

# TABLE OF CONTENTS

	Page
<b>1 Physics Building Blocks and their Reconstruction</b>	<b>1</b>
1.1 Charged-Particle Tracks and Primary Vertices . . . . .	2
1.2 Electrons and Muons . . . . .	4
1.2.1 Electrons . . . . .	4
1.2.2 Muons . . . . .	9
1.3 Jets . . . . .	12
1.3.1 Jet Reconstruction . . . . .	13
1.3.2 Jet Calibration . . . . .	16
1.3.3 Flavor Tagging of Jets . . . . .	21
1.4 The Missing Transverse Momentum . . . . .	27
1.5 Object-level Ambiguity Resolution . . . . .	27
<b>Bibliography</b>	<b>28</b>

# Chapter 1

## Physics Building Blocks and their Reconstruction

*If you can put your five fingers through it it is a gate, if not a door.*

–Stephen Dedalus, in James Joyce’s *Ulysses*

In order to convert the multitude of electrical signals read out by the subdetectors of ATLAS as a result of a successful trigger (c.f. Section ??) into well-defined and meaningful representations of the underlying physics process that initiated them, at the level required for performing high-quality physics analysis, several steps of reconstruction and identification must take place. The physics analyses presented in the current work involve the use of leptons, jets, and the so-called missing transverse momentum,  $\mathbf{p}_T^{\text{miss}}$ . The methods used to deduce the presence of these objects within the ATLAS detector will be discussed in this chapter. Section 1.1 introduces the reconstruction of charged-particle tracks and  $pp$  interaction vertices within the ID, both of which are used as low-level seeds or inputs to the reconstruction of the high-level physics objects to be discussed in the subsequent sections. Section 1.2 goes on to discuss the reconstruction of the charged leptons relevant to the current work: electrons and muons. Sections 1.3 and 1.3.3 describe the reconstruction of jet objects and the identification of jets arising from the decay of heavy-flavor hadrons, respectively. Section 1.4 then goes on to describe the reconstruction of  $\mathbf{p}_T^{\text{miss}}$ , which relies on an accurate description of leptons and jets. The methods used for reconstructing the leptons and jets are not one hundred percent accurate: detector information arising due to an electron may leave signatures similar to those of a jet, for example, and thus spoil their unambiguous description. Where relevant, in the following we will discuss the methods by which the reconstruction and identification of the physics objects is made more precise and how high levels of confidence about their actual presence within the detector are achieved. Section 1.5 will also introduce the notion of high-level object ambiguity resolution through the use of so-called *overlap removal* procedures.

## 1.1 Charged-Particle Tracks and Primary Vertices

The reconstruction of charged-particle tracks (‘tracking’) and primary interaction vertices (‘vertexing’) is based on information provided by the ID, primarily by the pixel and SCT subdetectors [1, 2, 3, 4, 5]. Charged-particles produced in  $pp$  collisions will leave signals — *hits* — on the different layers of the ID. The aim of tracking is to translate these layer hits into *spacepoints* which are then combined to form a track following the particle’s traversal through the ID. Given its highly granular readout, the pixel detector provides three dimensional spacepoints from each layer hit while the back-to-back readout strips on each layer of the SCT must be combined, using the stereo-angle information from the second set of strips, to give three dimensional spacepoint information. The hit information provided by the TRT straws is two-dimensional in nature, providing only  $r - \phi$  information in the barrel section and  $\phi - z$  information in the end-caps.

Within the solenoidal magnetic field of the ID, charged-particle tracks follow helical trajectories in the plane transverse to the beam-pipe ( $xy$ -plane) and can be fully characterised by five *track (perigee) parameters*:

$$(d_0, z_0, \phi, \theta, q/p), \quad (1.1)$$

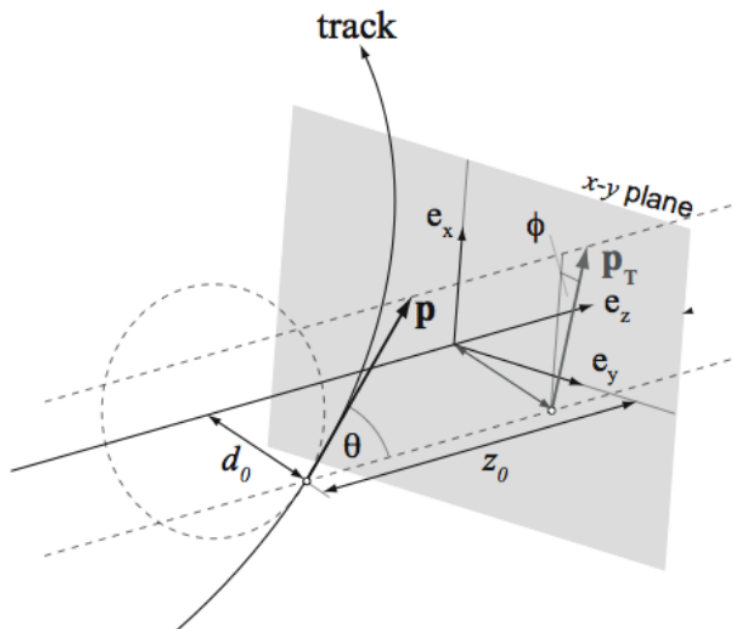
where  $d_0$  ( $z_0$ ) is the transverse (longitudinal) impact parameter,  $\phi$  and  $\theta$  are the azimuthal and polar coordinate, respectively, of the track at the point at which  $d_0$  and  $z_0$  are defined,  $q/p$  is the ratio of the particle charge to the magnitude of its momentum. The charge of a track is determined by its curvature within the magnetic field. The track parameters are defined with respect to their associated primary vertex, whose reconstruction will be described shortly. An illustration describing the track parameters is provided by Figure 1.1.

The primary track reconstruction algorithm used in ATLAS follows an *inside-out* pattern recognition procedure and first starts with information provided by track *seeds*, composed of a few spacepoints, in the silicon detectors which then are extended outwards into the TRT [1]. The inside-out approach accounts for the majority of tracks reconstructed in ATLAS but it is complemented by an *outside-in* approach that starts with the TRT hits and moves inwards [1]. This latter approach is useful in recovering those tracks with ambiguous or missing inner-layer pixel hits; for example, in the case of photon conversions or long-lived neutral particle decays.

The collection of reconstructed tracks is used as input to the primary vertex reconstruction. Primary vertex reconstruction follows a so-called *adaptive vertex fitting* (AVF) [3, 5] procedure and occurs in two steps: primary vertex finding, in which tracks are associated to a particular vertex candidate, and vertex fitting, which involves the reconstruction of the actual vertex position and its errors. After the vertex fitting stage, the tracks associated with a given vertex are refit with the constraint of the vertex position and its errors. The track refitting can update the track parameters (Eqn. 1.1) associated with the tracks. Only vertices with at least two charged particle tracks with  $p_T > 400$  MeV are considered.

In the high luminosity collisions at the LHC there will generally be multiple primary vertices

The presence of so-called *secondary*, *tertiary*, and so on..., vertices are also important and will be described in Section 1.3.3.



3

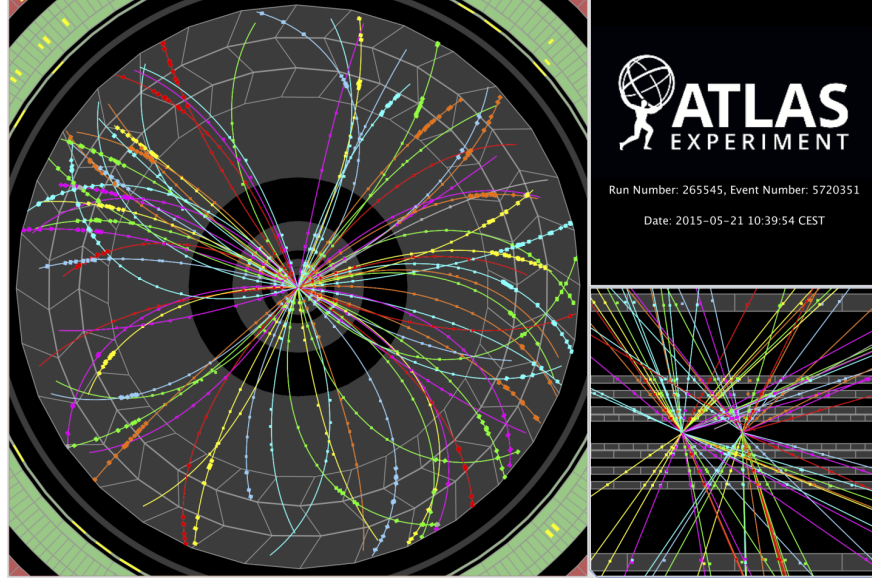


Figure 1.2: Event display of a low-pileup event recorded at the start of Run-II, in early 2015. *Left:* Transverse view of the ID. Seen in color are the reconstructed tracks traversing the inner layers of the pixel detector, SCT, and TRT. The colored dots are all reconstructed spacepoints used as input to the track fitting procedure. *Right, lower:* View in  $r - z$  of the same  $pp$  bunch-crossing event as on the left. Two reconstructed primary vertices are clearly observed. On average, in Run-II there were roughly 30 primary vertices reconstructed per event, with up to  $\approx 65$  occurring at maximum.

## 1.2 Electrons and Muons

Electrons and muons, being charged particles, leave identifiable tracks within the ID. As a result, their reconstruction involves the use of the tracks and vertices described in the previous section, using them essentially as initial seeds for their complete reconstruction. Electron reconstruction, described in Section 1.2.1, complements the track information provided by the ID with calorimetric information provided by the EM calorimeter (Section ??) and with knowledge about the pattern of transition radiation expected to occur in the TRT as a result of passing electrons. Muon reconstruction, described in Section 1.2.2, revolves around stitching together the tracks reconstructed in the ID with those tracks independently reconstructed in the MS layers at large radii.

### 1.2.1 Electrons

#### Electron Reconstruction

After 2016 they replaced sliding window algorithm with supercluster-based reco

The reconstruction of electron candidates is based on three components which characterise the signature of electrons: localised clusters of energy deposits found within the EM calorimeter, charged-particle tracks identified in the ID, close-matching (in  $(\eta, \phi)$ ) of the tracks to the clusters that form the final electron candidates [6]. It is generally possible to match multiple tracks to

the same electromagnetic cluster, all originating from the same primary electron produced in the hard-scatter. This is due to the fact that electrons lose significant amounts of energy to bremsstrahlung photons as they interact with and traverse the ID. These radiated photons can then undergo conversion to electron-positron pairs, which, too, can undergo further bremsstrahlung. The positrons, electrons, and photons are usually emitted in a very collimated fashion and thus deposit most of their energy in a localised fashion within the calorimeter.

