

Chapter 5

Event Reconstruction

Once recorded, the events are reconstructed using a wide variety of algorithms designed to identify the products of collisions at the center of the detector. The algorithms transform the raw data read out from the detector – hits in the inner detector silicon layers and TRT straws, energy deposits in calorimeter cells, and hits in the muon stations – into a list of physics objects and their energies or momenta. This section describes the techniques used to identify the objects used in the analyses described in chapters 7 and 8.

5.1 Track and Vertex Reconstruction

Tracks, nominally due to charged particles traversing the inner detector, are reconstructed using a series of algorithms based on hits in inner detector [1–3]. The primary “inside-out” algorithm begins by constructing three-dimensional space points associated with the hits in the pixel and SCT layers. Track seeds are formed from sets of three space points in the first four layers of the inner detector (three pixel layers and the innermost SCT layer), constrained to be consistent with a track originating from the interaction region. The track seeds are extended through the remaining SCT layers using a combinatorial Kalman filter. After screening the track candidates to reduce random coincidences and ambiguities from very close tracks, the tracks are extended through the TRT. Finally, the track is refitted using all of the associated hits.

A complementary “outside-in” algorithm identifies tracks with fewer or zero silicon hits, which can arise from hadron decays or photon conversions in the silicon layers, or from track seeds removed during the ambiguity resolution. The algorithm starts from TRT segments, and adds silicon hits proceeding inwards.

Vertices are points consistent with being the origin of a set of tracks. Primary vertices are vertices nominally due to pp collisions, while secondary vertices arise from other processes like hadron decays or photon conversions. The primary vertex reconstruction algorithm reconstructs vertices one by one, alternately finding a vertex seed and then fitting the corresponding tracks [3, 4]. For most applications, the tracks are required to have $p_T > 400$ MeV,

no missing hits along the track in the pixel layers¹ (*holes*), and at most two holes in the SCT layers. The seed finding identifies the global maximum in the distribution of z coordinates of the tracks. The vertex position is then determined by an adaptive vertex fitting algorithm [5], a χ^2 -based fit which suppresses the contribution from outlier tracks. Tracks whose impact parameter is inconsistent by more than 7σ with the vertex position are then reused to seed a new vertex, until no further seeds can be found.

For most applications, the vertex reconstruction is performed twice. After reconstructing an initial set of primary vertices, the position and size of the interaction region, or *beam spot*, is determined from a fit to the spatial distribution of vertices. The vertices are then reconstructed a second time using the beam spot as a constraint for seed finding and track fitting².

5.2 Electrons

The signature of an electron is an energy deposit in the electromagnetic LAr calorimeter and a track pointing at the energy deposit [6, 7]. The electron reconstruction algorithm begins by searching for clusters of energy in the calorimeter, based on a grid of $N_\eta \times N_\phi = 200 \times 256$ towers of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The tower energy is the sum of the cell energies in all longitudinal layers within the tower. Energy deposits with $E_T > 2.5$ GeV within a 3×5 window of towers form the seeds for both electrons and photons.

After passing loose shower shape requirements, the electron reconstruction algorithm searches for a track within a cone of radius $\Delta R = 0.3$ around the cluster barycenter. Two hypotheses are used for track pattern recognition and fitting: the standard pion hypothesis and an electron hypothesis that allows for larger energy losses due to bremsstrahlung. The cluster and track are required to satisfy one of the following two criteria:

- The barycenter of the cluster and the track extrapolation to the middle layer of the LAr calorimeter satisfy $\Delta\phi < 0.2$ in the direction of track bending or $\Delta\phi < 0.05$ in the other direction.
- The barycenter of the cluster and the track extrapolation to the middle layer of the LAr calorimeter, after rescaling the track momentum to the energy of the cluster, satisfy $\Delta\phi < 0.1$ in the direction of track bending or $\Delta\phi < 0.05$ in the other direction. This recovers low-energy electrons that potentially lose a significant fraction of their energy before reaching the calorimeter.

For tracks with at least four silicon hits, the track and the cluster must also satisfy $|\Delta\eta| < 0.05$. Finally, the cluster and track are rebuilt using algorithms optimized for measurement of the electron properties. The cluster is rebuilt sequentially in all four layers, using an

¹Excluding missing hits due to an inactive detector element.

²For the luminosity measurement using vertices, the beam spot constraint is not applied in order to avoid biases due to changes in the size of the interaction region

area of 3×7 layer-2 cells in the barrel or 5×5 layer-2 cells in the end-caps. The tracks of electron candidates are refit using an optimized electron track filter based on the Gaussian Sum Filter (GSF) algorithm [8]. The GSF track and the cluster must satisfy tighter spatial matching criteria: $\Delta\phi < 0.1$ in the direction of track bending, or $\Delta\phi < 0.05$ in the opposite direction. GSF tracks with less than four silicon hits are required to satisfy even tighter criteria: $|\Delta\eta| < 0.35$ or 0.2 in the TRT barrel or end-cap, and $\Delta\phi < 0.03$ in the direction of track bending or $\Delta\phi < 0.02$ in the other direction.

Identification

Further requirements can be imposed on electron candidates to suppress backgrounds from sources like misidentified hadronic jets, photon conversions, and electrons from hadron decays. Three increasingly stringent sets of cuts are defined, called loose, medium, and tight³. The analyses described in chapters 7 and 8 use the tight cuts to select signal electrons, and the medium and loose cuts to derive data-driven background estimates. The loose set of cuts impose requirements on the shower shape in the first and second calorimeter layers, the quality of the track, the fraction of energy in hadronic calorimeter cells behind the LAr cells, and the spatial match of the track and the calorimeter cluster. The medium set of cuts consist of more stringent versions of the loose cuts, and additionally require a small track impact parameter with respect to the primary vertex, a minimum number of high-threshold TRT hits associated with the track, and a hit in the innermost pixel layer. The tight set of cuts again consist of more stringent versions of the medium cuts, with an additional cut on the ratio of the cluster energy to the track momentum (E/p) and a veto of candidates associated to a photon conversion vertex.

Efficiency Measurements

The efficiency to detect an electron can be factorized into several components. The efficiency to detect a cluster in the electromagnetic calorimeter is very high, above 99% for electrons with $E_T = 15$ GeV and 99.9% for $E_T = 45$ GeV. The reconstruction efficiency, covering the matching of a good-quality track to the cluster, and the identification efficiencies, covering the identification cuts with respect to reconstructed electrons, are measured using tag-and-probe techniques targeting $Z \rightarrow ee$ and $J/\Psi \rightarrow ee$ events [7]. One electron, the tag, is required to satisfy strict selection criteria, while the second electron, the probe, is used for efficiency measurements. The combined reconstruction and identification efficiencies are shown in figure 5.1. The reconstruction efficiency makes up 1%-5% of the total efficiency loss for electrons with $E_T < 20$ GeV, and less than 1% for electrons with $E_T > 80$ GeV. The efficiencies are computed for both data and simulation, and the ratio between the two is used to correct the efficiency in simulation.

³An alternative method of identification using a likelihood-based multivariate method has been developed, but is not used in this dissertation

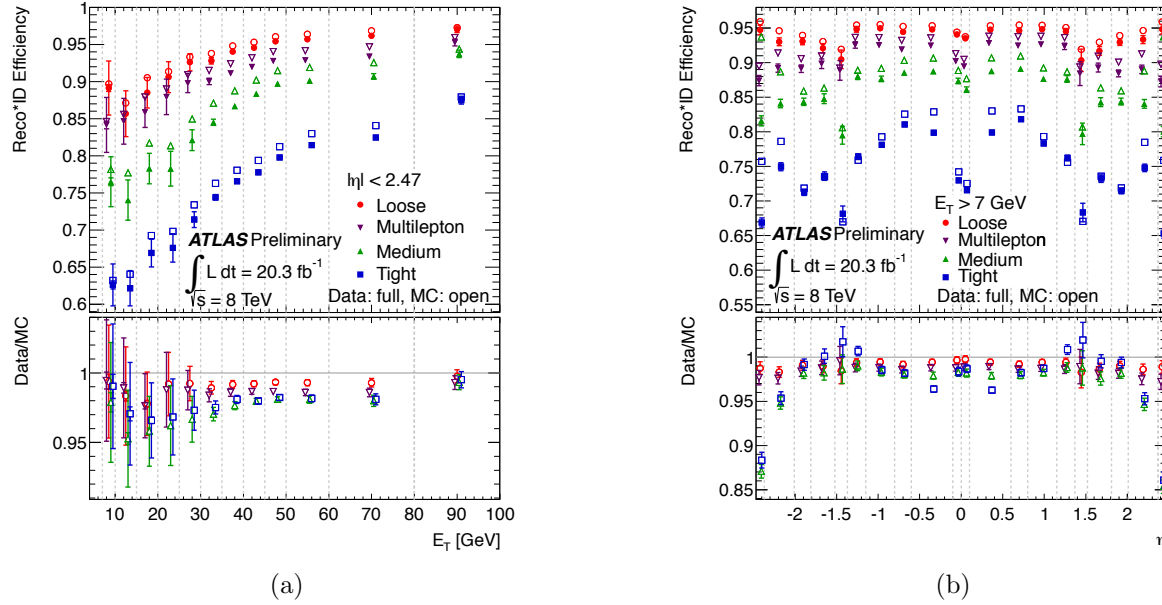


Figure 5.1: The combined reconstruction and identification efficiencies with respect to electrons detected as a cluster in the electromagnetic calorimeter, shown as a function of E_T (left) and η (right).

