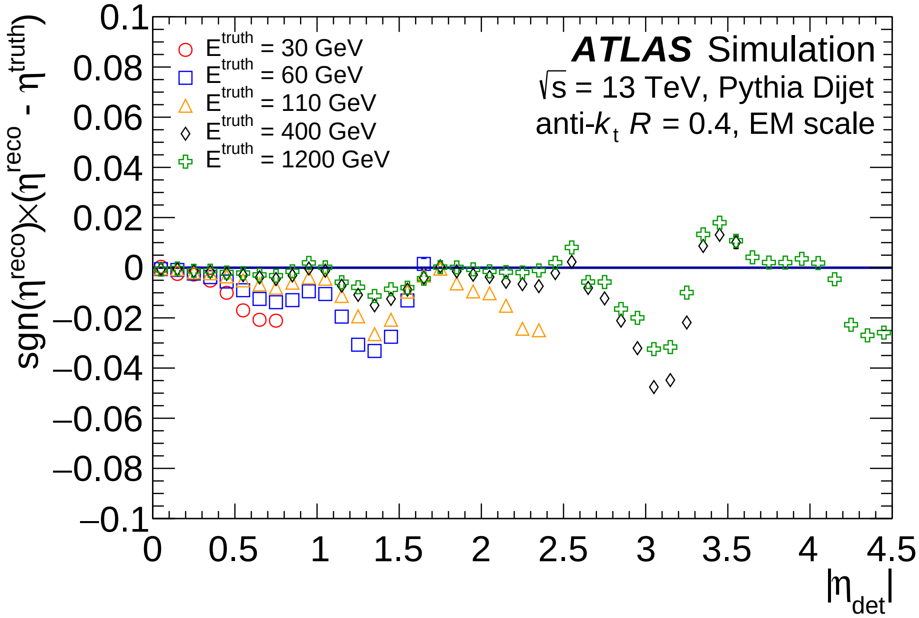


Figure 5.10: [16] The signed difference between truth jet η^{truth} and the reconstructed jet η^{reco} due to biases in the jet reconstruction. This bias is addressed with an η correction applied as a function of $|\eta_{\mathrm{det}}|$. The effect of changing geometries, such as barrel end-cap transition region around $|\eta_{\mathrm{det}} \sim 1.4|$ and endcap-forward transition region around $|\eta_{\mathrm{det}}| \sim 3.1$ can clearly be seen.

3. n_{trk}, $|\eta_{\mathrm{det}}| < 2.5$: the number of tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ associated to the jet

4. $\mathcal{W}_{\mathrm{trk}}$, $|\eta_{\mathrm{det}}| < 2.5$: the width of the tracks associated to the jet, weighted by their p_{T}

5. n_{segments}, $|\eta_{\mathrm{det}}| < 2.7$: the number of muon tracks associated to the jet

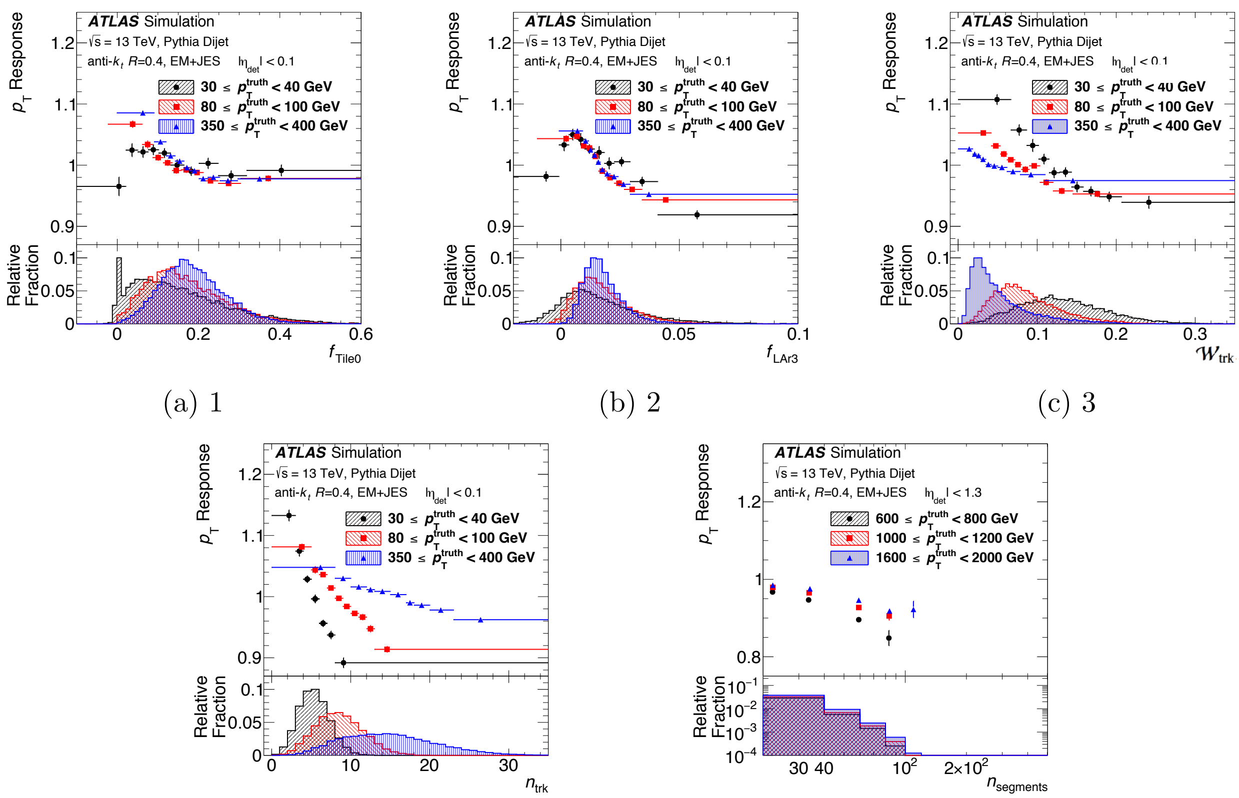
Figure 5.11 shows the distribution of the five GSC variables. The variables used are properties that characterize the logitudinal and transverse topology of the energy deposited by the jet [18]. For example, a large energy deposit in calorimeter layers closest to the interaction point indicates a shower that developed early, leading to a lower detector response in the calorimeters as a fraction of that jet energy would not have reached the calorimeters.

In-Situ Calibration

Following the JES and GSC calibration steps, a data-driven approach, in-situ calibration, is applied to account for differences in jet responses between data and Monte-Carlo simulation. The corrections are designed to correct for the energy scale differences between data and monte-carlo, as monte-carlo is already calibrated at this stage to have the correct energy scale with respect to truth jets. The differences can arise from imperfections in the simulation of the hard scatter event, pile-up, jet formation, and so on. These differences are quantified by a jet balancing approach, where the p_{T} of a jet is balanced against other well-measured reference objects, $\mathcal{R} = p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{object}}$. There are four in-situ calibrations performed:

1. η-intercalibration: correct the response of jets in the forward region $0.8 < |\eta| < 4.9$ using well-measured jets in the central region $|\eta| < 0.8$

2. jet-balance using Z+jet ($Z^0 \to e^+e^-, Z^0 \to \mu^+\mu^-$) with a well-reconstructed Z^0 boson



(a) 1

(b) 2

(c) 3

(d) 4

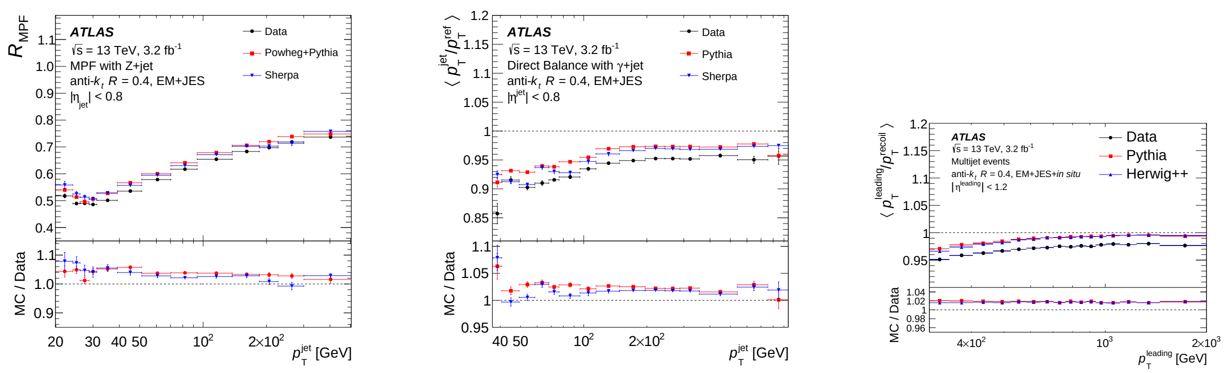
(e) 5

Figure 5.11: [16] The average jet response in Monte-Carlo simulation as a function of the GSC variables for three different ranges of $p_{\mathrm{T}}^{\mathrm{truth}}$. (a) and (b) are shown with no GSC corrections applied. (c) and (d) are shown with the first two corrections applied. (e) is shown with the previous four corrections applied. (a)-(d) are shown for $|\eta_{\mathrm{det}}| < 0.1$ while (e) is shown for $|\eta_{\mathrm{det}}| < 1.3$.

