Top quark pair production cross-section in proton-proton collisions at √s = 7 TeV

Tsiakiris, M.

Publication date
2012

Chapter 4

Reconstruction and selection of top quark-pairs.

"High thoughts must have high language."

Aristophanes - c. 446 - c. 386

In chapter 2 the ATLAS detector layout and its readout system were described. Here we focus on those elements that are necessary for detecting top quark-pair events ($t\bar{t}$) where one top quark decays to a leptonic final state and the other hadronically to jets (single-lepton). The $t\bar{t}$ single-lepton topology is a busy and complex signature with many different physics objects that need to be reconstructed. As a result, almost all the different sub-detectors of ATLAS are involved for detecting these interesting events. In this chapter, section 4.1 presents the reconstruction techniques of the objects involved in the $t\bar{t}(e\nu_e)$ and $t\bar{t}(\mu\nu_\mu)$ topologies, namely those of electrons, muons, missing transverse energy, and jets[1]. Section 4.2 discusses corrections that are applied on reconstructed observables for improving the agreement between data and Monte Carlo. The sections 4.3 and 4.4 present the requirements that are applied at both the physics object-level and the event level for selecting the $t\bar{t}$ single-lepton events for our analysis. Lastly, section 4.5 introduces us to the cross-section measurement that is presented in greater detail in the forthcoming chapters.

4.1 Reconstruction of physics objects

This section describes how the raw detector signals are transformed by specific algorithms to meaningful physics objects. The calibration of those objects is also discussed. For various objects, such as jets, several reconstruction possibilities exist. In these cases, the focus is on the choices made for the analysis presented in the later chapters.

[1]The $t\bar{t}(\tau\nu_\tau)$ final state where the τ -lepton decays hadronically is not considered explicitly as it is a special case when it comes to isolating candidate events. On the other hand, the case where the τ-lepton decays into an electron or a muon is practically indistinguishable from the actual $t\bar{t}(e\nu_e)$ or $t\bar{t}(\mu\nu_\mu)$ final states and enters the signal distributions even though kinematically they may differ due to the presence of the produced neutrino. Naturally, these events are not excluded.

4.1.1 Electrons

Reconstruction

For electron detection both the Inner Detector (ID) and the Electromagnetic (EM) calorimeter are used. The standard reconstruction starts from the EM calorimeter where a sliding window algorithm identifies clusters of cells[2] with a size of 3×5 cells and with at least 2.5 GeV of deposited energy (seed clusters). The seed clusters are then matched to tracks from the Inner Detector forming electron candidates. Multiple tracks may be matched to the seed cluster and in this case the best-match is selected as the one with the smallest distance in $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$; priority is given to tracks that have silicon hits. For each electron candidate the initial seed cluster is again scanned with a sliding window algorithm of size 3×5 cells for the barrel region, or 5×5 cells for the end-cap region. Finally, the electron four-momentum is calculated from the combination of the best-match track and the reconstructed cluster, the energy is recomputed as the weighted average between the cluster's energy and the track's momentum while the η and ϕ coordinates are taken from the track information. Specifically for lower-p_{T} electrons a better reconstruction efficiency is achieved if the track is used as the starting point. In this case, the track is extrapolated to the EM calorimeter and its entry point is used as the seed for the cluster reconstruction. Both algorithms are restricted by the pseudo-rapidity coverage of the ID ($|\eta| \leq 2.5$).

Identification requirements

The reconstruction algorithms are very likely to provide electron candidates which should be accounted as background instead, e.g. converted photons or jets (faking electrons). Depending on the analysis, a refinement of the objects is necessary with the goal to decrease the background rate while keeping real electrons. Three sets of identification criteria are established which define the following electron collections (a more detailed description is presented in [128]):

- The *"loose"* electrons which are identified using calorimeter information only. The energy deposition of the electromagnetic shower on the first layer of the hadronic calorimeter (hadronic leakage) is evaluated as well as the shower-shape from the second layer of the EM calorimeter.
- The *"medium"* electrons which require both calorimeter and tracking information. Selection cuts are applied on top of the "loose" information and utilize also the first layer of the calorimeter for shower-shape estimates as well as ID track quality and matching criteria.
- Finally the *"tight"* electrons keep the same calorimeter cuts as for the "medium" electrons but require tighter track matching and quality cuts as well as a number of hits on the Pixel B-Layer and the TRT sub-detectors. In addition, electrons that are matched to converted photons are also rejected.

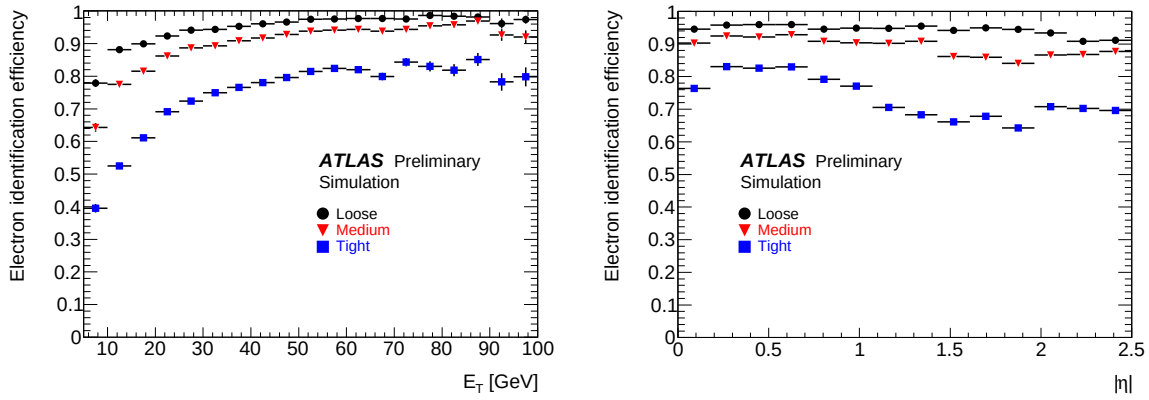
In figure 4.1 the binned efficiency is shown for each selection, as calculated from $Z \to e^+e^-$ simulated decays, with respect to the cluster's E_T and $|\eta|$. The efficiency is defined as the ratio of reconstructed signal electrons which pass the selection criteria and are matched to true electrons over the total number of true electrons. The apparent drop at $|\eta| > 1$ is attributed to the reduced tracking efficiency in the end-cap regions, where particles travel through more dead material, as well as to the lower efficiency in the transition region between barrel and

[2]The size of a cell is that of the middle layer cell and corresponds to $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

	Efficiency (%)		Jet Rejection
	$E_T \geq 17$ GeV	$E_T \geq 20$ GeV	
	$t\bar{t}(e\nu_e)$	$Z \to e^+e^-$	Di-jet QCD
Loose	89.73 ± 0.04	94.68 ± 0.03	614.3 ± 1.5
Medium	87.20 ± 0.05	89.61 ± 0.03	4435 ± 30
Tight	75.26 ± 0.06	72.77 ± 0.03	$(4.9 \pm 0.1) \cdot 10^4$

Table 4.1: *The expected integrated efficiency for "loose", "medium" and "tight" electron selections for the $Z \to e^+e^-$ and $t\bar{t}$ events as estimated from Monte Carlo. The E_T cut applied in each of the MC samples is slightly different. The expected jet rejection rate is shown in the last column. Numbers taken from [128].*

end-cap of the EM calorimeter [128].



(a) *Efficiency vs.* E_T (b) *Efficiency vs.* $|\eta|$

Figure 4.1: *Electron identification efficiency with respect to the cluster energy (a) and the absolute pseudo-rapidity values (b) for "loose", "medium" and "tight" selections as estimated from Monte Carlo. The efficiencies are calculated from a $Z \to e^+e^-$ sample using the tag-and-probe method. Plots taken from [128].*

Electrons originating from the Z boson decay are high-E_{T} and well isolated, similar to the electrons from the $t\bar{t}(e)$ decay. However, $t\bar{t}$ events have a large jet multiplicity and thus they are considered a more 'busy' environment. As a result, the probability of identifying a jet as an electron is increased. Table 4.1 shows the overall efficiency for the Z boson and the $t\bar{t}(e\nu_e)$ samples as well as the jet rejection rate which is calculated from QCD di-jet events. The jet rejection rate is defined as the ratio of the number of true jets in the sample over the number of electron candidates which are reconstructed and pass the selection criteria (fakes).

Isolation variables

Electron isolation cuts are explicitly added in the selection procedure if necessary. Two variables are defined, `etcone` and `ptcone`, which correspond to calorimetric and tracking isolation respectively [128]:

- The `etcone` corresponds to the reconstructed transverse energy on the calorimeter in a cone of opening angle of R_0 around the electron axis direction after excluding the intrinsic energy of the electron. The R_0 typical values, measured in the $\eta - \phi$ plane,

range from 0.2 to 0.6. Naturally, the `etcone` variables are dependent on the p_{T} of the object and in addition they are affected by the pile-up and the underlying event activity. For these effects corrections can be applied as explained in [129].

- The `ptcone` corresponds to the scalar sum of transverse momentum of the tracks which are within a cone of $R_0 = 0.3$ around the electron. Specific requirements are placed for the tracks used in this case. Namely, all tracks must have $p_{\mathrm{T}} \geq 1$ GeV, a hit at the inner layer of the Pixel detector as well as 7 hits on the SCT sub-detector. Lastly, both transverse and longitudinal impact parameters must have a value of at most 1 mm. Naturally, for this calculation neutral particles are not taken into account.

4.1.2 Muons

Reconstruction

The reconstruction of muons involves all layers of the detector. The Muon Spectrometer (MS) is the dedicated sub-detector for muon-finding but information from the Inner Detector (ID) as well as from the calorimeters is also exploited. Muon candidates are formed by any of the following identification strategies:

- *Standalone* muons (SA), which are reconstructed using only MS information and their track is extrapolated to the beam line in order to identify their trajectory and their impact parameter.
- *Combined* muons (CB), which are formed by combining the independently calculated tracks from the MS and the ID.
- *Segment-tagged* (ST) muons, which are ID-reconstructed muons for which the track is extrapolated to the MS and subsequently matched with straight-track segments.
- *Calorimeter-tagged* (CT) muons, which are ID-reconstructed muons that are extrapolated to the calorimeters.

For each of the above categories various algorithms are employed which are not necessarily mutually exclusive. Two major reconstruction chains exist in ATLAS, commonly referred to as 'collections', the `STACO` collection (or chain 1) and the `MuID` collection (or chain 2). For the analysis presented in this thesis the `MuID` collection is preferred as it provides a slightly better overall efficiency for high-p_{T} muons which is also relatively constant in each of the different pseudo-rapidity regions of the detector [130, 131].

In the `MuID` collection the SA muons are found by the `MOORE` [120] algorithm and the extrapolation of their tracks to the vertex is handled by `MuIDStandalone` [132]. Forming of CB muons is performed by `MuIDCombined` [132] where a global refit of ID tracks from `NewTracking` and SA tracks from `MOORE` is done. Segment-tagged muons are also reconstructed from two algorithms, namely the `MuGirl` [133] where a full refit is done between the seed ID track and the associated SA track, and the `MuTagIMO` which also uses ID track seed and associates it with `MOORE` segments.

The overall reconstruction efficiency for CB muons is measured from data with a tag-and-probe technique based on $Z \to \mu^+\mu^-$ events and is found to have an average value of $0.958 \pm 0.001\%$ [131]. The deviation of the efficiency from the ideal 100% is largely due to the Muon Spectrometer acceptance which is not uniform in $\eta - \phi$. The various detector regions are depicted in figure 4.2. In figure 4.3 the reconstruction efficiency is shown for CB muons for the various regions as well as the dependence with respect to the pseudo-rapidity. Clearly,

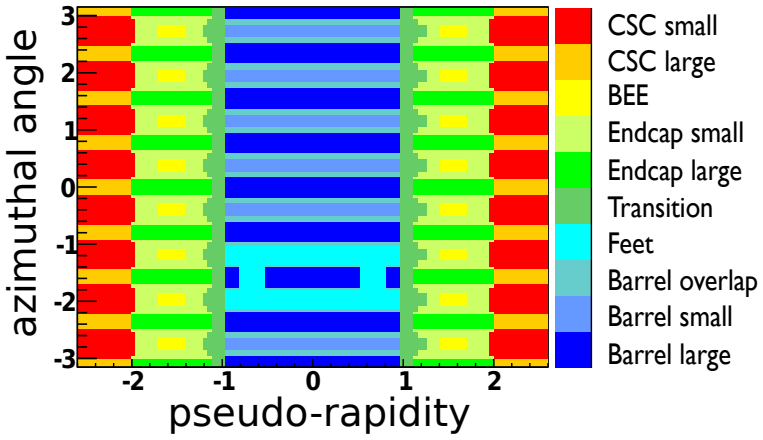
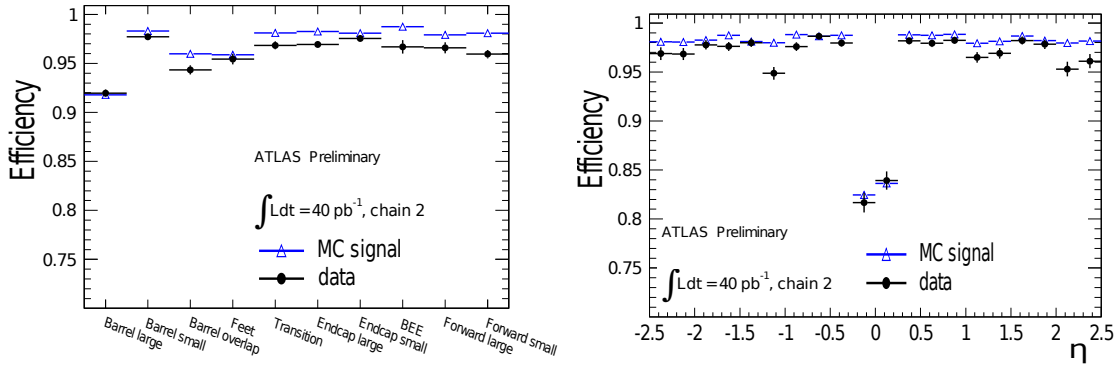


Figure 4.2: *The different detector regions with respect to $\eta - \phi$ coordinates.*



(a) *Efficiency vs. detector regions* (b) *Efficiency vs. η*

Figure 4.3: *The reconstruction efficiency (a) for each of the defined MS detector regions (b) as measured from $Z \rightarrow \mu^+\mu^-$ events using the tag-and-probe method. The muons probed are required to have $p_{\mathrm{T}} \geq 20$ GeV and lie within $|\eta| \leq 2.5$. The data are corrected for the background contribution. Plots taken from [131].*

the largest inefficiencies come from the 'large-barrel' stations, the 'feet' region and the 'barrel-overlap' region. The large dip in the 'large-barrel' region is because it covers the $\eta \approx 0$ area where a detector cabling hole exists, the dip in figure 4.3(b) justifies this effect; the 'barrel-overlap' region is also affected by the cabling hole. Similarly, the 'feet' region has less muon chambers in order to accommodate the structure for the detector's support, hence the quality of reconstruction is worse.