Energy and Momentum Measurement

For electron candidates with at least four silicon hits, the energy of the electron is taken from the calorimeter measurement, while the trajectory is taken from the GSF track. Candidates with fewer silicon hits are not used in this dissertation. The energy resolution of the calorimeter is parametrized as:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (5.1)$$

where a , b , and c are the sampling, noise, and constant terms, respectively, and generally vary with η . The design value of the sampling term is $a \approx 9 - 10\%$ in the central region, and worsens at higher pseudorapidities due to the increased material in front of the calorimeters. The noise term, due to electronics noise and pileup effects, is approximately $b = 350 \times \cosh \eta \text{ MeV}$. The constant term dominates the resolution at high energies, with a design value of $c = 0.7\%$.

As the ATLAS calorimeters are non-compensating, the energy measurement is calibrated

using a multivariate algorithm trained on single-electron simulation to determine the most probable electron energy. The method takes into account differences between data and simulation in the energy scales of each longitudinal layer and other detector effects not modeled in simulation.

After the initial simulation-based calibration, the electron energy scale and resolution are determined using $Z \rightarrow ee$ events. Residual differences in the energy scale between data and simulation are parametrized as:

$$E^{\text{data}} = E^{\text{simulation}}(1 + \alpha_i), \quad (5.2)$$

where the α_i quantify the energy scale difference in bins of pseudorapidity. The difference in energy resolution is derived assuming that the $Z \rightarrow ee$ invariant mass distribution is well-modeled up to a Gaussian constant term,

$$\left(\frac{\sigma_E}{E}\right)^{\text{data}} = \left(\frac{\sigma_E}{E}\right)^{\text{simulation}} \oplus c_i, \quad (5.3)$$

where i again denotes bins in pseudorapidity. Histograms of the $Z \rightarrow ee$ invariant mass distribution in simulation are then produced for a range of α_i and c_i values, and the optimal values are determined using a χ^2 minimization with respect to data. The energy scale corrections are found to be $\alpha_i \sim -1\%$ in the barrel, $\sim 1\%$ - 2% in the barrel/end-cap transition region, and up to $\sim -4\%$ in the end-caps. The resolution corrections are symmetrized about $\eta = 0$, with values of $c_i \approx 0.8\%$ in the barrel, $\sim 3.3\%$ in the barrel/end-cap transition region, and $\sim 0.5\%$ - 2.5% in the end-caps. The energy resolution is shown as a function of E_T in figure 5.2.

Sources of systematic uncertainty on the energy scale include the intrinsic accuracy of the $Z \rightarrow ee$ method, determined from closure tests where known scale variations are injected into simulation, calorimeter gain and pedestal dependence, data-simulation differences in the calibration of the calorimeter layers, and the modeling of material in front of and within the calorimeter. The total uncertainty ranges from 0.03%-0.22% for $E_T = 40$ GeV and 0.27%-2.25% for $E_T = 200$ GeV, with larger uncertainties in the pseudorapidity range $1.37 < |\eta| < 1.82$ corresponding to the transition region between the barrel and the end-cap.

The systematic uncertainty on the energy resolution is less than 10% for electrons with $E_T < 50$ GeV, and asymptotically approaches $\sim 40\%$ for high E_T . At low energies, the pileup contribution to the noise term, b , dominates the uncertainty; at higher energy, uncertainty is due to a mix of the sampling term, a , the pileup contribution to b , the modeling of material, and the intrinsic accuracy of the $Z \rightarrow ee$ method.

5.3 Muons

Muons are identified by matching tracks in the muon spectrometer to tracks in the inner detector. The requirements differ based on the instrumentation available in the vicinity of the muon candidate. The analyses described here use *combined* (CB) muons, consisting of

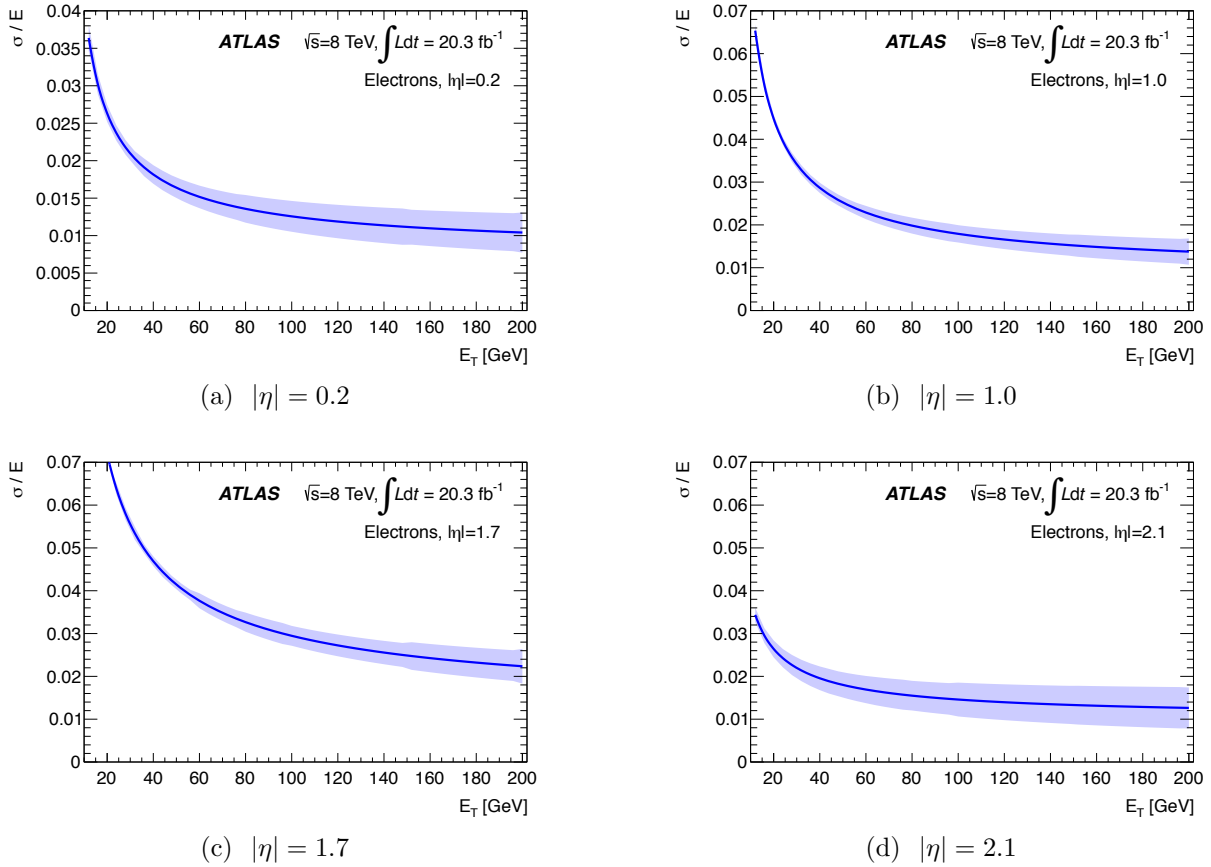


Figure 5.2: Electron energy resolution as a function of E_T for various values of $|\eta|$. The shaded band shows the uncertainty on the resolution [9].

matched tracks reconstructed independently in the inner detector and the muon spectrometer. The muon momentum is determined from a statistical combination of the two track's parameters and their corresponding covariance matrices. Combined muons have the highest purity, but suffer from a loss of acceptance near $\eta \sim 0$, where the muon spectrometer has gaps to accommodate services for the inner detector and calorimeters, and $1.1 < \eta < 1.3$, where some trajectories only pass through one muon station due to incomplete installation. The remaining categories are *standalone* (SA) muons, consisting of a track only in the muon spectrometer; *segment-tagged* (ST) muons, consisting of an inner detector track and one or more track segments in the MDT or CSC chambers; and *calorimeter-tagged* (CaloTag) muons, consisting of an inner detector track matched to a calorimeter energy deposit consistent with the passage of a muon. These categories recover efficiency in regions of the detector with less instrumentation at the cost of lower muon purity, and are not used in this dissertation.

For all categories of muons, the inner detector track is required to have at least 1 pixel

hit, at least 5 SCT hits, at most 2 pixel or SCT holes, and at least 9 TRT hits for $0.1 < |\eta| < 1.9$. Energy losses in the calorimeter due to ionization, bremsstrahlung, and electron pair production must also be taken into account.