3. jet-balance using γ+jet with a well-reconstructed photon

4. jet-balance using multijets with a high p_{T} jet balanced against a system of low p_{T} jets (multijet balance)

The jet p_{T} response of the three jet-balance in-situ calibrations are shown in fig. 5.12. Each of these techniques are statistically combined, in overlapping regions of jet p_{T}, into a total calibration as shown in fig. 5.13. Each method is weighted by p_{T} bin based on the statistical power, such that a method's weight is increased in p_{T} regions of smaller relative uncertainty and smaller bin size, in order to maximize the precision in each p_{T} region. The inverse of this ratio is taken as the **in-situ correction**.



(a) Z+jet (b) γ+jet (c) multijet

Figure 5.12: [16] The average jet p_{T} response of EM+JES jets calibrated up to the η-intercalibration as a function of jet p_{T} for (a) Z+jet events, (b) γ+jet events, and (c) multijet balance.

5.1.4 Uncertainties

At the end of the day, there are 79[9] JES systematic uncertainties propagated from the individual calibrations and studies described in section 5.1.2 [16].

[9]There are 80 total, but one of them is for a type of simulation not used in this thesis analysis and does not apply.

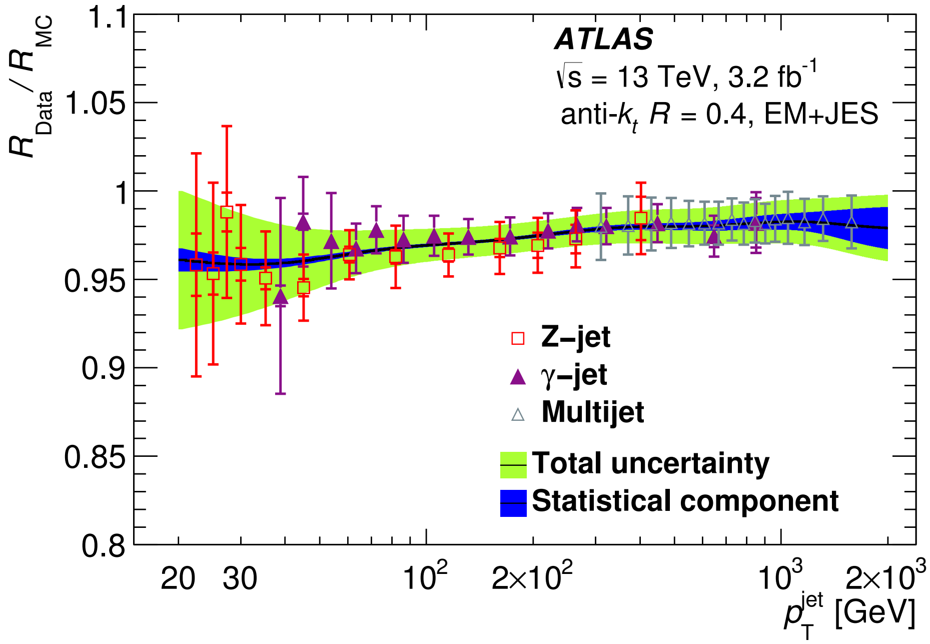
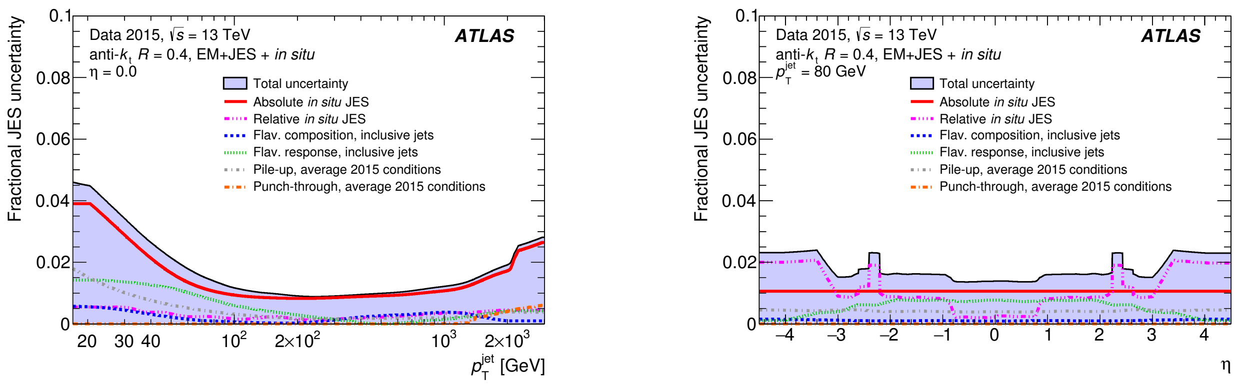


Figure 5.13: [16] The in-situ calibrations for Z+jet, γ+jet, and multijet balance are statistically combined to create a total calibration and uncertainty. The final derived correction (black line) and its statistical uncertainty band (dark blue) and total uncertainty band (light green) are shown.

- 67 come from the jet-balance in-situ calibration procedures in section 5.1.3;
- 3 from the η-intercalibration;
- 4 come from pile-up uncertainties in section 5.1.3;
- 3 come from differences in jet response and compositions of gluons, light-quarks, and b-quark initiated jets
- 1 comes from the GSC punch-through correction;
- and 1 comes from uncertainty on high jet $p_{\mathrm{T}} > 2\,\mathrm{TeV}$ beyond the reach of in-situ methods.

The full combination of uncertainties is shown in fig. 5.14 and is largest at low p_{T} starting at 4.5% decreasing to 1% at 200 GeV. It rises after due to the statistical uncertainties of the in-situ calibrations which end at 2 TeV, hence the sharp increase. In terms of η, the uncertainty is fairly constant across the detector and reaches a maximum of 2.5%[10] for the forward jets.



(a) p_T (b) η

Figure 5.14: [16] Combined uncertainty of JES calibration as a function of (a) jet p_T at $\eta = 0$ and (b) η at $p_T = 80$ GeV.

While the 79 uncertainties provide the most accurate understanding of the JES uncertainty, implementing and evaluating them all is computationally intensive. A reduced set of Nuisance Parameters (NPs) is made available through an eigen-decomposition in a way to preserve the correlations observed with all 79 [16]. Four reduced sets of NPs are formed after a global-reduction procedure and grouping in a way to preserve correlations in low-p_T (`JES2`), medium-p_T (`JES3`), and high-p_T (`JES4`) kinematic regimes, as well as one that provides general representation in all kinematic regimes (`JES1`).

5.1.5 *Jet Kinematics*

This section is meant to provide a brief summary of some of the kinematic properties of $R = 0.4$ anti-k_t jets formed from EM-scale topoclusters with both origin correction (section 5.1.3) and pile-up correction (section 5.1.3) applied to the reconstructed jets. A few representative

[10]There is a feature around $2.0 < |\eta| < 2.6$ due to the non-closure uncertainty of the η-intercalibration.

kinematic distributions of the topoclusters used as jet inputs and the kinematics of the jets themselves are shown, with more in [19]. Figure 5.15 shows the distribution of the number of jets as a function of η_{det} with disagreements observed in the gap region between the barrel and extended barrels of the Tile ($|\eta_{\text{det}}| \sim 1.4$) and the forward region ($3.1 < |\eta| < 4.9$). For example in the gap region, a larger number of high p_{T} jet events are observed in data than what is predicted by Monte-Carlo simulations. Figure 5.16 shows a distribution of the mean number of constituents for each jet binned in jet η. This plot is meant to emphasize the lower granularity in the forward regions and larger topocluster sizes. The primary differences between data and simulation is due to the modeling of the soft-energy components of the jets. And finally, fig. 5.17 shows the very nice modeling of the minimum ΔR between jets for which the bulk of the distribution agrees to better than 10%. This observable is used to define the isolation criteria for the MC-based calibration.

