Thesis

Search for heavy di-boson resonances decaying to boosted hadronic final states in proton-proton collision at √s=8 TeV with the ATLAS detector

PICAZIO, Attilio

Abstract

Cette thèse présente la recherche de résonances étroites de di-bosons avec un état final de deux gerbes de particules. L'analyse est basée sur 20.3 fb-1 de données de collisions proton-proton avec une énergie au centre de masse de 8 TeV, enregistrées avec le détecteur ATLAS au grand collisionneur de hadron (Large Hadron Collider). Les limites d'exclusion à un niveau confiance à 95 % ont été dérivées sur la section efficace de production multipliée par le rapport d'embranchement pour les états finaux WZ d'un nouveau boson de jauge lourd, W', et pour les états finaux WW et ZZ d'excitations Kaluza-Klein du graviton dans un modèle Randall-Sundrum, en fonction de la masse de la résonance hypothétique. Les constantes de couplage des bosons W' prédites par le modèle de jauge étendue dans la région de masse de 1.3 TeV à 1.5 TeV sont exclus à un niveau de confiance de 95 %.

Reference

PICAZIO, Attilio. *Search for heavy di-boson resonances decaying to boosted hadronic final states in proton-proton collision at √s=8 TeV with the ATLAS detector*. Thèse de doctorat : Univ. Genève, 2016, no. Sc. 4938

URN : urn:nbn:ch:unige-878277
DOI : 10.13097/archive-ouverte/unige:87827

UNIVERSITÉ DE GENÈVE
Département de physique nucléaire et corpusculaire

FACULTÉ DES SCIENCES
Professeur Giuseppe Iacobucci

Search for heavy di-boson resonances decaying to boosted hadronic final states in proton-proton collision at $\sqrt{s}$=8 TeV with the ATLAS detector

THÈSE

Présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention physique

par

Attilio Picazio

d'Italie

Thèse N° 4938

GENÈVE
Atelier d'impression ReproMail
2016

Abstract

This thesis presents the search for narrow di-boson resonances in a dijet final state. The analysis is based on 20.3 fb^{-1} of proton-proton collision data at a centre-of-mass energy of 8 TeV recorded with the ATLAS detector at the Large Hadron Collider. The search is focused on di-boson resonances with masses in the range from 1.3 to 3.0 TeV. The expected signal is a narrow resonance on a smoothly falling invariant mass distribution of QCD dijet events.
The two selected jets are required to be tagged as boson jets. A vector-boson tagging procedure has been implemented using jet filtering techniques, as well as jet mass and substructure properties. The largest deviation from the background-only hypothesis in the observed dijet invariant mass distribution occurs around 2 TeV in the WZ channel, with a global significance of 2.5 standard deviations. In order to exclude systematic effects that might produce statistically significant deviations in the QCD dijet invariant mass spectrum, extensive cross-checks have been performed on different areas of the analysis, i.e. data-taking and detector effects, jet reconstruction and boson tagging, and event selection. Exclusion limits at the 95 % confidence level have been set on the production cross section times branching ratio as a function of the resonance mass for the WZ final state of a new heavy gauge boson, W', and for the WW and ZZ final states of Kaluza-Klein excitations of the graviton in a bulk Randall-Sundrum model. W' bosons with couplings predicted by the Extended Gauge Model in the mass range from 1.3 to 1.5 TeV are excluded at 95 % confidence level.

Résumé

Cette thèse présente la recherche de résonances étroites de di-bosons avec un état final de deux gerbes de particules. L'analyse est basée sur 20.3 fb^{-1} de données de collisions proton-proton avec une énergie au centre de masse de 8 TeV, enregistrées avec le détecteur ATLAS au grand collisionneur de hadron (Large Hadron Collider).
La recherche du signal est concentrée sur les résonances de di-bosons pour des masses allant de 1.3 TeV à 3.0 TeV. Le signal attendu est une résonance étroite sur une distribution de masse invariante des événements QCD à deux gerbes chutant fortement .
Les deux gerbes sélectionnés sont requis d'être identifiées étant produite par des bosons vecteurs. Une procédure de marquage de boson vecteur a été implémentée en utilisant des techniques de filtrage de gerbes, les propriétés de structure sous-jacente et la masse de la gerbe. Sous l'hypothèse d'absence de signal la déviation statistique la plus importante dans le spectre de masse invariante, par rapport à l'estimation du bruit de fond, se manifeste autour de 2 TeV dans le canal WZ avec un variance globale de 2.5 écarts types.
Dans l'optique d'exclusion des effets systématiques qui pourraient produire des déviations statistiquement significatives dans le spectre de masse invariant des gerbes de QCD, des vérifications extensives ont été effectuées sur différentes parties de l'analyse, *i.e.* relatifs à la prise de données, aux effets du détecteur, à la reconstruction des gerbes et leur marquage d'origine de boson et à la sélection d'événements. Les limites d'exclusion à un niveau confiance à 95 % ont été dérivées sur la section efficace de production multipliée par le rapport d'embranchement pour les états finaux WZ d'un nouveau boson de jauge lourd, W', et pour les états finaux WW et ZZ d'excitations Kaluza-Klein du graviton dans un modèle Randall-Sundrum, en fonction de la masse de la résonance hypothétique. Les constantes de couplage des bosons W' prédites par le modèle de jauge étendue dans la région de masse de 1.3 TeV à 1.5 TeV sont exclus à un niveau de confiance de 95 %.