Efficiency Measurements

The efficiency of the reconstructing CB muons is measured as:

$$\epsilon(\text{CB}) = \epsilon(\text{CB}|\text{ID})\epsilon(\text{ID}|\text{MS}), \quad (5.4)$$

where $\epsilon(\text{CB}|\text{ID})$ is the probability that a muon reconstructed as an inner detector track is also reconstructed as a CB muon, and $\epsilon(\text{ID}|\text{MS}) \approx \epsilon(\text{ID})$ is the probability that a muon with a track in the muon spectrometer, i.e. a CB or SA muon, is also reconstructed as an inner detector track. The latter approximation is made because $\epsilon(\text{ID})$ is not directly accessible in data.

The efficiencies are measured using tag-and-probe techniques similar to those described in section 5.2, targeting $Z \rightarrow \mu\mu$ and $J/\Psi \rightarrow \mu\mu$ events. In this case, the tag muon is required to be a CB muon, and the probe muon is a CB or SA muon in the case of measuring $\epsilon(\text{ID}|\text{MS})$, and a CaloTag muon for the measurement of $\epsilon(\text{CB}|\text{ID})$. The efficiencies for all types of muon are shown in figure 5.3. CB muons have an efficiency of greater than 97% in most of the pseudorapidity range, except for significant inefficiencies due to gaps in the muon spectrometer in the ranges $|\eta| < 0.1$ and $1.1 < \eta < 1.3$. The measured efficiencies in data and simulation agree to within $\sim 2\%$, and the ratios in each pseudorapidity bin are used as scale factors to correct the efficiency in simulation. The systematic uncertainty on the scale factors is in the range 0.1-0.3%, rising near $|\eta| \sim 0$ and $|\eta| \sim 2.5$.

Energy Scale and Resolution

The muon momenta in simulation are scaled and smeared to match the momentum scale and resolution in data. The corrections are derived in bins of η and ϕ , with boundaries chosen to minimize the variation of the correction in each bin, and are applied separately to the transverse momenta measured by the inner detector (ID) and the muon spectrometer (MS). Specifically, the correction is implemented as:

$$p_{\text{T}}^{\text{Cor,Det}} = \frac{p_{\text{T}}^{\text{MC,Det}} + \sum_{n=0}^1 s_n^{\text{Det}(\eta,\phi)} (p_{\text{T}}^{\text{MC,Det}})^n}{1 + \sum_{m=0}^2 \Delta r_m^{\text{Det}}(\eta, \phi) (p_{\text{T}}^{\text{MC,Det}})^{m-1} g_m}, \quad (5.5)$$

where Det=ID or MS, the $\Delta r_m^{\text{Det}}(\eta, \phi)$ parametrize the momentum resolution smearing, the $s_n^{\text{Det}}(\eta, \phi)$ parametrize the scale corrections, and the g_m are normally-distributed random variable with mean 0 and width 1⁴. The constant scale correction term, s_0^{MS} , accounts for

⁴Note that this equation does not apply to cases where the resolution in data is better than that in simulation. In these cases, the resolution difference is included in the positive ID and MS variations, and the effect of the positive variation on the physical observables is symmetrized about the nominal value.

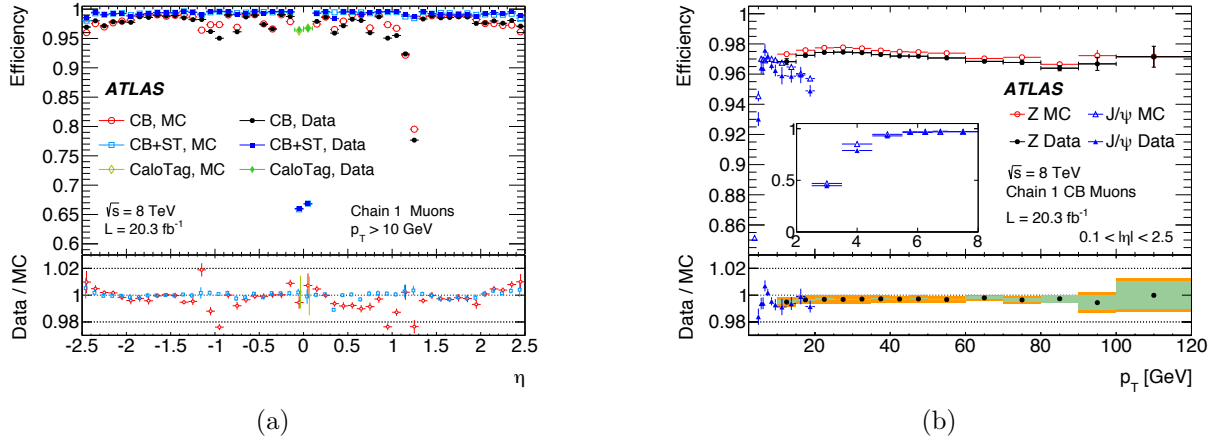


Figure 5.3: Muon reconstruction efficiencies as a function of η for muons with $p_T > 10$ GeV (left), and as a function of p_T for muons with $0.1 < |\eta| < 2.5$ (right). The uncertainty bars on the points indicate statistical uncertainties. The bottom plots show the ratio between the measured and simulated efficiencies, with the combination of statistical and systematic uncertainties indicated by the uncertainty bars.

the difference between data and simulation in the energy lost by muons before reaching the muon spectrometer. s_0^{ID} is set to zero due to the negligible energy loss before the inner detector. The linear terms, $s_1^{\text{ID,MS}}$, models discrepancies between data and simulation in the magnetic field integral and the radial dimension of the detector. The resolution corrections Δr_m^{Det} represent deviations from the resolution in data, which is parametrized empirically as:

$$\frac{\sigma(p_T)}{p_T} = \frac{r_0}{p_T} \oplus r_1 \oplus r_2 \cdot p_T, \quad (5.6)$$

where the r_0 term describes energy lost by muons as they traverse the material of the detector, the r_1 term describes multiple scattering, magnetic field inhomogeneities, and local radial displacements, and the r_2 term describes intrinsic resolution effects due to the spatial resolution of the hit measurements and residual misalignment.

The corrections are derived from $J/\Psi \rightarrow \mu\mu$, $\Upsilon \rightarrow \mu\mu$, and $Z \rightarrow \mu\mu$ events using a template maximum likelihood fit, similar to that described in section 5.2. The effect of the corrections on the invariant mass distribution of $Z \rightarrow \mu\mu$ events is shown in figure 5.4, along with the total systematic uncertainty. The muon resolution as measured in Z , Υ , and J/Ψ events is shown in figure 5.5.

5.4 τ Leptons

The signature of τ leptons is significantly more complex than electrons and muons due to the fact that they decay to a diverse set of final states. Tau leptons have a proper decay length