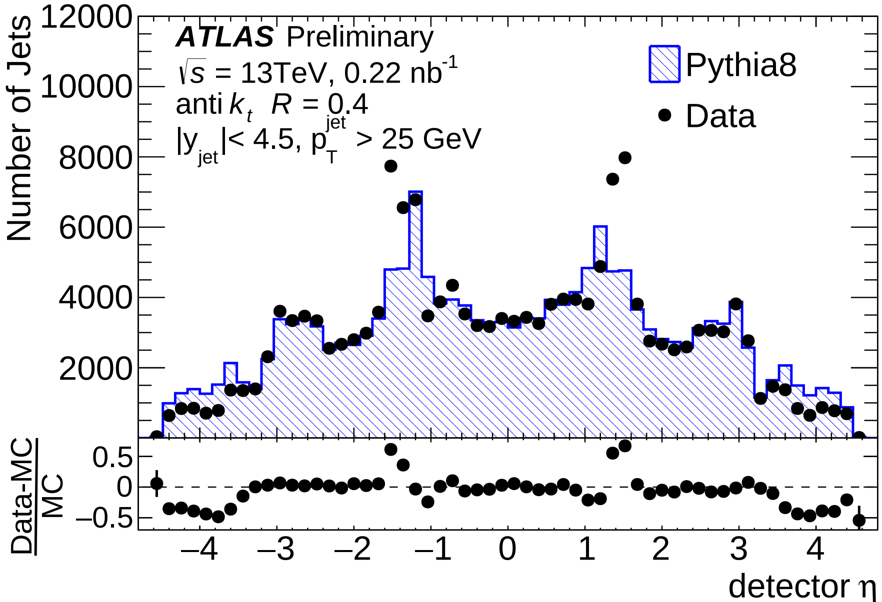


Figure 5.15: [19] Jet kinematic distributions for η_{det} for jets above $p_{\text{T}} > 25\,\text{GeV}$. The Monte-Carlo prediction is normalized to the data and the error s reflect the statistical uncertainty.

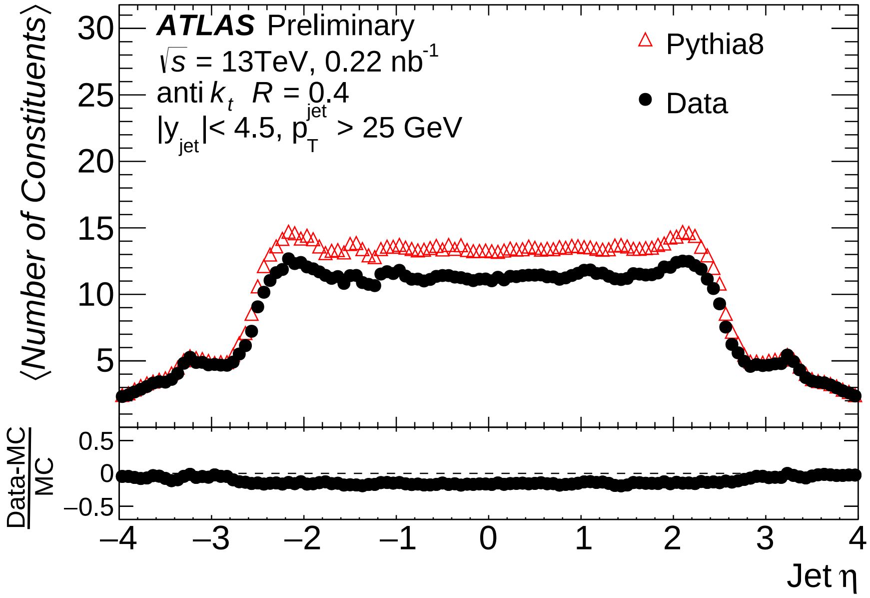


Figure 5.16: [19] Topocluster multiplicity for jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ is shown as a function of jet η. The error bars reflect the statistical uncertainty.

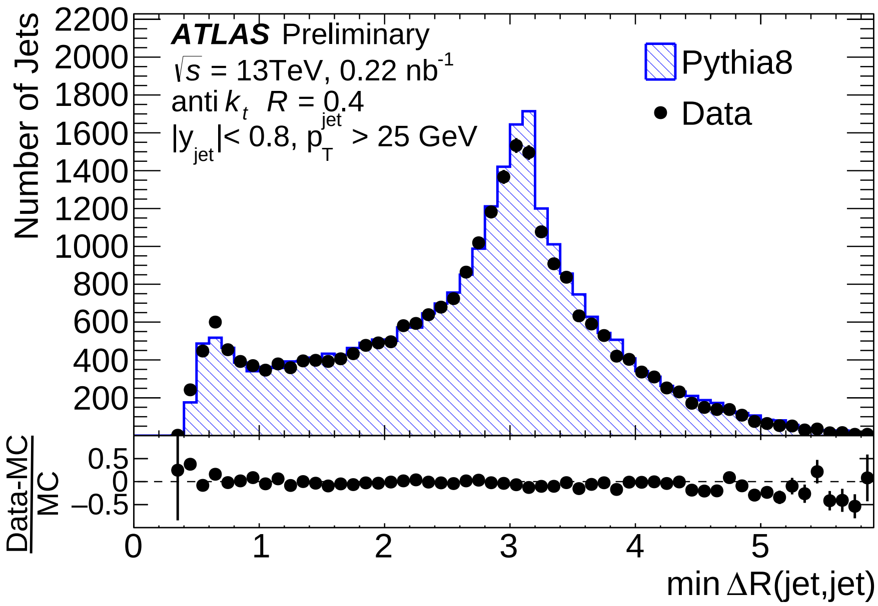


Figure 5.17: [19] The minimum distance in ΔR to a jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|y| < 0.8$. The Monte-Carlo prediction is normalized to the data. Requiring $R = 0.4$ jets to be isolated requires the minimum $\Delta R > 1.0 = 2.5 \times R$.

5.2 Flavour Tagging of Jets

In general, the jet reconstruction algorithms do not identify the type of parton that initiated a given jet. However, there are a few exceptions that rely on the tracking information from the ID. For example, one could roughly identify jets formed from hadronic τ decays since τ is at the mass of hadrons and can decay with one track (e.g. $\tau^- \to W^- \nu_\tau \to \pi^- \pi^0 \nu_\tau$) or three tracks (e.g. $\tau^- \to \pi^+ \pi^- \pi^- \nu_\tau$) [20]. The particular identification we are concerned with though are jets associated with B-hadrons. In particular, these are hadrons that are relatively long-lived[11] and decay primarily via weak interactions. This identification is known as b-tagging and is part of the flavor tagging efforts of the ATLAS collaboration. The most vital input required for b-tagging are the charged particle tracks reconstructed in the ID which has an acceptance $|\eta| < 2.5$. These B-hadrons will typically decay a few mm away from the primary vertex, inside the ID which has a radius of about 1 m. This secondary decay creates a **secondary vertex** and the b-tagging algorithms take advantage of this to identify the displaced tracks of B-hadrons [21, 22, 23, 24]. There are three basic algorithms whose outputs are used as inputs to the standard multi-variate discriminant algorithm (`MV2`) that is used in ATLAS analyses for Run II:

- an impact parameter-based algorithm (`IP2D`, `IP3D`),
- a secondary vertex reconstruction algorithm (`SV`),
- and a decay chain multi-vertex reconstruction algorithm, JetFitter (`JF`).

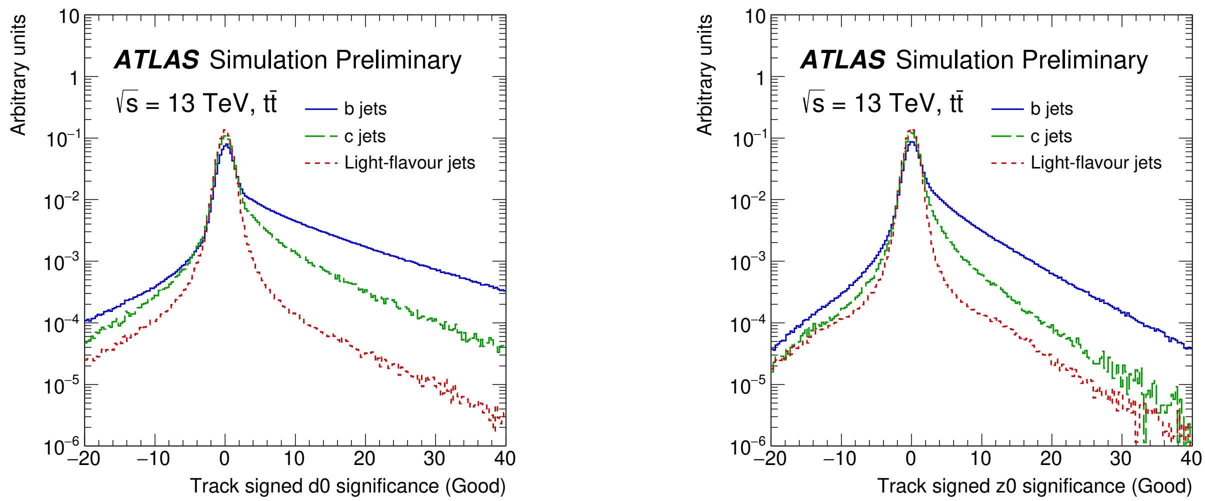
[11] B-hadrons have a lifetime $\sim 1.5\,\mathrm{ps}$ ($c\tau \sim 450\,\mu\mathrm{m}$) compared to top quarks with a mean lifetime $\sim 10^{-25}$s.