Improvements on the overall efficiency are achieved if ST muons are used together with CB muons [131, 105] since the former do not require a complete MS track and thus can increase the amount of successfully reconstructed muons. However, the trade-off for using ST algorithms is that the amount of non-prompt muons, e.g. muons coming from kaons or pions decaying in-flight, is significantly increased [105].

Quality definitions

Similar to the electron case, muons are categorized with respect to the quality of the reconstruction with a set of predefined selection requirements based on the algorithms used for the reconstruction. For muons in the `MuID` collection the categories are [134]:

- The *"tight"* muons, which are all the `MuID` combined muons and the `MuGirl` muons with

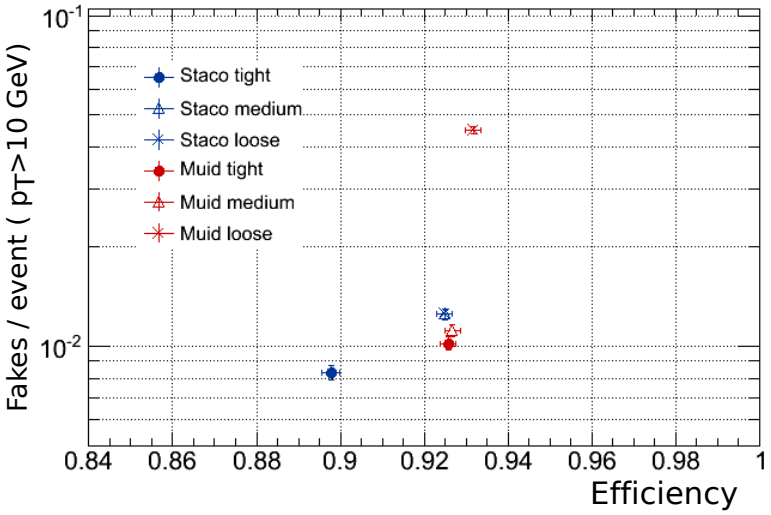


Figure 4.4: *The muon fake rate per event (purity) with respect to the efficiency for the various MuID (red) and STACO (blue) quality definitions as calculated for muons with* $p_{\mathrm{T}} \geq 10$ GeV. *Plot from [136].*

a successful combined fit.

- The *"medium"* muons, which are all muons with a successful standalone fit.
- The *"loose"* muons, which are all the muons found by segment-tagging algorithms and have Inner Detector tracks with silicon hits associated.
- Lastly, the *"very loose"* muons in which all muons that are reconstructed by `MuTagIMO` with only TRT seeds are included.

The requirements are inclusive for increasing tightness of the reconstruction quality e.g. "loose" muons contain "medium" and "tight" muons as well. In figure 4.4 the efficiencies of the quality definitions for `MuID` (red) are shown with respect to the relevant fake rate. It is clearly seen that for tighter requirements the fake rate is significantly reduced with a small trade-off to the overall efficiency. For completeness the `STACO` (blue) collection categories are also seen, for which the definitions are described in [135].

Isolation variables

As in the electron case, the calorimetric (`etcone`) and tracking (`ptcone`) isolation parameters are defined for muons [131]. The definitions of the variables are the same as before. The `etcone` variable gives the transverse energy deposited in the calorimeter in a cone of opening angle of 0.3 or 0.4 around the muon and after correcting for the muon energy loss. The `ptcone` variable is the scalar sum of the transverse momentum of particle tracks that reside within a cone of 0.3 or 0.4 around the muon axis.

4.1.3 Jets

At the LHC conditions the proton collisions result in an immense production of quarks and gluons which either come from the hard scattering or from the initial and final state radiation. These particles carry color charge and as a result of color confinement they cannot exist freely. Subsequently they hadronize[3] into a bunch of colorless and collimated stream of particles, usually hadrons, which can simply be seen as a jet.

For the identification of $t\bar{t}(\ell\nu_\ell)$ events, the jets are of particular importance since they can be used to directly reconstruct the hadronically decaying top quark mass. Detection

[3] With the exception of the top quark which decays much faster.

and reconstruction of jets is a challenging process that involves mainly the hadronic and electromagnetic calorimeters and from an experiment's point of view it can be split into three major ingredients: the input to the jet reconstruction, the jet finding (clustering) algorithm and the calibration that is used to correct the final jet energy back to the parton level.

Inputs to jet reconstruction

The inputs to jet reconstruction are the four-momentum elements that are used by a clustering algorithm to form a jet. Although at the Monte Carlo level it is straightforward to use stable particles, [4] with real data this is not possible and the calorimeter energy depositions must be utilized. Two main options exist in ATLAS, the *topological clusters* and the *calorimeter towers* [137]. Calorimeter towers are constructed by splitting the calorimeter in bins of fixed size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. All cells that lie within a bin and for the whole depth of the calorimeter are combined together to form a single four-vector object. The amount of energy that each cell contributes depends on the fraction of the cell that falls within the bin. The reconstructed energy at this level is at the electromagnetic scale, that is to say it corresponds only to the energy deposited from electromagnetic showers. Cells with no signal also take part in the reconstruction and therefore calorimeter towers maybe formed by using also noisy cells[5]. Given that no noise-suppression is applied, calorimeter towers can often result into having negative energies. When a negative energy tower is built it is typically combined with neighboring positive energy towers in order to cancel the effect. However, it is not unlikely especially in the presence of pile-up to have towers with significantly large negative energy where the above procedure is ineffectual and thus not being able to avoid the presence of negative energy towers.

A topological cluster on the other hand takes advantage of the fine granularity of the detector. It is a dynamically growing cluster of cells that is formed starting from a seed cell with a high signal-to-noise ratio. Once a cell is found that fulfills the requirement of $|E_{cell}|/\sigma_{cell}^{noise} > 4$ it serves as a seed, where the σ_{cell}^{noise} depends on the exact position of the cell in the calorimeter. At the next step, all cells neighboring the seed are added if they have $|E_{cell}|/\sigma_{cell}^{noise} > 2$. This continues for all neighbors of the neighbors as long as the requirement is fulfilled every time. The growing of the cluster stops when no cell passing the $2\sigma_{cell}^{noise}$ excess is found, where at this point all cells surrounding the cluster are also added. By construction, topological clusters are a noise-suppressed input since by taking the absolute value requirement the positive noise is on average cancelled by the negative noise effects. Thus this method is inherently dependent on the noise modeling of the detector. However, negative energy clusters are possible to emerge especially in the presence of large noise fluctuations. In this case those clusters are completely withdrawn from the reconstruction.

In an alternative approach a combination of the above two techniques can be made forming topological towers [138]. Topological towers are again tower objects of size $\Delta\eta\times\Delta\phi = 0.1\times0.1$ where only cells that belong to a topological cluster are combined. The benefit of this method is a noise-suppressed input that keeps a fixed geometrical size and therefore does not rely on the cluster-making algorithm.

[4] Typically, all particles with a decay length of more than 10mm excluding neutrinos and muons that result as decay products after the hadronisation of the partons.

[5] Calorimeter noise comes from the inherent electronic noise of the detector or from depositions due to pile-up activity, or from both. The noise modeling results in a gaussian distribution centered around zero and as such negative energy cells may exist.

Jet clustering

Clustering algorithms are assigned the task to collect the input objects and decide which and how they are going to be recombined in order to form a jet. Being a challenging field, many approaches have been implemented but two major algorithm classes can be distinguished, the *cone* algorithms and the *sequential recombination* algorithms.

In ATLAS, a fixed-size and seeded version of the cone algorithm has been extensively used in the past. It starts by first assembling all the input objects into a list with decreasing transverse momentum. A seed is then selected as the highest p_{T} object, with the lowest acceptable threshold being at the energy of 1 GeV. All other objects that are within $\Delta R \leq R_{cone}$ in the $\eta - \phi$ space from the seed are clustered together to form a jet. An iterative procedure is used thereafter, where the direction of the new jet becomes the central point and the objects around it are then re-clustered; the recalculation of the jet continues until its direction is considered stable. With the above approach it is possible that jets share the same constituents. As a result, a split-merge procedure is applied under which it is decided whether to merge two jets or split them by giving the overlapping constituents to the one of the highest p_{T}.

A significant problem with all common cone algorithms is that they are considered collinear and infrared unsafe hence they cannot be trusted for multi-jet systems. However, in a more advanced implementation, the Seedless Infrared Safe cone algorithm or `SISCone+` [139], these issues are resolved. A complete description of collinear and infrared safety effects is given in [140].

On the other hand, sequential recombination algorithms by construction do not suffer from the above issues. In ATLAS the main algorithm generally defines the following parameters which are calculated for each of the input objects i:

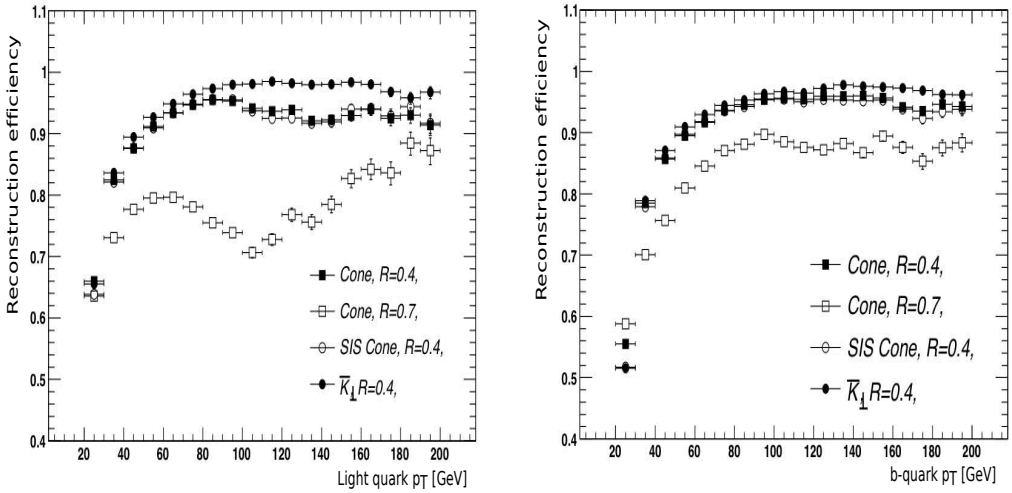
$$d_{ij} = min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{(\Delta R)_{ij}^2}{R^2} \ , \tag{4.1.1}$$

$$d_{iB} = k_{Ti}^{2p}. \tag{4.1.2}$$

The k_T variable refers to the transverse momentum of the object, ΔR_{ij} to the $\eta - \phi$ distance of two objects i and j, R is a measure of the jet size and the parameter p is the power of the energy scale. In practice, d_{ij} is the distance variable between two objects and d_{iB} is the distance variable between an object and the beam; for all the input objects the d_{ij} and d_{iB} are calculated and are put in a list. If the smallest entry of the list belongs to a d_{ij} category then i and j are combined into a single object and the distances are re-evaluated. If, however, it belongs to a d_{iB} category then the object is considered a jet and is removed from the list. This procedure continues until no objects remain in the list.

The jets used in this thesis follow the above clustering scheme where the parameter p is set to -1, referring to the `anti-`$k_\perp$ algorithm [126]. The main advantage, except from being infrared and collinear safe, is that it usually results in regular jet boundaries. In particular, it is by construction that the softer p_{T} particles will cluster with their neighboring hard p_{T} particles before they cluster with themselves. As a result, in the vicinity $\Delta R_{ij} \leq R$ of a single hard particle all soft candidates will be absorbed within the cone of size R. Although, typically this algorithm would require a significant amount of time, with the `FastJet` implementation this is drastically reduced [141].