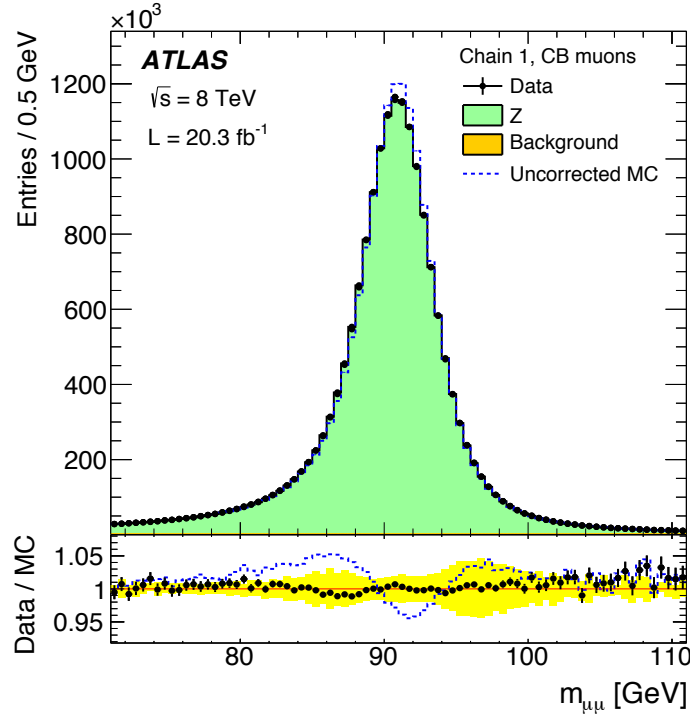
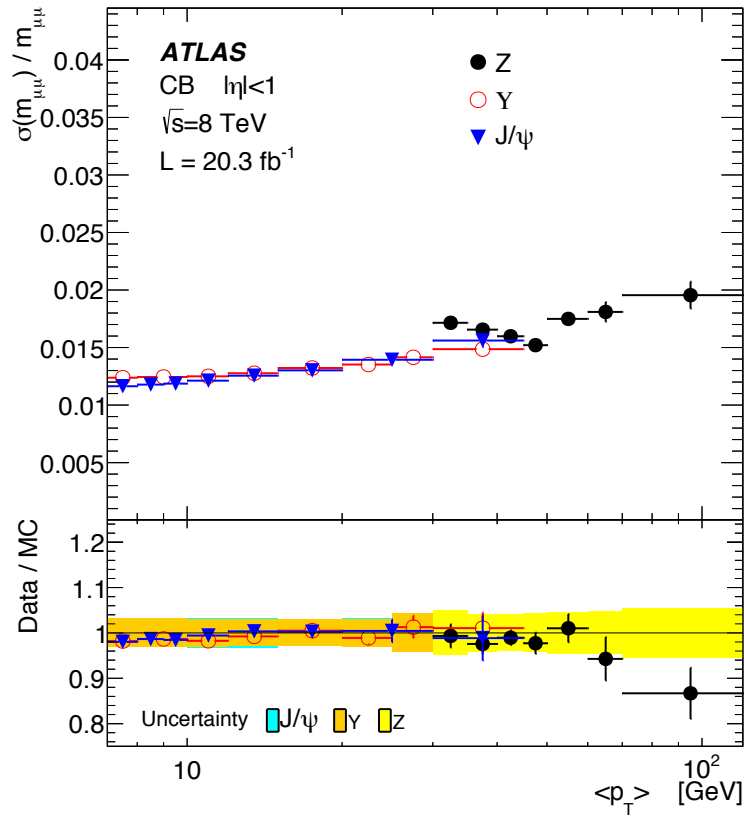
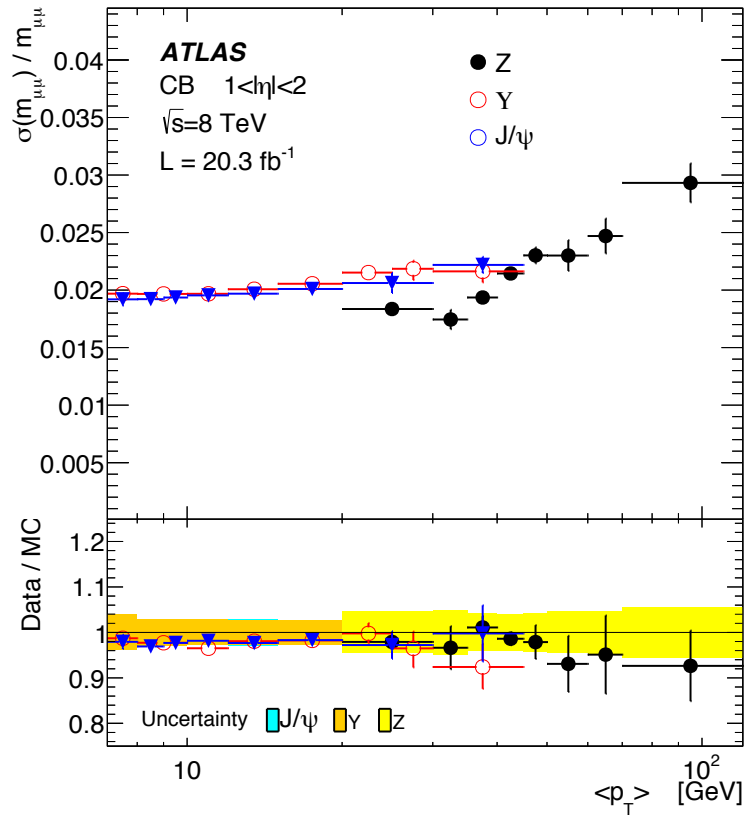


Figure 5.4: Dimuon invariant mass distribution of $Z \rightarrow \mu\mu$ events, using CB muons. The top panel shows the invariant mass distribution for data (black points), expected $Z \rightarrow \mu\mu$ signal from simulation (green), and expected background from simulation (yellow). The simulated background and signal have the muon momentum corrections applied, and their sum is normalized to the data. The dashed histogram shows the background plus signal without muon momentum corrections.

of $87\mu\text{m}$, and therefore typically decay before reaching the active layers of the detector. The leading decay modes are shown in table 5.1. 35.2% of τ leptons decay to an electron or muon plus two neutrinos. The remaining 64.8% decay to hadrons plus a neutrino. The hadronic decay modes contain one charged pion in 72% of the decays, and three charged pions in 22% of the decays; the majority of the remainder contain one or more kaons. The hadronic decay modes also frequently contain neutral pions, with 78% containing at least one neutral pion.

In this dissertation, no effort is made to identify or reconstruct leptonic τ decays (τ_{lep}), regarding them only as electrons or muons plus missing transverse energy (see section 5.6). Hadronic τ decays (τ_{had}) are identified using their visible decay products, namely the neutral and charged hadrons, which are collectively called $\tau_{\text{had-vis}}$. The signature consists of a narrow jet with one or three tracks, called one-prong or three-prong decays, respectively. Up to two $\pi^0 \rightarrow \gamma\gamma$ decays are also included. The reconstruction and identification proceeds as follows [11]:

- Jets are reconstructed using the anti- k_t algorithm with a distance parameter of $R =$

(a) $|\eta| < 1$ 

Decay	Branching Fraction [%]
$e^- \bar{\nu}_e \nu_\tau$	17.83 ± 0.04
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.41 ± 0.04
$\pi^- \nu_\tau$	10.83 ± 0.06
$\pi^- \pi^0 \nu_\tau$	25.52 ± 0.09
$\pi^- \pi^0 \pi^0 \nu_\tau$	9.30 ± 0.11
$\pi^- \pi^+ \pi^- \nu_\tau$	9.31 ± 0.06
$\pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	4.62 ± 0.06

Table 5.1: Leading branching fractions of the τ lepton to final states with leptons or pions. Most of the remaining decays are to final states with kaons. From [10].

0.4, built from TopoClusters calibrated with a local hadronic calibration. Jets with $p_T > 10$ GeV and $|\eta| < 2.5$ are used as seeds for τ lepton candidates.

- For each jet seed, the τ vertex is chosen to be the primary vertex with the greatest $\sum p_T$ of tracks in a cone of radius $\Delta R = 0.2$ around the jet seed. This vertex defines the $\tau_{\text{had-vis}}$ direction, i.e. is used to determine the η and ϕ of the τ candidate.
- π^0 candidates, consisting of a pair clusters within $\Delta R < 0.2$ of the τ candidate, are identified using a BDT-based algorithm. Up to two π^0 s are considered.
- A multivariate algorithm based on boosted decision trees (BDTs) is used to discriminate hadronic τ decays from the backgrounds, primarily due to jets with low track multiplicity. The BDTs use many input variables describing the energy cluster, the spatial arrangement and energy of the tracks, and the neutral pions. Separate BDTs are trained for one-prong and three-prong τ decays, using simulated $Z \rightarrow \tau\tau$, $W \rightarrow \tau\nu$, and $Z' \rightarrow \tau\tau$ decays for signal and collision data samples for the background. Three working points with different identification efficiencies are defined, called loose, medium, and tight. The performance of the identification algorithms is shown in figure 5.6, along with correction factors applied to simulated samples to equalize the efficiencies in simulation and real data. The correction factors are derived by measuring the efficiencies in data and simulation, using a tag-and-probe method targeting $Z \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ events.
- An additional BDT-based algorithm rejects one-prong τ candidates consistent with an electron. The most powerful discriminating variables are the ratio of high- to low-threshold TRT hits on the track and the ratio of energies deposited in the electromagnetic and hadronic calorimeters. The electron rejection power versus the τ efficiency, derived from simulated $Z \rightarrow ee$ events, is shown in figure 5.7.