Personal contributions

Due to the nature of particle physics experimental collaborations, the work in this thesis relies on the results of a large number of individual researchers and analysis subgroups. The leading contributions of the author to the work presented in this thesis are outlined below, subdivided by thesis chapter.

Chapter 6: Signal Monte Carlo samples validation and generation. The author was responsible of the production and validation of the EGM W' signal samples, using a MCValidation software, written and deployed during his PhD.

Chapter 6: Evaluation of jet systematic uncertainties using double-ratio technique. The author studied the jet related systematic uncertainties using the *in-situ* double-ratio technique. These results have been validated by the collaboration and represent one of the main source of systematic uncertainty for the analysis presented.

Chapter 6: Jet cleaning selection. The author studied an additional jet cleaning selection in order to treat the case of jets falling inside problematic regions of ATLAS hadronic calorimeter.

Chapter 7: Studies on systematic effects on jet reconstruction introduced by detector anisotropies. The author compared calorimeter jets, groomed track jets and standard small-radius jets to exclude any possible source of detector effect due to calorimeter anisotropies or to the jet filtering procedure.

Chapter 7: Studies on systematics effect on jet reconstruction introduced by data-taking instabilities or event selection. The author compared the full event selection in portions of data recorded in different periods, in order to exclude excesses of events caused by issues in the detector conditions. The author also made additional studies on the correlation of the event selection requirements.

Chapter 7: Limits setting procedure. The author worked on the frequentist interpretation of the results (CLs) using RooStats.

Contents

Introduction

Although the Standard Model (SM) predictions provide a very solid framework for experimental particle physics, there are still several missing pieces and unanswered questions that motivate searches for physics beyond the SM (BSM). Many extensions of the SM, such as Randall-Sundrum (RS) models with warped extra dimensions, Technicolor, and Grand Unified Theories, are predicted to manifest themselves as heavy resonances decaying to vector boson pairs.
The new energy regime explored at the LHC enables the production of Lorentz-boosted heavy particles, whose hadronic decay products can be reconstructed as one large-area jet. The opportunity of searching for fully hadronic signatures, if in control of the large QCD background, can enhance the sensitivity of these searches in the very high energy regime. This kind of final states presents many experimental challenges, because W and Z vector bosons produced at very high transverse momentum will often be reconstructed as a single jet (Lorentz-boosted topology). A fine granularity of the electromagnetic and hadronic calorimeters combined with highly-performant jet substructure techniques are the key to detect such kind of signatures and are extensively used here for the analysis of the 8 TeV collision data recorded by the ATLAS detector in 2012.

This thesis is focused on the search for narrow diboson resonances that should appear as a narrow peak in the di-jet invariant mass spectrum. The dominant background for this search is due to di-jet events from QCD processes. The application of jet substructure for jet grooming and boson tagging of the two most energetic jets in the events is able to significantly suppress the huge QCD di-jet background and provides high sensitivity for BSM resonances. Chapter 1 summarizes the key concepts of the SM and provides a description of the BSM models that are tested in this thesis. The main characteristics of the LHC accelerator system and of the ATLAS experiment are discussed in Chapter 2. In Chapter 3 and 4 the main concepts of QCD and jet reconstruction are provided. The details of boosted-boson-tagging techniques are described in Chapter 5. Chapters 6 and 7 are dedicated to the search for heavy di-boson resonances decaying to boosted hadronic final states. Conclusions are drawn in Chapter 8.

Chapter 1

Theory framework

The goal of particle physics is to identify the elementary components of matter and to understand the nature of the forces acting between them. The Standard Model (SM) of particle physics describes the strong, electromagnetic, and weak forces and provides our best explanation of the high energy physics phenomenology. Although it has been defined as the most successful theory of particle physics to date, the SM is not perfect and is an incomplete theory, because there are several physics phenomena in nature that it does not adequately explain. In the past years various extensions of the SM have been proposed.
In this chapter, a brief description of the SM will be presented, from the gauge principle to the description of the spontaneous symmetry breaking mechanism and the Higgs boson discovery from the ATLAS and CMS collaborations. A description of the Bulk Randall-Samdrum Graviton and of the Extended Gauge Model W' will be also provided.

1.1 The gauge principle

The four fundamental interactions can be described in terms of gauge theories. In particular, all can be derived from a principle, the *gauge principle*, introduced by H. Weyl in 1929 [1]. Physicists often describe system dynamics using *symmetry principles*. A symmetry is a mathematic operation (*trasformation*) that, if applied to the Lagrangian of a particular physics system, does not change the observables of the system. Lagrangians satisfying this requirement are referred to as *gauge invariant*. The term gauge refers to redundant degrees of freedom in the Lagrangian. A crucial distinction exists between *global* and *local* gauge invariance. In a global invariance the same transformation is carried out for all space-time points: it has an "everywhere-simultaneous" character. In a local invariance different transformations act on different individual space-time points. In general, a globally invariant Lagrangian will not be locally invariant. These ideas will be discussed in the context of the non-relativistic quantum mechanics and then generalised to the quantum field theory (QFT).