5.2.1 Impact Parameter Tagging Algorithms

The typical B-hadron usually has at least one vertex displaced from the primary vertex. To parameterize the minimum distance between the displaced track and the primary vertex, the transverse impact parameter d_0 is defined in the $r - \phi$ plane, while the longitudinal impact parameter $z_0 \sin\theta$ is defined in the logitudinal plane. B-hadrons will typically have large impact parameters due to their relatively long lifetimes. If a secondary vertex is identified as "behind" the primary vertex, the sign of the impact parameter is negative and is normally due to background and to calibrate the light quark tag rate [25]. Two impact parameter significances can be defined for transverse d_0/σ_{d_0} and longitudinal $z_0 \sin\theta/\sigma_{z_0 \sin\theta}$. Figure 5.18 shows distributions of the signed significances for each impact parameter, with well-measured tracks being weighted up and poorly-measured tracks being weighted down. A log-likelihood ratio discriminant shown in fig. 5.19 is computed using the number of tracks of a given jet along with the probability density functions (b-flavor, light-flavor) derived from distributions of the impact parameters. This log-likelihood ratio discriminant is used as input to the multivariate algorithm described later in this section.

5.2.2 Secondary Vertex Finding Algorithm

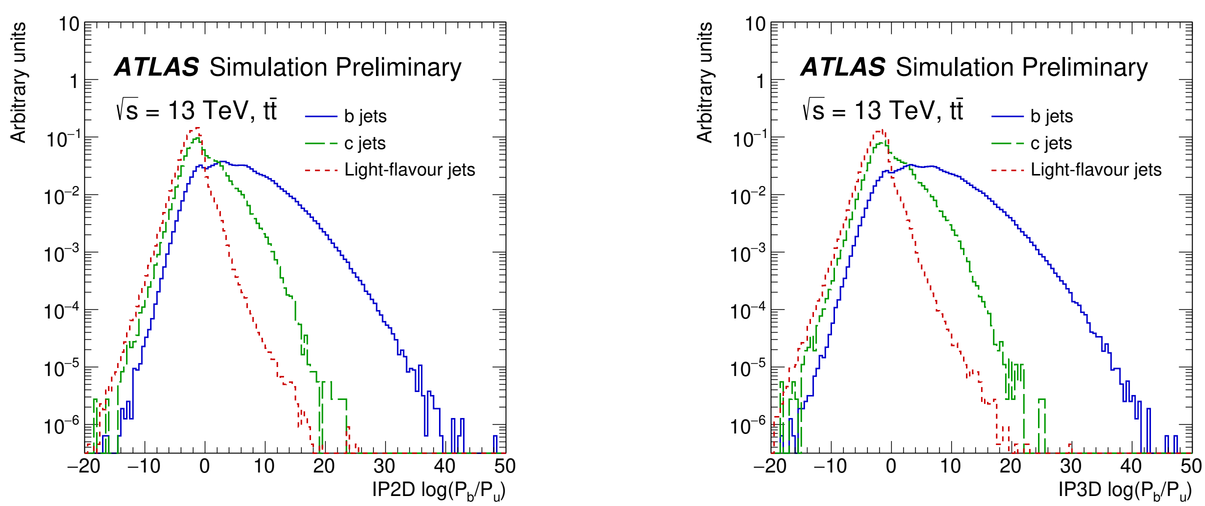
The secondary vertex finding algorithm [26] (`SV`) explicitely reconstructs a displaced secondary vertex within the jet. From all the vertices associated with a jet, these are filtered to only look at those with two tracks. Vertices with a pair of two tracks are rejected if they likely originate from the decay of some non B-hadron or hadronic interactions with the detector material. From the remaining tracks, all possible two-track vertices are formed and required to be signfiicantly displaced from the primary interaction vertex by requiring the sum of the impact parameter significances of the tracks in the two-track vertex to be higher than 2 [26]. Altogether, eight kinematic properties of the two-track vertices and the tracks



(a) transverse

(b) longitudinal

Figure 5.18: [22] The (a) transverse and (b) longitudinal signed impact parameter significance of tracks in $t\bar{t}$ events for b-flavor (blue), c-flavor (green), and light flavour (red) jets. The tail shown in the b jets line (blue) is due to the long lifetime of B-hadrons.

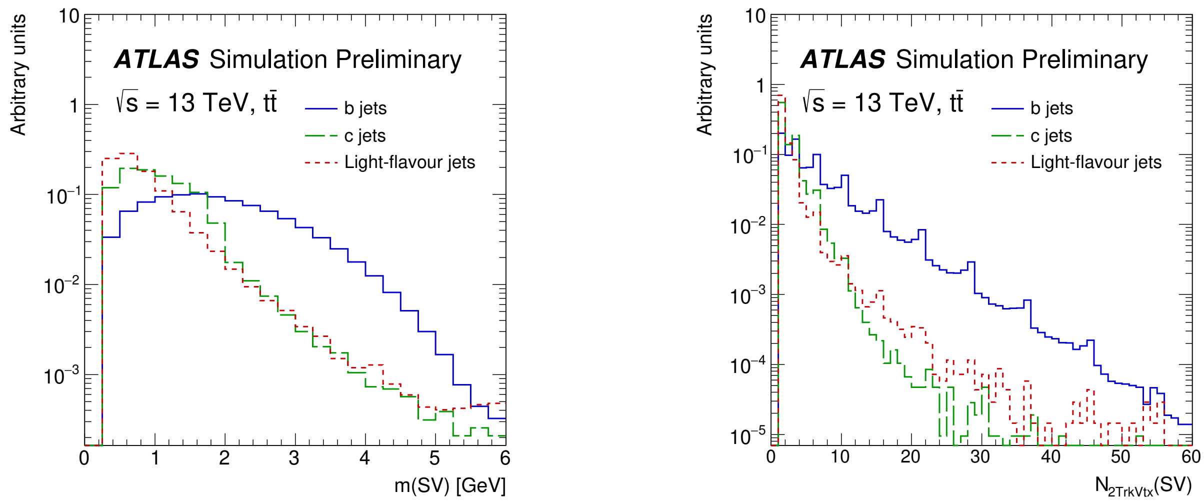


(a) IP2D

(b) IP3D

Figure 5.19: [22] The log-likelihood ratio for the (a) IP2D and (b) IP3D b-tagging algorithms in $t\bar{t}$ events for b-flavor (blue), c-flavor (green), and light flavour (red) jets. The log-likelihood is calculated as the ratio of b-flavor and light-flavor probability density functions. Jets with no tracks are assigned a large negative value in an underflow bin which is not shown on this plot.

at the reconstructed secondary vertex are used as inputs to the multivariate algorithm. Distributions of two of these inputs are shown in fig. 5.20 for the number of two-track vertices identified and the number of tracks at the reconstructed secondary vertex.



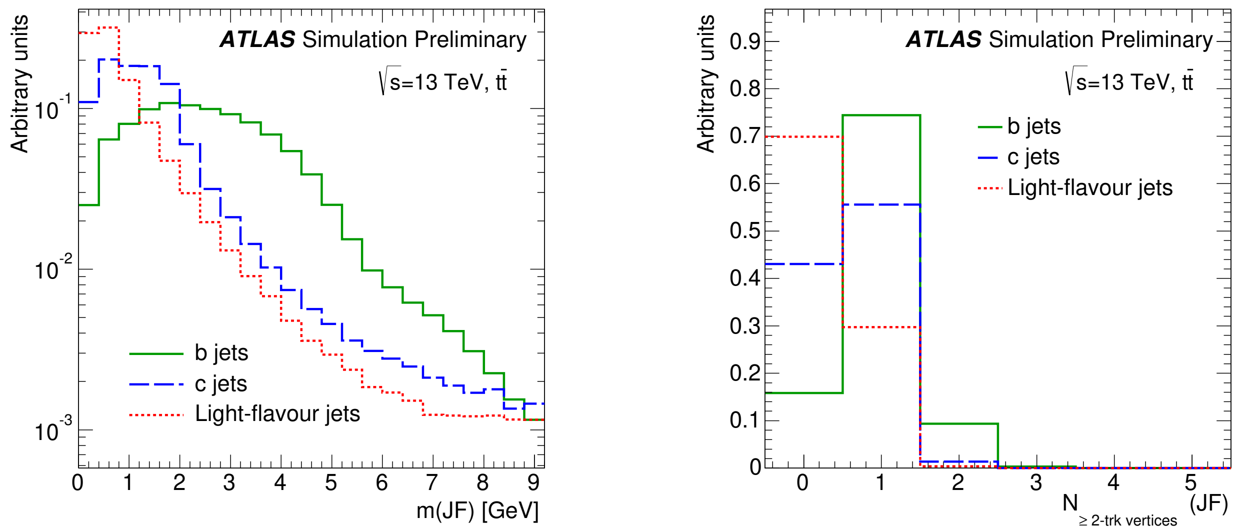
(a) Invariant mass

(b) Track multiplicity

Figure 5.20: [22] Properties of the secondary vertices reconstructed by the SV algorithm for b-flavor (blue), c-flavor (green), and light flavour (red) jets. Shown are (a) the invariant mass of the tracks associated with the reconstructed vertex and (b) the number of two-track vertices reconstructed within the jet.