Figure 4.5 shows the reconstruction efficiency of jets in the $t\bar{t}(e\nu_e)$ or $t\bar{t}(\mu\nu_\mu)$ topology with respect to the p_{T} of the mother quark (light-quark or b-quark). Events are preselected, requiring at least three jets of $p_{\mathrm{T}} \geq 40$ GeV or higher and at least one more with $p_{\mathrm{T}} \geq 20$ GeV or higher. All jets with $p_{\mathrm{T}} \geq 20$ GeV are considered and they are flagged as matched if they reside within $\Delta R \leq 0.3$ from the quark. The `anti-`$k_\perp$ algorithm with jet size $R = 0.4$ is



(a) *Jets from light quarks* (b) *Jets from bottom quarks*

Figure 4.5: *Reconstruction efficiency of light-quark (a) and b-quark (b) induced jets formed by the different algorithms, as a function of the p_{T} of the quarks. The* `anti-`$k_\perp$ *algorithm is denoted as $\bar{K}_\perp$. Plots taken from [142].*

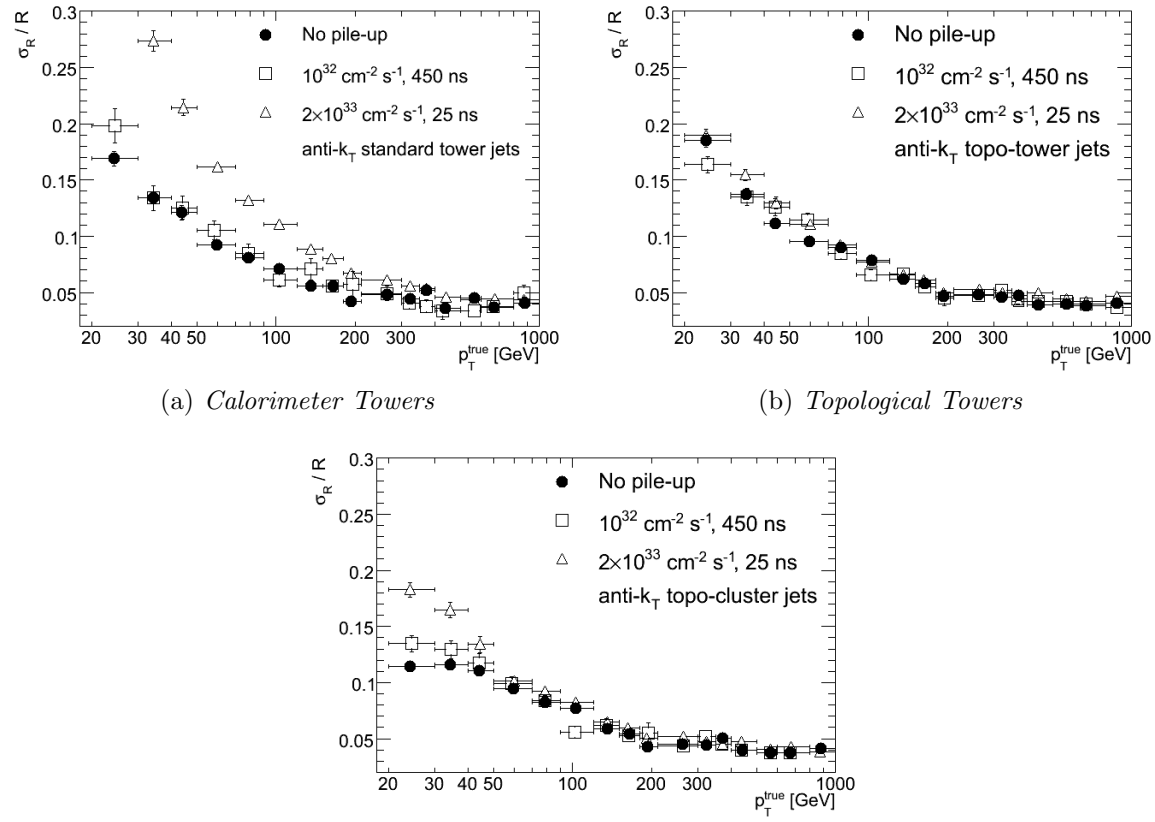
compared with the cone jet finder of ATLAS as well as with the `SISCone` algorithm. The large drop in efficiency of the cone jet finder with $R = 0.7$ for light quark p_{T} above 80 GeV is identified as an effect of the reconstruction which fails to identify a single jet at these energies and with the given R parameter. For the jets originating from b-quarks the effect is much less profound because typically these jets are more narrow. Evidently, the `anti-`$k_\perp$ achieves always a higher reconstruction efficiency.

Using the `anti-`$k_\perp$ algorithm, the different calorimeter inputs are compared in terms of the jet resolution performance using Monte Carlo simulations. Figure 4.6 shows the evolution of the jet resolution for each input case and for three different luminosity scenarios: the no pile-up case, a low luminosity scenario with a large bunch-spacing (425 ns) that corresponds to an early running period and a higher luminosity one with a nominal bunch-spacing (25 ns) that refers to a mid-term running period. The vertical axis shows the σ_R/R, where $R = p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{true}$, and the distribution is plotted against p_{T}^{true}. All jets are matched to true jets using a $\Delta R \leq 0.3$ requirement.

The conclusion that can be drawn from these plots is that using the topological clusters a better resolution is achieved in all scenarios. However, at the higher luminosity the performance is very similar with the topological towers which in fact show to be less affected by the increase in the pile-up activity.

Jet calibration

The energy deposited in the calorimeters corresponds to the energy of the showers. Eventually, the energy and the momentum of the jets must correspond to the hadronic scale and as a result a correction must be applied. This is an essential, yet challenging, step in the jet reconstruction and can take place either at the calorimeter reconstruction level, by correcting each input object to the hadronic scale and then letting the jet clustering algorithm act on them (local calibration), or at the jet reconstruction level, where jets are built from the input objects at their electromagnetic scale and afterwards they are adjusted as a whole (global



(a) *Calorimeter Towers*

(b) *Topological Towers*

(c) *Topological Clusters*

Figure 4.6: *Jet resolution comparison using the* `anti-`$k_\perp$ *algorithm with jet size* $R = 0.4$ *for the three different inputs variables and for three different luminosity scenarios as measured from Monte Carlo. Plots taken from [143].*

calibration). However, corrections do not only involve the hadronic energy deposition, but also effects from dead material in the detector, from energy leakage due to escaping particles or from losses due to reconstruction effects. All these are included in the jet calibration step.

A number of different methods are available for ATLAS [138] that are meant to be complementary for the understanding of the systematic uncertainties. However, as the research in this thesis takes place in the beginning of the data-taking era of the experiment, a simple yet robust method is used, the `EMJES` calibration [127]. This method corrects each jet depending on its transverse momentum and pseudo-rapidity and it gives a straighforward evaluation of the related systematic uncertainties. More advanced methods at the time of this thesis were still under commissioning.

The `EMJES` method has three major steps:

- First, the energy is corrected at the electromagnetic level for the pile-up activity in the event. The correction factors are derived by estimating from minimum bias data the average jet energy and the average towers per jet with respect to the number of primary vertices and the η of the jet.
- Second, the jet direction is corrected so that it points to the primary vertex of the event and not to the detector's geometrical center as it is by default.
- Last, a correction is applied in both the energy and direction of the jet using factors ob-

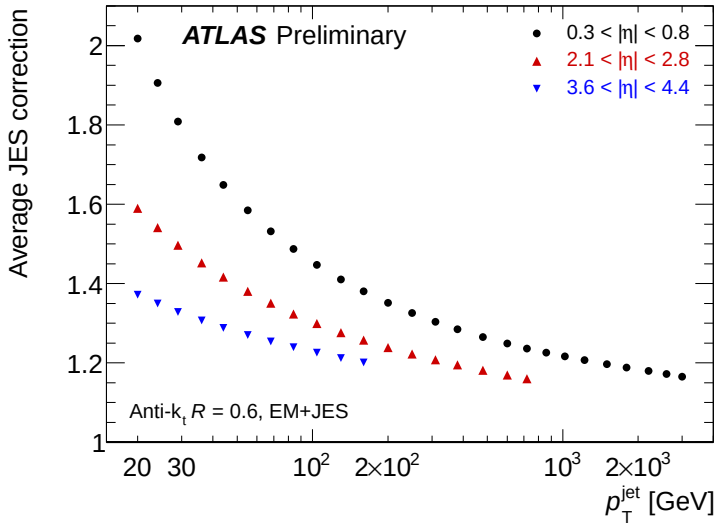
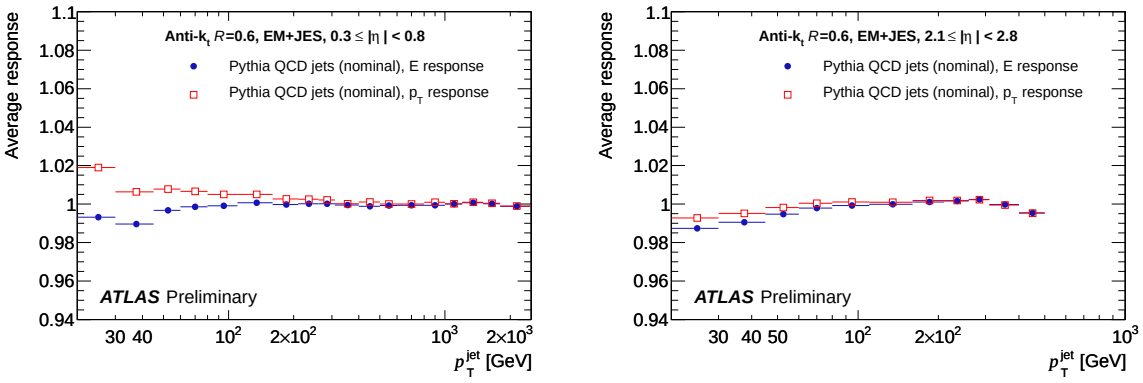


Figure 4.7: *The average jet energy scale correction applied with respect to the p_T of the jet after the calibration is applied. Plot taken from [127].*



(a) *Barrel* (b) *End-caps*

Figure 4.8: *The jet p_T and E response after the* `EMJES` *calibration is applied as calculated from Monte Carlo samples. Plots taken from [127].*

tained from Monte Carlo samples in which the reconstructed jet kinematics are compared with the true particle jets.

In figure 4.7 the average jet energy scale correction is shown with respect to the jet transverse momentum for three different calorimeter regions, while in figure 4.8 the ratio of the measured energy after the calibration step over the true jet energy with respect to the calibrated jet p_T is plotted. All jets used in these plots are coming from inclusive QCD Monte Carlo events and are reconstructed using the `anti-`$k_\perp$ algorithm with a jet size of $R = 0.6$ on topological clusters. For the jet response plots the deviations from unity, when observed, suggest a failure of restoring the kinematics of the calibrated jet to that of the particle jet. The main reason for this is that the same correction factor is applied in both the transverse momentum and the energy of the jet. Naturally, for when the jet mass is found to be non-zero and deviated from the true jet mass the correction only on the energy and pseudo-rapidity of the jet is not sufficient as it leads to a bias of its p_T. In addition to the above, the calibration algorithm assumes that all jet constituents should receive the same average compensation when deriving the calibration constants; this is not always the case.

4.1.4 B-jet finding

Top quarks decay predominantly into a W boson and a b-quark. As a result, being able to tag the jets that come from b-quarks provides a great tool in $t\bar{t}$ signal selection and eventually increases the signal-to-background ratio.

Identification of those jets is possible either by exploiting the relatively long lifetime of the B-meson or baryon (~ 1.5 ps) that the b-quark hadronizes to, or by identifying its decay products. In the former case, the resulting hadron travels a measurable distance from the primary vertex (PV) before it decays to other particles, creating a secondary vertex (SV) which can be distinguished in the inner tracker. In the latter case, when the hadron decays leptonically, the resulting charged lepton (electron or muon) can be utilized since due to the hard fragmentation and the high mass of its parent particle (≥ 5 GeV) it carries a relatively large transverse momentum and its track can be associated to the jet axis. Naturally, the Inner Detector and the calorimeters are the main detectors that take part in the process of B-jet finding.

A large number of algorithms are employed for measuring observables that are related to the above properties, but two major categories can be distinguished: the spatial taggers and the soft-lepton taggers. For the latter case, because the taggers utilize only properties of the resulting lepton, the efficiency of the algorithm is intrinsically limited by the branching ratio of the leptonic decay of the b-hadron. Hence, they suffer from a significant loss of statistics which is important at the early stages of the experiment. Consequently, they are not taken into account here and one can refer either to the b-tagging section of [137] or to [144] for more information. The spatial taggers on the other hand utilize information both from the lifetime of the hadron and the displaced vertex and several approaches have been considered for the initial running period.

Impact parameter taggers

One category of spatial taggers is dedicated in measuring the signed transverse (d_0) and the signed longitudinal (z_0) impact parameters (IP) of selected tracks. The former is defined as the closest distance between the track and the PV when projected in the $r-\phi$ plane, while the latter is defined as the z-coordinate of the track when projected to the $r-\phi$ plane. Typically, tracks originating from b-jets have larger positive values of d_0 and z_0 while the tracks that originate from lighter quarks are equally distributed around zero. The significance of these parameters, namely d_0/σ_{d_0} and z_0/σ_{z_0} are usually used as they give more discriminating power. Figure 4.9 shows an example of the transverse IP significance of selected tracks that are associated with b-jets, c-jets or lighter quark jets, as calculated from a sample of $t\bar{t}$ events.

Examples of taggers that take advantage of the impact parameter observables and are considered in the early data-taking are the `TrackCounting` [145] and the `JetProb` [137] algorithms.