1.1.1 Gauge invariance in quantum mechanics

The Schrödinger equation for a particle of charge q and mass m moving in an electromagnetic field is:

$$\left(\frac{1}{2m}(-i\nabla - q\mathbf{A})^2 + qV\right)\psi(\mathbf{x},t) = i\frac{\partial\psi(\mathbf{x},t)}{\partial t} \tag{1.1}$$

The solution $\psi(\mathbf{x},t)$ of the equation describes completely the state of the particle moving under the influence of the potentials V and $\mathbf{A}$. These potential are not unique because they can be changed by a gauge transformation:

$$V \to V' = V - \partial\chi/\partial t \tag{1.2}$$
$$\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \nabla\chi \tag{1.3}$$

and the Maxwell equations for the fields $\mathbf{E}$ and $\mathbf{B}$ will remain the same. Including the gauge transformed electromagnetic potentials in the Schrödinger equation (1.1), a new equation is obtained with solution:

$$\psi'(\mathbf{x},t) \neq \psi(\mathbf{x},t) \tag{1.4}$$

Unlike Maxwell's equations the Schrödinger equation is not gauge invariant. While $\mathbf{E}$ and $\mathbf{B}$ are directly observable quantities, only the squared modulus of the wave function, $|\psi|^2$, is a real number interpreted as the probability density of measuring a particle as being at a given place at a given time or having a definite momentum, and possibly having definite values for discrete degrees of freedom. This means that ψ does not need to remain unchanged when the potentials are changed by gauge transformations.
The crucial point for quantum mechanics is to allow a change in ψ, $\psi(\mathbf{x},t) \to \psi'(\mathbf{x},t)$, when a gauge transformation is applied to the Maxwell potentials. The new equation is:

$$\left(\frac{1}{2m}(-i\nabla - q\mathbf{A}')^2 + qV'\right)\psi'(\mathbf{x},t) = i\frac{\partial\psi'(\mathbf{x},t)}{\partial t}. \tag{1.5}$$

The *form* of (1.5) is exactly the same as (1.1), and this will effectively assure that both "*describe the same physics*". The equation (1.1) is *gauge covariant*, meaning that it maintains the same form under a gauge transformation.
Knowing the transformation for the electromagnetic potentials V and $\mathbf{A}$, $\psi'(\mathbf{x},t)$ can be derived in order that the equation (1.1) is consistent with (1.5). The required wave function is:

$$\psi'(\mathbf{x},t) = \exp[iq\chi(\mathbf{x},t)]\psi(\mathbf{x},t) \tag{1.6}$$

where χ is the same space-time-dependent function appearing in (1.2) and (1.3). This can be demonstrated if the action of the operator $(-i\nabla - q\mathbf{A}')$ on $\psi'(\mathbf{x},t)$ is considered. Although the space-time-dependent phase factor feels the action of the gradient operator ∇, it passes through the combined operator and by defining the following operator combinations:

$$\mathbf{D} \equiv \nabla - iq\mathbf{A} \tag{1.7}$$
$$D^0 \equiv \partial/\partial t + iqV \tag{1.8}$$

the relation

$$(-i\nabla - q\mathbf{A}')\psi' = \exp(iq\chi)(-i\nabla - q\mathbf{A}')\psi \tag{1.9}$$

may be written as:

$$(-i\mathbf{D}'\psi') = \exp(iq\chi)(-i\mathbf{D})\psi \tag{1.10}$$

$\mathbf{D}'\psi'$ has the same relation with $\mathbf{D}\psi$ of the one that ψ' has with ψ, therefore:

$$(iD^{0\prime}\psi') = \exp(iq\chi)(iD^0)\psi, \tag{1.11}$$

where equation (1.2) has been used for V'. Repeating the operation which led to equation (1.10), this leads

$$\begin{aligned} \frac{1}{2m}(-i\mathbf{D}')^2\psi' &= \exp(iq\chi)\cdot\frac{1}{2m}(-i\mathbf{D})^2\psi \\ &= \exp(iq\chi)\cdot iD^0\psi \\ &= iD^{0\prime}\psi' \end{aligned} \tag{1.12}$$

Equation (1.12) is just (1.5) written using the D operator. Thus (1.6) is the correct relationship between ψ' and ψ to ensure consistency between equations (1.1) and (1.5): hence equation (1.1) is gauge invariant.
The probability densities $|\psi|^2$ and $|\psi'|^2$ are equal, since the two wave functions are related by a *phase* transformation. However the current, which is $\psi^*(\nabla\psi) - (\nabla\psi)^*\psi$ and that involves derivative operators, is not invariant under the transformation (1.6). This happens because the phase $\chi(\mathbf{x}, t)$ is $\mathbf{x}$-dependent. But equations (1.10) and (1.11) show that to construct *gauge-invariant currents*, the operator ∇ must be replaced by $\mathbf{D}$ (and in general also $\partial/\partial t$ by D^0), so:

$$\psi^{*\prime}(\mathbf{D}'\psi') = \psi^* \exp(-iq\chi)\cdot\exp(iq\chi)\cdot(\mathbf{D}\psi) = \psi^*\mathbf{D}\psi \tag{1.13}$$