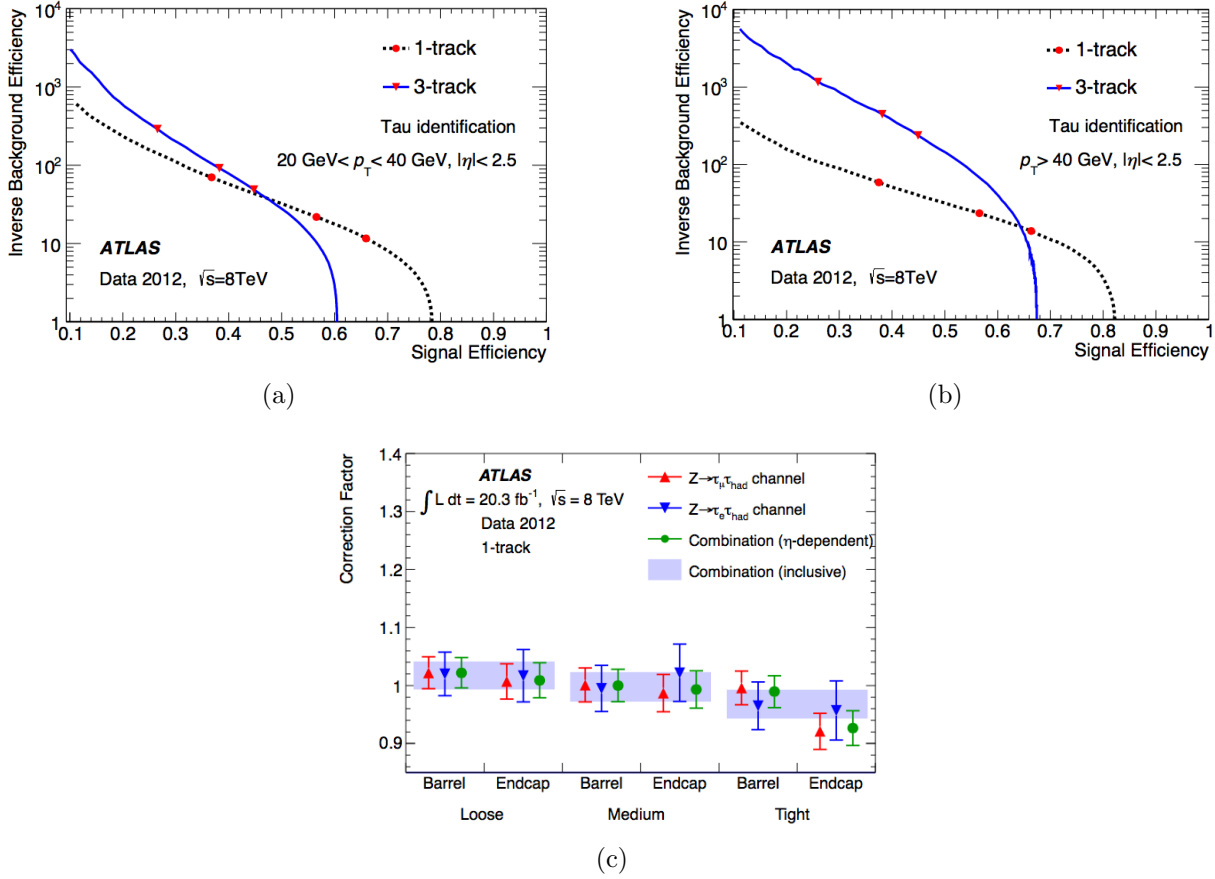


Figure 5.6: Top: inverse background efficiency (rejection power) versus signal efficiency for the BDT-based offline τ identification. Low p_T τ s ($20 \text{ GeV} < p_T < 40 \text{ GeV}$) are shown at left, and high p_T τ s ($p_T > 40 \text{ GeV}$) are shown at right. The background consists of simulated multijet events, while the signal consists of simulated Z , W , and Z' events decaying to τ leptons. Bottom: Correction factors applied to simulation to equalize the efficiency to that measured in data, as measured in Z tag-and-probe data.

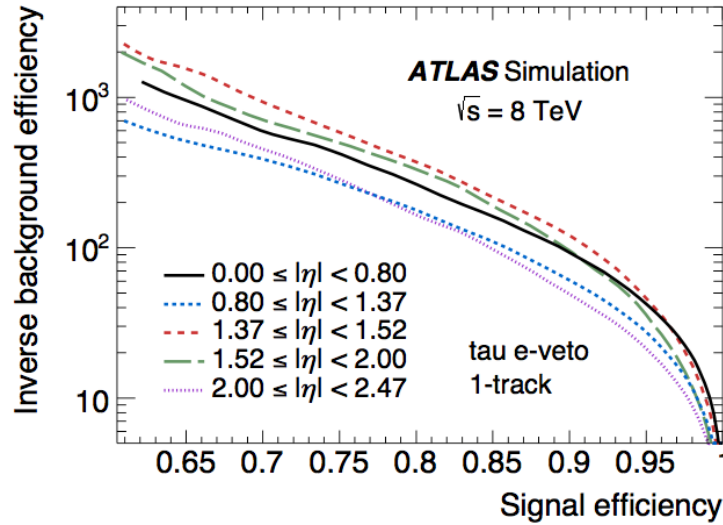


Figure 5.7: Electron rejection power versus 1-track τ_{had} efficiency for the electron rejection BDT.

Energy Scale and Resolution

The hadronic τ reconstruction and identification are based on calorimeter cells calibrated at the local hadronic scale. Several corrections are applied to correct the energy to a τ -specific energy scale (TES). First, corrections are derived using simulated $Z \rightarrow \tau\tau$, $W \rightarrow \tau\nu$, and $Z' \rightarrow \tau\tau$ events, generated with PYTHIA8. These corrections are determined as a function of the reconstructed $\tau_{\text{had-vis}}$ energy and pseudorapidity based on the medium identification working point. A small pseudorapidity correction, reaching up to $|\Delta\eta| = 0.01$, corrects a bias due to underestimated cluster energies in poorly-instrumented regions of the calorimeter. To account for pileup, 90-420 MeV per additional reconstructed vertex is subtracted from the reconstructed $\tau_{\text{had-vis}}$ energy, depending on η . The simulated $\tau_{\text{had-vis}}$ energy resolution after these corrections is shown in figure 5.8, and ranges between about 20% at very low energy to about 5% above a few hundred GeV.

Finally, data-driven corrections and systematic uncertainties on the τ energy scale are derived using a *deconvolution* method [ref-49-from-tau-paper]. The method combines the systematic uncertainties on the single-particle response of the calorimeters based on the well-known branching fractions of hadronically decaying τ s. A TES correction of about 1% is determined. Including additional systematic uncertainties covering the detector modeling, pileup, non-closure of the calibration method, and the hadronic shower model, the total TES uncertainty is 2 – 3% for one-prong decays and 2 – 4% for three-prong decays.

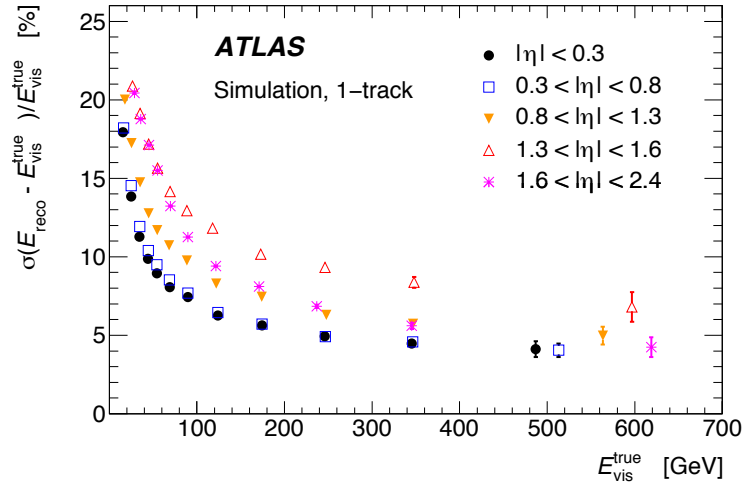


Figure 5.8: Energy resolution for hadronically decaying τ leptons with one associated track in various pseudorapidity regions. The resolution is the standard deviation of a Gaussian fit to the distribution of $\frac{E_{\text{reco}} - E_{\text{true-vis}}}{E_{\text{true-vis}}}$ in bins of $E_{\text{true-vis}}$ and $|\eta_{\text{true-vis}}|$.

5.5 Jets

Jets play an important complementary role to leptons in the analyses described in the following chapters. The new physics scenarios considered often produce jets in addition to the three required leptons. If the new particles are colored, as in strongly produced supersymmetry, then high- p_T jets can be produced in cascade decays; alternatively, if the new particles decay via the weak interaction, then the decays will often contain hadronically decaying weak bosons. From the background point of view, due to the colored initial state in pp collisions, jets are produced copiously at the LHC. Despite the stringent lepton identification cuts, jets misidentified as leptons or containing semileptonic decays can constitute a significant source of backgrounds to trilepton final states.

The reconstruction of jets in the calorimeter is based on topological clusters of energy [12, 13]. The reconstruction steps are shown in 5.9. Clusters are formed by grouping together calorimeter cells based on their signal-to-noise ratio, S/N , where the N includes electronic noise and contributions from pileup interactions. Cells with $S/N > 4$ form the cluster seeds, which are then expanded to include all connected cells with $S/N > 2$. Finally, cells along the perimeter with $S/N > 0$ are added to the cluster. The cluster energies are then calibrated according to two complementary scales, leading to two classes of jets: the electromagnetic (EM) scale assumes that the cell energy is due to an electromagnetic shower, while the local cell signal weighting (LCW) method classifies energy deposits as electromagnetic or hadronic in origin.

Jet finding algorithms then group the calibrated clusters into collections of jets. The algorithm used in this dissertation is the anti- k_t algorithm with a distance parameter of

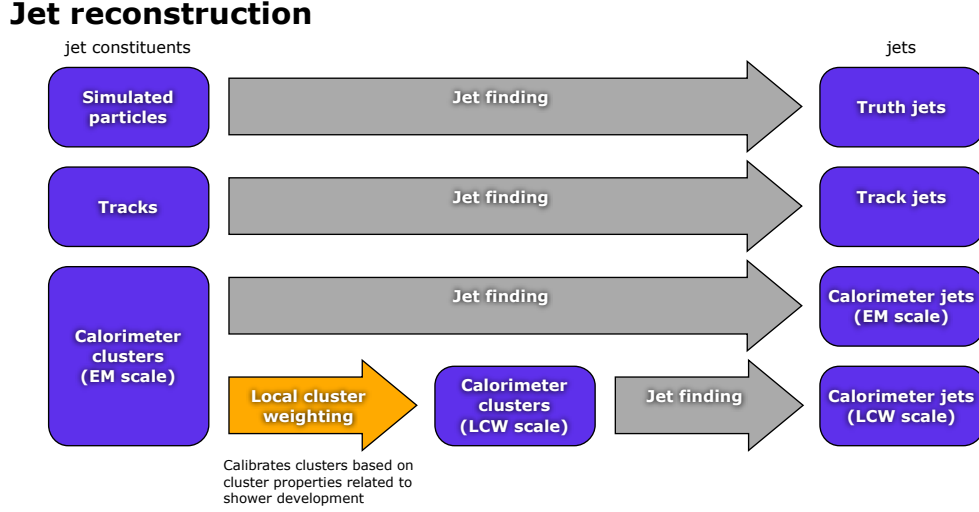


Figure 5.9: Overview of the jet reconstruction. The bottom two rows show the reconstruction of calorimeter jets, while the top two rows show jets built from truth particles in simulation and from tracks.