Secondary vertex reconstruction taggers

The other category of spatial taggers exploits information after the secondary vertex is fully reconstructed. The SV reconstruction first selects tracks that can be associated with a given jet depending on a ΔR threshold from the jet axis. Track-pairs are then formed using all tracks that satisfy $L_{3D}/\sigma_{L_{3D}} > 2$, where L_{3D} is the distance between the PV and the closest point of the track, and $\sigma_{L_{3D}}$ is its associated error. This requirement ensures that tracks originating from the PV are excluded from the search. Subsequently, all pairs are combined into a vertex and a fit is performed. If the fit satisfies a given χ^2 threshold, the vertex is kept as long as

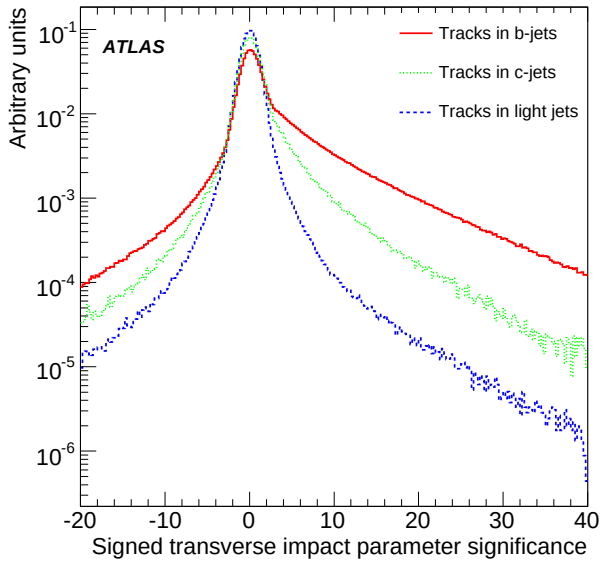


Figure 4.9: *The signed transverse impact parameter significance of tracks which are associated with b, c or lighter quark induced jets from a sample of Monte Carlo $t\bar{t}$ decays. Plot taken from [137].*

it is not likely to originate from K_s or Λ decays, or $\gamma \to e^+e^-$ conversions. Otherwise, the procedure is repeated after it removes the track with the largest χ^2 contribution.

Three observables can be defined after a secondary vertex is found:

- The invariant mass of all tracks in the vertex.
- The ratio of the energy sum of all tracks associated to the vertex over the energy sum of all tracks associated to the jet.
- The number of track-pairs that are used in the vertex.

As shown in figures 4.10 all of the above parameters are expected to have higher values for b-jets than for light jets and usually they are exploited in one of two ways: by retrieving information from the 2D-histogram of the first two variables and the 1D-histogram of the last (`SV1` tagger), or by retrieving information from the 3D-histogram of all the observables (`SV2` tagger) [137]. However, with this approach a significant amount of statistics is required and at the early stages of the experiment this leads to a large statistical uncertainty.

An alternative method instead of using the previous observables simply takes the signed decay length significance (L/σ_L), where the decay length (L) corresponds to the three-dimensional distance of the secondary vertex from the primary vertex. This algorithm which is named `SV0` is also considered for the early data-taking period [144].

In table 4.2, the selection purity as well as the rejection factors for light-jets and c-jets is shown for the 50% efficiency working point for the three algorithms that are considered for the early data taking period. The jet rejection factor is defined as the inverse of the tagging efficiency for the particular type of jets, namely $RF_{jet} = \frac{1}{\epsilon_{tag}} = N_{jet}/N_{jet}^{tagged}$.

The choice of the 50% efficiency is driven by the requirement to keep the statistics as high as possible while keeping low the fake rate. All jets that satisfy the $p_{\mathrm{T}} \geq 15$ GeV and $|\eta| \leq 2.5$ requirements are used. In the analysis presented hereafter the `SV0` algorithm is used since it has the largest purity and highest rejection factor.

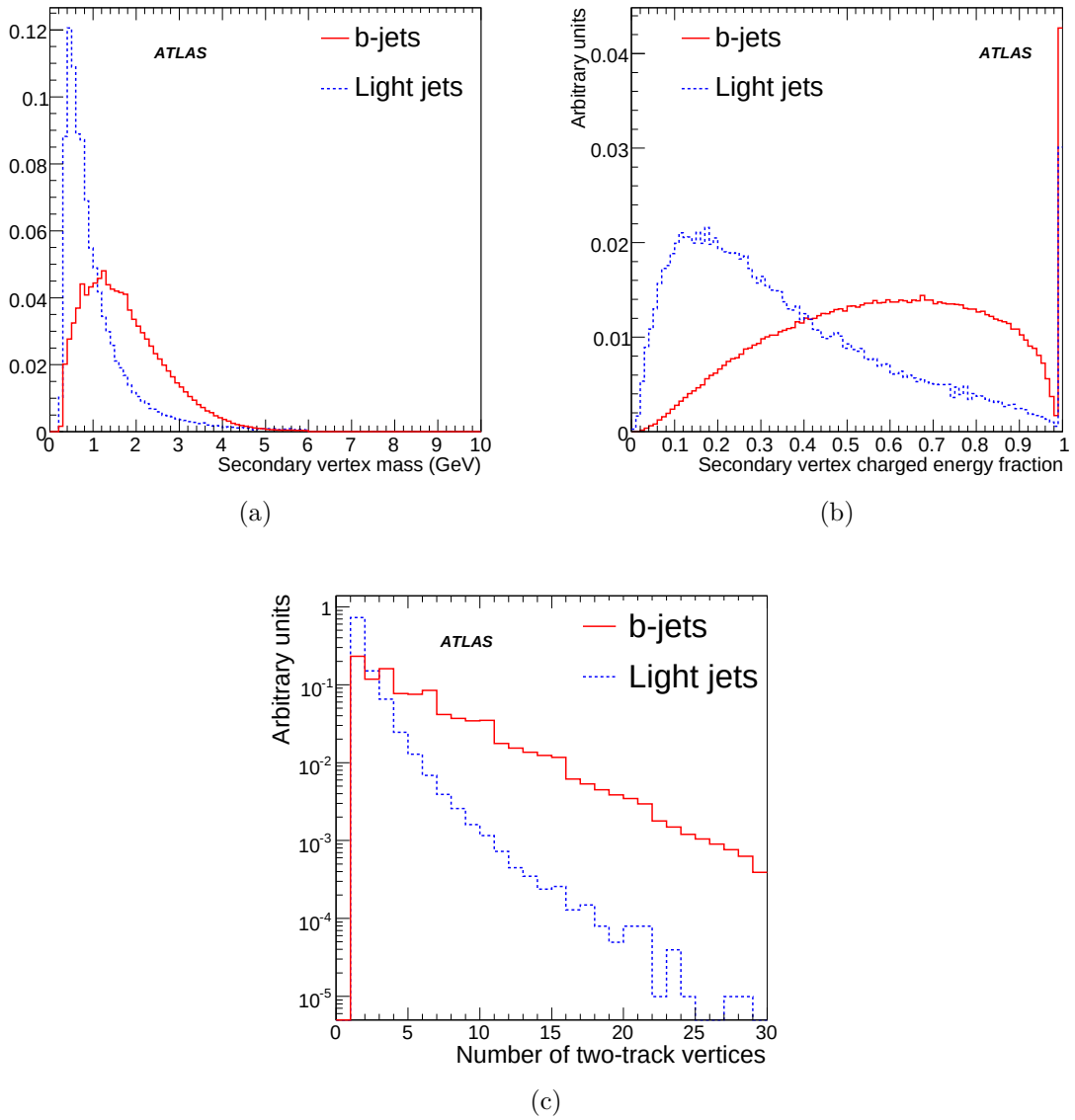


Figure 4.10: *The secondary vertex reconstructed mass (a) and charged tracks energy fraction (b), and the number of track-pairs used in the reconstruction (c). The distributions are obtained from $t\bar{t}$ Monte Carlo events. Plots taken from [137].*

	b-jet efficiency (%)	Purity(%)	Rejection factor		
			Light-jet	c-jet	τ-jet
`JetProb`	50.7	92.7	130	9	29
`TrackCounting`	50.2	92.5	140	8	34
`SV0`	50.1	93.9	271	9	38

Table 4.2: *Comparison of the purity and the rejection factors (RF) for each of the considered b-tagging algorithms at the 50% b-jet efficiency working point. Numbers estimated from Monte Carlo and taken from [146].*

4.1.5 Missing transverse energy - $E_{\mathrm{T}}^{\mathrm{miss}}$

In single-lepton $t\bar{t}$ decays a large part of the event's energy is carried away by the neutrino. This energy is not directly measurable since neutrinos do not leave a trace on any of the detector's components. However, it is possible to reconstruct the energy associated to all non-interacting particles in the event, by using the momentum conservation principle and balance the contribution of the other objects in the event. The "missing" energy, as it is referred, will correspond to the vector sum of the momenta of all non-interacting particles. It is easily perceived that this is not possible in the longitudinal direction where the exact initial momentum of the interacting partons is not known. Instead, the transverse $(x-y)$ plane is used as it can be assumed that in this direction the partons' momentum is zero, hence the missing transverse energy is calculated. This is then given by the following:

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}\ , \tag{4.1.3}$$

and the azimuthal angle (ϕ) of the vector is given by:

$$\tan\phi^{miss} = \frac{E_y^{miss}}{E_x^{miss}}. \tag{4.1.4}$$

Reconstruction

The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction sums up the deposited energy in the calorimeters (E^{Calo}) together with the energy of the muons (E^{Muon}) and in addition it corrects for the energy that is lost due to the inactive material in the cryostat region between the electromagnetic and the hadronic calorimeters (E^{Cryo}). The following relation holds for each of the two directions in the $x-y$ plane:

$$E_{x,y}^{miss} = -(E_{x,y}^{Calo} + E_{x,y}^{Muon} + E_{x,y}^{Cryo}), \tag{4.1.5}$$

where the minus sign is taken because of the energy conservation.

The calorimeter term

For the calorimeter term, in order to achieve noise-suppression, only cells that belong to topological clusters are used, although in this case all identified clusters are taken into account and not only the ones used in the jet reconstruction. A calibration scheme is also applied at the cell level.

This term can be refined into the individual contributions of the various objects that deposit energy in the calorimeters, e.g. electrons, jets etc. In this way a better calibration can be applied, based on the identified object, that replaces the previous. The assignment of cells to each reconstructed object is done by first identifying the latter and then moving back to the calorimeter components it traverses. It is possible that cells are shared between terms, in this case a priority list must be defined to avoid double counting of energy.

The cryostat term

The cryostat term in equation 4.1.5 effectively calculates the jet energy loss while it traverses the region between the two calorimeters. It is given by:

$$E_{x,y}^{Cryo} = \sum_{jets} E_{jet_{x,y}}^{Cryo} \tag{4.1.6}$$

where,

$$E_{jet}^{Cryo} = w^{Cryo}\sqrt{E_{EM3} \times E_{HAD}}. \tag{4.1.7}$$

The w^{Cryo} corresponds to a calibration weight, while the E_{EM3} and E_{HAD} are the amount of jet energy deposited to the third layer of the electromagnetic calorimeter and to the first layer of the hadronic calorimeter, respectively. The inclusion of the cryostat term is very important as for high-p_{T} jets it accounts for a 5% of its total energy [147].

The muon term

Lastly, the muon term takes into account the p_{T} of muons that reside within $|\eta| \leq 2.7$. For the region $|\eta| \leq 2.5$, only combined muons are accounted for, namely muons with a Muon Spectrometer track matched to an Inner Detector detector track, while for $|\eta| > 2.5$, where there is no Inner Detector coverage, muons from the Muon Spectrometer are used. Irrespectively of the region, the transverse momentum measurement that is used in the muon term is taken from the Muon Spectrometer component. The reason for this is that the energy deposition in the calorimeter is already included at the $E_{x,y}^{Calo}$ term and if combined muons where to be used then this would lead to double-count the energy loss.

4.2 Monte Carlo correction factors

One of the most important aspects in an experiment is the correct description of the detector performance by the Monte Carlo simulated events. Using well known physics processes it is possible to assess the level of agreement between data and Monte Carlo and apply corrections accordingly. For the results shown in the rest of the thesis the following corrections have been applied at the object reconstruction level by default[6].

4.2.1 Electron energy scale and resolution

Resonances such $Z \to e^+e^-$ or $J/\psi \to e^+e^-$ are used to determine the energy scale of electrons and correct the electromagnetic cluster energy [148]. Typically, the reconstructed invariant mass peak is required to match the well-known resonance line-shape of the respective particle. The extracted scale is quantified by the dimensionless parameter α as a function of pseudo-rapidity, and subsequently the correction is applied on data events, for each identified electron, using the following relation:

$$E_{new} = \frac{E}{1+\alpha}. \tag{4.2.1}$$

The values of α are shown in figure 4.11.

Both Z boson events and J/ψ events are examined in the di-electron final state with respect to their invariant mass distribution. Although for the latter the agreement between data and Monte Carlo is shown to be very good, for the Z boson events significant discrepancies have been observed. This effect is attributed to the modeling of the constant term in the electron energy resolution ($\sigma_E/E \approx 10\%/\sqrt{E} \oplus 1\%$) as it is the main contributing factor at the high electron energy region. As a result, a correction is applied to each electron in the Monte Carlo samples by smearing its energy according to new constant terms which have shown to achieve a good agreement. More details on the results for re-scaling and smearing the electrons can be found in [149].

[6] Unless stated otherwise within the text.

