# Universität Bonn

## Physikalisches Institut


# Measurement of jets production in association with a $Z$ boson and in the search for the SM Higgs boson via $H \to \tau\tau \to \ell\ell + 4\nu$ with ATLAS


Serena Psoroulas



Three measurements focussing on the understanding of jet final states in ATLAS, in di-jet, $Z$ and Higgs boson candidate events, using data corresponding to an integrated luminosity of 35 pb$^{-1}$ in 2010 and 4.7 fb$^{-1}$ in 2011, are presented.

In the first part, a calibration method, based on the transverse momentum balance in di-jet events, is described. The method is used to estimate the uncertainty of the jet energy scale in the forward region. The results show that the parton shower models are limited in reproducing the results in data, mostly for jets of low transverse momentum.

In the second part, the differential cross section measurement of the $Z \to \ell\ell +$ jets process is reported. Phase space regions not been previously studied at other experiments are investigated. The models used for the theory predictions provide a good description of the data, within the relative uncertainties.

In the last part, two contribution to the Higgs searches in the $H \to \tau\tau$ channel are shown: the modelling of the $Z \to \tau\tau$ background, and the modelling of jet final states. The $Z \to \tau\tau$ background is derived from data and validated in the $H \to \tau\tau \to \ell\ell + 4\nu$ channel. The modelling of jet final states in simulations is in good agreement with the data, when low-energy pile-up effects are subtracted.



CERN-THESIS-2012-149
28/09/2012



Physikalisches Institut der
Universität Bonn
Nußallee 12
D-53115 Bonn

BONN-IR-2012-11
Oktober 2012
ISSN-0172-8741

# Universität Bonn

## Physikalisches Institut


# Measurement of jets production in association with a $Z$ boson and in the search for the SM Higgs boson via $H \to \tau\tau \to \ell\ell + 4\nu$ with ATLAS


Serena Psoroulas



Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf der ULB Bonn `http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

| | |
|---|---|
| 1. Gutachter: | Prof. Dr. Norbert Wermes |
| 2. Gutachter: | Prof. Dr. Klaus Desch |
| Angenommen am: | 24.08.2012 |
| Tag der Promotion: | 28.09.2012 |

# Contents


**1 Introduction** **1**

**2 Theory predictions for physics at the LHC** **5**
2.1 The Standard Model of Particle Physics . . . 5
2.1.1 The Higgs mechanism . . . 7
2.2 QCD at high energy colliders . . . 8
2.2.1 The parton model of QCD . . . 9
2.2.2 QCD properties and jets . . . 10
2.2.3 Parton shower models . . . 11
2.3 $Z$ boson production in association with jets . . . 13
2.4 Higgs boson production at LHC . . . 14
2.4.1 Production and decay channels . . . 15
2.4.2 $H \rightarrow \tau\tau$ decay channel . . . 15
2.4.3 $Z$ boson as a background for $H \rightarrow \tau\tau$ searches . . . 18
2.4.4 Theoretical uncertainties on Higgs cross sections . . . 19

**3 The LHC and the ATLAS experiment** **21**
3.1 The Large Hadron Collider . . . 21
3.1.1 Pile-up . . . 23
3.2 The ATLAS detector . . . 25
3.2.1 Overall detector . . . 25
3.2.2 Coordinate system and conventions . . . 26
3.2.3 Inner Detector (Tracker) . . . 26
3.2.4 The LAr Calorimeter . . . 28
3.2.5 Hadron calorimeter . . . 29
3.2.6 Muon Spectrometer . . . 32
3.2.7 Trigger and Data Acquisition System . . . 32
3.2.8 Simulation of the ATLAS detector . . . 33

**4 Jet Algorithms in VBF Higgs searches** **35**
4.1 What is a jet? . . . 35
4.2 Jet reconstruction in ATLAS . . . 39
4.2.1 Input to jet reconstruction . . . 39
4.3 Jet Algorithm studies in Vector-Boson-Fusion simulated samples . . . 40
4.3.1 Motivation . . . 40
4.3.2 Analysis details . . . 42
4.3.3 Comparison of algorithms without pile-up . . . 44
4.3.4 Comparison of algorithms with pile-up . . . 49
4.4 Conclusions . . . 49

**5 Jet $\eta$-intercalibration in 2010 data at $\sqrt{s} = 7$ TeV** **51**
5.1 Strategy . . . 51
5.2 Jet energy scale calibration . . . 52
5.2.1 Hadronic showers in a calorimeter . . . 52
5.2.2 ATLAS calibration schemes . . . 54
5.3 *In-situ* $\eta$-intercalibration of the ATLAS detector using di-jet events . . . 58
5.3.1 Methods used in this analysis . . . 58
5.3.2 Data and Monte Carlo samples . . . 61
5.3.3 Di-jet events reconstruction and selection . . . 62
5.3.4 Results . . . 63
5.4 Outlook . . . 67

**6 Measurement of the production cross section for $Z/\gamma^*$ in association with jets** **71**
6.1 Data and simulation samples . . . 73
6.2 $Z \to \ell\ell$ + jets selection . . . 74
6.3 Detector level results . . . 78
6.3.1 Detector level validation . . . 78
6.3.2 Jet differential distributions . . . 80
6.4 Comparison with Monte Carlo LO and NLO pQCD predictions . . . 81
6.4.1 Jet systematic uncertainties . . . 81
6.4.2 Unfolding of detector results . . . 86
6.4.3 Combination of electron and muon channel . . . 88
6.5 Outlook . . . 89

**7 Search for the Standard Model Higgs boson in the $\tau^+\tau^-$ channel** **95**
7.1 Strategy . . . 95
7.2 Definition of analysis categories . . . 97
7.2.1 Jet topology in Higgs production channels . . . 97
7.2.2 Perturbative uncertainties in jet categories . . . 98
7.2.3 Definition of jet categories in the 2011 analysis . . . 100
7.3 Data and simulated samples . . . 101
7.4 Selection and reconstruction of physics objects . . . 103
7.4.1 Validation of the preselection in $Z \to \ell\ell$ + jets events . . . 104
7.5 Event selection . . . 110
7.6 Data-Driven Background Estimation . . . 116
7.6.1 $Z \to \tau\tau$ data-driven estimation: embedding method . . . 118
7.6.2 Validation of embedded $Z \to \tau\tau$ events in the $H \to \tau\tau \to \ell\ell + 4\nu$ analysis . . . 121
7.7 Systematic uncertainties . . . 122
7.8 Results . . . 125
7.9 Outlook . . . 126

**8 Outlook** **129**

**Bibliography** **131**

**List of Figures** **139**

**List of Tables** **143**

# Chapter 1

# Introduction

The main reason for the LHC physics program is the understanding of the mechanism of the Electroweak symmetry breaking. The Electroweak theory does not allow a mechanism to give to each particle its own mass. This is a fundamental problem, as not only we observe that particles do have mass, but the value of the masses of the force carriers influences the strength of the relevant force. The solution to this problem was provided by several authors [61, 70, 73–75, 82]. They showed that, if in nature an additional scalar field exists, with the characteristics of a non-zero vacuum expectation value, a spontaneous breaking of the Electroweak symmetry can occur. This spontaneous breaking gives a mass to the weak bosons, in such a way that the Electromagnetic and Weak interaction we experience is restored. The interactions of the fermions with this field also provides them with their masses. The breaking also produces a scalar boson, the Higgs boson.

This model explains the Electroweak symmetry breaking, but introduces an additional field, that depends on two fundamental parameters: the Higgs mass and the non-zero expectation value of the field. Because those parameters enter in the definition of the electroweak masses, it is possible to constrain them. The vacuum expectation value can be constrained with high precision using the measurement of the Fermi constant $G_F$: from this we obtain a value of 246 GeV. The Higgs mass, instead, can be constrained only from the radiative corrections to the top and the $W$ boson masses. The constraints are loose in this case, predicting the Higgs mass to be around 100 GeV. From other constraints, for example the breaking of the unitarity at 1 TeV that would occur in $WW$ scattering, we can expect the Higgs mass to be smaller than this value.

A "light" Higgs boson, i.e. a boson whose mass would be $\leq$ 200 GeV, is the favoured option considering the results of the electroweak fits, that include results from the previous generation of colliders[38]. For this reason, low-mass Higgs searches were carried out at the LEP and Tevatron experiments; but no evidence for the Higgs had been found in those searches.

The LHC was built to fully explore the mass range allowed by the unitarity constraint. A large effort has been directed towards the low-mass region, and this thesis is part of that effort. The latest results from ATLAS, that culminated with the discovery of a new particle consistent with the predictions for the Standard Model (SM) Higgs boson [21], are the outcome of this long process of preparatory studies. The discovery of a new particle in this energy regime will provide the informations we are missing to fully understand the SM, with possible consequences concerning cosmology as well. The consequences of this discovery go beyond the field of particle physics, involving all the studies about the origin and the fundamental behaviour of the universe.

The LHC explores an energy range not available at an accelerator before, and can deliver collisions at an incredibly high luminosity. Both features, however, require unprecedented precision both from the point of view of the operations, and from the point of view of the analysis. Studying physics at a new energy requires first of all the understanding of the physics at that energy. This is can be a difficult task: the models used are never perfect, and often tuned on some processes or energy ranges that don't ensure their reliability after an extrapolation to another regime. The

solutions to the issue might come either from tuning some parameters of the model, or from bigger changes to the model itself, that exploit the better understanding of the processes provided by the new measurements. The first task of the LHC experiments has been, thus, the measurement of known SM processes. In such processes, the reliability of the tools used by the analyses - the modelling of physics objects first, and then of fundamental interactions in pp collisions - has been tested.

**Outline** In this thesis, study on simulated data and three measurements, performed at the ATLAS experiment using the data from the first two runs, corresponding to an integrated luminosity of 35 $pb^{-1}$ in 2010 and 4.7 $fb^{-1}$ in 2011, are reported. The studies focusses on two topics: the understanding of Quantumchromodynamics (QCD) processes at the LHC, and the search for a light Higgs boson.

QCD is the theory that describes the interaction of hadronic particles. QCD processes are dominant in pp collisions; and because of their high cross section, the sample collected in 2010 is large enough to investigate them. Also, they are among the largest background for Higgs and new Physics searches: for this reason, they must be understood with a high precision.

In the context of the Higgs boson searches, I have focussed on the $H \rightarrow \tau\tau$ channel. This is a promising channel, because it provides a clean signature (provided that the $\tau$ leptons can be well reconstructed in the detector). The $H \rightarrow \tau\tau$ analysis relies on the excellent performance of the detector. The decay products are detected either in the tracking system (electrons, muons) or in the calorimeters (electrons, hadronic taus), while the contribution from the neutrinos is estimated from the missing transverse energy of the event ($E_{\mathrm{T}}^{\mathrm{miss}}$). The sensitivity relies on a precise prediction of the signal process as well as of the large background processes. Investigating the presence of additional jets (a bunch of hadronic particles; see Chap. 4) in the event, is useful to separate signal and background, and enhance the sensitivity. This stresses, again, the importance of understanding the hadronic interactions at the LHC energies, that had a central role in my whole thesis project.

**Jet algorithms studies in the context of the Higgs searches** The investigation of the Higgs boson production exploiting the jet characteristics started as a proof of principle in Aad et al. [9].That work investigated the possibility to extract the signal from the LHC data at 14 TeV. Since that publication, new technical development in the field of jet algorithms made more reconstruction algorithms available "on the market". A complete investigation has been carried on in all ATLAS analyses, to find the algorithm with the best performance. I have performed this study within the $H \rightarrow \tau\tau$ analysis, and the results are reported in Chap. 4. From this and other results, the default ATLAS algorithm has been replaced.

**Jet studies in ATLAS: calibration and uncertainties** During 2010, I have focussed on QCD understanding. I have first studied the performance of jet reconstruction in real collisions. Because there is no way to fully detect the energy deposited by hadronic particles in the detectors, we need to define a reference for the energy of a jet. The ATLAS calibration uses a reference from simulations; looking at the deviations between several simulation models and the measurements, it's possible to establish the precision of the calibration. This analysis, reported in Chap. 5, is a fundamental first step towards a successful physics program in ATLAS.

**Measurement of the production cross section for $Z/\gamma^*$ in association with jets** Studying the production of $Z$ bosons in association with jets is an important step in the understanding of

the underlying mechanisms of QCD. Also, the $Z$ is an important background for the $H \to \tau\tau$ decay channel, because it can lead to the same final state: the Higgs signal will show up as a small excess of events on the tail of the $Z$ resonance. I have focussed on this process in 2010, and the measurement of the differential cross section is reported in Chap. 6. This measurement provides the largest coverage in the phase space with respect to previous studies at hadronic colliders. My work has investigated the possibility to expand the rapidity reach of the measurement; the previous experience developed in the context of jet algorithms and calibration has been fundamental in reaching this goal. In addition, the measurement has investigated many variables that are important in a Higgs analysis, as the $Z \to \ell\ell$ + jets process can provide a clean control region for the $H \to \tau\tau$ channel.

**Search for the standard model Higgs boson in the $H \to \tau\tau \to \ell\ell + 4\nu$ channel** Finally, the results of the search for the Higgs boson in the $H \to \tau\tau \to \ell\ell + 4\nu$ decay channel are presented in 7. In addition to the investigation of the $Z \to \ell\ell$ + jets background, an investigation of the $Z \to \tau\tau$ background is performed. This is the main irreducible background to the $H \to \tau\tau$ process. Also, it provides additional challenges with respect to the $Z \to \ell\ell$ + jets background, due to the presence of the missing momentum carried away by the neutrinos in the $\tau$ decay. After summarizing the studies performed in the context of this search, the final limits on the production of the Higgs boson, presented as a ratio of the excluded cross section over the cross section predicted by the Standard Model, are shown.

**Public results** All the results obtained with ATLAS collisions data reported here have been published. The results on the jet calibration and uncertainties have been published in Aad et al. [10]. The measurement of the $Z \to \ell\ell$ + jets cross section is published in Aad et al. [17]. The results of the search for the Higgs boson in the $H \to \tau\tau$ channel have been submitted to JHEP; the preprint is public at `http://arxiv.org/abs/1206.5971v1`.

# Chapter 2

# Theory predictions for physics at the LHC

## 2.1 The Standard Model of Particle Physics

The current model of Particle Physics is known as the *Standard Model* (SM). It consists of a relativistic quantum field theory that includes all known forces of nature except gravity, since the latter is too weak at the energy scales typical for the other forces. The other forces are the Strong force, responsible for the structure of protons, neutrons and nuclei; the Electromagnetic force, that describes the interaction between electric charges; and the Weak force, that describes phenomena like the $\beta$ decay. According to the SM, the Universe is composed of elementary particles, that interact via one or more of the mentioned forces. In the theory, particles are described as fermion fields, and are divided into leptons (particles that don't feel the strong interaction) and quarks (particles that feel all interactions). Interactions can occur among these fermions, mediated via vector boson particles. A summary of all the particles described in the SM is shown in Fig. 2.1.

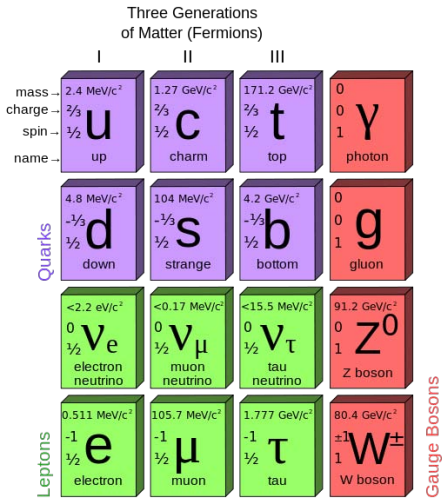



Figure 2.1: Scheme showing the elementary particles described in the Standard Model.

Quarks have been considered as elementary particles since the discovery of proton substructure. To reproduce the spins and charges of the known particles, each quark carries a fractional charge and is half-integer spin. Baryons, like the proton, are made of 3 valence quarks, while mesons, like the pion, are made of a quark-antiquark valence pair. The valence quarks determine the particle's quantum numbers, but the actual composition of the particle is only known on a statistical basis (see Sect. 2.2.1 for the proton case). Colour symmetry is responsible for the anti-symmetric property of the wave function: the quarks have an additional quantum number that represents their

behaviour under this symmetry. The colour symmetry is an exact symmetry, represented by the SU(3) symmetry group, and the generators for this group provide the mathematical form for the force carriers, the gluons. Gluons are massless neutral particles responsible only for the colour interaction. Because of the non-abelian nature of the symmetry, gluons interact not only with the quarks but also among themselves. The behaviour of quarks and gluons is described by Quantumchromodynamics (QCD); further comments on QCD, relevant for the analysis reported in this thesis, are in Sect. 2.2.

Concerning the electroweak interaction, experiments showed a pattern in the behaviour of the different particles; for example, the charges are the same for all neutrinos, charged leptons, and within two groups of quarks, and only some interactions between the particles are allowed. To reproduce these characteristics, they have been arranged in generations, as shown in Fig. 2.1. The particles inside a generation are represented as doublets (will then transform accordingly under $SU(2)_L$ transformations, that model the Weak interaction) and singlets (not affected by $SU(2)_L$). To reproduce the behaviour observed in nature, in particular the absence of flavour changing neutral currents and the neutral current observed for neutrinos, two additional effects are included: a mixing between quark families (the CKM matrix, that represents also the CP violating phase of quarks), and a mixing between the generators of the symmetry group $SU(2)_L \otimes U(1)_Y$ that represents the electroweak interaction[1], that provides two bosons charged under both electromagnetic and weak charge (representing the two charged $W$ bosons), a neutral weak boson (the $Z$ boson) and a neutral electromagnetic boson (the photon, affected only by the particle's charge). As SU(2) is non-abelian, also the weak bosons can interact among themselves.

The SM is a renormalizable theory. Local gauge invariance ensures renormalizability, because it fixes the form of the lagrangian density. It will include field terms, interaction terms between fields and particles, and because of the non-abelian structure of $SU(2)$ and $SU(3)$ also the interaction terms between the bosons related to those fields[60]. Local gauge invariance does not allow the introduction of a mass term for the gauge bosons; and this contradicts the experimental evidence. The fact that the electromagnetic and the weak interaction have a different strength at low energy must be included in the theory, and it is related to the finite mass of the weak bosons. The way those masses are introduced in the SM is via spontaneous symmetry breaking, i.e. a way to break the local gauge symmetry while still retaining renormalizability.

[1] Define $W_\mu^i$ ($i$ =1,2,3) and $B_\mu$ as the Electroweak gauge bosons ($SU(2)_L \otimes U(1)_Y$ respectively). When deriving the Lagrangian, the quadratic term in the boson fields is:

$$\mathscr{L} = \frac{v^2}{8}[(g_W W_\mu^3 - g'_W B_\mu)(g_W W^{3\mu} - g'_W B^\mu) + 2g_W^2 W_\mu^- W^{+\mu}]$$

where the choice of $W_\mu^\pm$ follows by the diagonalization of $W^{1,2}$: $W_\mu^\pm = \frac{1}{\sqrt{2}} W_\mu^1 \pm W_\mu^2$ . To take into account the difference in the neutral current interaction between neutrinos and charged leptons, the electrically neutral fields defined by $SU(2)_L \otimes U(1)_Y$ are diagonalized by a rotation:

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix}$$

The mixing angle value is fixed by the relative strenght of the coupling constants:

$$\sin\theta_W = \frac{g'^2_W}{g'^2_W + g_W^2} \simeq 0.23$$

### 2.1.1 The Higgs mechanism

One of the mechanisms to introduce spontaneous symmetry breaking in the SM is the Higgs mechanism. It introduces a single complex doublet of scalar fields:

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \tag{2.1}$$

It transforms as a doublet under $SU(2)_L$ and has a fixed hypercharge of $1/2$. The interaction between this field and the electroweak bosons is expressed by the following lagrangian density, invariant under $SU(2)_L \otimes U(1)_Y$:

$$\mathscr{L}_\phi = (\partial^\mu \phi^\dagger + i g_W W^\mu \cdot T \phi^\dagger + \frac{1}{2} i g'_W B^\mu \phi^\dagger)(\partial_\mu \phi + i g_W W_\mu \cdot T \phi + \frac{1}{2} i g'_W B_\mu \phi) - \mathscr{V}(\phi^\dagger \phi) \tag{2.2}$$

where $W^i_\mu$ ($i$ =1,2,3) and $B_\mu$ are the Weak and Electromagnetic gauge bosons respectively, $g_W$, $g'_W$ are the couplings to the two symmetry groups, and $T$ is the weak isospin matrix, responsible for the behaviour under $SU(2)_L$.

The potential that allows for spontaneous symmetry breaking is a potential where the field has a non-zero value in the vacuum state:

$$\mathscr{V}(\phi^\dagger \phi) = \lambda(\phi^\dagger \phi)^2 - \mu^2 \phi^\dagger \phi \tag{2.3}$$

as shown in Fig. 2.2.

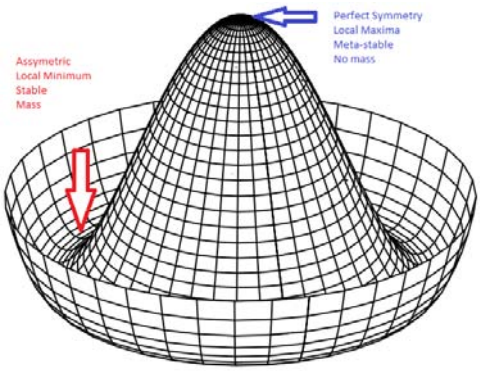



Figure 2.2: The “mexican hat” potential used in the Higgs field description.

The potential in Eq. 2.3 is invariant under rotation. In particular, there is a circle of degenerate minima, equal to:

$$|\phi| = \sqrt{\frac{\mu^2}{\lambda}} = \frac{\mathrm{v}}{\sqrt{2}} \tag{2.4}$$

Because of the invariance under $SU(2)$ rotations and the degeneracy of the potential, it is possible to parametrize the field as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ H(x) + \mathrm{v} \end{pmatrix} \tag{2.5}$$

where v is the expectation value at the minimum and $H(x)$ is a real function chosen for the parametrization of the field along the unbroken direction (i.e. the radial direction in Fig. 2.2).

In this parametrization, Eq. 2.2 shows quadratic terms that provide mass terms for the bosons:

$$\mathscr{L}_{\text{bosons}} = \frac{\mathrm{v}^2(\mathrm{g}_\mathrm{W}^2 + \mathrm{g}_\mathrm{W}'^2)}{8} Z_\mu Z^\mu + \frac{\mathrm{v}^2 \mathrm{g}_\mathrm{W}^2}{4} W_\mu^- W^{+\mu} \tag{2.6}$$

while no mass term exists for the $A_\mu$ field: the photon corresponds to a gauge transformation that leaves $\phi$ invariant. The masses for the charged and neutral bosons are slightly different, as confirmed in experiments, because the mass term for $Z^\mu$ includes also the U(1) coupling constant. The other quadratic terms in the lagrangian provide mass terms and self coupling terms for the Higgs:

$$\mathscr{L}_{\text{Higgs}} = -\mu^2 H^2 - \lambda \mathrm{v} \mathrm{H}^3 - \frac{\lambda \mathrm{H}^4}{4} = -\frac{1}{2}\mathrm{m}_\mathrm{H}\mathrm{H}^2 - \sqrt{\frac{\lambda}{2}}\mathrm{m}_\mathrm{H}\mathrm{H}^3 - \frac{\lambda \mathrm{H}^4}{4} \tag{2.7}$$

In the case of fermions, the Higgs field permits to introduce a Yukawa interaction of the form $g_f \bar{\psi}_f \psi_f \phi$, which will result in a mass term of the form $m_f = \frac{v g_f}{2}$ (as for the $W^\pm$ boson in Eq. 2.6). The values of the Yukawa couplings are determined from the masses measured in experiments, and don't fulfill a fundamental requirement.

The terms of the lagrangian density containing the $H(x)$ function from Eq. 2.5 are the interaction terms between the Higgs boson and the other particles. The coupling of the Higgs to the SM fermions can thus be written as:

$$g_{\text{Higgs}} = \frac{\sqrt{2} m_f}{v} = \frac{m_f g_W}{2\sqrt{2} m_W} \tag{2.8}$$

## 2.2 QCD at high energy colliders

The LHC is a proton-proton collider. QCD effects will be dominant and need to be understood for a reliable physics analysis. For this reason, part of this thesis investigated the features of two important QCD processes: di-jet production and the production of jets in association with a $Z$ boson. These two processes also highlight fundamental properties of QCD, as it will be explained in the next sections.

One fundamental property of QCD is that its coupling constant runs with energy, in a way such that it is relatively low at high energies and high at low energies. Because of this behaviour, quarks tend to form bound states very quickly, and are not observed as free particles in (low-energy) nature. Only baryons or mesons, i.e. color-neutral hadrons, are observed. However, the high energy interactions observed in the LHC can be considered as interactions between the protons constituents, because at those high energy the physics is sensitive to the proton structure. Proton-proton collisions are thus divided into two main steps:

1. the hard interaction is calculated (using perturbation theory) considering only the elementary partons, quarks or gluons, in the calculation of the matrix element

2. the result is then averaged over the composition of the proton, and taking into account higher order effects.

This separation is used extensively in QCD predictions: in the next sections, I will review its meaning and how it is implemented in particle physics studies.

### 2.2.1 The parton model of QCD

The parton model represents the proton as composed of several point-like constituents, that contribute to its total momentum. The interactions of hadrons can be explained as interactions of partons. The main assumption of this approach is that any interaction is much faster than eventual changes in the number and momenta of the partons: the structure is "fixed" during the interaction. Under this assumption, the parton structure is well defined and can be used in the calculations.

The $pp$ collision picture, according to the parton model, is sketched in Fig. 2.3. The cross section is defined as:

$$\sigma(P_1, P_2) = \sum_{i,j} \int f_i(x_1) f_j(x_2) \hat{\sigma}_{ij}(p_1 = x_1 P_1, p_2 = x_2 P_2) dx_1 dx_2 \tag{2.9}$$

In this equation, the cross section for the interaction, seen as a function of the momenta of the two colliding protons, is the sum over the constituents of the cross sections for the interaction between the constituents of proton 1 and 2, whose momenta are $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$, and integrated over the full momentum range, weighted by a function that express how likely it is to find a constituent of type $i$ and $j$ with fractions $x_1$ and $x_2$ of the proton momentum.

The interaction denoted by $\hat{\sigma}_{ij}$ is a "short-range" interaction ("short" here means on lenght scales much smaller than the nuclear dimensions). It can be computed in perturbation theory using the S-matrix formalism. The full interaction, however, can also be affected by "long-range" interactions, for example effects where the partons split into additional, softer partons, that move in collinear groups. Because of the running coupling constants, those long-range effects can be large, and can't be estimated by perturbation theory. In general, they cause terms that tend to infinity (singularities). They are called "long-range" because the time-scales of those effects are orders of magnitude larger than the short-range interactions. Eq. 2.9 shows how the interactions between hadrons can be expressed by the interactions between partons. This relation holds thanks to the weighting functions, called structure functions, that represent the partons' behavior, unknown from the experimental conditions.

The parton model has been accepted since the observation of Bjorken's scaling law in deep inelastic scattering experiments. The scaling law is however obeyed only approximately: higher order effects cause a scaling violation. The violation is related to the presence of effects that cannot be calculated in the frame of perturbation theory, as collinear gluon emission. The solution is to represent the structure functions as a calculable part and a "bare" part, that can only be measured from experiments. This separation is called *factorization* and can be defined with respect to an energy scale, that represents the limit value for the applicability of perturbation theory.

Factorization is valid at all orders in perturbation theory. It is a fundamental property of QCD. It means that it is possible to separate long- and short-range (i.e. non-perturbative and perturbative) contributions to a physics process, removing the singularities. At any given scale, the non-perturbative behaviour of the proton can be treated as the behaviour of its constituents, summed over the constituents distribution at the considered scale. These distributions are called "Parton Distribution Functions" (PDF). Also the singularities that occur in the calculation of QCD higher order diagrams can be treated in the same way. All collinear parton divergences that occur in the corrections of a physics process can be factorized into renormalized PDFs. This is a general feature

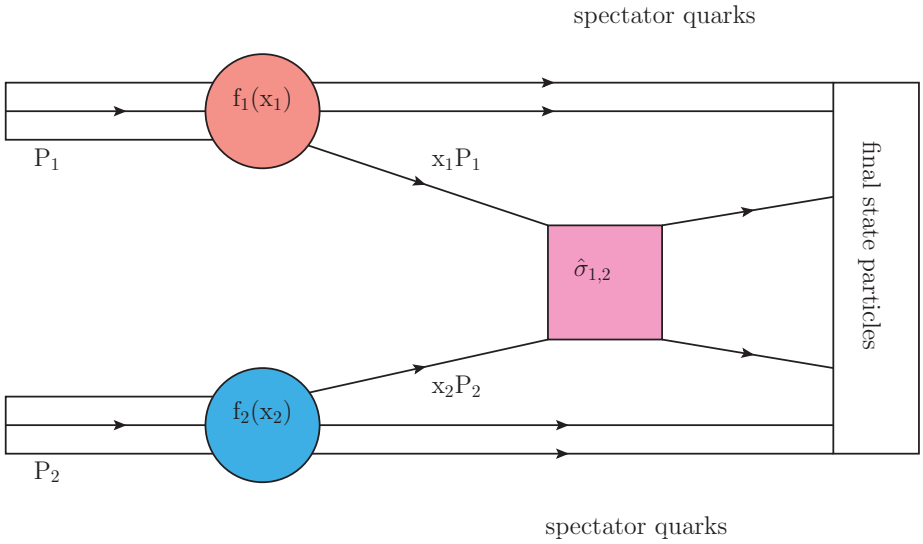



Figure 2.3: Parton model description of a hard scattering event.

of inclusive hard scattering processes in hadron-hadron collisions.

PDFs have been directly measured from fits to Deep-Inelastic-Scattering data. They show the distribution of partons as a function of the momentum. In Fig. 2.4, one of the latest results is reported.

### 2.2.2 QCD properties and jets

Partons don't exist as free particles in nature, but only inside colourless bound states. This property of "*colour confinement*" is one of the fundamental characteristics of QCD. The other main characteristic to bear in mind is the "*asymptotic freedom*", i.e. the fact that the coupling constant of QCD runs with energy, such that it is very small at high energies (and low distances) and increases at lower energies.

These two properties of QCD will obviously affect the partons that take part to an interaction, and will determine the final state observed in the detector:

- because of asymptotic freedom, the collision between two partons can be divided into a short- and long-range contribution. Asymptotic freedom is the physical reason behind factorization.
- the hard scattering is a short-range event, and can be treated with perturbative QCD.
- long-range effects are represented in the process of *fragmentation* of the partons. Additional partons will be produced. The treatment of these long-range contributions is analogous to the treatment of the initial state via PDFs.
- further decrease of the energy will cause the effects of confinement to become important. In this process, the quarks and gluons will combine into colourless hadrons (*hadronization*). Perturbation theory is not able to describe it; phenomenological models are used to make predictions for the final state as it could be observed in a detector.

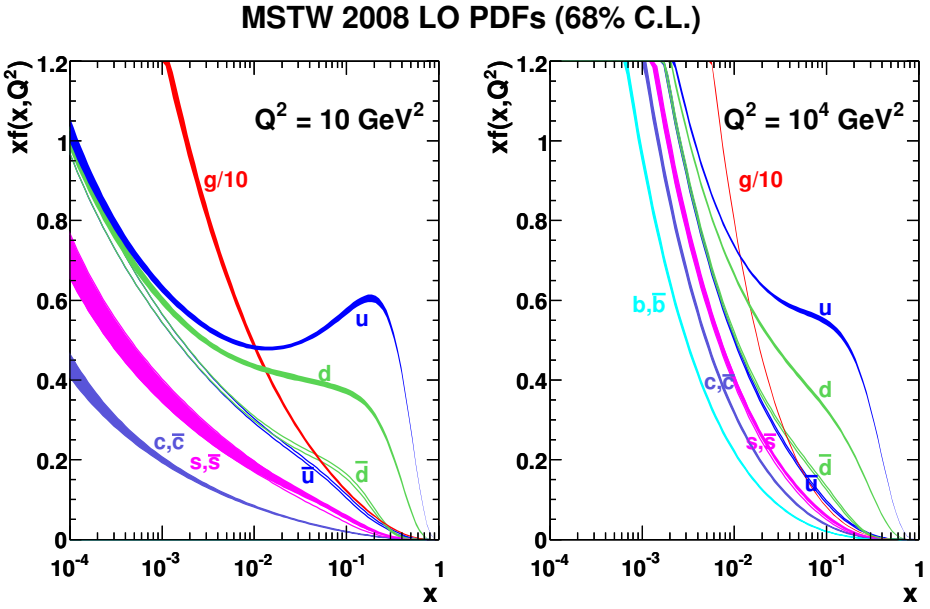



Figure 2.4: MSTW2008 Parton Distribution Function fit [91].

The result of a $pp$ collision are several particles, leptons and hadrons. Some of them will show a large transverse deviation with respect to the beam axis: they are coming from the partons from the hard scattering event, their large deviation due to the large momentum transfer of the scattering event. Other hadrons will be along the beamline, and make the beam remnants that will not be detected.

The easiest way to extract information about the hard scattering event from the hadrons in the final state is to build an algorithm that groups together hadrons in a meaningful way, such that the characteristics of the cluster are in a direct relation with the partons in the final state. This algorithm is called "*jet algorithm*" and the cluster is called "*jet*". A jet algorithm has to be well defined theoretically, and the requirements will be fully explained in Chap. 4. Fig. 2.5 shows schematically what is meant by a jet, from the initial partons through fragmentation and hadronization.

It is important to remark that we need to use "infrared safe" and "collinear safe" variables when defining a jet, as well as when defining any meaningful variables in QCD. Variables are safe when they are not sensitive to effects occurring in those regions of phase space where perturbation theory is not applicable, and that lead to infinities in fixed-order calculation. In the calculation of infrared safe variables, as inclusive cross sections, the infinities cancel out. We will treat in details this aspect in Chap. 4.

### 2.2.3 Parton shower models

QCD predictions are affected by large uncertainties. Complete perturbative calculations are available up to next-to-next-to-leading-order (NNLO) in most cases, and increasing the precision is not realistic, as the number of diagrams involved increases roughly factorially with the order in the perturbative expansion. In some regions of the phase space it is not possible to neglect higher orders, because they are fundamental to make sure that the resummation works and that the diagrams are finite: a typical example is collinear parton emission. The approach commonly used is the "*parton*

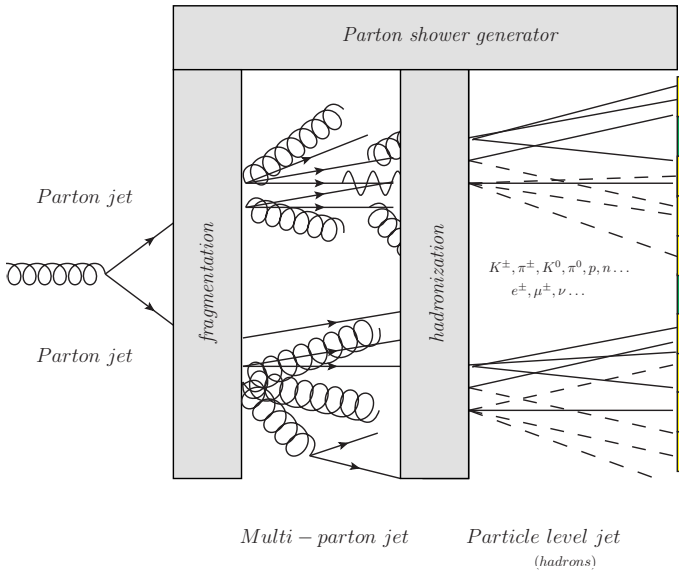



Figure 2.5: Schematic diagram showing the meaning of fragmentation and hadronization in a parton shower (for illustrative purpose only).

*shower*" approach: this approach is implemented into modern Monte Carlo generators, and takes into account these enhanced terms, regularizing infinities using a cutoff energy.

The parton shower approach is an approximate treatment of QCD, that can provide reliable predictions for the hadronic final state. The meaning of the cut-off energy is to separate the region where the perturbative approach is reliable from the one where its predictions are not reliable any more; in this way, the generator can treat the full evolution of the system, from the parton scattering event to the production of the hadrons. The perturbative part is described using Sudakov form factors and the evolution equations: in this way, fragmentation can be correctly described in the generator. When the initial partons are "split" into partons with energy of the order of the cut-off energy, the non perturbative treatment starts. This part of the parton shower algorithm represents the hadronisation using phenomenological models, because no satisfactory theoretical treatment can be provided by perturbative QCD[2].

The parton shower can also be used for the description of the fragmentation and hadronisation process alone, when it is interfaced with another generator, responsible for the matrix element calculation ($\hat{\sigma}_{ij}$ in Eq. 2.9). The partons produced in the hard scattering event are "matched" to those of the parton shower generators, and then the evolution and hadronisation are treated as in a normal parton shower generator. Such solutions are usually used when the prediction of the hard scattering event provided by the parton shower generator does not satisfy the precision requirements for the analysis. One example of such a "mixed" model will be used in Chap. 5 and 6, and is common in many analyses because of its advantages with respect to the simple parton shower model.

The parton shower approach takes into account both the hard scattering event and the PDFs, as well as the singularities that affect the fragmentation and hadronisation process. However, it

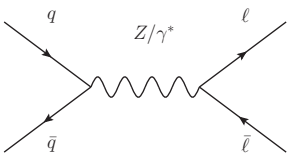
[2]Parton shower models need a tune to data to provide a reliable description of the event at a certain energy scale. Tunings of PYTHIA[105] and HERWIG[54], two of the most important parton shower generators, have been performed at LHC.

might be still limited in treating the non-singular part of the cross section. Higher order corrections in QCD are not small, but as said earlier often require too complex calculations to be estimated. Those effects may impact the value of the total cross section in a visible way. A common way to take into account those corrections is to define a "*K-factor*", i.e. a correction factor that corrects the cross section calculated with the Monte Carlo generator at a finite order to the value calculated at a higher order in the perturbative series. The K-factor can then be applied as a normalization correction to the predicted number of events.

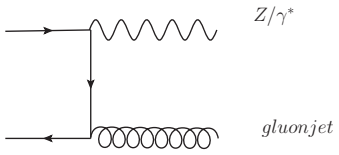
Parton shower generators are widely used in all HEP experiments. In ATLAS, the generator PYTHIA[105] has been considered as a Monte Carlo reference for the jet studies performed in 2010. In the thesis (see Chap. 5) I have investigated the precision of the jet modelling provided by PYTHIA comparing data and Monte Carlo events in an *in-situ* analysis, i.e. an analysis whose quantities don't need any Monte Carlo prediction or assumption to be derived. Also other parton generators (e.g. HERWIG[54]) have been compared to data in this analysis.

## 2.3 $Z$ boson production in association with jets

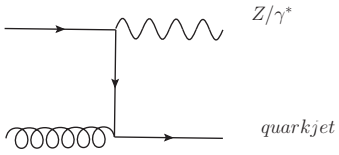
With respect to QCD studies performed with jets, as inclusive jet cross section measurements, measurements involving vector bosons can provide a cleaner final state, that permits a better understanding of the higher order QCD effects. Those are of fundamental importance in vector boson production, as they are responsible for the production of vector bosons with a high transverse momentum.



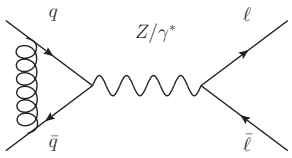

(a) Feynman diagram for Drell-Yan process.




(b) Feynman diagram for production with a gluon jet in the final state.




(c) Feynman diagram for production with a quark jet in the final state.




(d) Feynman diagram for QCD loop correction in Drell-Yan process

Figure 2.6: Feynman diagram for production and production of 1 additional jet in the final state.

The cross sections for multi-jet production in association with a $Z$, defined as $Z+\geq N_{\text{jets}}$, will in general have a perturbative expansion like this:

$$\sigma_{Z+\geq N_{\text{jets}}} = \sum_{N \geq N_{\text{jets}}} \alpha_S^N a_N \tag{2.10}$$

It is clear to see that the cross section for the different jet multiplicity is expected to scale as a function of $\alpha_S^{N_{jets}}$, the strong coupling constant at the power of the jet multiplicity considered. Terms with $N > N_{jets}$ are higher order QCD corrections.

When considering these corrections, the diagrams present infrared divergences. Fig. 2.6 shows the relevant diagrams at order $\alpha_S$. Divergences involve either virtual gluon corrections graphs, or real gluon corrections (emission of real gluons from one leg), or also quark gluon scattering process, where the $Z$ is radiated off the scattered quark. The factorization theorem, however, guarantees that for these corrections the cancellation occurs, at each perturbative order. While the proof of factorization at all orders is still discussed by some authors, the possibility to measure the cross sections at different orders (i.e. for different jet multiplicity) and compare it with the predictions would provide an indirect confirmation. Also, higher order corrections in QCD might be relatively large, and have a large impact on the precision of the prediction for this process.

The measurement of the $Z$ cross section in association with jets, and its comparison with several generator predictions, has been performed with 35 pb$^{-1}$ and is reported in Chap. 6. To study the process with high precision, the analysis has focussed on the di-electron and di-muon final state. The di-tau final state presents higher uncertainties related to the reconstruction of the tau leptons. In particular, I have focussed on the di-muon final state: the probability of mis-reconstruct a muon as a jet or vice-versa are very low and, for this reason, it is very useful to study the hadronic part of the final state.

Several Monte Carlo predictions have been compared with the measurements, in order to estimate how well commonly used generators can model this process. It is important to note that the most important generators used for these kind of processes (ALPGEN[90] and SHERPA[66]) are tree-level matrix-element generators: they compute only the first term of the perturbative series ($\alpha_S^{N_{jets}} a_{N_{jets}}$ from Eq. 2.10) for final states $Z+ \geq N_{jets}$. Up to $N_{jets} \geq 5$ has been considered. An agreement or disagreement with the data could shed light on both the tree-level calculation, as well as on the missing corrections. The measurement has also been compared with a NLO generator (BLACKHAT[41]).

Provided that the jet definition is infrared safe, the differential cross section as a function of the jet properties can provide important informations when compared with the theory. In this case, factorization ensures that the long-range effects show up only in collinear and infrared divergences which cancel because of unitarity of the time evolution operator of the partons that originate the jet. Also in this case, the accuracy of tree-level generators has been studied with respect to the measurement and to NLO calculations. Because jet production in $Z \to \ell\ell$ + jets events, as shown in Eq. 2.10, depends strongly on the order considered in the prediction, the phenomenological models used in parton shower generators might provide an inaccurate description of the process: the analysis confirmed this hypothesis.

## 2.4 Higgs boson production at LHC

The work presented in the last part of this thesis is part of the general ATLAS effort for the Higgs discovery. In 2011, the integrated luminosity was not high enough to ensure the discovery of the Higgs, and the results have been used to exclude a large part of the mass range. The results shown in Chap. 7 were also included in that combination.

The preliminary results on the combination of 2011 and 2012 data show a $5\sigma$ excess at a mass value of 126.5 GeV, shown in Fig. 2.7. The ATLAS results are consistent with the analogous results from CMS. The discovery of this new particle will be followed by additional studies targeting its

properties; first of all the decays into $b\bar{b}$ and tau pairs have to be observed to confirm or reject the SM Higgs boson hypothesis. In particular, this effort will still benefit from the work performed in 2011 shown in this thesis, that for the first time exploited the full potential of the $H \to \tau\tau$ channel to derive the exclusion limits.

Since the first studies of the ATLAS discovery potential, the Collaboration has particularly focussed on the understanding of the low mass region, that is favoured by the SM. The $H \to \tau\tau$ channel is one of the most important channels in that region, because it gives access to the Higgs coupling to the fermions, it has a relatively good signal to background ratio, and it is particularly sensitive with respect to the MSSM Higgs.

### 2.4.1 Production and decay channels

As shown in Eq. 2.6 and 2.8, the Higgs coupling constant to the other particles is proportional to their mass. This is fundamental for predicting the possible production and decay modes investigated in experiments.

At the LHC environment, the most likely production process are shown in Fig. 2.8, and their absolute contribution to the total cross-section at 7 TeV is shown in Fig. 2.9. The Higgs mass is a fundamental parameter of the SM: it can't be derived from other quantities, but only constrained from other measurements.

The gluon-fusion process ($gg \to H$) does not occur at leading order: its cross section nevertheless is very high because of the $x$ values sampled at LHC, and the large value of the gluon PDF at that scale. Similar argument holds for the Vector-Boson-Fusion (VBF) process: it is a leading-order electroweak process, with a relatively high cross section at the LHC because of the large sea quarks PDF. These two processes are dominant in the whole mass range investigated at the LHC, as shown in Fig. 2.9. Other processes have been investigated in the past but have a non-negligible impact only at low mass. One of them is the Higgs strahlung or associate production process, also called "$VH$" because in this process the Higgs is produced together with another vector boson.

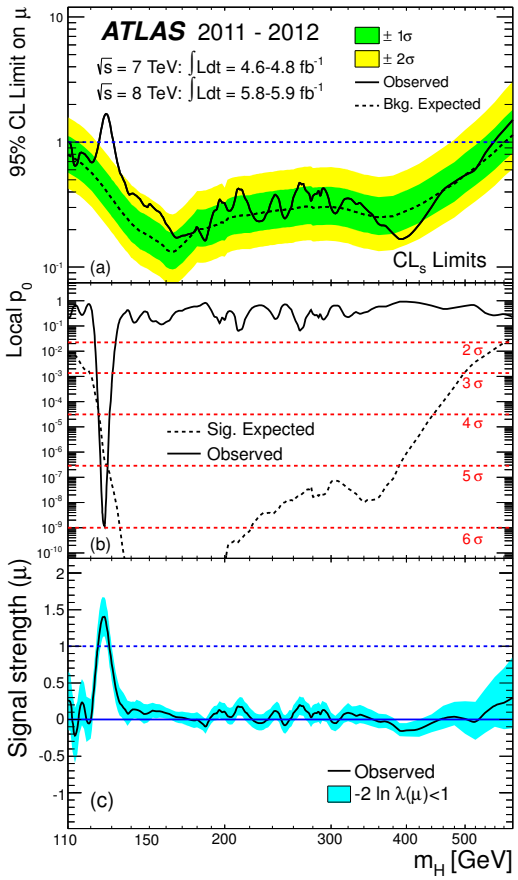
Because the decay is partly constrained by the possibility of producing the daughter particles, the Higgs decay modes depend on the Higgs mass as shown in Fig. 2.10.

For $M_H < 140$ GeV the main decay channel is in bottom pairs, i.e. the most massive particles available in that mass range. Also the decay into $\tau$ pairs provides a good experimental signature and a high branching ratio. The decays into charm and gluon pairs are not considered, despite their relatively high branching ratio, because of the experimental difficulties related to the discrimination of the signal from the background. The $H \to \gamma\gamma$ decay mode has a low branching ratio because it does not occur at leading order (as the photons are massless, they are not coupled to the Higgs), but only at higher order via a top or $W$ loop. It is anyway used in searches because it provides the best mass resolution and a very clean signature. Decay in two $W$ or $Z$ can also occur, one boson being off-shell.

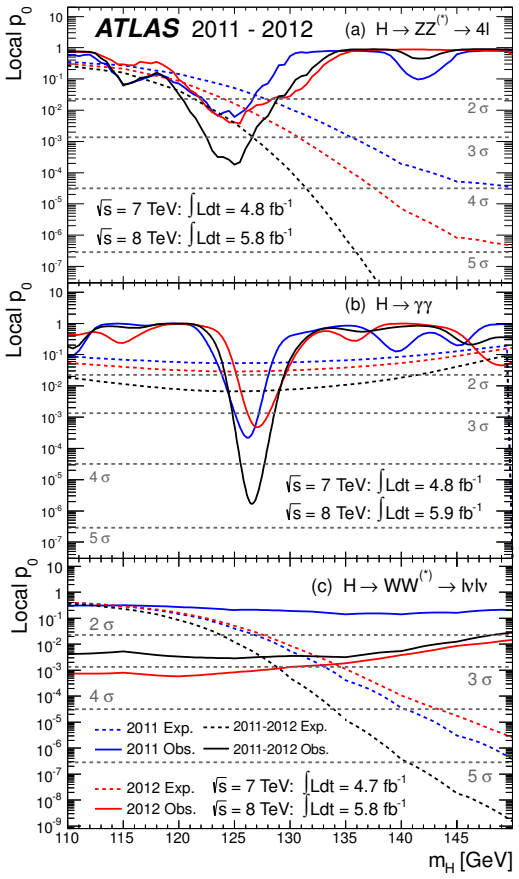
As soon as two real bosons can be produced, those decay channels become dominant. The ratio of the two decays depends on the ratio between the couplings to the Higgs and on the higher number of states available for the $WW$ case. At very high masses, as soon as top pairs can be produced, also this decay mode plays a role.

### 2.4.2 $H \to \tau\tau$ decay channel

The search presented in this work is performed in the $H \to \tau\tau$ final state. As shown in Fig. 2.10, the decay channels into $\tau$ pairs provide one of the highest branching ratio at low mass. In the Higgs




(a) Combined search results: (a) The observed 95% CL upper limit on the signal strength as a function of $m_H$ and the expectation under the background-only hypothesis. The dark and light shaded bands show the one sigma and two sigma uncertainties on the background-only expectation. (b) The observed local $p_0$ as a function of $m_H$ and the expectation for a SM Higgs boson signal hypothesis at the given mass. (c) The best-fit signal strength $\hat{\mu}$ as a function of $m_H$. The band indicates the approximate 68% CL interval around the fitted value.

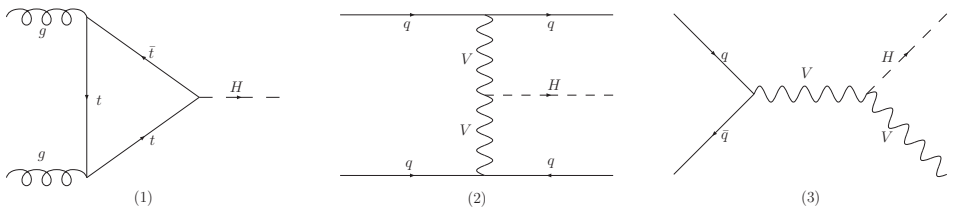



(b) The observed local $p_0$ as a function of the hypothesized Higgs boson mass for the (a) $H \to ZZ^{(*)} \to \ell\ell\ell\ell$, (b)$H \to \gamma\gamma$ and (c) $H \to WW^* \to \ell\nu\ell\nu$ channels. The dashed curves show the expected local $p_0$ under the hypothesis of a SM Higgs boson signal at that mass. Results are shown separately for the 7 TeV data (dark, blue), the 8 TeV data (light, red), and their combination (black).

Figure 2.7: Latest results from ATLAS: a) the exclusion limit, the local p0 and the signal strength, and b) the p0 values in the different channels, characterizing the observed 5-sigma excess at 126.5 GeV[21]. For an explanation of the kind of plots shown, see Sect. 7.8.

mass range $100 - 150$ GeV, the branching ratio goes from 8% to 1.8%; above 150 GeV drops so quickly that it does not contribute in any visible way to the total width any longer. In the SM, no other channels involving leptons are expected to be visible with a moderate integrated luminosity, because of the very low mass of the other leptons; for this reason, measuring the rates of this decay channel is the way to access the Higgs coupling to leptons. On the other hand, it is complementary to the other accessible channels in the same mass range, enhancing the overall sensitivity.

Tau leptons decay into other leptons or hadrons via weak interaction, via a $W$ boson. Due to lepton number conservation, every decaying tau produces a tau-neutrino in the final state. The




(a) $gg \to H$ (b) VBF (c) $VH$

Figure 2.8: Feynman diagrams at the leading order for the production of the Higgs boson at LHC.

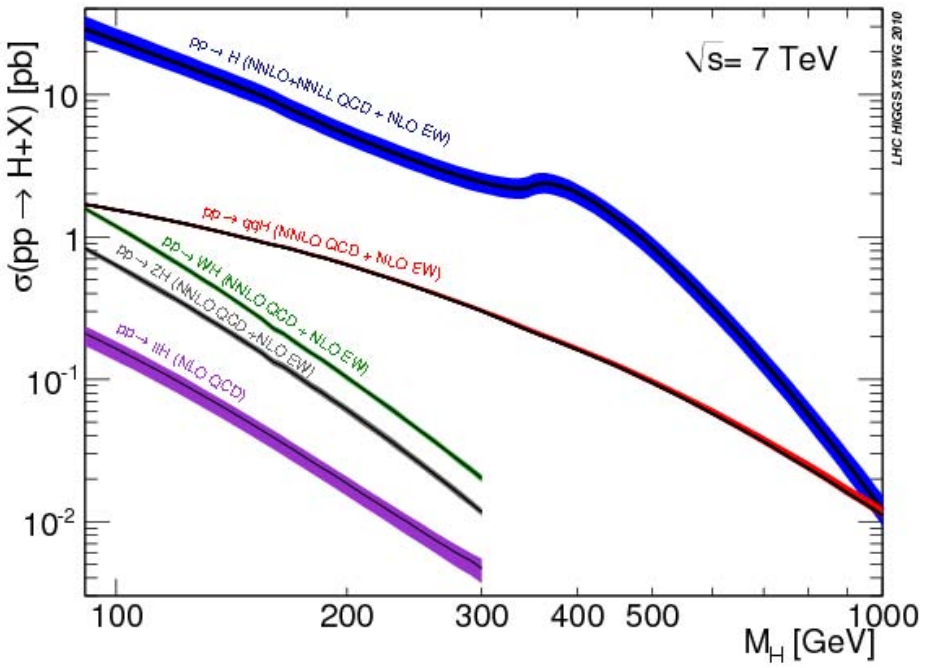



Figure 2.9: The SM Higgs production cross-section at $\sqrt{s} = 7$ TeV. [85]

neutrino will not be detected in a collider detector: for this reason, the reconstruction of the tau kinematics relies on the ability to estimate this missing momentum. At high energy colliders, the tau leptons are usually produced with a high Lorentz boost, resulting in the collinearity of the tau decay products, that helps in inferring the direction of the missing momentum from the neutrinos produced in the decay.

The decays into hadrons have the largest branching fractions, but the modelling and their detection efficiency can be affected by large uncertainties. Decays into leptons have a lower, but still large, branching fraction. These quantities are reported in Table 2.1. The final state into leptons is the cleanest to be detected in an experiment with good tracking performance, but it is difficult to distinguish from prompt lepton production if the missing momentum is not well estimated.

Thus, the $H \to \tau\tau$ analysis requires excellent performance of the experiment. Due to the small rate of signal production and large backgrounds, particle identification must be excellent and optimized specifically for this channel. A good estimate of the transverse momentum carried away by the neutrinos is needed as well.

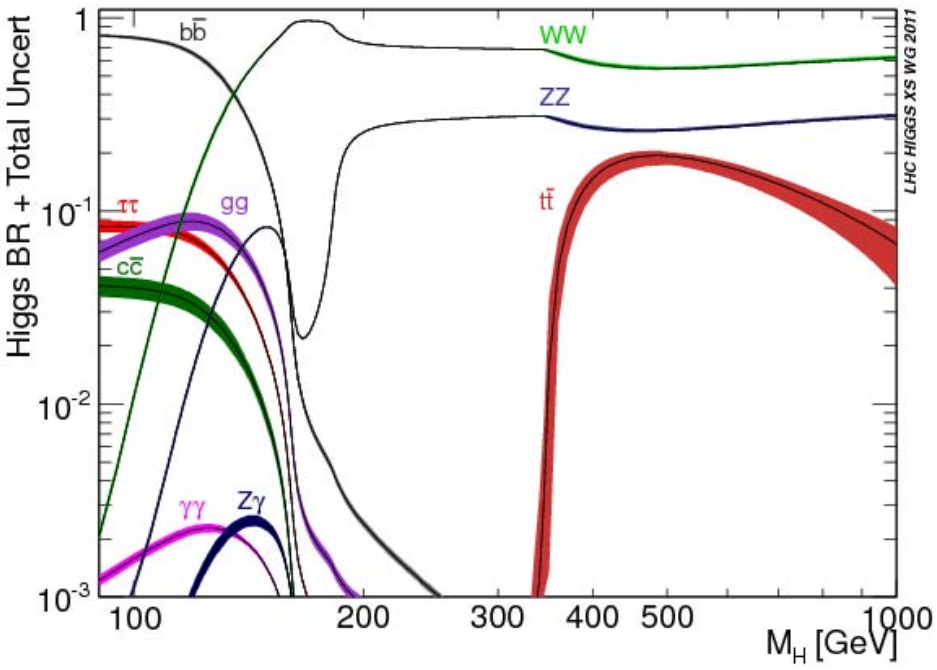



Figure 2.10: The SM Higgs branching ratios as a function of Higgs mass $M_H$ [86]

The low Higgs mass range where this decay mode is visible has two characteristics: a very small natural decay width for the Higgs, and the small difference with the $Z$ boson mass. The first leads to an invariant mass peak that is dominated by the experimental resolution; while the second introduces a potentially large source of uncertainty. Discriminating between $Z$ and Higgs processes is thus a central issue in this analysis.