The search for localised energy deposits in the EM calorimeter is performed by following a sliding window algorithm over the individual cells whose dimensions are defined by the second sampling layer of the EM calorimeter (Figure ??). Electron candidates are seeded by localised energy deposits whose summed transverse energy, across all layers of the EM calorimeter, is greater than 2.5 GeV [6]. These clusters act as seeds for the matching of reconstructed ID tracks. The reconstructed tracks are refit using a Gaussian Sum Filter (GSF) method [7] that accurately accounts for the bremsstrahlung energy losses characteristic of electron traversal and are then matched to the localised clusters using the cluster barycenter as the point of reference to match in  $\eta - \phi$ . If there is no GSF-track candidate matching to the EM calorimeter cluster seed, then the cluster is marked as an unconverted photon. The cluster is marked as a converted photon if a matched GSF-track candidate exists but is not associated with the primary hard-scatter vertex.

## Electron Identification

Once electron candidates are reconstructed, they are selected based on various levels of identification. A further set of identification criteria is required on top of the reconstruction so as to improve the selection of true electrons originating from the primary hard-scatter vertex — so-called *prompt* electrons — over *non-prompt* sources of reconstructed electrons such as those originating from photon conversions or the misidentification of charged pions that leave electron-like tracks in the ID. This identification criteria is based on the construction of a multivariate likelihood (LH) and is referred to as the *electron likelihood identification*. The inputs to the LH are listed in Table 1.1 and include measurements from the tracking system in the ID, calorimetric information, and quantities that combine the tracking and calorimetric information [6].

The electron LH is based on the products for the signal and background probability density functions (PDFs) associated with the set of inputs in Table 1.1:

$$L_{S(B)}(\mathbf{x}) = \prod_{i=1}^n P_{S(B),i}(x_i), \quad (1.2)$$

where  $\mathbf{x}$  is the vector of quantities listed in Table 1.1 and the  $P_{S(B),i}(x_i)$  are the values of the PDF for quantity  $i$  at value  $x_i$  for the signal ( $S$ ) and background ( $B$ ). The likelihoods are built using simulation and the signal is composed of samples of prompt electrons and the background is built from a combination of jets that mimic the signature of prompt electrons, electrons from photon conversions, and non-prompt electrons from the decay of hadrons containing heavy-flavours [6].

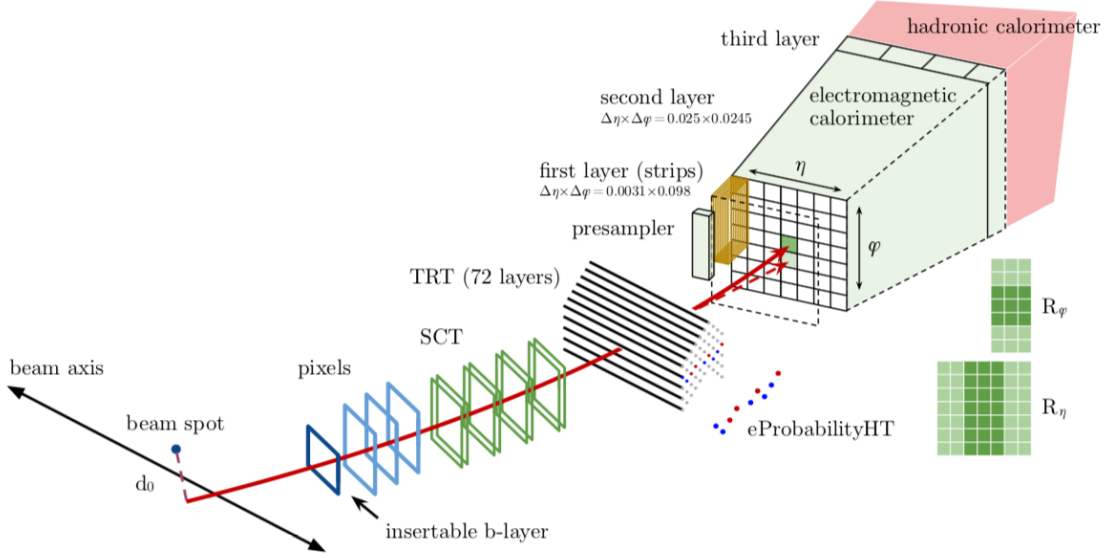


Figure 1.3

The final electron LH discriminant, shown in Figure 1.4, is based on a transformed version of the ratio,

$$d_L = \frac{L_S}{L_S + L_B}, \quad (1.3)$$

where the transformation acts to spread  $d_L$  to values not bounded by 0 and 1, motivated by the need to have well-defined working points based on selections on  $d_L$ .

There are four such fixed values of the final LH discriminant that are used to define four working points corresponding to increasing thresholds on the final LH discriminant: VERYLOOSE, LOOSE, MEDIUM, and TIGHT. The efficiencies to identify prompt electron candidates are measured using samples of  $Z \rightarrow ee$  and  $J/\psi \rightarrow ee$  following a tag-and-probe approach. They are found, for electron candidates with  $E_T > 40 \text{ GeV}$ , to be 93%, 88%, and 80% for the LOOSE, MEDIUM, and TIGHT working points, respectively [6].

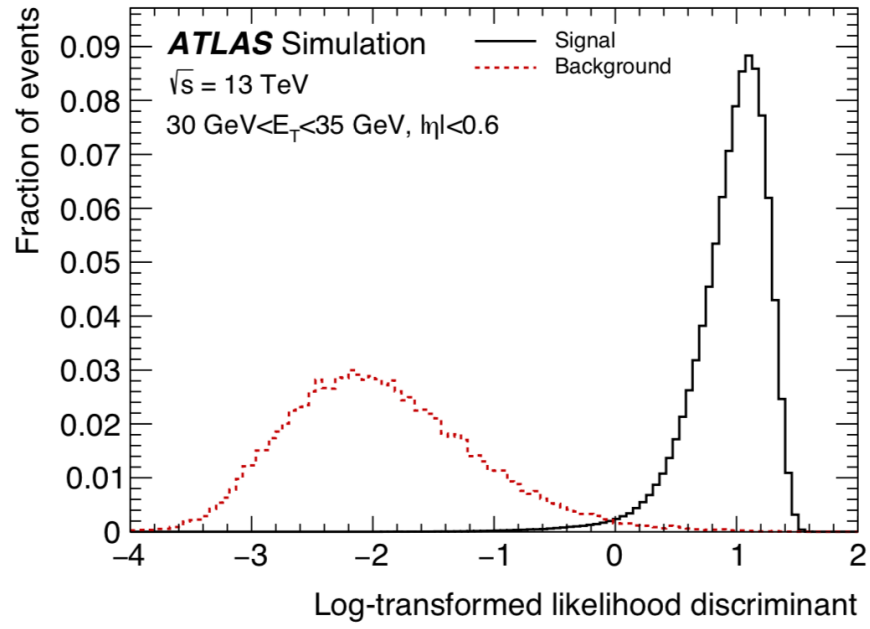


Figure 1.4: Transformed LH-based electron identification discriminant for electron candidates with  $30 \text{ GeV} < E_T < 35 \text{ GeV}$  and  $|\eta| < 0.6$ . From Ref. [6].



Table 1.1: From Ref. [6].

Input Type	Name	Description
Hadronic Leakage	$R_{\text{had1}}$	Ratio of $E_T$ in the first layer of the hadronic calorimeter to $E_T$ of the EM cluster
	$R_{\text{had}}$	Ratio of $E_T$ in the hadronic calorimeter to $E_T$ of the EM cluster
Third layer of EM calorimeter	$f_3$	Ratio of the energy in the third layer to the total energy in the EM calorimeter. Only used for $E_T < 30$ GeV and $ \eta  \leq 2.37$ .
Second layer of EM calorimeter	$w_{\eta 2}$	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$ , where $E_i$ is the energy and $\eta_i$ is the pseudorapidity of cell $i$ and the sum is calculated within a window of $3 \times 5$ cells centered at the electron cluster position.
	$R_\phi$	Ratio of the energy in $3 \times 3$ cells over the energy in $3 \times 7$ cells centered at the electron cluster position.
	$R_\eta$	Ratio of the energy in $3 \times 7$ cells over the energy in $7 \times 7$ cells centered at the electron cluster position.
First layer of EM calorimeter	$w_{\text{tot}}$	Shower width, $\sqrt{(\sum E_i (i - i_{\text{max}})^2)/(\sum E_i)}$ , where $i$ runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$ , corresponding typically to 20 strips in $\eta$ , and $i_{\text{max}}$ is the index of the highest-energy strip. Used only for $E_T > 150$ GeV.
	$E_{\text{ratio}}$	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies.
	$f_1$	Ratio of the energy in the first layer to the total energy in the EM calorimeter.
Track conditions	$n_{\text{Blayer}}$	Number of hits in the innermost pixel layer.
	$n_{\text{Pixel}}$	Number of hits in the pixel detector.
	$n_{\text{Si}}$	Total number of hits in the pixel and SCT detectors.
	$d_0$	Transverse impact parameter relative to the beam-spot.
	$ d_0/\sigma(d_0) $	Significance of transverse impact parameter defined as the ratio of $d_0$ to its uncertainty.
	$\Delta p/p$	Momentum lost by the track between the perigee and the last measurement point divided by the momentum at perigee.
TRT	eProbabilityHT	Likelihood probability based on transition radiation in the TRT.
Track-cluster matching	$\Delta\eta_1$	$\Delta\eta$ between the cluster position in the first layer and the extrapolated track.
	$\Delta\phi_{\text{res}}$	$\Delta\phi$ between the cluster position in the second layer of the EM calorimeter and the momentum-rescaled track, extrapolated from the perigee, times the charge $q$ .
	$E/p$	Ratio of the cluster energy to the track momentum. Used for $E_T > 150$ GeV.