$R = 0.4$, implemented in FASTJET [Cacciari:2011ma, 14]. A distance measure between objects, $d_{i,j}$, is defined as

$$d_{i,j} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad (5.7)$$

$$d_{i,B} = k_{t,i}^{2p} \quad (5.8)$$

where $k_{t,i}$ is the transverse momentum of object i , $\Delta_{i,j} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ is the distance in rapidity ($y_{i,j}$) and azimuthal angle ($\phi_{i,j}$), and $p = -1$. $d_{i,B}$ denotes the distance between objects and the beam. The algorithm combines objects sequentially by considering the smallest distance in the event. If the smallest distance is a $d_{i,j}$, then objects i and j are combined. If the smallest distance is a $d_{i,B}$, then object i is classified as a jet and removed from the object list. The algorithm continues until there are no objects left.

Loose quality cuts are applied to reject events with jets due to non-collision sources, such as beam-gas collisions between protons in one beam and the residual gas in the beam pipe, beam-halo events due to collisions with upstream collimators, muons from cosmic rays, and calorimeter noise. The cuts have an efficiency of above 99.8% for retaining real jets.

Energy Scale and Resolution

The jet energy is calibrated in several steps, as shown in figure 5.10. First, pileup and origin corrections are applied. Energy contributions from pileup interactions in the same or nearby bunch crossings are subtracted based on simulation, using the number of reconstructed

vertices in the event, the expected average number of interactions per crossing, the jet area, and the median p_T density of jets in the event. The geometry of the jet is corrected to point from the primary event vertex, rather than the nominal center of the ATLAS detector.

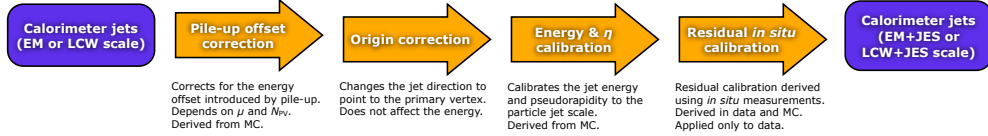


Figure 5.10: Overview of the jet calibration scheme.

The energy and pseudorapidity of the jet are initially calibrated based on the relationship between jets reconstructed from calorimeter clusters and jets reconstructed from truth particles in simulation. Then, in situ corrections are applied to the calibrated jets to correct for effects not described by the initial simulation-based calibration. The corrections are based on balancing the transverse momenta of jets against other well-calibrated objects. Jets in the central region, with detector rapidity $|\eta_{\text{det}}| < 1.2$, are calibrated against photons and leptonically decaying Z bosons. Jets with high transverse momentum, $p_T > 210$ GeV, are also calibrated using multijet events, balancing the high- p_T jet against several low- p_T jets. Jets with $1.2 < |\eta_{\text{det}}| < 2.8$ are calibrated against central jets with $|\eta_{\text{det}}| < 0.8$; due to a lack of statistics with which to derive the calibration, more forward jets, with $\pm 2.8 < \eta_{\text{det}} < 4.5$, are assigned the same calibration as jets with $\eta_{\text{det}} = \pm 2.8$. The in situ calibrations shift the jet energy in data down by about 2% for $p_T < 100$ GeV, and decreases with increasing p_T .

A large number of systematic uncertainties are assigned to the in situ calibration procedure, related to the modeling of the detector and physics processes in simulation and to the in situ methods themselves. The uncertainties are summarized in figure 5.11. Additionally, systematic uncertainties related to the pileup correction, high- p_T jets, jet flavor, and differences simulation settings are assigned where appropriate.

Pileup Suppression

Jets due to pileup interactions can be suppressed using the track information, at the cost of introducing some pileup-dependent jet reconstruction inefficiency. For a given jet and primary vertex, the *jet vertex fraction* (JVF) is defined as the ratio of the $\sum p_T$ of tracks associated with the jet and matched to the primary vertex to the $\sum p_T$ of all tracks associated to the jet, shown schematically in figure 5.12. Specifically, letting \mathcal{T} be the set of tracks associated to the jet, the JVF is defined as:

$$\text{JVF} = \frac{\sum_{i \in \mathcal{T}} p_{T,i} \Theta_V(i)}{\sum_{i \in \text{jet}} p_{T,i}}, \quad (5.9)$$

where $\Theta_V(i) = 1$ if the track i is matched to the primary vertex and 0 otherwise. If \mathcal{T} is empty, then $\text{JVF} \equiv -1$. In this dissertation, the primary vertex is chosen to have the highest $\sum p_T^2$ of tracks, and $\text{JVF} > 0.5$ is required for jets with $20 \text{ GeV} < p_T < 50 \text{ GeV}$.

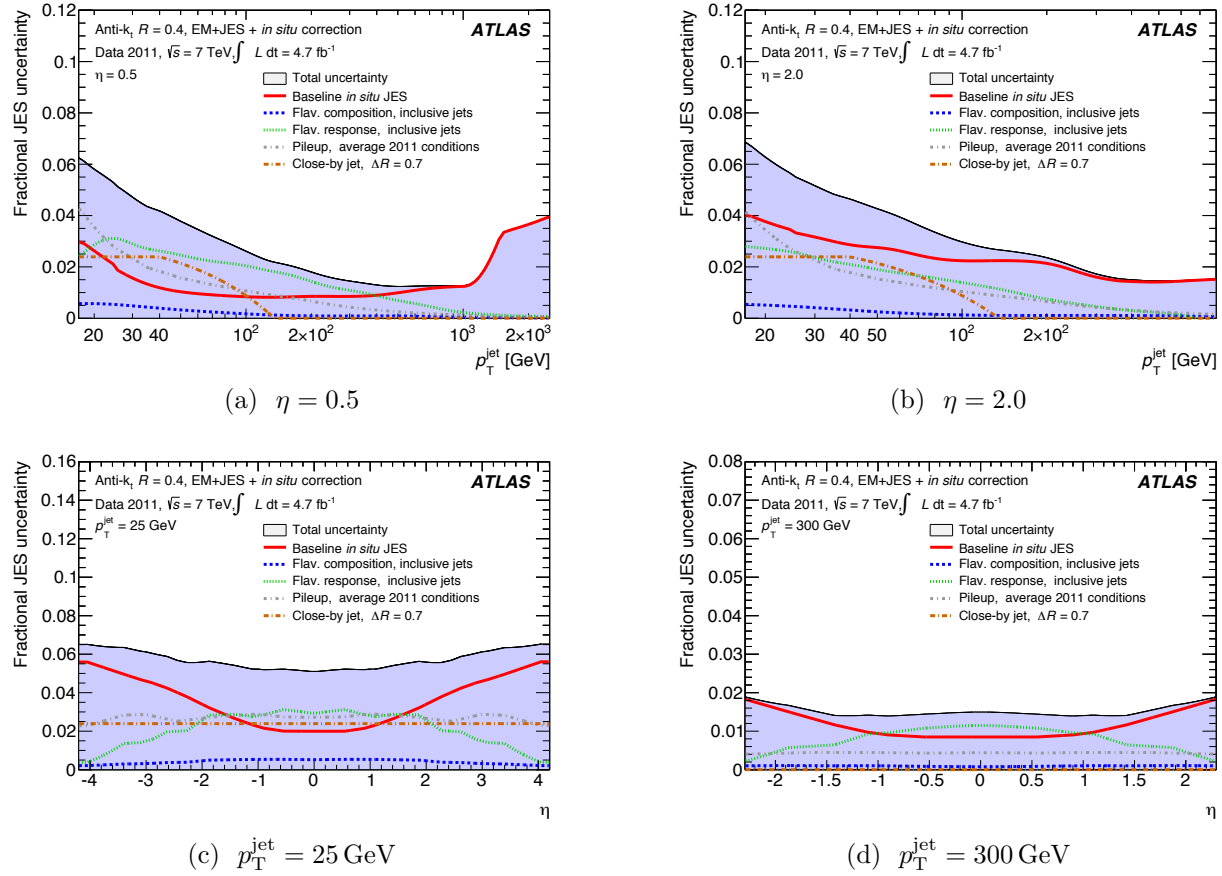


Figure 5.11: Sample-dependent fractional jet energy scale systematic uncertainty as a function of p_T^{jet} (top) and η (bottom). The jets are reconstruction using the anti- k_T algorithm with a radius parameter of $R = 0.4$ from clusters at the LCW scale, and are calibrated as described in the text.

b-tagging

The model-independent trilepton analysis (chapter 7) also makes use of the tagging of jets due to b quarks. Jets containing b quarks have several features which distinguish them from jets due to light quarks or gluons. B hadrons have proper decay lengths of approximate 0.5mm, long enough to be observed in the form of tracks or vertices reconstructed away from the primary interaction point. They also have large masses compared to other hadrons, leading to wider jets with more particles.