### 2.4.3 $Z$ boson as a background for $H \to \tau\tau$ searches

The SM Higgs boson has the same charge as the $Z$, but different spin and mass. The last two characteristics will impact the kinematics of the decay and might help in discriminating the two processes. On the other hand, the cross section for $Z$ boson production is higher than the one

| Properties | |
|---|---|
| Spin | 1/2 |
| Mass ( MeV) | $1776.82 \pm 0.16$ (PDG average) |
| Mean life ($10^{-15}$s) | $290.6 \pm 1.0$ (PDG average) |
| **Decays** | |
| $\tau^\pm \to \mu^\pm \nu_\mu \nu_\tau$ | $17.41 \pm 0.04\%$ (PDG fit) |
| $\tau^\pm \to e^\pm \nu_e \nu_\tau$ | $17.83 \pm 0.04\%$ (PDG fit) |
| $\tau^\pm \to$ hadrons$^\pm \nu_\tau$ | $64.68 \pm 0.16\%$ (from PDG values) |
| $\tau^\pm \to h^\pm$ neutrals $\nu_\tau$ ("1-prong") | $64.68 \pm 0.11\%$ (from PDG values) |
| $\tau^\pm \to h^\pm h^\pm h^\mp$ neutrals $\nu_\tau$ ("3-prong") | $14.57 \pm 0.07\%$ (PDG fit) |

Table 2.1: Tau lepton properties, from Beringer et al. [42].

for Higgs production by four orders of magnitude. Because of the similarities between the two processes, this process is one of the most important irreducible backgrounds for searches at low mass considering a fermionic decay of the Higgs boson.

As the suppression of the $Z$ background is not completely possible, it is of fundamental importance to carefully estimate it. To estimate the contribution of this process to the total yield of events, first we need a reliable value for the inclusive cross section, as well as the values for the differential cross sections. The inclusive cross section is known up to a precision of few percent[16], but the precision gets worse when the events are split according to the parton multiplicity. In the analysis reported in Chap. 6, several differential cross sections, used to highlight the difference between the Higgs signal and the background, have been carefully studied to check the agreement between the prediction and the data.

In the $H \to \tau\tau \to \ell\ell + 4\nu$ channel, studied in Chap. 7, the main background is indeed $Z$ boson production, with all lepton flavours to be taken into account ($\tau$ considered only in case of a leptonic decay). Even though the $Z \to \ell\ell$ + jets analysis shows a reasonable agreement between data and the simulations, the strategy is to not use the simulation predictions blindly in the analysis, because their uncertainties could impact strongly the analysis. An example is the possibility to distinguish between the $Z \to \tau\tau$ background and a Higgs signal: experimental effects can affect significantly the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, fundamental for the mass reconstruction and the separation between the two processes. A poorly modelled $E_{\mathrm{T}}^{\mathrm{miss}}$ can have a large impact on the limit derived. For this reason, the $Z \to \tau\tau$ background is derived from the data, using the so-called "embedding method" (see Sect. 7.6.2). The $Z \to \ell\ell$ + jets background, more independent from $E_{\mathrm{T}}^{\mathrm{miss}}$ modelling because no real $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected in the event, was instead estimated from simulations, and corrected using data-Monte Carlo comparisons performed in control (i.e. signal-free) regions.

### 2.4.4 Theoretical uncertainties on Higgs cross sections

The values predicted by the theory are affected by uncertainties. The uncertainties arise from two sources: the missing higher-order corrections yield the "theoretical" uncertainties, while the experimental errors on the SM input parameters, such as the quark masses or the strong coupling constant, give rise to the "parametric" uncertainties.

Concerning the branching ratio, the full calculation of the uncertainty is reported in LHC Higgs Cross Section Working Group et al. [86]. The relative importance of the theoretical and parametric uncertainties depends on the channel, with decay channels into quarks and gluons being more affected, on average, by parametric uncertainties than the other channels. On average, the uncertainties are of the order of 5%.

For the production processes, the treatment is more complicated, because it needs to include the limited knowledge of the PDFs. The production processes will be affected in a different way by the PDFs, because of the differences in the initial state. The full calculation of the uncertainties on the inclusive cross section is reported in LHC Higgs Cross Section Working Group et al. [85]. The uncertainty shows almost no trend with the Higgs mass, and are of the order of 20% for $gg \to H$, 3% for VBF, and $4-5$% for VH.

When considering the differential distributions, other uncertainties arise. In the case of $gg \to H$, the fact that the process is not a leading-order process, and the need for a high precision, suggests an effective field theory approach. This is in general reasonable, mostly for low $m_H$, but could introduce distortions in some of the distributions, most notably in case of large Higgs $p_{\mathrm{T}}$, where the process is sensitive to the bottom and top quark finite masses (not considered in an effective approach). This distortion can be of $\mathcal{O}(10\%)$ in these regions, while are within few percent in the

total cross section and in the rapidity distribution. The analyses fix this issue implementing a reweighting of the finite-order predictions to the values predicted at a higher order of the calculation.

Other very useful exclusive approaches are those that separate events with additional partons in the final state, from those with no additional partons. This is motivated by the background composition of the Higgs signal, that is different in the two cases. The prediction is accurate enough for VBF and VH. The effective approach used in $gg \to H$, instead, induces large logarithms to appear when the $gg \to H$ cross-section calculation is split according to the number of partons in the final state (they would cancel out only in the inclusive cross section). Those logarithms can't be resummed and their contribution must be evaluated for each case and then combined. The full treatment is reported in LHC Higgs Cross Section Working Group et al. [86]; the procedure to estimate those uncertainties reliably and the way it has been implemented in our analysis will be summarized in Sect. 7.2.

The last ingredient to take into account for a correct estimate of the theoretical uncertainty is the additional approximations that are done in the generators commonly used for the predictions at the detector level, and not only for the very precise calculations. For the experimental analysis, Monte Carlo generators that provide a finite-order calculation of the cross section are used. The Higgs cross section working group has carefully studied the differential distributions for the Higgs in several generators and compared this with the next-to-next-to-leading-order (NNLO) prediction [86], to estimate the uncertainty of the generators. In general, the results obtained with those generators have been reweighted to correct their output to the highest order level. This is particularly important for QCD effects, that are estimated via parton shower methods and provide a LO precision. Those effects affect largely all production processes, and have been carefully investigated for two processes: $gg \to H$ (where partons in the final state can be created only from QCD radiation) and VBF (where two leading partons in the final state are expected at leading order). The prescription of the LHC Cross Section Working Group has been implemented in the analysis, as will be explained in Sect. 7.3.

# Chapter 3

# The LHC and the ATLAS experiment

The Large Hadron Collider has been conceived as a global project. This effort has an importance beyond the understanding of the missing pieces of particle physics, and inspects other issues of modern physics, for example the lack of a candidate for dark matter[69]. The project was approved in 1991, as natural continuation of the LEP experiments.

## 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring-superconducting hadron accelerator and collider, installed in the existing 26.7 km tunnel that was constructed between 1984 and 1989 for the CERN

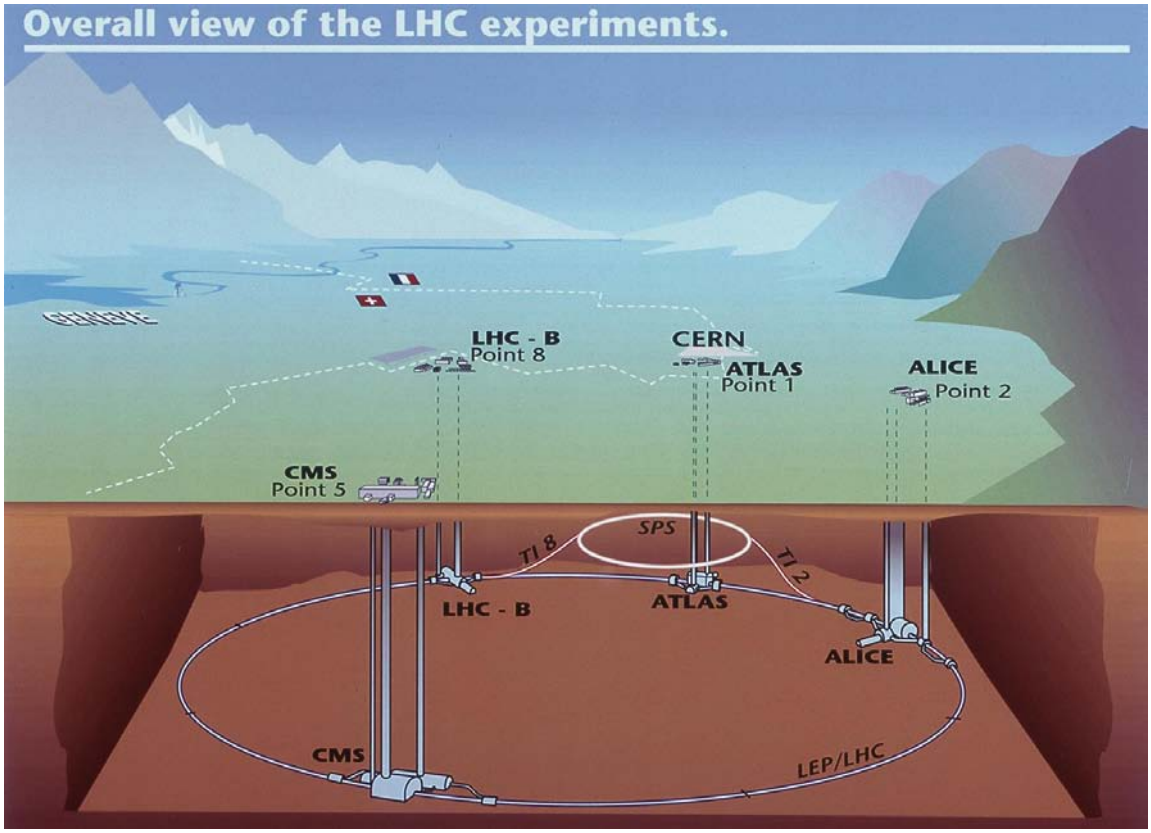



Figure 3.1: Representation of LHC.

LEP collider. The LEP tunnel has eight straight sections and eight arcs, and lies between 45 m and 170 m below the surface, near Geneva, across the Swiss-French border.

The tunnel in the arcs has an internal diameter of 3.7 m, which makes it extremely difficult to install two completely separated proton rings. This hard limit on space led to the adoption of the twin-bore magnet design[1], shown in Fig. 3.2. The extremely high energy target led to the use of superconducting magnets to reach the high currents needed. The LHC magnet system cools the magnets to a temperature below 2 K, using super-fluid helium, and could operate at fields above 8 T.

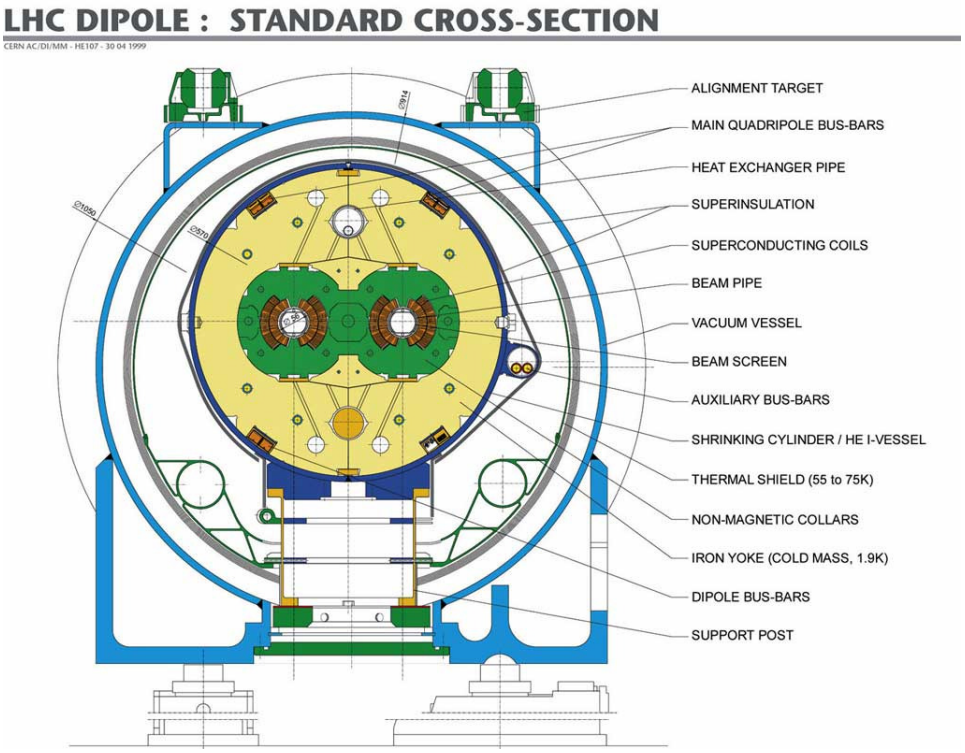



Figure 3.2: LHC dipole section, schematic drawing.[2]

Table 3.1 lists the fundamental parameters for the LHC machine, both for the pp and the heavy ion collisions.

Inside the LHC ring, bunches of up to $10^{11}$ protons collide, by design, every 25 ns (40 million times every second). Such a high interaction rate is needed because of the small cross sections expected for the process of interest. The inelastic pp cross section is 80 mb; thus the LHC is expected, at design energy and bunch settings, to produce a total rate of $10^9$ inelastic events per second. This puts a serious experimental difficulty, because it implies that interesting hard collisions will occur together with additional 23 inelastic events per bunch-crossing on average. This problem is called *pile-up*.

| Particles collided | pp | Pb Pb |
|---|---|---|
| Maximum beam energy (TeV) | 7.0 | 2.76 TeV/n |
| Luminosity ($10^{30}$ cm$^{-2}$ s$^{-1}$) | $1 \times 10^4$ | 0.001 |
| Time between collisions (ns) | 24.95 | 99.8 |
| Initial luminosity decay time, -L/(dL/dt) (hr) | 14.9 | 10.9 - 3.6 |
| Turn-around time (min) | 60 | 60 |
| Injection energy (TeV) | 0.450 | 0.1774 TeV/n |
| Particles per bunch (units $10^{10}$) | 11.5 | 0.007 |
| Bunches per ring per species | 2808 | 592 |
| Average beam current per species (mA) | 584 | 6.12 |
| Interaction regions | 2 high L + 1 | 1 dedicated + 2 |
| Dipoles in ring | 1232 main dipoles | |
| Quadrupoles in ring | 482 2-in-1<br>24 1-in-1 | |

Table 3.1: Fundamental parameters for the LHC machine.

### 3.1.1 Pile-up

The data collected in 2010 and 2011, even though not yet at the design energy, have allowed an extensive analysis of the detector performance. Fig. 3.3 shows the number of interactions per bunch crossing. It is clear that, most of all for the 2011 run, events with additional interactions are dominant. The mean number of interactions per crossing corresponds the mean of the poisson distribution on the number of interactions per crossing. It is calculated from the instantaneous luminosity as:

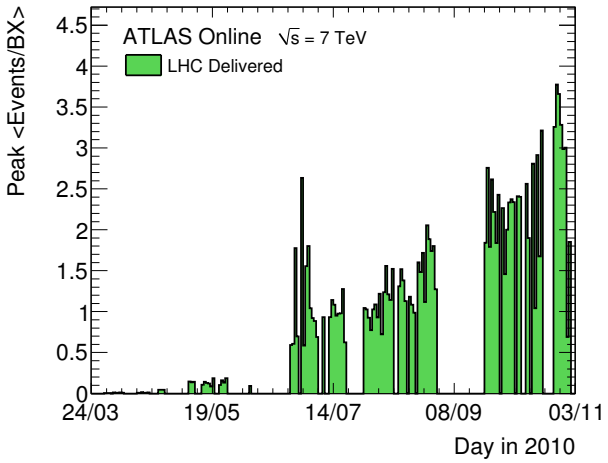
$$\mu = \frac{L * \sigma_{inel}}{n_{bunches} \nu_{rev}} \tag{3.1}$$

where $L$ is the instantaneous luminosity, $\sigma_{inel}$ is the inelastic cross section (71.5), $n_{bunches}$ is the number of colliding bunches and $\nu_{rev}$ is the LHC revolution frequency. More details can be found in Aad et al. [12].

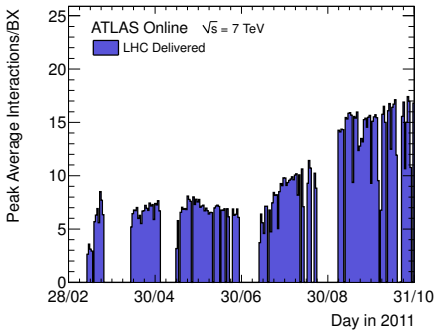
The so-called pile-up effects are the effects visible in runs where the value of $\mu$ is on average larger from one. These effects are due to either additional pp interactions in the same bunch crossing of the event of interest, or detector signals due to collisions which occured a bunch crossing before the event of interest, but that are reconstructed one bunch crossing later because of the large integration time of some detectors. The first case is called *in-time pile-up*, the second *out-of-time pile-up*. The latter effect usually affects the signal in the calorimeters, causing some cells to fire one bunch crossing later than the event they belong to. The former, instead, causes a larger number of particles to be produced, because the presence of many interactions in the same event will in general cause a larger multiplicity per event.

In 2010, only in-time pile-up affects the analysis: the strategy to correct for this effect was to apply an event-by-event correction to the signal recorded in the calorimeter, evaluated from

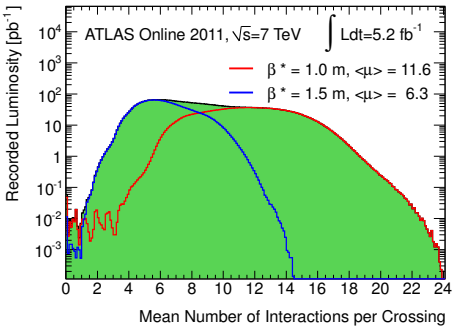
[1]Design proposed by John Blewett at the Brookhaven laboratory in 1971.




(a) **Number of Interactions per Crossing**. The average number of events per beam crossing at the start of an LHC fill versus day. The mean number of events per crossing is calculated from the per bunch luminosity assuming a total inelastic cross section of 71.5. The online luminosity measurement is used for this calculation as for the luminosity plots. Only the maximum value during stable beam periods is shown.




(b) **Number of Interactions per Crossing**. The maximum mean number of events per beam crossing versus day. This version shows the average value for all bunch crossings in a lumiblock. The online luminosity measurement is used for this calculation as for the luminosity plots. Only the maximum value during stable beam periods is shown.




(c) **Number of Interactions per Crossing**. Shown is the luminosity-weighted distribution of the mean number of interactions per crossing for 2011. The plot is shown for data taken before and after the September Technical Stop where the $\beta^*$ was reduced from 1.5m to 1.0m. The integrated luminosities and the mean $\mu$ values are given in the figure. The entries at $\mu \simeq 0$ arise from pilot bunches that were present during many of the early LHC fills. The luminosity in these bunches is >100 times smaller than in the main bunches resulting in values $\mu < 0.1$.

Figure 3.3: Luminosity plots for the 2010 and 2011 runs[1].

data. The $\mu$ values observed in 2010 are quite low, and no additional requirement is needed for the analysis. In 2011, instead, both in-time and out-of-time pile-up impact the measurement. To remove in-time pile-up, a check of the origin of the particles using tracks (see later) can be performed: only the particles coming from the collision of interest are considered. The out-of-time

pile-up effect on the reconstructed calorimeter energy was instead estimated as a function of the number of primary vertices in simulations, and cross-checked with data. Both corrections aim to remove any possible dependence of the measured quantities on the vertex multiplicity.

## 3.2 The ATLAS detector

The ATLAS detector started commissioning operations in autumn 2007 at the LHC Intersection Point 1, and started taking stable beams data[2] since beginning of 2010. The collaboration started in the early 90s, as the joined effort of two previous concepts for a new detector at the LHC accelerator, when the LEP collider at CERN was still in operation. In 2004 a combined test beam for the first time took data with all the detector components (small prototypes or single modules) working togheter, using a pion beam accelerated by the PS facility. The data have been used for initial commissioning and calibration of several parts of the detector; in particular, they have provided the initial calibration for the calorimeters.

In 2004 the first parts of the detector were installed in the cavern, while the last part was inserted in 2007. Because of problems at the LHC facilities, however, beam operations started only in late 2009.

For the main goals and a more complete description of the ATLAS experiment, we refer the reader to *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN* [4] and Aad et al. [31].

### 3.2.1 Overall detector

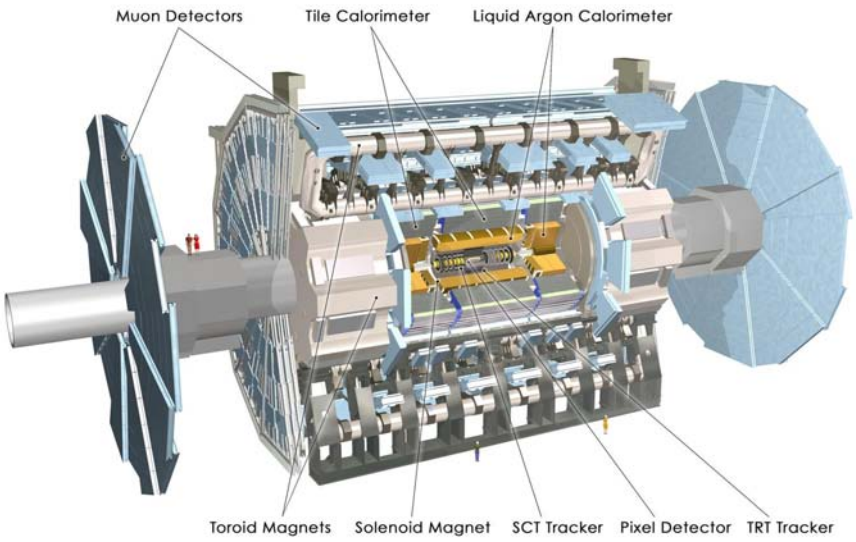



Figure 3.4: Cut-away representation of the ATLAS Detector. [31]

[2] "Stable beams" namely operations with beam conditions that are stable and safe enough to operate of the whole sub-detector components of ATLAS. All the data used in physics analyses come from a sub-set of stable beams data. The selection criteria used in each analysis will be summarized in the relevant chapters.

The ATLAS detector is composed of several stratified subsystems, responsible for different measurements, that will provide the complete information needed for the physics analysis.

The principal requirements for LHC detectors are mainly due to the high energy and luminosity. Thus the detector has strict constraints on resolution, timing performance, geometry, and radiation hardness; they can be mainly summarised as:

- Fast and radiation-hard on-detector electronics. The high fluxes require a high detector granularity.
- Large acceptance, almost all solid angle covered.
- Good charged particle momentum measurement and reconstruction efficiency, to be measured with a very precise tracker. To observe secondary vertices with high precision, the vertex detector should be as close as possible to the interaction region.
- Good calorimetry performance to identify particles and measure their energies; by the means of an electromagnetic calorimeter for electron and photon identification and measurements, and a hadronic calorimeter, to measure jets and missing transverse energy.
- Good muon identification and measurement - especially for high $p_T$ muons.
- Highly efficient trigger on low momentum objects, with sufficient background rejection, to achieve an acceptable trigger rate.

The ATLAS detector reaches 44 m in length along the beam axis, while its height is 25 m. It is composed of 6 different detector subsystems and 2 magnet subsystems.

### 3.2.2 Coordinate system and conventions

Conventionally the z direction is set along the beam axis, while x and y are in the transverse plane. Polar coordinates are often used in the transverse plane, with the $R$ coordinate describing the radial position from the beam axis, and the $\phi$ coordinate describing the azimuthal angle. The conventional coordinate for angular position with respect to the beam axis is the *pseudorapidity*, defined as $\eta = -\ln\tan(\theta/2)$, where $\theta$ is the angle with the beam axis.

Due to the shape of the detector, the central region is also called the *barrel*, while the two sides are called *end-caps*.

### 3.2.3 Inner Detector (Tracker)

The *Inner Detector* is the innermost part of the detector, and provides tracking measurements and reconstruction of secondary vertices. It is immersed in a 2 T magnetic field provided by a *superconducting solenoid magnet*, placed just outside the Inner Detector volume. It is sketched in Fig. 3.5, where also the dimension of the detector and its coverage in $\eta$ are highlighted. The maximum coverage of the detector is $|\eta| < 2.5$.

The dimensions and characteristics of such a system are determined by space requirements: on one side, the inner radius must be larger than the beam pipe radius, on the other side, the outer radius must be smaller than the solenoid needed to curve charged particles in the innermost part of the detector. The solenoid dimensions are limited by the calorimeter requirements, that are physics requirements related to the spatial resolution of electromagnetic showers. At the same

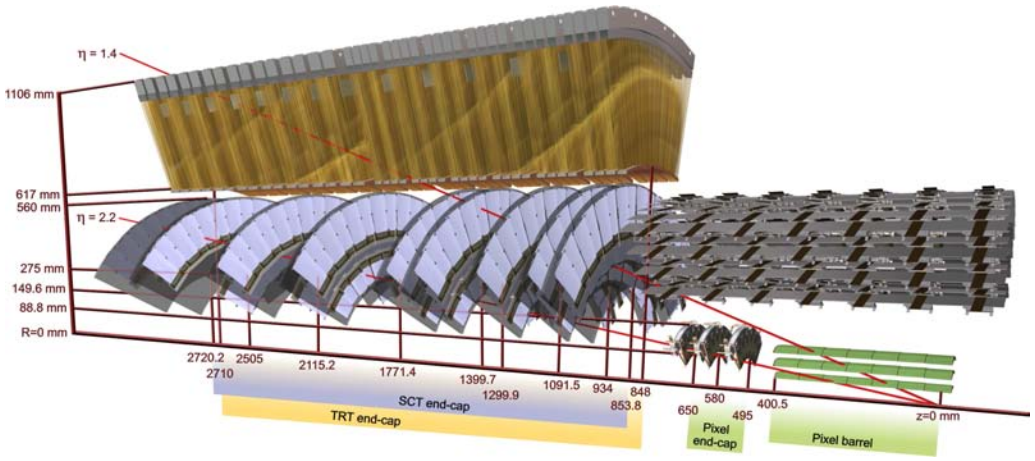



Figure 3.5: Drawing showing the sensors and structural elements traversed by two charged tracks of 10 GeV $p_T$ in the end-cap inner detector ($\eta = 1.4$ and 2.2). [31]

time, the need to minimize multiple scattering of the particles to avoid momentum resolution degradation requires to minimize as much as possible the amount of material in the detectors before the calorimeter. In addition, the occupancy conditions at the LHC put the detector electronics and its radiation hardness to the limit of the available technology.

To meet the low material and occupancy requirements, silicon detectors were chosen for the inner layers. The innermost section of the tracker consists of a *Pixel detector*, that provides precision measurements of the points along a track. This detector is made of 1744 modules, for a total of about 80 million channels. On-detector electronics permit a fast processing of the signal, necessary to cope with the high rate foreseen. Due to the high radiation dose, however, the innermost layer of this detector is expected to degrade in performance after few years of nominal operations; for this reason, an additional silicon pixel layer will be inserted into the detector during the first long LHC shutdown (planned in 2013), mitigating the inefficiency effects.

The second silicon system of the ATLAS tracker is a strip detector (*SCT, SemiConductor Tracker*), a 4-layered system where two strip silicon sensors, tilted by 40 *m*rad, are overlayed and equipped with fast on-detector electronics for signal processing. The reduction of the occupancy with the increase of the detectors inner radius allowed the usage of strips instead of pixels, without a degradation of the resolution performance.

The last detector is a *Transition Radiation Tracker* (TRT). It is a gaseous straw tube detector, that contributes to the electron identification thanks to the usage of transition radiation as a detection mechanism. Its occupancy as a standalone detector, however, is higher than the ATLAS requirements, as gaseous detectors tend to suffer from higher occupancies than silicon detectors. For this reason, the TRT provides a large number of hits along a particle track (about 36 hits), that, correlated with the high-precision measurements provided by the silicon detectors, allows a robust measurement of the track (improving the momentum resolution by a factor of 2) and reduces the total occupancy of the system.

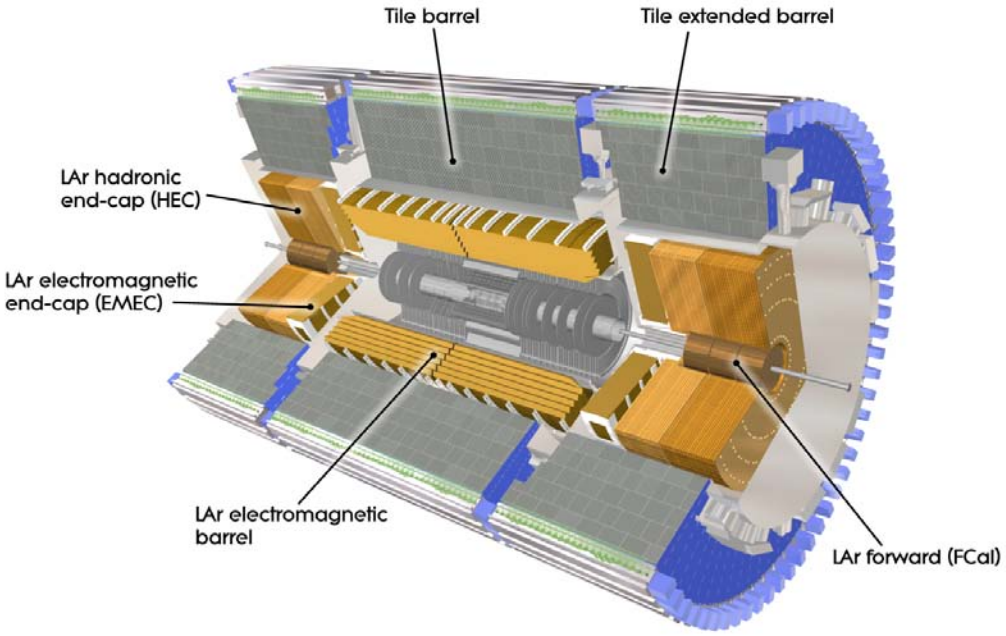



Figure 3.6: Cut-away representation of the calorimetry system. [31]

### 3.2.4 The LAr Calorimeter

The first detector outside the solenoid is an electromagnetic sampling calorimeter. It covers the pseudorapidity range $|\eta| < 3.2$; it is divided into a barrel part ($|\eta| < 1.475$) and two end-cap components ($1.475 < |\eta| < 3.2$). It is a sampling calorimeter with lead as absorber and Liquid Argon (LAr) as active volume; it shares the vacuum vessel with the central solenoid, using a single cryostat, minimizing the amount of material in front of the calorimeter and so multiple scattering effects that cause degradation of the resolution.

The Barrel calorimeter consists of two identical calorimeters separated at z=0 by a 4 mm gap. The end-caps are divided into two coaxial wheels, covering two different regions ($1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$). The region with $|\eta| < 1.8$ is also equipped with a *presampler* detector, to correct for the energy lost in the solenoid, before entering the calorimeter; it is built from an active LAr layer of 1.1 cm in the barrel region (0.5 cm in the end-cap region). The presampler is placed inside the cryostat, just before the solenoid.

The calorimeter has to provide precise measurements both in position and in energy, and it is optimized for the region $|\eta| < 2.5$. To reach the goal, it is layed out as accordion-shaped kapton electrodes with lead absorber plates: the accordion geometry provides $\phi$ symmetry without azimuthal cracks, and the lead thickness is optimised to maximise the electromagnetic energy resolution. In the precision region, the calorimeter is also segmented into three longitudinal sections, as shown in Fig. 3.7. The high granularity improves the precision of the measurement, and is higher in the first layer, to detect the point where the particle enters the calorimeter, and to be able to detect the shower shape (useful to separate single photons from neutral pion decays). The end-cap region has first of all to fulfil the environmental requirements, that constrain the dimension of the electrodes and influence the performance of the detectors. It provides a coarser granularity and

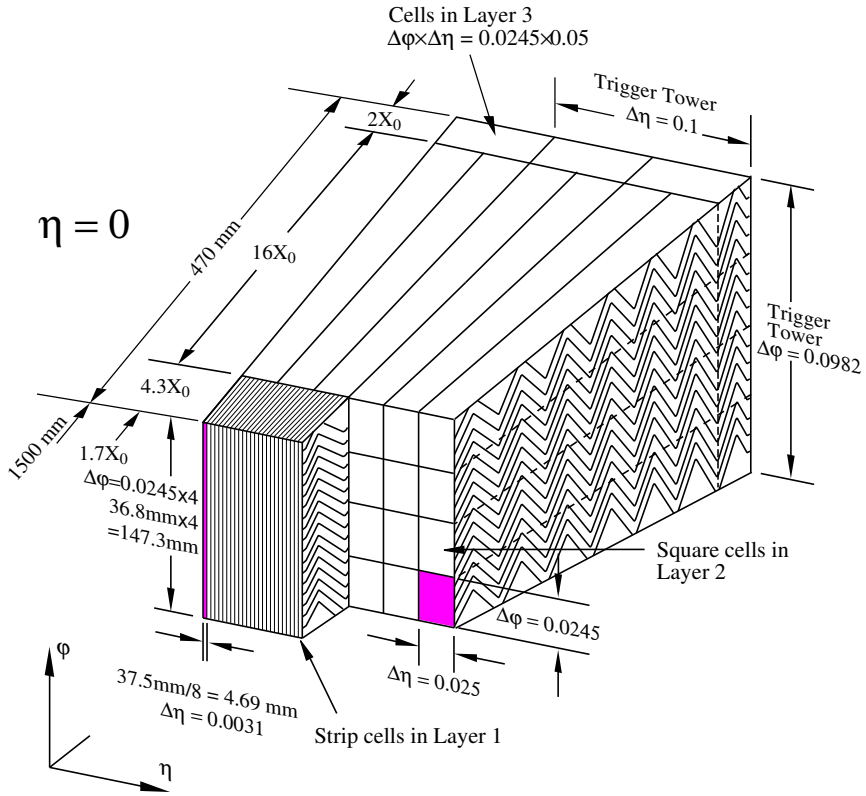



Figure 3.7: Sketch of a barrel module where the different layers are clearly visible with the ganging of electrodes in $\phi$. [31]

only two longitudinal sections.

The signal is formed in the calorimeter by ionization of the active medium. The signal readout is calibrated using test beam data, and after the first year of LHC operation also using the electron energy scale measurement using $Z \rightarrow ee$ events. The signal pulse of the LAr calorimeter consists of a triangular pulse that at the end of the readout chain is shaped in a long signal, with a rising edge, of amplitude proportional to the current measured by the electrodes (i.e. to the ionization produced by the particle), and with a long negative tail, that ensures that additional, low-energy deposits in the calorimeter don't disturb the detection of the first pulse. This bi-polar signal shape is the main characteristics of the ATLAS LAr Calorimeter, and ensures a robust energy measurement even in the high occupancy environment foreseen.

### 3.2.5 Hadron calorimeter

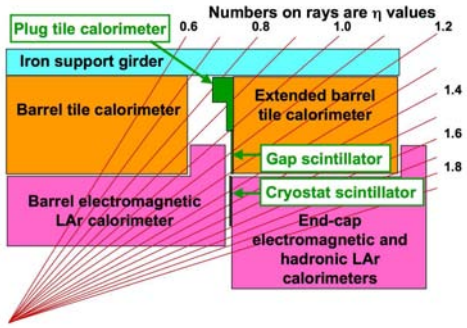
The Hadron calorimeter uses two different technologies: in the barrel region, scintillating tiles are used as active material, while in the encaps the LAr technology has been used to build the end-cap and forward calorimeters. A schematic representation of this arrangement is shown in Fig. 3.8.

The Tile Calorimeter is placed just outside the EM calorimeter envelope. Its barrel covers the region $|\eta| < 1.0$, and two extended barrels cover the range $0.8 < |\eta| < 1.7$; radially, it extends from an inner radius of 2.28 m to an outer radius of 4.25 m. It is a sampling calorimeter, using steel as absorber, and scintillating tiles as active material. It is divided in 64 modules in the azimuthal

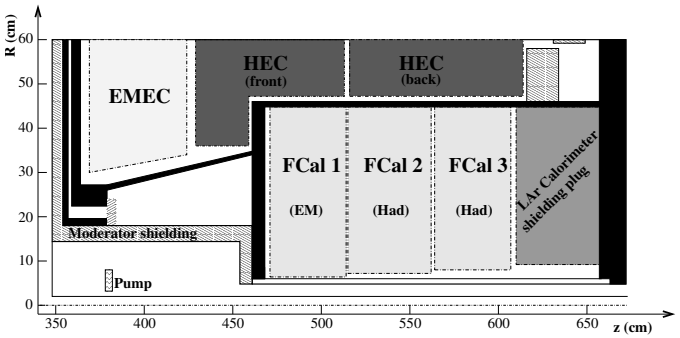
direction, and it is longitudinally segmented in three layers of approximatively 1.5, 4.1 and 1.8 radiation lenghts ($\lambda$). The total Tile Calorimeter thickness is 9.7 $\lambda$ at $\eta = 0$. The total thickness of the steel plates in one period is 14 mm, while the tiles are 3 mm thick. Two sides of the scintillating tiles are read out by wavelenght shifting fibres, into two separated photomultiplier tubes.

The End-cap Calorimeter (HEC) consists of two independent wheels per end-cap, located directly behind the end-cap electromagnetic calorimeter, and sharing the same cryostat. The HEC covers the region $1.5 < |\eta| < 3.2$, overlapping with the Tile extended barrel and the forward calorimeter, to avoid the drop in material density at the transition between the end-caps and the forward calorimeter on one side, and between the end-caps and the Tile extended barrel on the other side. Each wheel is built from 32 identical wedge-shaped modules, and divided into two longitudinal segments. Copper plates are the absorber layers, interleaved by LAr gaps, which provide the active medium for the sampling calorimeter.

The closest region to the beam pipe hosts the Forward Calorimeter (FCal), integrated into the end-cap cryostat. The front face of the FCal is recessed about 1.2 $m$ with respect to the EM calorimeter front face, to reduce neutron albedo in the ID cavity. This space limitation calls for a high-density design of this calorimeter. It is approximatively 10 interaction lenghts deep, and consists of three modules in each end-cap: the first, made from copper, is optimised for electromagnetic measurements; the second and the third, made from tungsten, measure predominantly the energy of hadronic interactions. Their geometry has been optimised to avoid problems due to buildup of the signal.




(a) Schematic of the transition region between the barrel and end-cap cryostats, where additional scintillator elements are installed to provide corrections for energy lost in inactive material (not shown), such as the liquid-argon cryostats and the inner-detector services.




(b) Schematic diagram showing the three FCal modules located in the end-cap cryostat. The material in front of the FCal and the shielding plug behind it are also shown. The black regions are structural parts of the cryostat. The diagram has a larger vertical scale for clarity.

Figure 3.8: Schematic pictures of the ATLAS Calorimeter system. The arrangement of the different detectors is shown, highlighting also the transition and dead material regions. [31]

The amount of material, which particles will be passing through before reaching the calorimeter, will have a large impact on its performance. In addition, the presence of cracks, i.e. non instrumented regions corresponding to detector discontinuity and space needed for the services, can degrade the performance of the measurement significantly. For this reason the ATLAS Calorimeter system is built with detectors partially overlapping in regions where additional gaps or dead material (due to readout or environmental services) is present (see Fig. 3.8). The material is carefully mapped, as an estimate of the energy lost in the material is of fundamental importance for calibra-

tion, and to reproduce the detector performance in the simulation of the detector. The estimate of the material as a function of $\eta$ is shown in Fig. 3.9.

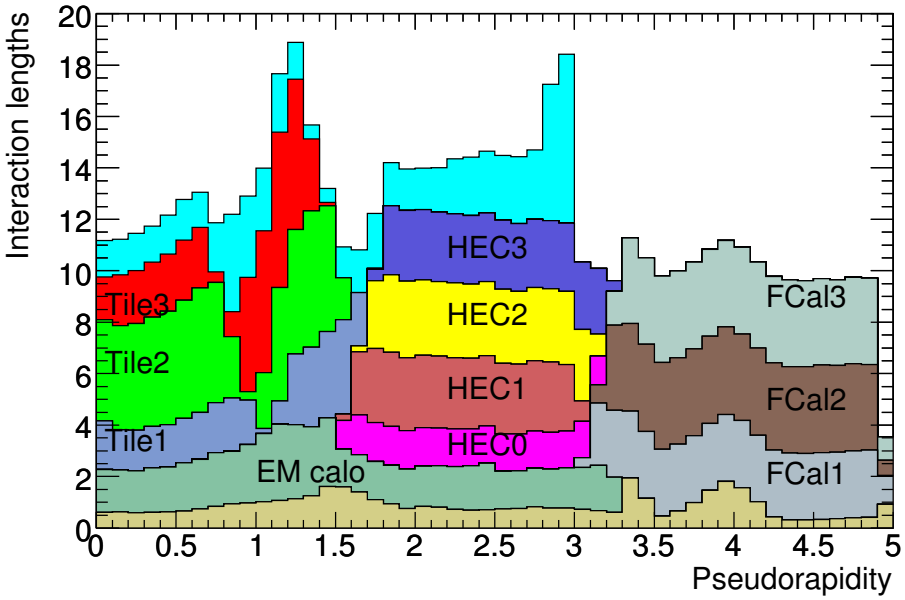



Figure 3.9: Cumulative amount of material, in units of interaction length, as a function of $|\eta|$, in front of the electromagnetic calorimeters, in the electromagnetic calorimeters themselves, in each hadronic compartment, and the total amount at the end of the active calorimetry. Also shown for completeness is the total amount of material in front of the first active layer of the muon spectrometer (up to $|\eta| < 3.0$). From [31].

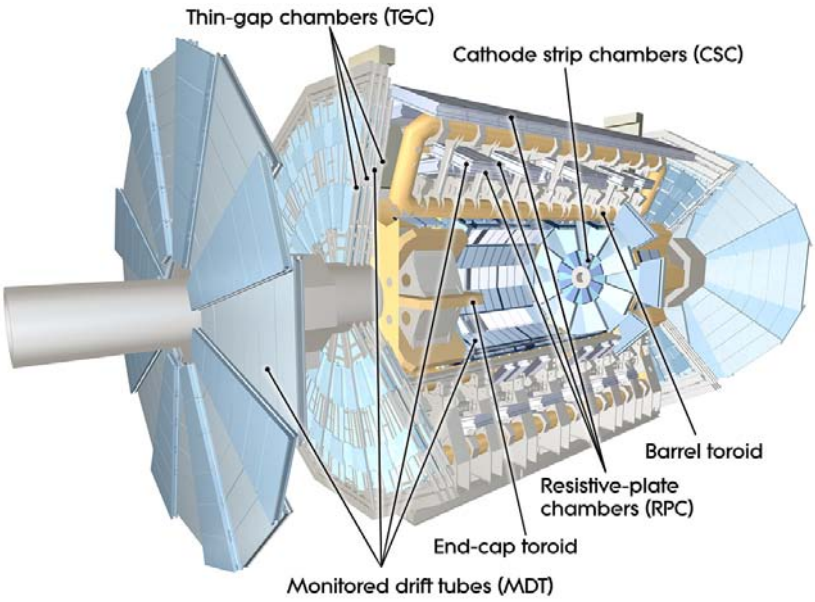



Figure 3.10: Cut-away representation of the Muon Spectrometer. [31]

### 3.2.6 Muon Spectrometer

The Muon Spectrometer aims to measure the magnetic deflections of muon tracks in the large superconducting air-core toroid magnets, instrumented with separated trigger and high-precision chambers. The magnetic field is provided in the range $|\eta| < 1.4$ by the large barrel toroid, while two smaller end-caps, inserted at both ends of barrel toroid, bend the muon tracks in the region $1.6 < |\eta| < 2.7$. In the transition region, the field is a superposition of the barrel and end-cap fields.

The magnet configuration provides a field which is mostly orthogonal to the muon trajectories, while minimising degradation of resolution due to multiple scattering. The barrel and end-cap regions have the best performance, while the transition region has a lower bending power, and also the lower performance.

Several types of muon chambers have been installed in the muon spectrometer, to provide measurement with the requested granularity at different values of $\eta$. Over most of the $\eta$ range, a precision measurement of the track coordinates in the principal bending direction of the magnetic field is provided by *Monitored Drift Tubes*; at larger pseudorapidities, *Cathode Strip Chambers* are used in the innnermost plane over $2 < |\eta| < 2.7$, to withstand the high rates and demanding background conditions. The latter are multiwire proportional chambers with cathodes segmented into stripes, providing a higher granularity in the innermost plane of the spectrometer. The stringent requirements over the relative alignment of the muon chambers layers are met by the combination of precision mechanical-assembly technologies, by an optical system installed within and between muon chambers, and the performance crosschecked with data.

The trigger system covers the pseudorapidity range $|\eta| < 2.4$. Two different types of chambers have been used: Resistive Plate Chambers (RPC) in the barrel, and Thin Gap Chambers (TGC) in the end-cap region. RPC is a gaseous detector providing a space-time resolution of 1 cm $\times$ 1 ns, and a rate capability of 1 kHz/cm$^2$. TGC is a gaseous detector with time resolution slightly worse than for the RPC. The trigger chambers installed in the Muon Spectrometer serve a three-fold purpose: provide bunch-crossing identification, provide well-defined thresholds in transverse momentum, and measure the muon coordinate in the direction orthogonal to that determined by the precision muon tracking chambers.

### 3.2.7 Trigger and Data Acquisition System

The high interaction rate demands an efficient trigger system, in order to handle the huge amount of hits expected in the detector every second.

The trigger is organised in three levels, each using a different subset of the data collected in the whole detector. The initial rate at design condition is 40 MHz; in the end, the trigger system reduces the rate to 200 Hz, with an event size of approximatively 1.3 Mbyte.

The first trigger level (LVL1, *Level 1 Trigger*) searches for high transverse-momentum muons, electrons, photons, jets, and $\tau$-leptons decaying into hadrons, as well as large missing and total transverse energy. Its selection is based on informations from a subset of detectors; it uses data from the trigger chambers in the muon spectrometer, and energy flow measurements from cells in the calorimeters. Tracking information are not used at the first stage of the trigger decision. Results from the L1 muon and calorimeter triggers are processed by the central trigger processor, which builds up a so-called menu made of a combination of different trigger selections. Event data from all the subdetectors wait in the Front-end electronics buffers until the trigger latency for a trigger signal; if they are selected, they are transmitted to the next stage of the detector-specific readout chain, otherwise they are discarded. In each event, the LVL1 also defines the so-called *Regions of*

*Interest* (RoI), i.e. the coordinates in $\eta$ and $\phi$ of those regions of the detector where its selection process has identified interesting features. The RoI data includes also the information on the type of interesting features, and the criteria passed (e.g. a threshold). The RoI data are used by the subsequent trigger stage.

The LVL1 reduces the rate from about 1 GHz to approximatively 75 kHz, with a processing time of less than 2.5$\mu s$.

The LVL2 selection is seeded by the RoI information. LVL2 uses all the available detector data within the RoI (approximatively 2% of the total event data). This second stage reduced the rate to approximatively 3.5 kHz, with an event processing time of about 40 ms.

The final stage is the *event filter* (EF), which reduces the rate to roughly 200 Hz. Its selections are implemented using offline analysis procedures, with an average processing time of the order of four seconds.

### 3.2.8 Simulation of the ATLAS detector

The GEANT4 software toolkit [34] within the simulation framework [32] propagates the generated particles through the detector and simulates their interactions with the detector material. The energy deposited by particles in the active detector material is converted into detector signals with the same format as the detector read-out. The simulated detector signals are in turn reconstructed with the same reconstruction software as used for the data.

In GEANT4 the model for the interaction of hadrons with the detector material can be specified for various particle types and for various energy ranges. For the simulation of hadronic interactions in the detector, the GEANT4 set of processes called `QGSP_BERT` is chosen [3]. In this set of processes, the Quark Gluon String model [35, 36, 46, 47, 62] is used for the fragmentation of the nucleus, and the Bertini cascade model [43, 71, 77] for the description of the interactions of hadrons in the nuclear medium.

The GEANT4 simulation and in particular the hadronic interaction model for pions and protons, has been validated with test-beam measurements for the barrel and endcap calorimeters. Agreement within a few percent is found between simulation and data for pion momenta between 2 GeV and 350 GeV.

Further tests have been carried out *in-situ* comparing the single hadron response, measured using isolated tracks and identified single particles. Agreement within a few percent is found for the inclusive measurement [5, 26] and for identified pions and protons from the decay products of kaon and lambda particles produced in proton-proton collisions at 7 TeV [25]. With this method particle momenta of pions and protons in the range from a few hundred MeV to 6 GeV can be reached. Good agreement between Monte Carlo simulation and data is found.

# Chapter 4

# Jet Algorithms in VBF Higgs searches

The most likely interaction expected in pp collisions at the LHC is parton scattering. Quantum chromo-dynamics (QCD) predicts that no free partons can exist in nature: only colourless particles can be expected in the final state, as explained in Sect. 2.2.2. Thus, the most likely final state to be observed in the detectors will show several hadrons, produced from the original parton interaction via hadronisation.

The hadrons detected in the final state carry the information about the hard scattering, but it is not trivial to extract it. The number and kind of particles that are detected in the final state might vary even if they originate from the same kind of parton interaction. To recover the link between the partons and the particles in the final state, it is useful to define a clustering algorithm (called *jet algorithm*): the output of such a procedure is a "macroscopic object", called a *jet*, that aims to describe the energy and kinematics of the partons in the final state. In theoretical calculations, a jet will be either a single parton (at the parton level), or made of several partons (after fragmentation) or hadrons (after hadronisation); in an experiment, it will be reconstructed from detector signals, or computed back at hadron or parton level using Monte Carlo methods (*unfolding*).

Jets are not real physical objects: they acquire a precise meaning only after an algorithm has been defined. Different definitions would produce different arrangements of the particles in the final states, and all of them are equally justified. As the jets carry fundamental information (the link to the partons), their representation of the interaction should be equivalent; however, the detailed implementation and definition depend also on the physics case, and different algorithms could provide a worse or better representation of different partonic final states.

The definition of a jet has been widely studied in the past[78] and the characteristics that have been agreed upon will be summarized in Sect. 4.1. In Sect. 4.2 the algorithms used in ATLAS will be reviewed. In particular, in the context of VBF Higgs searches the performances of the jet algorithms play an important role, as the possibility to detect and distinguish this production mechanism relies on jet reconstruction. The results obtained from the investigation of different jet algorithms in simulations of VBF Higgs events in ATLAS will be reported in Sect. 4.3.

## 4.1 What is a jet?

A good jet algorithm definition should allow a comparison between the theoretical and the experimental level, as well as between different experimental results. The so-called *Snowmass accord*[78] is the output of a discussion among theorists and experimentalists, aimed at defining a common ground for all high $p_{\mathrm{T}}$ QCD studies. According to it, a jet algorithm should:

- be simple to implement (both in a theoretical calculation and in an experimental measurement);

- be defined at any order in perturbation theory;
- yield finite cross section at any order of perturbation theory;
- yield a cross section that is relatively independent from hadronization.

The first condition requires the definition to be sufficiently general to be implemented, such that jet objects can be defined both at theoretical and experimental level. The second condition ensures that those jets can be compared with each other, because they are well defined regardless of the order considered for the calculation.

The third condition deals with the *collinear* and *infrared* safety. In both cases, the requirement for the algorithm is to be independent of variables that would lead to divergences when considered at a finite perturbative order.

The collinear unsafety arises when the jet algorithm makes a difference between two massless collinear particles and one single massless particle with energy equal to the combined energy of the former pair. In perturbative QCD, the cross section for the pair at a fixed order is divergent, but in the total cross section the divergence is cancelled by the contribution of the virtual correction at the vertex, consisting in a single particle with momentum equal to the combined momentum. If the algorithm does not respect this indifference, and instead represents these two cases differently, it is not collinear safe.

The infrared unsafety occurs when the algorithm is sensitive to soft radiation emission. In perturbative calculation, soft emission causes a divergence at low energy; this divergence is cancelled in the total cross section after renormalisation. If an algorithm is sensitive to these emissions, which could form for example new jets in the final state, the event description is not physical anymore, and the algorithm is infrared unsafe.

The last requirement states that the description of the event given by a jet algorithm should be mostly dependent on the primary partons, and only weakly on fragmentation and hadronization. This ensures that the global characteristics of the event depend largely on the partonic final state.

A jet algorithm which follows the Snowmass accord provides a link between the detector final state and the partonic final state as schematically represented in Fig. 4.1.

The main jet algorithms can be divided into two classes:

- *Cone algorithms*: define jets as the combination of input objects within a cone of radius R around the jet direction. The jet direction is the one that maximizes the energy flow through the cone. Objects belong to the jet if their distance (in $\eta, \phi$) is smaller than the cone radius. The four-momentum of the jet is computed using the energy of all the particles inside the cone, using conventional recombination schemes[1]. The cone radius can be optimized to minimize experimental effects.

  The regions of maximum energy flow can be defined in different ways, and this can be a potential drawback for this class of algorithms. To avoid this problem, the jets are built using an iterative procedure. Once the initial cone direction is defined, the energy-weighted central

[1] The most common recombination schemes are the *"E" scheme*, that adds the four vectors, and the *"$p_T$" scheme*, where each added quantity is weighted by the $p_T$ of the particle. In the $p_T$ scheme, for example:

$$E_{Tij} = E_{Ti} + E_{Tj} \qquad \eta_{ij} = \frac{E_{Ti}\eta_i + E_{Tj}\eta_j}{E_{Tij}} \qquad \phi_{ij} = \frac{E_{Ti}\phi_i + E_{Tj}\phi_j}{E_{Tij}}$$

that is equivalent to the simple vector sum if the opening angle between $i$ and $j$ is small.

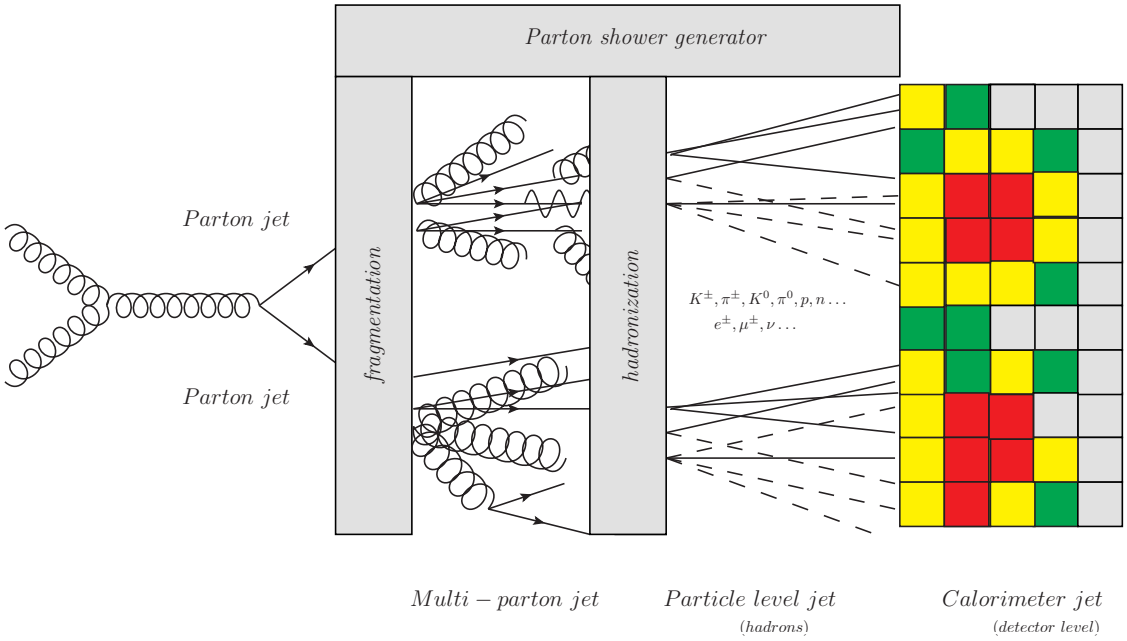



Figure 4.1: Schematic diagram showing the relationship between jet definitions at different levels. The colours represent the energy detected inside the calorimeter cells.

axis of the jet is computed, from the energy and position of all the cells inside the cone. If the jet axis coincides with the initial cone axis, the jet is considered as stable; otherwise, the jet axis is considered as the axis of a new cone. The computation is repeated until the position of the cone axis is stable. Also, a minimum separation distance between two jets is defined, to make the treatment of close or even partially overlapping jets consistent at different levels.

Two implementations are available in ATLAS: a *Seeded Cone Jet Finder*, that uses high $E_{\mathrm{T}}$ objects as a starting point for the iteration, and a *Seedless Cone Jet finder (SISCone)* [102]. The first implementation, because of the seed requirement, is not theoretically safe, as it could be affected by collinear splitting or soft radiation. The last implementation instead is theoretically safe, because it considers all possible configurations of objects in the final state inside cones of radius $R$. Overlapping stable cones are either split or merged, in such a way that the resulting final state picture represents the regions of maximum energy flow as by definition. Removing the seed makes the algorithm safe against infrared and collinear divergences. On the other hand, the price for this safety is high computing time. For this reason, the cone algorithm (in some improved versions, e.g. *Midpoint*, used at Tevatron) has been widely used, despite of its theoretical problems, until the FASTJET[50] implementation made the seedless algorithm competitive with the seeded one in terms of speed. This last implementation is used in the ATLAS software.