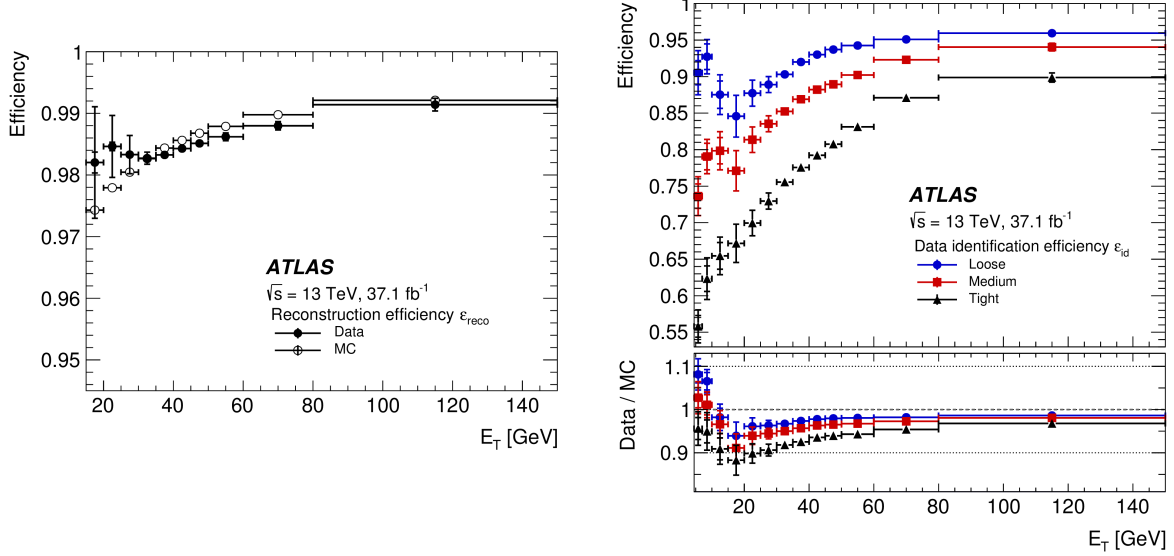


Figure 1.5: **2015-2016** *Left*: Electron candidate reconstruction efficiency, measured in simulation and in data, as a function of the candidate  $E_T$ . *Right*: Electron identification efficiency measured in data, as a function of electron  $E_T$ , for the three standard LH identification working points LOOSE (blue), MEDIUM (red), and TIGHT (black). The lower panel shows the ratio of the efficiency measured in data over that measured in simulation.

## 1.2.2 Muons

### Muon Reconstruction

The reconstruction of muon candidates is performed by combining the tracking capabilities of the ID and the MS [8]. Muon reconstruction first starts with the independent reconstruction of charged-particle tracks within the ID and the MS. The independently formed tracks are subsequently combined to form a complete track representing the traversal of a muon through the full detector. The muon track in the ID is reconstructed like any other charged-particle track (Section 1.1).

Muon reconstruction within the MS starts with a pattern finding phase, looking for hit patterns each of the muon chambers to then form track segments. The track segments between different MS layers are then fit together to form muon track candidates. At least two matching track segments are required in order to build a muon track candidate, except in the transition region between the barrel and end-cap where single track-segment candidates can be used. Once a muon track candidate is formed from the combined segments, a global  $\chi^2$  fit is performed to improve the association of hits to each muon candidate. The  $\chi^2$  is repeated several times, removing outlying hits as necessary, until a threshold is met for all associated hits.

There are several algorithms used to combine the muon track candidates in the ID and MS, each using different sets of information related to the ID, MS, and calorimeters. At the time of the current work, there are four standard combination algorithms used each based on the subdetectors

used in their construction:

- **Combined Muon (CB)** This type of muon is formed with a global refit using all muon track candidate hits in the ID and the MS. Hits may be added or removed from the MS track candidate during the refit. Muons are reconstructed following an outside-in pattern recognition algorithm, in which the muon is first reconstructed in the MS and extrapolated inwards to the ID hits. A complementary, albeit non-standard, inside-out algorithm also exists.
- **Segment-tagged Muon (ST)** An ID track is classified as a muon if, once extrapolated to the MS, it is associated with at least one local track segment in the MDT or CSC chambers. Segment-tagged muons are used when a muon candidate crosses only one layer of the MS chambers, either because of their low  $p_T$  or because they fall into un-instrumented regions of the MS.
- **Calorimeter-tagged Muon (CT)** An ID track is classified as a muon if it is matched to an energy deposit in the calorimeter that is compatible with a minimum ionising particle. Calorimeter-tagged muons have the lowest purity, but recover acceptance in regions of the MS that are only partially instrumented to allow for cabling and services to the calorimeter and ID systems, particularly in the region  $|\eta| < 0.1$ .
- **Extrapolated Muon (ME)** This type of muon is based only on the track candidates formed in the MS and a loose requirement that the track candidate be pointing back towards the IP. Extrapolated muons are mainly used to extend the acceptance of muon reconstruction into the region  $2.5 < |\eta| < 2.7$  that is not covered by the ID acceptance.

## Muon Identification

Muon identification refers to the act of applying additional quality criteria on the reconstructed muon candidates in order to mainly suppress contamination from background sources that mimic muon signatures, such as pion and kaon decays, while ensuring high rates for the acceptance of prompt muons. There are three standard muon identification working points in ATLAS, each a subset of the previous one, and are referred to as the LOOSE, MEDIUM, and TIGHT muon identification working points. MEDIUM muons are the default in ATLAS analyses and can only be composed of CB and ME muons. As all muons used in the present thesis are MEDIUM muons, only this identification working point will be described in detail.

Reconstructed muon candidates originating from non-prompt sources such as in-flight decays of charged hadrons in the ID, are often characterised by the presence of a ‘kink’ in the reconstructed muon’s track. It is therefore expected that the independent momentum measurements made in the ID and MS may be incompatible for non-prompt sources of muon candidates. The muon identification criteria, then, make use of quantities that relate the ID and MS muon track candidates. These quantities are described in Table 1.2. MEDIUM muons have a rather loose selection on the

compatibility between the ID and MS momentum measurements and, with respect to those quantities in Table 1.2, are only required to have a  $q/p$  significance less than 7. On top of requirements on those quantities described in Table 1.2, the muon identification working points place additional requirements on the number and type of hits in the ID and MS. All identification working points require, in the ID, that there be at least 1 hit in the pixel subdetector, at least 5 hits in the SCT subdetector, less than 3 silicon holes,<sup>1</sup> and at least 10% of the TRT hits originally assigned to the muon track candidate exist after the combined reconstruction. MEDIUM muons further require that the CB muons have at least 3 hits in at least two MDT layers, except in  $|\eta| < 0.1$  where tracks with at least one MDT layer but no more than one MDT hole are allowed. The ME Medium muons are required to have at least 3 MDT or CSC layer hits, and are employed only in  $2.5 < |\eta| < 2.7$ . [reference n MDT/CSC hits figure?](#)

Quantity Name	Measurement	Description
$q/p$ significance	$ (q/p)^{\text{ID}} - (q/p)^{\text{MS}} /\sqrt{\sigma_{p_T}^{\text{MS}} + \sigma_{p_T}^{\text{ID}}}$	Absolute value of the difference between the ratio of the charge and momentum of the muon candidates measured in the ID and MS, divided by the quadrature sum of the corresponding uncertainties.
$\rho'$	$ p_T^{\text{MS}} - p_T^{\text{ID}} /p_T^{\text{Combined}}$	Absolute value of the difference between the transverse momentum measurements in the ID and the MS, divided by that of the combined muon candidate.
$\chi_{\text{norm}}^2$	—	Normalized $\chi^2$ of the combined muon track fit

Table 1.2

<sup>1</sup>A missing hit is considered a ‘hole’ only if it falls between hits successfully assigned to a given track.

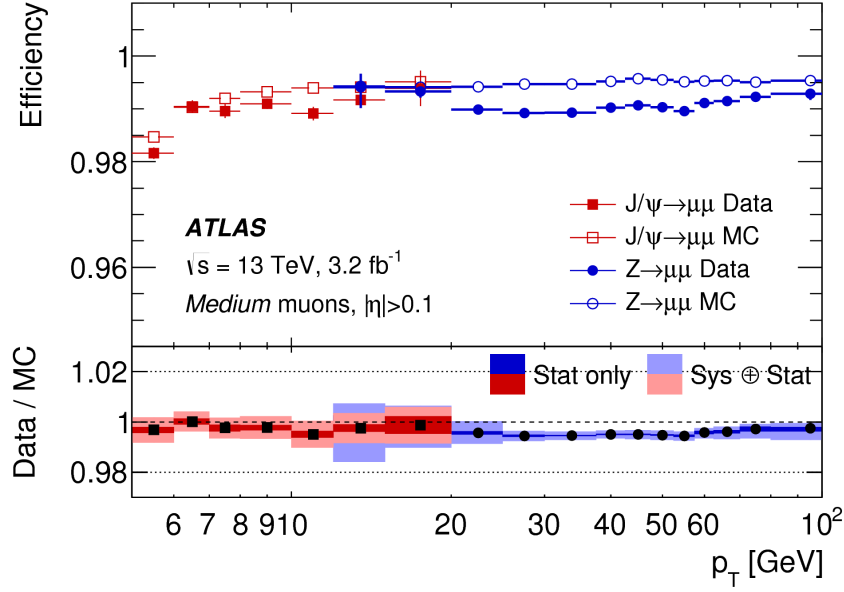


Figure 1.6: From Ref. [8].

### 1.3 Jets

Due to the confining nature of QCD, color-charged quarks and gluons produced in the initial  $pp$  interactions do not exist as free states for observably meaningful timescales and therefore do not leave unambiguous signatures in the detector. Instead, their production is characterised by the radiation of additional quarks and gluons roughly collinear with the initiating colored particles. The radiation pattern of these colored objects is dictated by the color field that binds them and eventually results in the production of color-neutral hadrons. The collimated spray of hadrons as a result of this *hadronisation* process leads to the phenomenology of *jets*, which are the macroscopically observable signature of produced quarks and gluons. The reconstruction of jets refers to any suitable, i.e. physically meaningful and stable, method for grouping together, or *clustering*, the end-products of the hadronisation process in such a way that the properties of the initiating quarks or gluons, such as their quantum numbers and/or kinematics, can be inferred from the resulting clustered object. The standard method for reconstructing jets in ATLAS will be introduced in Section 1.3.1. In Section 1.3.2, the steps taken to turn these reconstructed jets into accurate representations of the initiating quarks and/or gluons, so that they can be used meaningfully in physics analyses, will be discussed.

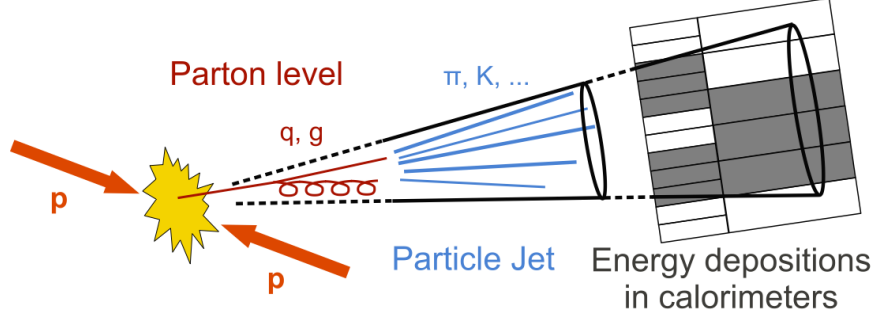


Figure 1.7: Illustration of the jet formation process, beginning with the initiating quark and/or gluons (partons) which hadronise to form particle jets discernible by the tracking detectors in the ID and calorimeter jets defined by energy depositions in the calorimeter systems.