The algorithm used to tag b -jets is an artificial neural network called the MV1 algorithm [MV1]. The neural network uses three simpler likelihood-based algorithms as inputs [15, 16]:

- **IP3D**: Uses the transverse and longitudinal impact parameters of the tracks associated

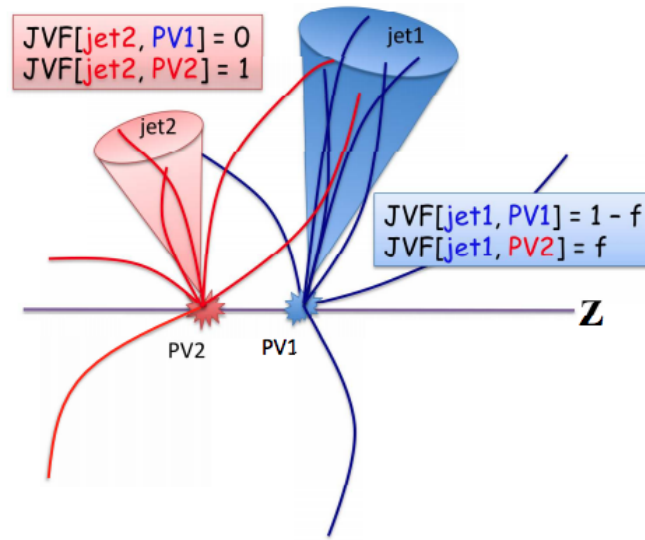


Figure 5.12: Schematic representation of the jet vertex fractions in the case of two jets and two primary vertices.

with a jet.

- **SV1:** Attempts to reconstruct a secondary vertex from the tracks associated with a jet. The most sensitive variable is the decay length significance between the secondary vertex and the primary event vertex. Additionally, the algorithm uses the invariant mass of tracks associated with the secondary vertex, the ratio of the energy of tracks assigned to the vertex to the energy of all tracks within the jet, and the number of two-track vertex candidates within the jet.
- **JetFitter:** Constructs a line connecting the primary vertex with one or more points associated with b - or c -hadron decays using a Kalman filter. The algorithm makes use of variables similar to the SV1 algorithm, along with the flight length significance between decays. Note that the algorithm does not require secondary vertices, allowing for the identification of decays with only one track.

The MV1 algorithm is trained on simulated data, using b -jets as signal and light quark jets as background, and returns a tag weight for each jet. The tag weight is used to establish working points with a given signal efficiency and background rejection power, as shown in figure 5.13a. The model-independent tripleton analysis uses the working point with 80% signal efficiency. The performance of the algorithm as a function of transverse momentum is shown in figure 5.13b.

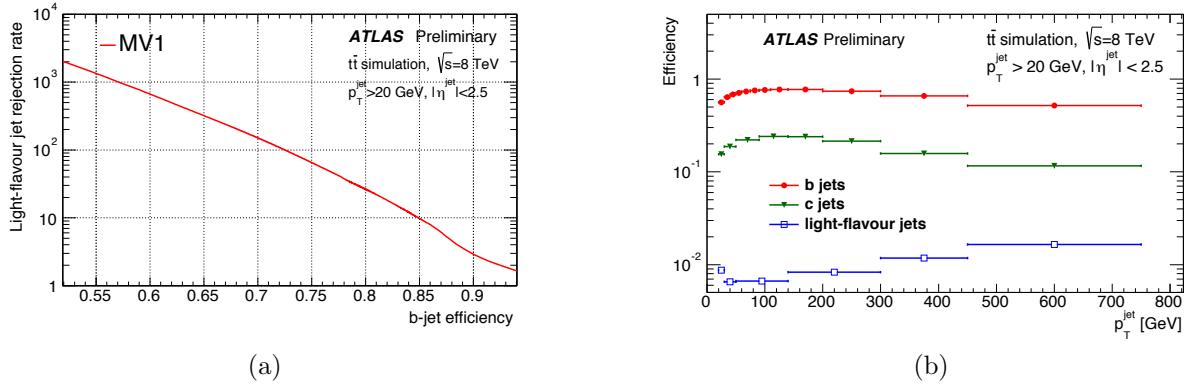


Figure 5.13: Performance of the MV1 b -tagging algorithm. The inclusive signal efficiency versus background rejection is shown at left. The signal and background efficiencies at the 70% working point are shown as a function of p_T at right.

5.6 Invisible Particles

Neutrinos interact only via the weak interaction, and hence escape the detector without interacting with any of the detector components. The same is true for any new stable, neutral, and colorless particles, such as the lightest supersymmetric particle in R parity-conserving scenarios or sterile neutrinos. The presence of such particles can only be inferred through the overall imbalance of the transverse momentum of the other visible collision products, $\mathbf{E}_T^{\text{miss}}$ ⁵.

The missing transverse momentum is defined as the negative vector sum of the visible objects in the event [TheATLASCollaboration:2012jy, TheATLASCollaboration:2013ua]. The energies are mostly determined from calorimeter measurements, except for muons. To avoid using calorimeter energy deposits multiple times, the algorithm assigns energy to deposits to a single object according to a strict order: electrons are identified first, followed by photons, hadronically decaying τ leptons, jets, and muons. Depending on the analysis requirements, the energy of a given physics object can be determined using the full reconstruction algorithm, or the energy deposits can be calibrated to the EM or LCW scales. Finally, topological clusters and tracks not assigned to physics objects are included in $\mathbf{E}_T^{\text{miss}}$ calculation as the so-called soft term. The missing transverse energy is given by:

$$E_{x(y)}^{\text{miss}} = - \left(E_{x(y)}^e + E_{x(y)}^\gamma + E_{x(y)}^\tau + E_{x(y)}^{\text{jets}} + E_{x(y)}^\mu + E_{x(y)}^{\text{soft}} \right), \quad (5.10)$$

where each term on the right represents the total momentum of the reconstructed objects in the x or y directions.

⁵The longitudinal momentum imbalance, E_z^{miss} , is not useful in pp collisions due to the unknown longitudinal momenta of the initial colliding partons.

The measured $\mathbf{E}_T^{\text{miss}}$ receives contributions from sources besides invisible particles, including calorimeter noise, particles falling in insensitive regions of the detector, energy mismeasurements, and pileup interactions. These effects are mitigable to varying degrees. Particles falling outside the detector acceptance contribute irreducibly to the $\mathbf{E}_T^{\text{miss}}$ resolution. For physics objects like electrons or jets, the noise and pileup contributions are suppressed by the reconstruction algorithms and identification cuts. To mitigate the noise and pileup contributions to the soft term, the cells are grouped into topological clusters. Rejecting events where the $\mathbf{E}_T^{\text{miss}}$ is parallel or antiparallel to the physics objects can mitigate the impact of energy mismeasurements.

5.7 Object Selection

The analyses described in this dissertation search for energetic leptons and jets produced the decays of new heavy particles. The leptons and jets typically have large transverse momentum and are well-separated from other objects in the event, due to the large difference in the mass scale between the final decay products (leptons or hadrons with mass $m \lesssim 10 \text{ GeV}$) and the parents (W/Z bosons or the new particles themselves, with mass $m \gtrsim 80 \text{ GeV}$). The new particles typically have short lifetimes, and hence the leptons and jets are produced promptly at the location of the initial proton-proton interaction, consistent with originating from the selected primary vertex. These properties can be used to suppress backgrounds due to Standard Model processes and detector effects, described in chapter 6.

Leptons

The events used in this dissertation are required to have at least three reconstructed electrons, muons, or hadronically decaying τ leptons, as described in section ???. A summary of the lepton selections is shown in tables 5.2 and ??. Leptons are required to satisfy the following requirements:

- **Transverse momentum:** Electrons and muons must have $p_T > 15 \text{ GeV}$, while hadronically decaying τ leptons must have $p_T > 20 \text{ GeV}$. The transverse momentum cut is driven by the availability of triggers with which to perform the data-driven reducible background estimate, described in section .
- **Geometrical acceptance:** Electrons are required to have $|\eta| < 2.47$, excluding the transition region $1.37 < |\eta| < 1.52$ between the barrel and end-cap calorimeters. Muons and τ leptons are required to have $|\eta| < 2.5$.
- **Particle identification:** To suppress the reducible backgrounds, the leptons must satisfy strict requirements related to particle identification, as described in section ??. Electrons candidates must satisfy the tight++ set of identification cuts. Electrons are neglected if they fall in a region affected by the presence of a dead front end board in

the first or second sampling layer, a dead high voltage supply, or a masked cell in the core. Muons are required to be *combined*, with associated hits in the inner detector and muon spectrometer. Specifically, associated inner detector track must have:

- A B-layer hit, unless the muon passes through a deactivated region of the B-layer.
- ≥ 1 pixel hit and ≥ 5 SCT hits, including any deactivated sensors along the trajectory.
- ≤ 2 total missing hits in the pixel and SCT, excluding deactivated sensors along the trajectory.
- A successful extension into the TRT, with ≥ 6 TRT hits, of which $< 90\%$ are classified as outlier hits.