Due to the geometrical definition of the jet objects, those algorithms provide very regular shapes in the $\eta$–$\phi$ plane (an example is shown in Fig. 4.2a). This feature is desirable from the experimental point of view, as the design of calorimeters, as well as of the calibrations of physics objects, benefit from this regular definition of the final state.

- *Clustering algorithms*: these algorithms cluster objects in the final state, using a custom defin-

ition of distance: "close" objects make a jet. To define a distance meaningful in the phase space, for every pair of particle $i$, $j$ the distance is defined in $\eta, \phi$ and weighted by the transverse momentum:

$$d_{i,j} = \min(E_{\mathrm{T}_i}^{2p}, E_{\mathrm{T}_j}^{2p}) \frac{\Delta^2}{\mathrm{R}^2} \qquad \text{with} \qquad \begin{array}{c}(\Delta^2 = \Delta\eta^2 + \Delta\phi^2) \\ p = \text{parameter}\end{array} \tag{4.1}$$

The relative power of the energy can be changed setting the $p$ parameter. R is a reference parameter (similar to the Cone radius for the Cone algorithms).

In the same way it is possible to define a reference distance:

$$d_{i,b} = E_{\mathrm{T}_i}^2 \tag{4.2}$$

The jet is built applying the following rules, for each pair (i,j) of particles:

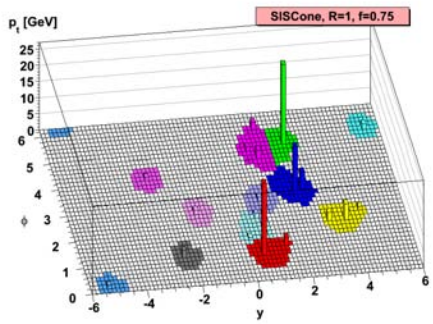
1. if $d_{i,j} < d_{i,b}$ then merge the two particles in a single particle (protojet)
2. if $d_{i,j} > d_{i,b}$ then the $i$ protojet is complete.

Because of this definition, this class of algorithms is completely seedless.

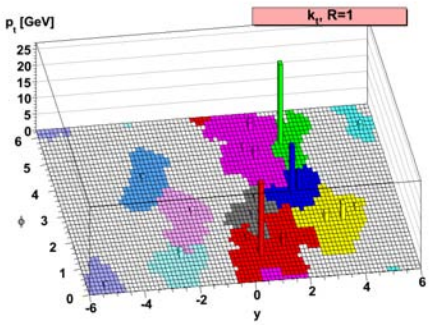
Depending on the definition of distance, the jets can show different characteristics, and, in general, no regular cone-shape with a fixed radius (as in the case of cone algorithms). The most important examples of this class are the $K_T$ algorithm[51] (shown in Fig. 4.2b), that uses a definition of distance proportional to the transverse momentum of a particle with respect to another particle ($p = 1$ in Eq. 4.1), and the anti$-K_T$ algorithm[51] (shown in Fig. 4.2c), which uses the inverse of the transverse momentum in the definition of distance ($p = -1$ in Eq. 4.1). Because of their definition, both algorithms are theoretically safe; also, there is no need for a splitting-merging step, because every object of the final state is associated with only one jet.

As for the SISCone algorithm, the $K_T$ and anti$-K_T$ algorithms were not used in physics analysis until the FASTJET library made their speed performance satisfactory. Before the LHC startup, they had been investigated as good candidates for replacement of the Seeded Cone algorithm within ATLAS and CMS. From the theoretical point of view, $K_T$ is very useful, as it could cluster all the particles coming from the same parton to the same jet, according to its transverse momentum with respect to the jet axis. However, it is quite sensitive to low-energy fluctuations, due to the definition of $d_{i,j}$ in Eq. 4.1; this is the reason for the irregular shapes visible in Fig. 4.2b. From the experimental point of view instead, the most attractive candidate is anti$-K_T$, because it leads to cone-shaped jets, at the same time providing theoretical safety and avoiding splitting-merging procedures.

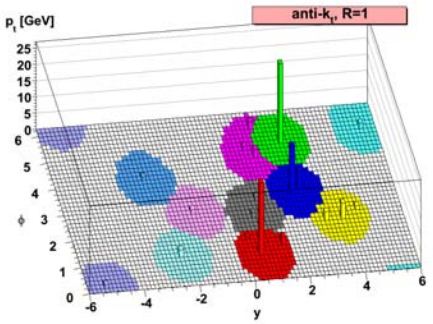
In Fig. 4.2 three algorithms are run on the same parton-level event, overlaid with additional soft radiation. The event is represented in a similar way by the three algorithm, though soft particles are clustered in a different way. These differences may play a large role on physics measurements and their investigation in ATLAS will be reported in Sect. 4.3.




(a) SISCone with radius 1 (f represents the split-merge parameter).




(b) $K_T$ with distance parameter 1.




(c) anti$-K_T$ with distance parameter 1.

Figure 4.2: The *active area* [52] of jets reconstructed with different algorithms. The event is generated at the parton level with HERWIG and overlaid with random "ghosts", soft particles, to mimic experimental effects in the final state on the three algorithms shown. The area shaded in the picture shows the area that on average contains soft particles that get clustered within the hard jets, and bears a close resemblance to the average susceptibility of the jet to a high density of soft radiation (pile-up). From Cacciari, Salam and Soyez [51].

## 4.2 Jet reconstruction in ATLAS

The ATLAS calorimeter was already described in 3.2. In the barrel, the LAr calorimeter is used for electromagnetic measurements and a Tile scintillator for hadronic measurements, while in the *end-cap* and *forward* region of the detector, at higher $\eta$ values, only the LAr technology is used, with different interacting material to allow the containment of the shower in the smaller space available at very low angles with respect to the beam pipe. The difference between the technology used, as well as the not uniform amount of dead material in front of the calorimeter, translates into different detection performance of the calorimeter subsystems.

### 4.2.1 Input to jet reconstruction

Particles produced in a collision event come from the beamspot, located approximately at the geometrical center of the detector, and because of diffraction will show a specific distribution with $\eta$. For this reason, calorimeter cells are approximately centred at the geometrical centre of the ATLAS system, and tend to increase in volume with $\eta$: this structure is called "projective". The cell geo-

metry is projective in the barrel, and only pseudo-projective in the hadronic end-cap (*HEC*), due to the compact size requirement for the detector. For the same reason, the forward calorimeter (*FCal*) is not projective, but the cells lie horizontally, to achieve a better performance in the very small region available to the detector.

The signal detected in the cells is then passed to the next step of the reconstruction. To feed the jet algorithm, three input objects have been defined:

- *Calorimeter towers*: projective regions in the $\eta - \phi$ plane, dimension $0.1 \times 0.1$. In projective calorimeters, a tower is consistent with a cell or a group of cells. For the non projective calorimeters, the grid of towers is overlaid on the cell structure in $\eta - \phi$, and the energy is computed from the cell energy times a weight that represents the fraction of the cell volume within the tower volume. Because of fluctuations in the pedestal of the signal, a tower can have either positive or negative energy, the negative energies representing only electronic noise.

- *Topological clusters* (*topoclusters*): three-dimensional topological calorimeter clusters [88] built from calorimeter cells. Each topocluster is built starting from a seed calorimeter cell with $|E_{\text{cell}}| > 4\sigma$, where $\sigma$ is the RMS of the electronic noise of the cell. Neighbouring cells are iteratively added to the topocluster if they have $|E_{\text{cell}}| > 2\sigma$. Finally, an outer layer of surrounding cells is added. Because of the threshold, a cluster can't show negative energies: the threshold selection suppresses electronic noise.

- *Topological towers (topotowers)*: after the calorimeter signal is reconstructed in topoclusters, the grid of towers is overlaid to the cluster distribution and towers are built from the cells inside the reconstructed topoclusters. The result are noise-suppressed towers. As for the topoclusters, also topotowers are always of positive energy.

The standard inputs in ATLAS are the topoclusters and the topotowers (usually simply called towers). In this study, only topoclusters will be used.

As mentioned earlier, a jet can also be defined at the particle and parton level. Jet at particle level are often compared to reconstructed jets, either to estimate the performance at the reconstruction level, or to extract the properties of an interaction by unfolding the measurement back to the particle level. In ATLAS, the particle level jets are called *Truth jets* and built from simulated particles after parton shower. For these jets, all stable particles with proper lifetimes longer than 10 ps are considered (excluding muons and neutrinos).

## 4.3 Jet Algorithm studies in Vector-Boson-Fusion simulated samples

### 4.3.1 Motivation

In Aad et al. [9], the ATLAS analysis strategies for different physics cases are tested using simulated data sets and the software framework designed for the data taking conditions. In that work, the ATLAS standard algorithm used is the Cone Algorithm. After the publication, several alternatives were proposed in theory papers, and, thanks to the development of the FASTJET code, could be implemented in the ATLAS framework. A global effort started in ATLAS, in order to understand whether a new algorithm could have provided better performance in the analysis than the standard

one. The work presented in this chapter is part of that effort, and the results contributed to the final ATLAS decision. I have used simulated Vector-Boson-Fusion Higgs boson production samples to test the performance of the algorithms, because the possibility to detect this production process strongly relies on the jet algorithm performance.

As mentioned already in Sect. 2.4.1, the VBF process predicts the interaction between two quarks via the exchange of two $W$ bosons, that fuse into a Higgs Boson. The Higgs boson decay products[2] can be detected in the transverse plane, while the two quarks will produce two jets in the final state, that on average will be detected in the forward region of the detector. As shown in Fig. 4.4, very little jet activity is expected in between the two high-$p_{\mathrm{T}}$ jets, because there is no colour flow between the partons. The typical background at LHC, instead, is characterised by high central activity: an example is given in Fig. 4.4a.

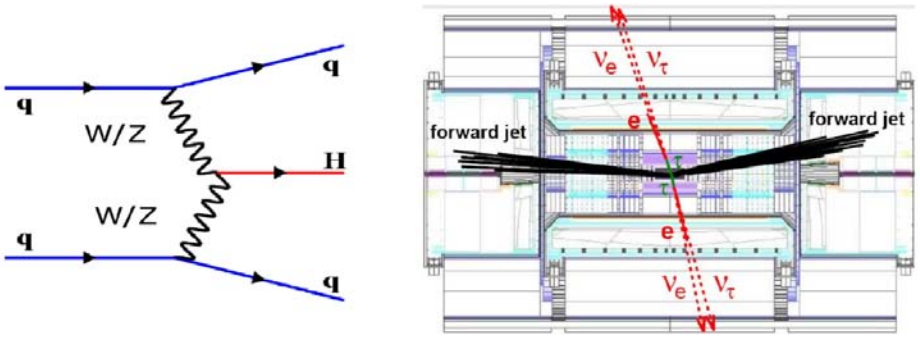
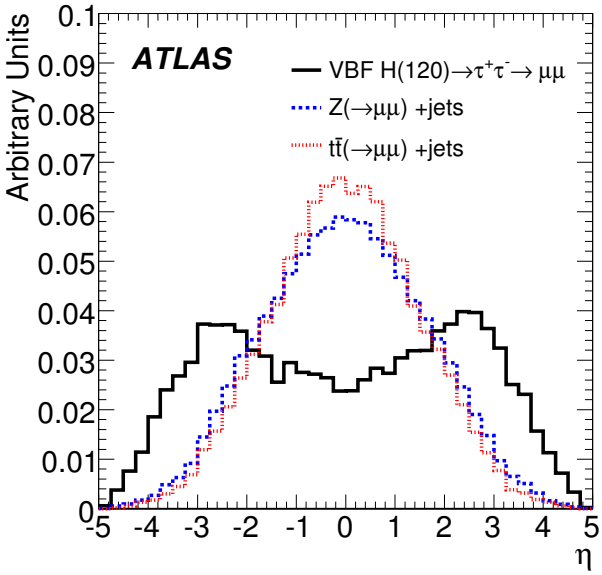



Figure 4.3: On the left, the Feynman diagram corresponding to the VBF process; on the right, a typical sketch of a VBF event in the detector.

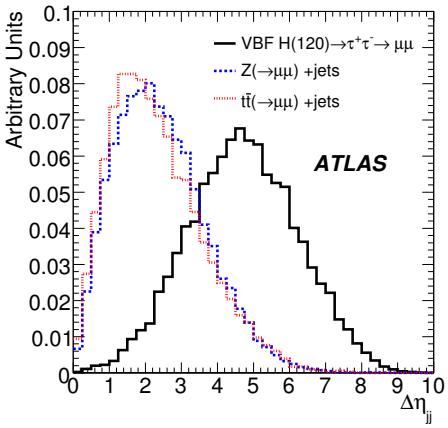
Thus, VBF events can be identified looking at the two highest $p_{\mathrm{T}}$ jets of the event, called *tagging jets*, widely separated in $\eta$, as shown in Fig. 4.4b. Because they will end up most likely in the End-cap and Forward Calorimeters of ATLAS, where no tracker coverage is provided, the analysis will be strongly affected by the calorimeter performance in those regions. Also, another important step in the event selection is the rejection of events with central jet activity using a *central jet veto*: it is usually implemented as a rejection of events with jets between the tagging jets, or with jets in the central region. Because of the different characteristics of signal and backgrounds, such a veto rejects background events very well. In real operation, however, the effect of *pile-up* (see Sect. 3.1.1) must be taken into account. Pile-up effects could lead to an increased jet multiplicity per event, that can affect the veto itself and lower the efficiency of the analysis.

In case of the VBF analysis, the performance of the algorithms needs to be evaluated with a special attention to the tagging jet selection efficiency and the central jet veto efficiency. The experimental conditions considered in this study were those expected at the end of 2008 for the first LHC run. In particular, the pile-up conditions considered in the simulation are mild with respect what has been observed later in data at 7 TeV. The conclusions of this study, anyway, are valid also at higher pile-up conditions and different energies, for reasons that will be highlighted below.

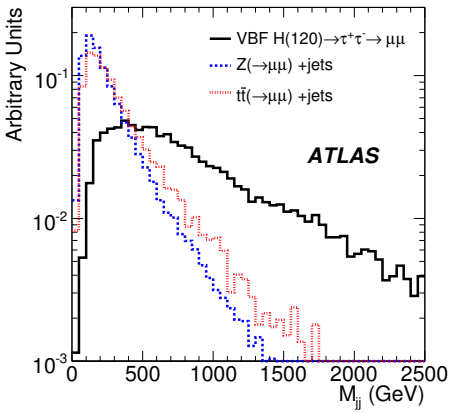
[2] In this work, only the Higgs decaying into two leptonically-decaying taus is considered. The final state can be selected looking for the two leptons from the tau decays.




(a) Pseudorapidity of the highest $p_T$ jets. Only $p_T$ cuts were applied to jets.




(b) $\Delta\eta$ between the two leading jets. A requirement $\eta_1 * \eta_2 < 0$ is used in addition to the cuts on jet $p_T$.




(c) Invariant-mass distributions of tagging jets

Figure 4.4: Tagging jets kinematic distributions in VBF $H \to \tau\tau \to \mu^+\mu^- + 4\nu$ events (hypothesis $M_H = 120$ GeV) and background. (a) shows the pseudo-rapidity of the two tagging jets, highlighting the double peak structure characteristics of VBF processes. (b) shows the difference in pseudo-rapidity between the tagging jets. (c) shows the invariant mass distribution of the di-jet pair formed by the tagging jets. Solid (black) histogram is for signal, dashed (red) histogram is for $t\bar{t} \to WW \to \mu^+\mu^-$, and dotted (blue) histogram is for $Z \to \mu^+\mu^-$ + jets. [9]

### 4.3.2 Analysis details

In the following, four algorithms are compared: the ATLAS Cone, the SISCone (defined in [102]), the $K_T$ and the anti$-K_T$ algorithm (defined in [51]). Jets are reconstructed at the electromagnetic

| Production | Dataset number | Transforms tags | Events | Pileup bunch spacing, $<\mu>$ |
|---|---|---|---|---|
| mc08 | 105333 | e357_s462_r635_t53 | 49670 | no |
| mc08 | 105333 | e357_s462_d150_r642_t53 | 49920 | 450 ns, 4.0 coll. |

Table 4.1: Dataset used in the analysis, in sect. 4.3.3 and 4.3.4.

| Object | 1st cut | 2nd cut | 3rd cut |
|---|---|---|---|
| Electrons | $p_T>$ 10, GeV $\|\eta\| < 2.7$ | ElectronAuthor= 1\|\|3 | ElectronMediumNoIso, ElectronEtcone20/pt $< 0.1$ |
| Muons | $p_T>$ 8 GeV, $\|\eta\| < 2.7$ | Staco muon | MuonEtCone20/pt $< 0.1$ |

Table 4.2: Lepton pre-selection used in the analysis.

scale *(EM-scale)*[3] with radius (for cone algorithms) and distance parameter (for $K_T$ and anti$-K_T$) $R = 0.4$ and full four-momentum recombination. In the jet reconstruction, each calorimeter cluster is considered as a massless particle with energy $E = \sum E_{\text{cell}}$, originating from the geometrical center of the ATLAS detector. In this study, for simplicity, the jet is also reconstructed with respect to the ATLAS geometrical centre, without taking into account the beamspot position.

The jet calibration will be treated in more details in Chap. 5. In this study, the calibration effects did not have an impact on the results and will not be treated in details.

To keep the highest statistics, only a subset of the standard cut selection to be used at 10 TeV[4] has been used, the same implemented in the HiggsValidation package in the ATLAS software framework [76]. To understand the tagging jets and the central-jet-veto performance, the tagging jets selection and the veto have been considered separately, as shown in Table 4.3, and the distributions have been compared before and after each step.

I compare the behaviour of the tagging jets reconstructed with several algorithms looking at the tagging jets distribution ($p_T$, $\eta$ and $\Delta\eta$) after tagging jets selection. To check the performance of the central jet veto, as mentioned in 4.3.1, the behaviour of the non-tagging jets is important. The behaviour of the central-jet-veto can be represented by the following variable:

$$Z^* = \frac{\eta_3 - \frac{(\eta_1+\eta_2)}{2}}{|\eta_1 - \eta_2|} \tag{4.3}$$

$Z^*$ relates the position of the tagging jets $\eta_1, \eta_2$ with the other jets ($\eta_3$) of the same event. All jets that fire the veto fall in a window $[-0.5, 0.5]$, by definition.

The shapes of the distributions are shown in Fig. 4.5 to 4.11. To allow a quick comparison with the non-pile-up case, the plots show the jets distributions for the same events without (left) and

[3] The electromagnetic scale is the basic calorimeter signal scale for the ATLAS calorimeters. It gives the correct response for the energy deposited in electromagnetic showers, while it does not correct for the lower hadron response. The EM scale, initially derived from test beam measurement, has been re-calibrated with ATLAS 7 TeV data using $Z \to ee$ events. In this chapter, the test beam EM scale was used, while later the recalibrated scale will be used. More informations about the jet calibration is in Sect. 5.2.

[4] This study was performed in early 2009, before the decision to run the LHC at 7 TeV. The samples used in this study have been simulated assuming the energy of 10 TeV.

| Cutflow | | | |
|---|---|---|---|
| Preselection | lepton selection | tagging jets selection | central-jet-veto |
| $p_T$> 20 GeV,<br>$\|\eta\| < 4.8$,<br>overlap w/ leptons,dR < 0.2 | 2 lepton ($p_T$> 15 GeV)<br>w/ opposite charge, | Njets $\gtrsim$2,<br>leading jet $p_T$> 40GeV,<br>$\eta_{j1} * \eta_{j2} < 0$,<br>$\Delta\eta_{j12} > 3.6$, | veto if<br>$\|\eta_{otherjets}\| < 3.2$ |

Table 4.3: Cutflow used in the analysis, with special care for tagging jets selection and central-jet-veto, used in sect. 4.3.3 and 4.3.4.

with (right) pile-up.

A more quantitative estimate of the efficiency of tagging jets selection and central-jet-veto is obtained by computing the efficiencies for every step of the cutflow. The efficiency is defined as:

$$\epsilon = \frac{\text{evts. after } n_{th} \text{ cut}}{\text{evts. before } n_{th} \text{ cut}} \tag{4.4}$$

Table 4.4 lists the efficiency for all the algorithms considered.

### 4.3.3 Comparison of algorithms without pile-up

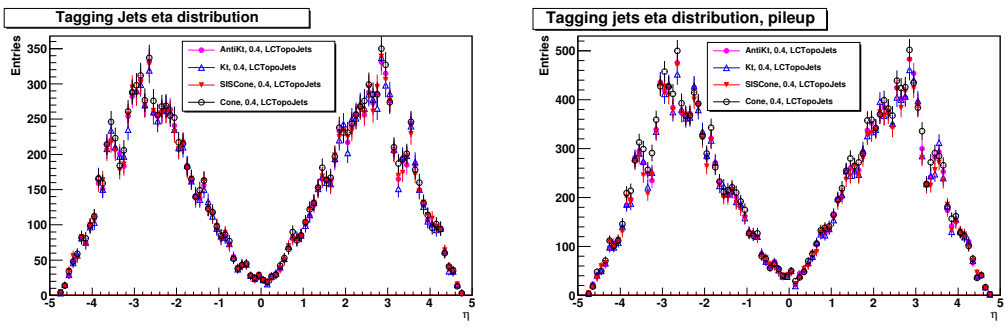
The distributions for $\eta$, $p_T$ and $\Delta\eta$ of the two tagging jets are shown in Figures 4.5a, 4.6a, 4.7a and 4.8a for a non-pile-up scenario. Our results show that the algorithm used for reconstruction does not affect the picture of the final state, and the selection used is robust.

Looking at the non-tagging jets of the event, Fig. 4.9a shows that the $p_T$ distribution of these jets is almost unchanged with the different algorithms.

Looking at the $Z^*$ distributions, no big difference is found between the algorithms before the central jet veto, shown in Fig. 4.10a, and, as a consequence, all of them are affected in a similar way by the veto, as shown in Fig. 4.11a.

Also, the efficiencies of the different algorithms at the several steps of the cutflow are compatible, within the fluctuations, as shown in Table 4.4.

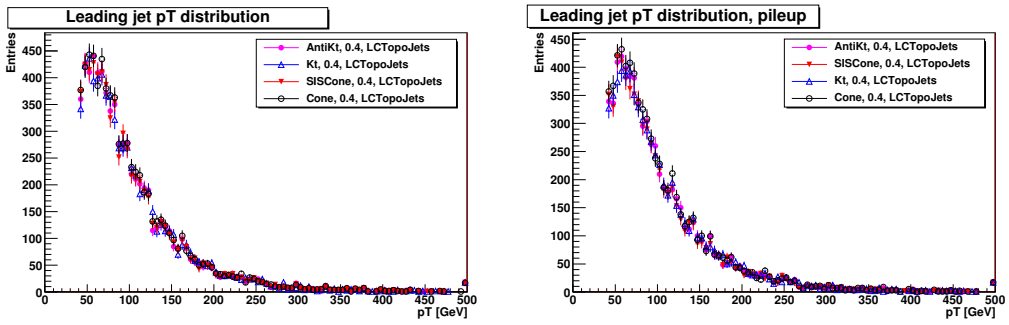
In conclusion, all the algorithms considered have similar performance, in absence of pile-up.




(a) Non-pile-up case

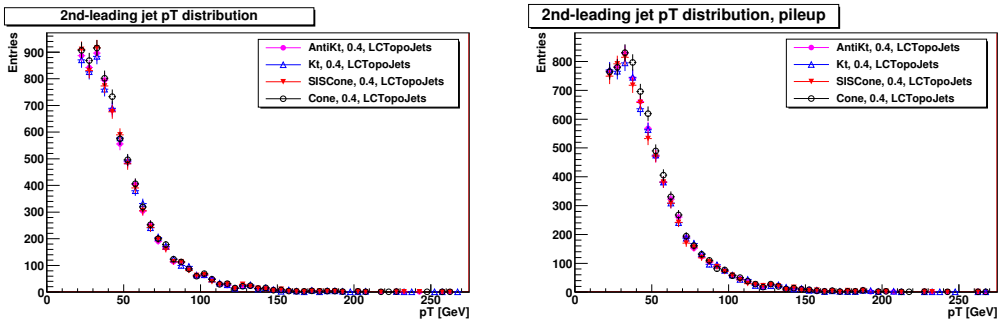
(b) Pileup case.

Figure 4.5: $\eta$ distribution of the two tagging jets, before the veto.



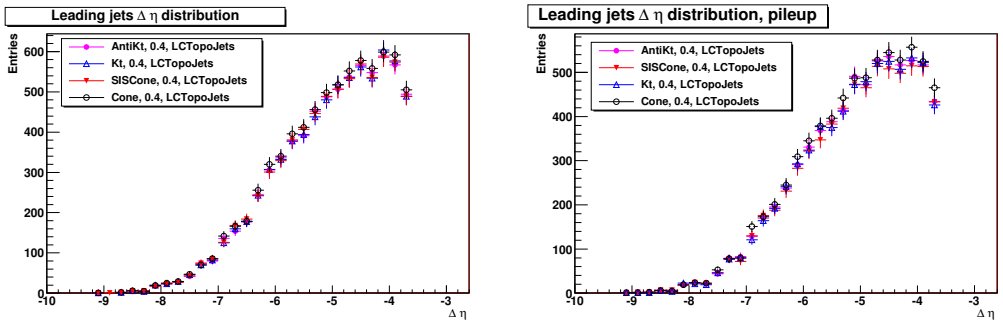

(a) Non-pile-up case (b) Pileup case.

Figure 4.6: Leading jet $p_{\mathrm{T}}$ distribution, before the veto.



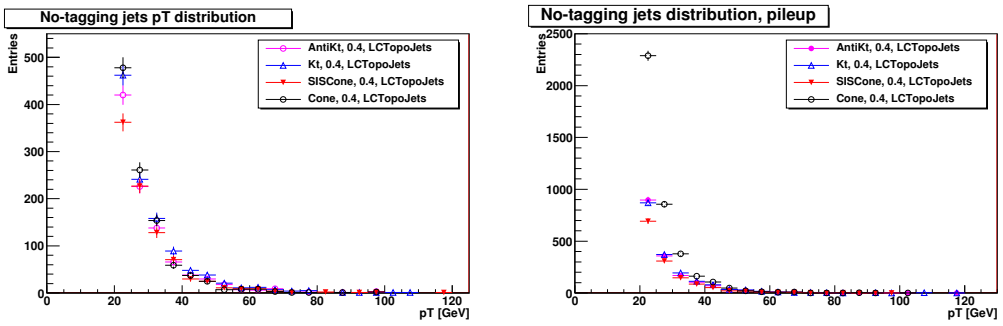

(a) Non-pile-up case. (b) Pileup case.

Figure 4.7: Second-leading jet $p_{\mathrm{T}}$ distribution, before the veto.



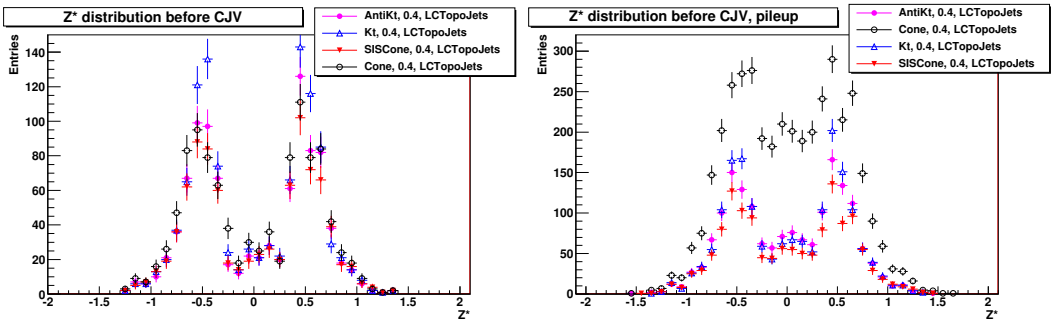

(a) Non-pile-up case. (b) Pileup case.

Figure 4.8: $\Delta\eta$ distribution of the two tagging jets, before the veto.



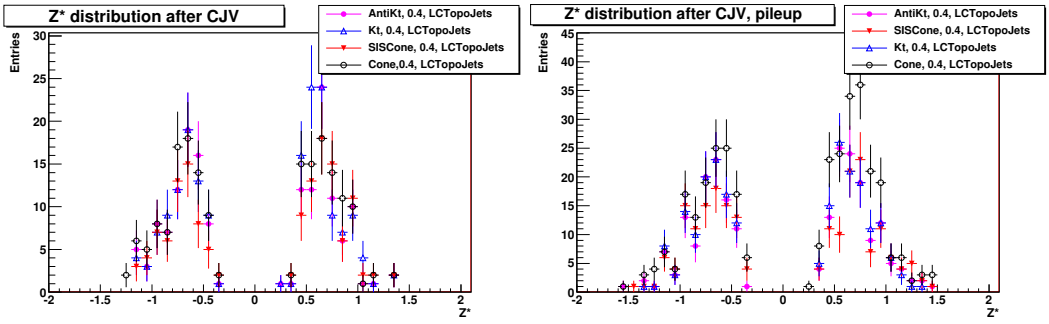

(a) Non-pile-up case. (b) Pileup case.

Figure 4.9: No-tagging jets $p_T$ distribution, before the veto.




(a) Non-pile-up case. (b) Pileup case

Figure 4.10: $Z^*$ distribution of jets before the veto, for all algorithms.




(a) Non-pile-up case. (b) Pileup case

Figure 4.11: $Z^*$ distribution of jets after the veto, for all algorithms.

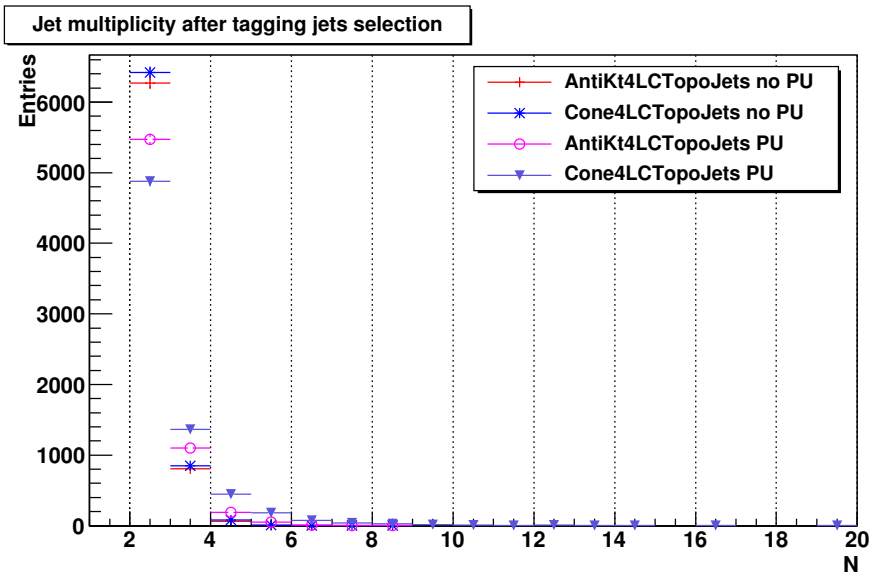



Figure 4.12: Number of jets after the tagging jets selection, for AntiKt and Cone algorithms, with and without pile-up .

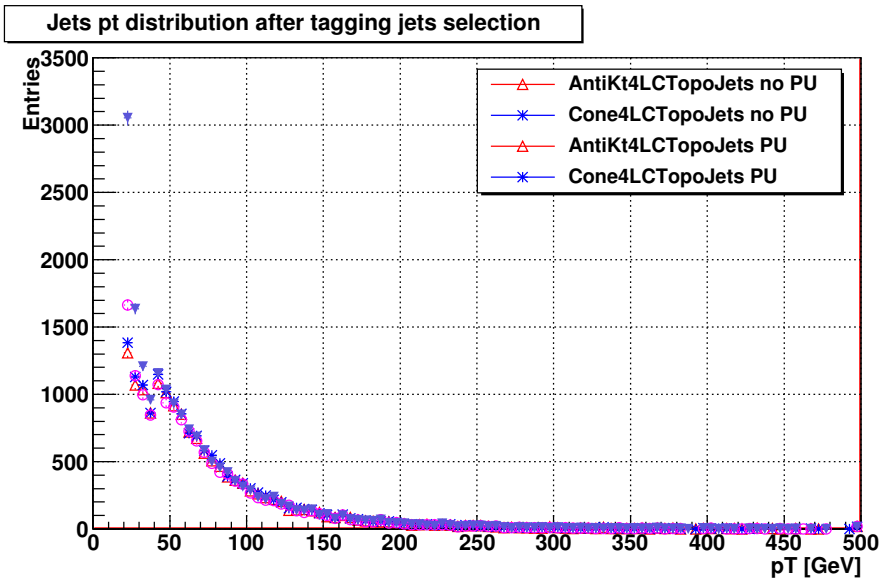



Figure 4.13: Inclusive jet pt distribution after the tagging jets selection, for AntiKt and Cone algorithms, with and without pile-up .

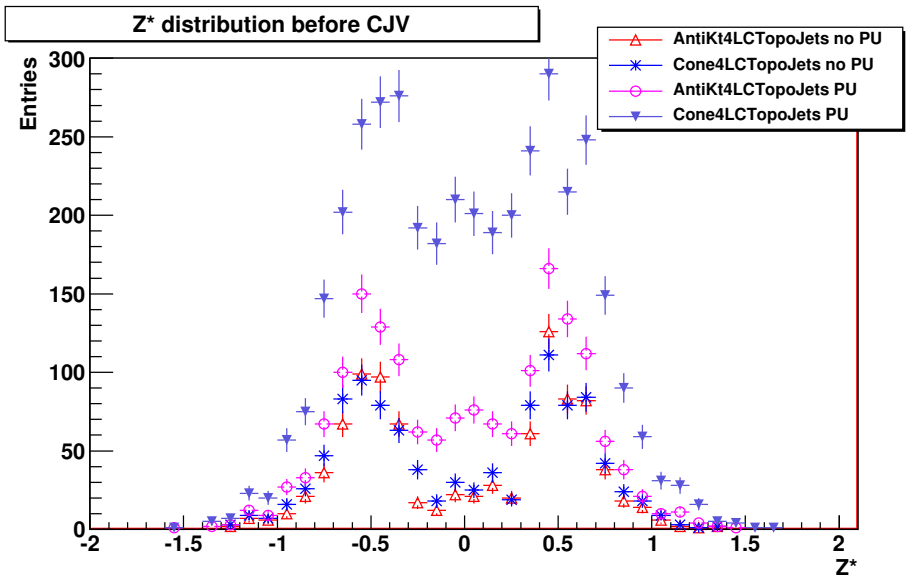



Figure 4.14: $Z^*$ distribution after the tagging jets selection, for AntiKt and Cone algorithms, with and without pile-up .

| Non-pile-up case | | | | |
|---|---|---|---|---|
| Cut | Cone | anti$-K_T$ | $K_T$ | SISCone |
| Preselection | $0.972 \pm 0.006$ | $0.969 \pm 0.006$ | $0.969 \pm 0.006$ | $0.969 \pm 0.006$ |
| lepton selection | $0.379 \pm 0.003$ | $0.379 \pm 0.003$ | $0.379 \pm 0.003$ | $0.379 \pm 0.003$ |
| tagging jets selection | $0.402 \pm 0.006$ | $0.392 \pm 0.005$ | $0.391 \pm 0.005$ | $0.392 \pm 0.005$ |
| central-jet-veto | $0.897 \pm 0.015$ | $0.899 \pm 0.015$ | $0.885 \pm 0.015$ | $0.907 \pm 0.016$ |
| Pileup case | | | | |
| Cut | Cone | anti$-K_T$ | $K_T$ | SISCone |
| Preselection | $0.984 \pm 0.006$ | $0.979 \pm 0.006$ | $0.978 \pm 0.006$ | $0.976 \pm 0.006$ |
| lepton selection | $0.332 \pm 0.003$ | $0.333 \pm 0.003$ | $0.332 \pm 0.003$ | $0.332 \pm 0.003$ |
| tagging jets selection | $0.432 \pm 0.006$ | $0.420 \pm 0.006$ | $0.417 \pm 0.006$ | $0.412 \pm 0.006$ |
| central-jet-veto | $0.734 \pm 0.013$ | $0.834 \pm 0.015$ | $0.824 \pm 0.015$ | $0.861 \pm 0.016$ |

Table 4.4: Efficiencies, as defined in eq. 4.4, for the VBF cutflow in the different scenarios and for different algorithms.

### 4.3.4 Comparison of algorithms with pile-up

The expected behavior in the pile-up case is a higher jet multiplicity per event, as shown in Fig. 4.12. Also other distributions could be affected by the increased activity per event. An example of the differences between the non-pile-up and the pile-up case is shown in fig. 4.13 and 4.14, for two of the considered algorithms.

The tagging jet distributions don't show big differences between the algorithms, as shown in fig 4.6b and 4.7b.

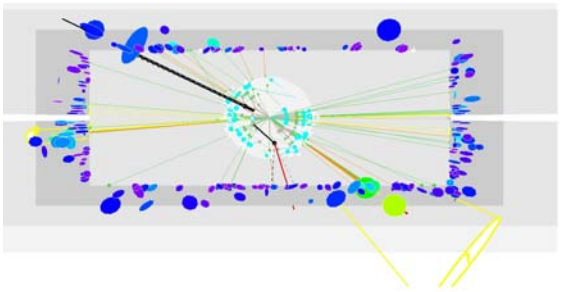
Looking at the non-tagging jets distributions, the differences get bigger. From the inclusive $p_{\mathrm{T}}$ spectrum (Fig. 4.13), the increased jet multiplicity is shown to consist mostly of jets at low $p_{\mathrm{T}}$; the second peak visible in the distribution, caused by the $p_{\mathrm{T}}^{\mathrm{jet}}$ cut on the leading jet at 40 GeV, is almost unaffected by the pile-up. The increase at low $p_{\mathrm{T}}^{\mathrm{jet}}$ is much higher for the Cone algorithm than for the other algorithms, as shown in fig. 4.9b. Also from the $Z^*$ distribution (Fig. 4.10b), is shown that these jets can enhance the probability to fire the veto, because the $Z^*$ distribution around 0 is more populated for the Cone algorithm. All the other algorithms show some differences, but their distributions are compatible within statistical fluctuations, hint that their representation of the final state is robust against pile-up.

The higher probability to fire the jet veto is reflected in the efficiencies reported in Table 4.4: the Cone algorithm shows a large efficiency drop in the pile-up scenario with respect to the non-pile-up case, and it is not compatible with the results from the other algorithms. This behaviour strongly discourages the use of the Cone algorithm in the analysis, in favour of the other infrared-collinear safe algorithms.

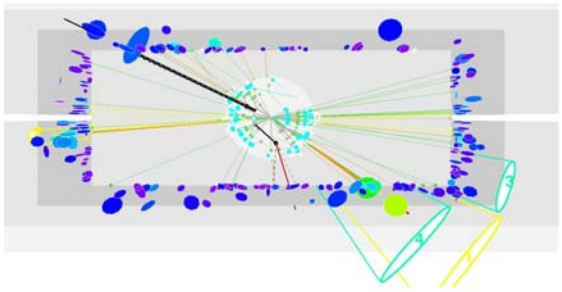
The behaviour observed is interpreted as a combined effect of the splitting-merging procedure and the intrinsic unsafety of the Cone algorithm. It seems likely that in some configurations the Cone algorithm splits one of the tagging jets into two or three jets: one of them still at high $p_{\mathrm{T}}$, with almost no effect on the leading jets distributions, and the others at lower $p_{\mathrm{T}}$, but still sometimes above the 20 GeV cut used in the jet selection. These additional jets will lie very close to the leading jets, and in some cases exactly inside the rapidity gap: in this case, the event will fire the veto. Such a behavior is an example of infrared unsafety. An example of such an event is shown in Fig. 4.15. In the event display with the Cone algorithm (Fig. 4.15b and 4.15d), two additional low-$p_{\mathrm{T}}$ jets are reconstructed, and one of them (labeled "4" in the picture) is very close to one of the tagging jets (labeled "1"). As the event display shows, those additional reconstructed jets can fall within the rapidity gap; thus the higher probability to fire the veto.

## 4.4 Conclusions

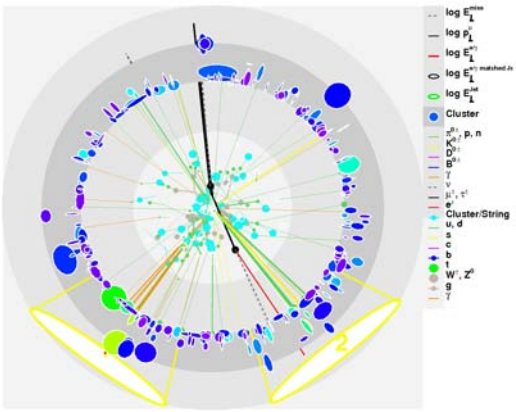
I have shown the results of the validation of the VBF analysis with respect to jet algorithms with simulated ATLAS data at 10 TeV. Some issues have been identified, already considering even only a very mild pile-up scenario. The results on the analysis using the Cone algorithm are not robust against pile-up; in the VBF case, this algorithm is less efficient than other algorithms after a central jet veto is applied, making this rejection technique less effective in enhancing the sensitivity to the signal. The other available infrared-collinear safe algorithms show no such behaviour, both with and without pile-up. Because the pile-up scenario considered is very mild in comparison to what has been already observed at LHC, this study strongly discourages the usage of the Cone algorithm in LHC physics analyses. The anti$-K_T$, $K_T$ and SISCone algorithms, meet the needs of the Higgs analysis designed for the VBF channel in ATLAS.



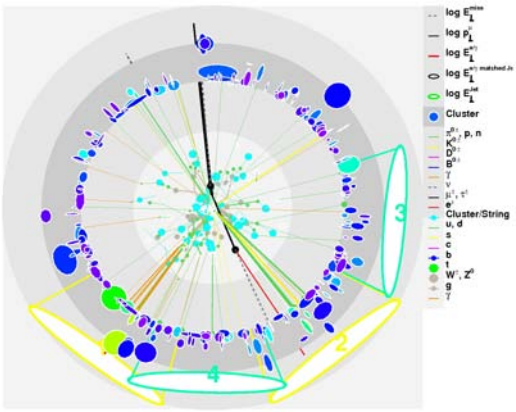
(a) Jets reconstructed with anti$-K_T$ algorithm, $R-Z$ view.



(b) Jets reconstructed with Cone algorithm, $R-Z$ view.




(c) Jets reconstructed with anti$-K_T$ algorithm, $R-\phi$ view.




(d) Jets reconstructed with Cone algorithm, $R-phi$ view.

Figure 4.15: Event display produced with ARA [37], showing the same VBF event reconstructed with the anti$-K_T$ (left, plots (a) and (c) ) and Cone (right, (b) and (d) ) algorithms. The energy deposited in the calorimeter are shown by the circles, with lighter colours identifying higher energies. The reconstructed jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV are shown as cones. The pictures shows the jet labeled as 4 in (b) and (d), not present in (a) and (c), that fires the central jet veto in this event, reconstructed with the Cone algorithm.

# Chapter 5

# Jet $\eta$-intercalibration in 2010 data at $\sqrt{s} = 7$ TeV

## 5.1 Strategy

The first ATLAS data collected at 7 TeV centre of mass energy have been used to establish the performance of the detector. From the point of view of jet reconstruction, the data provided a large sample to establish the performance of the jet algorithm and of the jet energy scale calibration (as well as the trigger performance[23], which will not be discussed). The strategy for 2010 data was designed as follows:

1. compare EM-scale jet properties observed in data with their description by simulation samples, looking at calorimeter variables, for validation purposes;
2. understand the non-collision backgrounds, effects expected to be not reproduced by the simulations used in the early data comparison;
3. validate the jet reconstruction and calibration performance using a preliminary jet energy scale calibration, based on test beam detector studies and simulations; after validation, rederive the jet energy scale calibration for the full 2010 dataset: the baseline calibration derived from simulation, the pile-up effects estimated from data (see Sect. 5.2.2);
4. use *in-situ calibrations* in data to perform a cross-check of the final jet energy scale (see Sect. 5.3).

*In-situ* calibrations are methods that permit to test or derive the energy calibration taking the reference and the object to be calibrated from the same dataset. The $\eta$-intercalibration is one of them: it consists in a tag-and-probe analysis in di-jet events, where one jet is considered as a reference and the other one is calibrated with respect to the reference. This calibration is designed to understand the jet energy scale in the entire calorimeter, and is thus of fundamental importance for measurements with jet final states. In particular, the $\eta$-intercalibration is important to calibrate the forward region, where the measurement from the calorimeter can't be cross-checked with measurements coming from other detectors. The challenges in describing and calibrating the different sections of the calorimeter will be summarized in Sect. 5.2.2.

The results obtained in 2010 are summarised in Aad et al. [10]. I have contributed to the ATLAS publication, focussing on the $\eta$-intercalibration. This has been the first attempt to understand the energy scale in the forward region, and reached already very good results with 35 pb$^{-1}$. The details of the analysis will be treated in Sect. 5.3. Results and open issues will be reviewed in Sect. 5.3.4. These results have been used to estimate the jet energy scale uncertainty in the forward region of the detector, as will be summarized in Sect. 5.3.4.

## 5.2 Jet energy scale calibration

### 5.2.1 Hadronic showers in a calorimeter

The definition of a jet has been explained in 4.1. In experiments, a jet is commonly built from detector measurements. The experimental input objects to the jet algorithm are tracks or calorimeter signals. We will focus on the second case in our analysis.

To detect the energy of a particle, the calorimeters use the principle of the energy loss in matter: a particle, during its passage, looses energy according to its charge, type, speed, and the stopping power of the material[84]. In addition, in case of hadronic particles also the probability of nuclear interactions with the nuclei of the material must be considered. If it is possible to measure the energy absorbed by the material, this corresponds to the energy of the particle at the entrance of the material itself.

The ATLAS calorimeter system consists of *sampling calorimeters*. A sampling calorimeter uses two types of material: a passive one, usually with high atomic number or with a high probability for nuclear interactions with hadronic particles, and an active material, with lower atomic number and able to detect a signal that is proportional to the energy of the particle (ionization in LAr and scintillation in Tile; see Sect. 3.2). The material is arranged in layers along the particle direction, such that each particle will pass through several layers of active and passive material. In each layer, the particle will loose energy as a result of the interaction with the material: at high energies the interaction will result in the production of additional particles with lower energies, from either hadronic interactions with nuclei, or pair production for photons, or bremsstrahlung for electrons; at lower energies, ionization and excitation of the atoms of the material become the dominant processes, and the production of additional particles does not occur anymore. The global process, from the entrance of the initial particle in the calorimeter to the thermalization of all produced particles in the material produces a so-called "shower particle" [113], as it is schematically shown in Fig. 5.1 for several particles. The kinetic energy of the incident particle is transferred to the shower particle, according to the interactions responsible for the process, and to the characteristics of the detector (not only the atomic numbers of the active and passive materials, but also their thickness, for example). In the active layers, the energy of the shower particle is sampled by measuring the energy deposited by the particles inside the shower, and this permits to reconstruct the energy of the initial particle.

Because the detection is based on the energy lost, the energy measurement has to take into account both the kind of interaction between the particle and the absorber, and the particle mass and energy. Some of the particles produced in the hard interaction and after hadronization will have a lifetime long enough to reach the calorimeter: protons, neutrons, pions and K mesons, photons and leptons.

For muons, calorimeters are not an efficient detector, because muons release only little energy in the layers. On the other hand, the presence of muons in a hadronic shower particle (where they could be produced by decays in flight) is negligible with respect to the global composition of the shower.

Electrons and photons loose energy via electromagnetic interaction. Photons mostly produce $e^+e^-$ pairs or interact via compton scattering (for moderate energies of the incident photon), while electrons radiate other photons via bremstrahlung. This creates an *electromagnetic shower* in the calorimeter. An electromagnetic shower is a relatively simple object: it is made only of electrons and photons, and as it depends only on quantum electrodynamics, its characteristics can be computed with a high accuracy. The shower develops quickly and in a small space, producing a very dense

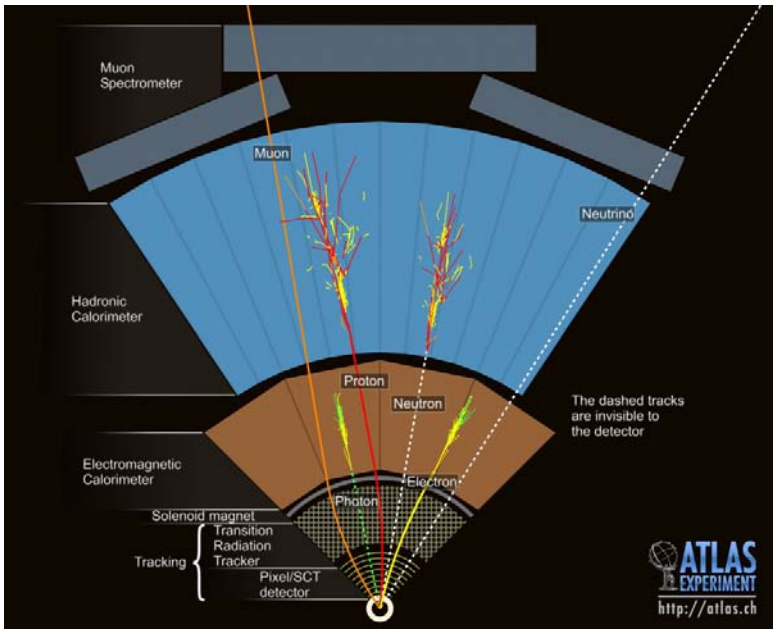



Figure 5.1: Schematic representation of particles detected in the different ATLAS subdetectors according to their nature.

energy deposition in the detector. The whole kinetic energy of the incident particle is transferred to the material, and, in principle, could all been detected. Because in a sampling calorimeter, however, only the active layers are responsible for the signal detection, only the amount of energy released by the shower particle into the active layers will be measured.

Hadronic particles will loose a significant amount of energy only in interactions with nuclei, because of their high mass (their interactions with electrons of the medium will only cause ionization and excitation). Through these nuclear interactions, they will produce a *hadronic shower*, schematically represented in Fig. 5.2. Because the interaction with the nuclei can occur via both strong and electromagnetic interaction, effects as nuclei excitations or break-up will have a significant impact. These effects will not cause any signal and could not be detected. In addition, a large number of pions and kaons is expected to reach the calorimeter or to be produced in an hadronic shower, and will decay in flight. This will produce either electrons or photons, that will originate an electromagnetic shower and thus will be detected, or muons and neutrinos, that won't leave energy in the calorimeter. A hadronic shower shows a slower development and a larger spread in the detector than an electromagnetic shower, and for this reason the energy deposition results less dense, as shown in Fig. 5.1.

In summary, a hadronic shower is composed of[113]:

- visible electromagnetic energy: the electromagnetic part of the shower particle, consisting of electrons and photons produced in the decay of hadronic particles inside the shower. It sums up to $40-60\%$ of the particle shower, with a strong dependence on the energy;
- visible non-electromagnetic energy: from the interaction of hadronic charged particles with the active layers (causing ionization of the atoms of the layer). It sums up to around 25% of the shower particle;
- invisible energy: energy absorbed by the material via nuclear interactions, usually causing

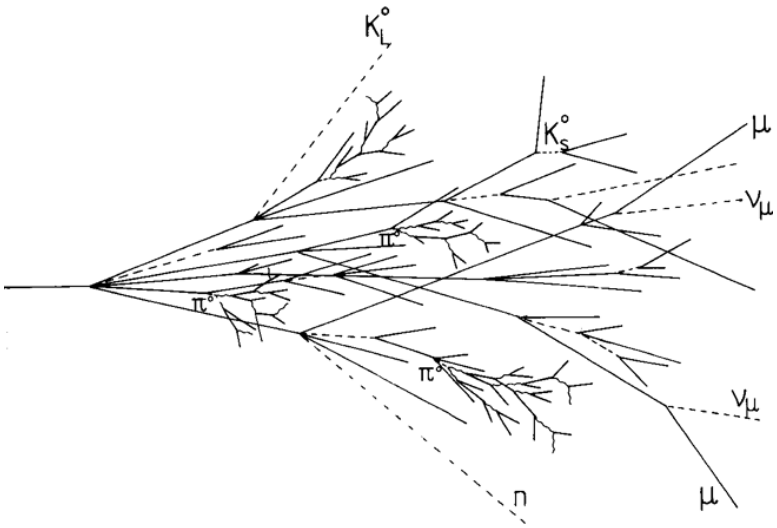



Figure 5.2: Schematic development of an hadronic shower.

nuclear excitations, that could not be detected. It is estimated from simulations to be between 20 and 25% of the shower particle;

- escaped energy, carried away by the muons and neutrinos and not absorbed by the material, of the order of 1% of the energy of the initial particle.

Because of the non-negligible invisible energy, a calorimeter is not able to fully detect the energy left in the material by the incident hadronic particle. An electron and a proton crossing a calorimeter with the same energy will give two different measurements, higher for the electron than for the proton. The reason is the lower response of the material to protons (with response calculated as the ratio of the measured energy to the particle energy[1]). The *jet energy scale calibration* is necessary to correct for this difference, bringing back the ratio of the energies to one.

As a jet does not exist without defining a jet algorithm, also the way it is calibrated will slightly depend on the algorithm. For this reason, the performance of the jet energy calibration are always studied with all the available algorithms, and the correction factors derived will show some small differences depending on the algorithm definition and on its parameters.

### 5.2.2 ATLAS calibration schemes

The shower characteristics, the energy density, as its width and lenght, will depend on several parameters, e.g. the energy and the kind of particle that are in the shower, and these characteristics might be used to derive the energy scale of a detector-level jet. On the other hand, jet algorithm at detector level should in principle collect all the shower particles corresponding to the same parton- or particle-level jet: this correspondence (represented by Fig. 4.1) can also provide a definition for the jet energy scale of detector-level jets. These principles are behind the two main calibration schemes for the jet energy scale (JES) used in ATLAS, explained in the following.

[1] In calorimetry, this is usually represented doing a double-ratio of the response of a material to electromagnetic particles and to hadronic particles. This ratio is called "*e/h ratio*". Estimating the e/h ratio for a detector is very complicated (see Wigmans [113] for more details). For a non-compensating calorimeter as the ATLAS calorimenter, the e/h ratio is always higher than 1, because of the lower response of hadronic particles with respect to electrons.

The response of the ATLAS calorimeters to jets depends on the jet direction, due to the changing calorimeter technology and to the varying amount of dead material in front of the calorimeters.

The jet response is defined as:

$$\mathscr{R}^{\text{jet}} = \frac{E^{\text{jet}}_{\text{EM-scale}}}{E^{\text{jet}}_{\text{truth}}} \tag{5.1}$$

for a jet matched to a truth jet. In the same way a $p_{\text{T}}^{\text{jet}}$ response can be defined. The jet energy calibration relates the jet energy measured with the ATLAS calorimeter to the true energy of the corresponding jet of stable particles entering the ATLAS detector, and for this reason the response is used as a measurement of the mis-calibration of the energy scale.

The estimate of the response at the EM-scale scale in simulated di-jet events is shown in Fig. 5.3. All ATLAS jet calibrations include corrections to remove this energy dependence on $\eta_{\text{jet}}$.

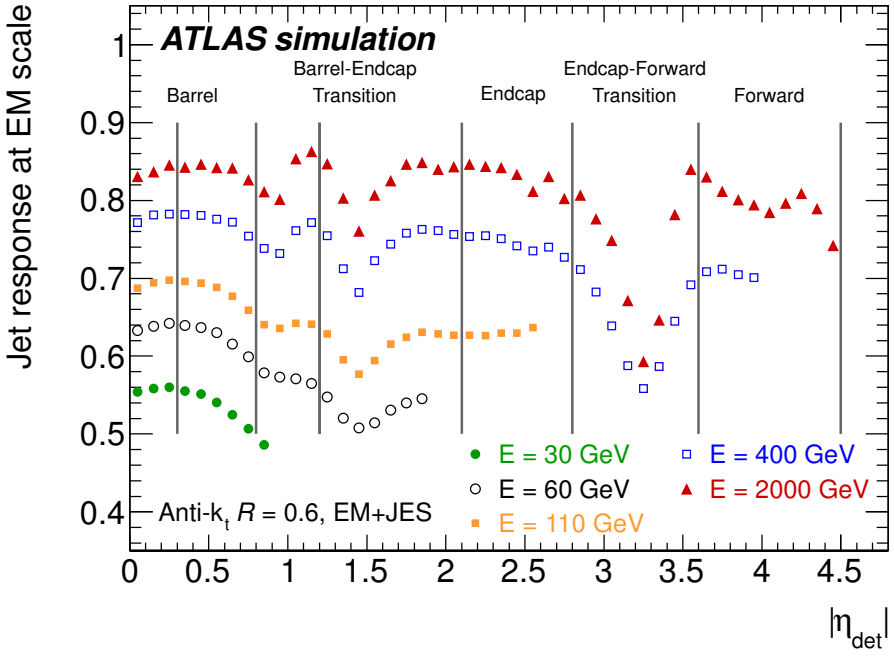



Figure 5.3: Average simulated jet response at the EM-scale in bins of EM+JES calibrated jet energy and as a function of the detector pseudorapidity $\eta_{\text{det}}$. The inverse of the response shown in each bin is equal to the average jet energy scale correction in the EM+JES calibration. From Aad et al. [10].