### 1.3.1 Jet Reconstruction

#### Topological Cell Clustering

The process of jet reconstruction begins first with the clustering of the lowest level calorimeter elements, *calorimeter cells*, corresponding to the readout channels in the LAr and tile calorimeters. Figure 1.8 gives an idea of the calorimeter cell granularity across the calorimeter system. The clustering algorithm used by ATLAS is a three-dimensional *topological clustering* algorithm [9, 10]. The highly granular calorimeter system used in ATLAS, with its finely segmented lateral readout and longitudinal sampling layers, allows for the subsequent topological clusters (‘topo-clusters’) to capture in detail the energy-flow details of jets. Topo-cluster formation begins by first identifying so-called *seed cells* which have a rather high signal-to-noise ratio ( $S/N$ ),  $S/N > 4$ . Here, the signal is defined as the absolute value of the calorimeter-cell energy measurement,  $|E|$ , and the noise is defined as the sum in quadrature of the RMS of the electronics and expected pileup noise contributions. Cells neighboring the seed cells satisfying  $S/N > 2$  are then collected into the topo-cluster. A neighboring cell is defined in three-dimensions as either the calorimeter cells directly adjacent within the same calorimeter layer as the seed cell, or, if in adjacent layers or in different calorimeter sub-systems, cells having at least partial overlap in the  $(\eta, \phi)$  plane with the seed. The final set of cells, the perimeter cells, satisfying  $S/N \geq 0$  are then collected. This last threshold essentially collects all those cells surrounding already-collected cells within each layer. Figure 1.9 illustrates the concept of topological cell clustering as described here.

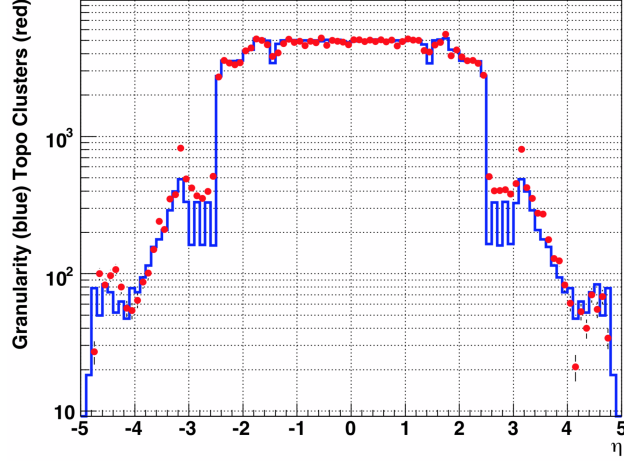


Figure 1.8: The blue histogram shows the average calorimeter cell granularity, i.e. number of calorimeter cells per  $\Delta\eta = 0.1$ , as a function of detector  $\eta$ . The red points show an approximation of the blue histogram based on calculations of the expected noise per calorimeter cell. From Ref. [9].

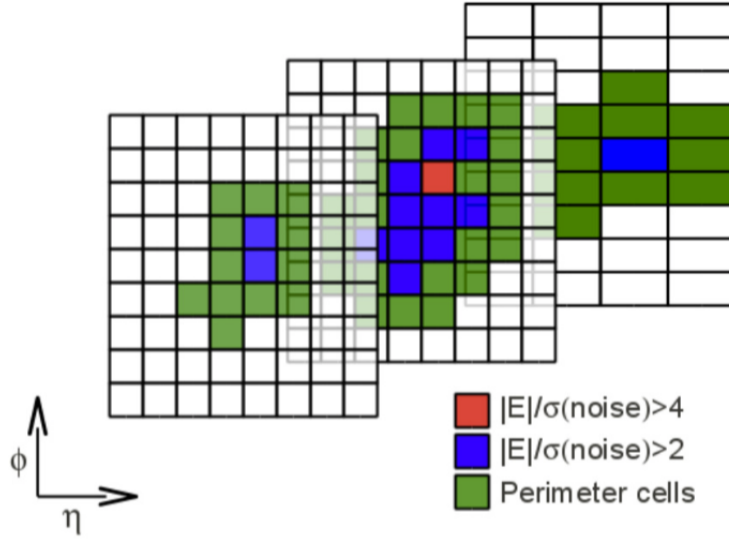


Figure 1.9: Illustration of calorimeter-cell topological clustering across the three layers of the hadronic calorimeter. Indicated are the cells satisfying the signal-to-noise requirements for the seed (red), neighbor (blue), and perimeter (green) cells that make up the final three-dimensional topo-cluster.

## Jet Finding

Once the set of topo-clusters is formed, the process of jet finding begins. As there is no single unique way to define a jet, there is a wide variety of jet finding algorithms whose purpose is to associate jet constituents — here, the calorimeter-cell topo-clusters — to form the final object representing the jet. The default jet finding algorithm used by ATLAS is the *anti- $k_t$*  jet clustering

algorithm [11]. The anti- $k_t$  algorithm belongs to the more general class of sequential recombination algorithms and is favored for its infrared and collinear (IRC) safe properties as well as the fact that it tends to produce rather simple jets, geometrically, that are circular in the  $\eta - \phi$  plane, as seen in Figure 1.10. IRC safety in jet finding algorithms refers to the property that neither additional collinear splitting of jet constituents (e.g. the initiating or radiated partons) nor soft emissions should change the clustered jet. IRC-safe jets are therefore robust against these divergent regimes of QCD, sensitive to arbitrary calculational choices made in perturbation theory, and makes them physically meaningful observable objects with which one can make predictions.

The anti- $k_t$  algorithm takes as input constituents the topo-clusters described in Section 1.3.1 and computes the quantities,

$$d_{ij} = \min \left( \frac{1}{k_{T,i}^2}, \frac{1}{k_{T,j}^2} \right) \frac{\Delta R_{ij}^2}{R^2}, \quad (1.4)$$

$$d_{iB} = \frac{1}{k_{T,i}^2}, \quad (1.5)$$

with  $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ , **its rapidity, right?**  $R$  is a parameter whose value regulates the radial extent of the jet, and  $k_{T,i}$  is the transverse momentum of the  $i^{th}$  constituent. The  $d_{ij}$  and  $d_{iB}$  quantities are ‘distance’ metrics used in the clustering of input topo-cluster constituents. The former represents the ‘distance’ between the  $i^{th}$  and  $j^{th}$  constituent while latter represents the ‘distance’ between the  $i^{th}$  constituent and the beam-line, introduced to distinguish between constituents originating from the primary hard-scatter vertex and those originating from soft proton interaction remnants. The work to be discussed in the present thesis sets  $R = 0.4$ , which is the standard used in ATLAS.

The anti- $k_t$  algorithm proceeds by clustering those constituents whose inter-distance is smallest, thereby tending to cluster higher- $p_T$  constituents together, which can be seen by inspection of Equation 1.4 and 1.5. If, of the set of input constituents, the smallest distance is a  $d_{ij}$ , the associated constituents indicated by  $i$  and  $j$  are recombined to form a single constituent in the list that replaces them both. If the smallest distance is a  $d_{iB}$ , then the constituent indicated by index  $i$  is removed from the set of constituents and is considered as a complete jet. This process repeats, starting with the now smaller (due to successful constituent recombination or removal) set of constituents, until no constituents are left. The result of this process is a set of recombined constituents that represent jets, as illustrated in Figure 1.10.



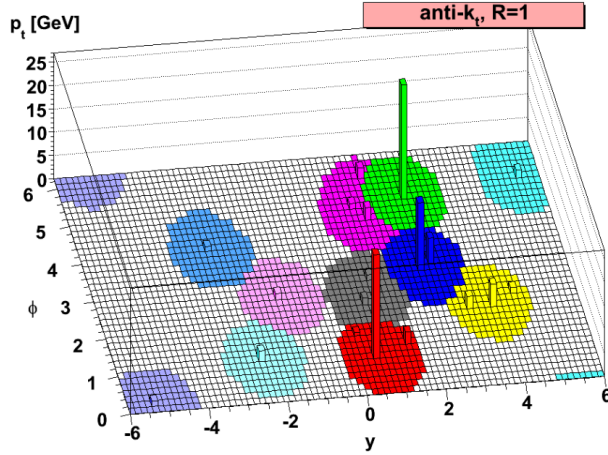


Figure 1.10: An illustration of jet constituents clustered by the anti- $k_t$  algorithm. Seen are the energetic constituents. The filled and colored circles represent areas populated by soft jet constituents, and represent the jet *catchment area* [12] whose size is dictated by the  $R$  parameter in the anti- $k_t$  algorithm (Equation 1.4). Figure taken from Ref. [11].

### 1.3.2 Jet Calibration

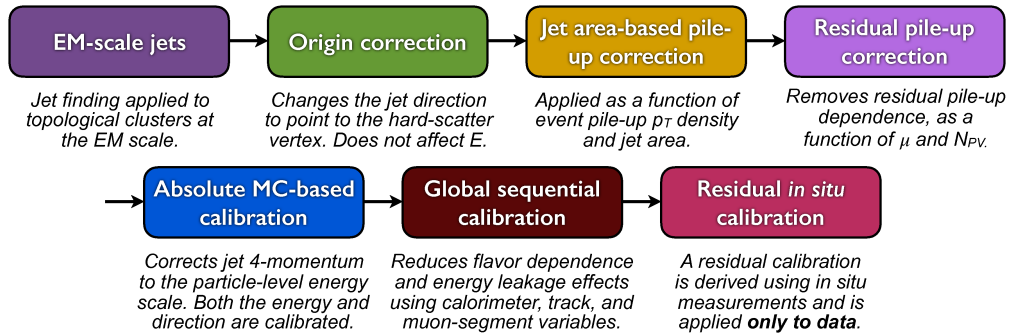


Figure 1.11: Flowchart representing the sequence of steps taken in the jet calibration. From Ref. [13].

The jets reconstructed following the steps described in Section 1.3.1 are objects clustered at the electromagnetic (EM) scale, which correctly measures the energy of electromagnetic showers but does not accurately account for energy depositions characteristic of hadronic particle decays and interactions. These jets are therefore referred to as ‘EM-scale’ jets. To correctly assign meaningful energy and momentum measurements to the reconstructed jets that correspond to the energies and momenta of the initiating, underlying particle-level jets, several *jet energy scale* (JES) calibration steps are taken [13]. The steps are detailed in the flowchart in Figure 1.11 and will be briefly described in the following text. The measurements made at each of these steps are subtle and account for many effects not present in the case of the reconstruction of electrons and muons, for example, due to the fact that jets are rather complicated collective phenomena whose measurements

rely primarily on a single subsystem (the calorimeters). Not only are electrons and muons generally simpler objects, seeded by comparatively unambiguous tracks within the ID, their energy and momentum measurements are the result of a combination of well-defined measurements made by two independent subsystems (ID measurements combined with EM calorimeter or the MS) that provide independent cross-checks on the validity of the measurements made by each. For the most part, this is not the case of the reconstruction of jets within ATLAS and as a result many of the choices made in the calibration of EM-scale jets have non-negligible impact in the analyses that will be discussed in this thesis, whereas the analogous choices for electrons and muons have minimal impact. For this reason the jet calibration procedure outlined in Figure 1.11 will be outlined briefly in the text that follows.