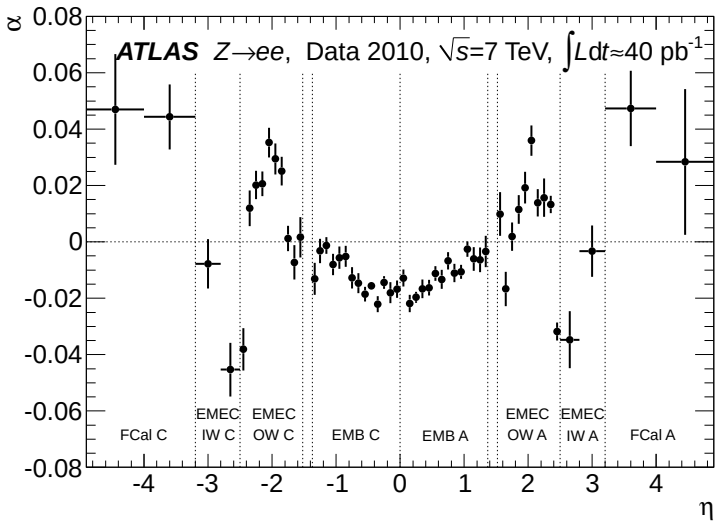


Figure 4.11: *The dimensionless electron energy scale parameter (α) as a function of pseudo-rapidity measured from $Z \to e^+e^-$ decays. Plot taken from [148].*

4.2.2 Muon momentum scale and resolution

Studies performed with the first data have shown discrepancies on the reconstructed Z boson invariant mass, obtained from di-muon events, when comparing to the Monte Carlo. This is attributed to inaccurate modeling of the material budget and the alignment of the different sub-detectors which affect the estimation of the muon's momentum and resolution. In order to improve the agreement the following smearing term is applied on each muon on the Monte Carlo:

$$\left(\frac{1}{p_{\mathrm{T}}}\right)_s = \frac{1}{C_1} \times \left(\frac{1}{p_{\mathrm{T}}}\right)_{MC} \times (1 + G(0,1) \times C_2)\,, \tag{4.2.2}$$

where $\left(\frac{1}{p_{\mathrm{T}}}\right)_s$ is the muon curvature after the smearing, $(\frac{1}{p_{\mathrm{T}}})_{MC}$ is the muon curvature before the smearing, and C_1 and $C2$ are the momentum scale and resolution smearing terms respectively. The function $G(0,1)$ corresponds to a gaussian with mean $\mu = 0$ and width $\sigma = 1$ which is used to randomly generate a number. Both the C_1 and C_2 are tuned based on the Z boson line-shape and are binned in different η bins.

For combined muons, the Inner Detector and the Muon Spectrometer components are treated independently. Once the smeared p_{T} is estimated for each component, a weighted average is taken for its expected resolution. This is done because the Inner Detector and the Muon Spectrometer components contribute in a different way to the combined muon's resolution, depending to a large extend on the the total momentum of the muon. A detailed discussion of the muon momentum scale and resolution can be found here [150].

4.2.3 Jet energy resolution

The jet resolution term is written in the following form:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \otimes \frac{S}{\sqrt{p_{\mathrm{T}}}} \otimes C, \tag{4.2.3}$$

where N is a term that parametrizes noise fluctuations and offset energy from multiple pile-up activity (noise term), S parametrizes stochastic fluctuations in the energy measured from the

hadronic shower (stochastic term), and C parametrizes constant fluctuations (constant term).

Using di-jet QCD events the energy resolution of jets was examined directly from data and a good agreement was in general observed [151]. However minor corrections are still applied to the Monte Carlo to further improve their description of data. The transverse momentum of each jet is corrected by a scale value following the relation:

$$(p_{\mathrm{T}})_s = (p_{\mathrm{T}})_{MC} \times (1 + G(0, S)), \tag{4.2.4}$$

where $(p_{\mathrm{T}})_{MC}$ and $(p_{\mathrm{T}})_s$ are the p_{T} of the jet before and after the smearing, and $G(0, S)$ is a gaussian random number generator with a mean of zero and a width taken from S, the smearing term. The S is typically obtained from look-up tables and it depends on the p_{T} and the rapidity (y) of the jet.

4.2.4 Heavy flavor composition in W+jets events

The composition of W+jets events in the Monte Carlo includes those events with jets induced by heavy flavor quarks, namely $Wb\bar{b}$, $Wc\bar{c}$ and Wc; extra jets from lighter quarks may also be present in these event types. The exact heavy flavor fraction in those events is not known with good accuracy as it depends on the heavy flavor Parton Distribution Functions (PDFs) and for this reason a better approximation is obtained using a data-driven approach. The method estimates the fraction of b-tagged events in the kinematic region where exactly two jets are identified; it is explained in detail in [152]. The result shows that no correction for Wc events is required but the simulation underestimates the fraction of $Wb\bar{b}$ and $Wc\bar{c}$ by 30% and therefore a per event scale factor should be applied.

4.2.5 Flavor tagging

For analyses that depend on event selection based on the identification of b-jets, the calibration of the flavor tagging algorithms is an important step. The b-tagging efficiencies for the `SV0` tagger have been measured in data using an enriched $b\bar{b}$ di-jet sample where a significant muon content is observed. In addition, the mis-tag rate, namely the rate at which light flavored jets are mistakenly seen as b-jets, is also examined for `SV0`. Further details on these methods can be found in [144].

The comparison of the data-driven measurements with the Monte Carlo shows differences which are parametrized in the form of scale factors. These scale factors are applied on the Monte Carlo in order to improve the agreement with the data and they depend on the flavor of the jet (b-jet, c-jet or light jet) as well as on its p_{T} and η. Summarizing, the scale factor for b-jet tagging efficiency is given by:

$$SF_{tagged}(p_{\mathrm{T}}, |\eta|) = \frac{\epsilon^{data}(p_{\mathrm{T}}, |\eta|)}{\epsilon^{MC}(p_{\mathrm{T}}, |\eta|)}. \tag{4.2.5}$$

While the scale factor for b-jets that are not tagged is given by:

$$SF_{not-tagged}(p_{\mathrm{T}}, |\eta|) = \frac{1 - SF_{tagged}(p_{\mathrm{T}}, |\eta|)\epsilon^{MC}(p_{\mathrm{T}}, |\eta|)}{1 - \epsilon^{MC}(p_{\mathrm{T}}, |\eta|)}. \tag{4.2.6}$$

The above formulation is also used for c-jets that are either tagged or not-tagged with the only difference that the estimated uncertainty in the scale factors is doubled. Lastly, and in addition to the above, a scale factor for the mis-tag rate of light jets is obtained. A complete list of the values of the scale factors for the `SV0` tagger at the 50% working point, which is of interest for this thesis, is provided in table 4.3.

p_{T}**range (GeV)**	b**-tagging Scale Factor**	**Mis-tag rate Scale Factor**	
		$\lvert\eta\rvert < 1.2$	$1.2 < \lvert\eta\rvert < 2.5$
20-25	0.872 ± 0.208	1.19 ± 0.43	1.36 ± 0.49
25-40	0.925 ± 0.105	1.36 ± 0.35	1.39 ± 0.32
40-60	0.942 ± 0.074	1.03 ± 0.19	1.20 ± 0.19
60-75	0.947 ± 0.102	1.07 ± 0.12	1.01 ± 0.22
75-90	0.947 ± 0.150	idem	idem
90-140	0.947 ± 0.200	0.95 ± 0.15	1.01 ± 0.19
140-200	idem	1.05 ± 0.14	0.99 ± 0.28
200-300	idem	1.29 ± 0.30	0.92 ± 0.60
300-500	idem	0.85 ± 0.51	1.70 ± 1.35

Table 4.3: *Scale factors for the tagging efficiency of b-jets and the mis-tag rate of light jets. The scale factor for the tagging efficiency of c-jets is the same but with a double uncertainty. All numbers correspond to the* 50% `SV0` *tagger working point. Numbers taken from [153].*

4.3 "Good" object definitions

To increase the quality of the $t\bar{t}$ single-lepton signature the reconstructed objects are required to pass a number of precise identification criteria which are based on the characteristics of the $t\bar{t}$ topology. The selected objects are defined as "good" and the event selection is applied later on these.

"Good" electrons

Selecting good electrons is essential for the search in the $t\bar{t}(e\nu_e)$ channel. The electron must satisfy the following requirements:

- Be reconstructed by the algorithms presented in section 4.1.1.
- Be identified as a tight electron. This is motivated by the fact that although the efficiency of signal electrons drops to 75%, the expected fake rate and consequently the background acceptance is significantly decreased.
- Must lie within $|\eta_{cluster}| \leq 2.47$, where $\eta_{cluster}$ is the pseudo-rapidity of the calorimeter cluster, but if it falls within $1.37 < |\eta_{cluster}| < 1.52$ it is rejected. The upper limit of $|\eta_{cluster}| \leq 2.47$ is set because of the acceptance of the Inner Detector which is up to $|\eta| \leq 2.5$ and is an essential part of the reconstruction algorithms. The small difference between the actual cut and the ID acceptance is because the pseudo-rapidity of the electron is not coming from the reconstructed track but from the calorimeter. On the other hand, the rejected region corresponds to uninstrumented parts of the detector which are not well modeled in the Monte Carlo.
- Be isolated. The calorimetric isolation parameter (`etcone`) is used in this case (see section 4.1.1) with an R_0 parameter set to 0.2. The energy deposition in the calorimeter and within the defined cone should not exceed 4 GeV.
- Finally, the electron should have transverse energy (E_T) of at least 20 GeV.

The above criteria take into account the distinct characteristics of electrons that result from the W decay in a $t\bar{t}$ event. However, detector problems such as: dead or non-nominal high-voltage regions, channels with very high noise which are masked by the reconstruction or dead

front-end boards, must also be taken into account. For this reason, dedicated $\eta-\phi$ maps of the calorimeters are created for each data-taking run where such problems are documented. For each electron the cluster coordinates $(\eta_{cluster}, \phi_{cluster})$ are checked in the map and a decision is made on whether it will be rejected or not. This is applied not only in data but also in Monte Carlo to get a better agreement. For the electrons selected here the map corresponding to the highest integrated luminosity is used.

"Good" jets

Motivated by the performance comparisons shown in section 4.1.3, all jets that are considered in this analysis use as input topological clusters and are reconstructed with the `anti-`$k_\perp$ algorithm with an R parameter of 0.4. In addition, as mentioned above, the `EMJES` calibration scheme is used. At the software level these jets are denoted as `AntiKt4TopoEMJets`.

For the definition of a good jet, the following requirements are applied:

- It does not overlap within $\Delta R \leq 0.2$ with a "good" electron. The coordinates used for the electron are taken from its track component $(\eta_{track} - \phi_{track})$ and for the jet they are obtained before applying the calibration $(\eta_{em} - \phi_{em})$. However, for each electron only one jet, the closest, is removed.
- Unless the top quark-pair is boosted, the jets are produced in the central region of the detector since they are the result of the decay of a heavy object. Thus, it is required that the jet lies within $|\eta| \leq 2.5$.
- Lastly, the transverse momentum is required to be $p_{\mathrm{T}} \geq 20$ GeV.

In addition to the above cuts, a good jet is flagged as b-tagged if the weight assigned to it by the `SV0` algorithm is greater than 5.82. This weight corresponds to the working point of 50% efficiency as shown in section 4.1.4.

"Good" muons

As mentioned in section 4.1.2 all muons used in this thesis come from the `MuID` collection. Specifically for the analysis presented here, a good muon should:

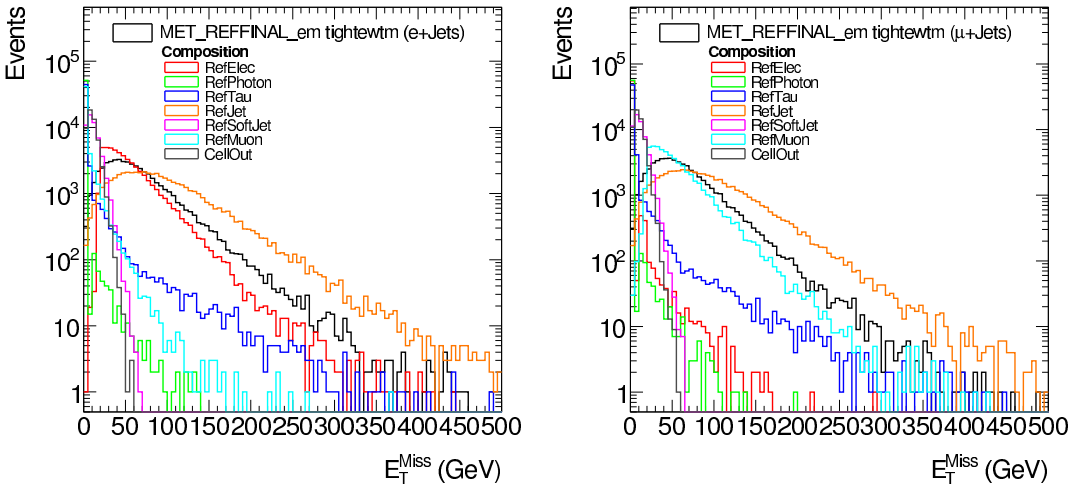
- Be a tight muon. As shown in figure 4.4 the tight requirements select muons with an efficiency above 90% and have a better fake rejection.
- Be a combined `MuID` muon. Hence, `MuGirl` muons are not considered in this analysis.
- The muon must satisfy the quality requirements for the Inner Detector track quality which include the detector status during data-acquisition. Documented in detail in [154].
- It must be within $|\eta| \leq 2.5$ which is the Inner Detector's acceptance region.
- The muon must be isolated. Both calorimetric and tracking isolation is used in this case (see section 4.1.2). For both terms, the cone parameter is set at 0.3 and the energy content within the cone should not exceed 4 GeV.
- To improve the isolation and to avoid selecting mainly QCD events in which a b-quark decayed leptonically, any muon that is within $\Delta R \leq 0.4$ from a jet with $p_{\mathrm{T}} \geq 20$ GeV is rejected.
- Lastly, the transverse momentum of the muon must be at least 20 GeV.