Thus the identity between the physics described by ψ and ψ' is ensured and the *equality* between the first and last terms in (1.13) is indeed a proof of *(gauge) invariance.*
These considerations are summarized by the statement that the gauge invariance of the Maxwell equations re-emerges as covariance in quantum mechanics provided that the combined transformations of the potentials and of wave function are made.
As already mentioned, the Schrödinger equation is non-relativistic, but the prescriptions showed

here are valid in a relativistic regime. The D operators can be written in a manifestly covariant form as

$$D^\mu \equiv \partial^\mu + iqA^\mu \tag{1.14}$$

where $\partial^\mu = (-\nabla, \frac{1}{c}\frac{\partial}{\partial t})$ and $A^\mu = (\mathbf{A}, \frac{V}{c})$. (1.10) and (1.11) become then

$$-iD'^\mu \psi' = \exp(iq\chi) \cdot (-iD^\mu \psi). \tag{1.15}$$

It follows that any equation involving the operator ∂^μ can be made gauge invariant under the combined transformation

$$\begin{aligned} A^\mu &\rightarrow A'^\mu = A^\mu - \partial^\mu \chi \\ \psi &\rightarrow \psi' = \exp(iq\chi)\psi \end{aligned}$$

if ∂^μ is replaced by D^μ. Using this prescription, the wave equation for a particle in the presence of an electromagnetic field is obtained from the corresponding *free particle* wave equation.
The next Section will show how this mathematical operation is the basis of the so-called "gauge principle" whereby the form of the *interaction* is determined by the strong requirement of (local) gauge invariance.
In conclusion, the new derivative operator (1.14) is of fundamental importance. It generalizes from the (Abelian) phase symmetry of Quantum Electro-Dynamics (QED) to the (non-Abelian) phase symmetry of weak and strong interaction theories and it is called *gauge covariant derivative.*

1.1.2 The gauge principle

In the previous Section, the Schrödinger equation for a charged particle in an electromagnetic field (1.1) was written, and its gauge invariance under the combined transformation of electromagnetic potentials and of the wave function were checked. In this section, the argument will be reversed requiring that the theory is invariant under the *space-time-dependent phase transformation*

$$\psi(\mathbf{x}, t) \rightarrow \psi'(\mathbf{x}, t) = \exp[i\alpha(\mathbf{x}, t)]\psi(\mathbf{x}, t) \tag{1.16}$$

The free-particle Schrödinger equation is:

$$\frac{1}{2m}(-i\nabla)^2\psi(\mathbf{x}, t) = i\frac{\partial\psi(\mathbf{x}, t)}{\partial t} \tag{1.17}$$

This equation is invariant under a global phase transformation of the wave function $\psi(\mathbf{x}, t)$, because both ∇ and $\partial/\partial t$ do not act on $\alpha =$ *constant* in the phase factor. This is not true for the local phase transformation (1.16), where $\alpha = f(\mathbf{x}, t)$. Thus local phase invariance is not an invariance of the free-particle wave equation. To satisfy the strong requirement of a local

phase invariance, the equation (1.17) needs to be modified into something for which there is a corresponding covariance. The covariance will manifest in the inability to distinguish observationally between the effect of making a local change in phase convention and the effect of some new field in which the particle moves.
In order to understand what is the field that can allow a local phase covariance in the described theory, the concept of gauge invariance, discussed in Sec. 1.1.1, needs to be considered. The local phase transformation (1.16) with $\alpha = q\chi$ is just the transformation associated with electromagnetic gauge invariance. Thus the Schrödinger equation (1.17) must be modified to (1.5), satisfying the local phase invariance by demanding that $\mathbf{A}$ and V transform as follows:

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + q^{-1}\nabla\alpha \tag{1.18}$$
$$V \rightarrow V' = V - q^{-1}\partial\alpha/\partial t \tag{1.19}$$

when $\psi \rightarrow \psi'$. The modified wave equation is precisely the Schrödinger equation describing the interaction of the charged particle with the electromagnetic field described by $\mathbf{A}$ and V.
The vector and the scalar electromagnetic potentials can be regarded as part of a 4-vector A^μ, just as the operator $-\nabla$ and $\partial/\partial t$ are part of the operator ∂^μ. Thus the presence of the vector field A^μ, interacting in a "universal" prescribed way with any particle of charge q, is dictated by local phase invariance. A vector field introduced to guarantee local phase invariance, is called *gauge field*. The principle stating that the interaction should be generated by the phase (or gauge) invariance is called the *gauge principle*. This fundamental principle allows to write the wave equation for the interaction directly via the replacement

$$\partial^\mu \rightarrow D^\mu \equiv \partial^\mu + iqA^\mu \tag{1.20}$$

in the free-particle equation.

1.2 The Standard Model

According to classical physics, matter and force are clearly separated. Nature of matter is intuitive, based on everyday macroscopic experience; force, however, is more problematic. The dualism between matter and force can be overcome in the context of the modern quantum field theory (QFT).
In 1935, Yukawa proposed a theory describing the strong interaction between a proton and a neutron, and also considered its possible extension to neutron β-decay. He built his theory by analogy with electromagnetism, postulating a new force with an associated new field quantum similar to the photon. In doing so, he showed that, in quantum field theory, *particles interact by exchanging virtual quanta*, which *mediate* the force.
The fundamental interactions discovered and studied in detail so far by physics are four: strong, electromagnetic, weak and gravitational. The reduced strength of the gravitational interaction produces negligible effects on the behavior of sub-atomic particles, and plays no role in determining the internal properties of matter.
QFT represents the fundamental formal and conceptual framework of the SM, which describes both particles and forces in term of quantum fields. The particles are described by semi-integer

spin fields (fermions), and forces by integer spin fields (bosons). Depending on the way the elementary particles interact, they can be divided into two families: quarks and leptons. Both families are composed by spin 1/2 (in $\hbar$ unit) particles [1] that can be arranged in three generations, as it is shown in Table 1.1. Leptons are subject to weak and electromagnetic forces (except neutrinos which are neutrals and so are pure probes of the weak interaction), while quarks can interact with each other also via the strong force, creating bound states, the hadrons.
In the SM, weak, electromagnetic and strong interactions are described in terms of a local gauge symmetry group:

$$SU(3)_{col} \otimes SU(2)_L \otimes U(1)_Y \tag{1.21}$$

This set of local transformations is a set of *gauge* transformations, meaning that to preserve symmetry under a given transformation it requires the introduction of additional fields via the application of the gauge principle (Sec. 1.1.2). These are spin 1 gauge fields in the Lagrangian, that are associated with new particles. In particular, $SU(3)_{col}$ is a non abelian[2] gauge symmetry group which describes the strong interactions. The generators of the symmetry group are eight independent matrices. This geometrical relation reflects the fact that the strong interaction is carried by eight massless particles, the gluons. In this interpretation the gluons are the mediators or the strong interaction and they have a strong charge, known as "color". Gluons and quarks strong interactions are described by the Quantum Cromodynamic (QCD) theory. $SU(2)_L \otimes U(1)_Y$ is the weak isospin symmetry group, introduced by Glashow-Weinberg-Salam (GWS) [4] [5] [6], which describes the unified electromagnetic and weak (EWK) interactions. The mediators of the EWK interaction are three massive vector bosons W^+,W^- and Z, plus the photon, γ.

	Family 1	2	3	Q/e
Leptons	$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}$	$\begin{matrix} 0 \\ -1 \end{matrix}$
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} t \\ b \end{pmatrix}$	$\begin{matrix} 2/3 \\ -1/3 \end{matrix}$

Table 1.1: Quarks and leptons generations and their charges.

The symmetry group in (1.21) extends the gauge invariance U(1) of the Quantum Electro Dynamics (QED) to the EWK and strong interactions. In Table 1.2 the four interactions are shown together with the corresponding mediators.

Provided this set of matter fields and gauge fields, the resulting Lagrangian is composed of the set of terms invariant under this gauge group and Lorentz transformations. However, with

[1] To each particle corresponds an anti-particle, which has the same mass of the particle, but opposite electric charge and additive quantum numbers.

[2] A non abelian symmetry group is a group $G(G, *)$ in which there are at least two elements a and b of G such that $a * b \neq b * a$. A common example is the rotation group SO(3) in three dimensions.

Interaction	Mediator	Symbol	Intensity
Strong	8 gluons	g	10^{38}
Electromagnetic	photon	γ	10^{36}
Weak	IVB	W^+, W^-, Z	10^{29}
Gravity	graviton	G	1

Table 1.2: Fundamental interactions with their mediators and relative intensity.

only these components, all particles of the Standard Model would be massless. While this is true for the photon, the W and Z bosons have masses nearly 100 times that of a proton. The Brout-Englert-Higgs *spontaneous symmetry breaking* mechanism prescribes [7, 8] the addition of a spin 0 scalar field, the Higgs field, which acquires a vacuum expectation value through electroweak symmetry breaking, giving rise to the masses of the fermions and bosons.

1.2.1 Renormalisation

A charged particle scattered by an infinitely massive (static) point-like source with same electric charge is a process described by QFT by a potential $V(\mathbf{r})$ in the Schrödinger equation. The scattering problem can be treated to its lowest perturbative order in $V(\mathbf{r})$ and it can be interpreted as one-quantum exchange mechanism: the exchange of a mediator. However, this approximation does not include effects of multi-particle exchanges.
Such multi-particle exchange amplitudes are given by integrals over the momenta of the exchanged particles, constrained only by the four-momentum conservation. These integrals often *diverge* as the momenta of the exchanged particles tend to infinity. Nevertheless, this theory can be reformulated, by a process called *renormalisation*, in such a way that all multi-particle processes (higher-order) become finite and calculable. In the case of the SM interactions, this is a crucial requirement. The relevant data are precise enough to test the accuracy of the theory well beyond lowest order, particularly in the case of QED. The basic parameters of the theory, such as masses and coupling constants, have to be treated as parameters to be determined by comparison to the data, and cannot themselves be calculated.
There are also some theories that are *non-renormalisable*. This is the case of theories in which the coupling constant has dimensions of a mass to an inverse power, like the original Fermi theory of the weak interactions (the coupling constant $[G_F] = [m]^{-2}$). The essential point of the Fermi theory is that the dimensionful coupling constant introduces an *energy scale* into the problem, namely $G_F^{-1/2} \sim 300$ GeV. Therefore, for much lower energies than $G_F^{-1/2}$ the effective strength will be very weak, and the lowest order term in perturbation theory provides reliable predictions; this is how the Fermi theory was used for many years. But as the energy of the experiments increases and approaches $G_F^{-1/2}$, the theory becomes totally non-predictive and breaks-down. Thus renormalisability is essential in a theory.

1.2.2 Higgs sector of the Standard Model

The Lagrangian for the *unbroken* $SU(2)_L \otimes U(1)_Y$ gauge theory of vector bosons and fermions describes all the particles of the SM as massless. In the electroweak interactions, according