**EM+JES calibration scheme** The standard calibration scheme in 2010 is called *EM+JES*. It calibrates the jets reconstructed at the EM-scale using energy and $\eta$ dependent correction factors derived from simulated PYTHIA[105] events.[2]

The EM+JES calibration scheme consists of three subsequent steps:

1. **Pile-up correction**: The average additional energy due to additional proton-proton interactions is subtracted from the energy measured in the calorimeters using correction constants obtained from in-situ measurements.

[2]Details of the PYTHIA event simulation are given in Sec. 5.3.2.

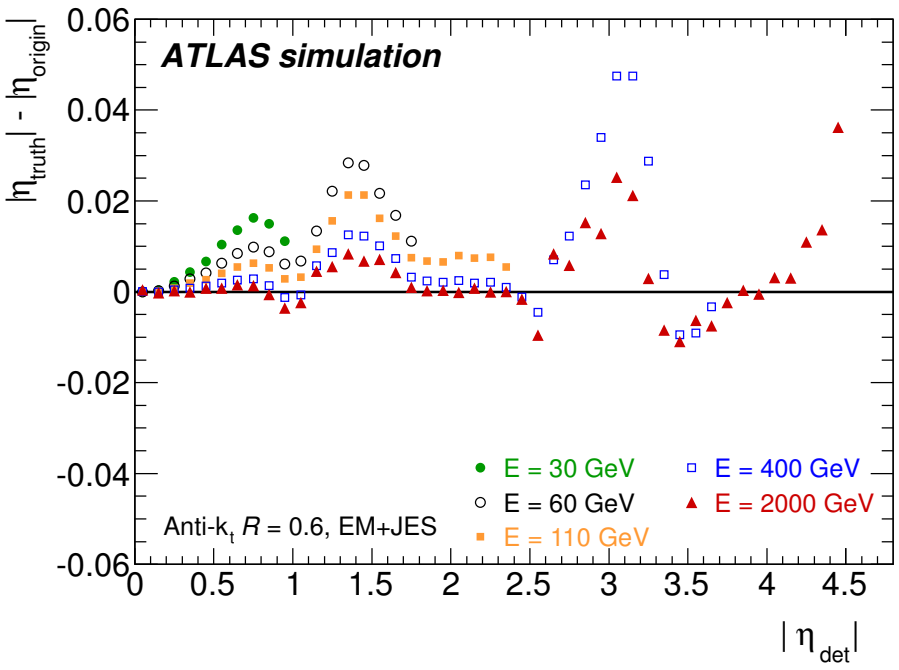



Figure 5.4: Difference between the jet pseudorapidity calculated using an origin correction and the true jet pseudorapidity in bins of the calorimeter jet energy calibrated with the EM+JES scheme as a function of the detector pseudorapidity $|\eta_{\text{det}}|$. From Aad et al. [10].

2. **Vertex correction**: The direction of the jet is corrected such that the jet originates from the primary vertex of the interaction instead of the geometrical centre of the detector.

3. **Jet energy and direction correction**: The jet energy and direction as reconstructed in the calorimeters are corrected using constants derived from the comparison of the kinematic observables of reconstructed jets and those from truth jets in Monte Carlo simulation.

The vertex correction is applied to each jet by correcting the direction of each calorimeter cluster back to the primary hard-scattering vertex. The raw jet four-momentum is thereafter redefined as the vector sum of the clusters four-momenta. This correction improves the angular resolution and results in a small improvement ($< 1\%$) in the jet $p_{\text{T}}$ response. The jet energy is unaffected.

To derive the energy correction factors, truth particle jets are matched[3] with jets reconstructed in the calorimeter. The calibration function is derived in each $\eta$ bin as a function of $E^{\text{jet}}_{\text{truth}}$, and the correction corresponds approximately to the inverse of the response[4].

On top of this, an $\eta$-dependent correction is applied to remove a bias in the reconstructed $\eta$ of jets, that occurs when the jet falls in specific regions of the calorimeter with a much lower response than the regions nearby. The reason for this bias is that the jet direction is reconstructed using the $p_{\text{T}}$-weighted sum of the constituents directions. In these lower response regions, topoclusters are reconstructed with a lower energy with respect to better instrumented regions. This causes the jet direction to be biased towards the better instrumented calorimeter regions. The effect of the

[3]The matching considers only a spatial matching, i.e. considering the $\Delta\text{R} = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ difference between the two jet axis of the truth and of the detector level jet.

[4]This correction is derived for jets with $p_{\text{T}} > 10$ GeV at the EM scale and is parameterized as a function of jet $p_{\text{T}}$ and $|\eta|$. For jets with $p_{\text{T}} < 10$ GeV, the correction factor at each value of $|\eta|$ is set to the value obtained for $p_{\text{T}} = 10$ GeV.

bias on the reconstructed $\eta$ position is shown in Fig. 5.4. The regions affected by this bias are the regions of discontinuity in the calorimeter system (*cracks*).

**LCW+JES calibration scheme** Another calibration scheme investigated in 2010 is the *Local calibration* (LCW). This scheme has been specifically designed for topoclusters. Based on test beam data and Monte Carlo simulation, the LCW aims to correct the energy deposited in a cluster to the "true energy" released by the particle shower. This calibration scheme does not calibrate jets as a global objects, but identifies and calibrates individually the topoclusters inside a jet, separating electromagnetic and hadronic components of a hadronic particle shower starting from the topoclusters characteristics as measured in pion test beams. The response of the LCW calibration scheme estimated from simulation is shown in Fig. 5.5. As the calibration already takes into account the different response and dead material effects, no $\eta$ dependent correction is applied to LCW calibrated jets; however, additional corrections are needed, to compensate for example for topoclusters misidentification. Because it calibrates directly calorimeter clusters, it is already used for the calibration of the calorimeter component of the $E_{\mathrm{T}}^{\mathrm{miss}}$ (for more details, see Sect. 7.4).

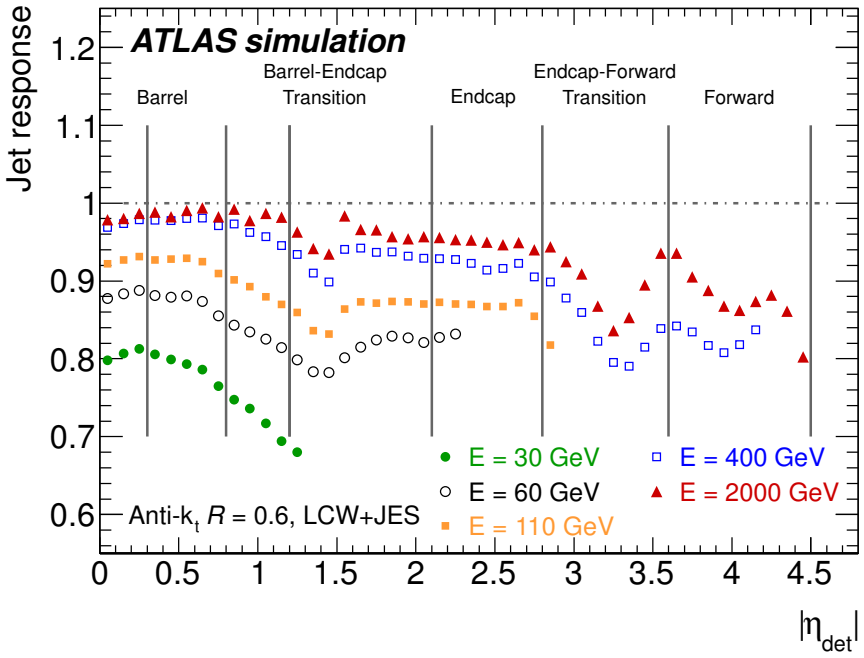



Figure 5.5: Average simulated jet energy response at the LCW scale in bins of the LCW+JES calibrated jet energy and as a function of the detector pseudorapidity $|\eta_{\mathrm{det}}|$. From Aad et al. [10].

**In-situ calibration schemes** Jets can also be calibrated from measurements in data, by comparing their energy (at the EM-scale) to those of other physics objects. An alternative is to use the same *in-situ* methods to test the jet energy calibration, using a well-calibrated object as reference and comparing data to the nominal PYTHIA Monte Carlo simulation.

The following *in-situ* techniques have been used by ATLAS:

1. **Comparison to the momentum carried by tracks associated to a jet:** The mean transverse momentum sum of tracks that are within a cone with size $R$ provides an independent

test of the calorimeter energy scale over the entire measured $p_{\mathrm{T}}^{\mathrm{jet}}$ range within the tracking acceptance. The comparison is done in the jet $\eta_{\mathrm{jet}}$ range $0 \leq |\eta| < 2.1$.

2. **Direct $p_{\mathrm{T}}$ balance between a photon and a jet:** Events with a photon and one jet at high transverse momentum are used to compare the transverse momentum of the jet to that of the photon. To account for effects like soft QCD radiation and energy migrating out of the jet area the data are compared to the Monte Carlo simulation. The comparison is done in the jet $\eta$ range 1.2 and for photon transverse momenta $25 \leq p_{\mathrm{T}}^{\gamma} < 250$ GeV.

3. **Photon $p_{\mathrm{T}}$ balance to hadronic recoil:** The photon transverse momentum is balanced against the full hadronic recoil using the projection of the missing transverse momentum onto the photon direction (MPF). This method does not explicitly involve a jet algorithm. The comparison is done in the same kinematic region as the direct photon balance method.

4. **Balance between a high-$p_{\mathrm{T}}$ jet and low-$p_{\mathrm{T}}$ jet system:** If jets at low transverse momentum are well-calibrated, jets at high transverse momentum can be balanced against a recoil system of low transverse momentum jets. This method can probe the jet energy scale up to the TeV-regime. The $\eta$ range used for the comparison is 2.8.

5. **$\eta$-intercalibration:** The relative jet calorimeter response is studied by comparing the transverse momenta of a well-calibrated central jet and a jet in the forward region in events with only two jets at high transverse momenta (dijets).

All methods are applied to data and Monte Carlo simulation.

These *in-situ* techniques usually rely on assumptions that are only approximately fulfilled. An example is the assumption that the jet to be calibrated and the reference object are balanced in transverse momentum. This balance can be altered by the presence of additional high-$p_{\mathrm{T}}$ particles. For the determination of the JES uncertainties the modelling of physics effects has to be disentangled from detector effects.

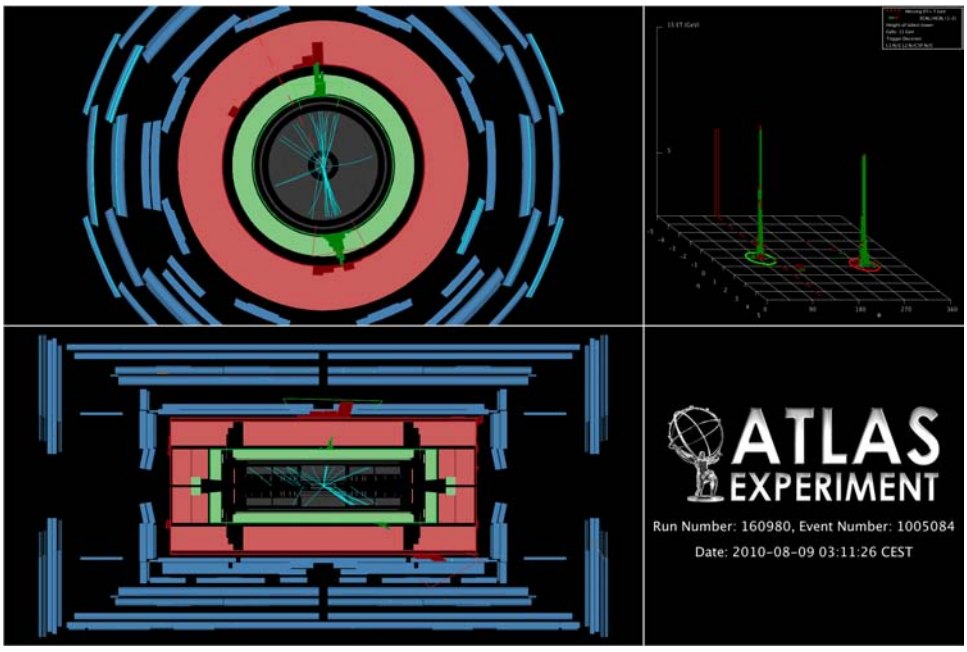
In 2010 the *in-situ* techniques have been used to validate the systematic uncertainty in the jet energy measurement. The $\eta$-intercalibration, because of its characteristics, is instead used to estimate the systematic uncertainty of the forward region with respect to the central region, and for this reason is very important for 2010 analyses using the full calorimeter coverage.

In order to be used as calibration techniques, a larger integrated luminosity is needed. This option has been used in 2011 for the first time. Some comments on this will be given in Sect. 5.4.

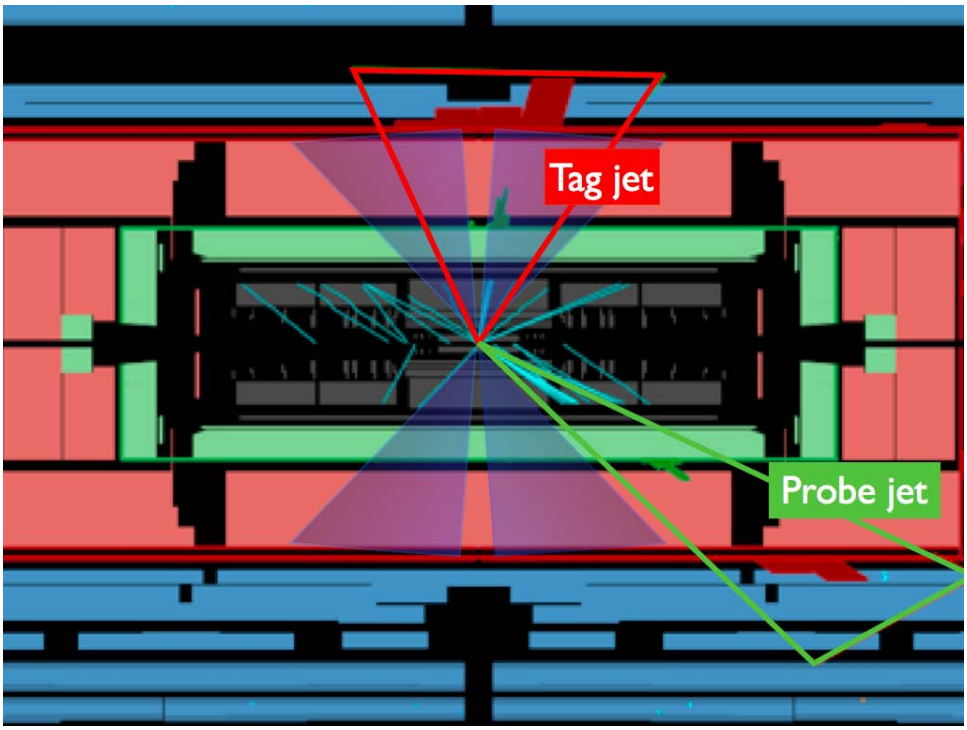
## 5.3 *In-situ* $\eta$-intercalibration of the ATLAS detector using di-jet events

### 5.3.1 Methods used in this analysis

The $\eta$-intercalibration is a "tag-and-probe" analysis using di-jet events. A true di-jet event consists of two jets that are balanced in $p_{\mathrm{T}}^{\mathrm{jet}}$ and 180° apart from each other in the transverse plane, because of transverse momentum conservation. Such an event, reconstructed in a detector, could eventually show a difference in the two $p_{\mathrm{T}}^{\mathrm{jet}}$. The $p_{\mathrm{T}}^{\mathrm{jet}}$ imbalance could occur because of the different energy response and resolution of the different regions of the calorimeter. For this reason, we expect the effect to be dependent on $\eta_{\mathrm{det}}$ (detector pseudo-rapidity) and $p_{\mathrm{T}}^{\mathrm{jet}}$.




(a) Event display showing the $r-\phi$, the $\eta-\phi$, and the $z-r$ view of the ATLAS detector. The di-jet event candidate is clearly visible from the balance in the $r-\phi$ view.




(b) Zoom in the $z-r$ view. The di-jet event is shown overlaid on a view of the calorimeter that shows the pseudorapidity region taken as a reference in the Central reference method for the $\eta$-intercalibration analysis.

Figure 5.6: Event display of a di-jet event in ATLAS 7 TeV data.

**Central reference region method** The traditional approach for $\eta$-intercalibration with dijet events is to use a fixed central region of the calorimeters as the reference region. For calibration purposes in ATLAS, this method has been investigated in Weber [112]. The relative calorimeter response to jets in other calorimeter regions is then quantified by the $p_{\mathrm{T}}$ balance between the reference jet and the probe jet, exploiting the fact that these jets are expected to have equal $p_{\mathrm{T}}$ due to transverse momentum conservation. The $p_{\mathrm{T}}$ balance can be characterised by the asymmetry $\mathscr{A}$, defined as

$$\mathscr{A} = \frac{p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{ref}}}{p_{\mathrm{T}}^{\mathrm{avg}}}, \tag{5.2}$$

with $p_{\mathrm{T}}^{\mathrm{avg}} = (p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}})/2$. The reference region is chosen as the central region of the barrel: $|\eta| < 0.8$, shown in the event display in Fig. 5.6b. If both jets fall into the reference region, each jet is used, in turn, as the reference jet. As a consequence, the average asymmetry in the reference region will be zero by construction.

The asymmetry is then used to measure an $\eta$-intercalibration factor $c$ for the probe jet, or its response relative to the reference jet $1/c$, using the relation

$$\frac{p_{\mathrm{T}}^{\mathrm{probe}}}{p_{\mathrm{T}}^{\mathrm{ref}}} = \frac{2 + \mathscr{A}}{2 - \mathscr{A}} = 1/c. \tag{5.3}$$

The asymmetry distribution is calculated in bins of jet $\eta_{\mathrm{det}}$ and $p_{\mathrm{T}}^{\mathrm{avg}}$: the bins are labeled $i$ for each probe jet $\eta_{\mathrm{det}}$ and $k$ for each $p_{\mathrm{T}}^{\mathrm{avg}}$-bin. Intercalibration factors are calculated for each bin according to Equation (5.3):

$$c_{ik} = \frac{2 - \langle \mathscr{A}_{ik} \rangle}{2 + \langle \mathscr{A}_{ik} \rangle}, \tag{5.4}$$

where the $\langle \mathscr{A}_{ik} \rangle$ is the mean value of the asymmetry distribution in each bin. The uncertainty on $\langle \mathscr{A}_{ik} \rangle$ is taken to be the RMS$/\sqrt{N}$ of each distribution, where $N$ is the number of events per bin.

**Matrix method** A disadvantage of the method outlined above is that all events are required to have a jet in the central reference region. This results in a significant loss of event statistics, especially in the forward region, where the dijet cross section drops steeply as the rapidity interval between the jets increases. In order to use the full event statistics, the default method can be extended by replacing the "probe" and "reference" jets by "left" and "right" jets defined as $\eta^{\mathrm{left}} < \eta^{\mathrm{right}}$. Equations (5.2) and (5.3) then become:

$$\mathscr{A} = \frac{p_{\mathrm{T}}^{\mathrm{left}} - p_{\mathrm{T}}^{\mathrm{right}}}{p_{\mathrm{T}}^{\mathrm{avg}}} \quad \text{and} \quad \mathscr{R}_{\mathrm{lr}} = \frac{p_{\mathrm{T}}^{\mathrm{left}}}{p_{\mathrm{T}}^{\mathrm{right}}} = \frac{c^{\mathrm{right}}}{c^{\mathrm{left}}} = \frac{2 + \mathscr{A}}{2 - \mathscr{A}}, \tag{5.5}$$

where the term $\mathscr{R}_{\mathrm{lr}}$ denotes the ratio of the responses, and $c^{\mathrm{left}}$ and $c^{\mathrm{right}}$ are the $\eta$-intercalibration factors for the left and right jets, respectively.

In this approach there is a response ratio distribution, $\mathscr{R}_{ijk}$, whose average value $\langle \mathscr{R}_{ijk} \rangle$ is evaluated for each $\eta^{\mathrm{left}}$-bin $i$, $\eta^{\mathrm{right}}$-bin $j$ and $p_{\mathrm{T}}^{\mathrm{avg}}$-bin $k$. The relative correction factor $c_{ik}$ for a given jet $\eta$-bin $i$ and for a fixed $p_{\mathrm{T}}^{\mathrm{avg}}$-bin $k$, is obtained by minimising a matrix of linear equations:

$$S(c_{1k},...,c_{Nk}) = \sum_{j=1}^{N}\sum_{i=1}^{j-1}\left(\frac{1}{\Delta\left\langle\mathscr{R}_{ijk}\right\rangle}\left(c_{ik}\left\langle\mathscr{R}_{ijk}\right\rangle - c_{jk}\right)\right)^2 + X(c_{ik}), \tag{5.6}$$

where $N$ denotes the number of $\eta$-bins, $\Delta\left\langle\mathscr{R}_{ijk}\right\rangle$ is the statistical uncertainty of $\left\langle\mathscr{R}_{ijk}\right\rangle$ and the function $X(c_{ik})$ is used to avoid the trivial solution[5]. Note that if the jet response does not vary with $\eta_{\text{jet}}$, then the relative response will be unity for each $(\eta_{\text{left}}, \eta_{\text{right}})$-bin combination (see Equation 5.5). A perfect minimization $S = 0$ is achieved when all correction factors equal unity.

The minimisation of Equation 5.6 is done separately for each $p_{\text{T}}^{\text{avg}}$-bin $k$, and the resulting calibration factors $c_{ik}$ (for each jet $\eta$-bin $i$) are scaled such that the average calibration factor in the reference region $|\eta| < 0.8$ equals unity.

The advantages and disadvantages of the two methods are summarized in Table 5.1.

| **Method** | Central reference method | Matrix method |
|---|---|---|
| **type** | calibration wrt "reference region" (usually: inside barrel) | calibration factors obtained via minimization of matrix of lin. eqs., rescale barrel to 1 for comparison |
| **tag jet** | in reference region (=tag jet) | from any detector region |
| **probe jet** | from any detector region | from any detector region |
| **pro** | easy computing and uncertainties | high statistics |
| **contra** | low statistics | CPU time (for minimization procedure) uncertainties |

Table 5.1: Comparison between $\eta$-intercalibration methods used in 2010.

In early data, both methods have been used, to have an independent cross-check of the results. No disagreement between them was found.

### 5.3.2 Data and Monte Carlo samples

The data analysed were collected in 2010, and the total amount of data recorded is $38 \pm 4$ pb$^{-1}$. The exact amount of data in the different bins is different and depends on the evolution, in the different data periods, of the triggers used for the selection of events in each bin.

The standard MC samples used for the comparison with data have been generated with PYTHIA 6.4.21 [105]. PYTHIA implements a leading-order matrix element generation for QCD $2 \to 2$ processes, followed by fragmentation, calculated in leading-logarithm approximation, and hadronisation, performed using the Lund String model. The underlying event is multiple-parton interaction interleaved with the initial state parton shower.

[5] $X(c_{ik}) = K\left(N_{\text{bins}}^{-1}\sum_{i=1}^{N_{\text{bins}}} c_{ik} - 1\right)^2$ is defined with $K$ being a constant and $N_{\text{bins}}$ being the number of $\eta$-bins (number of indices $i$). This term prevents the minimisation from choosing the trivial solution: all $c_{ik}$ equal to zero. The value of the constant $K$ does not impact the solution as long as it is sufficiently large ($K \approx 10^6$).

Other samples have been produced to permit a more detailed study of the uncertainties. To study the uncertainty within the PYTHIA simulation, two tunes have been used: the standard ATLAS MC10 tune[6], and the PERUGIA2010 tune[106]. By comparing the standard tune to the PYTHIA PERUGIA2010 tune, the effects of soft physics modelling are tested. The PERUGIA2010 tune provides, in particular, a better description of the internal jet structure[30].

Other generators have been considered to assess the systematics related to the choice of the generator. The HERWIG[39] generator has similar leading-order matrix computation as PYTHIA, but uses an angle-ordered fragmentation and a cluster hadronization model. The ALPGEN[90] generator provides instead a leading-order calculation for up to six partons in the final state, and is then matched to HERWIG for parton shower and JIMMY[49] for the underlying event. The ALPGEN Monte Carlo uses different theoretical models for all steps of the event generation and therefore gives a reasonable estimate of the systematic variations, in particular in the comparison with PYTHIA.

All generators are passed through the standard GEANT4 ATLAS simulation [32], and then fully reconstructed and analysed with the same software used for the last data processing in 2010.

### 5.3.3 Di-jet events reconstruction and selection

Jets are reconstructed from calorimeter signals at the electromagnetic scale *(EM-scale)* using the anti$-K_T$ algorithm[6] with distance parameter $R = 0.6$[7] and full four-momentum recombination. Each calorimeter cluster is considered as a massless particle with energy $E = \sum E_{\text{cell}}$, originating from the geometrical center of the ATLAS detector, and set as standard input to the jet algorithm[8]. The jet is then corrected using one of the calibration schemes explained in Sect. 5.2.2.

Pile-up effects will affect the measurement of the energy in the calorimeter, as already mentioned in Chap. 4. The effects on the EM-scale energy as a function of the number of primary vertices in the event have been evaluated and separately corrected for in physics analyses. In this way, the jet energy scale calibration derived for the non-pile-up case was still applicable to the data, regardless of their pile-up conditions. For the same reason, in the *in-situ* studies only events with a single vertex (no-pile-up events) have been considered.

Events are retained if there were at least two jets above the jet reconstruction threshold of $p_{\text{T}}^{\text{jet}} > 7$ GeV. The event is rejected if either of the two leading jets did not satisfy the standard jet selection criteria, consisting in quality requirements aimed to reject the non collision background (for more details, see Aad et al. [10]).

Events are required to satisfy a specific logic using one central and one forward jet trigger, which select events based on jet activity in the central ($|\eta| < 3.2$) and forward ($|\eta| > 3.2$) trigger regions, respectively [23]. The requirements are chosen such that the trigger efficiency, for a specific region of $p_{\text{T}}^{\text{avg}}$, was greater than 99% and approximately flat as a function of the pseudorapidity of the probe jet.

To cover the region $p_{\text{T}}^{\text{avg}} < 45$ GeV, events triggered by the minimum bias trigger scintillators[9]

[6] Details on this algorithm and the reasons for this choice are reported in Chap. 4.

[7] The value of the distance parameter chosen is the value recommended for QCD studies dominated by gluon jets, as estimated from simulations. A smaller value of 0.4 is instead preferred in physics measurements dominated by quark jets. Both values have been used for the estimation of the performance, but, for simplicity, we will report only results on $R = 0.6$ jets.

[8] Also tower jets have been investigated, but, as they won't be used in the analyses reported in this thesis, we are neglecting them here.

[9] The minimum bias triggers consists of a system of trigger scintillators placed at the two edges of the inner detector, covering the region $2.09 < |\eta| < 3.84$. Any event that fires at least one channel is detected in this stream. These triggers were prescaled very early in the run, and were used only for the data collected in the first three monts of

were used.

To enhance events which have only two jets at high $p_{\mathrm{T}}$, the following selection criteria are applied;

$$p_{\mathrm{T}}^{\mathrm{avg}} > 20 \text{ GeV}, \qquad \Delta\phi(\mathrm{j}_1,\mathrm{j}_2) > 2.6 \text{ rad}, \tag{5.7}$$

$$p_{\mathrm{T}}(\mathrm{j}_3) < \max(0.15\, p_{\mathrm{T}}^{\mathrm{avg}}, 7 \text{ GeV}), \tag{5.8}$$

where $\mathrm{j}_i$ denotes the $i^{\mathrm{th}}$ highest $p_{\mathrm{T}}$ jet in the event and $\Delta\phi(\mathrm{j}_1,\mathrm{j}_2)$ is the azimuthal angle between the two leading jets.

The lowest $p_{\mathrm{T}}^{\mathrm{avg}}$-bins are likely to suffer from biases. At very low $p_{\mathrm{T}}^{\mathrm{avg}}$, it is expected that this technique may not measure accurately the relative response to jets, because the assumption of dijet balance at hadron level may start to fail. First, there are residual low-$p_{\mathrm{T}}$ jet effects since the selection criterion on the third jet, which is used to suppress the unbalancing effects of soft QCD radiation, is not as efficient due to the jet reconstruction threshold of 7 GeV. Second, the jet reconstruction efficiency is worse for low-$p_{\mathrm{T}}$ jets.

### 5.3.4 Results

In Fig. 5.7, a comparison between the data and the PYTHIA simulation is shown. The comparison between the particle-level jets (labeled as "Dijet Truth") and the reconstructed jets (labeled as "Dijet MC") in the PYTHIA samples shows already the effects expected in the analysis. In general, the relative response is expected to be close to 1 for $|\eta| < 2$, with fluctuations mostly located in the points where the calorimeter structure, and so the response, changes (see Fig. 5.3). At large pseudo-rapidities ($|\eta| > 3$, roughly corresponding with the FCal system) both the reconstructed and the truth jets show a relative response smaller than one: this means that not only the calorimeter response, but also a physics effect is playing a role, as will be discussed later in the text.

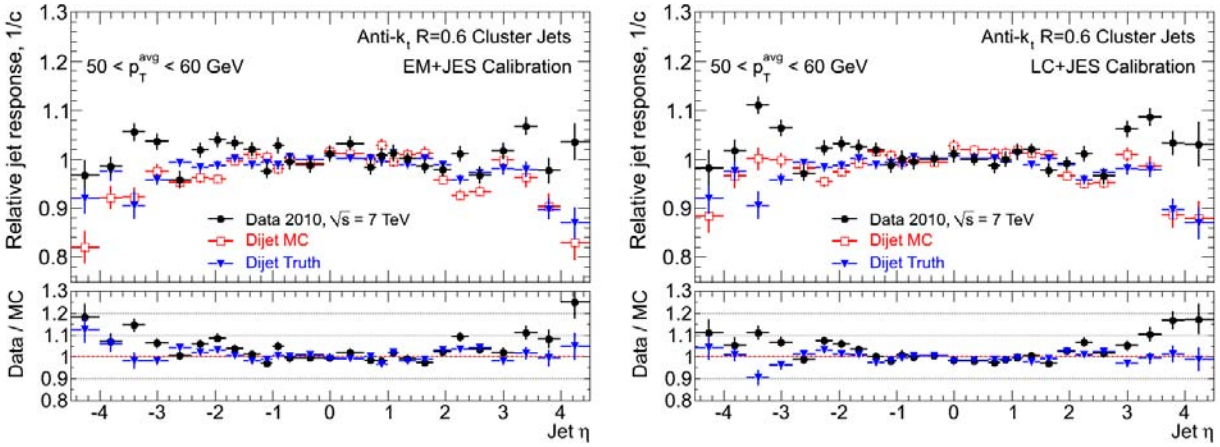
When compared with the data, the most striking feature is that the data does not follow the simulation behavior at high $|\eta|$. The comparison between Fig. 5.7a and 5.7b shows that the calibration is not responsible for this discrepancy. The results shown are produced with the central reference region method and cross-checked with the matrix method; the two methods agree within the statistical uncertainties.

To interpret the observations, the comparison has been performed[10] with the set of samples mentioned in 5.3.2. In Fig. 5.8 the behavior of the four systematic samples is overlaid on the data, for four different $p_{\mathrm{T}}^{\mathrm{avg}}$ bins. The behavior is different mostly at low $p_{\mathrm{T}}^{\mathrm{avg}}$: the PYTHIA samples (both the standard and the PERUGIA2010 tune) tend to have a lower response than the one seen in data, while the HERWIG sample (and the ALPGEN+HERWIG as well) tend to have a higher response. The difference can be of the order of 10% at 20 GeV and large $\eta$. The effect has been tested against calibration effects, uncertainties on the soft radiation[11] and underlying event simulation, and in all cases the results were consistent with the picture given. This result has been interpreted as an effect of the different parton shower. This could explain the large difference between HERWIG and

operations (2010 run period A, B and C).

[10] In this comparison, due to the lower statistics of the systematic samples, only the matrix method has been used. Also, the binning in $\eta$ is slightly coarser, in order to provide a result more robust against statistical fluctuations.

[11] This uncertainty is related to the selection cuts listed in Eqs. 5.8. To assess the uncertainty, the selection criteria are varied for both data and Monte Carlo. The resulting uncertainty is negligible with respect to the modelling uncertainty.




(a) EM+JES calibration

(b) LCW+JES calibration

Figure 5.7: Relative response from $\eta$-intercalibration in data, di-jet PYTHIA reconstructed events and di-jet PYTHIA particle-level jets (*Truth*) for a given $p_{\mathrm{T}}^{\mathrm{avg}}$ bin and the two calibration schemes: EM+JES in (a) and LCW+JES in (b) (calibration schemes defined in Sect. 5.2.2). Only events with a single vertex are included (no pileup correction applied). Central reference region method results, shown at the ATLAS Hadronic calibration workshop.

PYTHIA, but also between HERWIG and the data: other studies have shown that HERWIG does not reproduce well some variables, as the jet shapes, predicting a higher value with respect to what is seen in data [30]. In PYTHIA, similar effects are probably among the reasons for the smaller response at high $|\eta|$. The study will continue with larger statistics and also with the update of the ATLAS standard Di-jet generator to PYTHIA 8, that will be fully implemented in 2012 in the ATLAS simulation infrastructure.

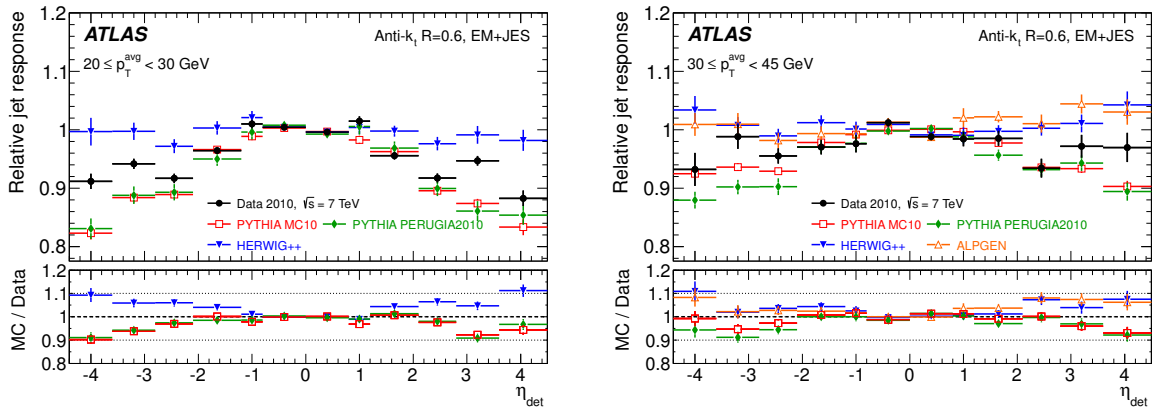
Fig. 5.9 shows the performance as a function of $p_{\mathrm{T}}^{\mathrm{avg}}$. At high pseudorapidity and low $p_{\mathrm{T}}^{\mathrm{avg}}$, no prediction is satisfactory. Because the jet energy scale calibration is based on simulation, these results show that the calibration at low $p_{\mathrm{T}}$ will be affected by such a mismodelling. For this reason, these results have been included in the calculation of the uncertainties, and provide the dominant uncertainty term for the end-cap and forward regions.

### Derivation of jet energy scale uncertainties

The jet energy scale (JES) uncertainty is determined in the central detector region using the estimate of single particle response from data, and comparing systematic variations of the Monte Carlo simulations with the nominal PYTHIA sample. The systematic variations take into account eventual under- or over-estimate of the dead material in the detector simulation, the resolution effects in the threshold used in the cells of the calorimeter for noise suppression, and a small non-closure[12] effect of the EM+JES scheme. The result is extended to the forward regions using the $\eta$-intercalibration analysis, taking the central region as a reference for the JES.

The contributions to the JES uncertainty from the modelling of the fragmentation, the underlying event and other choices in the event modelling of the Monte Carlo event generator are obtained

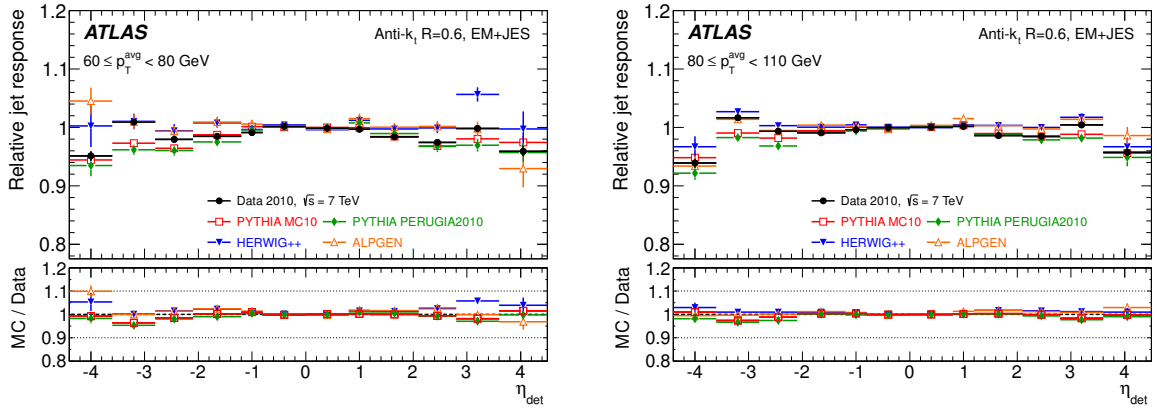
[12]The *non-closure* of the calibration estimates the residual difference in the jet response after calibration; for a more detailed analysis, see Aad et al. [10].




(a) $20 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 30$ GeV




(b) $30 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 45$ GeV

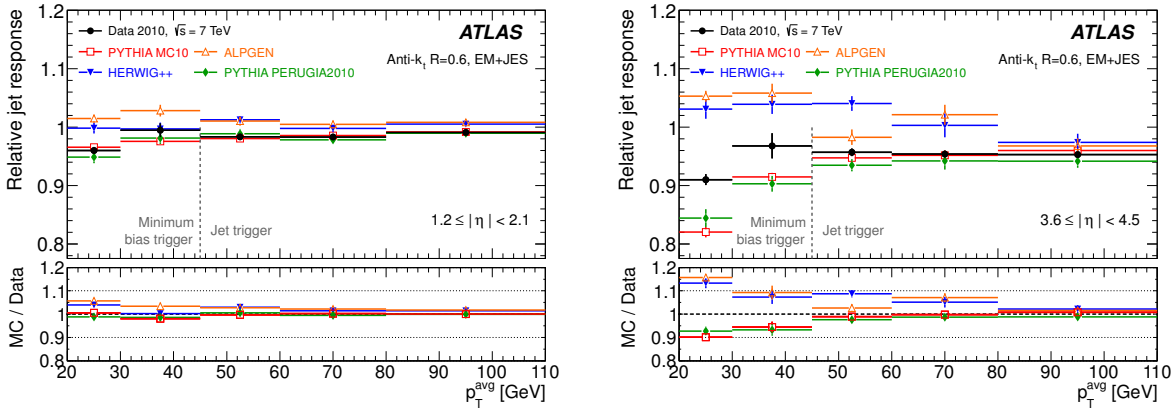



(c) $60 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 80$ GeV




(d) $80 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV

Figure 5.8: Relative jet response, $1/c$, of anti$-K_T$ jets with $R = 0.6$ as a function of the jet pseudorapidity measured using the matrix $\eta$-intercalibration method in bins of the average $p_{\mathrm{T}}$ of the two leading jets (a) $20 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 30$ GeV, (b) $30 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 45$ GeV, (c) $60 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 80$ GeV and $80 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV. The lower part of each figure shows the ratio of Monte Carlo simulation to data. Only statistical uncertainties are shown. From Aad et al. [10].




(a) $1.2 \leq |\eta| < 2.1$ (b) $3.6 \leq |\eta| < 4.5$

Figure 5.9: Relative jet response, $1/c$, of anti$-K_T$ jets with $R = 0.6$ as a function of $p_{\mathrm{T}}^{\mathrm{avg}}$ found using the matrix $\eta$-intercalibration method for (a) $1.2 \leq |\eta| < 2.1$ and (b) $3.6 \leq |\eta| < 4.5$. For $p_{\mathrm{T}}^{\mathrm{avg}} < 45$ GeV, the data are collected using the minimum bias trigger stream. For $p_{\mathrm{T}}^{\mathrm{avg}} > 45$ GeV, the data are collected using the calorimeter trigger stream. The lower part of each figure shows the ratio of Monte Carlo simulation to data. Only statistical uncertainties are shown. From Aad et al. [10].

from the systematic samples based on ALPGEN+HERWIG+JIMMY and the PYTHIA PERUGIA2010 tune[13] discussed in Section 5.3.2.

For each ($p_{\mathrm{T}}^{\mathrm{jet}}$, $\eta_{\mathrm{jet}}$)-bin, the uncertainty contributions from the calorimeter, the jet calibration non-closure, and systematic Monte Carlo simulation variations are added in quadrature.

The final uncertainty is estimated as follows:

1. The total JES uncertainty in the central region $0.3 \leq |\eta| < 0.8$ is kept as a baseline.

2. The uncertainty from the $\eta$-intercalibration is taken as the RMS deviation of the MC predictions from the data and is added in quadrature to the baseline uncertainty.

The measurements in the $\eta$-intercalibration are performed for transverse momenta in the range $20 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV. The uncertainty for jets with $p_{\mathrm{T}} > 100$ GeV is taken as the uncertainty of the last available $p_{\mathrm{T}}$-bin[14]. The uncertainties are evaluated separately for jets reconstructed with distance parameters $R = 0.4$ and $R = 0.6$, and are in general found to be slightly larger for $R = 0.4$.

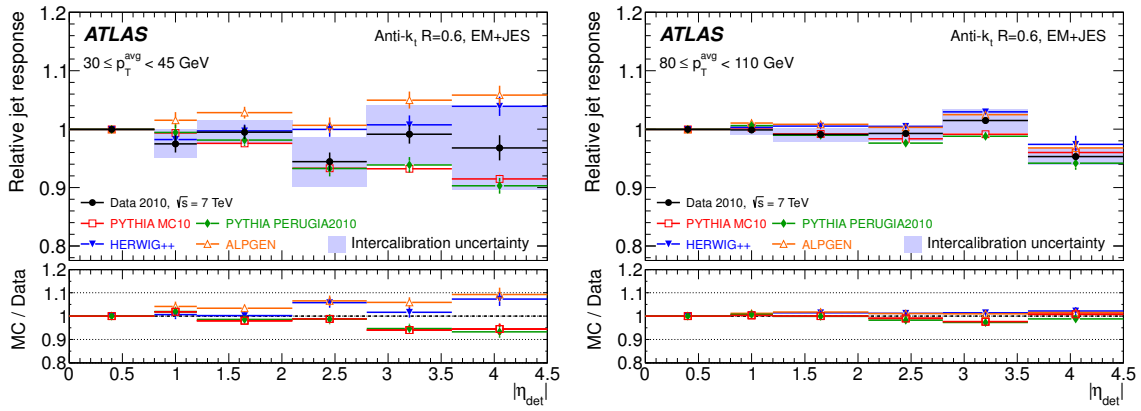
Figure 5.10 shows the relative jet response, and the associated intercalibration uncertainty calculated as above, as a function of jet $|\eta|$ for two representative $p_{\mathrm{T}}^{\mathrm{avg}}$-bins. The dependence of the uncertainty on $p_{\mathrm{T}}$ is visible, and, mostly in the forward region, is dominated by the choice of the hadronization model used in the generator.

Figure 5.11 shows the final fractional jet energy scale systematic uncertainty and its individual contributions as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ for three selected $\eta_{\mathrm{jet}}$ regions.

The fractional JES uncertainty in the central region amounts from 2% to 4% for $p_{\mathrm{T}}^{\mathrm{jet}} < 60$ GeV, and it is between 2% and 2.5% for $60 \leq p_{\mathrm{T}}^{\mathrm{jet}} < 800$ GeV. For jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 800$ GeV, the uncertainty

[13] The comparison of the standard PYTHIA samples with these two different models should give a reasonable estimate of the systematic variations. However, the possible compensation of modelling effects that shift the jet response in opposite directions cannot be excluded.

[14] This is justified by the decrease of the intercalibration uncertainty with $p_{\mathrm{T}}$, but cannot completely exclude the presence of calorimeter non-linearities for jet energies above those used for the intercalibration.




(a) $30 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 45$ GeV (b) $80 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV

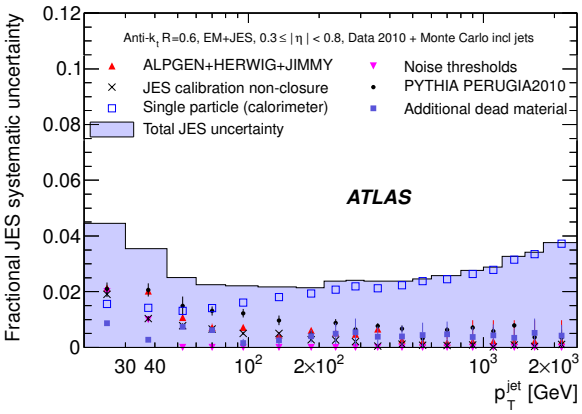
Figure 5.10: Average jet response for anti$-K_T$ jets with $R = 0.6$ calibrated with the EM+JES scheme measured relative to a central reference jet within $|\eta| < 0.8$ in data and various Monte Carlo generator samples as a function of $|\eta_{\mathrm{jet}}|$ for $p_{\mathrm{T}}^{\mathrm{avg}}$ in the ranges $30 - 45$ GeV (a) and $80 - 110$ GeV (b). The resulting systematic uncertainty component is shown as a shaded band around the data points. The errors bars on the data points only show the statistical uncertainties. From [10].

ranges from 2.5% to 4%. The uncertainty amounts to up to 7% and 3%, respectively, for $p_{\mathrm{T}}^{\mathrm{jet}} <$ 60 GeV and $p_{\mathrm{T}}^{\mathrm{jet}} > 60$ GeV in the endcap region, where the central uncertainty is taken as a baseline and the uncertainty due to the intercalibration is added. In the forward region, a 13% uncertainty is assigned for $p_{\mathrm{T}}^{\mathrm{jet}} = 20$ GeV. The increase in the uncertainty is dominated by the modelling of the soft physics in the forward region that is accounted for in the $\eta$-intercalibration contribution, as explained in Sect 5.3.4. This uncertainty contribution is estimated conservatively.

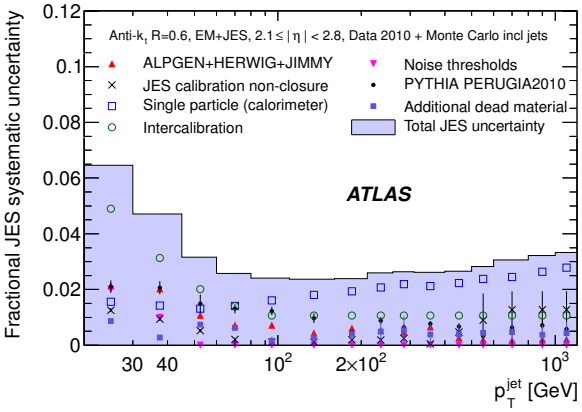
## 5.4 Outlook

The precision of the jet energy measurement with the ATLAS detector has been established in the first year of proton-proton collisions at the LHC. In the central detector the jet energy can be measured with a precision of about 2 to 3% over a wide transverse momentum range. The $\eta$-intercalibration, designed for calibrating the jet energy scale over the full $\eta$ range, has been used in 2010 to estimate the precision of the jet energy measurement in the forward region of the detector. This has been estimated to be below 5% for $p_{\mathrm{T}}^{\mathrm{jet}} > 50$ GeV, and between 5 and 12% for $20 < p_{\mathrm{T}}^{\mathrm{jet}} < 50$ GeV. This excellent performance would not have been possible without a very detailed understanding of the detector and sophisticated calorimeter calibration procedures as well as the good description of the ATLAS detector in the simulation.

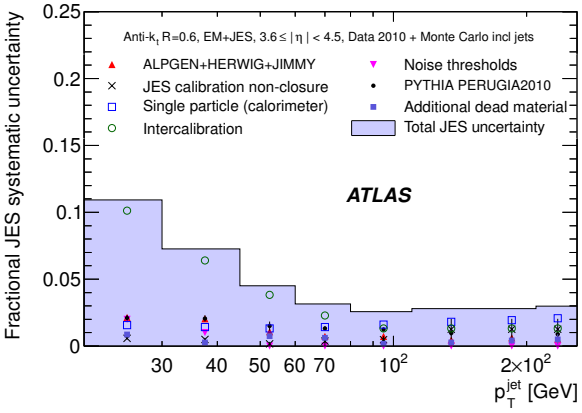
**Jet calibration development in 2011** Further improvements are possible if the *in-situ* calibrations, presented in Sect. 5.2.2, are used to derive energy scale corrections. This is the calibration strategy used in 2011; the data collected in 2010 were not sufficient to permit an *in-situ* calibration with low uncertainties. The improvement expected by using the *in-situ* calibrations is expected to be very large, and the results on jet calibration on 2011 data (paper in preparation) show that the JES uncertainty is below 3% for central jets, and reaches 6% for jets with $p_{\mathrm{T}}^{\mathrm{jet}} = 20$ GeV and $\eta = 4.5$, for both EM+JES and LCW jets corrected with factors derived from the combination of the




(a) $0.3 \leq |\eta| < 0.8$




(b) $2.1 \leq |\eta| < 2.8$




(c) $3.6 \leq |\eta| < 4.5$

Figure 5.11: Fractional jet energy scale systematic uncertainty as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ for jets in the pseudorapidity region $0.3 \leq |\eta| < 0.8$ in the calorimeter barrel (a), $2.1 \leq |\eta| < 2.8$ in the calorimeter endcap (b), and in the forward pseudorapidity region $3.6 \leq |\eta| < 4.5$. The total uncertainty, derived from the sum of the single contributions as explained in the text, is shown as the solid light shaded area. The individual sources are also shown together with uncertainties from the fitting procedure if applicable. From Aad et al. [10].

different *in-situ* calibrations. The improvement in the overall uncertainty is about a factor 2 with respect to 2010 uncertainties.

In the 2011 calibration the $\eta$-intercalibration has been used to derive calibration factors to extend the central ($|\eta| < 1.2$) JES to the forward region. These factors, and their relative uncertainties, have been produced with the matrix method, while the central reference region method has provided a cross-check of the results. The matrix method has been preferred for calibration purposes to the central reference region method, because it needs lower statistics to calibrate a large phase-space region. Instead, the central reference region method is used as a baseline for the measurement of the jet energy resolution *in-situ*, where a clearly defined reference region is needed.

**2012 goal: 1% jet energy scale uncertainty** The JES uncertainty is among the dominant uncertainties in physics analyses in ATLAS. Lowering further this uncertainty is the long term goal of the ATLAS effort; the goal would be to reach the 1% level in 2012. This is a challenging goal, and the main reasons are: the impact of the higher pile-up conditions on the measurement, and the modelling issues.

In this chapter, I haven't investigated the pile-up effects on the calibration, as the ATLAS strategy foresees that those effects should be corrected before the calibration is applied. However, in 2010 the amount of 1-vertex events was still quite high, and the data and Monte Carlo samples have been separated into two subsets: 1-vertex events, where the calibration performance have been studies, and events with more than one vertex, where the pile-up effects have been studied, and the pile-up corrections have been derived. This approach is not feasible when looking at the 2011 pile-up conditions (reported in Fig. 3.3); and pile-up can be expected only to increase with higher energies and luminosities. Every time the bunches conditions will change, a first investigation of the calorimeter performance will always be needed, and will be followed by the analysis of the performance at the jet level. This task is very challenging in particular for the $\eta$-intercalibration: pile-up effects will impact the central and forward calorimeters in a different way, and, if not understood and corrected properly, might introduce artificial effects in the analysis. The central reference region method, explained in Sect. 5.3.1, might provide anyway useful inputs, because it permits to clearly separate the different detector regions in the analysis.

This work has shown how the $\eta$-intercalibration can succeed in extending the understanding of the energy scale up to the full calorimeter coverage. The traditional bottle-neck of the *in-situ* calibrations, i.e. the need of high statistics, has been solved by the matrix method explained in Sect. 5.3.1, developed in ATLAS to overcome this difficulty. But the results are still heavily dominated by the modelling uncertainty. This is related to the difference in the Monte Carlo predictions, and not at all to the calibration. The only way to solve this issue would be to investigate the generators themselves and find the reason for their difference. As mentioned in Sect. 5.3.4, the parton shower is thought to be responsible for the large uncertainty. A better understanding of PYTHIA and HERWIG in physics analyses, together with the update of the PYTHIA code to the latest version, foreseen in 2012, is expected to shed light on this issue.

# Chapter 6

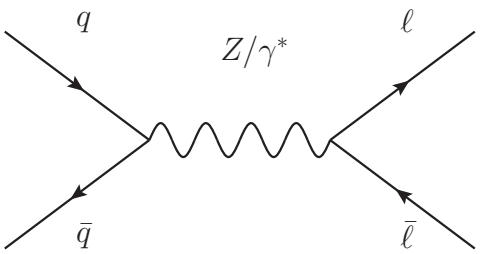
# Measurement of the production cross section for $Z/\gamma^*$ in association with jets

The main goal of the measurement performed in ATLAS is the understanding of high-order QCD effects in the hadronic final state. The theoretical background of this phenomenon was already highlighted in Sect. 2.3. The *Drell-Yan* process (Feynman diagram represented in Fig. 6.1a) provides a good test case for QCD. It provides a clean experimental signature, the leptons coming from the $Z/\gamma^*$ intermediate boson, that can be identified with high precision. The production of jets in association with a $Z$ boson (first order Feynman diagrams in Fig. 6.1b and 6.1c) occurs as a higher order QCD correction of the Drell-Yan process. The higher order corrections to the cross section predictions are non-negligible, and impact the differential distributions, making the prediction highly sensitive to the precision used in the calculation.

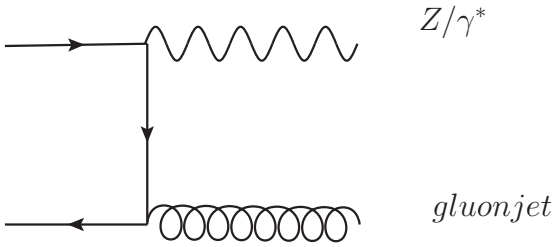
At the moment, NLO predictions are available up to $Z$ boson $+4$ jets [79], and LO results for higher orders. Because of the structure of the perturbative expansion (see Sect. 2.3 and Eq. 2.10), it is possible to predict the leading term of the cross section using leading order calculations at each order, i.e. per jet multiplicity in the final state; however, the results must be carefully merged in the inclusive predictions, to avoid double counting of terms and ensure that cancellations predicted by theory occur. This affects most of all the precision of other variables, for example the $p_{\mathrm{T}}^{\mathrm{jet}}$ distribution. A measurement of the differential cross section from the experiments is fundamental to test our understanding of theory, that still is affected by such large uncertainties.

With an integrated luminosity of 36 pb$^{-1}$, corresponding to the 2010 dataset, a measurement of the production cross section for $Z$ boson in association with jets in data at the ATLAS experiment in proton-proton collisions at 7 TeV has been performed. $Z$ bosons are identified using di-electron and di-muon decay modes, the final states with lower uncertainties in the lepton identification. Jets are reconstructed with the anti$-K_T$ algorithm. The measurements include jet multiplicities and differential cross sections with respect to the jet transverse momenta, the jet rapidity, and observables describing the mass and angular relations between the two leading jets.

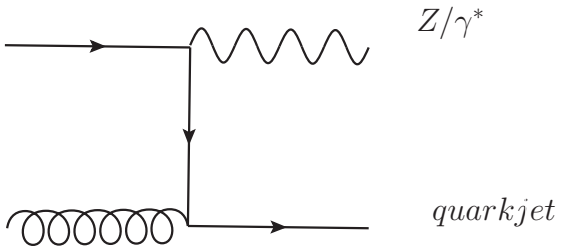
The data are compared with predictions from different generators: ALPGEN[90], SHERPA[66], PYTHIA[105] and BLACKHAT[41]. ALPGEN is a tree-level fixed-order generator able to generate events with the $Z$ boson produced in association with a large number of partons; the parton level event is then matched to HERWIG[54] for parton showering and JIMMY[49] for the underlying event simulation. SHERPA is also a tree-level fixed order generator matched with its own parton shower implementation, that uses a modified clustering algorithm, and its own underlying event model. PYTHIA is a multipurpose generator; the settings used for $Z \to \ell\ell +$ jets production are a $2 \to 2$ computation, and for this reason it can predict reasonably well events with up to 1 jet in the partonic final state, while jets beyond the leading jet are generated only during the parton shower evolution. BLACKHAT is a parton-level NLO generator implemented with the purpose of high precision predictions for QCD processes. All the generators mentioned are commonly used to estimate




(a) Feynman diagram for Drell-Yan process.




(b) Feynman diagram for $Z$ boson production with a gluon jet in the final state.




(c) Feynman diagram for $Z$ boson production with a quark jet in the final state.