5.2.3 Decay Chain Multi-Vertex Algorithm

The decay chain multi-vertex reconstruction algorithm, JetFitter [27], uses the topologies of B-hadron and C-hadron decay vertices inside the jet to reconstruct the full B-hadron decay chain. A Kalman filter is used to identify a line along which the B-hadron and C-hadron decay vertices lie along to approximate the B-hadron flight path and identify potential secondary vertices. This algorithm seeks to identify the two tracks of the B-hadron and C-hadron decays. The eight kinematic properties of the reconstructed two-track vertices and tracks at the reconstructed secondary vertex are used as inputs to the multivariate algorithm. Distributions of two of these inputs are shown in fig. 5.21 for the number of two-track vertices identified and the number of tracks at the reconstructed secondary vertex.



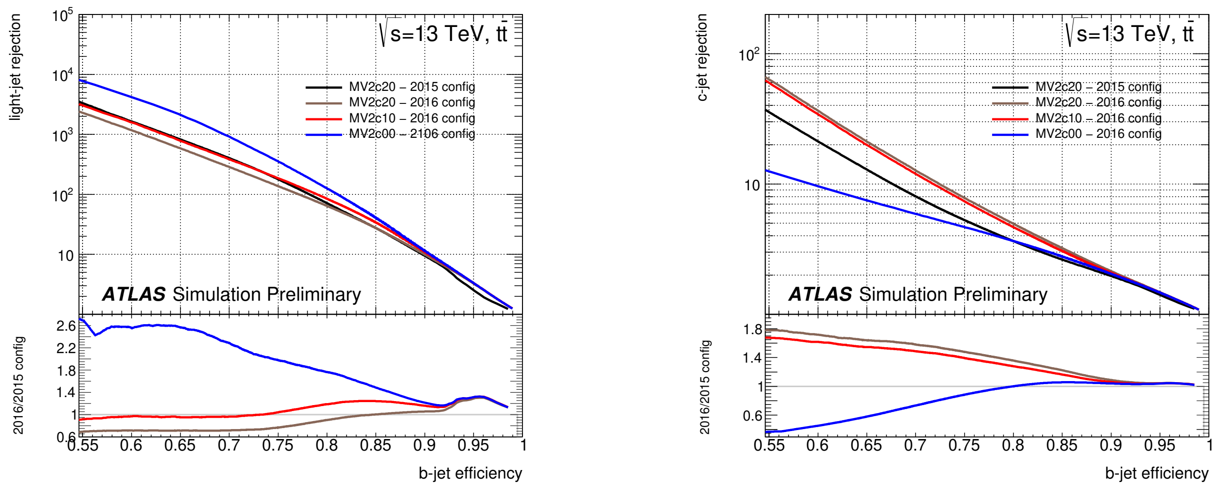
(a) Invariant mass

(b) 2-Track vertex multiplicity

Figure 5.21: [22] Properties of the secondary vertices reconstructed by the JF algorithm for b-flavor (blue), c-flavor (green), and light flavour (red) jets. Shown are (a) the invariant mass of the tracks associated with the reconstructed vertex and (b) the number of two-track vertices reconstructed within the jet.

5.2.4 *Multivariate Algorithm*

The p_{T} and η of the jet, along with the three outputs from IP2D and IP3D, eight outputs from SV and from JF, make up the 24 input variables that go into a multivariate classifier [28], using a Boosted Decision Tree (BDT) algorithm. This achieves a better discrimination than any of the basic algorithms described previously. The tagger, called `MV2`, is trained on jets from $t\bar{t}$ monte-carlo simulation. Three such taggers were developed for ATLAS in Run II called `MV2c00`, `MV2c10`, `MV2c20` with the names indiciating the c-jet fraction in the background. `MV2c10` is the tagger used in this thesis analysis and indicates that the tagger was trained on a sample whose background composition is 10% c-flavor jets and 90% light-flavor jets. Figure 5.22 shows the performance of the optimized MV2 algorithms in rejecting light-flavor jets and c-flavor jets as a function of the b-jet efficiency. It should be noted that while the MV2c20 algorithm does provide better c-flavor jet rejection, it does so at the expense of a reduced light-flavor jet rejection. Because of this consideration and that this thesis analysis is sensitive to light jets, the MV2c10 tagger was chosen as the standard b-tagging discriminant.



(a) light-jet rejection

(b) c-jet rejection

Figure 5.22: [22] (a) Light-flavor jet and (b) c-flavor jet rejection versus b-flavor jet efficiency for the 2015 and 2016 configurations of the MV2 b-tagging algorithm is shown evaluated for $t\bar{t}$ events.

From this, one can define four sets of standard working points shown in table 5.1 for b-tagging identified by picking a b-jet efficiency. Figure 5.23 shows the output of the MV2c10. Operating points are defined by a single cut value on the discriminant output distribution and are chosen to provide a specific b-jet efficiency on an inclusive $t\bar{t}$ sample. For example, the 77% working point has a rejection factor of 6 and of 134 on charm and light-jets, respectively.

Finally, in order for this tagger to be useful for physics analyses, scale factors need to be derived using data to account for differences between simulation and data. For Run I, this was done in [29, 30, 31, 32]. Correction factors are applied to the simulated event samples to compensate for differences between data and simulation in the b-tagging efficiency for b, c and light-jets. The correction for b-jets is derived from $t\bar{t}$ events with final states containing

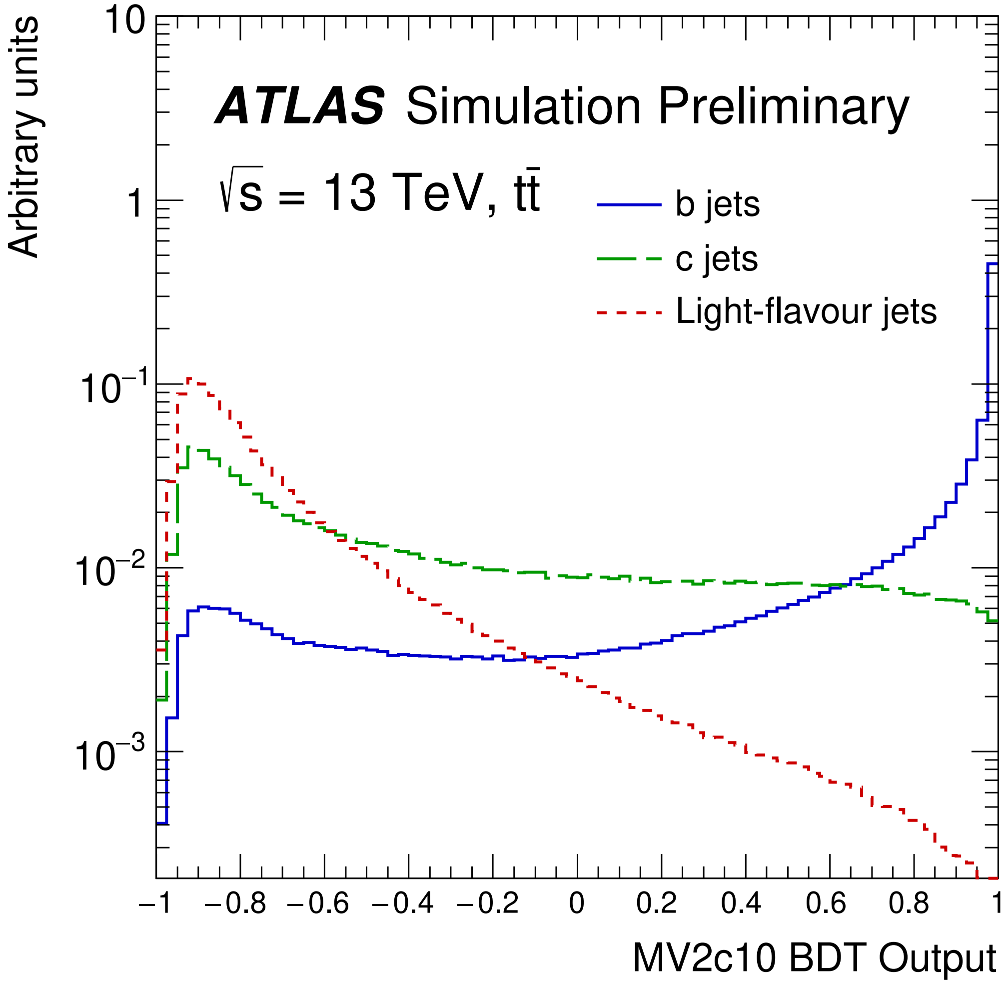


Figure 5.23: [22] The `MV2c10` output for b-flavor (blue), c-flavor (green), and light flavour (red) jets evaluated with $t\bar{t}$ events.

	Efficiency [%]	Rejection [%]		
Cut Value	b-jet	c-jet	light-jet	τ
0.9349	60	34	1538	184
0.8244	70	12	381	55
0.6459	77	6	134	22
0.1758	85	3.1	33	8.2

Table 5.1: [22] Operating points for the `MV2c10` b-tagging algorithm, including benchmark numbers for the efficiency and rejection rates.

two leptons, and the corrections are consistent with unity with uncertainties at the level of a few percent over most of the jet p_{T} range. An example of these correction factors for the Run-I algorithm[12] are shown in fig. 5.24 along with a total systematic uncertainty that needs to be considered in this thesis analysis which is very sensitive to the b-jet tagging algorithm due to the large number of expected b-jets in the final state.