to the experiments, neither the mediating quanta (apart from the photon) nor the fermions are massless. They acquire mass via a mechanism that does not break the gauge symmetry of the Lagrangian, or else the renormalization of the theory is destroyed. In superconductors such a breaking of a gauge symmetry can happen dynamically. According to this idea, *some* additional fields must be present in order to give mass to the originally massless gauge bosons. Brout-Englert-Higgs [7, 8] assumed that in the SM a suitable scalar ("Higgs") field exists, with a potential which causes the vacuum (the ground state) to break the symmetry spontaneously. Furthermore, the fermion masses are put in the theory "by hand" via Yukawa-like coupling to the Higgs field.
In particular, while the coupling of the Higgs filed to the gauge fields are all determined by gauge symmetry, the Higgs self-couplings (trilinear and quadrilinear) are not gauge interactions and are unrelated to anything else in the theory. Moreover, the Yukawa-like fermion couplings are not gauge interactions either. They are both unconstrained and uncomfortably different in orders of magnitude. On the other hand, all these are renormalisable couplings. Thus the coupling constant with the Higgs field λ, which determines m_H (the Higgs boson mass) given the known value of v[3] is undetermined in the SM, and can be only measured by experiments.
The "Higgs sector" of the SM can be considered more as a "phenomenology" than a "theory", much as the current-current Fermi model for the weak interactions was. The Lagrangian of the Higgs sector of the Standard Model leads to many precise predictions which may be compared with experiment.
After three decades of experiments, at the beginning of 2012, the combined results of the CDF and D0 [9] experiments at Tevatron, and the ATLAS and CMS experiments at the LHC, excluded an m_H value in the interval 130 GeV to 600 GeV, at the 95 % C.L.. Finally, in July 2012 the ATLAS [10] and the CMS [11] collaborations announced the discovery, with a significance of 5σ, of a neutral boson within the mass range 125-126 GeV. Its production and decay rates are broadly compatible with the prediction for the SM Higgs boson (Fig. 1.1).

This long-anticipated discovery opens a new era in particle physics and pushes forward our understanding of the nature. The measured value of the Higgs boson mass leaves many questions unanswered. In the next Section, some of the limitations of the SM will be discussed. They represent the starting point for searches "beyond" SM.

1.3 Models for physics beyond Standard Model

The SM, including the neutrino masses, provides a remarkably successful description of presently known phenomena.

Figure 1.2 shows several SM total production cross section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations (calculated at NLO or higher). It shows a very good agreement between the SM predictions and the experimental measurements. Despite its success within its domain, the SM has a number of limitations.
A new framework will be certainly required at the Planck scale $M_{Pl} = (8\pi G_{Newton})^{-1/2} = 2.4 \times 10^{18}$ GeV, where quantum gravitational effects become important. The fact that the ratio $M_{Pl}/M_W = 10^{16}$, where M_W sets the electroweak scale, provides a strong indication of the

[3] $v/\sqrt{2}$ is the (tree-level) Higgs vacuum expetation value. The relation between the value of v and the Higgs boson mass m_H is $m_H = \sqrt{2}\mu = \sqrt{\lambda}v/\sqrt{2}$

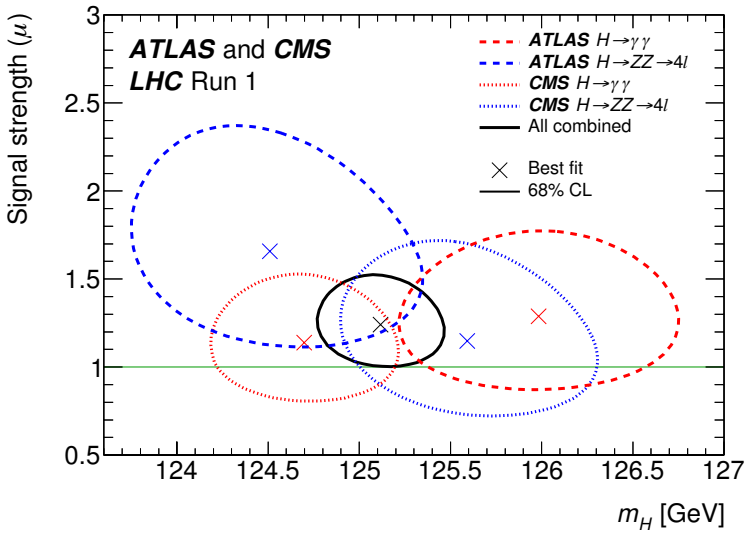


Figure 1.1: Summary of likelihood scans in the 2D plane of signal strength μ versus Higgs boson mass m_H for the ATLAS and CMS experiments [12]. The 68 % C.L. confidence regions of the individual measurements are shown by the dashed curves and of the overall combination by the solid curve. The markers indicate the respective best-fit values. The SM signal strength is indicated by the horizontal line at $\mu = 1$.

character of physics beyond the SM, because of the "hierarchy problem", that is related to the presence of the Higgs scalar field.
In the SM, particle mass values are affected by radiative corrections. The Higgs boson mass m_H receives large quantum corrections from every particle which couples to the Higgs field. Equation (1.22) shows the relation between the observable Higgs mass, m_H, the bare mass m_H^0 of the Higgs boson, and the radiative corrections to the Higgs mass, Δm_H.

$$m_H^2 = (m_H^0)^2 + \Delta m_H^2. \tag{1.22}$$

The latter term of (1.22) can be written as

$$\Delta m_H^2 = -\frac{\lambda_f^2}{16\pi^2}\left(2\Lambda^2 + \mathcal{O}\left[m_f^2 \ln\left(\frac{\Lambda}{m_f}\right)\right]\right), \tag{1.23}$$

where λ_f and m_f are the Yukawa couplings and masses of the fermions, and Λ is the energy cutoff which is interpreted as the energy scale up to which the SM is valid. Since the m_H in the first term of (1.22) is the measured value of the Higgs boson mass, the value of m_H^0 is dependent on the scale Λ. This means that if the SM needs to describe the nature up to the Planck scale, the quantum correction Δm_H^2 has to be 32 orders of magnitude larger than m_H^2. Without considering an automatic cancellation of these corrections, only a very large "fine tuning" could make it possible and this is considered highly *unnatural*. In other words, the theory must be extended beyond the SM to correctly describe these higher energy regimes, and to recover the very large separation of scales between the Planck scale and the electroweak scale, at which the Higgs boson is observed. In quantum field theory, this is accomplished by introducing new particles using the *naturalness* argument. Several extensions of the SM explain this unnatural