$E_{\mathrm{T}}^{\mathrm{miss}}$ *definition*

The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction used in this thesis uses the definition given earlier with a refined calorimeter term which is derived as follows:

$$E_{x,y}^{Calo} = E_{x,y}^{RefElec} + E_{x,y}^{RefPhoton} + E_{x,y}^{RefTau} + E_{x,y}^{RefJet} + \\ + E_{x,y}^{RefSoftJet} + E_{x,y}^{RefMuon} + E_{x,y}^{CellOut}, \tag{4.3.1}$$

where the parameters refer respectively to the contribution from electrons, photons, muons that deposit energy in the calorimeter, taus that decay hadronically, jets with $p_{\mathrm{T}} \geq 20$ GeV, soft jets with $5 \leq p_{\mathrm{T}} \leq 20$ GeV, and finally calorimeter cells that reside within the topological cluster but do not contribute to any of the above objects. The inclusion of the last term is proven to give a better absolute $E_{\mathrm{T}}^{\mathrm{miss}}$ value and a better resolution [147].



(a) *Electron channel* (b) *Muon channel*

Figure 4.12: *The total $E_{\mathrm{T}}^{\mathrm{miss}}$ and the E^{Calo} term components for single-lepton $t\bar{t}$ events. Plots taken from [155].*

A detailed description of the object selection requirements as well as the relevant calibration schemes that are used for the calorimeter term in this thesis are documented in [155]. Figure 4.12 shows the components of the calorimeter term together with the total $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement - denoted as `MET_REFFINAL_em_tightewtm` - for $t\bar{t}(e)$ and $t\bar{t}(\mu)$ events as calculated for this thesis.

4.4 Baseline top quark-pair event selection

For each of the $t\bar{t}(e)$ or $t\bar{t}(\mu)$ final states we apply certain requirements. These involve: the online selection criteria, namely the trigger used for accepting the events to be registered in the data sample; the data quality criteria, which are requirements related with the status of the detector and which affect the object reconstruction; the analysis requirements, which are based on the object definitions given in section 4.3 and are motivated by the topology of the events.

4.4.1 Data sample

As mentioned already in 3.2.1, the data recorded by ATLAS are separated in periods and further into sub-periods. Because the 2010 physics run was the very first physics data-taking of ATLAS, it served to a large extend for commissioning and understanding of the detector. For the $t\bar{t}$ measurement, the data from sub-period E4 and after are used as earlier periods suffer from technical shortcomings that are difficult to treat in the analysis. However, the data content of the excluded periods does not represent a significant fraction of the data. The integrated luminosity that was registered by the detectors is shown in table 4.4. The relative uncertainty on the integrated luminosity collected is measured in [156] and is found to be $\pm 3.4\%$. For the present analysis only data from the electron and the muon streams are used and the total events that are processed are about 16.8 million from the first and around 13.5 million from the second stream.

Period	$\int \mathcal{L} dt$ (pb^{-1})
A-E3	0.75
E4-E7	0.51
F (1-2)	1.53
G (1-6)	5.53
H (1-2)	6.99
I (1-2)	20.73
All Periods	36.05
excl. A-E3	35.30

Table 4.4: *Total integrated luminosity registered by ATLAS in 2010 with respect to the various data-taking periods.*

4.4.2 Trigger selection

The first step in the selection procedure of a data sample is to require that the events have been accepted by a certain trigger. From all the distinctive characteristics of the $t\bar{t}$ single-lepton topology, the simplest and most straightforward approach is to use a single-lepton trigger for each of the channels that are analyzed. One of the main considerations in the decision of a trigger is its energy or transverse momentum threshold. Having a threshold as low as possible increases the overall acceptance of the signal. However, this decision depends largely on the instantaneous luminosity delivered by the LHC which, as it increases, also increases the rate of the triggers; if the threshold is too low, triggers might need to be pre-scaled[7]. In addition, it must also be considered that not all the possibilities can be exploited at the beginning of the experiment. For example, leptons coming from $t\bar{t}$ are also expected to be well isolated with no jet activity around their trajectory. However, using such isolation criteria at the trigger level requires a better understanding and calibration of the detector. In the following, the triggers used for each of the channels of this analysis are discussed.

[7] A pre-scaled triggered performs the same task as a normal trigger with the only difference that it will allow a fraction of the total events to be accepted. For example a pre-scale of 1000 for a trigger will only give a positive decision for 1 out of 1000 fired events.

Electron channel

For the $t\bar{t}(e)$ channel the `Egamma` stream is used and the selection trigger is `EF_e15_medium` which is used consistently in both Monte Carlo and data. This trigger follows the chain:

$$\texttt{L1\_EM10} \rightarrow \texttt{L2\_e15\_medium} \rightarrow \texttt{EF\_e15\_medium}.$$

At the L1, at least one calorimeter cluster with at least 10 trigger counts is required, where 1 trigger count corresponds to approximately 1 GeV in the calorimeters. At the L2, the electron is required to have an $E_T \geq 15$ GeV and at the Event Filter, where a better resolution can be achieved, it must still pass the 15 GeV threshold but also be identified as a "medium" electron. The identification requirements for "medium", as well as for "loose" and "tight", at the trigger level are not exactly the same as in the offline reconstruction; they are documented in [157].

Table 4.5 shows the integrated efficiency for each trigger level as calculated using the tag-and-probe method on identified $Z \rightarrow e^+e^-$ events [157]. The overall trigger efficiency for an electron of $E_T > 20$ GeV is well close to 100%.

Trigger	**Efficiency (%)**	
	Data	**Simulation**
`L1_EM10`	$99.90 \pm 0.03(stat.) \pm 0.02(syst.)$	$99.995 \pm 0.001(stat.)$
`L2_e15_medium`	$99.60 \pm 0.06(stat.) \pm 0.05(syst.)$	$99.699 \pm 0.004(stat.)$
`EF_e15_medium`	$98.97 \pm 0.09(stat.) \pm 0.09(syst.)$	$99.445 \pm 0.006(stat.)$

Table 4.5: *Integrated efficiency for each trigger level of the* `e15_medium` *chain at* $E_T > 20$ GeV, *using tag-and-probe on* $Z \rightarrow ee$ *events. Numbers taken from [157].*

Muon channel

For the $t\bar{t}(\mu)$ channel the `Muon` stream data are used and three different triggers are utilized, each corresponding to a different data-taking period. In section 3.1 a detailed description of the muon online reconstruction was given. From period E4 until and including period F the `EF_mu10_MSonly` trigger is used. The chain followed is:

$$\texttt{L1\_MU0} \rightarrow \texttt{L2\_mu10\_MSonly} \rightarrow \texttt{EF\_mu10\_MSonly}.$$

At the L1, the trigger item uses the 'open' road configuration, thus a pass-through threshold of approximately 4 GeV is effectively applied. At the L2, the trigger uses the `muFast` algorithm (see section 3.1.2) to reconstruct the track, while at the Event Filter the result is more refined with the use of the `TrigMuonEF` algorithm but without using the `Combiner` step (see section 3.1.3). The event is accepted if at each level the muon satisfies the momentum threshold requirement that were given in table 3.4. For the periods G1 to G5 the `EF_mu13` trigger is used following the chain:

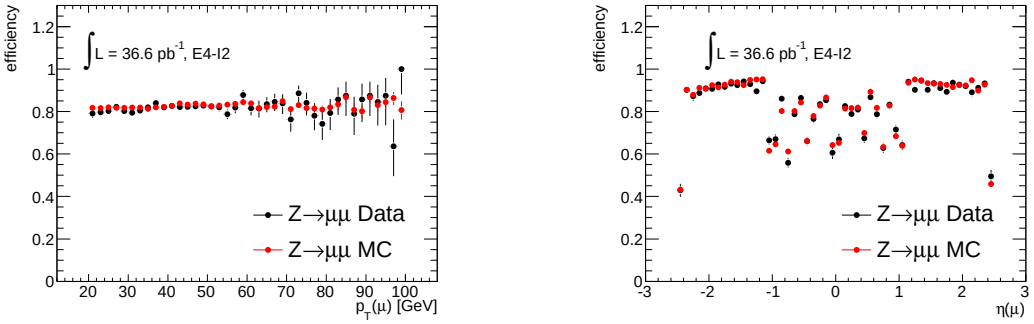
$$\texttt{L1\_MU0} \rightarrow \texttt{L2\_mu13} \rightarrow \texttt{EF\_mu13}.$$

The Level-1 trigger is the same as as before. At the HLT the Inner Detector information is used and a combined track is fitted. For the L2 trigger the `muComb` algorithm is used and for the Event Filter the `TrigMuonEF` including the `Combiner` step. The transverse momentum thresholds applied at L2 and Event Filter are showed in table 4.6

Lastly, from period G6 and until the end of the 2010 proton-proton collisions the trigger used is `EF_mu13_tight` which is similar to the `EF_mu13` with the exception that at Level-1 it is seeded by `L1_MU10`. Most of the statistics used in the analysis are collected during this

$\|\eta\|$ bin	p_{T} threshold (GeV)	
	Level-2	Event Filter
$[0.00, 1.05]$	12.6	12.67
$(1.05, 1.50]$	12.2	12.55
$(1.50, 2.00]$	12.2	12.49
$(2.00, \infty)$	12.4	12.46

Table 4.6: *Level-2 and Event Filter p_{T} thresholds with respect to $|\eta|$ for the* `mu13` *chain. Numbers taken from [122].*



(a) *Efficiency vs.* p_{T} (b) *Efficiency vs.* η

Figure 4.13: *Muon trigger efficiency distributions with respect to the muon p_{T} (a) and the pseudo-rapidity (b) calculated with tag-and-probe from $Z \to \mu^+\mu^-$ events. The result is overlaid with the Monte Carlo. Plots taken from [148].*

period, thus in order to retain simplicity only the `EF_mu13_tight` is used for the Monte Carlo samples. The efficiency of these triggers has been extracted using the tag-and-probe method on $Z \to \mu^+\mu^-$ events. This method has been discussed in detail in section 3.3 where the results of the study on the `EF_mu10_MSonly` trigger were shown. Here we use the results as obtained from the analysis documented in [148]. Figure 4.13 shows the efficiency binned as a function of p_{T} and η. A good agreement between data and Monte Carlo is evident.

4.4.3 Offline event selection

Before an event is evaluated for the physics analysis cuts, it must fulfill certain requirements that take into account the machine conditions and the quality of the reconstructed data. These data-quality cuts are:

- **Non-collisions background cut:** To reject events that are not related to a collision, a reconstructed primary vertex is required with at least 5 tracks associated to it.
- **Jet-cleaning cut:** The quality of the jet reconstruction is ensured with a set of cuts which are referred to as "jet-cleaning" and are documented in detail in [158]. These guarantee that out-of-time activity or defects in the calorimeters during data-taking do not affect the analysis. The event is eventually rejected if a jet with $p_{\mathrm{T}} \geq 20$ GeV is flagged as `LooseBad`. This particular cut is not applied in Monte Carlo.
- **Electron-muon overlap cut:** Finally, an event is rejected when a "good" electron and a "good" muon, but without the muon-jet overlap removal step (see section 4.3), are

identified as sharing the same Inner Detector track.

If the event is not rejected by the data-quality cuts it is tested against the following cuts, which also classify the event as belonging to the electron channel (e-channel) or to the muon channel (μ-channel):

- **Trigger cut:** The electron or the muon Event Filter trigger must have passed for the event. The relevant triggers are presented in section 4.4.2.
- **Charged lepton cut:** Exactly one "good" lepton, either an electron or a muon must be identified. In addition, the lepton needs to be matched with the corresponding Event Filter object using a ΔR cut in $\eta - \phi$ space of 0.15. The electron is matched using the calorimeter coordinates ($\eta_{cluster} - \phi_{cluster}$) with the Region-of-Interest of the trigger, while the muon matches its reconstructed track with the track of the trigger object.