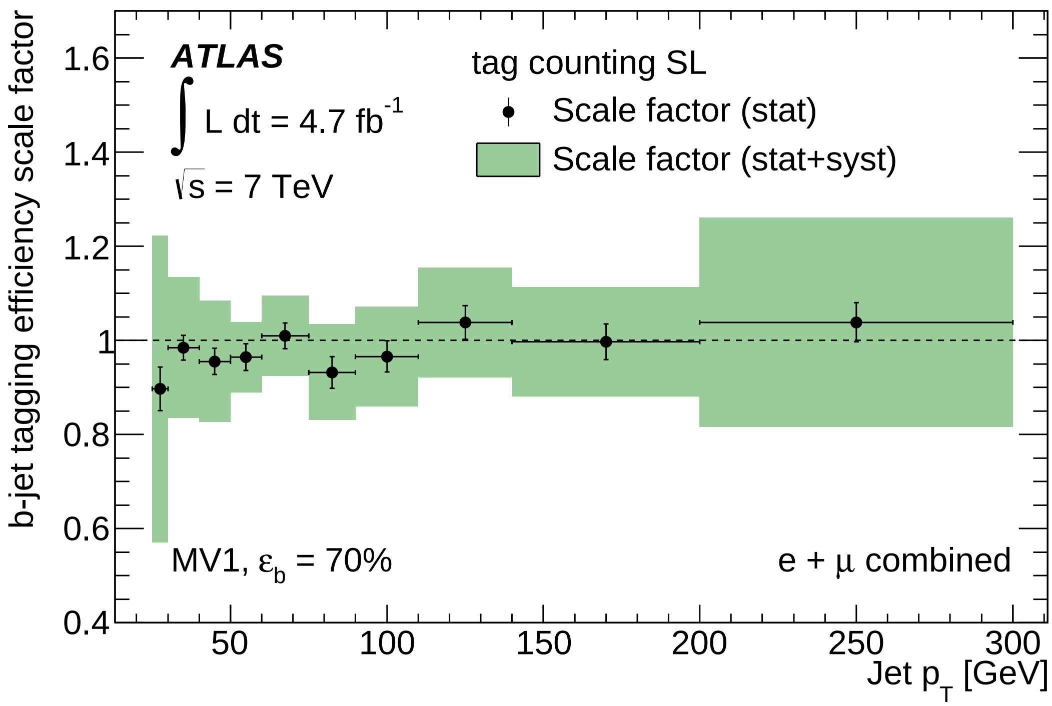


Figure 5.24: [33] The data/simulation scale factors for the `MV1` algorithm at 70% b-jet tagging efficiency. The error bars show the statistical uncertainties while the green band indicates the total uncertainty. Please note that the algorithm here is different from the `MV2` algorithm used in this thesis analysis. The performance studies for the current algorithm are still being done.

5.3 Muons

Muons are one of the simplest particles to identify in the ATLAS detector. As muons traverse the entire detector, reconstructed tracks from both the ID and the muon spectrometer (MS)

[12]The Run-2 algorithm performance is still undergoing study and will not be public in time for this thesis.

are used. Four different muon types are defined depending on which subdetectors are used in the reconstruction in order of decreasing priority:

1. Combined (CB) muon: tracks are reconstructed independently in the ID and MS, and a combined track fit is performed by adding or removing tracks from the MS to improve the fit quality.

2. Segment-tagged (ST) muons: a track in the ID is classified as a muon if it associated with a track segment in the Monitored Drift Tubes (MDT) or the Cathode Strip Chambers (CSC). This is primarily used for low p_{T} muons that don't traverse the entire MS.

3. Calorimeter-tagged (CT) muons: a track in the ID is classified as a muon if it is associated with an energy deposit in the calorimeter compatible with a low-ionizing particle.

4. Extrapolated (ME) muons: the muon trajectory is reconstructed based only on the MS track in at least two layers and ensuring that it originates from the interaction point. This is mainly used to extend the acceptane for muon reconstruction in the region outside the ID from $2.5 < |\eta| < 2.7$.

From the muon-classified tracks, **muon quality** requirements are placed on tracks from each portion of the subdetector which amount to requiring a specific number of hits in each subcomponent [34]. Four muon quality identificates are used:

medium Default selection for muons in ATLAS. This is the quality criteria used in the thesis analysis for identifying muons. Only CB and ME tracks are used with at least 3 CB track hits and at least 3 ME layers.

loose Designed to maximize reconstruction efficiency, primarily for reconstructing Higgs boson candidates in the four-lepton final state [35]. All muon types are used.

tight Designed to maximize the purity of muons. Only CB muons with hits in at least two layers of the MS and satisfying the "medium" selection are used.

high-p_{T} Designed to maximize the momentum resolution for tracks with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$, primarily for high-mass W' and Z' resonances [36, 37]. CB muons passing the "medium" selection and having at least three layers of the MS are selected.

Muons are further calibrated to data using the well-studied resonances $J/\Psi \to \mu\mu$ and $Z \to \mu\mu$. Figure 5.25 shows the combined uncertainty in quadrature of this calibration effort as a function of the p_{T} of the reconstructed muons. $J/\Psi \to \mu\mu$ targets the low-p_{T} region with $Z \to \mu\mu$ targeting the high-p_{T} region. The total systematic uncertainty in muon reconstruction is less than 2% across the board.

As muons from the decay of heavy particles such as the W, Z, h are often produced isolated from other particles, a set of **muon isolation** requirements are also placed around each muon candidate. Two primary isolation-based variables are used, one is track-based ($p_{\mathrm{T}}^{\mathrm{varcone30}}$) and one is calorimeter-based ($E_{\mathrm{T}}^{\mathrm{topocone20}}$). Each isolation criteria looks at the scalar sum of transverse momentum in a cone of $\Delta R < 0.2$ or $\Delta R < 0.3$[13] and compares it to the transverse momentum of the muon candidate. The isolation is effective at suppressing muons produced from processes such as meson decay in flight and heavy-flavor decay. There are seven isolation working points that are defined for use by analyzers which differ primarily on a cut of the ratio between the energy of the muon candidate and the surrounding "background" energy. This isolation requirement is defined to ensure a flat efficiency of around 99% across the whole electron transverse energy and muon transverse momentum ranges. This is described

[13] Look at the name of the isolation variable to know the cone size.

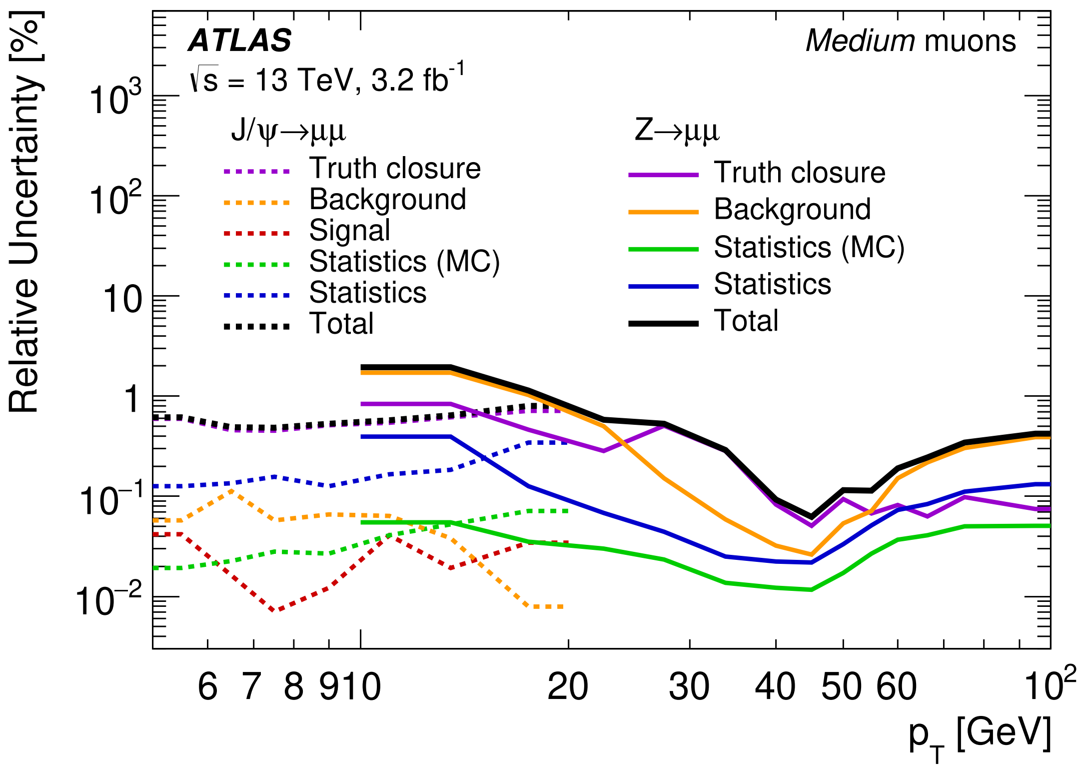


Figure 5.25: [34] Total uncertainty in the efficiency scale factor for "medium" muons as a function of p_{T} as obtained from $Z \to \mu\mu$ (solid lines) and $J/\Psi \to \mu\mu$ (dashed lines) decays. The combined uncertainty is the sum in quadrature of the individual contributions.

in more detail in ??.