### Jet Origin Correction

The reconstructed EM-scale jets are built with the assumption that they originate from the geometric center of the detector, as opposed to the primary hard-scatter vertex from which the initiating partons arise. The so-called jet origin correction, therefore, refers to recalculating the jet four-momentum vector by adjusting it in such a way that it points to the primary hard-scatter vertex. This correction acts primarily to improve the  $\eta$  resolution of jets. This procedure is only one hundred percent accurate, of course, under the assumption that all of the jet constituents going into the EM-scale jet reconstruction originated from the hard-scatter vertex, as opposed to some fraction having come from a pileup vertex, for example.

### Pileup Corrections

Jets are extended object with relatively large *catchment areas* [12] that make them susceptible to pileup effects. Several corrections, therefore, to the jet energy are taken in order to account for contributions to the EM-scale jet reconstruction arising from both in-time and out-of-time pileup interactions.

The first pileup correction is an area-based correction which subtracts the per-event pileup contribution to the  $p_T$  of each jet based on the jet’s area, where the jet area is defined as in Ref. [12]. This pileup contribution is taken as the median  $p_T$  density,  $\rho$ , of jets in the  $\eta - \phi$  plane and can be thought of as a baseline ‘noise’ term distributed evenly throught the calorimeter that contributes to a jet’s reconstructed energy.

After the area-based pileup correction is made, there still remains residual dependence of the reconstructed jet  $p_T$  on the number of reconstructed primary vertices and on the number of interactions,  $\mu$ . These dependences are measured by performing linear fits of the jet  $p_T$  as a function of each quantity, binned as a function of the detector  $\eta$ ,  $\eta_{\text{det}}$ .

The final, pileup-corrected jet  $p_T$  is given by Equation 1.6:

$$p_T^{\text{Corr}} = p_T^{\text{Reco}} - \underbrace{\rho \times A}_{\text{Area-based}} - \underbrace{\alpha \times (N_{\text{PV}} - 1) - \beta \times \mu}_{\text{Residual}}, \quad (1.6)$$

where  $A$  is the jet area, and the  $\alpha$  and  $\beta$  terms are derived from the linear fits mentioned above and are  $\alpha = \partial p_T / \partial N_{PV}$  and  $\beta = \partial p_T / \partial \mu$ , respectively. The former accounts for effects arising as a result of in-time pileup and the latter for those due to out-of-time pileup. The effect of the pileup corrections is shown in Figure 1.12, where it can be seen that the area-based correction is an overall offset, as expected, after which residual pileup dependencies of the jet  $p_T$  as a function of  $|\eta_{det}|$  are still observed. This residual pileup dependence is greater in the forward regions of the detector where pileup and background activity is largest.

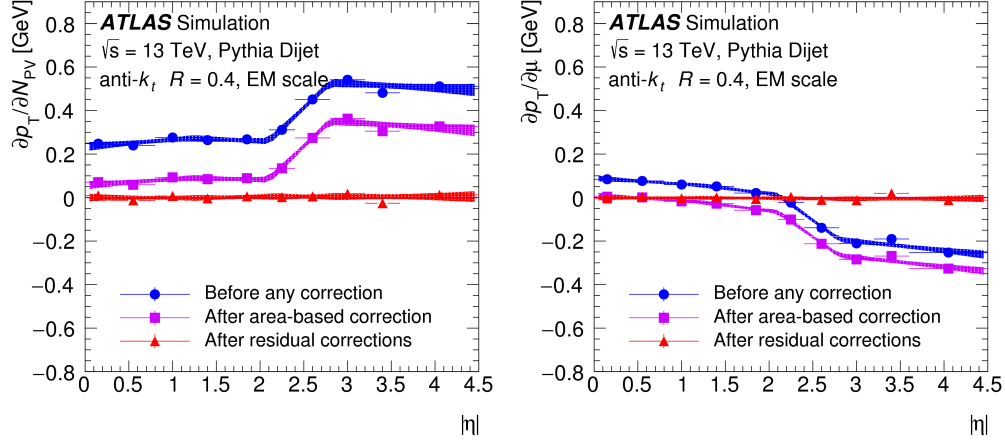


Figure 1.12: Dependence of the  $p_T$  of EM-scale reconstructed jets on  $N_{PV}$  (in-time pileup) (*left*) and on  $\mu$  (out-of-time pileup) (*right*). The blue curves show the dependence prior to any pileup corrections, the purple curves are after the area-based correction, and the red curves are the final dependence after the full pileup correction described in Equation 1.6 is taken into account. Figures taken from Ref. [13].

### Absolute Jet Energy Scale and $\eta$ Correction

This correction corrects the EM-scale reconstructed jet to the true energy scale based on particle-level jets and is therefore purely MC-based. Particle-level jets are jets clustered using the anti- $k_t$  algorithm using the generator-level particles at the end of the hadronisation step as the input constituents, and therefore represent the reconstructed jet prior to its interaction with the calorimeter (see Figure 1.7). The correction accounts for mismodelling of the inactive material within the detector, radiation not accounted for in the reconstructed calorimeter-based jet due to the clustering algorithm not accepting it (‘out-of-cone radiation’), non-compensation of the hadronic calorimeters<sup>2</sup>, and for effects related to detector geometry or transitions between calorimeter technologies.

The correction is derived by matching the EM-scale reconstructed jets, in simulation, to the particle-level jets and deriving the average energy response,  $E^{\text{Reco}}/E^{\text{Truth}}$ , where  $E^{\text{Truth}}$  is the energy of the particle-level jet. The inverse of this energy response is taken as a correction to the

<sup>2</sup> Compensating...

EM-scale reconstructed calorimeter jets in simulation. An additional correction accounts for biases observed in the EM-scale reconstructed jet  $\eta$ , which is largest in regions wherein a jet is likely to encompass two calorimeter regions or technologies which result in different energy responses. The  $\eta$  correction is derived as the difference between the reconstructed and particle-level jet  $\eta$  values and applied to the EM-scale reconstructed calorimeter jets as in the case of the energy-response correction. The average energy response, as a function of EM-scale jet  $p_T$ , and the  $\eta$  correction is shown in Figure 1.13. After this correction stage, jets are referred to as being at the EM+JES scale.

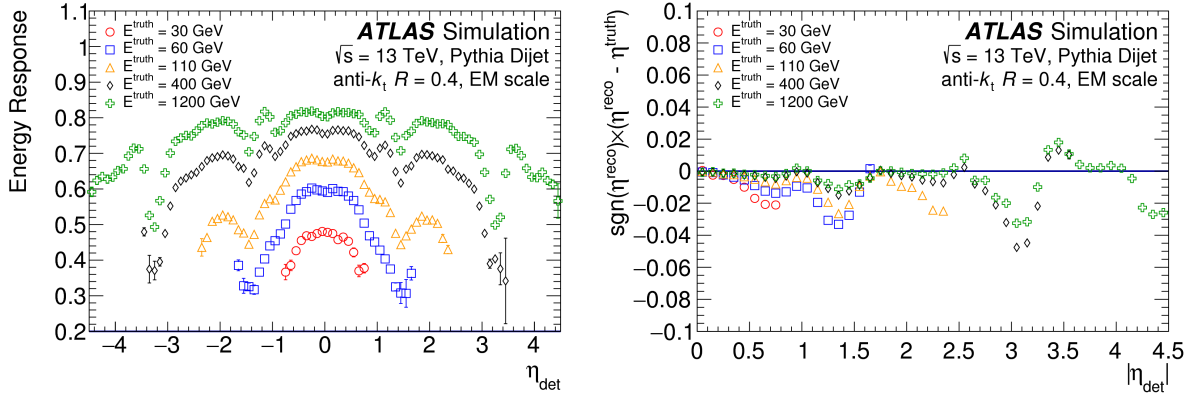


Figure 1.13: *Left:* Average energy response as a function of jet detector  $\eta$ ,  $\eta_{\text{det}}$ . The colors correspond to different energy regimes for the particle-level jet to which the EM-scale reconstructed jet is matched. The inverse of the response is the final correction and can be seen to be largest for lower- $p_T$  jets. *Right:* Difference in  $\eta$  for the EM-scale reconstructed jet and the particle-level jet to which it is matched. The bias is clearly seen, with values typically negative, and it being largest for  $|\eta_{\text{det}}| \sim 1.4$  ( $\sim 3.1$ ), corresponding to the barrel-endcap (endcap-forward) transition regions.

## Global Sequential Calibration

The so-called Global Sequential Calibration (GSC) is a catch-all correction to account for remaining dependencies of the EM+JES-scale jets on the jet shower shapes as well as fluctuations in the jet flavor composition and inter-jet energy distribution. This correction improves the handling of fluctuations in the composition of the particles that initiate the jet; for example, correcting for the differences expected between quark- and gluon-initiated jets. The former (quark-initiated) jets are typically more collimated with fewer, but higher- $p_T$ , hadronic constituents. The latter (gluon-initiated) jets typically contain many more, softer- $p_T$ , particles and have wider transverse profiles (and therefore do not traverse as far into the calorimeter). The GSC has five stages, each following a numerical inversion of a corresponding jet response as in the case(s) described in Section 1.3.2, but are based on observables sensitive to the jet shower profile and growth within the calorimeter as well as on the number and type of tracks associated with the reconstructed jet. The use of tracking information from muons in this correction additionally helps with correcting the energy response

of jets that are not fully contained in the calorimeter but leak into the MS, so-called *punch-through jets*. Figure 1.14 illustrates the concept of jet punch-through.