Figure 6.1: Feynman diagram for $Z$ boson production and production of 1 additional jet in the final state.

signals and irreducible backgrounds in Standard Model and New Physics searches; it is important to test the quality of their prediction because they still can limit a measurement with their large theoretical uncertainties.

In the complete analysis [17] the data are unfolded to the particle level (see Fig. 4.1) and compared the prediction from the multi-purpose generators mentioned above; the BLACKHAT prediction, at parton level, is corrected to the particle level using corrections for the radiation effects derived from the parton shower evolution computed with PYTHIA. The choice of the particle level as a reference for the measurement is not the only one possible, but has been preferred to the parton level,

| Trigger | data period | integrated luminosity [pb$^{-1}$] |
|---|---|---|
| L1_EM_14 | A-E3 | 0.7 |
| EF_e15_medium | E4-I2 | 35.5 |
| EF_mu10_MG | E4-G1 | 2.9 |
| EF_mu13_MG | G2-I1 (run 166788) | 15.3 |
| EF_mu13_MG_tight | I1, run167607 ff | 17.3 |

Table 6.1: Integrated luminosities for the various data taking periods.

that requires an additional care in treating the effects of gluon radiation in the unfolding.

In the context of this thesis, particular attention will be devoted to the $\mu^+\mu^-$ final state, that provides the cleanest leptonic final state, and to the problematics of the detector level measurement. As the interaction between well isolated muons and jets is negligible, this final state can be expecially used to investigate the behavior of jet reconstruction at the detector level. Many variables useful for the Higgs analysis will be investigated in Sect. 6.3.2. The impact of the jet energy scale uncertainties and resolution will also be treated in detail in Sect. 6.4.1.

The analysis is performed within a private framework developed in ROOT [99]. The framework has been developed in the Bonn group by G. Gaycken *et al.* [65], and the event selection has been then adapted to the $Z \to \ell\ell$ + jets selection, following the guidelines highlighted in Sect. 6.2.

## 6.1 Data and simulation samples

The analysis is based on a sample of $\sqrt{s} = 7$ TeV proton-proton collisions collected in 2010. We use data collected during periods of stable beam operation, and restrict to data acquired with all ATLAS sub-detectors at nominal bias voltages. We apply a Good Run List (GRL) criterion to select luminosity blocks that satisfy our quality criteria. We also require that data quality (DQ) shifters certify the data as usable for analysis based on electrons, muons, jets and missing transverse energy.

Events are selected with a combination of a hardware-based L1 trigger and a software-based higher-level trigger (see Sect. 3.2). The trigger strategy, for both electrons ("EM" and "e") and muons ("mu") triggers used, is listed in Table 6.1. All these triggers are unprescaled and events with trigger problems are discarded by the GRL. The total amount of data collected sums up to an integrated luminosity of 36.2 pb$^{-1}$ in the electron channel and 35.5 pb$^{-1}$ in the muon channel.

The data are compared to simulation both at detector and at particle level.

Signal events ($Z \to \ell\ell$ + jets) are generated using ALPGEN v2.13 interfaced with HERWIG v6.510 for parton shower and fragmentation into particles and to JIMMY v4.31 to model underlying event contributions (using the AUET1 (LO*)tune [18]). The leading-order parton density functions (PDFs) CTEQ61L [109] are used.

Similar samples are generated using SHERPA v1.2.3with $m_T$ scale and the NLO PDFs CTEQ6.6. In addition, $Z \to \ell\ell$ + jets samples are produced ($q\bar{q} \to Z/\gamma^* g$ and $qg \to Z/\gamma^* q$) using PYTHIA v6.421 and HERWIG plus JIMMY with the modified LO PDF set MRST2007LO* [104].

Background samples from $W$ and $Z \to \tau\tau$ production, and from diboson processes ($WW$, $WZ$ and $ZZ$) are generated similarly to the signal samples, using ALPGEN with HERWIG and JIMMY with CTEQ61L PDFs, normalized to NNLO and NLO pQCD predictions, respectively.

Top-quark pair production samples are generated using the NLO generator MC@NLO [63] to-

gether with the NLO PDF set CTEQ6.6.

The generated samples are passed through a full simulation of the ATLAS detector and trigger, based on GEANT4. The simulated events are then reconstructed and analyzed with the same analysis chain as for the data, and the same trigger and event selection criteria. Details of these samples are summarized in Table 6.2. All samples are part of the ATLAS MC10 production.

The program PHOTOS [68] is used to simulate final state QED radiation in ALPGEN and PYTHIA samples. SHERPA uses a self designed parton shower algorithm.

The W and Z samples are normalized globally to next-to-next-to-leading order (NNLO) pQCD inclusive Drell-Yan predictions computed by the FEWZ [64] program. For hadron-level comparisons, the PYTHIA $Z \to \ell\ell$ + jets cross sections are later scaled to the cross section measured in data. The resulting global scale factor is $1.186 \pm 0.021$.

The Pile-up simulation corresponds to a bunch-train setup with double trains with 225 ns (= 9 BC) separation, each containing 8 filled bunches with 150 ns bunch separation. The train arrangement reproduces the accelerator settings used from Run period D onwards, i.e. in the data used for this analysis. The number of minimum-bias interactions follows a Poisson distribution with a mean of approximately two. These samples have been re-weighted such that the distributions of the primary vertices follows the distribution observed in data. The weights have been derived at an early preselection stage, to avoid selection biases.

The multi-jets background, instead, is extracted from data. The presence of this background is due to the mis-identification of a jet as a lepton. In the electron channel, is dominated by jets faking electrons in the final state, thus by a calorimeter mis-measurement or by a mis-identification; in the muon channel, mainly originates from heavy-flavour jet production processes, with muons from bottom and charm quark decays, as well as from the decay-in-flight of pions and kaons, which are highly suppressed by the isolation requirement applied to the muon candidates (see Sect. 6.2). For this reason, the influence of this background is higher in the electron channel than in the muon channel.

## 6.2 $Z \to \ell\ell$ + jets selection

The selection requirement for event selection are compatible with all ATLAS analyses. In particular, data quality requirements on the luminosity blocks considered for the analysis are applied before the object selection.

### $Z$ boson selection

Quality requirements on the leptons are applied, and they are compatible with those applied in the other $W$ and $Z$ boson analyses. Those requirements are treated in detail in [8] for electrons and in [19, 20] for muons. As this chapter will mostly focus on muons, the details of the muon selection and requirements are also reported in Table 6.3. For the details about the other channel, we refer the reader to Aad et al. [17].

The $Z \to \mu^+\mu^-$ + jets sample is collected online using a trigger that requires the presence of at least one muon candidate reconstructed in the muon spectrometer, consistent with having originated from the interaction region with $p_T > 10$ GeV or $p_T > 13$ GeV, depending on the data period, and with the majority of the data taken with the higher threshold, and $|\eta| < 2.4$. The muon candidates are associated with track segments reconstructed in the inner detectors which, combined

| process | data set | generator | cross section (nb) | note |
|---|---|---|---|---|
| $Z \to ee$ | 107650-107655 | ALPGEN (+HERWIG+JIMMY) | 1.07 | $m_{\ell\ell} > 40$ GeV |
| $Z \to ee$ | 109126 | SHERPA v1.2.3 | 1.07 | $m_{\ell\ell} > 40$ GeV |
| $Z \to ee$ | 106050 | PYTHIA | 0.99 | $m_{\ell\ell} > 60$ GeV, 1lep |
| $Z \to \mu\mu$ | 107660-107665 | ALPGEN (+HERWIG+JIMMY) | 1.07 | $m_{\ell\ell} > 40$ GeV |
| $Z \to \mu\mu$ | 109127 | SHERPA v1.2.3 | 1.07 | $m_{\ell\ell} > 40$ GeV |
| $Z \to \mu\mu$ | 106051 | PYTHIA | 0.99 | $m_{\ell\ell} > 60$ GeV, 1lep |
| $W \to e\nu$ | 107680-107685 | ALPGEN (+HERWIG+JIMMY) | 10.46 | |
| $W \to \mu\nu$ | 107690-107695 | ALPGEN (+HERWIG+JIMMY) | 10.46 | |
| $Z \to \tau\tau$ | 107670-107675 | ALPGEN (+HERWIG+JIMMY) | 1.07 | $m_{\ell\ell} > 40$ GeV |
| $t\bar{t}$ | 105200 | MC@NLO | 0.165 | not full-hadr.: $\epsilon = 0.543\%$ |
| $WZ$ | 107104-107109 | ALPGEN (+HERWIG+JIMMY) | 0.00167 | lepton filter |
| $ZZ$ | 107108-107111 | ALPGEN (+HERWIG+JIMMY) | 0.00103 | lepton filter |
| $WW$ | 107100-107103 | ALPGEN (+HERWIG+JIMMY) | 0.00445 | lepton filter |

Table 6.2: Monte Carlo samples used in the analysis. The cross-sections quoted are the ones used to normalize estimates of expected number of events. The cross-sections for the QCD sample, the $b\bar{b}$ and the $c\bar{c}$ sample are directly from PYTHIA and the cross sections for the Di-Boson samples are taken from ALPGEN at LO, scaled with a global k-factor of 1.21. Sources for the other cross-sections are discussed in the text.

| | |
|---|---|
| **Collision event selection** | |
| Primary vertex | $N_{\text{vtx}} \geq 1$ with $N_{\text{tracks}} \geq 3$ |
| **Good-muon selection** | |
| Phase space | $p_{\text{T}} > 20$ GeV, $\lvert\eta\rvert < 2.4$ |
| Muon ID | Staco, combined |
| | author 6 |
| Muon cleaning | $z_0 < 10$ mm (wrt. the primary vertex) |
| | $d_0/\sigma(d_0) < 3.0$ ($d_0$ wrt. the primary vertex) |
| | $N_{\text{B-layer}} > 0$ (if expected) |
| | $N_{\text{PIXhits}} + N_{\text{crossed-dead-pixel-sensors}} \geq 2$ |
| | $N_{\text{SCThits}} + N_{\text{crossed-dead-SCT-sensors}} \geq 6$ |
| | $N_{\text{Pixelholes}} + N_{\text{SCTholes}} < 2$ |
| | if $\lvert\eta\rvert < 1.9$ then $N_{\text{TRThits}} \geq 6$ and $N_{TRT}^{\text{outliers}}/N_{\text{TRT}} < 0.9$ |
| | if $\lvert\eta\rvert \geq 1.9$ and $N_{\text{TRThits}} \geq 6$ then $N_{\text{TRT}}^{\text{outliers}}/N_{\text{TRT}} < 0.9$ |
| | with $N_{\text{TRT}} = N_{\text{TRT}}^{\text{outliers}} + N_{\text{TRThits}}$ |
| Track isolation | $\Sigma p_{\text{T}} < 0.1 \times p_{\text{T}}(\text{muon})$ in $\Delta R < 0.2$ around the muon track |
| **$Z \to \mu\mu$ event selection** | |
| | Al least 2 good muons |
| Charge | Opposite sign |
| Invariant Mass | $66 < m_{\mu\mu} < 116$ GeV |

Table 6.3: Event selections for the $Z \to \mu^+\mu^-$ + jets analysis.

with the muon spectrometer information, define the final muon track. Combined muon tracks with $p_{T\mu} > 20$ GeV and $|\eta_\mu| < 2.4$ are selected.

The quality requirements for muons are: the associated inner detector track segment is required to have a minimum number of hits in the pixel, SCT and TRT detectors, and the muon transverse and longitudinal impact parameters, $d_0$ and $z_0$, with respect to the reconstructed primary vertex[1] are required to be $d_0/\sigma(d_0) < 3$ and $z_0 < 10$ mm in the $r-\phi$ and $r-z$ planes, respectively, where $\sigma(d_0)$ denotes the $d_0$ resolution.

The muons are required to be isolated: the scalar sum of the transverse momenta of the tracks in an $\eta-\phi$ cone of radius 0.2 around the muon candidate is required to be less than 10% of the muon $p_T$. Events are selected with two oppositely charged muons and an invariant mass 66 GeV $< M_{\mu^+\mu^-} <$ 116 GeV.

The muon tracks selected are not matched with the trigger muons, which are reconstructed with a different combined algorithm ("MuGirl"). The muons used for the reconstruction of the $Z$ boson are the two leading muons. The possibility that either only one of the offline muons, or both, or neither of them[2] has triggered the event, is treated in the Monte Carlo analysis, in order to ensure that the trigger vs. offline selection efficiency is consistent between data and simulation.

The muons are required to be on trigger plateau, i.e. in the $p_T$ range where the trigger selection is fully efficient.

In both the electron and muon channel, events are required to have a reconstructed primary vertex of the interaction with at least 3 tracks associated to it, which suppresses beam-related background contributions and cosmic rays. The selected dilepton samples contain a total of 9705 and 12582 events for the electron and muon channels, respectively.

### Jet and event selection

In contrast to the first $Z$ boson analyses in ATLAS and previous measurements at other colliders, this analysis uses the full calorimeter coverage for jet reconstruction. The events are selected according to general requirements on the status of the calorimeter in those lumiblocks, and then the jets are first reconstructed, as described already in Chap. 5.

The jet algorithm used in the analysis is anti$-K_T$, calibrated with the standard 2010 calibration (EM+JES), explained in 5.2.2. In particular, as the analysis uses the full 2010 dataset, an energy correction has been applied to make the energy measurement in the calorimeter independent on the pileup conditions[3]. The investigation of the jet performance has shown that this subtraction

[1] "Primary vertex" is defined as the good vertex with the highest value of $\sum_{\text{tracks}} p_{T,\text{tracks}}^2$. Good vertices are all vertices that pass an analysis-dependent selection criteria. For analysis considering electrons and muons, vertices are considered as good if associated with at least 3 tracks; for other objects, as hadronic taus or jets, where the number of tracks defining the object is by definition higher than 3, the requirement on the number of tracks for a good vertex is 4 or higher.

[2] This third case is a case where more than two good muons are present in the event: the muon track that caused the trigger to fire is not matched to any of the two leading muons. 3 events of this kind have been found in the 2010 analysis.

[3] This correction is called *Jet Offset Correction*. It uses an estimation of the average energy per cell inside a topocluster jet due to pileup, to subtract this energy already at EM-scale. Then, the new jet is calibrated using the EM+JES scheme. The amount of energy to be subtracted has been estimated from 2010 data, and is applied as a function of the number of vertices with at least 5 tracks associated in the event. The additional energy offset measured is of $(500 \pm 160)$ MeV within the jet cone per each extra interaction. The derivation of this correction, as well as the closure tests, showing that the corrections restore the JES removing the NPV dependence, are reported in Aad et al. [10].

technique was adequate for the needs and the precision of the 2010 analysis, and that no additional pileup treatment was needed.

As the presence of a jet and an electron close by can bias the measurement of both objects, jets overlapping with electrons have been removed. Similarly, jets overlapping with muons have been removed, as well as non-isolated muons, as the probability that such muons come from a secondary vertex inside the jet is quite high.

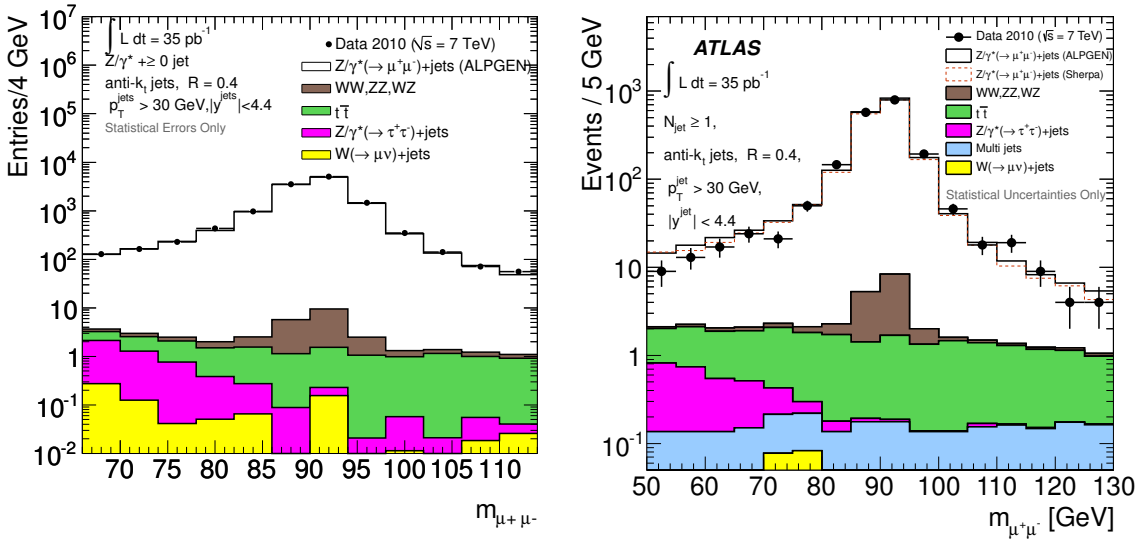
Also jets that could be due to fake signals (due e.g. to bad calorimeter conditions, or beam halo or cosmic particles) are removed. The final jet collection obtained is then passed through a requirement $p_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV.

## 6.3 Detector level results

### 6.3.1 Detector level validation

To validate the selection, data/Monte Carlo comparisons, as well as comparisons with the results of the ATLAS inclusive $Z$ boson analysis[16], have been performed. The results were found to be consistent.

After the event selection, 1514 (1885) events in the $Z \to ee$ ($Z \to \mu\mu$) channel have at least one associated jet. In Fig. 6.2 the invariant mass of all the candidates selected, and of the $Z \to \mu\mu$ candidates with at least one associated jet, are reported. The description of the Z peak is found to be very accurate, indicating an excellent understanding of the detector performance. In Fig. 6.3a, instead, the jet multiplicity is shown.

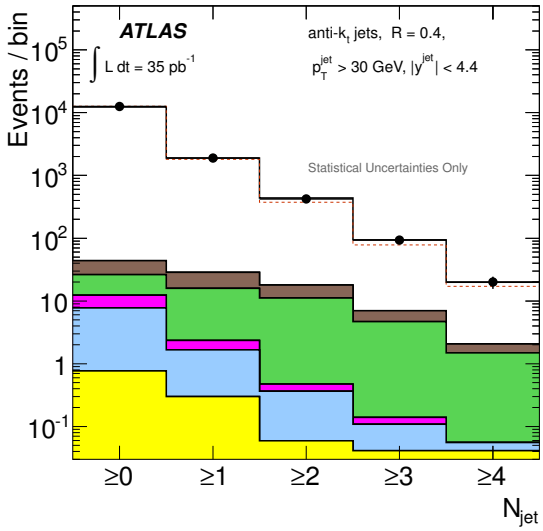



(a) Inclusive invariant mass plot.

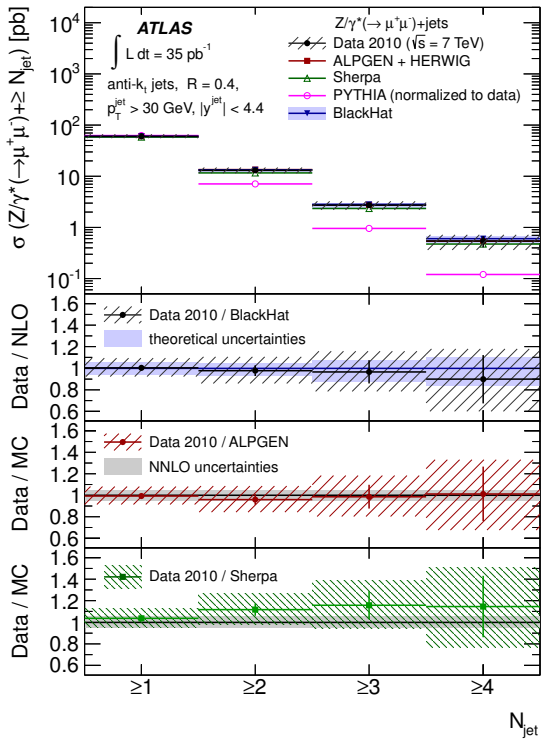
(b) Main $Z$ boson distributions [17]

Figure 6.2: Di-muon distributions in (a) the inclusive sample and (b) in events with at least one associated jet.

Fig. 6.2b and 6.3a show the comparison at the detector level with SHERPA, which has been investigated as an alternative model to ALPGEN. These inclusive distributions show no issues in the description that the two generators give of the process. A more detailed investigation though, that




(a) Inclusive jet multiplicity at detector level.




(b) Unfolded multiplicity distribution.

Figure 6.3: Inclusive jet multiplicity (a) at detector level and (b) after unfolding [17].

has also taken into account the systematic uncertainties, is reported in Sect. 6.4.3.

### 6.3.2 Jet differential distributions

The rapidity region considered in the analysis goes up to $|y| < 4.4$: jets with higher rapities suffer from the limited calorimeter coverage, because part of the belonging particles are lost in the beampipe, and their energy can't be measured with high precision.

The analysis has used the full calorimeter coverage, because the performance of the forward region, reported in Chap. 5, were sufficiently good for the measurement. I have studied in details the impact of the inclusion of the forward jets in the analysis, in order to avoid a loss of precision in case the jet description in that region was not satisfactory. Even though forward jets suffer from a large uncertainty, they are not the dominant component in this analysis. The pileup conditions, that worsen the performance of forward jets, were simple enough to avoid the need for additional strategies (on top of the offset correction) in dealing with pile-up effects in the forward region. When considering forward jets, there is no possibility to measure the exact vertex those jets belong to, because of the limited tracker acceptance. I have studied the impact of a mis-association and found it negligible with respect to the statistical uncertainty in the forward region in 2010.

In general, the description of jet variables shows good agreement with what observed in data. Picture 6.4 shows the rapidity distribution of all jets, in the events that passed the full cutflow. In particular, it is very interesting to show that the agreement is good also in the FCal, even though the amount of data did not permit any more precise investigation of the performance.

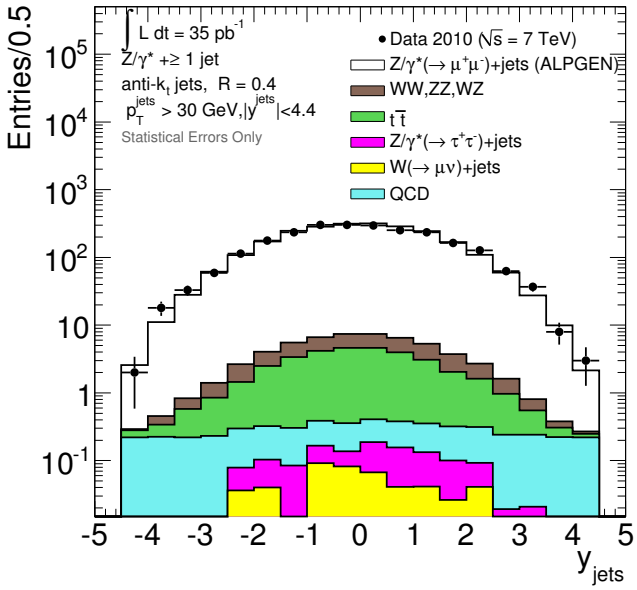



Figure 6.4: Rapidity of all jets considered in the $Z \to \ell\ell$ + jets analysis, at detector level.

The sample is large enough to allow a quite precise measurement of the characteristics of the leading jets[4] of the event. The jets in a $Z+2$ jets events can be either produced in the fragmentation process, or be produced independently in the hard interaction. The fact that PYTHIA, that reproduces quite well the production of jets in $Z+1$ jet events, fails in describing higher multiplicities, as shown in Fig. 6.3b, proofs that the production of jets in the hard scattering can't be reproduced within

[4]The jets are ordered using a $p_T$ criterion. The $p_T$ considered for the ordering is the $p_T$ at the EM+JES scale (considering also the Offset correction).

a parton shower generator. The measurement of $\Delta y$ between the leading jets, a quantity highly sensitive to hadronization effects, shows that even those jets tend to be produced with a small separation in rapidity, as expected due to the color flow between the two partons in the production process. In Fig. 6.5a the $\Delta y$ is shown at detector level for a large range; Fig. 6.5b instead shows the result at particle level, where the behavior of the jets produced only from a parton shower (PYTHIA) shows clearly a not satisfying description of the data. The matrix element generators (ALPGEN and SHERPA), instead, provide a more precise description of the jets in the event. The difference between SHERPA and the data at the core of the distribution provides also new inputs for the SHERPA hadronization model, and is being studied by the authors. The difference at high rapidities will instead be investigated further in future measurements.

In addition to those variables, the investigation of the rapidity gaps in $Z \to \mu^+\mu^- +$ jets events has considered also possible additional jets. Unfortunately, the statistics in 2010 was too low to permit a more sophisticated study. Fig. 6.6 shows one of these variables, $Z^*$(defined in Eq. 4.3), that puts in relation the rapidity of the leading jets with the rapidity of the third jet. Events with a low value of $|Z^*|$ have additional jets in between the two leading jets of the event.

Consistent results have also been obtained in the electron channel. The two channels differ mostly for the reconstruction and identification efficiencies, that explains the difference already mentioned in the number of $Z \to ee$ and $Z \to \mu\mu$ events collected. Also the acceptance of the calorimeter and the muon spectrometer are different. For this reason, the analysis is kept separated into two channels at the particle level: because of the large difference in the efficiencies, the unfolding corrections for electrons are higher than those for muons. After unfolding, the two measurements have been combined (after extrapolation to a common acceptance).

## 6.4 Comparison with Monte Carlo LO and NLO pQCD predictions

Several uncertainties contribute to the measurement, dominated by the jet energy uncertainties. For the purpose of this thesis, they are also the most important and will be treated in more details. The details for the other uncertainties are in Aad et al. [17]; they consist of:

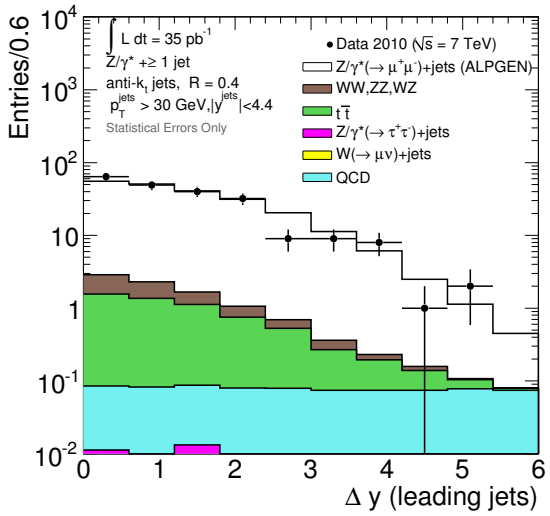
- leptons selection and reconstruction efficiencies: it includes uncertainties on the absolute energy scale and energy resolution, the uncertainty on the identification efficiency (only for electrons), and the uncertainties on the reconstruction scale factors applied to the MC simulation. The trigger efficiency uncertainty has been found to be negligible;
- the QCD background uncertainty for $Z \to ee$ (for $Z \to \mu\mu$ is found to be negligible).

### 6.4.1 Jet systematic uncertainties

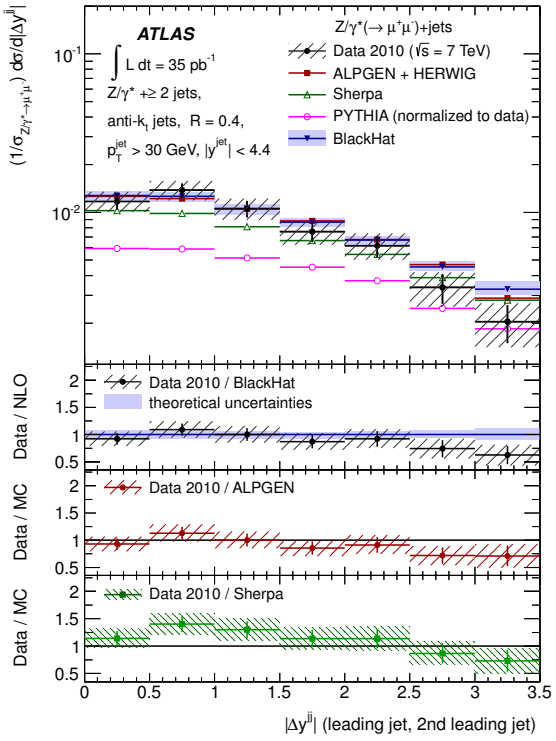
#### Jet energy scale uncertainty

For the $Z \to \ell\ell +$ jets analysis the uncertainty in the central and forward region needs to be understood. This is provided by the jet in-situ analyses, as shown in Chap. 5. The results have shown that the modelling, mostly for $p_{\mathrm{T}} < 50$ GeV, is suffering from large uncertainties, larger when the jet is reconstructed at high $|\eta|$.

The effect of a jet energy scale mismeasurement is visible in the number of jets counted in the different jet bins. Given that the energy scale uncertainty for jets is moreover larger than the




(a) Detector level $\Delta y$ distribution




(b) Unfolded $\Delta y$ distribution

Figure 6.5: Jet distributions from the 2010 analysis [17].

uncertainties for the other objects in the event, this uncertainty is also the dominant one in the 2010 analysis.

An example at detector level of the kind of effect the jet energy scale uncertainty can have on the analysis is shown in Fig. 6.7, only on the signal sample.

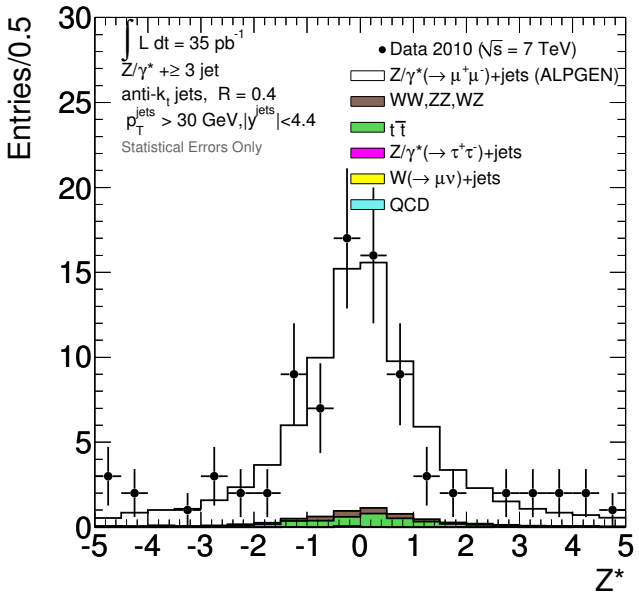
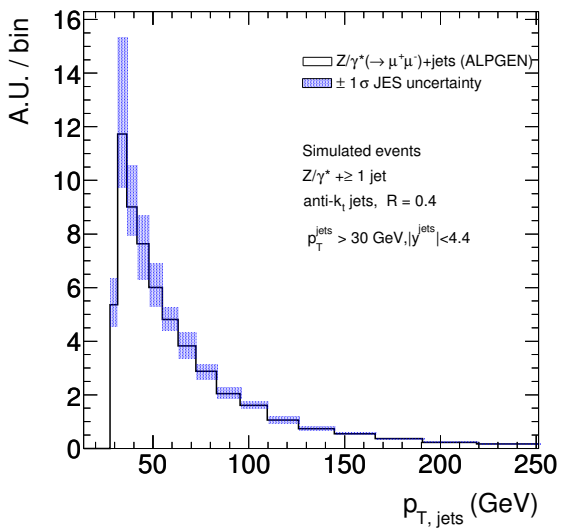
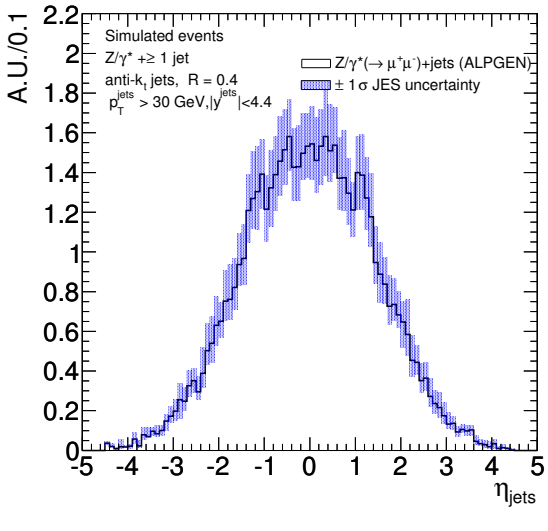



Figure 6.6: $Z^*$distribution observed at detector level.




(a) Detector level distribution from the Monte Carlo sample used in the analysis, and the same sample rerun with all jet energies scaled up and down by $1\sigma_{\text{JES}}$.




(b) Detector level $\eta_{\text{jets}}$ distribution from the Monte Carlo sample used in the analysis, and the same sample rerun with all jet energies scaled up and down by $1\sigma_{\text{JES}}$.

Figure 6.7: Comparison between the Monte Carlo sample used in the analysis and the same sample rerun with all jet energies scaled up and down by $1\sigma_{\text{JES}}$, for (a) and (b) $\eta_{\text{jets}}$. The uncertainty used in these plots were still preliminary and are not the final uncertainties used in ATLAS for 2010. Plots for illustrative purpose only.

### Jet energy resolution

An *in-situ* method similar to the $\eta$-intercalibration, but optimized for the measurement of the jet energy resolution, is implemented in ATLAS. With this method, the width of the asymmetry distributions, extracted from the comparison between two jets in a di-jet event as explained in Sect. 5.3.1, is used to estimate the $p_{\mathrm{T}}$ resolution [98]. In 2010, the jet energy resolution observed in data is reproduced well by the simulations used.

The uncertainties of the method used for the measurement in data, as well as any additional effect (for example, the fact that the resolution changes with rapidity, due to the different cell sizes in the calorimeter), need to be estimated to correctly take into account the resolution effects in the unfolding. These uncertainties are known for $|y| < 2.8$.

For high rapidities ($|y| > 2.8$), the data samples collected in 2010 are not large enough to allow an estimate of the resolution with the in-situ methods. As the jet energy resolution method assumes that the asymmetry distributions can be modeled with a gaussian distribution, and the extraction of the resolution is performed via a gaussian fit, any distribution where the statistics is not high enough to permit a good fit is discarded. This affects jets with $p_{\mathrm{T}} < 30$ GeV and/or $|y| > 2.8$.

The solution proposed by my studies, in collaboration with the Jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ ATLAS group and other ATLAS analyses, is to extrapolate the resolution measurements to the forward region and to low $p_{\mathrm{T}}$, using the knowledge developed in the central region and additional data–Monte Carlo comparison.

The good agreement between the resolution obtained in data and in simulations for the central region is the first point supporting this approach. The parametrization used is derived from simulation studies: thus, the comparison with data, and with results from different methods, consists basically in a validation of the resolution of the Monte Carlo sample.

To extrapolate in rapidity, I compared the resolution obtained in the forward region from the simulation and the measured resolution curves in the central region. To take into account the detector effects on the resolution, the region $|y| > 2.8$ has been divided into two subregions ($y \in [2.8, 3.6]$ and $y \in [3.6, 4.5]$, roughly reproducing the structures of the calorimeter). The comparison with the simulations shows that the resolution curves in those two regions agree well, within the uncertainties, with the resolution curve in the region $y \in [2.1, 2.8]$, the last region where the in-situ measurements are available. This is related to the energy range considered. At high $\eta$ even low-$p_{\mathrm{T}}$ jets are highly energetic: this compensates for the higher electronic noise expected from the larger cell size, and globally the performance of the forward region are compatible with more central regions, at least in the range $p_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV.

I have tested an extrapolation of the uncertainty, from the region $y \in [2.1, 2.8]$ to the region $|y| > 2.8$, as we did not observe any difference in the resolution curves between these regions of the detector in the simulation. Also, the $\eta$-intercalibration analysis shows that the dominant uncertainty in the forward region was due to the modeling (see Sect. 5.3.4 for details); for the energy resolution, modeling effects were shown to be rather constant in $p_{\mathrm{T}}$ and $y$. The results show that the extrapolation at high $\eta_{\mathrm{det}}$ is reasonable solution to the problem.

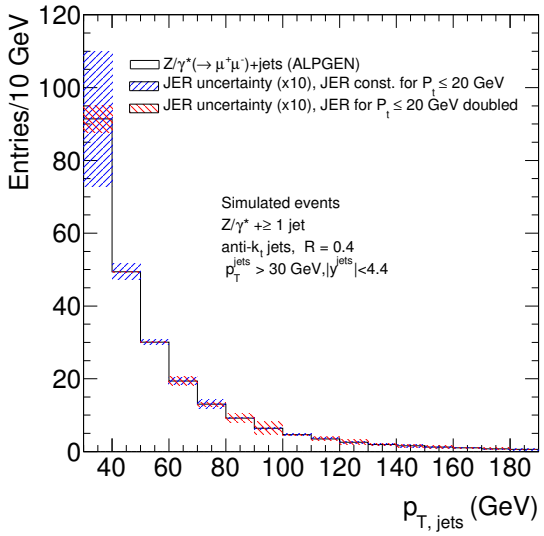
A different case is the extrapolation at low $p_{\mathrm{T}}$. In this case, the resolution due to detector electronics becomes very large, and becomes dominant for $p_{\mathrm{T}} \lessapprox 15$ GeV. The resolution effects on the analysis will be likely dominated by jets whose $p_{\mathrm{T}}$ is around the $p_{\mathrm{T}}$ threshold for the different $p_{\mathrm{T}}$ bins. Because of resolution effects in the calorimeter, the jet $p_{\mathrm{T}}$ can migrate from one bin to another one. As the resolution tends to increase with lower $p_{\mathrm{T}}$, the migration from a lower $p_{\mathrm{T}}$ bin to a higher $p_{\mathrm{T}}$ bin will be larger than the fluctuation in the opposite direction.

The $p_{\mathrm{T}}$ bins used in the $Z \to \ell\ell$ + jets 2010 analysis are designed to minimize this effect: in this way, the migration should be not negligible, but contribute with a small term to the overall

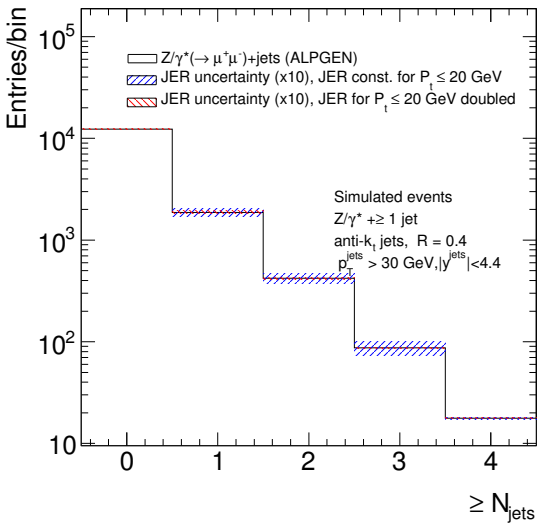
uncertainty, due to a sort of "balancing" effect of up-side and down-side migration. The only case where this migration cannot be balanced is the $p_T$ threshold for jet selection. As only the upwards fluctuations will show up in the analysis, and won't be balanced by the downward fluctuations, the first $p_T^{jet}$ bin will show the largest effects.

The region below the $p_T^{jet}$ threshold is also the region where no in-situ measurement of the resolution is available. To perform an extrapolation there, I first investigated the impact of such extrapolation could have on the analysis, and second tried a conservative estimate for an extrapolation factor, from the results available.

The first scenario I considered is the case where the resolution is constant for jets with $p_T \leq 30$ GeV. As this approximation is definitely too simple for low $p_T$ jets, I have also considered the case where the uncertainty below 20 GeV becomes twice as large as between 20 and 30 GeV. The results of these initial studies are shown in Fig. 6.8.




(a) JER impact on the $p_T^{jet}$ distribution. The bands show the JER uncertainty times 5.




(b) JER impact on the multiplicity distribution. The bands show the JER uncertainty times 5.

Figure 6.8: Comparison between nominal Monte Carlo and the same sample rerun with all jet energies smeared according to the two scenarios mentioned in Sect. 6.4.1. The bands show the JER uncertainty times 5 (to make the effect visible in the plot). The final JER uncertainty is computed after several iterations. Plots for illustrative purpose only.

The results show that the analysis is not much affected by this kind of fluctuations, also in the lowest $p_T^{jet}$ bin, as shown in 6.8.

The strategy for the ATLAS 2010 data is agreed to perform the extrapolation in rapidity as mentioned above, and to perform the extrapolation below 30 GeV using the 30 GeV uncertainty as a baseline. The uncertainty for low $p_T$ increases smoothly with decreasing $p_T$, ending with being three times higher than the baseline at 10 GeV. This assumption is very conservative when compared with the results from simulations, and is expected to cover any uncertainty related to the loose constraint of the noise term from the in-situ measurements.

The final results are included in Aad et al. [17]. Some examples of the resolution uncertainties, obtained with the mentioned assumptions, are shown in Fig. 6.9.

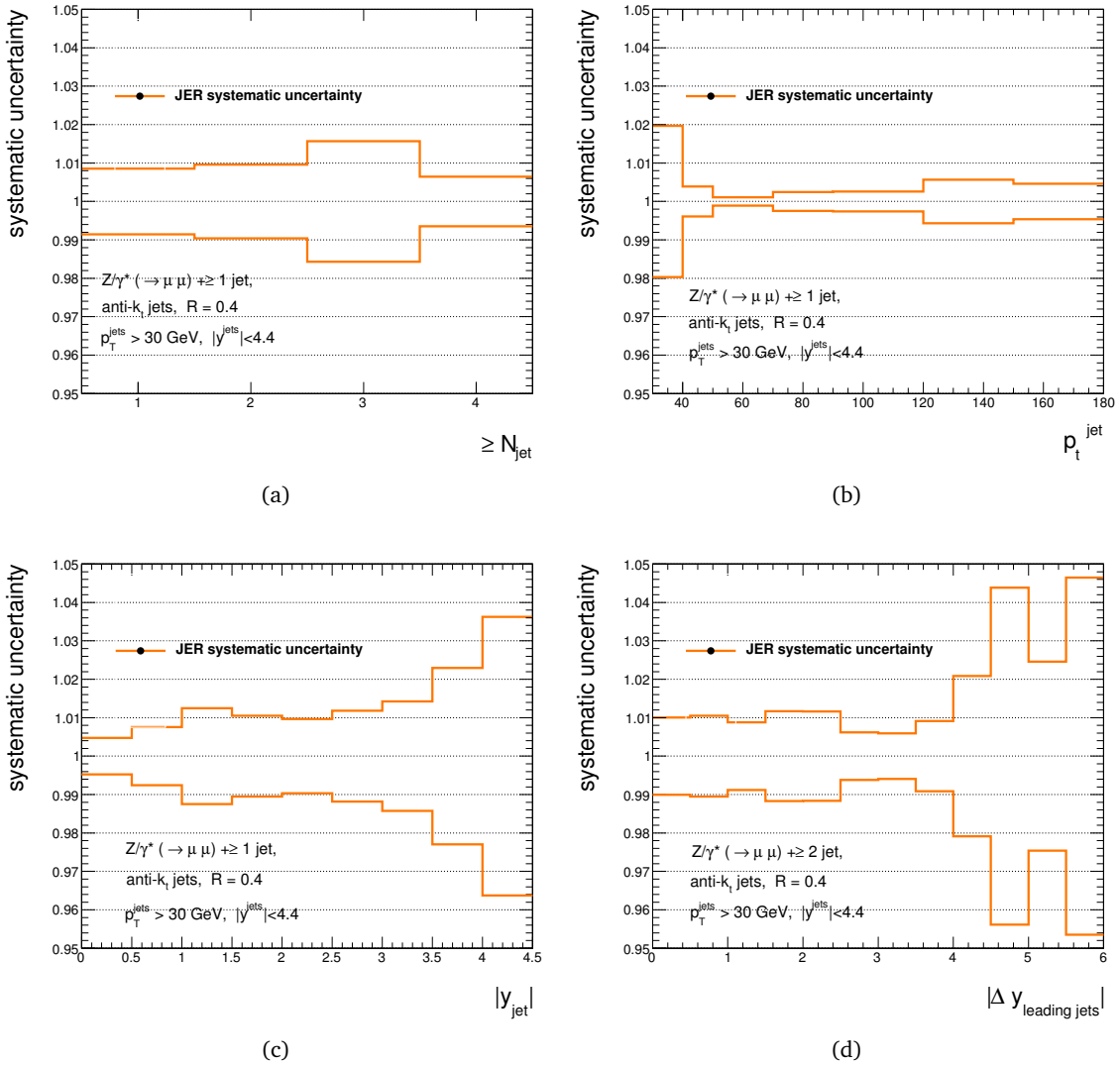



Figure 6.9: Uncertainty of the measured differential cross section due to the jet energy resolution [17], measured as the relative systematic variation between the sample used in the analysis and the variated samples.

### 6.4.2 Unfolding of detector results

The results shown in Fig. 6.10–6.13 summarize the measurement performed in 2010. The results are corrected to the particle level using a bin-by-bin unfolding technique. The central values of the bin-by-bin corrections are derived from the fully-reconstructed $Z \to \ell\ell$ + jets sample generated with ALPGEN as described in Sect. 6.1. In this method, the cross section for particle level jets $\frac{d\sigma}{d\alpha}$ with respect to a jet distribution in the variable $\alpha$ is derived using the unfolding correction $U(\alpha)$ which is the ratio of the number of events $N_{\text{particle level}}$ and after reconstruction $(N_{\text{reconstructed}})$ $for\ a\ given\ bin\ in\ \alpha$:

$$U(\alpha) = \frac{N_{\text{particle level}}}{N_{\text{reconstructed}}} \tag{6.1}$$

The ratio U($\alpha$) is used as follows to derive the central value of the final particle level cross section:

$$\frac{d\sigma}{d\alpha} = \frac{N_{\mathrm{DATA}} - N_{\mathrm{BKG}}}{\int \mathscr{L} dt} * U(\alpha) \tag{6.2}$$

The particle level state corresponds to *dressed* muons and electrons, and jets clustered on all final-state particles excluding the dressed Z decay products. The 4-vector of a dressed lepton is calculated by adding the 4-vectors of all photons from the Z decay in a radius of $\Delta R < 0.1$ around the lepton to the 4-vector of the lepton. The cross sections are measured within the fiducial acceptance region, corresponding to exactly two muons or exactly two electrons, which fulfill the same phase space requirements as the detector level objects in the selection detailed in Sect. 6.2:

| Selection | Acceptance cuts |
|---|---|
| $Z \rightarrow \mu\mu$ | 2 OS $\mu$, $p_{\mathrm{T}} > 20$ GeV, $\lvert\eta\rvert < 2.4$, |
| | $m_{\mu\mu} = 91$ GeV $\pm$ 25 GeV |
| $Z \rightarrow ee$ | 2 OS el., $p_{\mathrm{T}} > 20$ GeV, $0 < \lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.47$, |
| | $m_{ee} = 91$ GeV $\pm$ 25 GeV |

Particle level jets are reconstructed from final state hadrons and their decay products, electrons and tau-leptons, excluding the dressed Z-decay products. The same algorithms and the same phase space requirements as in the detector level jet reconstruction, detailed in Sect. 6.2, are used, including the isolation requirement in $\Delta$R with respect to the leptons.

The unfolding procedure accounts for efficiencies, non-linearities and resolution effects of both the lepton and the jet reconstruction. It also accounts for the lepton triggers, averaged over the lepton phase space, as a function of the jet kinematics. By unfolding each jet observable separately, the possible variations in the lepton performance connected to the jet kinematics, to the extend to which they are modeled in Monte Carlo, are taken into account.

Biases are further reduced by correcting the simulated sample which is used to provide the unfolding correction by the extend to which discrepancies between data and simulated events are visible. Corrections are applied for reconstruction, identification and trigger efficiencies. The cross-check of the Monte Carlo description is performed first on the inclusive sample, looking for example at the data-Monte Carlo agreement in describing the $Z$ boson peak (shown previously in Fig. 6.2a).

The bin-by-bin unfolding is in general susceptible to a wrong modelling of the jet kinematics in the generator used for the unfolding, since it affects the extend to which migration towards larger jet $p_{\mathrm{T}}$ is favored. The bin-by-bin migration is reduced substantially by choosing the $p_{\mathrm{T}}$ binning such that the bin width corresponds to roughly twice the size of the jet $p_{\mathrm{T}}$ resolution expected for the center value of the bin. Further generator dependencies are introduced by the underlying-event tune, the fragmentation model and, to a minor extend, by the heavy-flavour fraction within the jets. The systematic uncertainty connected to the generator is estimated by comparing for each observable the unfolding corrections derived with full-simulated samples generated with ALPGEN and SHERPA (see Sect. 6.1).

As a second cross-check, the bin-by-bin unfolding in the electron channel is compared with the iterative Bayesian unfolding technique [56] implemented in the package RooUnfold [33]. All these checks have been performed using the ALPGEN $Z \rightarrow e^+e^- +$ jets sample, and the systematics estimated with toy-Monte Carlo.

For each observable considered, cross sections are computed from the number $N_{\mathrm{DATA}}$ of events selected in data according to the following prescription for the central value of a bin:

$$\frac{\Delta\sigma}{\Delta\alpha} = \frac{N_{\mathrm{DATA}} - N_{\mathrm{BKG}}}{\int \mathscr{L} dt} * U(\alpha) * \frac{1}{\Delta\alpha} \tag{6.3}$$

where $U(\alpha)$ is the bin-by-bin correction, which accounts for the unfolding from detector to particle level and $N_{\mathrm{BKG}}$ is the predicted number of background events. The normalization to the bin-width $\Delta\alpha$ is also applied.

The correction factors vary between 1.3 and 1.6 in the electron channel, and between 1.1 and 1.2 in the muon channel. The uncertainties range from 0.4 to 7%, increasing with higher jet multiplicity, higher $p_{\mathrm{T}}^{\mathrm{jet}}$ or higher rapidity.

The measured cross sections are compared at the particle level to predictions of the generators ALPGEN and SHERPA, both normalized to the respective NNLO cross sections, and to predictions by PYTHIA $Z \to \ell\ell$ + jets, normalized to the average $Z$ boson cross section in the electron and muon channel observed in data. The data is also compared to perturbative QCD predictions at NLO by BLACKHAT (more details in Aad et al. [17]).

We observe an overall agreement of the ALPGEN and SHERPA predictions with the results observed in data. The pQCD predictions of BLACKHAT describe well the inclusive and differential cross sections measured in data.

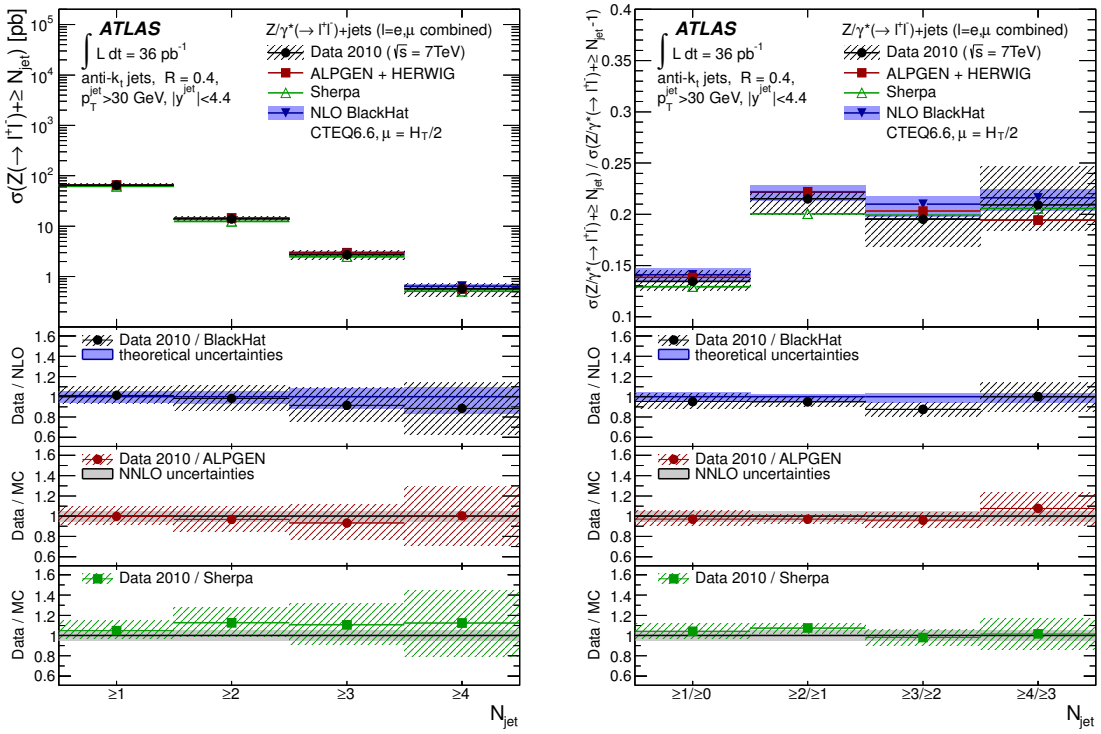
### 6.4.3 Combination of electron and muon channel

The measured cross section distributions for the $Z \to ee$+jets and $Z \to \mu\mu$+jets analyses are combined. The results present precise absolute jet cross section measurements.

The electron and muon measurements are performed in different fiducial regions for the rapidity of the leptons in the final state (see Table 6.4.2). In addition, the QED radiation effects are different in both channels. For each measured distribution, bin-by-bin correction factors, as extracted from ALPGEN $Z \to ee$+jets and $Z \to \mu\mu$+jets MC samples, are used to extrapolate the measurements to the region $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ for the leptons, where the lepton kinematics are defined at the decay vertex of the $Z$ boson. The increased acceptance in the lepton rapidities translates into about a 14% and a 5% increase of the measured cross sections in the electron and muon channels, respectively. The correction for QED effects increases the cross sections by about 2%. The uncertainties on the acceptance corrections are at the per mille level, as determined by using SHERPA instead of ALPGEN, and by considering different PDFs.

A $\chi^2$ test is performed for each observable to quantify the agreement between the electron and muon results before they are combined, where the statistical and uncorrelated uncertainties are taken into account. The statistical tests lead to probabilities larger than 60% for the electron and muon measurements to be compatible with each other, consistent with slightly conservative systematic uncertainties. The electron and muon results are then combined using the BLUE (Best Linear Unbiased Estimate) [89] method, which considers the correlations between the systematic uncertainties in the two channels.

The measurements are well described by the BLACKHAT NLO pQCD predictions, and by the predictions from ALPGEN and SHERPA. The corresponding $\chi^2$ tests relative to the different predictions, performed separately in each channel and for each observable, lead to $\chi^2$ per degree of freedom values in the range between 0.05 and 2.70. All correction factors and $\chi^2$ values are reported in Aad et al. [17].




(a) (b)

Figure 6.10: Particle level cross section results as function of jet multiplicity [17]. (a) shows the absolute cross section, (b) shows the cross section of each bin normalized to the value of the previous bin, showing the scaling of the cross section, expected from Eq. 2.10.

## 6.5 Outlook

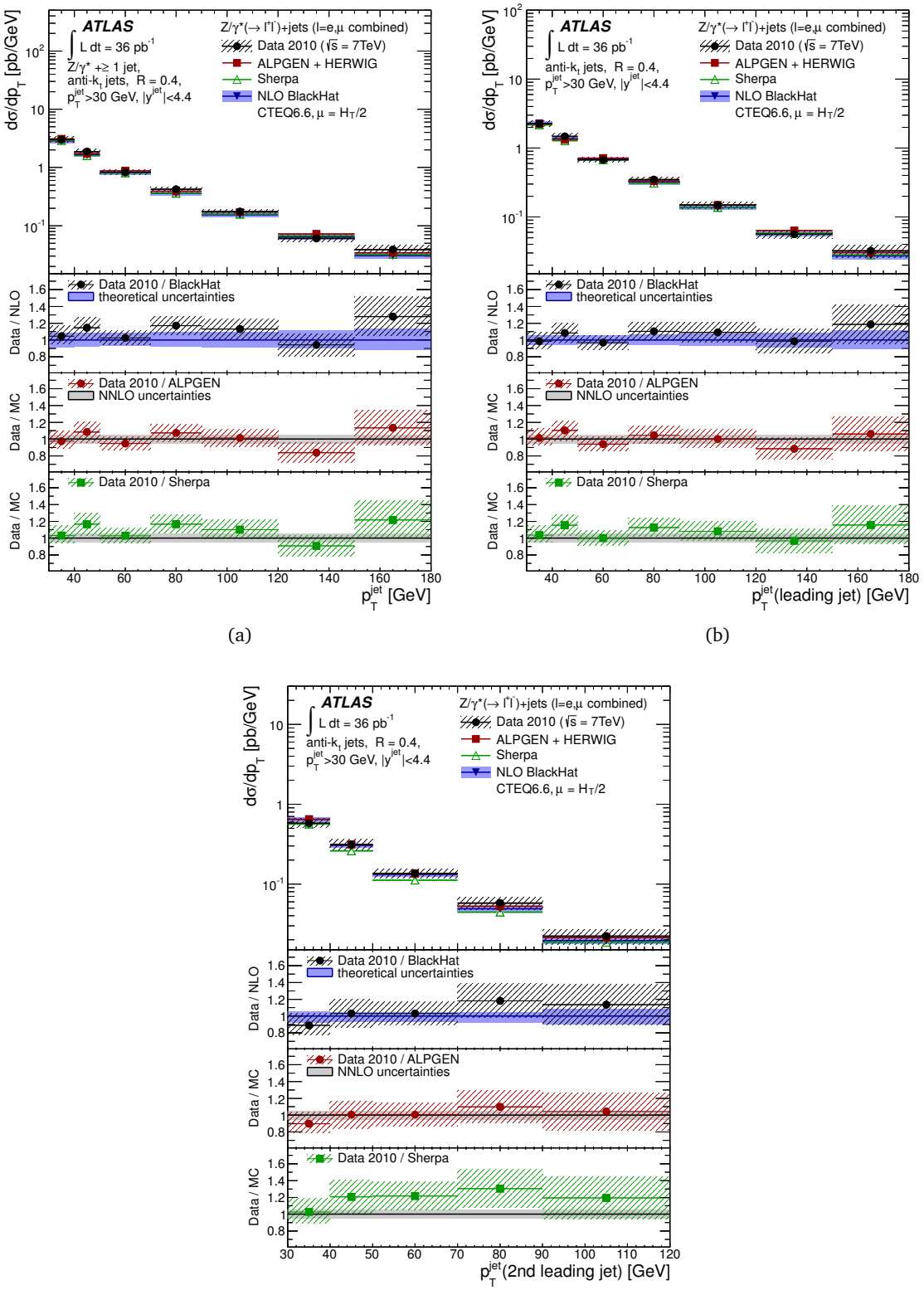
The measurement performed in 2010 has verified the precision of the current Monte Carlo predictions with respect to higher order corrections. The fact that both ALPGEN and SHERPA reproduce the data reasonably well ensures that those generators can be used in analysis where a more precise QCD prediction is needed, with respect to the predictions given by parton shower generators. However, the investigation of this process at high energies has only started, and will continue with the next runs.