5.4 Electrons and Photons

Both electrons and photons are reconstructed based on the electromagnetic shower in the LAr electromagnetic calorimeter described in ??. The electromagnetic calorimeter is divided into a grid of 3×5 towers of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ to scan for areas of local maxima which are used to seed clusters. The clusters are then matched to a well-reconstructed ID track. The existence of this matched track and its properties are used to subsequently identify the cluster as being consistent with a prompt electron, a photon conversion, or an unconverted photon [38, 39]. A photon carries no electric charge and will not produce a track in the ID: label as an unconverted photon. A converted photon is one which has a secondary vertex because of the decay into an electron-positron pair: label the cluster as a converted photon if the matched track is extrapolated from a secondary vertex. Otherwise, an electron would have a matched track that is extrapolated from the primary vertex and so is labeled as a prompt electron.

Similar to muons, electrons have **electron quality** and **electron isolation** identifications. Three levels of identification working points are provided for **electron quality** called `Loose, Medium, Tight` which are selections on the discriminant of a multivariate analysis to provide electron identification using a likelihood-based method. In addition to the identification criteria, isolation criteria are defined using two primary isolation-based variables: one is track-based ($p_{\mathrm{T}}^{\mathrm{varcone20}}$) and one is calorimeter-based ($E_{\mathrm{T}}^{\mathrm{topocone20}}$). Each isolation criteria looks at the scalar sum of transverse momentum in a cone of $\Delta R < 0.2$ or $\Delta R < 0.3$[14] and compares it to the transverse momentum of the electron candidate. The isolation is effective

[14]Look at the name of the isolation variable to know the cone size.

at disentangle prompt electron candidates[15] from other non-isolated electron candidates[16].

Electrons and photons are calibrated in a similar procedure to muons (section 5.3) to derive data-driven scale factors using $J/\Psi \to ee$, $Z \to ee$, and $Z \to \ell\ell\gamma$ processes. These corrections ensure uniformity in the electromagnetic response across separate regions of the detector and will introduce systematic uncertainties including the mismodeling between simulation and data.

5.5 Taus

While tau leptons are not used directly in this thesis or the analysis in this thesis, I would like to devote a short section to explaining their relevance. Tau leptons are charged leptons, but they are very different from electrons and muons. From an experimental point of view, tau leptons decay into other types of particles before entering the detector. They can decay hadronically around 60% of the time into hadrons plus neutrinos and 40% of the time leptonically to electrons or muons plus neutrinos. The leptonic decays are not distinguishable from electrons and muons described in sections 5.3 and 5.4; the hadronic decays present as multiple hadronic showers matched to tracks in the ID. As taus can present with a secondary vertex, they can fake b-tagged jets and this is described more previously in section 5.2.

5.6 Missing Transverse Momentum

The last crucial object is the missing transverse momentum which represents the overall transverse momentum imbalance in the event, commonly written out as $E_{\mathrm{T}}^{\mathrm{miss}}$ or colloquially

[15] Prompt electron candidates come from heavy-resonance decays such as $W \to e\nu_e, Z \to ee$.

[16] Non-isolated candidates include electrons from photon-conversion, from heavy-flavor hadron decays, and light hadrons mis-identified as electrons.

"MET"[17]. Conservation of momentum in the plane transverse to the beam axis implies that the vector transverse momenta of the collision products should sum to zero. An imbalance imlpies the existence of weakly-interacting stable particles, such as neutrinos in the SM or many supersymmetric particles such as neutralinos in the BSM. The measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$ is also affected by poorly-reconstructed objects, visible particles that escape the detector unseen, or particles that otherwise fail to be reconstructed. While the $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed offline, it can and often is reconstructed again at an analysis-level with extra refinement such as the precise removal of objects that overlap or to specify the inputs of all visible particles identified in an analysis. For this thesis analysis, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed again and the procedure is described in more detail in **??**. In general, $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined in eq. (5.10)

$$-E_{\mathrm{T}}^{\mathrm{miss}} = \sum_{e} p_{\mathrm{T}} + \sum_{\gamma} p_{\mathrm{T}} + \sum_{\tau} p_{\mathrm{T}} + \sum_{\mu} p_{\mathrm{T}} + \sum_{j} p_{\mathrm{T}} + \sum_{\mathrm{soft}} p_{\mathrm{T}}. \tag{5.10}$$

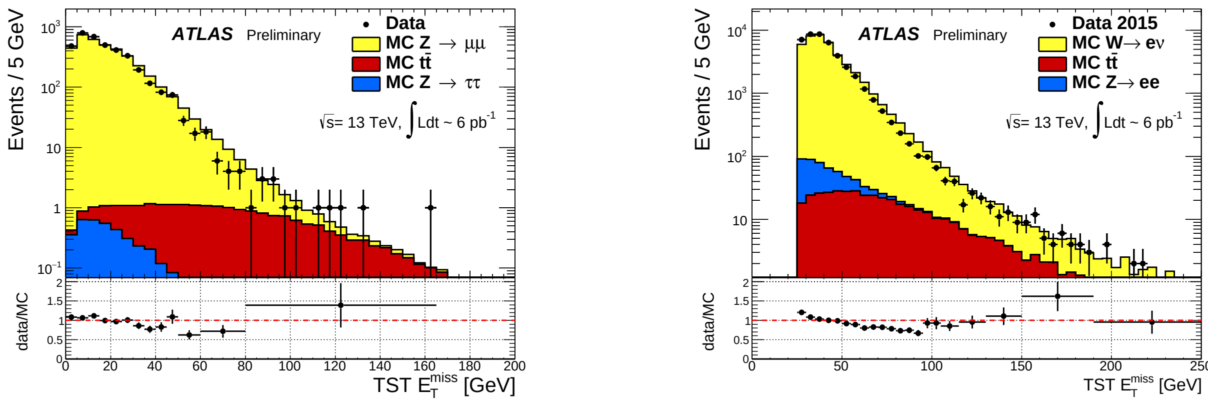
As described in [40, 41], a baseline set of selection critera are applied to all the visible, reconstructed objects that enter the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, such as quality criteria on the leptons and p_{T} cuts on the objects. However, one special term is the "soft" term which comes in two different forms: track and cluster. Track soft-terms (TST) are ID tracks, extrapolated from the primary vertex, which are not within $\Delta R < 0.05$ of an electron or photon, $\Delta R < 0.2$ of a tau, or matched with a combined muon (section 5.3) or jet. Similarly, Cluster soft-terms (CST) are built from energy depositions in the calorimeter not associated with reconstructed physics objects. By construction[18], this is sensitive to pile-up which makes it a less suitable choice for high-luminosity environments. For the thesis analysis, as described in **??**, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed using track soft-terms.

The performance of $E_{\mathrm{T}}^{\mathrm{miss}}$ shown in fig. 5.26 is studied by comparing to $Z \to \mu\mu$ and

[17]Not to be confused with the New York Mets.

[18]A pun!

$W \rightarrow e\nu_e$ processes [41]. $Z \rightarrow \mu\mu$ has no real $E_{\mathrm{T}}^{\mathrm{miss}}$ and is a good choice to study the performance of $E_{\mathrm{T}}^{\mathrm{miss}}$ due to the precise measurements of the kinematics of the Z boson. Likewise, $W \rightarrow e\nu_e$ is provides genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ through the neutrinos from the hard-scatter interaction and helps validate the scale and direction of reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ [41]. Systematics from all the included input objects are propagated through the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. The soft-term has systematics associated with the data/simulation scale-factor that is derived using comparisons with $Z \rightarrow \ell\ell$.



(a) $Z \rightarrow \mu\mu$

(b) $W \rightarrow e\nu_e$

Figure 5.26: [41] The distribution of reconstructed track soft-term $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown for (a) $Z \rightarrow \mu\mu$ topologies and (b) $W \rightarrow e\nu_e$ topologies. The agreement between data and simulation for $W \rightarrow e\nu_e$ is notably worse with respect to $Z \rightarrow \mu\mu$ in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ region likely due to the missing QCD multijet background not included in the studies.