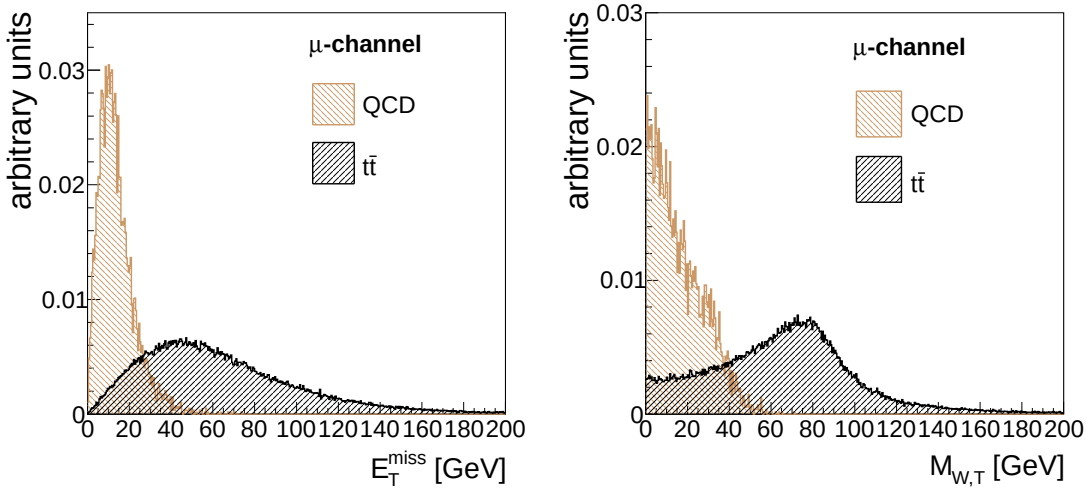
One of the most prominent backgrounds are the QCD multi-jet final states due to their large cross-section. Naturally, these topologies do not result in energetic and isolated leptons but given the large number of events it is possible that a considerable amount of them survives the trigger and charged lepton requirements. To reduce this background two observables can be used: the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ component of the events and the transverse mass of the W boson ($M_{W,T}$). The former is motivated by the presence of the energetic neutrino in $t\bar{t}$ final states which results in a peak at high $E_{\mathrm{T}}^{\mathrm{miss}}$ values. The latter exploits the fact that the leptonically decaying W boson in the $t\bar{t}$ events can be probed by its decay products. The $M_{W,T}$ is defined by the following:

$$M_{W,T} = \sqrt{(E_{\mathrm{T}}^{\mathrm{miss}} + p_{\mathrm{T},\ell})^2 - \sum_{i=x,y} (E_i^{\mathrm{miss}} + p_{i,\ell})^2} \; . \tag{4.4.1}$$

The discriminating power of these observables is shown in figure 4.14 where they are plotted for the QCD multi-jet and the $t\bar{t}$ samples after applying the trigger and charged lepton requirements for the μ-channel. In addition, figure 4.15 shows the two-dimensional $E_{\mathrm{T}}^{\mathrm{miss}} - M_{W,T}$ plot for both channels after the trigger and charged lepton cuts. Utilizing these two observables the following cuts are applied at the analysis level:

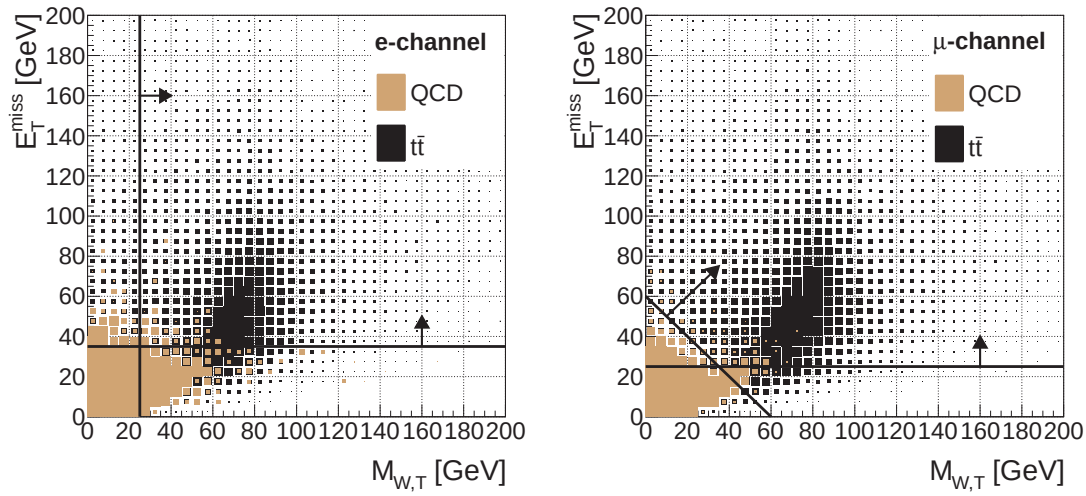
- $E_{\mathrm{T}}^{\mathrm{miss}}$ **cut:** For the e-channel the missing transverse energy component is required to be $E_{\mathrm{T}}^{\mathrm{miss}} > 35$ GeV while for the μ-channel it is $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV.
- W **boson transverse mass cut:** For the μ-channel the measurement of $M_{W,T}$ is summed with the $E_{\mathrm{T}}^{\mathrm{miss}}$ to form the *triangular* cut; the cut requires $E_{\mathrm{T}}^{\mathrm{miss}} + M_{W,T} > 60$ GeV. For the e-channel no triangular cut is used but it is still required that $M_{W,T} > 25$ GeV.

The considered single-lepton $t\bar{t}$ final states have four jets in their final topology with two originating from the hadronization of a b-quark. Additional jets can be produced through initial or final state radiation (ISR/FSR). This is different from the leptonically decaying W+jets events where jets are exclusively produced by ISR/FSR. In addition, jets in $t\bar{t}$ events being the product of the decay of heavy particles have on average higher transverse momenta. In figure 4.16 the $t\bar{t}$ and W+jets are compared in the μ-channel for their "good" jet content in terms of multiplicity and of the p_{T} of the fourth highest-p_{T} "good" jet. Evidently, the $t\bar{t}$ is dominant in higher multiplicities and high transverse momenta. Therefore the following cuts can be applied which also define the signal region of the analysis:



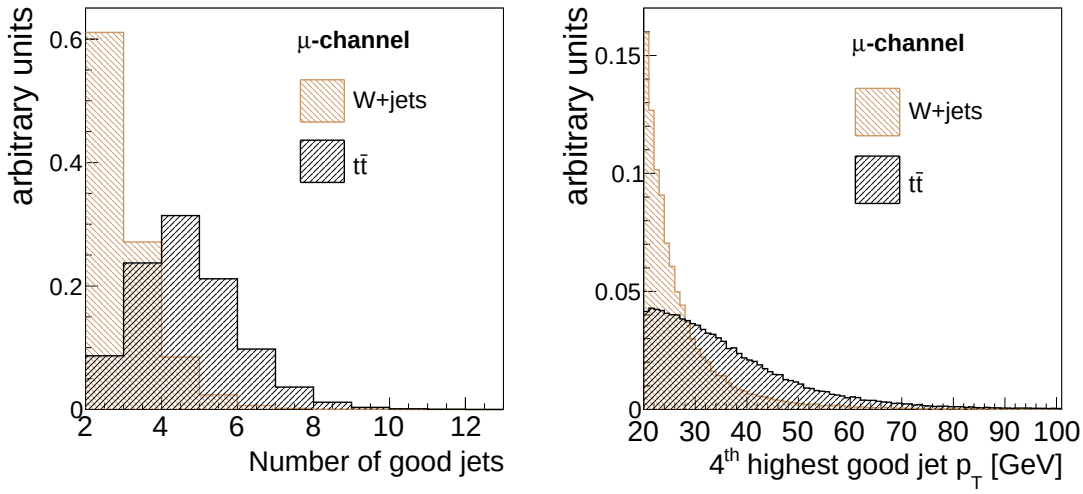
(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ *component* (b) *W boson transverse mass*

Figure 4.14: *Comparison of $t\bar{t}$ events with the QCD multi-jet events after the trigger and charged lepton cuts in the μ-channel. The reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ component (a) and the $M_{W,T}$ (b) are shown. The behavior also holds for the e-channel cuts. Histograms normalized to unit area.*



(a) *e-channel* (b) *μ-channel*

Figure 4.15: *$E_{\mathrm{T}}^{\mathrm{miss}}$ versus $M_{W,T}$ for the e-channel (a) and the μ-channel after trigger and lepton requirements. The triangular cut is defined by the sum of these two variables. In this analysis it is used for the μ-channel. The lines with the arrows indicate the phase space that is selected.*



(a) (b)

Figure 4.16: *The "good" jet multiplicity (a) and the p_{T} of the fourth highest-p_{T} jet (b) in the $t\bar{t}$ and W+jets events. The trigger and charged lepton requirements for the μ-channel selection are applied as well as a preselection of at least two good jets. The results are similar for the e-channel cuts. Histograms are normalized to unit area.*

- **Jet cut:** The event must have at least four good jets with $p_{\mathrm{T}} \geq 25$ GeV.
- **b-jet cut:** At least one good jet with $p_{\mathrm{T}} \geq 25$ GeV must be identified as a b-jet, that is to say it must have a weight of 5.85 from the `SV0` algorithm. This weight corresponds to the 50% efficiency working point, see section 4.1.4.

4.4.4 Cut-flow

In table 4.7 the cut-flow is given for each channel and for each Monte Carlo sample for all but the jet cuts[8]. Clearly, the charged lepton cut as well as the missing transverse energy and the transverse W mass cuts reduce significantly the QCD background. On the other hand, the rest of the background processes, which can easily provide an isolated lepton, are affected less and in particular the W+jets which still contributes significantly. Table 4.8 shows the surviving Monte Carlo and data events after applying the jet cut broken down into two bins: one where at least one jet is b-tagged, which corresponds to the signal region, and one where no b-tagged jet exists. The combination of the two bins defines the pre-tagged sample. The background contribution, and especially the W+jets events, is largely reduced in the signal region ($N_{j,25} \geq 4$, $N_{bj,25} \geq 1$) and the signal-to-background ratio (S/B), as given purely by the Monte Carlo, is at 4.7 for the μ-channel and 5.7 for the e-channel. For the pre-tagged sample the S/B is only 0.7 for the μ-channel and 0.8 for the e-channel, highlighting the importance of b-tagging.

[8]It should be reminded that the QCD Monte Carlo that are used are heavily filtered as discussed in the first chapter, and the filters are effectively cancelled out only at the last steps of the event selection procedure.

μ-channel

	Trigger & Data Quality cuts	Charge Lepton cut	$E_{\mathrm{T}}^{\mathrm{miss}}$ cut	Triangular cut
$t\bar{t}$	1172.5±2.0	624.6±1.8	545.3±1.7	522.3±1.7
W+jets	225296±143	172272±128	144261±117	143641±116
Z+jets	37991±57	15297±37	6036±24	5721±23
Single-top	360.8±0.7	210.0±0.8	177.3±0.7	170.0±0.7
Dibosons	418.6±0.8	218.0±0.7	170.5±0.6	164.5±0.6
QCD	$2.5 \cdot 10^6 \pm 3.0 \cdot 10^3$	19519±259	1906±80	333.7±34.1
Data	$4.818 \cdot 10^6$	323024	187246	181042

e-channel

	Trigger & Data Quality cuts	Charge Lepton cut	$E_{\mathrm{T}}^{\mathrm{miss}}$ cut	$M_{W,T}$ cut
$t\bar{t}$	1031.6±2.0	568.9±1.7	429.1±1.5	377.5±1.4
W+jets	220692±142	131356±104	52370±65	51736±65
Z+jets	34565±55	16984±39	655.0±7.6	234.2±4.5
Single-top	315.1±0.7	183.0±0.8	124.3±0.6	113.2±0.6
Dibosons	431.5±0.2	183.7±0.6	100.4±0.4	93.0±0.4
QCD	317293±1060	5229±141	366.2±37.3	113.4±20.9
Data	$8.119 \cdot 10^6$	372085	68101	63881

Table 4.7: *Event cut-flow from left to right for the Monte Carlo samples and the collected data. All numbers correspond to* 35.3 pb^{-1} *of integrated luminosity and all uncertainties are due to statistics.*

	μ-channel		**e-channel**	
	$N_{j,25} \geq 4$		$N_{j,25} \geq 4$	
	$N_{bj,25} = 0$	$N_{bj,25} \geq 1$	$N_{bj,25} = 0$	$N_{bj,25} \geq 1$
$t\bar{t}$	70.7±0.6	193.5±1.0	51.3±0.5	136.7±0.8
W+jets	270.1±3.1	25.5±0.9	160.4±2.0	14.7±0.6
Z+jets	21.6±1.2	2.9±0.4	18.8±1.2	2.1±0.3
Single-top	5.2±0.1	9.1±0.1	3.8±0.1	6.7±0.1
Dibosons	3.4±0.1	0.0±0.0	2.2±0.1	0.2±0.0
QCD	14.3±7.1	3.5±3.4	19.9±8.7	0.1±0.1
Total MC	385.3±7.9	234.5±3.7	256.4±9.1	160.5±1.1
Data	373	234	239	157

Table 4.8: *Event yields of the two channels after all the cuts including the jet cut, separated in an exclusive zero b-jets bin and an inclusive one b-jet bin. All numbers correspond to* 35.3 pb^{-1} *of integrated luminosity and all uncertainties are due to limited statistics.*

4.5 A preamble to the cross-section analysis

The cross-section analysis performed in this thesis will be presented in detail in chapter 6. However, some key aspects upon which the analysis depends are presented here.

The following general relation holds for any cross-section calculation:

$$\sigma = \frac{N_{signal}}{\int \mathcal{L} dt \cdot \epsilon_{signal}} \ , \tag{4.5.1}$$

where the N_{signal} is the number of identified signal events, $\int \mathcal{L} dt$ is the integrated luminosity and ϵ_{signal} is the selection efficiency of the signal events. In the most straight-forward method, the number of signal events can be extracted by subtracting from the collected data the expected background contribution (cut-and-count):

$$N_{signal} = N_{observed} - N_{background} \ ,$$

where the $N_{background}$ can be calculated by determining the normalization of each contributing process. For following this approach it is preferable to benefit from using a kinematic region where the S/B ratio is large so as to reduce the contribution on the uncertainty coming from $N_{background}$.

4.5.1 The template fit approach

A more evolved method is implemented for the analysis presented in this thesis. Firstly, an extended kinematic region is used that includes all events that pass the baseline selection but requiring at least three "good" jets with $p_{\mathrm{T}} \geq 25$ GeV. In this way, information from a significantly larger dataset is exploited, as opposed to the one from the signal region only. Based on this selection, four distinct regions are defined according to their jet multiplicity, all being mutually exclusive:

- Bin-30 : Exactly three jets ($N_{j,25} = 3$) with no b-jets ($N_{bj,25} = 0$).
- Bin-31 : Exactly three jets ($N_{j,25} = 3$) with at least one b-jet ($N_{bj,25} \geq 1$).
- Bin-40 : At least four jets ($N_{j,25} \geq 4$) with no b-jets ($N_{bj,25} = 0$).
- Bin-41 : At least four jets ($N_{j,25} \geq 4$) with at least one b-jet ($N_{bj,25} \geq 1$), which is the signal region defined previously.

Table 4.9 shows the event yields for the exclusive three-jet bins; the inclusive four-jet bins are already shown in 4.8.