The final goal of this effort would be to provide high-precision QCD measurements. Additional variables, that are sensitive also to the leading logarithms corrections at the different orders, will be considered in the future. The most recent NLO predictions include calculations up to 4 jets [79] in the final state. The LHC experiments have now a large statistics sample of high multiplicity events too, and the comparison between data and predictions should provide interesting results, if the experiments will be able to show that the experimental uncertainties are well under control. In particular, the largest uncertainty in the $Z \rightarrow \ell\ell$ + jets analysis is due to the jet energy scale: an improvement in the calibration, using the results of the *in-situ* methods, has been achieved in 2011, but is still not high enough to reach the level needed by a high-precision QCD measurement.

Another interesting improvement would be the unfolding of the measurement not only at particle level, but also at parton level. This would allow a better comparison with the theoretical predictions

and with the measurement performed at other experiments.

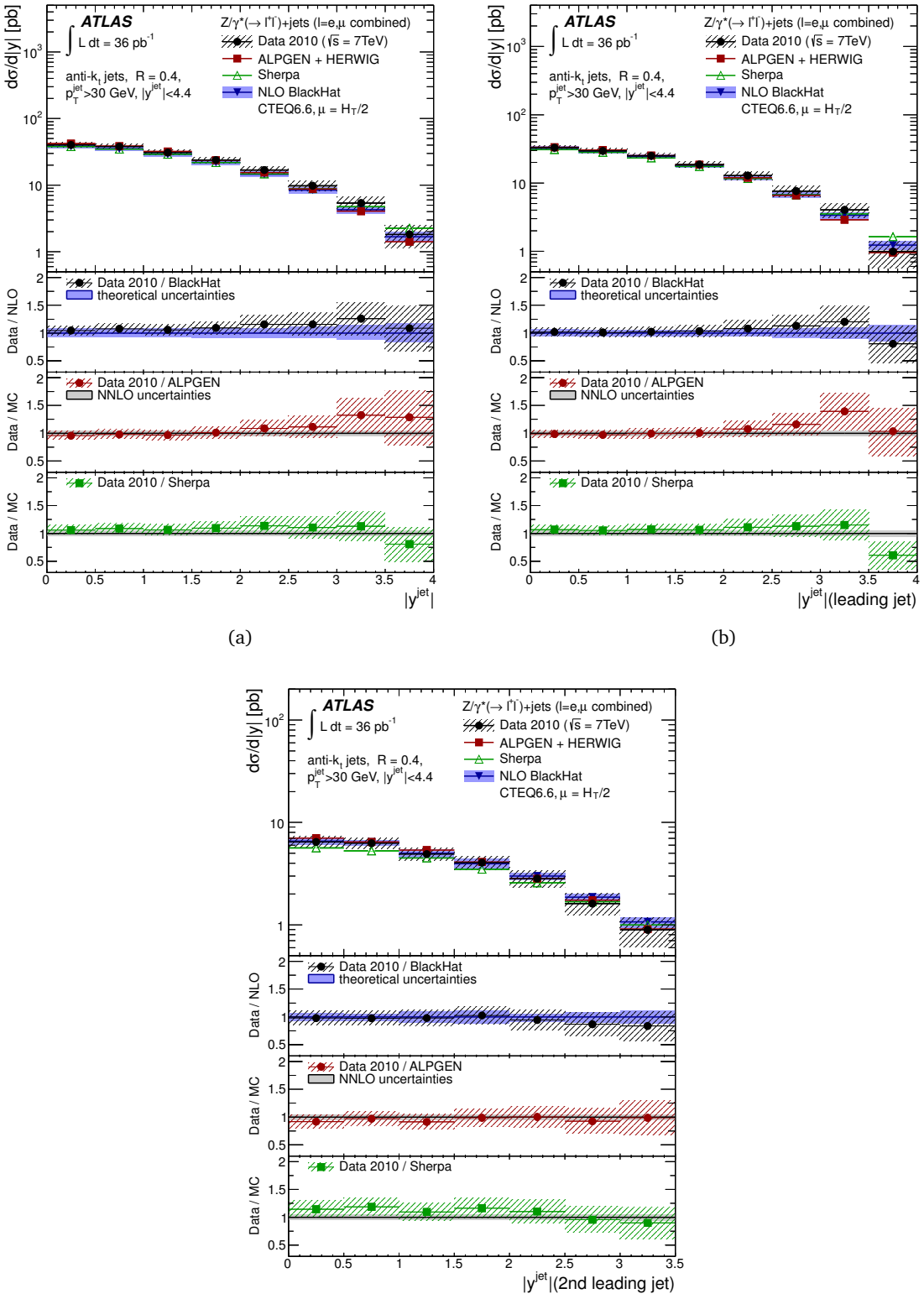
A comparison between the two general-purpose experiments at the LHC, ATLAS and CMS, would be very useful also for Monte Carlo tuning. At the moment, the measurements at the two experiments are only performed within the acceptance region, and use different generators for the data-Monte Carlo comparison. The first could be overcome by an extrapolation, and the latter, in principle, could be an advantage, since a combination or a comparison between the two experiments could shed light on the behaviour of different parton shower and underlying event generators in a similar environment. The comparison between the two experiments is however not straightforward. The physics objects are defined differently, as expected from two different detectors. In some cases (e.g. the leptons), the difference could be absorbed into the acceptance corrections. In contrast, the choice of a different size of the jet algorithm (anti$-K_T$ with parameter 0.4 for ATLAS, 0.5 for CMS [111]) can't be easily re-absorbed and makes the comparison harder. It is clear that all the quantities measured have a dependence on the algorithm definition, while in the pQCD regime, that is the final goal of a precision $Z \rightarrow \ell\ell$ + jets measurement, any such dependence should not be visible. Even defining a "pure" pQCD regime will be challenging: the analysis should be able to demonstrate that it is not sensitive to fragmentation, underlying event, and other effects that occur at lower energy scales. Reaching such a precision will not be matter of integrated luminosity, but it will need a joint effort of both the experimental and the theory community working on QCD studies at LHC.




(a) (b)

(c)

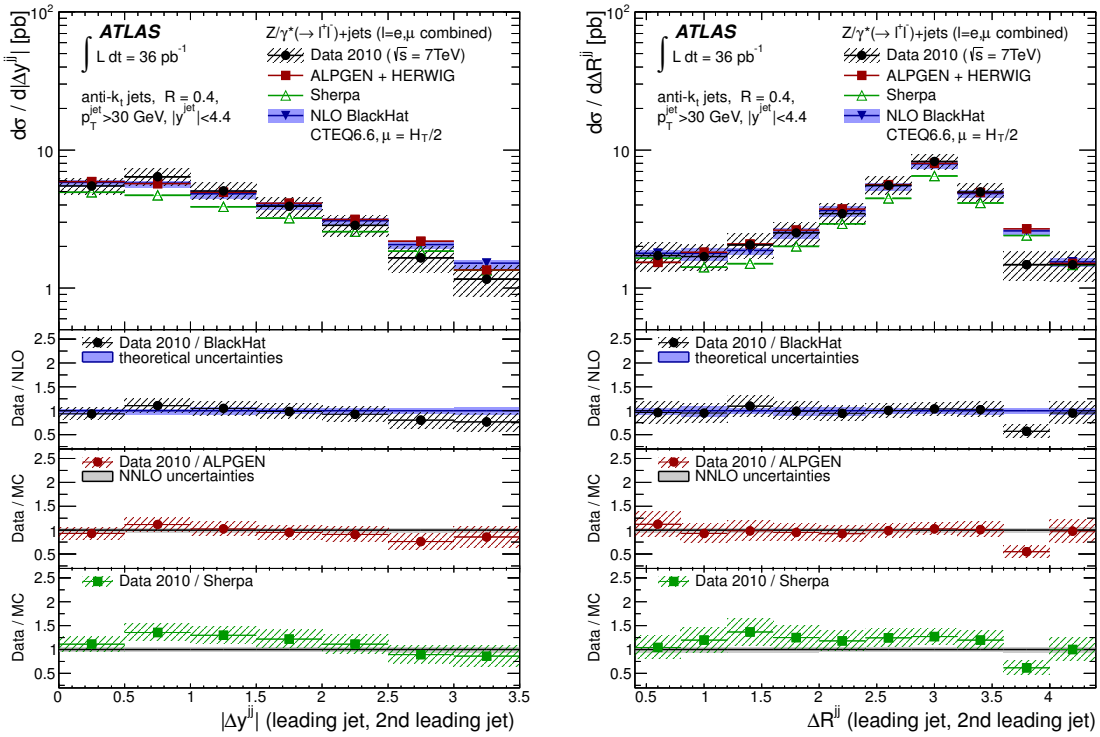
Figure 6.11: Particle level differential cross section results as function of (a) inclusive $p_{\mathrm{T}}$, (b) leading jet $p_{\mathrm{T}}$, and (c) 2nd leading jet $p_{\mathrm{T}}$ [17].




(a) (b)

(c)

Figure 6.12: Particle level differential cross section results as function of (a) inclusive $|y|$, (b) leading jet $y$, and (c) 2nd leading jet $y$ [17].




(a) (b)

Figure 6.13: Particle level cross section results as function of (a) jet $\Delta y$ and (b) jet $\Delta$R [17].

# Chapter 7

# Search for the Standard Model Higgs boson in the $\tau^+\tau^-$ channel

## 7.1 Strategy

In this chapter, the results of the searches for the Higgs boson performed in 2011 with the ATLAS detector in the $H \to \tau\tau$ channel, on a sample corresponding to an integrated luminosity of 4.7 fb$^{-1}$, are reported. I have contributed to the search in the final state with electrons and muons, the $H \to \tau\tau \to \ell\ell + 4\nu$ process.

The importance of this channel in the SM Higgs searches has already been highlighted in Sect. 2.4. $\tau$ leptons have a large branching fraction both in leptons and in hadrons, thus the possibility to detect this channel with good efficiency are high. The final state into electrons and muons is in particular very clean, and can benefit from the good ATLAS reconstruction performance for these physics objects. But, as mentioned in Sect. 2.4.2, several other effects play a role in the experimental search, and can make the extraction of the signal harder.

The dominant backgrounds for the $H \to \tau\tau \to \ell\ell + 4\nu$ analysis are:

- $Z \to \tau\tau$: this is the dominant irreducible background to the $H \to \tau\tau$ signal. The final state characteristics are the same, and the main visible difference between the two events will be depending on the characteristics of the production channels, different for $Z \to \tau\tau$ and $H \to \tau\tau$, and on the Higgs mass. Exploiting the first information requires knowing how the parton level final state can propagate to the detector level according to the production mechanism and additional effects, as hadronization (see Sect. 7.2). The detection of the Higgs mass peak on the $Z \to \tau\tau$ peak tail relies on the mass reconstruction and resolution. Techniques that allow a good mass reconstruction using the $E_{\mathrm{T}}^{\mathrm{miss}}$ information are used: in particular, we have used the *collinear approximation* (see Sect. 7.5). For this reason, in $Z \to \tau\tau$ the modelling of the $E_{\mathrm{T}}^{\mathrm{miss}}$ description is of fundamental importance.

- $Z \to \ell\ell$ $(e, \mu)$: the difference between this background and the $H \to \tau\tau$ signal is related to the $E_{\mathrm{T}}^{\mathrm{miss}}$, because true $E_{\mathrm{T}}^{\mathrm{miss}}$ is present only in $H \to \tau\tau$ events and not in $Z \to \ell\ell$ events. For this reason, any experimental effect on the $E_{\mathrm{T}}^{\mathrm{miss}}$ description will be crucial in the discrimination between this background and the signal. Concerning the description of the $Z \to \ell\ell$ kinematics, the modelling given by Monte Carlo is in agreement with the measurements shown in Chap. 6, and no issues are expected in this case.

- $W \to \ell\nu$ + jets: this background is important because it has a large cross section and its probability to pass the selection depends on the mis-identification rate. In these events one lepton comes from the leptonic decay of the $W$ boson, while the other one comes from a jet identified as a lepton, or from leptons produced in hadronic cascades (as caused by heavy

flavour production) but mis-identified as isolated leptons because of a bad reconstruction of the cascade. The mis-identification rate is difficult both to measure with high precision and to reproduce with Monte Carlo.

- top quark production: these events can provide a quite large source of background, because they can include both real jets and real leptons coming from the $W$ boson decay. The possibility to reject this background, however, are quite high because of the good performance of the b-taggers used in ATLAS.

Considering these large backgrounds, it is of fundamental importance to develop a strategy that allows a successful rejection of the background while keeping a high signal efficiency. The strategy used in the ATLAS analysis is based on two different goals, highlighted in the following.

**Higher discrimination between signal and background in events with jets:** the $H \to \tau\tau$ final state is characterised by a back-to-back configuration of the two $\tau$ leptons in the Higgs boson rest frame. The leptons originating in the leptonic decay are boosted in the direction of the decaying $\tau$ and have small opening angles with respect to such direction. If the Higgs boson is produced in absence of additional particles with high transverse momentum, the event is characterised by a small $E_{\mathrm{T}}^{\mathrm{miss}}$ in the laboratory rest frame. In this case, the signal to background ratio will be too small for a successful search, because of the large $Z \to \ell\ell$ background and the difficulties in distinguishing between the two topologies.

If in the event extra jets with high transverse momentum are produced, the Higgs boson acquires a boost in the transverse plane. This enhances the transverse momenta of the Higgs decay products and, as a consequence, the $E_{\mathrm{T}}^{\mathrm{miss}}$ of the event. This topology can be better distinguished from background processes, as shown by Mellado, Quayle and Wu [93]. However, previous analyses in ATLAS [29], simply requiring the presence of an additional jet in the event, did not show a very large improvement of the significance; a more careful investigation of the jet topology was needed to achieve the goal. This is the main aim of the investigation performed on 2011 data.

The presence of additional radiation, and its characteristics, depend on the production process of the Higgs boson, as shown in Sect. 7.2. Using the information on both the production and decay of the Higgs, we can optimize the selection and enhance the sensitivity of the analysis. The 2011 ATLAS analysis has decided to consider all production processes of the Higgs relevant for the low mass region and to combine the results in the limit, instead of focussing on a single production channel (as was done in preliminary studies, e.g. on simulations [9], and in Chap. 4). Exploiting the jet multiplicity, it is possible to divide the analysis in different jet bins, and to optimize the selection in each bin according to the dominant signal and background processes.

Using my previous experience with the $Z \to \ell\ell$ + jets analysis, I have investigated the description of several jet variables important for the Higgs analysis. $Z \to \ell\ell$ + jets events provide a good control sample, as the signal contamination is expected to be negligible. At the same time, I have also studied how to make the jet selection robust with respect to experimental effects, such as pile-up, that are difficult to reproduce in Monte Carlo and are highly affecting the 2011 run, as shown in Fig. 3.3. The results are reported in Sect. 7.4.1, and provide an important input to the definition of the event selection cuts used in this analysis, summarized in Sect. 7.4.

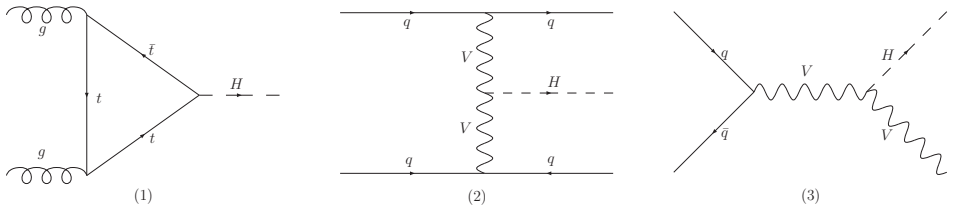
**$E_{\mathrm{T}}^{\mathrm{miss}}$ modelling in the background:** as stated already, the background description needs to be cross-checked in data. The most careful checks concern the $E_{\mathrm{T}}^{\mathrm{miss}}$ description, as the estimate of the invariant di-$\tau$ mass depends on its performance. A summary of the methods used for $Z \to \ell\ell$, $t\bar{t}$,

$W \to \ell\nu$ and backgrounds due to misreconstructed objects (*fake leptons* background) is reported in Sect. 7.6.

Estimating the $Z \to \tau\tau$ background requires special care. This channel is very similar to the signal, and it is very difficult to define a completely signal-free region where a data-driven estimate can be performed reliably. High precision is needed both in the normalization and in the shape of the background. To solve the issue, a data-driven technique that estimates the background using an "hybrid" method has been developped: the signal is simulated, while the environment is extracted from data in a signal-free region. This technique is called *embedding* and its validation will be reported in details in Sect. 7.6.2. I have focussed on the embedding validation in my work, as it complements the work already done with the $Z \to \ell\ell$ sample validating the description of the $Z$ boson background and the experimental environment for the Higgs search.

Sect. 7.2 and 7.6.2 explain in details how these issues have been treated in the analysis; the summary of the analysis is reported in Sect. 7.5, and the resulting $H \to \tau\tau$ cross section will be reported in Sect. 7.8. This analysis is included in the combined $H \to \tau\tau$ search, reported in Aad et al. [28].

## 7.2 Definition of analysis categories



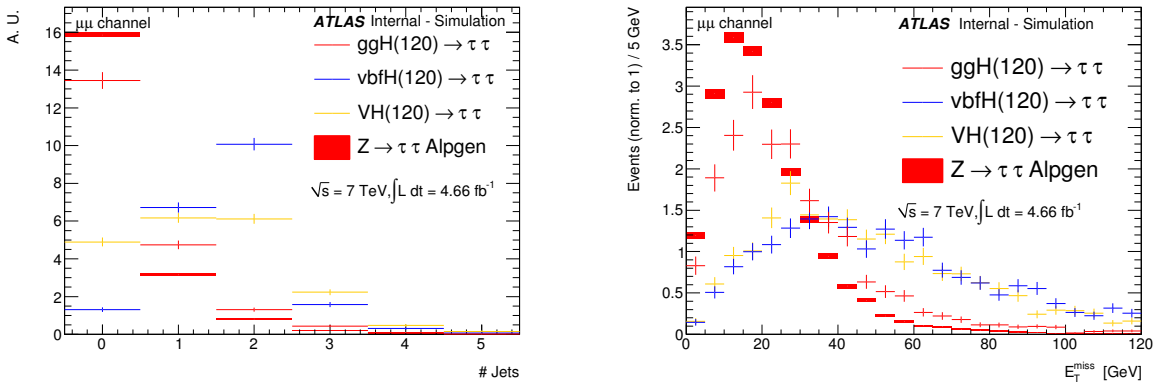

(a) $gg \to H$ (b) VBF (c) VH

Figure 7.1: Feynman diagrams at the leading order for the production of the Higgs boson at LHC.

### 7.2.1 Jet topology in Higgs production channels

The different production processes can lead to different jet multiplicities in the detector final state. The Feynman diagrams for the different processes considered in this analysis are shown in Fig. 7.1, and the cross sections for these processes at 7 TeV are reported in Sect. 2.4.1. In the $gg \to H$ process additional jets are produced only by QCD radiation, and this will lead to a characteristic jet multiplicity and $p_{\mathrm{T}}$ distribution in $gg \to H$ events. VBF processes, instead, will be dominant with respect to the other processes if two additional jets, widely separated in $\eta$, and a large boost of the Higgs boson are required. The analysis of the jet distributions is thus a very useful way in tagging Higgs candidates.

Fig. 7.2a shows the number of jets in events from different processes: $gg \to H$, VBF, VH, and $Z \to \tau\tau \to \ell\ell + 4\nu$ for comparison with background events. VBF and VH show a large jet multiplicity (together with a large boost of the Higgs system, as shown in Fig. 7.2b from the large $E_{\mathrm{T}}^{\mathrm{miss}}$ values), while $gg \to H$ is more similar to the background, with only a small fraction of events showing additional jets. $gg \to H$ is the dominant production process for events with 0 or 1 jet in the final




(a) Number of jets

(b) $E_{\mathrm{T}}^{\mathrm{miss}}$

Figure 7.2: Shape comparison between $H \to \tau\tau \to \ell\ell + 4\nu$ and $Z \to \tau\tau \to \ell\ell + 4\nu$, motivation for the categorization used in the analysis. All distributions are normalized to unit area. For details on the sample used, see Sect. 7.3.

state, and with relatively low $E_{\mathrm{T}}^{\mathrm{miss}}$; VBF and VH instead are the dominant production process for events with 2 jets, and show a large $E_{\mathrm{T}}^{\mathrm{miss}}$. Because of its high cross section, however, the gluon-fusion process produces a non-negligible number of 2-jet events. Taking this contribution into account is particularly important when computing the theory systematic uncertainties.

### 7.2.2 Perturbative uncertainties in jet categories

Higher-order effects and reconstruction inefficiencies, and their systematic uncertainties, must be taken into account to give a reliable prediction of the expected number of events as a function of the jet multiplicity. In Sect. 2.4, the main theoretical uncertainties related to the Higgs production cross section have been discussed: as mentioned, the differential distributions can be affected by large uncertainties, mostly in the case of $gg \to H$. This was taken into account in the definition of the categories used in the analysis.

When considered inclusively, the theoretical uncertainties on the $gg \to H$ production cross section are $\approx 20\%$[85]. When considering the different exclusive cross sections, as a function of the jet multiplicity, for example, the uncertainty increases up to $\approx 70\%$ in the case of $\geq 2$ jets events. A way to achieve a better prediction for the signal of interest was studied within the LHC Higgs Cross Section Working Group[86] and is briefly outlined in the following.

The selection of Higgs candidates according to the jet multiplicity corresponds to the following definition of an exclusive and inclusive contribution to the inclusive cross section:

$$\sigma_{\mathrm{total}} = \int_0^{p^{\mathrm{cut}}} dp \frac{d\sigma}{dp} + \int_{p^{\mathrm{cut}}} dp \frac{d\sigma}{dp} \equiv \sigma_0(p^{\mathrm{cut}}) + \sigma_{\geq 1}(p^{\mathrm{cut}}) \tag{7.1}$$

with $p$ being the variable used for the cut. In the analysis reported here, $p$ corresponds to the $p_{\mathrm{T}}^{\mathrm{jet}}$ and $\eta_{\mathrm{jet}}$ cuts used in the selection.

This selection causes a restriction of the phase space for the exclusive cross section ($\sigma_0$), that will change its perturbative structure with respect to the inclusive ones ($\sigma_{\mathrm{total}}$ and $\sigma_{\geq 1}$). For the 0-jet bin, this is due to the restriction that the selection cuts on the allowed additional radiation

applies to the collinear initial-state radiation as well as the overall soft radiation in the event. This restriction causes large Sudakov-logarithms to appear: the cancellation that occurs in the inclusive cross sections, which ensures the convergence of the series (as for example in the case treated in Sect. 2.3), does not happen any longer. Similar effects play a role in any definition of an exclusive jet cross section, independently from the number of jets considered.

If we try then to estimate the uncertainty on the cross section due to the missing higher order diagrams via scale variations (varying the renormalization and factorization scale in the fixed order prediction, as it is commonly done for QCD processes), we might not get meaningful results due to the influence of these logarithms on the calculations. This issue affects strongly the $gg \to H$ production channel, while, for example, has a negligible effect on the $Z \to \ell\ell$ + jets process.

Because quantities are well defined in an inclusive cross section, the solution of the problem lies in the possibility of using the uncertainties on $\sigma_{\text{total}}$ and $\sigma_{\geq 1}$ (that can be estimated by scale variations) to extract the uncertainty on $\sigma_0$. It can be shown that, in the simple example mentioned, the uncertainty on $\sigma_0$ can be estimated by error propagation from the uncertainty on the two inclusive cross sections.

This simple example can easily be extended to additional jet-bins and cuts, provided that:

1. the inclusive cross sections are known up to the order needed to the calculation and are all treated consistently;
2. the cuts used divide the phase space in mutually exclusive regions.

While the second item is easily satisfied in the analysis, the first one is challenging. In fact, for a $gg \to H$ analysis:

- $\sigma_{\text{total}}$ is known at NNLO;
- $\sigma_{\geq N}$ is known at $N^{2-\text{N}}LO$;
- because NNLO corrections are large, there will be a non-negligible ambiguity in the definition of the exclusive cross section rates: predictions show large differences depending if they are defined from $\sigma_{\geq N}$ or $\sigma_{\geq N-1}$.

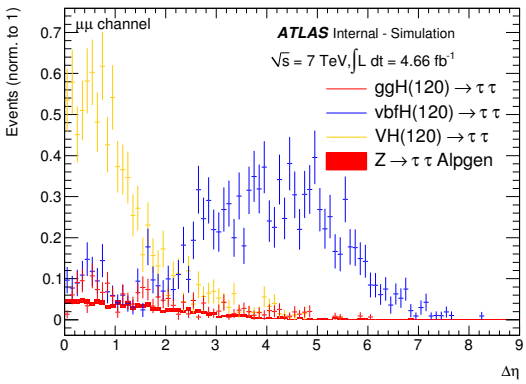
From these considerations, it is clear where the large increase in the uncertainty of the $gg \to H$ cross section comes from: if $\{\sigma_0,\sigma_1,\sigma_{\geq 2}\}$ are the different jet categories, the uncertainty on $\sigma_{\geq 2}$, available only at LO, will largely affect the result, as well as introducing a dangerous source of ambiguity in the definition. All these effects are related to the poor convergence of the perturbative series. To avoid those large uncertainties and improve the discrimination of the signal, additional selection cuts as rapidity gaps and invariant mass cuts have been used to enhance the contribution of VBF and VH production processes.

Exploiting VBF and VH topologies, which are predicted with a better precision in those jet bins, increases the sensitivity to the signal. The uncertainties available for the H+2 jets cross section are available at NLO (see Sect. 7.3). $gg \to H$ events will still be present in those jet bins, but they will not dominate the signal distribution. The gluon fusion uncertainty will be used in treating the correlation with the other exclusive bins, as the definition of the 1-jet bin and of the 2-jet bins are anti-correlated. But if 2-jet bins are well defined in the phase space via additional cuts, and all events that don't pass those cuts are recovered in the 1-jet bin, no ambiguity is left in the definition of the cross sections, and the uncertainty can be treated consistently using the $gg \to H$ uncertainties at NLO.

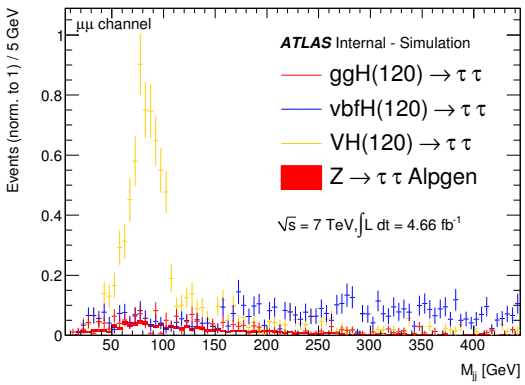
### 7.2.3 Definition of jet categories in the 2011 analysis

According to the theoretical reasons explained above, these are the categories designed for the analysis:

- The first two categories, called *2-jet VBF* and *2-jet VH*, require the presence of 2 jets in two different kinematic regions: those cuts are optimized for the VBF and VH final state respectively. Fig. 7.3a shows the pseudorapidity difference between the leading jets, a good discriminating variable for VBF events. Fig. 7.3b instead shows the main selection criterion of the VH category, related to the decay of the additional vector boson in the event, in two quarks. In the first case, the backgrounds (including $gg \to H$) have a different behaviour from VBF, while in the second case are constant in the region of the ZH peak. The optimization of the selection cuts for these two processes results in an improved signal-to-background ratio. The VBF category is the most sensitive one; in 2011 the statistical uncertainties, however, still limit its power.




(a) $\Delta\eta$ between the two leading jets in $H \to \tau\tau \to \ell\ell + 4\nu$ events and in $Z \to \tau\tau \to \ell\ell + 4\nu$ background




(b) Dijet mass of the system of the two leading jets in $H \to \tau\tau \to \ell\ell + 4\nu$ events and in $Z \to \tau\tau \to \ell\ell + 4\nu$ background

Figure 7.3: Comparison between $H \to \tau\tau \to \ell\ell + 4\nu$ and $Z \to \tau\tau \to \ell\ell + 4\nu$, motivation for the categorization used in the analysis. For details on the sample used, see Sect. 7.3. Private plots.

- The *1-jet* category requires the presence of at least one high $p_{\mathrm{T}}$ jet and the main signal is $gg \to H$, with the Higgs boost allowing a good discrimination between the signal and the irreducible $Z \to \ell\ell$ background. As mentioned earlier, it is mutually exclusive with the 2-jet categories, i.e. events that fail the 2-jet VBF or VH selection are recovered in the 1-jet category and passed through the 1-jet selection cuts.
- The *0-jet* category uses a more inclusive selection to collect the signal event topologies not selected by the other categories, and tend to collect mostly events where the Higgs has little or no boost at all. Only the $H \to \tau\tau \to e\mu + 4\nu$ decay is considered, to avoid the contamination from the dominant $Z \to \ell\ell$ ($e^+e^-, \mu^+\mu^-$) background in the signal region.

To keep a high signal efficiency in the event selection, the categories are defined as $N_{\mathrm{jet}} \geq 0, 1, 2$. This definition is analogous to the one used in Sect. 7.2.2. The orthogonality requirement is ensured via a set of kinematic cuts (see Sect. 7.5 and Table 7.2). The categories fulfil the theoretical

requirement highlighted earlier, thus ensure a better control of the prediction, and provide a better sensitivity with respect to an inclusive analysis.

## 7.3 Data and simulated samples

**Data** The results are based on proton-proton collision data collected with ATLAS in 2011, at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. Only data taken with all sub-systems relevant to this analysis operational are used. This results in an integrated luminosity of 4.7 fb$^{-1}$ for the full 2011 data sample.

**Higgs boson production:** The cross-sections for Higgs boson production are calculated following the prescriptions of the LHC Higgs cross-section working group [86]. For the gluon fusion process, it has been calculated using HIGLU [107] and ggh@nnlo [72] at next-to-next to leading order (NNLO). For the vector boson fusion, the NLO calculation was performed with VBF@NLO [44, 86]. For the associated production with vector bosons, the cross-section calculation is performed at NNLO QCD + NLO electro-weak (EW) scale [48, 53].

The SM Higgs boson production via gluon fusion is simulated with POWHEG [97] and the $gg \to H$ events were re-weighted as a function of the Higgs transverse momentum as prescribed by the LHC Higgs cross-section working group [86]. The VBF production of the Higgs boson was also simulated with POWHEG, while the associated production with vector boson was simulated with PYTHIA [105].

**$Z \to \ell\ell$ + jets and $Z \to \tau\tau \to \ell\ell + 4\nu$ production:** The cross-section for $Z \to \ell\ell$ + jets is calculated in NNLO QCD using the FEWZ code [64]. The $Z \to \ell\ell$ background is is simulated with the ALPGEN [90] generator, using the same settings used in Chap. 6. This background is also checked on data, in particular for checking the normalization and the modelling of $E_{\mathrm{T}}^{\mathrm{miss}}$ along the tails.

A data-driven approach is instead used to determine the $Z \to \tau\tau \to \ell\ell + 4\nu$ contribution, as explained in Sect. 7.6.2. The validation is performed against the relevant ALPGEN sample.

**$W \to \ell\nu$ + jets production:** The cross-section is calculated in NNLO QCD using the FEWZ code. However, because this background is dominated by mis-identification effects, known with little accuracy, a data-driven approach is used to determine the contribution of this background. It is included in the so-called *Fake leptons* background.

**Di-boson production:** The electroweak production of pairs of vector bosons ($WW$, $WZ$, $ZZ$) can lead to final states with two or more charged leptons from the leptonic decays of the $W$ and $Z$ bosons. The cross-section is calculated at NLO with MC@NLO[63]. The di-boson ($WW$, $WZ$, $ZZ$) production processes are generated with HERWIG [54].

**$t\bar{t}$ production:** The cross-section is calculated in NLO+Next-to-Next-to-Leading Logarithmic (NNLL) QCD. The $t\bar{t}$ process is generated with MC@NLO, and a cross-check of the normalization is done on data, looking at a $t\bar{t}$ enriched control region.

**Single-$t$ production:** The production of single top quarks via $t$ or $s$-channel production or in association with a $W$ boson contributes to the background if one $W$ boson decays leptonically and one of the leptons is either due to a mis-identified hadronic jet or, for $Wt$ production, comes from the decay of the second $W$ boson. The cross-section is calculated in NLO+NNLL QCD. The single top quark processes are generated with AcerMC[81].

**Multi-jet processes:** Multi-jet processes have large production rates at the LHC and the mis-identification of jets as leptons has large uncertainties. For these reasons, a data-driven approach is used to determine its contribution. It is included in the Fake leptons background.

The cross sections used in this analysis for the above processes are summarised in Table 7.1. Relevant branching ratios (BR) are taken into account, as written in the table. The following corrections are used in the cross-section calculations: NNLO for $W/Z$+jets, NLO + Next-to-Next-to-Leading Logarithmic (NNLL) for $t\bar{t}$ and single-$t$, NLO for di-boson, and NNLO for $gg \to H$ ($\to \tau^+\tau^-$).

For all Mc@Nlo and Alpgen samples, parton showers and hadronisation are simulated with Herwig and the activity of the underlying event with Jimmy[49], while for AcerMC and Powheg the hadronisation is simulated with Pythia.

The programs Tauola [67, 80] and Photos[68] are used to model respectively the decay of $\tau$ leptons and additional photon radiation in decays produced in Pythia, Mc@Nlo and Herwig.

As mentioned earlier, the effects of pile-up in data are reproduced by the simulation. Before event reconstruction the Monte Carlo events have been overlaid with minimum-bias simulated events reproducing the bunch train structure and spacing of the LHC beams (50 ns bunch spacing). Residual differences between data and Monte Carlo simulation have been corrected by re-weighting the Monte Carlo events to reproduce the distribution of the number of interactions per luminosity block observed in data (reported in Fig. 3.3). This procedure reproduces by construction the expected "in-time" and "out-of-time" effects in data. Tests on the number of primary vertices per events as well as on calorimeter variables in simulation after the re-weighting are performed.

| Process | Cross-section (pb) [× BR] |
|---|---|
| SM $gg \to H$ ($\to \tau^+\tau^-$), $m_H = 120$ GeV | 1.18 |
| SM VBF $H$ ($\to \tau^+\tau^-$), $m_H = 120$ GeV | $9.0 \times 10^{-2}$ |
| SM $WH$ ($\to \tau^+\tau^-$), $m_H = 120$ GeV | $4.7 \times 10^{-2}$ |
| SM $ZH$ ($\to \tau^+\tau^-$), $m_H = 120$ GeV | $2.6 \times 10^{-2}$ |
| $W(\to \ell\nu)$+jets ($\ell = e, \mu$ or $\tau$) | $31.5 \times 10^3$ |
| $Z/\gamma^* \to \ell\ell$+jets ($m_{\ell\ell}$ >10 GeV, $\ell = e, \mu$ or $\tau$) | $15.0\times 10^3$ |
| $t\bar{t}$ | 167 |
| Single top quark $t, s$ and $Wt$-channels | 64.6, 4.6, 15.7 |
| $WW$, $WZ$ and $ZZ$ (66 GeV < $M_Z$ < 116 GeV) | 44.9, 18.0, 5.6 |

Table 7.1: Cross-sections used in the analysis.

## 7.4 Selection and reconstruction of physics objects

**Electrons:** Electron candidates are formed from an energy deposit in the electromagnetic calorimeter and associated to a track measured in the inner detector. They are selected if they have a transverse energy $E_{\mathrm{T}} > 15$ GeV, lie within $|\eta| < 2.47$ but outside the transition region between the barrel and end-cap calorimeters ($1.37 < |\eta| < 1.52$), and meet quality requirements based on the expected shower shape [8].

**Muons:** Muon candidates are formed from a track measured in the inner detector and linked to a track in the muon spectrometer [14]. They are required to have a transverse momentum $p_{\mathrm{T}} > 10$ GeV and to lie within $|\eta| < 2.5$. Additionally, the difference between the $z$-position of the point of closest approach of the muon inner detector track to the beam-line and the $z$-coordinate of the primary vertex is required to be less than 1 cm. This requirement reduces the contamination due to cosmic ray muons and beam-induced backgrounds. Muon quality criteria based on, e.g., inner detector hit requirements are applied in order to achieve a precise measurement of the muon momentum and reduce the misidentification rate. The preselection requirements are very similar to those used in the $Z \rightarrow \ell\ell$ + jets analysis; only some cuts have been relaxed to improve the efficiency of the preselection.

**Isolation:** Identified electrons and muons are required to be isolated: the additional transverse energy in the electromagnetic and hadronic calorimeters must be less than 8% (4%) of the electron transverse energy (muon transverse momentum) in a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2$ around the electron (muon) direction. The sum of the transverse momenta of all tracks with $p_{\mathrm{T}}$ above 1 GeV located within a cone of radius $\Delta R = 0.4$ around the electron (muon) direction and originating from the same primary vertex must be less than 6% of the electron transverse energy (muon transverse momentum).

**Jets:** Jets are reconstructed using the anti$-K_T$ algorithm with a distance parameter value of $R = 0.4$, taking topoclusters as input. Reconstructed jets with $p_{\mathrm{T}} > 20$ GeV and within $|\eta| < 4.5$ are selected. Events are discarded if a jet is associated with out-of-time activity or calorimeter noise. The jets are calibrated using calibration constants derived from the combination of the data/Monte Carlo comparison of the *in-situ* calibrations.

**Jet vertex fraction:** After having associated tracks to jets by requiring $\Delta R < 0.4$ between tracks and the jet direction, a jet-vertex fraction (JVF) is computed for each jet as the scalar $p_{\mathrm{T}}$ sum of all associated tracks from the primary vertex divided by the scalar $p_{\mathrm{T}}$ sum of all tracks associated with the jet. Conventionally, JVF $= -1$ is assigned to jets with no associated tracks. Jets with $\left|\eta\right| < 2.4$ are required to have |JVF| $> 0.75$ in order to suppress pileup contributions. The reasons behind this selection criterion are reported in Sect. 7.4.1.

**B-Jets rejection:** In the pseudorapidity range $\left|\eta\right| < 2.5$, $b$-jets are identified using a tagging algorithm based on the discrimination power of the impact parameter information and of the reconstruction of the displaced vertices of the hadron decays inside the jets [7]. The $b$-tagging algorithm has an average efficiency of 58% for $b$-jets in $t\bar{t}$ events [15]. The corresponding light-quark jet misidentification probability is 0.1–0.5%, depending on the jet $p_{\mathrm{T}}$ and $\eta$ [13].

**Overlap removal:** When different objects selected according to the above criteria overlap with each other geometrically (within $\Delta R < 0.2$), only one of them is considered for further analysis. The overlap is resolved by selecting muon, electron, and jet candidates in this order of priority.

**Missing transverse momentum:** The magnitude of the $E_{\mathrm{T}}^{\mathrm{miss}}$ [22] is reconstructed including contributions from muon tracks and energy deposits in the calorimeters. Calorimeter cells belonging to three-dimensional noise-suppressed clusters are used and they are calibrated taking into account the reconstructed physics object to which they belong.

### 7.4.1 Validation of the preselection in $Z \to \ell\ell$ + jets events

The 2011 run shows different conditions with respect to the 2010 run, due to a general increase in the intensity of the beam and in the number of bunches circulating in the beams. This results in an overall increase of the level of pile-up. I have carried out an investigation of the experimental conditions in $Z \to \mu^+\mu^-$ + jets events, to understand how pile-up effects might impact the analysis.

Additional events in the same bunch crossing will produce additional primary vertices; in case of a $Z \to \ell\ell$ + jets event, it will be important to select two leptons coming from the same vertex, in order to avoid a mis-identification of the signal. This can be done with a high precision with the ATLAS tracking system, also under high pile-up. The association of tracks with vertices allows already a very good rejection of pile-up events.

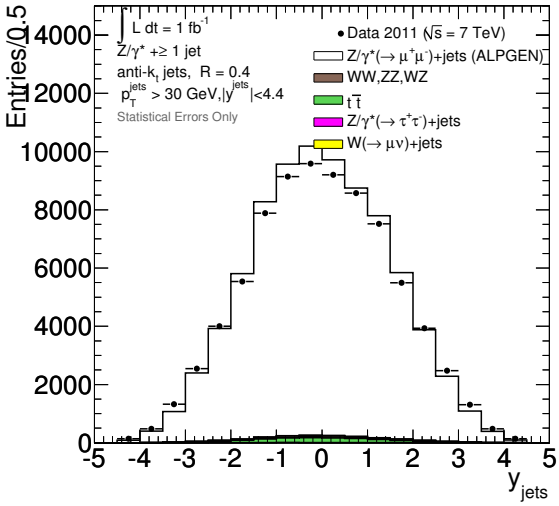
In the calorimeters, the large integration time of the signal might lead to the detection and association of a signal with the wrong bunch crossing. This means that both in-time pile-up and out-of-time pile-up can be observed (see Sect. 3.1.1). The signal shape and fitting is designed to avoid as much as possible this situation, using a bipolar shape (with null area) and a large number of sampling along the analog signal to permit a precise association; in general, the effects on reconstructed objects are small.

Only a part of the 2011 dataset is considered in this study, corresponding to an integrated luminosity of about 1 fb$^{-1}$. The Monte Carlo samples are the same used in the $Z \to \ell\ell$ + jets 2010 analysis (see Sect. 6.1), but simulated according to the 2011 environment. The reconstruction and calibration used in the Monte Carlo samples are the same used for the first 2011 data.The jets in this study are still calibrated with the EM+JES calibration, and no offset corrections are applied[1].

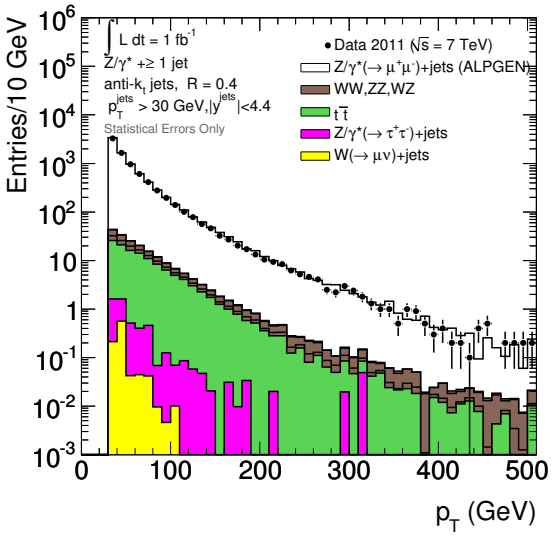
Jets are reconstructed starting from topoclusters seeds; with higher pile-up, a higher number of seeds is expected, and as the jet reconstruction relies only on calorimeter signals, they will be reconstructed regardless of their origin, if they are from a primary vertex or from pile-up.

The $p_{\mathrm{T}}^{\mathrm{jet}}$ distributions (shown in Fig. 7.4b for all jets, and in 7.5a, 7.5b and 7.5c for the first, second and third leading jet respectively) show an agreement between data and simulations consistent with what was shown already in 2010. As the calibration is very similar to the one used in 2010, the jet energy scale uncertainty is the same: for $p_{\mathrm{T}}^{\mathrm{jet}} = 30$ GeV, the jet energy scale uncertainty (without taking into account pile-up effects) is of the order of 3%. Such uncertainty covers most of the differences observed between data and Monte Carlo in the $p_{\mathrm{T}}^{\mathrm{jet}}$ distribution.

[1] The average increase of the energy in the calorimeter, due to pile-up, is treated in 2011 using new offset corrections, derived on a simulated sample. The usage of a Monte Carlo to derive the corrections is the main difference with respect to the 2010 offset correction. The correction computed on simulation is validated against the truth jet energy in Monte Carlo, and with the response of tracks or to single particles in data. These corrections perform very well in removing the dependence of the detected jet energy on the number of collisions in the event, but as the validation required a long time, were used only in a later stage of the 2011 analyses. For this reason this correction is not implemented in the analysis shown in this section, but is included in the Higgs analysis.




(a) Detector level inclusive $y$ distribution in 2011.



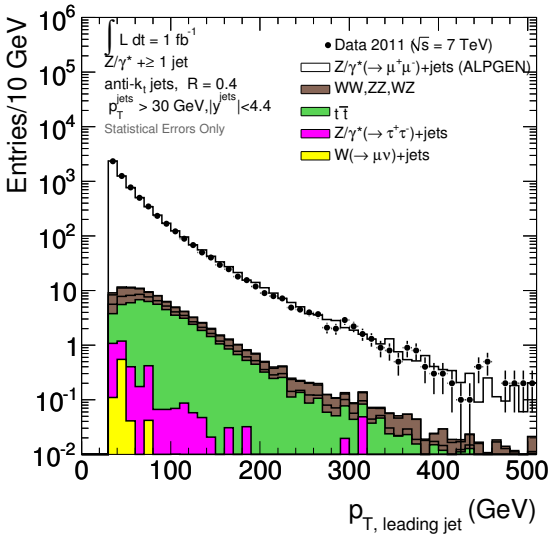

(b) Detector level inclusive $p_{\mathrm{T}}$ distribution in 2011.

Figure 7.4: Jet distributions from the 2011 analysis. Only statistical uncertainties are shown.

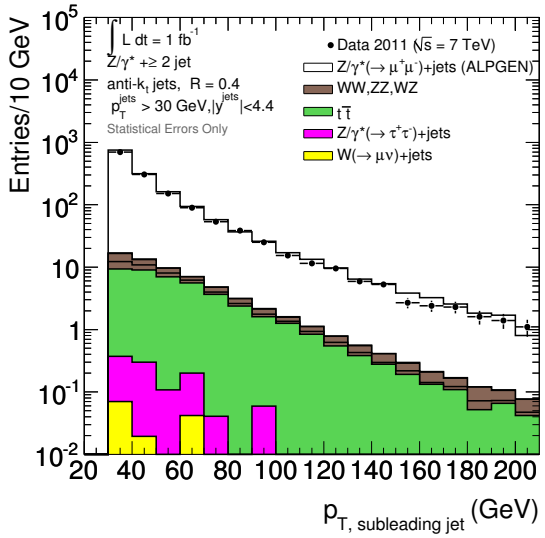
The difference between data and Monte Carlo tends to be larger in the subleading jets distributions, and at low $p_{\mathrm{T}}^{\mathrm{jet}}$, as shown in Fig. 7.5. When looking at the inclusive rapidity distribution, shown in Fig 7.4a, the disagreement in the central region looks larger with respect to what is observed in the 2010 studies. The correlation of this effect with the presence of pile-up jets can be understood by checking the behaviour of the charged particles contributing to the jets in the event.

Using tracks, it is possible to associate a jet measured in the calorimeter with the tracks of the charged particles that contribute to its energy[2], and estimate how much of the $p_{\mathrm{T,track}}$ observed

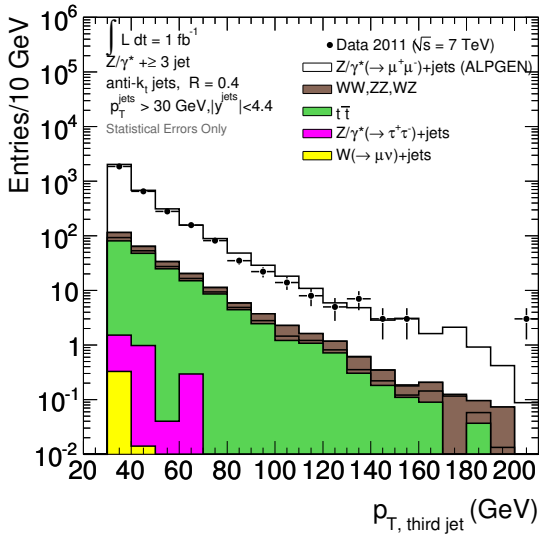
[2]This is done using a central ATLAS tool, named "Jet vertex association". Additional informations in Ruwiedel, Wermes and Ströher [101] and Miller, Dong and Schwartzman [94]. The tracks considered must fulfil quality requirement on $p_{\mathrm{T}}$, the number of hits in the inner detector, and the perigee parameters. They are matched with the jet, requiring a




(a) Detector level leading jet $p_T$ distribution in 2011.




(b) Detector level sub-leading jet $p_T$ distribution in 2011.




(c) Detector level third-leading jet $p_T$ distribution in 2011.

Figure 7.5: Jet distributions from the 2011 analysis. Only statistical uncertainties are shown.

is due to tracks coming from the primary vertex or from tracks coming from other vertices. This information obviously exists only for the central region of the detector, as a reliable detection of the tracks has to be done within the tracking acceptance. This quantity is named *Jet vertex fraction* (JVF) and is defined as:

$$\mathrm{JVF} = \frac{\sum_{\mathrm{associated}} p_{T,\mathrm{track}}}{\sum_{\mathrm{all}} p_{T,\mathrm{track}}} \tag{7.2}$$

radial distance of less than 0.4 between the track and the jet axis.

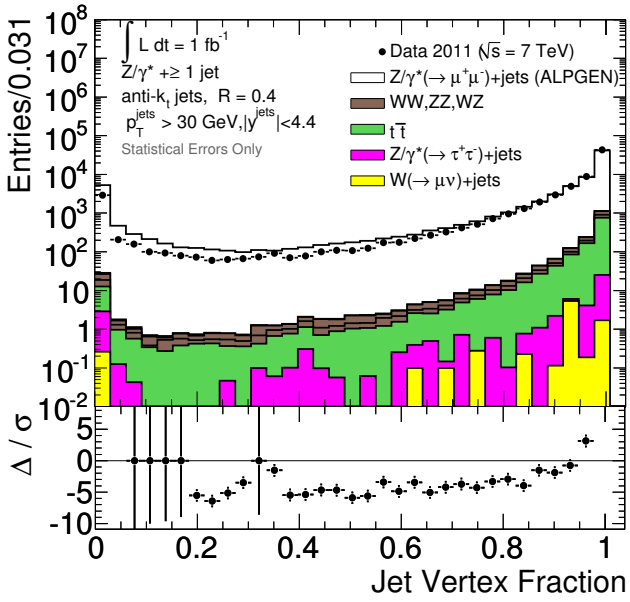



Figure 7.6: Detector level Jet-Vertex-Fraction (JVF) distribution in 2011. Only statistical uncertainties are shown.

The distribution in 2011 data of the JVF is shown in Fig. 7.6.

The JVF is usually computed with respect to the primary vertex of the event. From Eq. 7.2, it is clear that jets made of particles coming from the primary vertex will tend to have a JVF close to one, while lower JVF values mean that the number of tracks contributing to the jet from the other vertices is very high, and their contribution to the total $p_{\mathrm{T}}$ is higher than the one of the particles coming from the primary vertex. As the method relies on tracking, two things must be taken into account:

- the neutral particle contribution is not considered
- the JVF is limited to the tracking acceptance, $|\eta| < 2.5$
- jets with $\eta_{\mathrm{jet}} > 2.5$ and jets within the tracking acceptance but that could not be associated to any good quality track automatically get JVF $= -1$ (not shown in Fig. 7.6).

Even considering these limitations, the JVF is a useful quantity to understand the effects of pile-up in the central region of the detector. Fig. 7.4a shows a disagreement between data and the Monte Carlo predictions that is not fully consistent with the one in the $p_{\mathrm{T}}$ distributions showed in Fig. 7.5a, 7.5b and 7.5c. Looking at the JVF distribution in Fig. 7.6, the difference seems to be dominating in the low JVF distribution, where pile-up effects play a very large role. This means that the description of pile-up in 2011 is not completely reproducing the data; but, as those effects seems to not to show up much in the $p_{\mathrm{T}}$ distribution of the leading jets of the event, this means that the badly modeled population is mostly composed of low-$p_{\mathrm{T}}$ jets.

I checked this hypothesis in data and developed a strategy for improving the selection using the JVF variable. A cut of |JVF| $> 0.75$ was added to the selection cuts, chosen in a region where the JVF is well described and the difference between data and Monte Carlo is roughly flat, such

that the uncertainties on the estimation of the JVF would have not impacted the results. All jets with JVF$= -1$ are kept. The rapidity distribution of the jets passing the JVF cut (in events taken from period G-H only, higher pile-up conditions wrt 2010 data) is shown in Fig. 7.7, and shows an improved agreement with respect to Fig. 7.4a. A jet population composed of only high-JVF jets is better described by the simulation.

Because a JVF cut makes the measured jet energy more stable against the number of interactions per bunch crossing, it opens the possibility to reduce the $p_{\mathrm{T}}^{\mathrm{jet}}$ cut in the analysis. The argument is that the low-$p_{\mathrm{T}}$ jets that survive a JVF cut are less influenced by pile-up effects. After the JVF selection also the uncertainty related to the energy scale is under control with respect to pile-up effects. Considering $Z \to \ell\ell$ + jets events in the central region ($|y| < 2.8$), the number of jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 25$ GeV is better described for jets with JVF $> 0.75$ as shown in Fig. 7.8.

I have also checked that the rapidity distributions of jets in events with high number of vertices becomes compatible with the same distribution in events with a low number of vertices. For all these reasons, the JVF cut I proposed and validated in this study has also been used in many analyses with jets in 2011.

The most significant effect of a JVF cut, however, is concerning the application of a jet veto. As mentioned already in Chap. 4, a jet veto could be compromised by the presence of additional jet activity, not related to the hard scattering event, in the central region. In the case studied here, the problem is not the jet reconstruction itself, but in the selection: in the case of a background rejection strategy that considers a jet veto, also signal events with central jets originated only from pile-up vertices would be rejected. I have investigated the $Z^*$ distribution, whose definition has been provided in Chap. 4, to estimate the impact of the JVF cut on a central jet veto. The distribution in $Z \to \mu^+\mu^-$ + jets events for jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 25$ GeV and $|y_{\text{leading jet}} - y_{\text{subleading jet}}| > 2$ is shown in Fig. 7.9. The small rapidity gap used is chosen to optimise the amount of statistics available. The better description of the $Z^*$ peak, shown in Fig. 7.9b, shows how the JVF can significantly improve

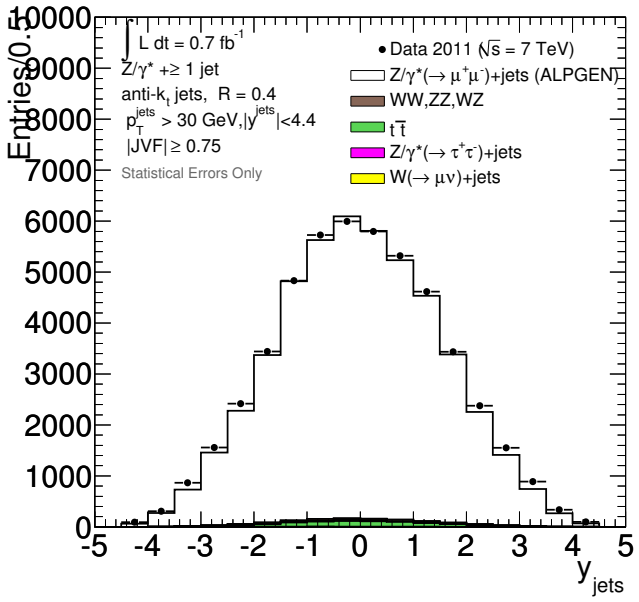
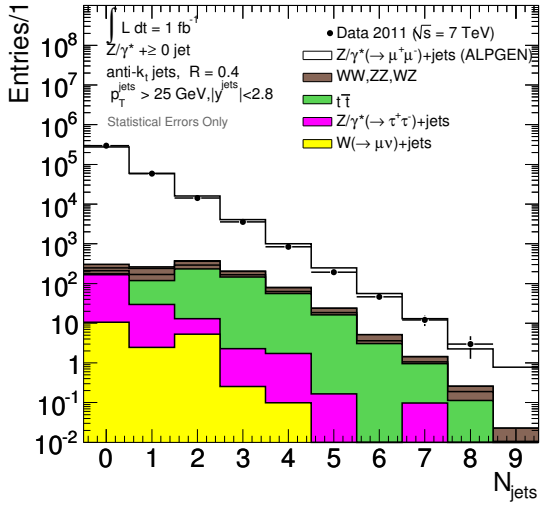
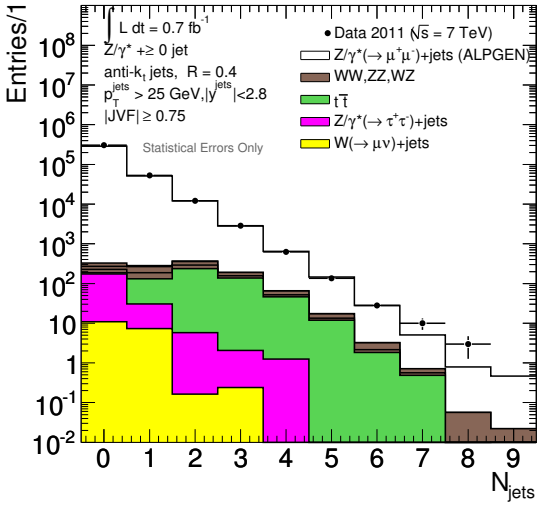



Figure 7.7: Detector level inclusive rapidity distribution in period G-H in 2011. Only jets passing a cut JVF $> 0.75$ are considered. Only statistical uncertainties are shown.