Finally, τ leptons must satisfy the `BDT-tight` selection criteria.

- **Impact parameter:** The inner detector track associated with electrons and muons must be consistent with originating from the event primary vertex. The transverse impact parameter significance, defined as the transverse impact parameter d_0 divided by its uncertainty σ_{d_0} , is required to satisfy $\frac{d_0}{\sigma_{d_0}} < 3$. Similarly, the longitudinal impact parameter z_0 is required to satisfy $z_0 \sin \theta < 0.5$ mm. These requirements suppress leptons from semileptonic heavy flavor decays.
- **Isolation:** To further reduce the impact of non-prompt and misidentified leptons, the leptons are required to be isolated from other activity in the event. The cuts on electrons and muons are similar, and limit the amount of nearby activity as measured by inner detector tracks and calorimeter energy deposits:
 - For both electrons and muons, a cut is applied on `ptcone30`, the sum of transverse momenta of tracks associated to the same primary vertex as the lepton within a cone of $\Delta R < 0.3$.
 - For muons, a cut is applied on `Etcone30`, the scalar sum of transverse energies of calorimeter cells within $\Delta R < 3.0$ of the muon track.
 - For electrons, a cut is applied on `TopoEtcone30`, the sum of topological calorimeter clusters within a cone of $\Delta R < 3.0$. The use of topological clusters reduces the impact of pileup and out-of-cone leakage.

All isolation variables are required to be less than 10% of the lepton transverse momentum for leptons with $p_T < 100$ GeV, and less than $10 \text{ GeV} + 0.01 \times p_T$ for leptons with $p_T \geq 100$ GeV.

Cut	Electrons	Muons
Object ID	Tight++	Combined Tight
Leading (trigger) E_T/p_T	$E_T > 26$ GeV	$p_T > 26$ GeV
Subleading E_T/p_T	$E_T > 15$ GeV	$p_T > 15$ GeV
Trigger Acceptance	$(\eta < 2.47) \&\& !(1.37 < \eta < 1.52)$	$ \eta < 2.4$
Acceptance	$(\eta < 2.47) \&\& !(1.37 < \eta < 1.52)$	$ \eta < 2.5$
Calo. Isolation ($E_T/p_T < 100$ GeV)	$\text{TopoEtcone30} < 0.1 \times E_T$	$\text{Etcone30} < 0.1 \times p_T$
Calo. Isolation ($E_T/p_T > 100$ GeV)	$\text{TopoEtcone30} < 10 \text{ GeV} + 0.01 \times E_T$	$\text{Etcone30} < 10 \text{ GeV} + 0.01 \times p_T$
Track Isolation ($E_T/p_T < 100$ GeV)	$\text{ptcone30} < 0.1 \times E_T$	$\text{ptcone30} < 0.1 \times p_T$
Track Isolation ($E_T/p_T > 100$ GeV)	$\text{ptcone30} < 10 \text{ GeV} + 0.01 \times E_T$	$\text{ptcone30} < 10 \text{ GeV} + 0.01 \times p_T$
Track d_0	$\frac{d_0}{\sigma_{d_0}} < 3$	$\frac{d_0}{\sigma_{d_0}} < 3$
Track z_0	$z_0 \sin \theta < 0.5 \text{ mm}$	$z_0 \sin \theta < 0.5 \text{ mm}$

Table 5.2: Electron and muon selection criteria.

Cut	Taus
Object ID	BDT Tight
p_T	$p_T > 26$ GeV
Acceptance	$ \eta < 2.5$

Table 5.3: Tau selection criteria.

Jets and Missing Transverse Energy

Jets are required to have $p_T > 30$ GeV, in order to limit the presence of pileup jets. For the geometrical acceptance, jets must lie in the range $|\eta| < 4.5$, so that the jet falls within instrumented regions of the detector. Pileup jets are additionally suppressed with a cut on the jet vertex fraction: the $\sum p_T$ of tracks within the jet cone and associated with the selected primary vertex must be at least 50% of the $\sum p_T$ of all tracks within the jet cone [jvf].

Jets consistent with originating from the decay of a b -hadron are identified using the MV1 algorithm [MV1], with an efficiency of 80%.

The missing transverse momentum, E_T^{miss} , is calculated using the MET_Egamma10NoTau_RefFinal algorithm. Calorimeter cells associated with electrons or photons with $p_T > 10$ GeV are calibrated specifically to that object; cells associated with τ leptons are not calibrated as τ leptons due to the change in the energy calibration for the objects used in the data-driven reducible background estimate 5.7.

Overlap Removal

Objects are frequently reconstructed as multiple objects; for example, a muon with a hard bremsstrahlung emission might be reconstructed as a muon, an electron, and a jet. In order to resolve ambiguities, the following overlap removal procedure is applied:

- If $\Delta R(e, e) < 0.1$, remove lower p_T electron, to avoid “a potential bias in the simulation of the reconstruction efficiency for two real, close-by same-flavour leptons” [17].
- If $\Delta R(e, \text{jet}) < 0.2$, remove jet. This addresses the ambiguity between electrons and jets.

- If $0.2 < \Delta R(\text{jet}, e) < 0.4$ AND $p_T(\text{jet}) > 30 \text{ GeV} + 0.05 * p_T(e)$, remove electron. This reduces the reducible electron backgrounds.
- If $\Delta R(\mu, e) < 0.1$, remove electron. This addresses cases where a muon radiates a hard photon, which is then identified as an electron.
- If $\Delta R(\mu, \text{jet}) < 0.1$, and:

$$\begin{aligned} p_T^{\text{jet}} < 0.5 p_T^\mu & : p_T^\mu < 200 \text{ GeV, or} \\ p_T^{\text{jet}} < 100 \text{ GeV} & : p_T^\mu \geq 200 \text{ GeV,} \end{aligned} \quad (5.11)$$

remove the jet. This is intended to reduce efficiency loss (in the next bullet point) from jets induced by muons at high muon p_T .

- If $\Delta R(\text{jet}, \mu) < 0.3$, remove muon. This reduces the reducible muon backgrounds.

Bibliography

- [1] T. Cornelissen et al.,
Concepts, Design and Implementation of the ATLAS New Tracking (NEWT) (2007).
- [2] The ATLAS Collaboration, *Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at $\sqrt{s} = 7$ TeV* (July 2010) pp. 1–14,
URL: <https://cds.cern.ch/record/1281363>.
- [3] The ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment* (Mar. 2012) pp. 1–20.
- [4] The ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment* (July 2010) pp. 1–10, URL: <http://cds.cern.ch/record/1281344>.
- [5] W. Waltenberger, R. Frühwirth, and P. Vanlaer, *Adaptive vertex fitting*,
Journal of Physics G: Nuclear and Particle Physics **34**.12 (Nov. 2007) N343–N356,
URL: <http://stacks.iop.org/0954-3899/34/i=12/a=N01?key=crossref.1572e0c569f8431ab1d721e989956caa>.
- [6] T. A. Collaboration,
Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data,
The European Physical Journal C **74**.7 (July 2014) pp. 2941–38.
- [7] The ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data* (June 2014) pp. 1–50.
- [8] The ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung* (May 2012) pp. 1–22.
- [9] T. A. Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*,
The European Physical Journal C **74**.10 (Oct. 2014) pp. 3071–48.
- [10] K. Olive et al., *Review of Particle Physics*, Chin.Phys. **C38** (2014) p. 090001.
- [11] The ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV*,
arXiv.org (Dec. 2014), arXiv: 1412.7086 [hep-ex].

- [12] The ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, English, arXiv.org 3 (Dec. 2011) pp. 2304–118, arXiv: 1112.6426 [hep-ex], URL: <http://link.springer.com/10.1140/epjc/s10052-013-2304-2>.
- [13] T. A. Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, English, The European Physical Journal C **75.1** (Jan. 2015) pp. 17–101, URL: <http://link.springer.com/10.1140/epjc/s10052-014-3190-y>.
- [14] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, arXiv **0804.04** (2008) pp. 063–063, URL: <http://stacks.iop.org/1126-6708/2008/i=04/a=063?key=crossref.2ffef838d94c2b899ee27950c18ee8d2>.
- [15] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics* (July 2009) pp. 1–1852.
- [16] The ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data* (July 2011) pp. 1–22, URL: <https://cds.cern.ch/record/1369219>.
- [17] D. L. Adams et al., *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014* (July 2014) pp. 1–41, URL: <https://cds.cern.ch/record/1700874>.