The method, instead of simply relying on the counting of data events and subtracting the background, determines the contributions of both signal and background events, in each bin, from the data; practically constraining the result to the number of observed events. To achieve this, the signal and background are characterized by a single observable: the reconstructed invariant mass of the top quark. This is a straightforward choice considering that its distribution is expected to peak in the region of the true mass of the top providing a distinctive shape with respect to the shape of the expected background processes. Eventually, the shape of the observable is obtained individually for signal and background and each remains fixed to serve as template in the fitting procedure. The relative normalization between signal and background shapes becomes then the floating parameter.

	μ-channel		e-channel	
	$N_{j,25} = 3$		$N_{j,25} = 3$	
	$N_{bj,25} = 0$	$N_{bj,25} \geq 1$	$N_{bj,25} = 0$	$N_{bj,25} \geq 1$
$t\bar{t}$	53.6±0.5	103.8±0.7	38.2±0.4	76.3±0.6
W+jets	952.5±7.1	49.9±1.3	529.9±3.8	28.2±0.8
Z+jets	65.0±2.3	4.0±0.4	31.7±1.8	1.8±0.3
Single-top	12.8±0.2	17.6±0.2	9.1±0.1	12.2±0.2
Dibosons	13.8±0.2	1.2±0.05	8.3±0.1	0.7±0.03
QCD	28.4±7.1	3.6±3.6	27.0±10.0	11.5±6.7
Total MC	1126.1±10.3	180.1±3.9	644.2±10.8	130.7±6.8
Data	1027	213	594	173

Table 4.9: *Event yields of the two channels after all the cuts but with the jet cut requiring exactly three good jets. The events are separated in an exclusive zero b-jets bin and an inclusive one b-jet bin. All numbers correspond to* 35.3 pb^{-1} *and all uncertainties are statistical only.*

Calculating the M_{jjj} observable

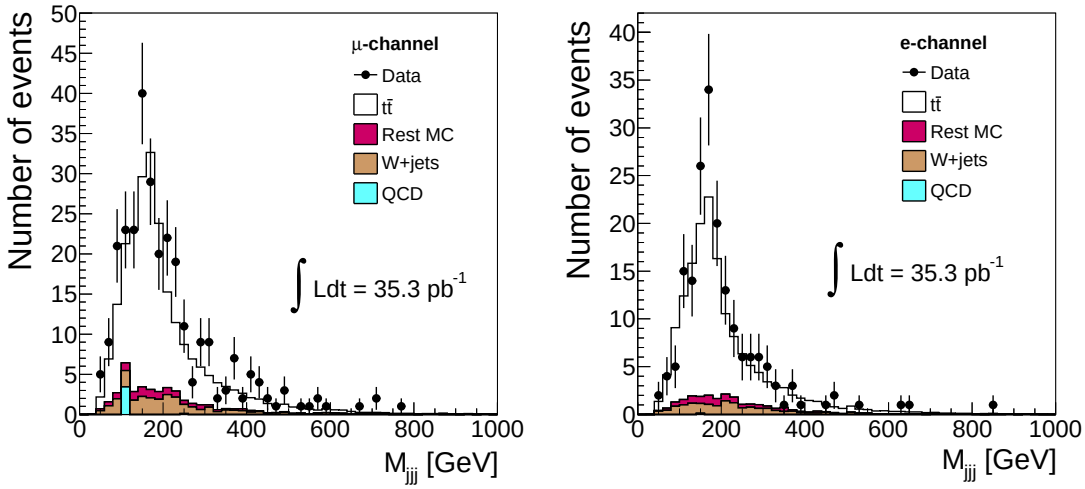
In the single-lepton topology the invariant mass of the top quark is derived from the hadronically decaying side of the event which, consisting of three jets, is fully reconstructable. However, given the increased jet multiplicity of $t\bar{t}$ events, especially due to ISR/FSR effects, it is possible to have several three-jet combinations and as a result further refinement is needed. Eventually, the M_{jjj} observable is created from those three jets that maximized the vector-sum of their transverse momenta. This choice is motivated by the fact that the top and anti-top quarks are created back-to-back in the transverse plane and any directional information must be retained in the daughter particles. Hence, the three jets originating from the hadronically decaying top are most likely to be boosted to the same direction, fulfilling the above requirement. In the present analysis, all "good" jets are taken into account for the above implementation. Figure 4.17 shows the M_{jjj} distributions for both channels after four "good" jets with $p_{\mathrm{T}} \geq 25$ GeV are requested and with at least one being identified as a b-jet.

4.5.2 The background contribution

From what discussed previously, it is evident that estimating the contribution of the background in the selected dataset is very important for any cross-section measurement. For the fit method, as it is followed in this thesis, the shape of the background is the crucial element.

Regardless of the choice, in the simplest approach the background can be estimated by employing the Monte Carlo samples. However, this also introduces large uncertainties for the following reasons:

- Firstly, as explained in greater detail in section 1.3.3, the cross-sections of the W+jets and QCD samples are known from theory with a significant uncertainty. Especially for the large jet multiplicity final states, the relevant normalization is not well known.
- Secondly, the estimated QCD cross-section is extremely large. As a result, the number of events produced does not exceed the equivalent of 10 pb^{-1} in integrated luminosity and subsequently the overall statistical uncertainty is large. Naturally, this affects also the shape of event-level observables since after the baseline cuts, which are shown earlier, only a handful of events remains which have typically large weights to account for the



(a) *μ-channel* (b) *e-channel*

Figure 4.17: *The three-jet invariant mass observable in the signal region, namely the Bin-41 as described in the beginning of section 4.5.1. All other requirements of the baseline selection for $t\bar{t}$ events are applied.*

normalization.

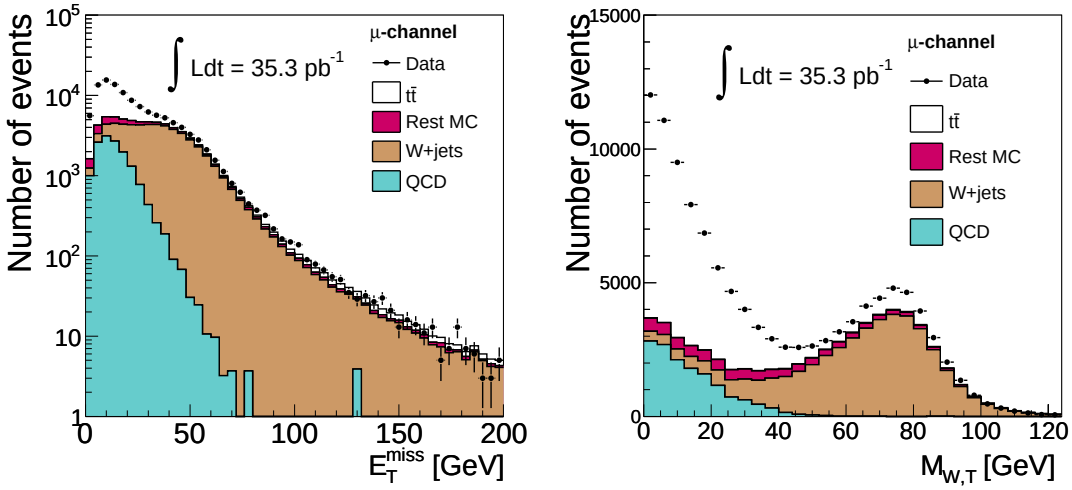
- Lastly, both the QCD and W+jets samples are produced using the `AlpGen` generator. As this is a leading order generator the shape of observables cannot be trusted a priori. A systematic uncertainty must, therefore, be introduced.

Besides from the large uncertainties, the lack of understanding is clearly demonstrated when comparing the Monte Carlo to the data. Figure 4.18 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $M_{W,T}$ distributions for the μ-channel after requiring the trigger, the data-quality and the charged lepton cuts, as well as at least one jet with $p_{\mathrm{T}} \geq 20$ GeV and $|\eta| \leq 2.5$. The rest of the top selection specific cuts are not applied. It is clear that in the lower values of both distributions, a region where the QCD is expected to be prominent, the Monte Carlo underestimates the data. Additionally, a non-negligible discrepancy exists at the region of the Jacobian peak of the $M_{W,T}$ distribution, which suggests that the W+jets normalization is not accurate.

The aforementioned issues, related to the use of Monte Carlo, can effectively be evaded if we adapt a method in which these background contributions are estimated from the data itself. This also has the advantage that the statistical uncertainty can only but decrease as more data are collected. In the following chapter, a method is implemented to extract the background shape directly from the data. This shape is then used for the cross-section measurement that is presented in detail in chapter 6.

Normalization of QCD

The normalization of the QCD after applying the event selection for the single-lepton $t\bar{t}$ topologies can be evaluated with data-driven methods. Typically, such methods are used for estimating the cross-section with the cut-and-count approach and in our analysis, as it will be presented, they are not needed since the total background normalization is the result of the fit. However, we do make use of these methods in order to provide an initial estimate of the background contribution of the QCD in parameters that serve as input to our fit, therefore reducing the dependance on Monte Carlo QCD. Two methods are implemented:



(a) *Missing transverse energy* (b) *Transverse W mass*

Figure 4.18: *The $E_{\mathrm{T}}^{\mathrm{miss}}$ and $M_{W,T}$ distributions comparison between Monte Carlo and data showing significant discrepancy especially in the lower values. Events are selected using the trigger, data-quality and charged lepton cuts, for the μ-channel and with at least one jet within $|\eta| \leq 2.5$ being above 20 GeV.*

- The **matrix method**: Implemented for the μ-channel. The QCD that is identified as signal in the μ-channel is mainly due to real muons that are coming from heavy flavor decays, hence they are not well isolated. The matrix method uses the intrinsic differences between the QCD muons and the isolated muons, such as the ones found from W or Z decays. Two samples are defined, one where the track isolation and calorimetric isolation criteria, as defined in section 4.3, are dropped (loose), and one where they are kept as in the "good" muon definition (tight). The following holds:

$$N_{sig}^{tight} = \frac{\epsilon_{sig}}{\epsilon_{sig} - \epsilon_{QCD}}(N^{tight} - \epsilon_{QCD} N^{loose}), \text{ and} \tag{4.5.2}$$

$$N_{QCD}^{tight} = \frac{\epsilon_{QCD}}{\epsilon_{sig} - \epsilon_{QCD}}(\epsilon_{sig} N^{loose} - N^{tight}), \tag{4.5.3}$$

  where N^{tight} and N^{loose} are the total amount of events in the tight and loose samples respectively, ϵ_{sig} and ϵ_{QCD} are the respective efficiencies for the selection of the two samples, and the final numbers N_{sig}^{tight} and N_{QCD}^{tight} are the events with a real isolated muon (signal) and with a non-isolated muon respectively. For the estimation of ϵ_{sig}, the tag-and-probe method on $Z \to \mu^+\mu^-$ decays is implemented, while the ϵ_{QCD} is calculated by measuring the ratio of tight events over loose events in a control region rich in QCD events and orthogonal to the final analysis sample which is then extrapolated to the signal region.
- The **template-fit method**: Implemented for the e-channel. The method uses a binned template likelihood fit on the $E_{\mathrm{T}}^{\mathrm{miss}}$ observable at the sideband region of $E_{\mathrm{T}}^{\mathrm{miss}} \leq 35$ GeV. The signal template is estimated from Monte Carlo, while the QCD template from a QCD-rich sample obtained after inverting the electron identification criteria. The selected electrons in this case (anti-electrons) would typically be accounted as jets in the baseline $t\bar{t}$ selection and consequently the $E_{\mathrm{T}}^{\mathrm{miss}}$ must be re-evaluated for the QCD-rich sample. The result of the fit in the sideband region is extrapolated to the signal region ($E_{\mathrm{T}}^{\mathrm{miss}} \geq 35$ GeV).

	μ-channel		e-channel	
	Pre-tagged	**Tagged**	**Pre-tagged**	**Tagged**
$N_{j,25} = 3$	121.4 ± 8.4	24.2 ± 3.4	62.0 ± 11.4	10.8 ± 8.6
$N_{j,25} \geq 4$	51.3 ± 5.6	13.0 ± 2.5	22.0 ± 8.0	8.6 ± 9.2

Table 4.10: *Data-driven estimates of QCD contribution per kinematic region after the baseline $t\bar{t}$ event selection for the μ-channel (matrix method) and e-channel (template fit method). Numbers taken from [159].*

Both the matrix and the template fit methods are documented in detail in [159]. The estimates on the QCD content for each is given in table 4.10.

4.6 Summary

In this chapter we discussed those ingredients that are needed to perform the $t\bar{t}$ cross-section measurement. We explained the reconstruction techniques for the various objects of interest and we highlighted the important parameters which allow us to efficiently select the $t\bar{t}(e)$ and $t\bar{t}(\mu)$ events. In addition, we presented the methodology, based on data-driven methods, for correcting certain parameters at the Monte Carlo level, hence improving the agreement between data and simulation. Subsequently, we showed and justified the requirements that were placed, at both particle-level and event-level, for selecting the events for our measurement. We separated the phase space of interest in four kinematic regions based on their "good" jets multiplicity, considering all events with three or more "good" jets, and whether they contain a b-tagged jet or not. In this way, we included, but kept exclusive, kinematic regions that are background dominated in order to utilize both the greater statistics and the background characteristics of the event sample. This is a necessary step for our cross-section measurement that is presented later on. We concluded the chapter with pointing out the importance of understanding the background contribution in our method. Our measurement has minimum reliance on the exact normalization of the background however the shape is regarded as an important element. The following chapter discusses a method with which the background shape is obtained from data, in view of the cross-section measurement that is presented in chapter 6.