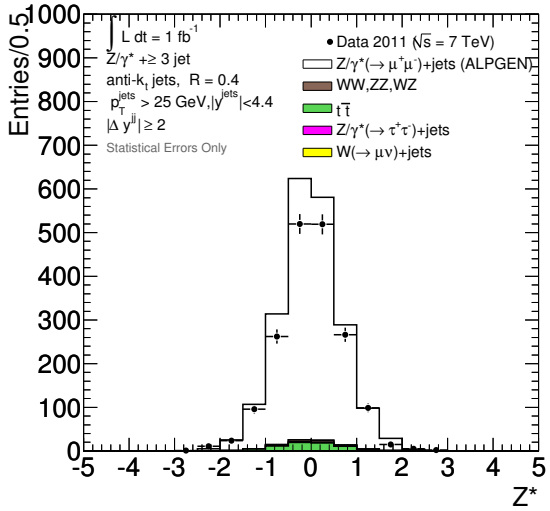



(a) Detector level $N_{jet}$ distribution.

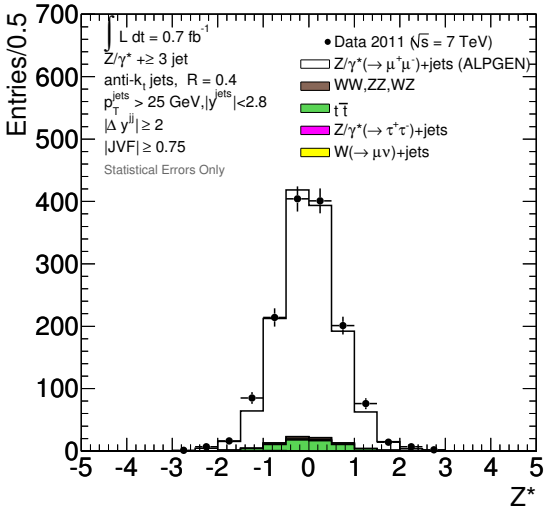



(b) Detector level $N_{jet}$ distribution. Only jets passing a cut JVF > 0.75 are considered.

Figure 7.8: $N_{jet}$ distribution within $|y| < 2.8$ in 2011, without and with a JVF cut applied. Only statistical uncertainties are shown.




(a) Detector level $Z^*$ distribution in 2011.




(b) Detector level $Z^*$ distribution in 2011. Only jets with JVF > 0.75 are considered.

Figure 7.9: $Z^*$ distributions from the 2011 analysis. Only statistical uncertainties are shown.

the performance of a jet veto.

## 7.5 Event selection

All cuts presented in this section have been optimised on simulation only, comparing the characteristics of the signal with those of the background. The data have been blinded during the definition of the cuts, and the studies for the assessment of detector performance have included data-simulation comparisons in control regions only.

The event selection is summarised in Table 7.2. Two isolated leptons of opposite charge are the first requirement for reconstructing a $H \to \tau\tau \to \ell\ell + 4\nu$ event. Depending on the flavour of the leptons, there is a small difference in the following selection, as events with same flavour (SF) leptons ($e^+e^-$ and $\mu^+\mu^-$ ) largely suffer from the $Z \to \ell\ell$ background, that instead is negligible for events with different flavours (DF) leptons. One of the cuts that shows a difference between the SF and DF case is the invariant mass cut, that in the SF case excludes the $Z$ boson peak from the signal region. The distribution of the invariant mass of the two leptons is shown in Fig. 7.10.

After the invariant mass cut, a first separation between events with and without additional jets is performed. The leading jet is required to have $p_{\mathrm{T}}^{\mathrm{jet}} > 40$ GeV. In addition, if the jet is within $|\eta_{\mathrm{det}}| < 2.4$, also a cut $|\mathrm{JVF}| > 0.75$ is applied.

**0-jet bin definition** Events without a leading jet are classified as "0-jet" events. Because of the large $Z \to \ell\ell$ background expected, only DF events are considered in the 0-jet category. DF events are still very common in other background samples, most of all in the top backgrounds. For this reason, additional requirements on the angle between the two leptons and on low $H_{\mathrm{T}}^{\mathrm{Lep}}$[3] are added. The characteristics of the last two variables used for the discrimination are shown in Fig. 7.12a and 7.11, where both signal and background are compared to the distribution in data.

**Candidates with jets** Events with at least an additional jet are treated as boosted events. This means that an additional requirement on $E_{\mathrm{T}}^{\mathrm{miss}}$ ($E_{\mathrm{T}}^{\mathrm{miss}} > 40$ (20) GeV for SF (DF) events) as well as the collinear approximation are applied. This approach is in particular very powerful in rejecting successfully the background while keeping most of the signal.

In the collinear approximation [59], it is assumed that the neutrinos coming from the $\tau$ decay are aligned with the visible leptons. If the $E_{\mathrm{T}}^{\mathrm{miss}}$ of the event is due only to the undetected neutrinos of the $\tau$ decay, the fractions $x_1$ and $x_2$ of the undetected momenta carried by the decay products of each of the $\tau$ can be calculated:

$$x_{1,2} = \frac{p_{\mathrm{vis1,2}}}{(p_{\mathrm{vis1,2}} + p_{\mathrm{mis1,2}})} \tag{7.3}$$

where $x_1$ denotes the undetected fraction of the momentum associated with the leading lepton and $x_2$ the undetected fraction of the momentum associated with the sub-leading lepton. $p_{\mathrm{vis1,2}}$ are the momenta of the leptons and $p_{\mathrm{mis1,2}}$ are the momenta associated with the invisible decay products of the $\tau$ leptons. The distribution for the two variables is shown in Fig. 7.13. The values of $x$ allowed in the event candidates are $0.1 < x_1, x_2 < 1.0$.

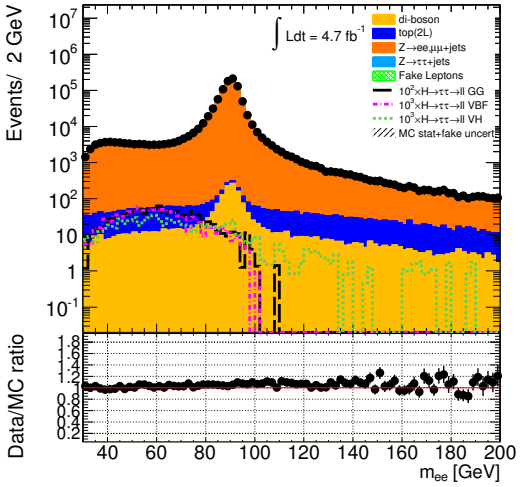
In this approximation the four-momentum of the $\tau$ pair can be reconstructed and the invariant mass of the di-$\tau$ system, $m_{\tau^+\tau^-}$, is derived as:

$$m_{\tau^+\tau^-} = \frac{m_{\ell\ell}}{\sqrt{x_1 x_2}} \tag{7.4}$$

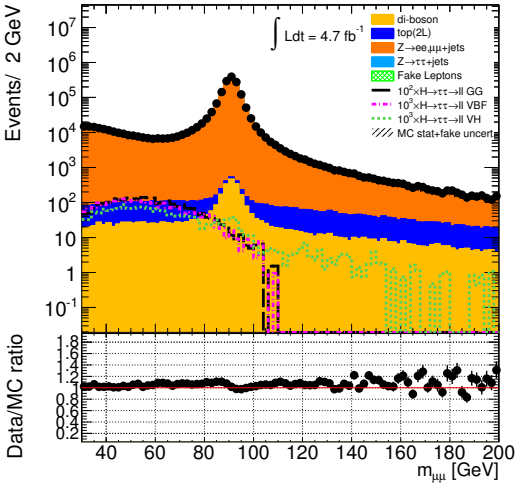
[3] $H_{\mathrm{T}}^{\mathrm{Lep}} = p_{\mathrm{T}}(\ell_1) + p_{\mathrm{T}}(\ell_2) + E_{\mathrm{T}}^{\mathrm{miss}} < 120$ GeV: this cut rejects the top backgrounds, characterised by a large value of $H_{\mathrm{T}}^{\mathrm{Lep}}$ because of the true $E_{\mathrm{T}}^{\mathrm{miss}}$ and the topology of the event.

Table 7.2: Cutflow used in the analysis, highlighting the most important cuts for defining the jet bins.

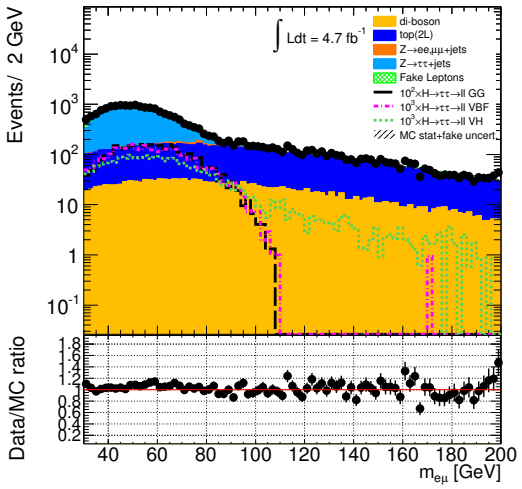
| **0-jet** | **1-jet** | **2-jet VBF** | **2-jet VH** |
|---|---|---|---|
| **Event preselection** | | | |
| Apply GRL<br>Apply trigger selection<br>Select events with at least 1 vertex with more than 2 associated tracks<br>LAr quality cuts<br>Lepton preselection<br>Jet selection<br>Remove e if overlaps with $\mu$, remove jet if overlaps with e and $\mu$<br>Remove event if it contains at least one bad jet | | | |
| **Event selection** | | | |
| Exactly two leptons opposite sign<br>**Mass window cut:** e$\mu$ events with $30 < m_{\text{Eff}} < 100$ GeV, ee $+\mu\mu$ events with $30 < m_{\text{Eff}} < 75$ GeV | | | |
| $\mathbf{\Delta\phi_{\ell\ell} > 2.5}$ | $\mathbf{N^{jet} \geq 1}$ with $p_{\text{T}}^{\text{jet}} > 40$ GeV (\|JVF\| $> 0.75$ if $\|\eta\| < 2.4$)<br>$E_{\text{T}}^{\text{miss}} > 40$ GeV for ee $+\mu\mu$ events $\quad$ $E_{\text{T}}^{\text{miss}} > 20$ GeV for e$\mu$ events<br>$0.1 < x_1, x_2 < 1.0$ (the collinear approximation)<br>$0.5 < \Delta\phi_{\ell\ell} < 2.5$ | | |
| $H_{\text{T}}^{\text{Lep}} = p_{\text{T}}(\ell_1) + p_{\text{T}}(\ell_2) + E_{\text{T}}^{\text{miss}} < 120$ GeV | **event fails the VBF or VH cuts**<br>$\mathbf{m_{\tau\tau j} > 225}$ **GeV** | $\mathbf{N^{jet} \geq 2}$ with $\mathbf{p_{\text{T}}^{\text{jet}} > 25}$ **GeV**<br>$\mathbf{\Delta\eta_{jj} > 3.0}$<br>$\mathbf{m_{jj} > 350}$ **GeV** | $\mathbf{\Delta\eta_{jj} < 2.0}$<br>$\mathbf{50 < m_{jj} < 120}$ **GeV** |
| | $b$-tag veto | | |




(a) $ee$ channel




(b) $\mu^+\mu^-$channel




(c) $e\mu$ channel

Figure 7.10: Comparison of the $m_{\ell\ell}$ distribution between data and MC after preselection and opposite sign requirement for $e^+e^-$, $\mu^+\mu^-$ and $e\mu$ channels. Backgrounds with fake leptons and $Z \to \tau\tau$ are estimated with data-driven methods. All other contributions are estimated using simulated event samples.

This approximation is useful only in the boosted case[4], and the mass estimator in Eq. 7.4 (so-called *collinear mass*) is used only in the 1- and 2-jet categories.

**2-jet bin definition** If additional jets are present in the event, the second leading jet must fulfill both the JVF and a $p_{\mathrm{T}}^{\mathrm{jet}}$ requirement to classify the event as a 2-jet event. A $p_{\mathrm{T}}^{\mathrm{jet}} > 25$ GeV cut has been chosen, following the results shown in Sect. 7.4.1. Lowering the $p_{\mathrm{T}}^{\mathrm{jet}}$ cut improves the

[4]In the unboosted case, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is almost undetectable and the estimate of the invariant mass can only rely on the visible variables, i.e. the lepton kinematics. The difference in the lepton topology between the two cases is clearly shown in Fig. 7.12, where the events in the 0-jet and in the 1- and 2-jet bin are compared, just before the selection on the angular distribution of the leptons.

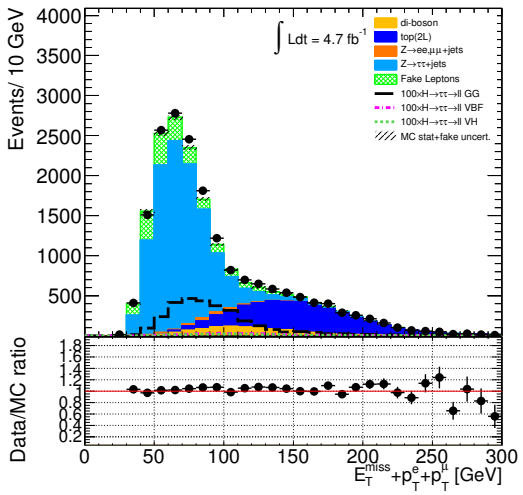
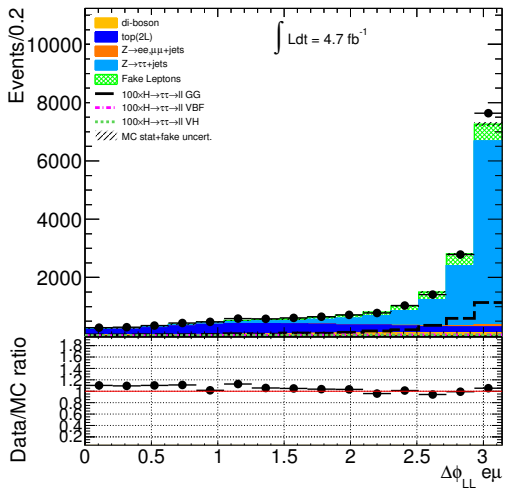
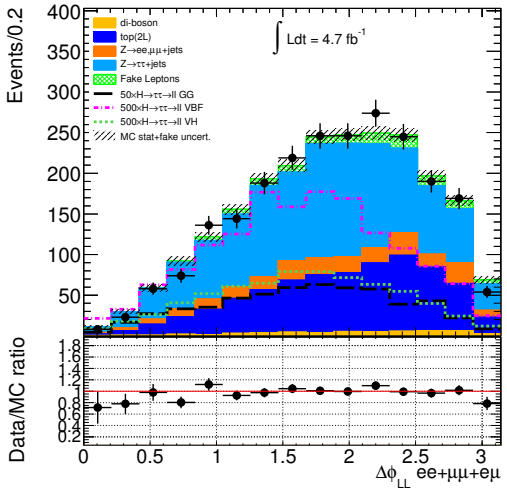



Figure 7.11: Comparison of the $H_{\mathrm{T}}^{\mathrm{Lep}}$ distributions after invariant mass cut for $e\mu$ events in the 0-jet category.




(a) $\Delta\phi_{\ell\ell}$ in the 0-jet bin, after the invariant mass cut

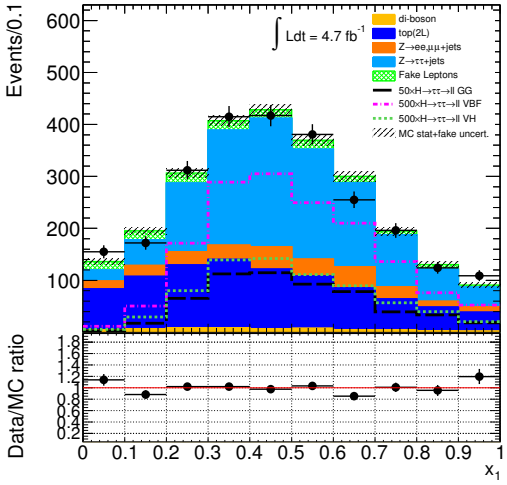



(b) $\Delta\phi_{\ell\ell}$ in the 1- and 2-jet bins, after the cut on $x_1$ and $x_2$.

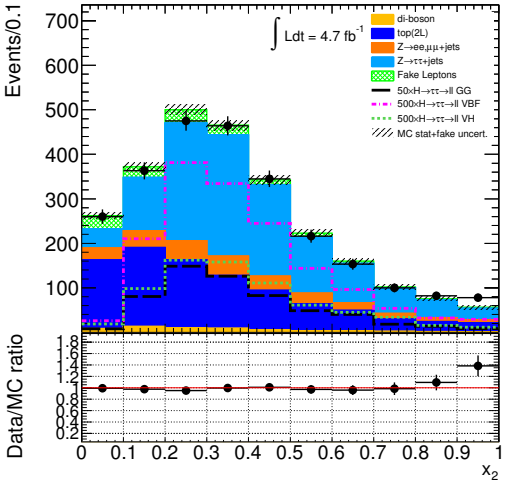
Figure 7.12: Comparison of the leptons $\Delta\phi_{\ell\ell}$ before the cut on $\Delta\phi_{\ell\ell}$, highlighting the difference between the events in the 0- and in the 1- and 2-jet bin.

efficiency of the selection. Fig. 7.14 shows the jet multiplicity for both $p_{\mathrm{T}}^{\mathrm{jet}}$ cuts: the excellent description of these distributions is also an indirect confirmation of the precision of the background estimation methods. The differences seen in the plots are usually negligible when the systematic uncertainties (not shown in the plots) are taken into account.

Events with at least two jets are then separated into two additional bins, depending on their main characteristics, as explained in Sect. 7.2.3.

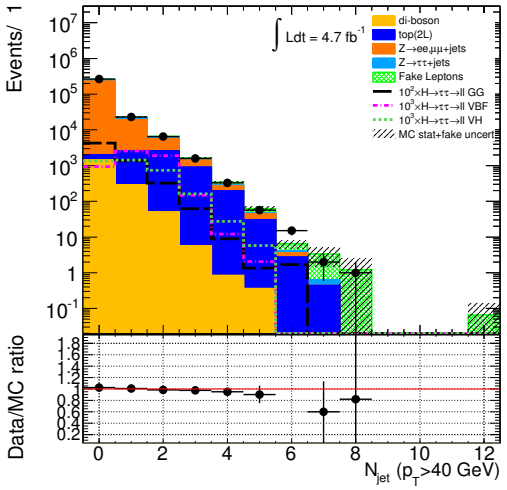



(a) $x_1$

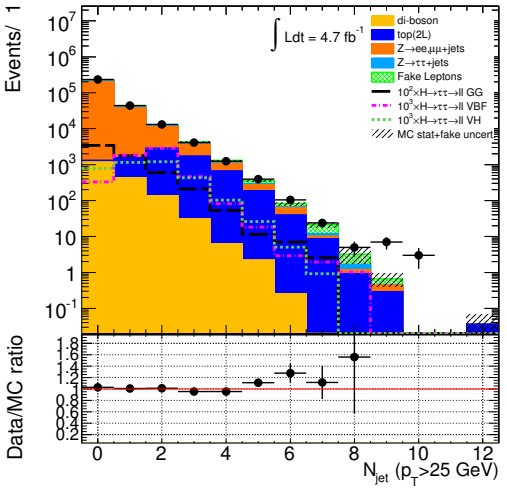



(b) $x_2$

Figure 7.13: Comparison of the leptons $x_1$ and $x_2$ distributions between data and MC after cuts 1-4 for $ee$+$\mu\mu$+$e\mu$ events.



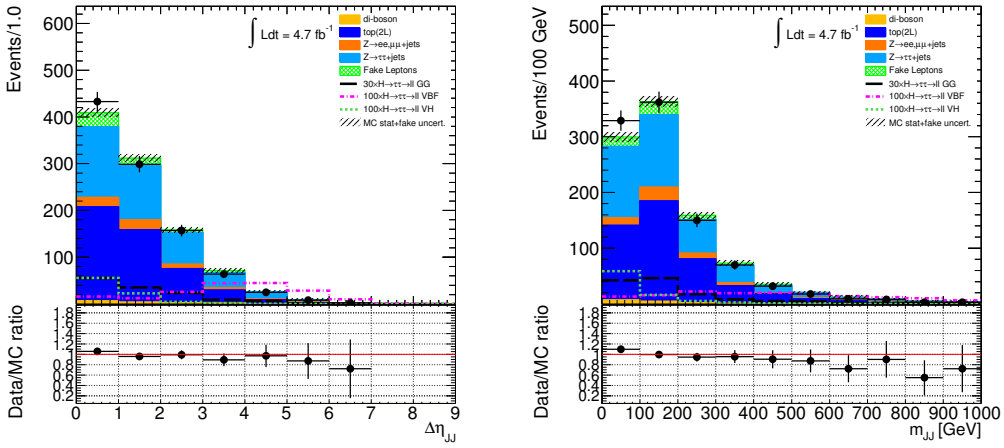

(a) Comparison between data and Monte Carlo of the jet multiplicity distribution with $p_{\mathrm{T}}^{\mathrm{jet}} > 40$ GeV after invariant mass cut for all channels.




(b) Comparison between data and Monte Carlo of the jet multiplicity distribution with $p_{\mathrm{T}}^{\mathrm{jet}} > 25$ GeV after invariant mass cut for all channels.

Figure 7.14: Comparison of $p_{\mathrm{T}}^{\mathrm{jet}}$ distribution. Backgrounds with fake leptons and the $Z \to \tau\tau$ contribution are obtained with data-driven methods. All other contributions are estimated using simulated event samples.

**2-jet VBF category** According to the main VBF characteristics (see Sect. 4.3.1 and 7.2.3), events showing a large pseudorapidity gap ($\Delta\eta_{jj} > 3.0$), as well as a large invariant mass ($m_{jj} > 350$ GeV) are classified as "VBF-like" events. The two values are chosen to provide a large rejection against the $t\bar{t}$ background. In order to avoid additional contamination from top backgrounds, also a b-jet veto has been implemented.




(a) $\Delta\eta_{jj}$ (b) $m_{jj}$

Figure 7.15: Comparison of the jets $\Delta\eta_{jj}$ and $m_{jj}$ distributions after requiring a second jet in the event for $ee{+}\mu\mu{+}e\mu$ events in the 2-jet (VBF+$VH$) bin.

**2-jet VH category** Events with at least two jets but with small pseudorapidity gaps are more similar to $VH$ and $gg \to H$ events (see Sect. 7.2.3). In this case, Higgs candidate events are selected within a window in the invariant mass distribution of the di-jet system. This enhances the selection of ZH events over the background. Also an additional b-jet veto to reject the top background is applied.

**1-jet category definition** In case the event fails the requirement for VBF and VH selection, it is classified as a "1-jet event". The boosted topology is required to satisfy a cut on the invariant mass of the $\tau^+\tau^-$ pair, calculated in the collinear approximation, and the leading jet ($m_{\tau^+\tau^-,j}$). This requirement on large invariant mass of the di-$\tau$ system and the leading jet reduces the background from $Z \to \ell\ell$ process, that does not show such a large boost [93]. Figure 7.16 shows the invariant mass distribution of the leading jet and the di-$\tau$ system before the cut on $m_{\tau^+\tau^-,j}$.

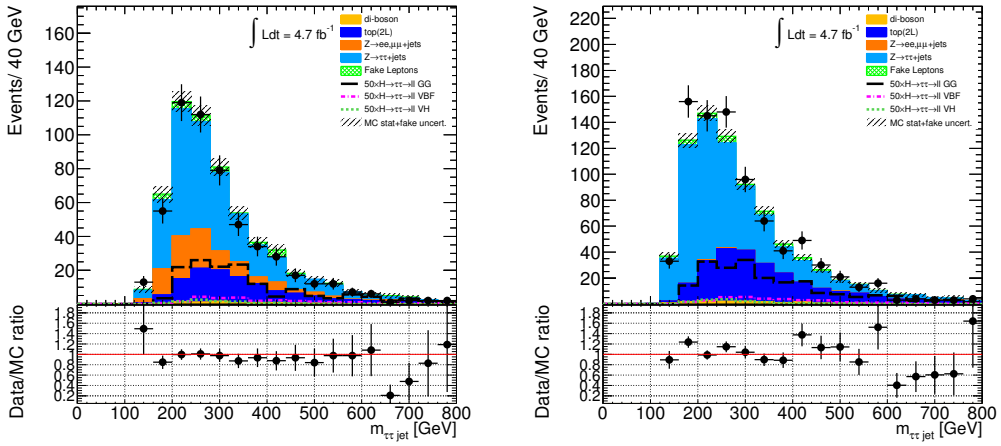
To avoid contamination from the top background, also an additional b-jet veto is applied on the leading jet.

**Invariant mass distributions** Figure 7.17 shows the $m_{\tau^+\tau^-}$ distribution of events passing the selection cuts for the four categories. For the categories where the collinear approximation is used, the collinear mass (Eq. 7.4) is used as mass estimator; for the 0-jet category instead, the effective mass $m_{\text{Eff}}$[5] is used. All the events in the mass distribution are considered in the limit setting procedure.

Table 7.3 shows the event yields and estimated number of background events for 4.7 fb$^{-1}$ after the selection criteria in the four bins.

The backgrounds shown in the plots and quoted in the table are where possible extracted directly from data, or otherwise estimated from simulation. More details on the data-driven estimation, in particular for the $Z \to \tau\tau$ background, are reported in Sect. 7.6.

[5]The effective mass is defined as the invariant mass formed by the $E_{\text{T}}^{\text{miss}}$ and the two lepton four-momenta.




(a) $ee$+$\mu\mu$ (b) $e\mu$

Figure 7.16: Comparison of the invariant mass distribution of the Higgs boson candidate and leading jet between data and MC after preselection cuts and event selection, before the $m_{\tau^+\tau^-,j}$ cut, for $ee$+$\mu\mu$ and $e\mu$ events in the 1-jet category.

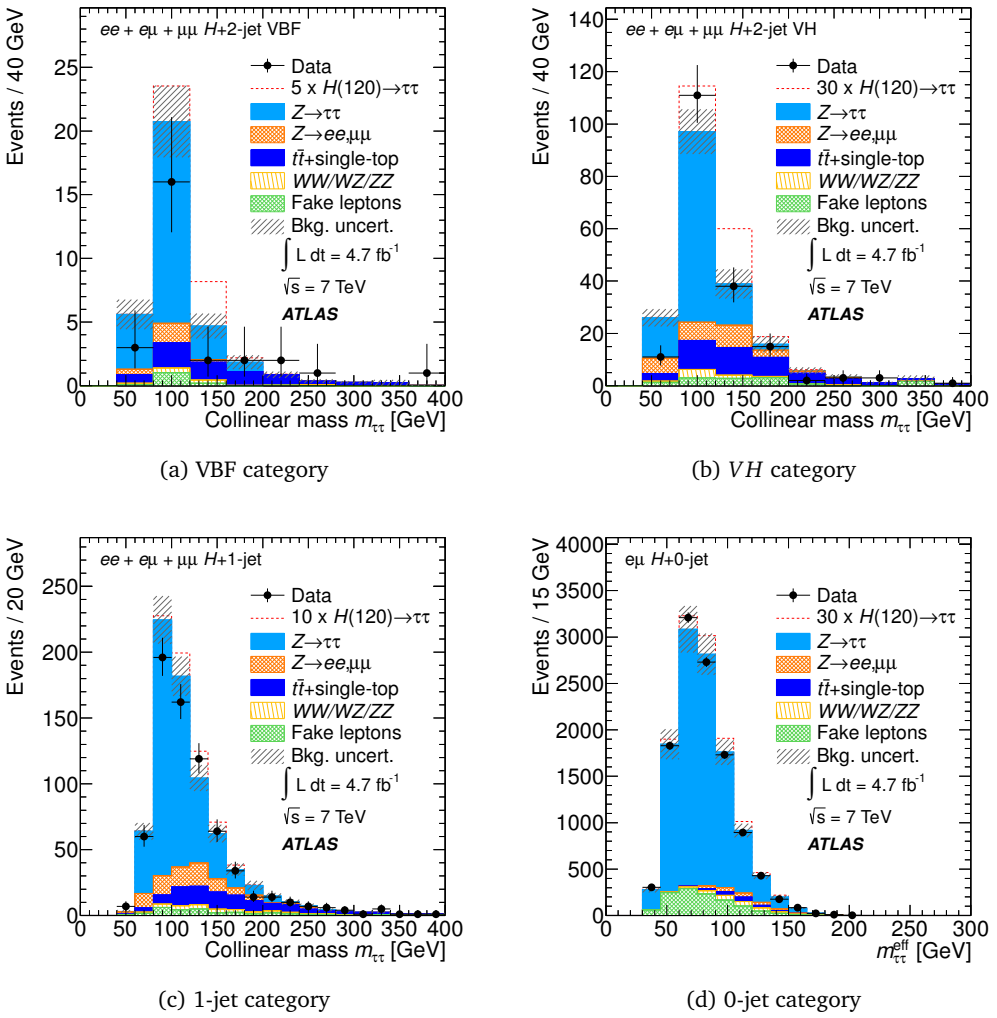
Table 7.3: Number of events after the $H \to \tau\tau \to \ell\ell + 4\nu$ selection for the four categories in data and predicted number of background events, for an integrated luminosity of 4.7 fb$^{-1}$. Expectations for the Higgs boson signal ($m_H = 120$ GeV) are also given. Statistical and systematic uncertainties are quoted, in that order.

| | $ee + \mu\mu + e\mu$ $H$ +2-jet VBF | $ee + \mu\mu + e\mu$ $H$ +2-jet $VH$ | $ee + \mu\mu + e\mu$ $H$ +1-jet | $e\mu$ $H$ +0-jet |
|---|---|---|---|---|
| $gg \to H$ signal | 0.26±0.06±0.10 | 0.8 ±0.1 ±0.2 | 3.9 ±0.2 ±1.0 | 23 ±1 ±3 |
| VBF $H$ signal | 1.08±0.03±0.11 | 0.10±0.01±0.01 | 1.15±0.03±0.01 | 0.75±0.03±0.06 |
| $VH$ signal | 0.01±0.01±0.01 | 0.53±0.02±0.07 | 0.40±0.02±0.03 | 0.52±0.02±0.04 |
| $Z/\gamma^* \to \tau^+\tau^-$ | 24 ± 3 ± 2 | 107 ±12 ± 9 | (0.52±0.01±0.04)·10$^3$ | (9.68±0.05±0.07)·10$^3$ |
| $Z/\gamma^* \to \ell^+\ell^-$ ($\ell$=e,$\mu$) | 2 ± 1 ± 1 | 25 ± 4 ± 9 | 83 ±10 ±30 | 185 ±11 ±14 |
| $t\bar{t}$+single top | 7 ± 1 ± 2 | 42 ± 2 ± 6 | 98 ± 3 ±12 | 169 ± 4 ±14 |
| $WW/WZ/ZZ$ | 0.9± 0.3± 0.3 | 6 ± 1 ± 1 | 21 ± 1 ± 3 | 221 ± 3 ±18 |
| Fake leptons | 1.3± 0.8± 0.6 | 13 ± 2 ± 5 | 30 ± 4 ±12 | ( 1.2±0.5)·10$^3$ |
| Total background | 35 ± 3 ± 4 | 193 ± 7 ±20 | (0.75±0.01±0.05)·10$^3$ | (11.4±0.5)·10$^3$ |
| Observed data | 27 | 185 | 702 | 11420 |

## 7.6 Data-Driven Background Estimation

**$Z$ boson production** The contribution from $Z \to \ell\ell$ $(e,\mu)$ is determined by scaling the yields in the Monte Carlo simulation using correction factors obtained by comparing data to simulation in low- and high-$E_{\mathrm{T}}^{\mathrm{miss}}$ control regions enriched in these backgrounds. In this way, the modelling of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is ensured to be correct. The correction factors are obtained separately for $Z \to ee$ and $Z \to \mu\mu$ and for the different analysis categories and are on the order of 10%.

The $Z \to \tau\tau$ background is instead modelled using the embedding method. Because this background is crucial in the analysis, it is described in more details in Sect. 7.6.2. The results of the




(a) VBF category

(b) $VH$ category

(c) 1-jet category

(d) 0-jet category

Figure 7.17: Comparison of the $m_{\tau^+\tau^-}$ invariant mass after the full cutflow in the four defined jet bins. In the case of the 0-jet category the effective mass $m_{\text{Eff}}$ is shown. The backgrounds with fake leptons and the $Z \to \tau\tau$ contribution are estimated from data. All other contributions are estimated using simulated event samples.

validation of the embedding sample used in the $H \to \tau\tau \to \ell\ell + 4\nu$ analysis is reported.

**Fake leptons** The fake lepton background consists of events that have a reconstructed lepton that did not originate from the decay of a $\tau$ lepton or the leptonic decay of a $W$ or $Z$ boson. For example, a di-jet event with two jets faking a lepton, but also a  event with an additional jet reconstructed as a lepton contribute to this background. The normalisation and shape of relevant distributions are obtained from data with a template method using a control region in which the lepton isolation requirement is reversed. The chosen template shape is the $p_{\text{T}}$ distribution of the sub-leading lepton. For this method to be applied, it is first verified that the template shapes of the fake lepton distribution in the control and signal regions agree within uncertainties. This is performed at intermediate steps of the event selection where the data sample is dominated by

background events and where the number of expected signal events is negligible. After subtracting the simulated backgrounds, the template shape in a given distribution is obtained from the control region, while the normalisation is obtained from a fit of the distribution of the events in the signal region with the template shape.

**Top and di-boson background** The contributions of the $t\bar{t}$, single top quark and electroweak di-boson backgrounds are estimated from simulation. The Monte Carlo description of the top quark backgrounds has been validated using data by selecting control regions enriched in top quark background processes. The control regions are defined by inverting the b-jet selection for the categories with jets, and by inverting the $H_{\mathrm{T}}^{\mathrm{lep}}$ selection for the 0-jet category.

Table 7.3 displays the number of events expected and observed in the four categories after all selection criteria including all systematic uncertainties described in Sect. 7.7. The estimated combined background contributions are found to give a good description of all quantities relevant to the analysis.

### 7.6.1 $Z \to \tau\tau$ data-driven estimation: embedding method

The goal of a data-driven estimation of the $Z \to \tau\tau$ background is to reduce the experimental uncertainties related to the event modelling in the simulation. This improved estimation should ensure that the description provided by the background model matches the conditions in data. This means reproducing the behaviour of calorimeter variables, jet production and topology in association with jets (affected by non-negligible uncertainties, as shown in Chap. 6), and the $E_{\mathrm{T}}^{\mathrm{miss}}$ effects related to mis-measurements and to the environmental conditions. To do this, an hybrid sample, where those variables are directly taken from data $Z \to \mu\mu$ events, is used. The procedure has been studied in detail on simulated 14 TeV signal samples in Schmitz [103] and Möser [96]. It has been further developed for the data operation in the context of other Bonn theses (Schwindt, in preparation), which developed also the validation and systematics procedures. The embedding method has been already used in previous searches in the context of MSSM Higgs [27].

$Z \to \tau\tau$ and $Z \to \mu\mu$ events show the same lepton and jet topology, because of the universality of the weak interaction. For this reason, $Z \to \mu\mu$ events can provide an excellent model for the $Z \to \tau\tau$ events in data. The only difference between these two processes is due to the $\tau$ and $\mu$ mass difference, that has to be taken into account separately. Because the Higgs coupling to muons in the SM is negligible, $Z \to \mu\mu$ events are also completely signal free, providing an excellent control region. $Z \to \tau\tau$ events from data, instead, because of the presence of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event, are always affected by the uncertainty in the reconstruction of the invariant mass, that does not permit to define a completely signal-free region. All those assumptions on the process hold at particle level; at detector level, a validation of the description of $Z \to \tau\tau$ events has to be performed.

In $Z \to \mu\mu$ data events detected in ATLAS, the embedding algorithm replaces muons with taus. The main assumption of the method is that, even though $Z \to \mu\mu$ events in data will show some characteristics related to the muon system responsible for the detection, as well as for trigger requirements and inefficiencies, those effects are not playing a role for the final $Z \to \tau\tau$ quantities used for the limit extraction, as invariant masses. I have carried out an investigation of the lepton kinematics, because the description of their topology (as shown, e.g., in Chap. 6) is already good in Monte Carlo simulations. For this reasons, $Z \to \tau\tau$ simulated samples can provide a reference

for the lepton description in the embedding sample, and I have thus used the Monte Carlo sample for the validation and normalization of the embedding sample.

### Definition of the hybrid embedding sample

The procedure of embedding simulated $\tau$ decays into $Z \to \mu\mu$ data events is briefly described below and illustrated in Fig. 7.18. Corresponding example event displays are shown in Fig. 7.19.

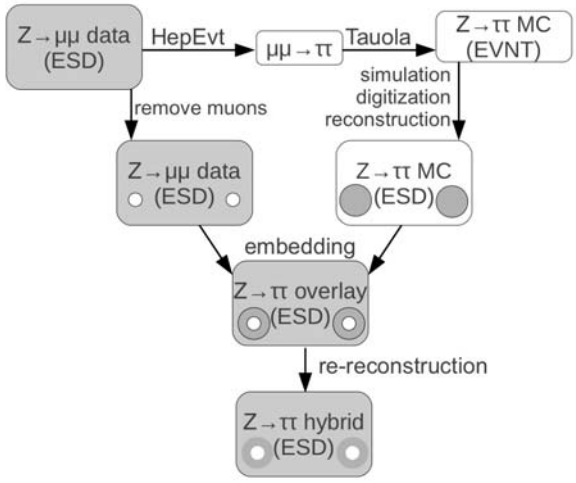



Figure 7.18: Flowchart of the embedding procedure.

(a) data event (b) *mini event* (c) embedded event

Figure 7.19: Example event displays of the embedding steps for a single $Z \to \mu\mu$ to $Z \to \tau\tau \to \tau_h\tau_h$ event: (a) after the selection, (b) after the Monte Carlo simulation, and (c) after the re-reconstruction.

$Z \to \mu\mu$ events are selected using cuts as those already listed in Chap. 6. To optimize the amount of $Z \to \mu\mu$ events collected, all the events that fire a single muon trigger in data are considered. The stream is stored as *ESD*, a special data format that contains more details about the detector condition with respect to the streams usually used in the analysis. The full calorimeter information, at cell level, is needed to remove the muon and its energy deposits from the event, and substitute it with a $\tau$ lepton decay simulated in ATLAS, in the same position as the original muon track.

The muons are selected applying a $p_T$ cut at 15 GeV and a tracking isolation requirement[6]. To

[6]The isolation requirement is done on tracks in order not to bias the procedure, that also considers calorimeter energy

fulfill the $Z \to \mu\mu$ requirement, these muons are in addition required to have a common primary vertex and an invariant mass $m_{\mu\mu} > 55$ GeV.

The $Z$ is reconstructed and the characteristics of the $Z \to \mu\mu$ decay are written to an ascii file in the HEPEVT format. The rest of the event is neglected now. In the ascii file, the production vertex of the $\tau$ leptons is set to the production vertex of the input muons. The muons are then replaced by $\tau$ leptons on truth level, with their four-momenta rescaled in the $Z$ rest frame, so that $p_\tau = \sqrt{E_\mu^2 - m_\tau^2}$ in order to account for the mass difference between the muons and $\tau$ leptons

The HEPEVT file with the $Z \to \tau\tau$ kinematics extracted from the reconstructed $Z \to \mu\mu$ events is then processed with TAUOLA and PHOTOS, resulting in a pure Monte Carlo decay without an underlying event. Here, the decay of the $\tau$ leptons into different configurable final states by TAUOLA is taking into account the polarisation of the $\tau$ leptons while PHOTOS adds the final state radiation from the decay products. This Monte Carlo event is then processed by the full ATLAS detector simulation, digitisation and reconstruction. In order to avoid double counting in the later merging with the corresponding data event, the calorimeter noise is switched off during the digitisation. The output of this simulation step is called a *mini event*, as it does not include the other ingredients of the original $Z \to \mu\mu$ event.

In order to replace the muons in the data event with the correspondingly simulated $\tau$ leptons all associated muon tracks are removed in each selected $Z \to \mu\mu$ data event. To subtract the energy deposition of the muons in the calorimeter, a second *mini event* with the initial $Z \to \mu\mu$ kinematics is produced, and the simulated calorimeter energy is subtracted from the $Z \to \mu\mu$ data event at cell level. Then all calorimeter cell energies from the simulated $Z \to \tau\tau$ event are added to the corresponding data, and all tracks are copied. This inserts the pure $Z \to \tau\tau$ decay into the data environment keeping the event properties as close to data conditions as possible.

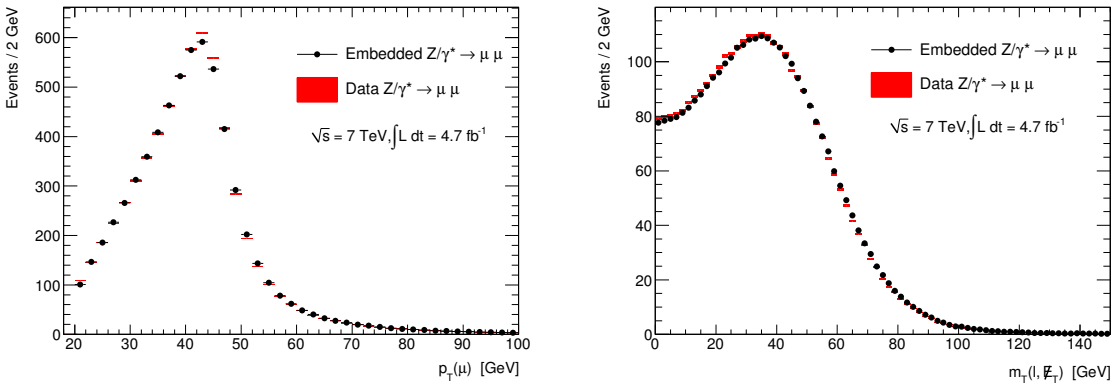
The resulting $Z \to \tau\tau$ hybrid events are then submitted to a full event reconstruction, and all objects and the missing transverse energy are recomputed from the cells and tracks of the hybrid event.

### Validation of embedding method in $Z \to \mu\mu$ to $Z \to \mu\mu$ embedding

The first step in validating the method is to replace the $Z \to \mu\mu$ data event with a $Z \to \mu\mu$ simulated event and check the event description against the original event. This procedure is exactly identical to the one used for the $Z \to \tau\tau$ embedding, but instead of simulating a $\tau$ decay, a mini-event with muons is simulated and used to replace the muons in data. As the two events used in the comparison are exactly the same, except for the muon tracks, we can identify possible biases of the method itself.

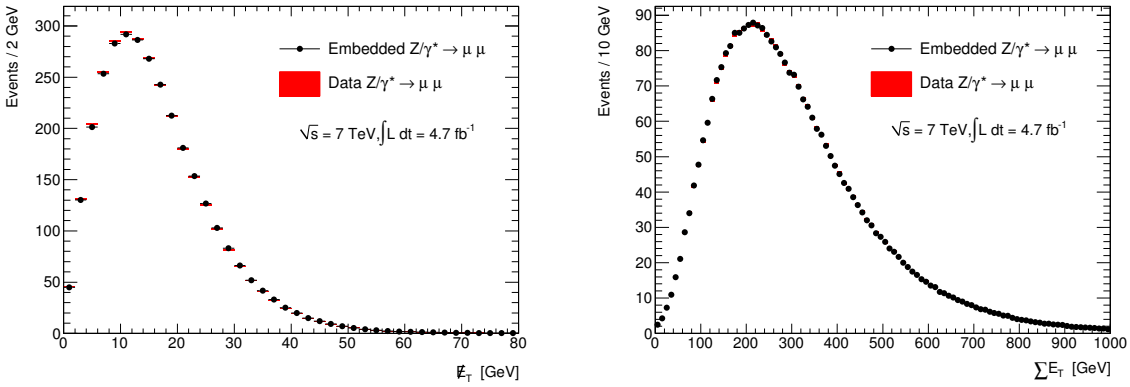
Figures 7.20 and 7.21 show similar kinematic distributions before and after embedding. The difference is mainly due to the resolution of the simulated muons, which is folded with the resolution already present in the selected data. This validation study shows that no bias is introduced by the embedding in the $E_{\mathrm{T}}^{\mathrm{miss}}$ and in sum of transverse energy distributions, that are in good agreement with the distributions observed in the $Z \to \mu\mu$ data sample, as shown in Fig. 7.21. A perfect description would only be achieved by unfolding the resolution of the selected muons. All visible effects are however negligible in embedded $\tau$-decays, especially regarding the applied systematic variations of the method.

deposit in the simulation of the $Z \to \tau\tau$ event. The requirement is $\sum_{\mathrm{tracks}}^{\Delta R \leq 0.2} \frac{p_{\mathrm{T\,tracks}}}{p_{\mathrm{T}}} < 0.2$




(a) (b)

Figure 7.20: Comparison of the transverse momentum (left) and transverse mass distribution (right) from the selected data and the embedded $Z \rightarrow \mu\mu$ sample. From Schwindt (in preparation).




(a) (b)

Figure 7.21: Comparison of the missing transverse energy (left) and the total sum of energy (right) from the selected data and the embedded $Z \rightarrow \mu\mu$ sample. From Schwindt (in preparation).

### 7.6.2 Validation of embedded $Z \rightarrow \tau\tau$ events in the $H \rightarrow \tau\tau \rightarrow \ell\ell + 4\nu$ analysis

The definition of a reference for the validation of the embedded $Z \rightarrow \tau\tau$ events need some initial approximations.

Because the trigger can't be simulated in the embedding framework used, and because the sample is produced from data $Z \rightarrow \mu\mu$ events which are affected in a non-trivial way by the trigger selection, real data events are not a good reference.

Assuming that the lepton experimental uncertainties in Monte Carlo simulations are well described, simulated samples can be considered as a reasonable reference for the lepton kinematics. Also, the trigger effects on the variables of interest (e.g., the leptons $p_T$) can be studied, and those variables can be then compared to the distributions obtained with the embedding sample. In addition, because of the complicated composition of the embedding sample, it is not possible to define an absolute normalization for it. The Monte Carlo sample provides a reference for the normalization

of the embedded sample. Defining a Monte Carlo reference for the embedding sample means that some of the uncertainties that affect the simulations (e.g. the cross-section uncertainty) affect the embedded sample as well; but on the other side, the uncertainty on the event modelling is reduced using this procedure, bringing a clear advantage over using a pure Monte Carlo background.

The embedded data distributions are normalised to the $Z \to \tau\tau$ visible cross section after the opposite sign cut. The embedding visible cross section is defined as:

$$\sigma_{\text{embedding}} = \sigma_{Z\to\tau\tau,\text{NLO}} \cdot C_{\text{eff}}(\ell\ell) \tag{7.5}$$

The correction factor takes into account the fact that the selection on the embedding sample and on the $Z \to \tau\tau$ Monte Carlo sample is not identical, due to the effects of the trigger selection on the lepton kinematics. The factor is defined as a double ratio of the selection efficiency in the two samples:

$$C_{\text{eff}}(\ell\ell) = \frac{\epsilon_{\text{MC}}}{\epsilon_{\text{embedding}}}(\ell\ell) = \frac{\mathrm{N}_{\text{MC}}^{\text{after sel}}}{\mathrm{N}_{\text{MC}}^{\text{total}}} \frac{\mathrm{N}_{\text{embedding}}^{\text{total}}}{\mathrm{N}_{\text{embedding}}^{\text{after sel}}} \tag{7.6}$$

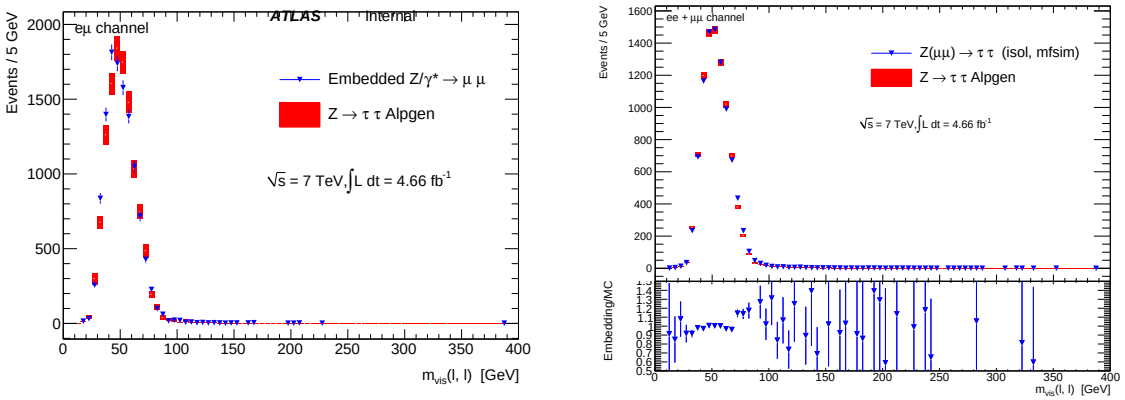
Fig. 7.22 and 7.23 compare the distributions of some kinematic quantities for $Z \to \tau\tau$ MC and the embedded data sample: the visible mass in Fig. 7.22, and the effective mass in Fig. 7.23. In Fig. 7.24 the nominal embedding sample is shown in the plots together with its systematic variations:

- default sample: a relatively soft isolation of ptcone20/pt<0.2 ("`isol`") is applied to the selected muons. Also, to subtract the muon energy from the calorimeter a muon with the same characteristics is simulated in a clean event (without PHOTOS) and the simulated cell energies are subtracted from data ("`mfsim`").
- Two different requirements: no isolation at all ("`noisol`"), or a tighter version ptcone40/pt< 0.06 AND etcone20/pt< 0.04 ("`tightisol`"), define the systematic variations for the embedding sample.
- in addition, the treatment of the calorimeter cell subtraction during replacement of the muons is varied: as a conservative estimation of the systematic error, the energy of each cell is scaled up/down by 30% before subtraction in the corresponding "`mfsup`"/"`mfsdn`" samples. These two variations define two additional systematic samples with respect to the default procedure.

Another important quantity for the definition of $m_{\tau^+\tau^-}$ is the $E_{\text{T}}^{\text{miss}}$. In this case, the validation is more complicated. A simple comparison with the reference Monte Carlo could not provide the answer: any mis-modeling of the $E_{\text{T}}^{\text{miss}}$ in the Monte Carlo would result in a difference between the two samples, that is not at all related to the embedding method. Such a difference is visible in our plots, shown in Fig. 7.25; this is actually related to a mis-modelling of the $E_{\text{T}}^{\text{miss}}$ in the PYTHIA sample used for the pile-up simulation, as confirmed by other performance studies. Only the comparison of $E_{\text{T}}^{\text{miss}}$ in $Z \to \mu\mu$ events, shown in Fig. 7.21, can confirm that no biases are introduced by the embedding procedure. Also, the impact of the $\tau$ leptons decay on the $E_{\text{T}}^{\text{miss}}$ distribution is not negligible, and leaves a modelling uncertainty in the embedding sample.

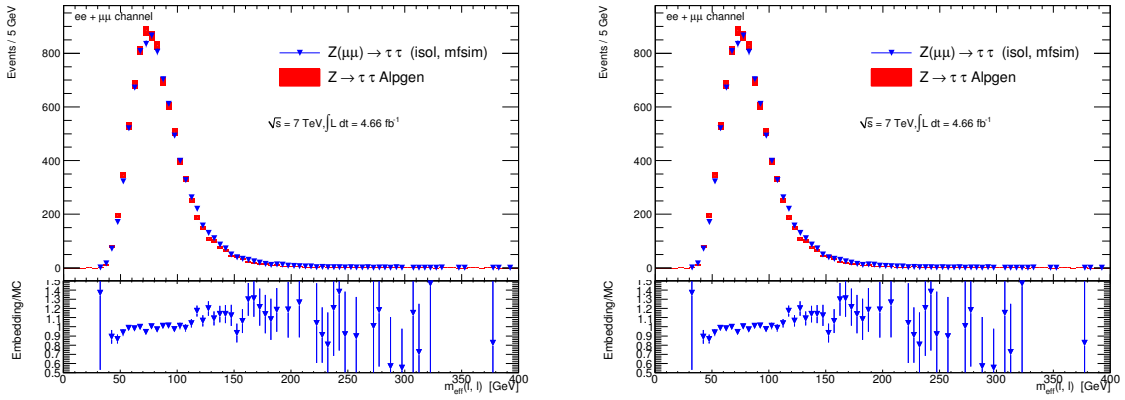
## 7.7 Systematic uncertainties

Systematic uncertainties on the normalisation and shape of the signal and background mass distributions are taken into account. These are treated either as fully correlated or uncorrelated across




(a) (b)

Figure 7.22: Comparison of the visible mass distribution from the $Z \rightarrow \tau\tau$ MC and the embedding sample, before the visible mass cut.

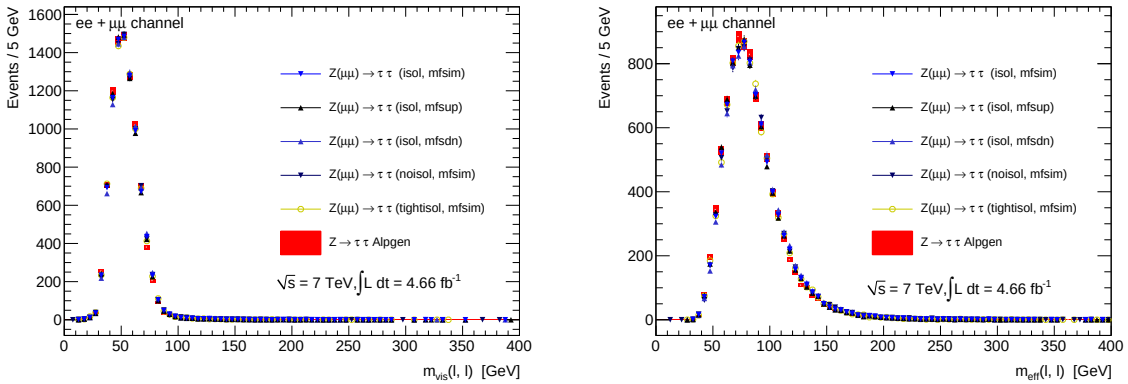



(a) (b)

Figure 7.23: Comparison of the transverse mass distribution from the $Z \rightarrow \tau\tau$ MC and the embedding sample, before the visible mass cut.

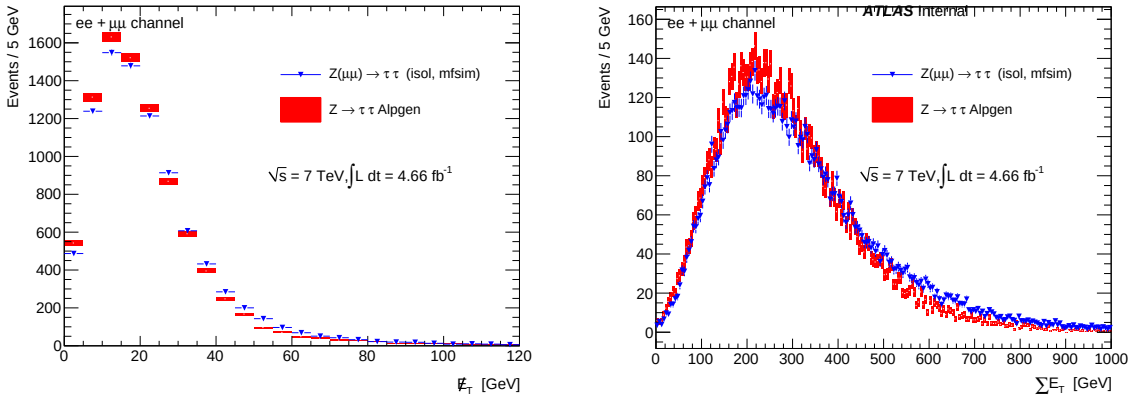
categories. In the case of partial correlations, the source is separated into correlated and uncorrelated components. The dominant correlated systematic uncertainties are those on the measurement of the integrated luminosity and on the theoretical predictions of the signal production cross sections and decay branching ratios, as well as those related to detector response that impact the analyses through the reconstruction of electrons, muons, hadronic $\tau$ decays, jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and b-tagging.