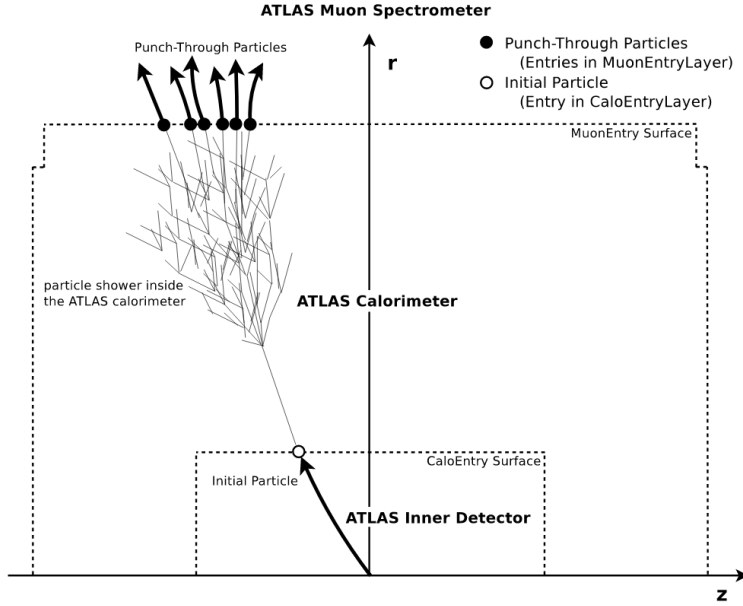


Figure 1.14: Illustration of jet punch-through. High momentum particles produced in the shower of an energetic jet within the hadronic calorimeter escape into the muon spectrometer, leaving detectable signatures in the muon chambers. It can be seen that energy and/or momentum may be left unaccounted for in the final event reconstruction or be assigned to separate muon objects as opposed to the initiating jet from which the punch-through particles arise. Such effects disrupt the proper assignment of energy-momentum to the jet and can spoil the overall momentum balance in the event.

## In-situ Corrections

The last stage of the jet calibration accounts for remaining differences in the EM+JES jet response between data and simulation and is primarily derived using events from data, as opposed to MC simulation.<sup>3</sup> There are two classes of correction, the first being the so-called  $\eta$ -intercalibration step which corrects the average response of forward jets ( $0.8 < |\eta| < 4.5$ ) to that of well-measured central jets using dijet events in which the two jets (the one in the forward region and the other in the central region) are back-to-back in  $\phi$ . The second class of corrections is based on the method of balancing an EM+JES scale jet against a well-measured reference object. The balance methods use only central jets ( $|\eta| < 0.8$ ) and the choice of the reference object provides the  $p_T$  scale to which the response correction applies. Balance methods using well-measured photons and  $Z$  boson decays to leptons ( $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$ ) as the reference objects are used to derive the response corrections for jets with  $p_T$  up to 950 GeV. Response corrections for jets covering  $p_T$  ranges up to

<sup>3</sup>In this case, the term ‘in-situ’ means that the corresponding corrections are derived using data, as opposed to using events from Monte-Carlo simulation.

2 TeV are derived using a multijet balance method in which a high- $p_T$  central jet is balanced against a reference system of lower- $p_T$  jets. The final response correction as a function of jet  $p_T$  derived using these in-situ methods is shown in Figure 1.15.

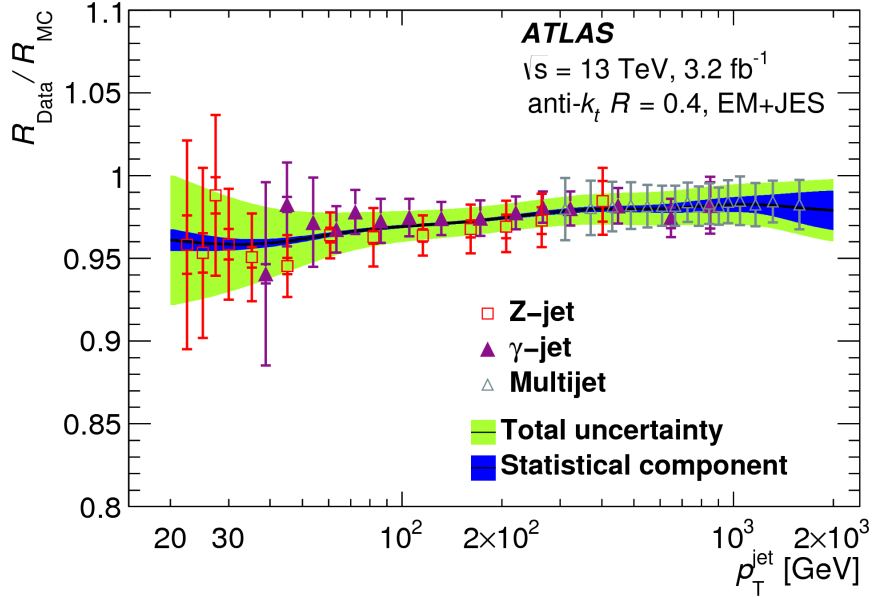


Figure 1.15: Ratio of the EM+JES jet response in data to that in MC simulation as a function of jet  $p_T$ . The different markers indicate the measurement contributions from the different reference-object used to derive the in-situ response corrections. The final response correction is given by the black line. The total uncertainty on the response is given by the green band and the blue band indicates only the statistical component of the uncertainty. Figure from Ref. [13].

### 1.3.3 Flavor Tagging of Jets

The ability to identify jets containing heavy-flavored hadrons, i.e. jets containing  $b$ - and  $c$ -flavored hadrons, is an important aspect to many of the most critical measurements and analyses being done at the large LHC experiments. The SM top-quark is the heaviest known elementary particle, the second-most recent elementary particle to be discovered, and, given its importance to electroweak and Higgs physics, is an object subjected to high levels of precise study at the LHC. The top-quark decays before hadronisation timescales and therefore its decay products, which are a  $W$ -boson and a  $b$ -quark nearly 100% of the time, carry away information characterising its properties. Being able to characterise the jets initiated by the hadronisation of these  $b$ -quarks, then, is of critical import if the physics of the top-quark wish to be understood. The recent discovery of an SM-like Higgs boson, with a mass of  $m_h = 125 \text{ GeV}$ , decays to a pair of  $b$ -quarks nearly 60% of the time. Without question, then, the thorough study of the Higgs boson necessitates the ability to precisely identify the presence of the pair of  $b$ -initiated jets from its decays. Additionally, we will see in subsequent chapters that the presence of  $b$ -quark initiated jets is a characteristic signature of many BSM physics scenarios. The identification of these types of jets, a process referred to as

‘flavor tagging’, is of the utmost importance to analyses performed with the ATLAS detector as well as in the work to be presented in this thesis. Jets tagged as having likely arisen as a result of the hadronisation of an initiating  $b$ -( $c$ -)quark are referred to as ‘ $b$ -tagged jets’ (‘ $c$ -tagged jets’), or simply as ‘ $b$ -jets’ (‘ $c$ -jets’). All other jets then are assumed to have arisen from the decay of light-flavor quarks and are referred to as ‘light-flavor jets’ (‘light-jets’).

Heavy-flavor tagging of jets relies on the relatively long lifetimes of the  $b$ - and  $c$ -hadrons which initiate them. The typical  $b$ -hadron lifetime is  $\tau \approx 1.6$  ps ( $c\tau \approx 450$   $\mu$ m), which leads to  $b$ -hadrons traversing typically macroscopic distances away from the primary hard-scatter vertex before they decay. As seen in Figure 1.16,  $b$ -hadrons with transverse momenta on the order of 50 GeV will travel nearly half a centimeter before decaying. Also seen in Figure 1.16, those  $b$ -hadrons with  $p_T$  values nearing 250 GeV will actually decay outside of the beam-pipe within the region of the IBL. Given the high spatial resolution of the ID pixel and SCT detectors discussed in Section ??, the presence of these long-lived particles should be detectable via the presence of at least one secondary decay vertex corresponding to the point at which the heavy-flavored hadron decays. Figure 1.17 illustrates the standard topology of a heavy-flavor initiated jet with a secondary decay vertex that is displaced with respect to the primary hard-scatter vertex and is the source of displaced tracks.

The algorithms used to identify heavy-flavor initiated jets, then, rely on information characterising the long lifetimes of the initiating particles and on the presence of secondary decay vertices within the jet. These high-level algorithms used to identify these jets take as input information provided by taggers that rely on low-level information based on the presence of displaced tracks and secondary vertices. These low-level taggers will be introduced in Section 1.3.3 and the construction of the high-level tagger, used in the analyses to be presented in this thesis, will be presented in Section 1.3.3.

## Low Level Taggers and Inputs

The final algorithm used to identify  $b$ -tagged jets take as input the outputs from several low-level  $b$ -tagging algorithms. There are two classes of low-level algorithms: those that rely on the impact parameter information of the tracks associated with the jets and those that rely on the explicit reconstruction of secondary decay vertices within the jet [14]:

- **IP2D and IP3D** The IP2D and IP3D algorithms make use of the signed transverse impact parameter significance of tracks to construct discriminating variables. The IP3D algorithm additionally makes use of the longitudinal impact parameter significance. The algorithms rely on constructing log-likelihood ratios (LLR) taking as inputs probability density functions (PDFs) for  $b$ -,  $c$ -, and light-flavor jet probabilities on a per-track basis. See Figure 1.17 for an illustration of the signed impact parameter. Tracks associated with secondary decay vertices in heavy-flavor jets tend to have positively signed impact parameters since they are produced *within* the jet cone. Signed impact parameter quantities tend to be symmetric for light-flavor jets. See Figure 1.18.

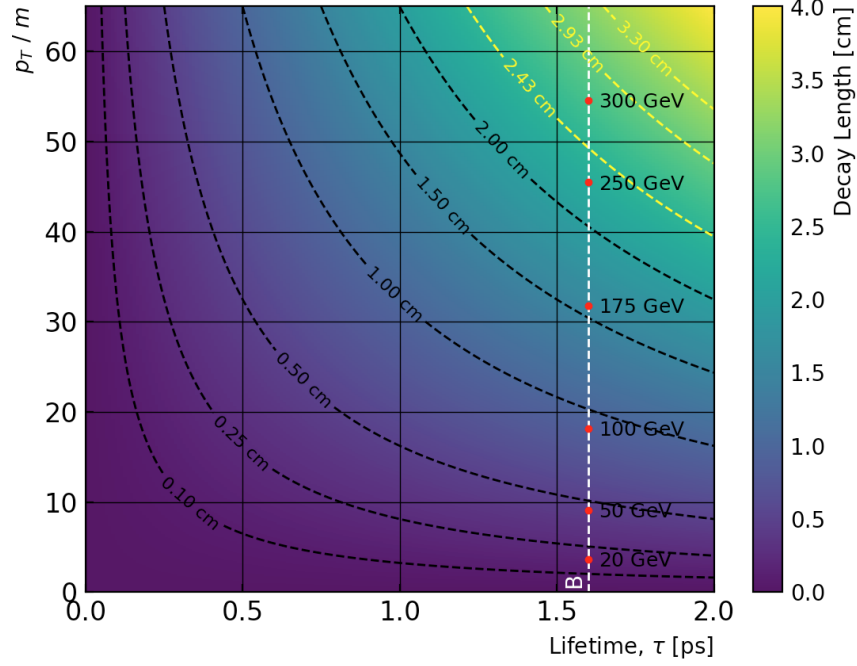


Figure 1.16: Particle decay length as a function of its lifetime and transverse momentum normalised to its rest mass. The white-dashed line indicates the average lifetime of  $B$ -hadron species, taken as 1.6 ps, with a mass taken to be 5.5 GeV. The red-dots along the  $B$ -hadron line indicate locations for specific transverse momenta for the decaying  $B$ -hadron. The yellow contours indicate the locations of the IBL (Figure ??), with 2.43 cm corresponding to the beam-pipe radius. *Perhaps move this plot elsewhere? Add references to PDG*