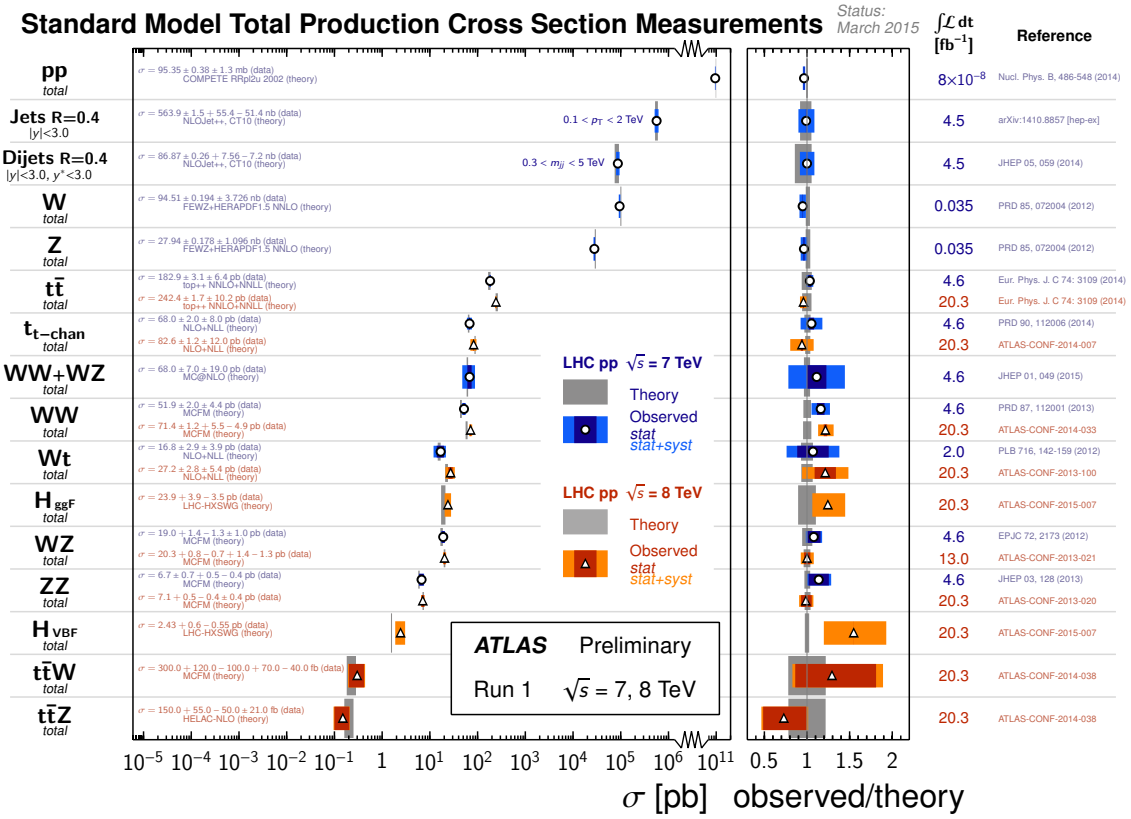


Figure 1.2: A detailed summary of several Standard Model total production cross section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations. All theoretical expectations were calculated at NLO or higher orders. The W and Z vector-boson inclusive cross sections were measured with 35 pb^{-1} of integrated luminosity from the 2010 dataset. All other measurements were performed using the 2011 dataset or the 2012 dataset. The dark-color error bar represents the statistical uncertainly. The lighter-color error bar represents the full uncertainty, including systematics and luminosity uncertainties. The data/theory ratio, luminosity used and reference for each measurement are also shown. Uncertainties for the theoretical predictions are quoted from the original ATLAS papers. They were not always evaluated using the same prescriptions for PDFs and scales. For the dijets measurement, $y^\star = |y_1 - y_2|/2$.

hierarchy by postulating new physics at the energy scale of the order of the TeV, that can be probed by the LHC.

A description of the gravitational force is also not included in the SM, because a quantization of general relativity would be needed. Therefore, the SM cannot be a theory that would unify gravity and the other three forces, a "theory of everything".

Another important missing piece in the SM is a particle description of the dark matter [13] and dark energy [14], whose existence can be inferred by several cosmological observations, measuring the Cosmic Microwave Background and "missing mass" in the orbital velocities of galaxies and clusters, among others.

These and other fundamental problems provide motivations for beyond SM searches. The SM could be only an effective model of a more general theory and only further experimental results

will be able to rule out or to confirm some of these models, and then to drive science into the right direction.