**Theoretical uncertainties** The theoretical uncertainties considered in this analysis have been already explained in Sect. 2.4.4. The QCD scale uncertainties on the signal cross sections have been computed following the prescription explained in Sect. 7.2. They are of the order of 1% for the VBF and $VH$ production modes, and in the range of 8–25% for $gg \rightarrow H$ depending on jet multiplicity [24, 108]. An uncertainty of 4–5% is assumed for the inclusive cross section of the single vector boson and di-boson production mechanisms and a relative uncertainty of 24% is added in




(a) (b)

Figure 7.24: Comparison of the mass distributions from the embedding sample, and the systematic variations, before the visible mass cut.




(a) (b)

Figure 7.25: Comparison of the missing transverse energy (left) and the total sum of energy (right) from the $Z \rightarrow \tau\tau$ MC and the embedding sample, before the visible mass cut.

quadrature per additional jet. For both $t\bar{t}$ production and single top quark production, the QCD scale uncertainties are in the range of 3–6% [83, 95]. The uncertainties related to the PDF amount to 8% for the predominantly gluon-initiated processes, $gg \rightarrow H$ and $t\bar{t}$, and 4% for the predominantly quark-initiated processes, VBF, $VH$, single vector boson and di-boson production [40, 45, 87, 92]. The systematic uncertainty arising from the choice of different sets of PDF is included. In addition to the theoretical errors considered in LHC Higgs Cross Section Working Group et al. [86], other effects are taken into account. Uncertainties related to hadronisation effects are estimated by replacing PYTHIA with HERWIG. Effects due to initial and final state radiation are assessed with variated PYTHIA. The effect of a different choice of parton shower and underlying event parametrisation yields a total uncertainty of about 10% on the acceptance of the Higgs boson produced via the VBF mechanism in the H+2jet VBF channel.

**Detector-related uncertainties** The uncertainty on the integrated luminosity is considered as fully correlated across all analysis categories and amounts to 3.9% [11, 12]. The effect of pileup on the signal and background expectations is modelled in the Monte Carlo simulations and the corresponding uncertainty is taken into account.

Appropriate scale factors for the trigger efficiencies of electron and muon triggers are obtained from data and applied to the Monte Carlo simulations. The associated systematic uncertainties are typically 1–2%. Differences between data and Monte Carlo simulations in the reconstruction and identification efficiencies of electrons and muons are taken into account, as well as the differences in the momentum scales and resolutions.

The systematic uncertainties related to the jet energy scale, resolution and b-veto are modelled as functions of $\eta$ and $p_{\mathrm{T}}$. The uncertainty on the energy scale of the different physics objects is propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$. Uncertainties associated with the remaining pileup noise and cluster activity in the calorimeters are also considered as independent $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties.

The detector-related uncertainties depend on the event topology and are typically small compared to the theoretical uncertainties. The main exceptions is the jet energy scale uncertainty, which reaches up to 12%.

**Background modelling uncertainties** The uncertainty of the embedding sample are estimated as explained in Sect. 7.6.2, using the variations shown in Fig. 7.24. The variated samples provide the uncertainty related to the event selection and the muon subtraction procedure.

The uncertainty related to the estimation of backgrounds with fake leptons is calculated from the uncertainty on the subtraction of other processes from Monte Carlo simulation and from the difference in the $p_{\mathrm{T}}$ shape of the events in the control region and signal regions. Such systematic uncertainties lie in the range of 30–40%.

## 7.8 Results

No significant excess is observed in the data compared to the SM expectation. From the comparison between data and predictions in the mass distributions reported in Fig. 7.17, an upper limit on the Higgs cross section as a function of the Higgs mass has been derived. The amount of integrated luminosity is not enough to derive an exclusion or a discovery of the Higgs boson in a certain mass range with this channel alone. The results have thus been combined with the other $\tau^+\tau^-$ channels, and included in the global ATLAS combination.

The limit has been derived using a binned likelihood method. The limit calculation uses the number of events, for the signal (at different Higgs masses) and the backgrounds, in the mass distribution used in each of the four categories. The systematic uncertainty have been considered for each sample, and included in the statistical analysis as nuisance parameters.

The statistical analysis of the data employs a binned likelihood function constructed as the product of the likelihood terms for each category. The likelihood in each category is a product over bins in the distributions of the collinear mass or effective mass shown in Fig. 7.17.

The expected signal and background, as well as the observed number of events, in each bin of the mass distributions enter in the definition of the likelihood function $\mathscr{L}(\mu, \boldsymbol{\theta})$. A "signal strength" parameter ($\mu$) multiplies the expected signal in each bin. The signal strength is a free parameter in the fit procedure. The value $\mu = 0$ ($\mu = 1$) corresponds to the absence (presence) of a Higgs boson signal with the SM production cross-section.

Signal and background predictions ($s$ and $b$) depend on systematic uncertainties that are parametrised by nuisance parameters $\boldsymbol{\theta}$, which in turn are constrained using Gaussian functions. The correlation of the systematic uncertainties across categories are taken into account:

$$\mathscr{L}(\mu,\theta) = \prod_{\text{bin } j} \text{Poisson}(N_j|\mu(s_j)+b_j) \prod_{\theta} \text{Gaussian}(\theta|0,1). \tag{7.7}$$

The expected signal and background event counts in each bin are functions of $\boldsymbol{\theta}$. The parametrisation is chosen such that the rates in each channel are log-normally distributed for a normally distributed $\boldsymbol{\theta}$.

The test statistic $q_\mu$ is defined as:

$$q_\mu = -2\ln\left(\mathscr{L}(\mu,\hat{\boldsymbol{\theta}}_\mu)/\mathscr{L}(\hat{\mu},\hat{\boldsymbol{\theta}})\right), \tag{7.8}$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ refer to the global maximum of the likelihood (with the constraint $0 \leq \hat{\mu} \leq \mu$) and $\hat{\boldsymbol{\theta}}_\mu$ corresponds to the conditional maximum likelihood of $\boldsymbol{\theta}$ for a given $\mu$. This test statistic is used to compute exclusion limits following the modified frequentist method known as $\text{CL}_s$ [100]. The asymptotic approximation [55] is used to evaluate the Gaussian probability density functions rather than performing pseudo-experiments and the procedure has been validated using ensemble tests.

The profile likelihood formalism used in this statistical analysis incorporates the information on the observed and expected number of events, nuisance parameters, probability density functions and parameters of interest. The statistical significance of an excess is evaluated in terms of the same profile likelihood test statistic. The expected sensitivity and the $\pm 1,2\,\sigma$ bands are obtained for the background expectation in the absence of a Standard Model Higgs boson signal. The consistency with the background-only hypothesis is quantified using the $p$-value, the probability that the test statistic of a background-only experiment fluctuates to at least the observed one.

No significant excess is observed in the data compared to the SM expectation. Exclusion limits at the 95% confidence level, normalised to the Standard Model cross section times the branching ratio of $H \to \tau^+\tau^-$ ($\sigma_{\text{SM}}$), are set as a function of the Higgs boson mass. The limit plot is shown in Fig. 7.26.

The expected limits vary between 7 and 17 times the predicted Standard Model cross section times branching ratio for the mass range $100-150$ GeV. The most sensitive categories is the VBF category, that is limited by statistics in the 2011 analysis. The observed limits are in the range between 5 and 15 times the predicted Standard Model cross section times branching ratio for the same mass range. No significant deviation from the background-only hypothesis is observed.

## 7.9 Outlook

This analysis brought a large improvement in the $H \to \tau\tau$ analysis via the consistent treatment of all production and decay processes considered, and by using their characteristics to enhance the sensitivity to the signal. Selecting events according to the jet multiplicity in the final state increases the rejection power with respect to a simple, inclusive analysis.

The analysis performed in 2011 is still limited by statistics in its most sensitive bin, the 2-jet VBF category. The possibility to use a central jet veto has not been used yet, because of the low statistics in the bin before the eventual veto; with higher statistics, it would provide an even better separation of signal from background. The JVF, used in the jet selection, ensures a reliable operation of the

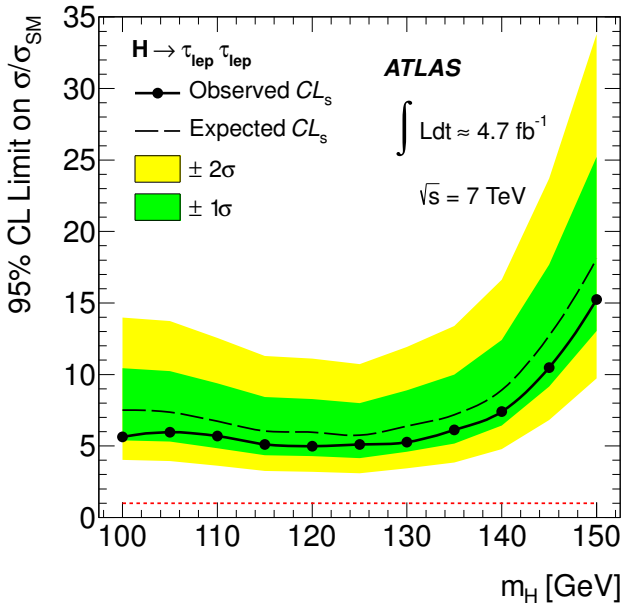



Figure 7.26: Expected (dashed) and observed (solid) 95% confidence level upper limits on the Higgs boson cross section times branching ratio of $H \to \tau^+\tau^-$, normalised to the Standard Model expectation, as a function of the Higgs boson mass. Expected limits are given for the scenario with no signal. The bands around the dashed line indicate the expected statistical fluctuations of the limit.

central jet veto, as shown in Sect. 7.4.1.

Additional studies are ongoing. First of all, the new data taking conditions (the increase in energy and luminosity in 2012 data) require a careful re-investigation of the selection cuts. This could also be an opportunity to improve the sensitivity. This investigation could also be done on 2011 data, to check on an already well established dataset the kind of improvements that could be achieved.

Additional categories can be added to the investigation. The aim of these additional categories would be to introduce additional discrimination criteria between the signal and the background. As I have remarked, the background composition is different in the different categories; this means, that the more exclusive a category is, the better it could reject a background process. For example, to reject the $Z \to \tau\tau$ background the categories containing events with at least one jet could be divided into high- and low-Higgs-$p_{\mathrm{T}}$ categories. Because the Higgs $p_{\mathrm{T}}$ distribution is expected to show a larger tail with respect to the $Z$ $p_{\mathrm{T}}$ distribution, selecting those events at high di-tau $p_{\mathrm{T}}$ and separate them in an independent category would enhance the global sensitivity without affecting the sensitivity of the existing categories. Another important improvement would be to recover also the $WH$ events, where the Higgs is produced in association with a $W$ boson, in the analysis. Those events where at the beginning neglected, because the need to achieve a high precision in selecting such a challenging signature at the LHC leads to a very low efficiency of the selection. With the increase in luminosity in 2012, however, this channel should provide enough statistics, and be useful for the combination.

Improvements could come also from the background estimation. A data-driven technique is used in 2011 to cross-check the top normalization in data: this could be extended to provide a full data-driven background estimation of the top background. The fake-leptons contribution

is already estimated data-driven, and an improvement of this technique is needed to reduce its systematic uncertainty, that dominates the background uncertainty. For the $Z \to \ell\ell$ background, it is already under control; further studies of the $Z \to \ell\ell$ + jets background could anyway lower the uncertainty on the normalization, by providing more precise predictions, and this would also impact the normalization of the $Z \to \tau\tau$ background. For the embedding procedure, development and improvements will also need to take into account the changes in the data-taking conditions, that affect the sample selection.

Finally, an improvement planned in ATLAS is to unify the mass definitions used in the analysis. The $H \to \tau\tau \to \tau_h\ell$ analysis uses already a sophisticated mass estimator, the *Missing Mass Calculator* [58]. Such mass estimator achieves a better mass resolution with respect to the mass estimators used in the $H \to \tau\tau \to \ell\ell + 4\nu$ analysis, and, as already mentioned, the mass resolution is crucial for the extraction of the signal. Several studies in ATLAS are ongoing, related to the use and calibration of such estimator, and a non-negligible improvement is expected in 2012 also from the implementation of this estimator.

# Chapter 8

# Outlook

An observed excess of events, corresponding to the hypothesis of a new particle with mass $126 \pm 0.4$ (stat.) $\pm 0.4$ (syst.), has now (July 2012) been reported by the ATLAS collaboration [21]. Also the CMS collaboration at LHC has reported a similar result [110]. The results in the $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ and $H \to WW^* \to \ell\nu\ell\nu$ channels are compatible with a Higgs boson hypothesis, the observed deviations from the SM expectations being still small to be statistically significant. However, many measurements are needed in order to assess the nature of this particle; and further developments, starting from the results presented in this work, will be very important in this effort.

**$H \to \tau\tau$ studies in 2012** The hypothesis of a Higgs boson could not be verified without checking that this particles decays into fermions. With respect to earlier studies [57], that predicted a visible $H \to \tau\tau$ at higher integrated luminosity and only in the VBF channel, the 2011 ATLAS analysis has shown that the $H \to \tau\tau$ analysis can provide an input to the Higgs combination already with a low integrated luminosity. The increase in luminosity will definitely provide the possibility to improve the limit and approach the $\sigma/\sigma_{\mathrm{SM}}$ level, with the possibility of setting direct limits from this channel alone. Also, the possible observation of an excess of events could be expected.

Observing or not an excess would be a fundamental input to the current Higgs studies. The $\tau^+\tau^-$ channel is the cleanest channel that provides a Higgs decay to fermions: observing it is fundamental both in the study of the Higgs characteristics in general, as well as for understanding the Yukawa coupling of the Higgs to fermions, that is included "ad-hoc" in the SM to reproduce the particle masses. For this reason, the $H \to \tau\tau$ analysis is expected to play a very important role within the Higgs analyses. Not only the SM Higgs analysis will be affected: many BSM scenarios predict a different coupling of the Higgs to the tau lepton. In general, measuring any possible deviation from the SM expectation of the Higgs decay rate would be the only way at the LHC to get informations about the total Higgs width around $m_H = 126$ GeV, and on any possible new physics process that could influence the observed values.

Neglecting or not carefully estimating the uncertainties would yield, in the case of negative search results, to an overestimate of the excluded regions of the parameter space. In case of a Higgs signal these uncertainties are crucial to perform a reliable and accurate determination of $m_H$ and of the Higgs boson couplings. As already mentioned in the thesis, both experimental and theoretical uncertainties need to be carefully understood and estimated to ensure the success of the analysis.

**Development in modelling and uncertainties** Experimental uncertainties have been reduced during the first two LHC runs, thanks to a precise estimate of the performance. Still, some of these uncertainties are dominating the analyses, and limit the possibility for additional improvements. One of such examples is the $Z \to \ell\ell$ + jets analysis, where the main limit is now the jet energy scale uncertainty. The first results, documented here, from the $Z \to \ell\ell$ + jets cross section measurement

showed that the current understanding of the process, provided by matrix element generators interfaced with a parton shower generator is satisfactory; but deviations have been observed, mostly in the di-jet quantities shown in Fig. 6.13, that are at the moment under investigation, both in the ATLAS collaboration and by the authors of the Monte Carlo generators.

Another experimental improvement in the JES calibration has been brought in 2011 by the implementation of the *in-situ* calibrations. To keep detector effects and physics effects well separated, and to have a well defined reference, the jets measured in data have been calibrated using correction factors derived from the data-Monte Carlo comparison. In this way, the calibration is still based on Monte Carlo, but the combination of the corrections coming from the *in-situ* techniques should better address several uncertainties of the older calibrations (e.g. the quark-gluon effects in the EM+JES, and the jet-level corrections in the LCW calibration). The improvements in 2011 have been already very large, with the largest uncertainty (the uncertainty at low $p_{\mathrm{T}}^{\mathrm{jet}}$ in the forward region) dropping to 6% from the previous 12%. But to get at the level of the other experimental uncertainties, the aim is to bring it down to 1%, and for this the current strategy needs still to be improved. Because of the large impact of the Monte Carlo predictions on the ATLAS jet energy scale determination, any improvement in the Monte Carlo modelling will propagate to the JES calibration and uncertainty.

One of the top priorities of the experiment, however, will be the understanding of the experimental conditions. In this thesis, the impact of pile-up effects on the measurement has been shown in several examples; and, with the LHC increasing in instantaneous luminosity, the need to update both the performance studies and the Monte Carlo simulations is crucial. In 2012, with a higher energy and higher pile-up conditions, the results have already been successful.

**Developments and the long-term goals** The theoretical modelling will still be one of the top priorities in the Higgs searches. The $Z \to \ell\ell$ + jets analysis has shown that the theoretical models available for that process are satisfactory; but the $Z \to \ell\ell$ + jets process is a process whose perturbative series converges quite well, in contrast with the Higgs production channels, in particular $gg \to H$[86]. The ATLAS strategy has carefully addressed this point, and several discussions have been ongoing between ATLAS and CMS experimentalists, and theorists. Any improvement or optimized selection of the phase space needs to be cross-checked with the theory uncertainty, in order not to fall into ill-defined regions. This effort will be even more important if the results from the different decay channels will be combined to extract the Higgs couplings: for such an analysis, a good knowledge of the production processes is needed. In the past, the need for such a precision has brought the collaborations to neglect some cases, as for example the $gg \to H$ production in the $H \to \tau\tau$ channel. A full combination of the LHC results will be an outstanding results for the scientific community, showing how the understanding of both the theoretical predictions and the experimental conditions have been successful at the LHC.

# Bibliography


[1] [Online]. Available: https://twiki.cern.ch/twiki/bin/view/AtlasPublic.

[2] 2010. [Online]. Available: atlas.ch.

[3] A. Ribon et al. (2010). Status of Geant4 hadronic physics for the simulation of LHC experiments at the start of the LHC physics program.

[4] *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN*, ser. LHC Tech. Proposal. Geneva: CERN, 2010.

[5] G. Aad et al., 'ATLAS Calorimeter Response to Single Isolated Hadrons and Estimation of the Calorimeter Jet Scale Uncertainty', CERN, Geneva, Tech. Rep. ATLAS-CONF-2011-028, 2011.

[6] —, 'Charged particle multiplicities in pp interactions for track $p_T > 100$ MeV at $\sqrt{s} =$ 0.9 and 7 TeV measured with the ATLAS detector at the LHC', CERN, Geneva, Tech. Rep. ATLAS-CONF-2010-046, 2010.

[7] —, 'Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data', CERN, Geneva, Tech. Rep. ATLAS-CONF-2011-102, 2011.

[8] —, 'Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data', Oct. 2011. eprint: 1110.3174v2. [Online]. Available: http://arxiv.org/abs/1110.3174v2.

[9] —, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics*. Geneva, 2008.

[10] —, 'Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} =$ 7 TeV', 2011. arXiv:1112.6426 [hep-ex].

[11] —, 'Luminosity Determination in pp Collisions at $\sqrt{(s)} = 7$TeV using the ATLAS Detector in 2011', CERN, Geneva, Tech. Rep. ATLAS-CONF-2011-116, 2011.

[12] —, 'Luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC', *The European Physical Journal C - Particles and Fields*, vol. 71, no. 4, p. 1630, 2011. DOI: 10.1140/epjc/s10052-011-1630-5.

[13] —, 'Measurement of the Mistag Rate with 5 $fb^1$ of Data Collected by the ATLAS Detector', CERN, Geneva, Tech. Rep. ATLAS-CONF-2012-040, 2012.

[14] —, 'Measurement of the W -> lnu and Z/gamma* -> ll production cross sections in proton-proton collisions at sqrt(s) = 7 TeV with the ATLAS detector', *JHEP*, vol. 12, p. 060, 2010. DOI: 10.1007/JHEP12(2010)060. arXiv:1010.2130 [hep-ex].

[15] —, 'Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb1 of Data from the ATLAS Detector', CERN, Geneva, Tech. Rep. ATLAS-CONF-2012-043, 2012.

[16] —, 'Measurement of the inclusive W+- and Z/gamma cross sections in the electron and muon decay channels in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector', 2011. arXiv:1109.5141 [hep-ex].

[17] —, 'Measurement of the production cross section for $Z/\gamma*$ in association with jets in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector', *Phys. Rev. D*, vol. 85, p. 032 009, 3 2012. DOI: 10.1103/PhysRevD.85.032009. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.85.032009.

[18] —, 'Measurement of underlying event characteristics using charged particles in $pp$ collisions at $\sqrt{s} = 900$ GeV and 7 TeV with the ATLAS detector', *Phys. Rev. D*, vol. 83, p. 112 001, 11 2011. DOI: 10.1103/PhysRevD.83.112001. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.83.112001.

[19] —, 'Muon Reconstruction Performance', CERN, Geneva, Tech. Rep. ATLAS-CONF-2010-064, 2010.

[20] —, 'Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector', CERN, Geneva, Tech. Rep. ATLAS-CONF-2011-063, 2011.

[21] —, 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC', Jul. 2012. eprint: 1207.7214v1. [Online]. Available: http://arxiv.org/abs/1207.7214v1.

[22] —, 'Performance of missing transverse momentum reconstruction in proton-proton collisions at 7 TeV with ATLAS', *Eur. Phys. J.*, vol. C72, p. 1844, 2012. DOI: 10.1140/epjc/s10052-011-1844-6. arXiv:1108.5602 [hep-ex].

[23] —, 'Performance of the ATLAS Trigger System in 2010', *Eur. Phys. J.*, vol. C72, p. 1849, 2012. DOI: 10.1140/epjc/s10052-011-1849-1. arXiv:1110.1530 [hep-ex].

[24] —, 'Procedure for the LHC Higgs boson search combination in summer 2011', CERN, Geneva, Tech. Rep. ATL-PHYS-PUB-2011-011, 2011.

[25] —, 'Response of isolated particles identified using resonances in proton-proton collisions at $sqrts$=7 TeV with the ATLAS detector', CERN, Geneva, Tech. Rep. ATLAS-CONF-2011-019, 2011.

[26] —, 'Response of the ATLAS calorimeters to single isolated hadrons produced in proton proton collisions at a center of mass energy of sqrts = 900 GeV', CERN, Geneva, Tech. Rep. ATLAS-CONF-2010-017, 2010.

[27] —, 'Search for neutral MSSM Higgs bosons decaying to $\tau^+\tau^-$ pairs in proton-proton collisions at sqrt(s) = 7 TeV with the ATLAS detector', *Phys. Lett.*, vol. B705, pp. 174–192, 2011. DOI: 10.1016/j.physletb.2011.10.001. arXiv:1107.5003 [hep-ex].

[28] —, 'Search for the Standard Model Higgs boson in the H to $tau^+tau^-$ decay mode in $\sqrt(s) = 7$TeV pp collisions with ATLAS', Jun. 2012. eprint: 1206.5971v1. [Online]. Available: http://arxiv.org/abs/1206.5971v1.

[29] —, 'Search for the Standard Model Higgs boson in the decay mode $H \to \tau^+\tau^- \to$ ll + 4 neutrinos in Association with jets in Proton-Proton Collisions at $\sqrt(s) = 7$ TeV with the ATLAS detector', CERN, Geneva, Tech. Rep. ATLAS-CONF-2011-133, 2011.

[30] —, 'Study of Jet Shapes in Inclusive Jet Production in pp Collisions at sqrt(s) = 7 TeV using the ATLAS Detector', *Phys. Rev.*, vol. D83, p. 052 003, 2011. DOI: 10.1103/PhysRevD.83.052003. arXiv:1101.0070 [hep-ex].

[31] —, 'The ATLAS Experiment at the CERN Large Hadron Collider', *JINST*, vol. 3, no. 08, S08003, August 2008.

[32] —, 'The ATLAS Simulation Infrastructure', *Eur. Phys. J.*, vol. C70, pp. 823–874, 2010. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv:1005.4568 [physics.ins-det].

[33] T. Adye, 'Unfolding algorithms and tests using roounfold', May 2011. eprint: 1105.1160v1. [Online]. Available: http://arxiv.org/abs/1105.1160v1.

[34] S. Agostinelli et al., 'Geant4 a simulation toolkit', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 –303, 2003, ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[35] N. Amelin", 'Transverse flow and collectivity in ultrarelativistic heavy-ion collisions', *Physical Review Letters*, vol. 67, no. 12, pp. 1523–1526, 1991. DOI: 10.1103/PhysRevLett.67.1523.

[36] N. Amelin et al., 'Collectivity in ultra-relativistic heavy ion collisions', *Nuclear Physics A*, vol. 544, no. 1â"2, pp. 463 –466, 1992, ISSN: 0375-9474. DOI: 10.1016/0375-9474(92)90598-E. [Online]. Available: http://www.sciencedirect.com/science/article/pii/037594749290598E.

[37] *AthenaROOTAccess*, documentation available from the ATLAS Protected wiki. [Online]. Available: twiki.cern.ch/twiki/bin/view/AtlasProtected/AthenaROOTAccess.

[38] M. Baak et al., 'Updated Status of the Global Electroweak Fit and Constraints on New Physics', 2011. arXiv:1107.0975 [hep-ph].

[39] M. Bahr et al., 'Herwig++ Physics and Manual', *Eur.Phys.J.*, vol. C58, pp. 639–707, 2008. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv:0803.0883 [hep-ph].

[40] R. D. Ball et al., 'Impact of heavy quark masses on parton distributions and LHC phenomenology', *Nucl.Phys.*, vol. B849, pp. 296–363, 2011. DOI: 10.1016/j.nuclphysb.2011.03.021. arXiv:1101.1300 [hep-ph].

[41] C. Berger et al., 'Vector Boson + Jets with BlackHat and Sherpa', *Nucl.Phys.Proc.Suppl.*, vol. 205-206, pp. 92–97, 2010. DOI: 10.1016/j.nuclphysbps.2010.08.025. arXiv:1005.3728 [hep-ph].

[42] J. Beringer et al., 'Review of particle physics', *Physical Review D*, vol. 86, no. 1, 2012. DOI: 10.1103/PhysRevD.86.010001.

[43] H. Bertini, 'Intranuclear-Cascade Calculation of the Secondary Nucleon Spectra from Nucleon-Nucleus Interactions in the Energy Range 340 to 2900 MeV and Comparisons with Experiment', *Physical Review*, vol. 188, no. 4, pp. 1711–1730, 1969. DOI: 10.1103/PhysRev.188.1711.

[44] P. Bolzoni et al., 'Higgs production via vector-boson fusion at NNLO in QCD', Mar. 2010. eprint: 1003.4451v2. [Online]. Available: http://arxiv.org/abs/1003.4451v2.

[45] M. Botje et al., 'The PDF4LHC working group interim recommendations', 2011. arXiv:1101.0538 [hep-ph].

[46] L. Bravina, 'Scaling violation of transverse flow in heavy ion collisions at AGS energies', *Physics Letters B*, vol. 344, pp. 49 –54, 1995, ISSN: 0370-2693. DOI: 10.1016/0370-2693(94)01560-Y. [Online]. Available: http://www.sciencedirect.com/science/article/pii/037026939401560Y.

[47] L. Bravina et al., 'Fluid dynamics and Quark Gluon string model: What we can expect for Au+Au collisions at 11.6 GeV/c', *Nuclear Physics A*, vol. 566, no. 0, pp. 461 –464, 1994, ISSN: 0375-9474. DOI: 10.1016/0375-9474(94)90669-6. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0375947494906696.

[48] O. Brein, A. Djouadi and R. Harlander, 'NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders', *Phys.Lett.*, vol. B579, pp. 149–156, 2004. eprint: hep-ph/0307206. [Online]. Available: http://arxiv.org/abs/hep-ph/0307206.

[49] J. Butterworth, J. R. Forshaw and M. Seymour, 'Multiparton interactions in photoproduction at HERA', *Z.Phys.*, vol. C72, pp. 637–646, 1996. DOI: 10.1007/s002880050286. arXiv:hep-ph/9601371 [hep-ph].

[50] Cacciari, Matteo and Salam, Gavin P., 'Dispelling the N**3 myth for the k(t) jet-finder', *Physics Letters B*, vol. 641, pp. 57–61, 2006.

[51] M. Cacciari, G. P. Salam and G. Soyez, 'The Anti-k(t) jet clustering algorithm', *JHEP*, vol. 0804, p. 063, 2008. DOI: 10.1088/1126-6708/2008/04/063. arXiv:0802.1189 [hep-ph].

[52] —, 'The catchment area of jets', *Journal of High Energy Physics*, vol. 2008, no. 04, p. 005, 2008. [Online]. Available: http://stacks.iop.org/1126-6708/2008/i=04/a=005.

[53] M. L. Ciccolini, S. Dittmaier and M. Krämer, 'Electroweak Radiative Corrections to Associated WH and ZH Production at Hadron Colliders', *Phys.Rev.*, vol. D68, p. 073003, 2003. eprint: hep-ph/0306234. [Online]. Available: http://arxiv.org/abs/hep-ph/0306234.

[54] G Corcella et al., 'Herwig 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)', *Journal of High Energy Physics*, no. 1, p. 010, Jan. 2001.

[55] G. Cowan et al., 'Asymptotic formulae for likelihood-based tests of new physics', *The European Physical Journal C - Particles and Fields*, vol. 71, no. 2, p. 1554, 2011. DOI: 10.1140/epjc/s10052-011-1554-0.

[56] G. D'Agostini, 'Bayesian reasoning versus conventional statistics in high energy physics', eprint: physics/9811046v1. [Online]. Available: http://arxiv.org/abs/physics/9811046v1.

[57] M. Duhrssen et al., 'Extracting Higgs boson couplings from CERN LHC data', *Physical Review D*, vol. 70, no. 11, 2004. DOI: 10.1103/PhysRevD.70.113009.

[58] A. Elagin et al., 'A New Mass Reconstruction Technique for Resonances Decaying to ditau', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 654, no. 1, pp. 481 –489, 2011, ISSN: 0168-9002. DOI: 10.1016/j.nima.2011.07.009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900211014112.

[59] R. Ellis et al., 'Higgs decay to $\pi^+\pi^-$: A possible signature of intermediate mass Higgs bosons at high energy hadron colliders', *Nuclear Physics B*, vol. 297, no. 2, pp. 221 –243, 1988, ISSN: 0550-3213. DOI: 10.1016/0550-3213(88)90019-3. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0550321388900193.

[60] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and Collider Physics*, ser. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge: Cambridge Univ. Press, 2003, Photography by S. Vascotto.

[61] F. Englert and R. Brout, 'Broken symmetry and the mass of gauge vector mesons', *Phys.Rev.Lett.*, vol. 13, pp. 321–323, 1964. DOI: 10.1103/PhysRevLett.13.321.

[62] G. Folger and J. Wellisch, 'String Parton Models in Geant4', eprint: nucl-th/0306007. [Online]. Available: http://arxiv.org/abs/nucl-th/0306007.

[63] S. Frixione and B. R. Webber, 'The MC@NLO 3.4 Event Generator', 2008. arXiv:0812.0770 [hep-ph].

[64] R. Gavin et al., 'FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order', Nov. 2010. eprint: 1011.3540v1. [Online]. Available: http://arxiv.org/abs/1011.3540v1.

[65] G Gaycken, S Psoroulas et al., *SVN Bonn framework area*. [Online]. Available: svnweb.cern.ch/trac/atlasgrp/browser/Institutes/Bonn/Zmumu/Overkill.

[66] T. Gleisberg et al., 'Event generation with SHERPA 1.1', *JHEP*, vol. 0902, p. 007, 2009. DOI: 10.1088/1126-6708/2009/02/007. arXiv:0811.4622 [hep-ph].

[67] P Golonka et al., 'The tauola-photos-F environment for the TAUOLA and PHOTOS packages, release II', *Computer Physics Communications*, vol. 174, no. 10, pp. 818–835, May 2006. DOI: DOI10.1016/j.cpc.2005.12.018.

[68] P. Golonka and Z. Was, 'PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays', *Eur.Phys.J.C*, vol. 45, pp. 97–107, 2006. eprint: hep-ph/0506026. [Online]. Available: http://arxiv.org/abs/hep-ph/0506026.

[69] D. Green, *At the leading edge: the ATLAS and CMS LHC experiments*. Singapore: World Scientific, 2010.

[70] G. Guralnik, C. Hagen and T. Kibble, 'Global conservation laws and massless particles', *Phys.Rev.Lett.*, vol. 13, pp. 585–587, 1964. DOI: 10.1103/PhysRevLett.13.585.

[71] M. Guthrie, R. A. Jr. and H. Bertini, 'Calculation of the capture of negative pions in light elements and comparison with experiments pertaining to cancer radiotherapy', *Nuclear Instruments and Methods*, vol. 66, no. 1, pp. 29 –36, 1968, ISSN: 0029-554X. DOI: 10.1016/0029-554X(68)90054-2. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0029554X68900542.

[72] R. V. Harlander and W. B. Kilgore, 'Next-to-Next-to-Leading Order Higgs Production at Hadron Colliders', eprint: hep-ph/0201206v2. [Online]. Available: http://arxiv.org/abs/hep-ph/0201206v2.

[73] P. W. Higgs, 'Broken symmetries and the masses of gauge bosons', *Phys.Rev.Lett.*, vol. 13, pp. 508–509, 1964. DOI: 10.1103/PhysRevLett.13.508.

[74] —, 'Broken symmetries, massless particles and gauge fields', *Phys.Lett.*, vol. 12, pp. 132–133, 1964. DOI: 10.1016/0031-9163(64)91136-9.

[75] —, 'Spontaneous symmetry breakdown without massless bosons', *Phys.Rev.*, vol. 145, pp. 1156–1163, 1966. DOI: 10.1103/PhysRev.145.1156.

[76] *HiggsValidation package - HSG4*, documentation available from the ATLAS Protected wiki. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HSG4Validation.

[77] Hugo W. Bertini and Miriam P. Guthrie, 'News item results from medium-energy intranuclear-cascade calculation', *Nuclear Physics A*, vol. 169, no. 3, pp. 670 –672, 1971, ISSN: 0375-9474. DOI: 10.1016/0375-9474(71)90710-X. [Online]. Available: http://www.sciencedirect.com/science/article/pii/037594747190710X.

[78] *Toward a standardization of jet definitions*, "FERMILAB-CONF-90-249-E", 1990.

[79] H. Ita et al., 'Precise Predictions for Z + 4 Jets at Hadron Colliders', Aug. 2011. eprint: 1108.2229v2. [Online]. Available: http://arxiv.org/abs/1108.2229v2.

[80] S JADACH, J. KUHN and Z WAS, 'TAUOLA - A LIBRARY OF MONTE-CARLO PROGRAMS TO SIMULATE DECAYS OF POLARIZED TAU-LEPTONS', *Computer Physics Communications*, vol. 64, no. 2, pp. 275–299, May 1991.

[81] B. P. Kersevan and E. Richter-Was, 'The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5', eprint: hep-ph/0405247v1. [Online]. Available: http://arxiv.org/abs/hep-ph/0405247v1.

[82] T. Kibble, 'Symmetry breaking in non-Abelian gauge theories', *Phys.Rev.*, vol. 155, pp. 1554–1561, 1967. DOI: 10.1103/PhysRev.155.1554.

[83] N. Kidonakis, 'Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production', Mar. 2011. eprint: 1103.2792v1. [Online]. Available: http://arxiv.org/abs/1103.2792v1.

[84] G. F. Knoll, *Radiation detection and measurement*, Chow, C.-Y., Ed. 1979.

[85] LHC Higgs Cross Section Working Group et al., 'Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables', *CERN-2011-002*, CERN, Geneva, 2011. arXiv:1101.0593 [hep-ph].

[86] —, 'Handbook of LHC Higgs Cross Sections: 2. Differential Distributions', *CERN-2012-002*, CERN, Geneva, 2012. arXiv:1201.3084 [hep-ph].

[87] H.-L. Lai et al., 'New parton distributions for collider physics', *Phys.Rev.*, vol. D82, p. 074 024, 2010. DOI: 10.1103/PhysRevD.82.074024. arXiv:1007.2241 [hep-ph].

[88] W Lampl et al., 'Calorimeter clustering algorithms: description and performance', CERN, Geneva, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, 2008.

[89] L. Lyons, D. Gibaut and P. Clifford, 'How to combine correlated estimates of a single physical quantity', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 270, no. 1, pp. 110 –117, 1988, ISSN: 0168-9002. DOI: 10.1016/0168-9002(88)90018-6. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0168900288900186.

[90] M. L. Mangano et al., 'ALPGEN, a generator for hard multiparton processes in hadronic collisions', *JHEP*, vol. 0307, p. 001, 2003. arXiv:hep-ph/0206293 [hep-ph].

[91] A. Martin et al., 'Parton distributions for the LHC', Jan. 2009. eprint: 0901.0002v3. [Online]. Available: http://arxiv.org/abs/0901.0002v3.

[92] —, 'Parton distributions for the LHC', *Eur.Phys.J.*, vol. C63, pp. 189–285, 2009. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv:0901.0002 [hep-ph].

[93] B. Mellado, W. Quayle and S. L. Wu, 'Prospects for the observation of a Higgs boson with $H \to \tau^+\tau^- \to l^+l^- \not{p}t$ associated with one jet at the LHC', *Phys.Lett.*, vol. B611, pp. 60–65, 2005. DOI: 10.1016/j.physletb.2005.02.021. arXiv:hep-ph/0406095 [hep-ph].

[94] D. W. Miller, S. Dong and A. Schwartzman, 'Measurement of Hadronic Event Shapes and Jet Substructure in Proton-Proton Collisions at 7.0 TeV Center-of-Mass Energy with the ATLAS Detector at the Large Hadron Collider. ', Presented 03 Jun 2011, PhD thesis, Stanford University, Menlo Park, CA, 2011.

[95] S. Moch et al., 'Theoretical status and prospects for top-quark pair production at hadron colliders', *Physical Review D*, vol. 78, no. 3, 2008. DOI: 10.1103/PhysRevD.78.034003.

[96] N. Möser, 'A Sensitivity Study for Higgs Boson Production in Vector Boson Fusion in the $H \to \tau\tau \to lh + 3\nu$ Final State with ATLAS', PhD thesis, University of Bonn, 2011.

[97] P. Nason, 'A New method for combining NLO QCD with shower Monte Carlo algorithms', *JHEP*, vol. 0411, p. 040, 2004. DOI: 10.1088/1126-6708/2004/11/040. arXiv:hep-ph/0409146 [hep-ph].

[98] S Psoroulas, 'Measurement of the jet energy resolution in ATLAS', CERN, Geneva, Tech. Rep. ATL-PHYS-PROC-2011-156, 2011.

[99] *Root homepage*. [Online]. Available: {root.cern.ch}.

[100] A. L. Read, 'Presentation of search results: the CL s technique', *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, no. 10, p. 2693, 2002. [Online]. Available: http://stacks.iop.org/0954-3899/28/i=10/a=313.

[101] C. Ruwiedel, N. Wermes and H. Ströher, 'Identification of hadronic $\tau$ decays using the $\tau$ lepton flight path and reconstruction and identification of jets with a low transverse energy at intermediate luminosities with an application to the search for the Higgs boson in vector boson fusion with the ATLAS experiment at the LHC.', Presented 22 Jun 2010, PhD thesis, University of Bonn, Bonn, 2010.

[102] Salam, Gavin P. and Soyez, Gregory, 'A practical seedless infrared-safe cone jet algorithm', *JHEP*, vol. 5, 2007.

[103] M. Schmitz, 'Sensitivity of the ATLAS Experiment to discover the Decay $H \to \tau\tau \to \ell\ell + 4\nu$ of the Standard Model Higgs Boson produced in Vector Boson Fusion', PhD thesis, University of Bonn, 2011. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:hbz:5N-25545.

[104] A. Sherstnev and R. Thorne, 'Parton Distributions for LO Generators', *Eur.Phys.J.*, vol. C55, pp. 553–575, 2008. DOI: 10.1140/epjc/s10052-008-0610-x. arXiv:0711.2473 [hep-ph].

[105] T. Sjostrand, S. Mrenna and P. Z. Skands, 'PYTHIA 6.4 Physics and Manual', *JHEP*, vol. 0605, p. 026, 2006. DOI: 10.1088/1126-6708/2006/05/026. arXiv:hep-ph/0603175 [hep-ph].

[106] P. Z. Skands, 'Tuning Monte Carlo Generators: The Perugia Tunes', May 2010. eprint: 1005.3457v4. [Online]. Available: http://arxiv.org/abs/1005.3457v4.

[107] M. Spira, 'HIGLU: A Program for the Calculation of the Total Higgs Production Cross Section at Hadron Colliders via Gluon Fusion including QCD Corrections', eprint: hep-ph/9510347. [Online]. Available: http://arxiv.org/abs/hep-ph/9510347.

[108] I. W. Stewart and F. J. Tackmann, 'Theory uncertainties for Higgs and other searches using jet bins', *Phys.Rev.*, vol. D85, p. 034011, 2012. DOI: 10.1103/PhysRevD.85.034011. arXiv:1107.2117 [hep-ph].

[109] D. Stump et al., 'Inclusive Jet Production, Parton Distributions, and the Search for New Physics', eprint: hep-ph/0303013v1. [Online]. Available: http://arxiv.org/abs/hep-ph/0303013v1.

[110] The CMS collaboration, 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC', Jul. 2012. eprint: 1207.7235v1. [Online]. Available: http://arxiv.org/abs/1207.7235v1.

[111] The CMS collaboration et al., 'Jet production rates in association with W and Z bosons in pp collisions at $\sqrt{s} = 7$TeV', *Journal of High Energy Physics*, vol. 2012, no. 1, p. 10, 2012. DOI: 10.1007/JHEP01(2012)010.

[112] P. Weber, 'ATLAS Calorimetry: Trigger, Simulation and Jet Calibration', PhD thesis, Kirchhoff-Institut für Physik (KIP), 2008. [Online]. Available: http://www.kip.uni-heidelberg.de/Veroeffentlichungen/details.php?id=1824.

[113] R. Wigmans, *Calorimetry: Energy Measurement in Particle Physics*, ser. International series of monographs on physics. Oxford: Clarendon Press, 2000.

# List of Figures


2.1 Scheme showing the elementary particles described in the Standard Model. . . . . 5
2.2 The "mexican hat" potential used in the Higgs field description. . . . . 7
2.3 Parton model description of a hard scattering event. . . . . 10
2.4 MSTW2008 Parton Distribution Function fit [91]. . . . . 11
2.5 Schematic diagram showing the meaning of fragmentation and hadronization in a parton shower (for illustrative purpose only). . . . . 12
2.6 Feynman diagram for production and production of 1 additional jet in the final state. 13
2.7 Latest results from ATLAS: a) the exclusion limit, the local p0 and the signal strength, and b) the p0 values in the different channels, characterizing the observed 5-sigma excess at 126.5 GeV[21]. For an explanation of the kind of plots shown, see Sect. 7.8. 16
2.8 Higgs boson production modes . . . . . 17
2.9 The SM Higgs production cross-section at $\sqrt{s} = 7$ TeV. [85] . . . . . 17
2.10 The SM Higgs branching ratios as a function of Higgs mass $M_H$ [86] . . . . . 18

3.1 Representation of LHC. . . . . 21
3.2 LHC dipole section, schematic drawing.[2] . . . . . 22
3.3 Luminosity plots for the 2010 and 2011 runs[1]. . . . . 24
3.4 Cut-away representation of the ATLAS Detector. [31] . . . . . 25
3.5 Drawing of the inner detector traversed by charged tracks [31]. . . . . 27
3.6 Cut-away representation of the calorimetry system. [31] . . . . . 28
3.7 Sketch of a LAr barrel module [31]. . . . . 29
3.8 Schematic pictures of the ATLAS Calorimeter system. The arrangement of the different detectors is shown, highlighting also the transition and dead material regions. [31] . . . . . 30
3.9 Cumulative amount of material, in units of interaction length, as a function of $|\eta|$, in front of the electromagnetic calorimeters, in the electromagnetic calorimeters themselves, in each hadronic compartment, and the total amount at the end of the active calorimetry. Also shown for completeness is the total amount of material in front of the first active layer of the muon spectrometer (up to $|\eta| < 3.0$). From [31]. . . . . 31
3.10 Cut-away representation of the Muon Spectrometer. [31] . . . . . 31

4.1 Schematic diagram showing the relationship between jet definitions at different levels. The colours represent the energy detected inside the calorimeter cells. . . . . 37
4.2 The *active area* [52] of jets reconstructed with different algorithms [51]. . . . . 39
4.3 VBF process. . . . . 41
4.4 Tagging jets kinematic distributions in VBF $H \to \tau\tau \to \mu^+\mu^- + 4\nu$ events (hypothesis $M_H = 120$ GeV) and background [9]. . . . . 42
4.5 $\eta$ distribution of the two tagging jets, before the veto. . . . . 44
4.6 Leading jet $p_T$ distribution, before the veto. . . . . 45
4.7 Second-leading jet $p_T$ distribution, before the veto. . . . . 45

4.8 $\Delta\eta$ distribution of the two tagging jets, before the veto. . . . 45
4.9 No-tagging jets $p_{\mathrm{T}}$ distribution, before the veto. . . . 46
4.10 $Z^*$ distribution of jets before the veto, for all algorithms. . . . 46
4.11 $Z^*$ distribution of jets after the veto, for all algorithms. . . . 46
4.12 Number of jets after the tagging jets selection, for AntiKt and Cone algorithms, with and without pile-up . . . . 47
4.13 Inclusive jet pt distribution after the tagging jets selection, for AntiKt and Cone algorithms, with and without pile-up . . . . 47
4.14 $Z^*$ distribution after the tagging jets selection, for AntiKt and Cone algorithms, with and without pile-up . . . . 48
4.15 Event display produced with ARA [37], showing the same VBF event reconstructed with the Cone and the anti$-K_T$ algorithms. . . . 50

5.1 Schematic representation of particles detected in the different ATLAS subdetectors according to their nature. . . . 53
5.2 Schematic development of an hadronic shower. . . . 54
5.3 Average simulated jet response at the EM-scale in bins of EM+JES calibrated jet energy and as a function of the detector pseudorapidity $\eta_{\mathrm{det}}$ [10]. . . . 55
5.4 Difference between the jet pseudorapidity calculated using an origin correction and the true jet pseudorapidity in bins of the calorimeter jet energy calibrated with the EM+JES scheme as a function of the detector pseudorapidity $|\eta_{\mathrm{det}}|$. From Aad et al. [10]. . . . 56
5.5 Average simulated jet energy response at the LCW scale in bins of the LCW+JES calibrated jet energy and as a function of the detector pseudorapidity $|\eta_{\mathrm{det}}|$. From Aad et al. [10]. . . . 57
5.6 Event display of a di-jet event in ATLAS 7 TeV data. . . . 59
5.7 Relative response from $\eta$-intercalibration in data, di-jet PYTHIA reconstructed events and di-jet PYTHIA particle-level jets (*Truth*) for a given $p_{\mathrm{T}}^{\mathrm{avg}}$ bin and the two calibration schemes (see Sect. 5.2.2). . . . 64
5.8 Relative jet response, $1/c$, of anti$-K_T$ jets with $R = 0.6$ as a function of the jet pseudorapidity measured using the matrix $\eta$-intercalibration method in bins of the average $p_{\mathrm{T}}$ of the two leading jets (a) $20 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 30$ GeV, (b) $30 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 45$ GeV, (c) $60 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 80$ GeV and $80 \leq p_{\mathrm{T}}^{\mathrm{avg}} < 110$ GeV. From Aad et al. [10]. . . . 65
5.9 Relative jet response, $1/c$, of anti$-K_T$ jets with $R = 0.6$ as a function of $p_{\mathrm{T}}^{\mathrm{avg}}$ found using the matrix $\eta$-intercalibration method for (a) $1.2 \leq |\eta| < 2.1$ and (b) $3.6 \leq |\eta| < 4.5$. From Aad et al. [10]. . . . 66
5.10 Average jet response for anti$-K_T$ jets with $R = 0.6$ calibrated with the EM+JES scheme measured relative to a central reference jet within $|\eta| < 0.8$ in data and various Monte Carlo generator samples as a function of $|\eta_{\mathrm{jet}}|$ for $p_{\mathrm{T}}^{\mathrm{avg}}$ in the ranges $30-45$ GeV (a) and $80-110$ GeV (b). The resulting systematic uncertainty component is shown as a shaded band around the data points. The errors bars on the data points only show the statistical uncertainties. From [10]. . . . 67

5.11 Fractional jet energy scale systematic uncertainty as a function of $p_T^{jet}$ for jets in the pseudorapidity region $0.3 \leq |\eta| < 0.8$ in the calorimeter barrel (a), $2.1 \leq |\eta| < 2.8$ in the calorimeter endcap (b), and in the forward pseudorapidity region $3.6 \leq |\eta| < 4.5$. The total uncertainty, derived from the sum of the single contributions as explained in the text, is shown as the solid light shaded area. The individual sources are also shown together with uncertainties from the fitting procedure if applicable. From Aad et al. [10]. . . . . 68

6.1 Feynman diagram for $Z$ boson production and production of 1 additional jet in the final state. . . . . 72
6.2 Di-muon distributions in (a) the inclusive sample and (b) in events with at least one associated jet. . . . . 78
6.3 Inclusive jet multiplicity (a) at detector level and (b) after unfolding [17]. . . . . 79
6.4 Rapidity of all jets considered in the $Z \to \ell\ell$ + jets analysis, at detector level. . . . . 80
6.5 Jet distributions from the 2010 analysis [17]. . . . . 82
6.6 $Z^*$distribution observed at detector level. . . . . 83
6.7 Comparison between the Monte Carlo sample used in the analysis and the same sample rerun with all jet energies scaled up and down by $1\sigma_{JES}$. . . . . 83
6.8 Comparison between nominal Monte Carlo and the same sample rerun with all jet energies smeared according to the two scenarios mentioned in Sect. 6.4.1. The bands show the JER uncertainty times 5. . . . . 85
6.9 Uncertainty of the measured differential cross section due to the jet energy resolution [17], measured as the relative systematic variation between the sample used in the analysis and the variated samples. . . . . 86
6.10 Particle level cross section results as function of jet multiplicity [17]. (a) shows the absolute cross section, (b) shows the cross section of each bin normalized to the value of the previous bin, showing the scaling of the cross section, expected from Eq. 2.10. . . . . 89
6.11 Particle level differential cross section results as function of (a) inclusive $p_T$, (b) leading jet $p_T$, and (c) 2nd leading jet $p_T$ [17]. . . . . 91
6.12 Particle level differential cross section results as function of (a) inclusive $|y|$, (b) leading jet $y$, and (c) 2nd leading jet $y$ [17]. . . . . 92
6.13 Particle level cross section results as function of (a) jet $\Delta y$ and (b) jet $\Delta$R [17]. . . . 93

7.1 Higgs boson production modes . . . . . 97
7.2 Comparison between $H \to \tau\tau \to \ell\ell + 4\nu$ and $Z \to \tau\tau \to \ell\ell + 4\nu$ . . . . . 98
7.3 Comparison between $H \to \tau\tau \to \ell\ell + 4\nu$ and $Z \to \tau\tau \to \ell\ell + 4\nu$ . . . . . 100
7.4 Jet distributions from the 2011 analysis. Only statistical uncertainties are shown. . . 105
7.5 Jet distributions from the 2011 analysis. Only statistical uncertainties are shown. . . 106
7.6 Detector level Jet-Vertex-Fraction (JVF) distribution in 2011. Only statistical uncertainties are shown. . . . . 107
7.7 Detector level inclusive rapidity distribution in period G-H in 2011. Only jets passing a cut JVF > 0.75 are considered. Only statistical uncertainties are shown. . . . . 108
7.8 $N_{jet}$ distribution within $|y| < 2.8$ in 2011, without and with a JVF cut applied. Only statistical uncertainties are shown. . . . . 109
7.9 $Z^*$ distributions from the 2011 analysis. Only statistical uncertainties are shown. . . 109
7.10 Comparison of the $m_{\ell\ell}$ distribution . . . . . 112

7.11 Comparison of the $H_{\mathrm{T}}^{\mathrm{Lep}}$ distributions . . . 113
7.12 Comparison of the leptons $\Delta\phi_{\ell\ell}$ . . . 113
7.13 Comparison of the leptons $x_1$ and $x_2$ distributions . . . 114
7.14 Comparison of $p_{\mathrm{T}}^{\mathrm{jet}}$ distribution . . . 114
7.15 Comparison of the jets $\Delta\eta_{jj}$ and $m_{jj}$ distributions . . . 115
7.16 Comparison of the $m_{\mathrm{H,jet}}$ distribution . . . 116
7.17 Comparison of the $m_{\tau^+\tau^-}$ distribution . . . 117
7.18 Flowchart of the embedding procedure. . . . 119
7.19 Example event displays of the embedding steps for a single $Z \to \mu\mu$ to $Z \to \tau\tau \to \tau_{\mathrm{h}}\tau_{\mathrm{h}}$ event: (a) after the selection, (b) after the Monte Carlo simulation, and (c) after the re-reconstruction. . . . 119
7.20 Comparison of the transverse momentum (left) and transverse mass distribution (right) from the selected data and the embedded $Z \to \mu\mu$ sample. From Schwindt (in preparation). . . . 121
7.21 Comparison of the missing transverse energy (left) and the total sum of energy (right) from the selected data and the embedded $Z \to \mu\mu$ sample. From Schwindt (in preparation). . . . 121
7.22 Comparison of the visible mass distribution from the $Z \to \tau\tau$ MC and the embedding sample, before the visible mass cut. . . . 123
7.23 Comparison of the transverse mass distribution from the $Z \to \tau\tau$ MC and the embedding sample, before the visible mass cut. . . . 123
7.24 Comparison of the mass distributions from the embedding sample, and the systematic variations, before the visible mass cut. . . . 124
7.25 Comparison of the missing transverse energy (left) and the total sum of energy (right) from the $Z \to \tau\tau$ MC and the embedding sample, before the visible mass cut. . . . 124
7.26 95% confidence level upper limits on the Higgs boson cross section times branching ratio . . . 127

# List of Tables


2.1 Tau lepton properties, from Beringer et al. [42]. . . . . 18

3.1 Fundamental parameters for the LHC machine. . . . . 23

4.1 Dataset used in the analysis, in sect. 4.3.3 and 4.3.4. . . . . 43
4.2 Lepton pre-selection used in the analysis. . . . . 43
4.3 Cutflow used in the analysis, with special care for tagging jets selection and central-jet-veto, used in sect. 4.3.3 and 4.3.4. . . . . 44
4.4 Efficiencies, as defined in eq. 4.4, for the VBF cutflow in the different scenarios and for different algorithms. . . . . 48

6.1 Integrated luminosities for the various data taking periods. . . . . 73
6.2 Monte Carlo samples used in the analysis. The cross-sections quoted are the ones used to normalize estimates of expected number of events. . . . . 75
6.3 Event selections for the $Z \to \mu^+\mu^- +$ jets analysis. . . . . 76

7.1 Cross-sections used in the analysis. . . . . 102
7.2 Cutflow used in the analysis, highlighting the most important cuts for defining the jet bins. . . . . 111
7.3 Number of events after the $H \to \tau\tau \to \ell\ell + 4\nu$ selection for the four categories in data and predicted number of background events, for an integrated luminosity of $4.7\,\text{fb}^{-1}$. Expectations for the Higgs boson signal ($m_H = 120\text{GeV}$) are also given. Statistical and systematic uncertainties are quoted, in that order. . . . . 116

# Acknowledgements

I would like to thank first of all Prof. Dr. Norbert Wermes, that gave me the opportunity to join his group and work in ATLAS in such an exciting period. I would also thank Dr. Jürgen Kroseberg, for his help and understanding, with all the troubles I have encountered in the last years, both in and outside work. They both have been extremely supportive in the hardest period of my life, and I could not express in words how grateful I am for this.

Several other people have helped me during these years of research, and I hope I will not forget some in this few lines of acknowledgements.

Thanks to the Bonn group, and in particular to Kristof Schmieden, Stephan Hageböck, Jan Theerhag, Dennis Hellmich and Klemens Müller for the coffee, cakes, and chats; concerning chats, I would also like to thank Marc Lehmacher for its patience, in the (not so) few cases our chats just next to his desk were a bit too long or noisy! And a huge thank you to Thomas Schwindt, for his help with the code developed for the Higgs analysis.

Thanks to the all the people from the OOSAS-AW[1] groups, those here in Bonn, and those based at CERN. My stay in both towns could have not been the same without you. I will bring with me a wonderful memory, and a much better knowledge of cinematography.

Thanks to my beta-readers - Jens Dopke and Nik Patel - that helped me pinning down the typos and unclear points of my thesis, not to mention how good friends they are outside work. Thank you also to the other people I have met during the last years, in summer schools and conferences: Valentina, Ferdinando, Francesca, Daniele, Paolo, Vincent, Michele, Tom . . . You made this period special for me. I love you.

A big thank you to my family, for their support, in the past as well as now. To Ms. Dr. Schubert, as her help has been fundamental to every thing I did in the last months.

Simo, “ce lo siamo presi” questo Ph.D., alla fine.

[1] Organization-of-social-activity-After-work