- **Secondary Vertex Finding Algorithm (SV1)** The SV1 algorithm [15] reconstructs a single displaced secondary vertex within a jet, starting from the set of all possible two-track vertices while rejecting tracks likely to be associated with non-heavy-flavored long-lived particles ( $K_s$  or  $\Lambda$ ), photon conversions, or vertices due to detector material interactions. The inclusive secondary vertex and associated tracks are then used to construct discriminating observables sensitive to the differences between  $b$ -,  $c$ -, and light-flavor jets.
- **Multi-vertex Finding Algorithm (JetFitter, JF)** The JetFitter algorithm [16] attempts to reconstruct the full  $b$ -hadron decay chain using a modified Kalman filter [17] which assumes a common line on which the primary,  $b$ -hadron, and  $c$ -hadron decay vertices lie. As with SV1, the construction of many discriminating observables related to the reconstructed set of vertices are used to build templates for  $b$ -,  $c$ -, and light-flavor jets.

Figure 1.18 provides an example of a few observables provided by the low-level tagging algorithms.



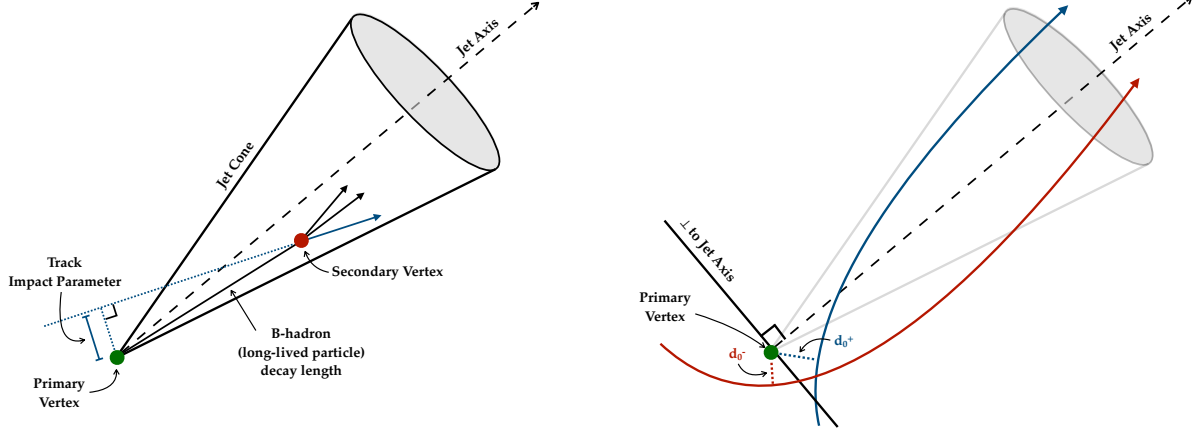


Figure 1.17: *Left*: Topology of a  $b$ -jet. The  $b$ -hadron produced near the primary hard-scatter vertex (green dot), initiating the  $b$ -jet, has a long lifetime and decays a macroscopic distance away from the primary hard-scatter vertex to produce a secondary vertex (red dot) from which additional tracks are produced and subsequently reconstructed. The tracks originating from the secondary vertex will have larger impact parameters relative to the primary hard-scatter vertex as compared to tracks originating from the primary hard-scatter vertex. *Right*: Illustration of signed impact parameter, specifically the signed transverse impact parameter  $d_0$ . The blue track has a transverse impact parameter whose projection onto the jet axis is in the direction in which the jet momentum points and is given a positive  $d_0$ . The red track's transverse impact parameter's projection onto the jet axis points opposite to the jet momentum and is given a negative  $d_0$ .

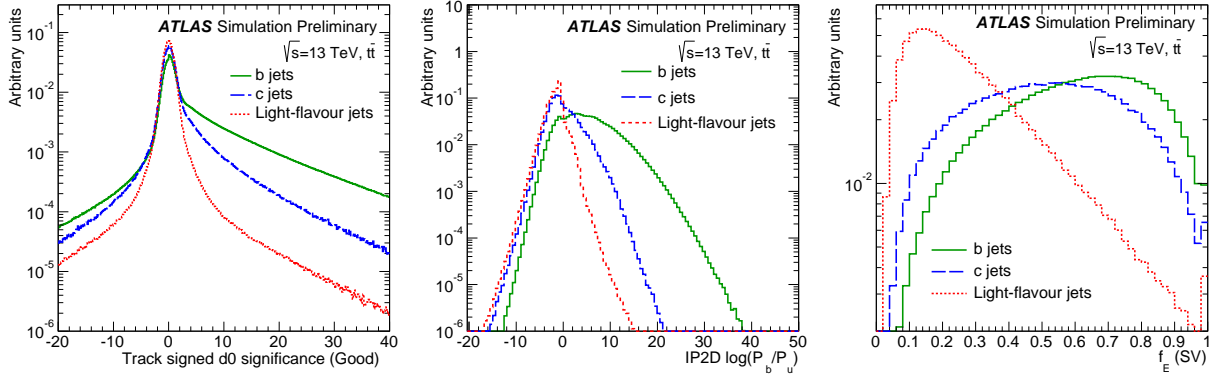


Figure 1.18: Examples of a few low-level quantities used in the ATLAS flavor tagging algorithms. The blue histograms are distributions associated with  $b$ -jets, green are those of  $c$ -jets, and red are those of light-flavor jets. *Left*: Two-dimensional (signed)  $d_0$  significance for tracks matched to jets. *Middle*: IP2D  $b$ -jet log-likelihood ratio. *Right*: Energy fraction, defined as the energy of the tracks in the displaced vertex reconstructed by the SV1 algorithm relative to the energy of all tracks in the jet.

## High Level Tagger: MV2

The low-level taggers discussed in the previous section provide a set of many useful and complementary observables. In an attempt to make the most efficient use of all the information provided by

this set of observables, a boosted decision tree (BDT) algorithm is used to combine the outputs of these low-level algorithms. This algorithm, referred to as the MV2  $b$ -tagging algorithm, is trained using the ROOT Toolkit for Multivariate Data Analysis (TMVA) [18]. In the analyses based on data recorded by ATLAS between 2015–2016, the MV2 algorithm was trained using jets from a simulated sample of top-quark pair production events. For the analyses based on the full Run-II data recorded by ATLAS, up to and including the year 2018, the MV2 algorithm was retrained using a sample composed of jets both from top-quark pair production events as well as from simulated events of a BSM physics scenario of a heavy  $Z'$  decaying to  $b\bar{b}$ . The latter was included in the retraining so as to allow the algorithm to have in its training sample high- $p_T$  jets that are not present in the SM top-quark pair production events, as illustrated in Figure 1.19.

Figure 1.19 shows a distribution of the MV2 algorithm’s output. The MV2 score is computed on a per-jet basis, using the set of low-level inputs listed in Table 1.3. Working points, defined with different target-efficiencies for accepting  $b$ -tagged jets, are defined by selection thresholds on the MV2 discriminant. The standard ATLAS  $b$ -tagging working points are defined for accepting  $b$ -jets with  $p_T > 20$  GeV with average efficiencies of 60%, 70%, 77%, and 85%. The working points are based on selections made on the MV2 output score and are defined in Table 1.4, along with the rejection factors<sup>4</sup> for  $c$ -jets,  $\tau$ -jets<sup>5</sup>, and light-flavor jets.

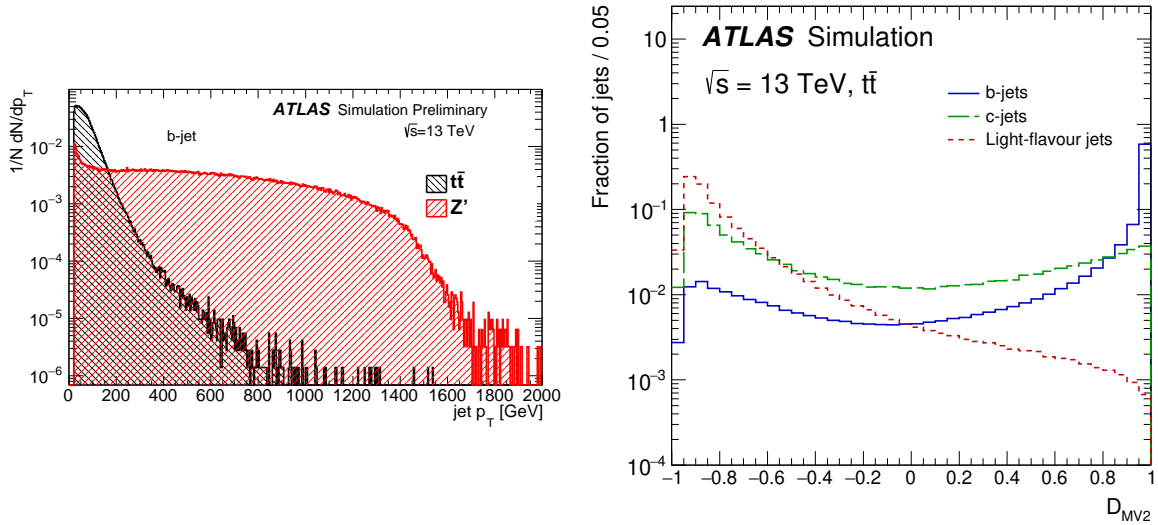


Figure 1.19: *Left*: Distribution of jet  $p_T$  for jets used in the training of the MV2 BDT algorithm. The  $Z'$  sample of jets is only included in the MV2 training relevant to the analyses based on the full Run-II dataset collected by ATLAS. *Right*: Distribution of the BDT-based MV2  $b$ -tagging algorithm output score, shown for  $b$ -jets (blue),  $c$ -jets (green), and light-flavor jets (red).

<sup>4</sup>The rejection factor is defined as the inverse of the efficiency. A rejection factor of 100 means, therefore, that the associated object is accepted — on average — 1 out of every 100 times that it appears.