Glossary

ATLAS a general-purpose detector at the LHC. 1, 3, 7–9, 13, 16, 26, 31, 33, 35

BC Bunch Crossing. 2

BSM Beyond the Standard Model. 1, 39

CSC Cathode Strip Chambers. 33

EMCal electromagnetic calorimeter. 2, 9, 19

FCal forward calorimeter. 9, 10

GSC Global Sequential Calibration. 17, 19, 20

HCal hadronic calorimeter. 2

HEC LAr Hadronic End-Cap Calorimeter. 9

ID Inner Detector. 4, 8, 24, 26, 33, 35, 37, 39, 40

JES jet energy scale. 15, 17, 19, 22, 23

LAr Liquid Argon Calorimeter. 4, 9, 10, 37

LHC Large Hadron Collider. 1

MDT Monitored Drift Tubes. 33

MS muon spectrometer. 33, 35, 36

NP Nuisance Parameter. 23

punch-through For jets at very high transverse momentum it is possible that part of the energy is not deposited in the calorimeter, but leaks out to the detector components beyond the calorimeter. This leads to a systematic reduction in the measured jet energy. Jets that deposit energy beyond the hadronic Tile calorimeter and in the muon system are called punch-through jets. [18]. 13

QCD Quantum Chromodynamics. A theory describing the strong interactions of SM particles.. 1, 16, 41

SM Standard Model. 1, 39

SUSY Supersymmetry. 2

Tile Tile calorimeter. 4, 9, 19, 24

Bibliography

[1] Pavel Weber. "ATLAS calorimetry: Trigger, simulation and jet calibration". PhD thesis. Kirchhoff Inst. Phys., 2008. URL: http://www.ub.uni-heidelberg.de/archiv/8170/ (cit. on p. 3).

[2] ATLAS Collaboration. "Search for Supersymmetry in final states with missing transverse momentum and multiple b-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: (2017). arXiv: 1711.01901 [hep-ex] (cit. on p. 4).

[3] John E. Huth et al. "Toward a standardization of jet definitions". In: *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990*. 1990, pp. 0134–136. URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-249 (cit. on p. 2).

[4] Gavin P. Salam. "Towards Jetography". In: *Eur. Phys. J.* C67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph] (cit. on pp. 5, 6).

[5] Jessie Shelton. "Jet Substructure". In: *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012): Boulder, Colorado, June 4-29, 2012*. 2013, pp. 303–340. DOI: 10.1142/9789814525220_0007. arXiv: 1302.0260 [hep-ph]. URL: http://inspirehep.net/record/1217434/files/arXiv:1302.0260.pdf (cit. on p. 5).

[6] Bora Isildak. "Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ tev". PhD thesis. Bogazici U., 2011. arXiv: 1308.6064 [hep-ex]. URL: http://inspirehep.net/record/1251416/files/arXiv:1308.6064.pdf (cit. on p. 6).

[7] S. Catani et al. "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions". In: *Nucl. Phys.* B406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M (cit. on p. 7).

[8] Yuri L. Dokshitzer et al. "Better jet clustering algorithms". In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph] (cit. on p. 7).

[9] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph] (cit. on pp. 7, 8).

[10] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph] (cit. on p. 7).

[11] Matteo Cacciari and Gavin P. Salam. "Dispelling the N^3 myth for the k_t jet-finder". In: *Phys. Lett.* B641 (2006), pp. 57–61. DOI: 10.1016/j.physletb.2006.08.037. arXiv: hep-ph/0512210 [hep-ph] (cit. on p. 7).

[12] Peter Loch et al. *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run I.* Tech. rep. ATL-COM-PHYS-2014-1439. Geneva: CERN, Nov. 2014. URL: https://cds.cern.ch/record/1967028 (cit. on pp. 8, 9).

[13] ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex] (cit. on pp. 9, 10).

[14] ATLAS Collaboration. *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV.* ATL-PHYS-PUB-2015-015. 2015. URL: https://cds.cern.ch/record/2037613 (cit. on p. 11).

[15] ATLAS Collaboration. “Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 75 (2015), p. 17. DOI: 10.1140/epjc/s10052-014-3190-y. arXiv: 1406.0076 [hep-ex] (cit. on p. 11).

[16] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 96 (2017), p. 072002. DOI: 10.1103/PhysRevD.96.072002. arXiv: 1703.09665 [hep-ex] (cit. on pp. 13, 15, 17, 18, 20–23).

[17] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. In: *Phys. Lett.* B659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph] (cit. on p. 14).

[18] ATLAS Collaboration. “Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 73 (2013), p. 2304. DOI: 10.1140/epjc/s10052-013-2304-2. arXiv: 1112.6426 [hep-ex] (cit. on pp. 17, 19, 42).

[19] ATLAS Collaboration. *Properties of jets and inputs to jet reconstruction and calibration with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*. ATL-PHYS-PUB-2015-036. 2015. URL: https://cds.cern.ch/record/2044564 (cit. on pp. 24, 25).

[20] C. Patrignani et al. “Review of Particle Physics”. In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001 (cit. on p. 26).

[21] ATLAS Collaboration. *Expected performance of the ATLAS b-tagging algorithms in Run-2*. ATL-PHYS-PUB-2015-022. 2015. URL: https://cds.cern.ch/record/2037697 (cit. on p. 26).

[22] ATLAS Collaboration. *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*. ATL-PHYS-PUB-2016-012. URL: https://cds.cern.ch/record/2160731 (cit. on pp. 26, 28–32).

[23] ATLAS Collaboration. *Proton tagging with the one arm AFP detector*. ATL-PHYS-PUB-2017-012. 2017. URL: https://cds.cern.ch/record/2265187 (cit. on p. 26).

[24] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017. URL: https://cds.cern.ch/record/2273281 (cit. on p. 26).

[25] ATLAS Collaboration. "Performance of b -jet identification in the ATLAS experiment". In: *Journal of Instrumentation* 11.04 (2016), P04008. URL: http://stacks.iop.org/1748-0221/11/i=04/a=P04008 (cit. on p. 27).

[26] ATLAS Collaboration. *Commissioning of the ATLAS high performance b-tagging algorithms in the* 7 *TeV collision data*. ATLAS-CONF-2011-102. 2011. URL: https://cds.cern.ch/record/1369219 (cit. on p. 27).

[27] Giacinto Piacquadio and Christian Weiser. "A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS". In: *J. Phys. Conf. Ser.* 119 (2008), p. 032032. DOI: 10.1088/1742-6596/119/3/032032 (cit. on p. 29).

[28] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039 [PHYSICS] (cit. on p. 30).

[29] ATLAS Collaboration. *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with* 5 fb^{-1} *of data from the ATLAS detector*. ATLAS-CONF-2012-043. 2012. URL: https://cds.cern.ch/record/1435197 (cit. on p. 31).

[30] ATLAS Collaboration. *Measurement of the Mistag Rate of b-tagging algorithms with* 5 fb^{-1} *of Data Collected by the ATLAS Detector*. ATLAS-CONF-2012-040. 2012. URL: https://cds.cern.ch/record/1435194 (cit. on p. 31).

[31] ATLAS Collaboration. *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*. ATLAS-CONF-2014-004. 2014. URL: https://cds.cern.ch/record/1664335 (cit. on p. 31).

[32] ATLAS Collaboration. *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*. ATLAS-CONF-2014-046. 2014. URL: https://cds.cern.ch/record/1741020 (cit. on p. 31).

[33] ATLAS Collaboration. "Performance of b-jet identification in the ATLAS experiment". In: *JINST* 11 (2016), P04008. DOI: 10.1088/1748-0221/11/04/P04008. arXiv: 1512.01094 [hep-ex] (cit. on p. 33).

[34] ATLAS Collaboration. "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 76 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex] (cit. on pp. 34, 36).

[35] ATLAS Collaboration. "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev. D* 91 (2015), p. 012006. DOI: 10.1103/PhysRevD.91.012006. arXiv: 1408.5191 [hep-ex] (cit. on p. 35).

[36] ATLAS Collaboration. "Search for high-mass dilepton resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Phys. Rev. D* 90 (2014), p. 052005. DOI: 10.1103/PhysRevD.90.052005. arXiv: 1405.4123 [hep-ex] (cit. on p. 35).

[37] ATLAS Collaboration. "Search for new particles in events with one lepton and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *JHEP* 09 (2014), p. 037. DOI: 10.1007/JHEP09(2014)037. arXiv: 1407.7494 [hep-ex] (cit. on p. 35).

[38] ATLAS Collaboration. "Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data". In: *Eur. Phys. J. C* 74 (2014), p. 3071. DOI: 10.1140/epjc/s10052-014-3071-4. arXiv: 1407.5063 [hep-ex] (cit. on p. 37).

[39] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*. ATLAS-CONF-2016-024. 2016. URL: https://cds.cern.ch/record/2157687 (cit. on p. 37).

[40] ATLAS Collaboration. *Expected performance of missing transverse momentum reconstruction for the ATLAS detector at* $\sqrt{s} = 13$ *TeV*. ATL-PHYS-PUB-2015-023. 2015. URL: https://cds.cern.ch/record/2037700 (cit. on p. 39).

[41] ATLAS Collaboration. *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at* $\sqrt{s} = 13$ *TeV*. ATL-PHYS-PUB-2015-027. 2015. URL: https://cds.cern.ch/record/2037904 (cit. on pp. 39, 40).