<sup>5</sup>Hadronically decaying  $\tau$  leptons are accepted by the  $b$ -tagging algorithms at rates higher than light-flavor jets due to the non-negligible decay length of the  $\tau$  lepton which give them  $b$ -like characteristics.

Table 1.3: Variables used as input to the high-level tagger MV2c10.

Input Source	Input Name	Description
Kinematics	$p_T(\text{jet})$	Jet transverse momentum
	$\eta(\text{jet})$	Jet pseudorapidity
IP2D, IP3D	$\log(p_b/p_{\text{light}})$	Likelihood ratio between the $b$ - and light-jet hypotheses
	$\log(p_b/p_c)$	Likelihood ratio between the $b$ - and $c$ -jet hypotheses
	$\log(p_c/p_{\text{light}})$	Likelihood ratio between the $c$ - and light-jet hypotheses
Secondary Vertex (SV1)	$m_{\text{SV}}$	Invariant mass of tracks at the secondary vertex assuming pion masses
	$f_E$	Fraction of the charged jet energy in the secondary vertex
	$N_{\text{TrkAtVtx}}$	Number of tracks used in the secondary vertex
	$N_{2\text{TrkVtx}}$	Number of two-track vertex candidates
	$L_{xy}$	Transverse distance between the primary and secondary vertices
	$L_{xyz}$	Distance between the primary and secondary vertices
	$S_{xyz}$	Distance between the primary and secondary vertices divided by its uncertainty
JetFitter (JF)	$\Delta R(\text{jet}, \text{SV})$	$\Delta R$ between the jet axis and the direction of the secondary vertex relative to the primary vertex
	$N_{2\text{TrkVtx}}$	Number of two-track vertex candidates (prior to JetFitter decay-chain fit)
	$m$	Invariant mass of tracks from displaced vertices assuming pion masses
	$S_{xyz}$	Significance of the average distance between the primary and displaced vertices
	$f_E$	Fraction of the charged jet energy in the secondary vertices
	$N_{1\text{-trk vertices}}$	Number of displaced vertices with one track
	$N_{\geq 2\text{-trk vertices}}$	Number of displaced vertices with more than one track
	$N_{\text{TrkAtVtx}}$	Number of tracks from displaced vertices with at least two tracks
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})$	$\Delta R$ between the jet axis and the vectorial sum of the momentum of all tracks attached to displaced vertices

### $b$ -jet Identification Calibration

As the efficiencies for the MV2 algorithm detailed in Table 1.4 are based entirely on MC simulation, a calibration procedure is performed to correct the MC-based efficiencies to those observed in data. This is necessary to get an accurate prediction in simulation of the rate of  $b$ -tagged jets occurring in data. The efficiencies (and rejection factors) in Table 1.4 are therefore measured in data. The result of the calibration is a correction scale-factor, applied on a per-jet basis, that is defined as  $\text{SF} = \varepsilon_{\text{Data}}/\varepsilon_{\text{MC}}$ , where  $\varepsilon_{\text{Data(MC)}}$  are the measured efficiencies in data (MC) for a jet to be tagged as a  $b$ -jet. The  $b$ -tagging SF are derived using a sample enriched in  $b$ -tagged jets arising from SM top-quark pair production [14]. The  $b$ -tagging efficiencies in MC and data as well as the data-to-MC efficiency SF are shown in Figure 1.20. The rate of  $c$ -,  $\tau$ -, and light-flavor jets to be identified as  $b$ -tagged jets corresponding to the rejection factors listed in Table 1.4, also have corresponding correction scale-factors using events in data enriched in  $c$ -jets and mis-tagged light-flavor jets [14,

Table 1.4: Working points defined for the MV2  $b$ -jet identification algorithm. The cut thresholds on the MV2 discriminant associated with a given  $b$ -jet efficiency (working point) are given in the second column. The rejection factors for  $c$ -,  $\tau$ -, and light-flavor jets are shown in the three right-most columns. The precise MV2 discriminant thresholds are dependent on the calibration, and differ between the analyses based only on the data collected in the years 2015-2016 and the full Run-II dataset including the years 2017 and 2018. The MV2 threshold values shown here are those corresponding to the full Run-II dataset.

$b$ -jet efficiency	MV2 selection	Rejection Factor		
		$c$ -jet	$\tau$ -jet	Light-flavor jet
60%	$> 0.94$	23	140	1200
70%	$> 0.83$	8.9	36	300
77%	$> 0.64$	4.9	15	110
85%	$> 0.11$	2.7	6.1	25

19, 20]. These latter scale-factors correct the rate of mis-tagging (i.e. identifying jets as  $b$ -jets when they are not initiated by  $b$ -hadrons).

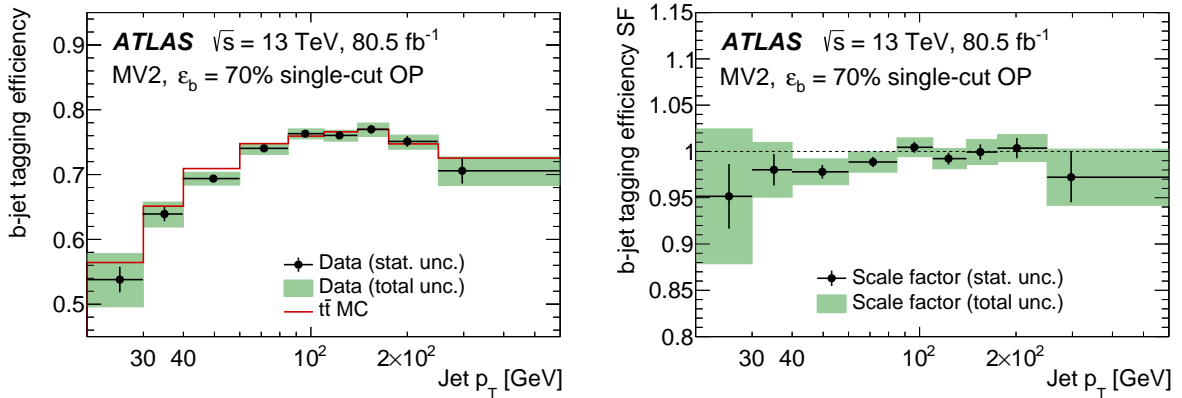


Figure 1.20: *Left*:  $b$ -jet tagging efficiency as a function of jet  $p_T$  for the 70% WP of the MV2  $b$ -jet tagging algorithm in MC (top-quark pair production, in red) and data (black points). Efficiency values below and above 70% occur, but the efficiency averaged over the full range shown is roughly 70%. *Right*:  $b$ -jet tagging efficiency correction scale-factors for the 70% WP of the MV2  $b$ -jet tagging algorithm as a function of jet  $p_T$  for the same samples as on the *left*.

## 1.4 The Missing Transverse Momentum

## 1.5 Object-level Ambiguity Resolution

# Bibliography

- [1] T Cornelissen et al. “Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)”. In: ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002 (Mar. 2007). URL: <https://cds.cern.ch/record/1020106> (cit. on p. 2).
- [2] “The Optimization of ATLAS Track Reconstruction in Dense Environments”. In: ATL-PHYS-PUB-2015-006 (Mar. 2015). URL: <https://cds.cern.ch/record/2002609> (cit. on p. 2).
- [3] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *Eur. Phys. J. C* 77.5 (2017), p. 332. DOI: 10.1140/epjc/s10052-017-4887-5. arXiv: 1611.10235 [physics.ins-det] (cit. on p. 2).
- [4] “Performance of primary vertex reconstruction in proton-proton collisions at  $\sqrt{s}=7$  TeV in the ATLAS experiment”. In: ATL-CONF-2010-069 (July 2010). URL: <https://cds.cern.ch/record/1281344> (cit. on p. 2).
- [5] G Piacquadio, K Prokofiev, and A Wildauer. “Primary vertex reconstruction in the ATLAS experiment at LHC”. In: *Journal of Physics: Conference Series* 119.3 (July 2008), p. 032033. DOI: 10.1088/1742-6596/119/3/032033. URL: <https://doi.org/10.1088%2F1742-6596%2F119%2F3%2F032033> (cit. on p. 2).
- [6] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data”. In: (2019). arXiv: 1908.00005 [hep-ex] (cit. on pp. 4, 5, 6, 7, 8).
- [7] “Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung”. In: ATL-CONF-2012-047 (May 2012). URL: <https://cds.cern.ch/record/1449796> (cit. on p. 5).
- [8] Georges Aad et al. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s}=13$  TeV”. In: *Eur. Phys. J. C* 76.5 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex] (cit. on pp. 9, 12).
- [9] W. Lampl et al. “Calorimeter clustering algorithms: Description and performance”. In: (2008) (cit. on pp. 13, 14).
- [10] Georges Aad et al. “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 p. 13).

- [11] 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. 15, 16).
- [12] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Catchment Area of Jets”. In: *JHEP* 04 (2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph] (cit. on pp. 16, 17).
- [13] M. Aaboud et al. “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* 96.7 (2017), p. 072002. DOI: 10.1103/PhysRevD.96.072002. arXiv: 1703.09665 [hep-ex] (cit. on pp. 16, 18, 21).
- [14] ATLAS Collaboration. “ATLAS  $b$ -jet identification performance and efficiency measurement with  $t\bar{t}$  events in  $pp$  collisions at  $\sqrt{s} = 13$  TeV”. In: (2019). arXiv: 1907.05120 [hep-ex] (cit. on pp. 22, 26).
- [15] “Secondary vertex finding for jet flavour identification with the ATLAS detector”. In: ATL-PHYS-PUB-2017-011 (June 2017). URL: <https://cds.cern.ch/record/2270366> (cit. on p. 23).
- [16] “Topological  $b$ -hadron decay reconstruction and identification of  $b$ -jets with the JetFitter package in the ATLAS experiment at the LHC”. In: ATL-PHYS-PUB-2018-025 (Oct. 2018). URL: <https://cds.cern.ch/record/2645405> (cit. on p. 23).
- [17] R. Fruhwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl. Instrum. Meth.* A262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4 (cit. on p. 23).
- [18] “TMVA – Toolkit for Multivariate Data Analysis”. In: (). arXiv: physics/0703039 [physics.data-an] (cit. on p. 25).
- [19] *Calibration of the ATLAS  $b$ -tagging algorithm in  $t\bar{t}$  semi-leptonic events*. Tech. rep. ATLAS-CONF-2018-045. Geneva: CERN, Sept. 2018. URL: <https://cds.cern.ch/record/2638455> (cit. on p. 27).
- [20] *Calibration of light-flavour jet  $b$ -tagging rates on ATLAS proton-proton collision data at  $\sqrt{s} = 13$  TeV*. Tech. rep. ATLAS-CONF-2018-006. Geneva: CERN, Apr. 2018. URL: <https://cds.cern.ch/record/2314418> (cit. on p. 27).