Typically, these models are predicted to manifest themselves as distinctive *signatures* in the LHC collisions through the production and decay of new, and often massive, particles. A "signature" is a particular final state topology that can be detected by the experiments. By performing a signature-based search in proton-proton collision events at the LHC, the rates of production of such final states are measured, and compared to SM predictions. Deviations from these predictions can indicate the existence of possible new physics models. It is important to remark that there is not a one-to-one correspondence between models and signatures. In many cases one signature can be expected in multiple possible models. This thesis is focused on signatures involving intermediate state topology consisting of pairs of W and Z bosons (e.g. WW, WZ, ZZ) that result from the decay of a massive particle. These final-state topologies are fully defined by the subsequent decay of the intermediate vector bosons into the hadronic channel ($W/Z \to q\bar{q}$). The hadronic decay mode has the benefit that the branching fraction of W/Z to quarks ($BR(W \to q\bar{q}) \sim 68\%$ and $BR(Z \to q\bar{q}) \sim 69\%$) is much larger than the branching fraction to leptons ($BR(W \to l\nu) \sim 10.8\%$, $BR(Z \to l^+l^-) \sim 11\%$, and $BR(Z \to \nu\bar{\nu}) \sim 20\%$) due to the three quark color charges. Although providing a higher branching fraction, the fully hadronic di-boson channel presents many experimental challenges that will be discussed in Chapter 5 and 6.
Two specific benchmark models have been chosen to interpret the results and determine (in the absence of an excess) the sensitivity of this new physics search. The first is the production of Kaluza-Klein modes of a graviton within the context of the bulk Randall-Sundrum framework of extra dimensions (Sec. 1.3.1). The second is the production of a W' boson resulting from a simple extension of the SM, the extended gauge model (EGM) (Sec. 1.3.2).

1.3.1 The Bulk Randall-Sundrum Graviton

As discussed in the previous Section, a *natural* solution to the hierarchy problem between the Planck and EWK scales invokes the existence of new particles with a mass scale between the EWK and $\bar{M}_{Pl}$ at the TeV scale. If the new states have non-negligible coupling with the SM particles, new physics should be detected at the LHC. A solution to the hierarchy problem is based on the Randall-Sundrum (RS) framework with a warped extra dimension [15–17]. The most distinctive feature of this scenario is the existence of spin-2 Kaluza-Klein (KK) gravitons whose masses and couplings to the SM are set at the TeV scale. KK gravitons are expected in experiments as resonances widely separated in mass.
A solution of the Einstein equations with the addition of this extra dimension yields the space-time metric

$$ds^2 = e^{-2kr_c|\phi|}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2 \tag{1.24}$$

where there are two free parameters, k and r_c, representing respectively a common energy scale of the theory and the radius of curvature of the warped extra dimension, and $0 \leqslant \phi \leqslant \pi$ is the coordinate for an extra dimension, which is a finite interval whose size is set by r_c. The warping of the extra dimension is analogous to the warping of spacetime in the vicinity of a massive object. In this space, four-dimensional mass scales are related to five-dimensional input

mass parameters and the warp factor, $e^{-2kr_c\pi}$. This extra dimension provides a bulk separation between two 3+1 dimensional branes, on which dynamics occurs at the ultraviolet Planck scale and the infrared TeV scale respectively (Fig. 1.3). This space has boundaries at $y = 1/k$

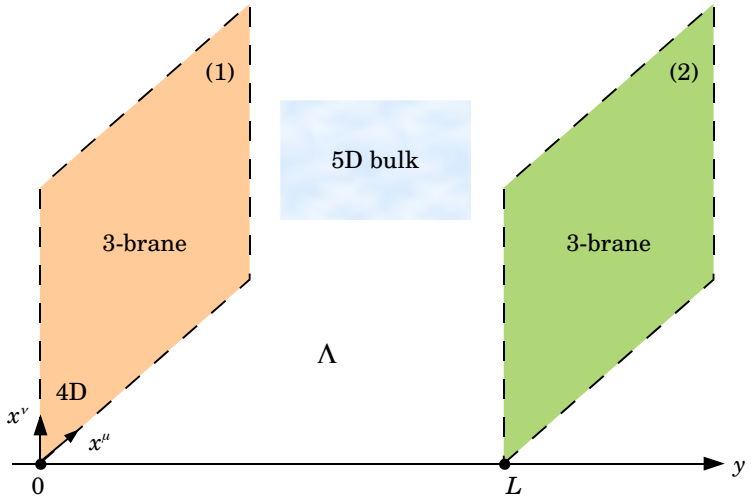


Figure 1.3: Randall-Sundrum setup. This construction entails two fixed points, one at the origin $y = 0$ and one at the other extremity of the circle, at $y = 1/e^{-2kr_c\pi}k \equiv L$. On each of these boundaries stands a four-dimensional world like the one we live in. By analogy with membranes enclosing a volume, these worlds with $3 + 1$ dimensions enclosing the 5D bulk have been called 3-branes. The picture is then two 3-branes, at a distance L one from another, enclosing a 5D bulk [18].

and $y = 1/e^{-2kr_c\pi}k \equiv L$, with $0 \leqslant \frac{1}{k} \leqslant \frac{1}{e^{-2kr_c\pi}k}$ where k is of the order of the Planck scale and $e^{-2kr_c\pi}k$ is of the TeV scale. The boundary at $y = 1/k$ is called the *Planck brane* and the boundary at $y = 1/e^{-2kr_c\pi}k$ is called the *TeV brane*. SM particles reside on the TeV brane. The physical scales in case of matter fields confined on the TeV brane are investigated in this setup, and considering the metric of the RS model. Through an analysis of the Higgs contribution to the Lagrangian in this framework [18], it can be seen that the vacuum expectation value (VEV) of the Higgs on the TeV brane, v, is related to that on the Planck brane, v_0, as $v = e^{-kr_c\pi}v_0$. As the Higgs VEV sets all the mass parameters in the SM, this means that all mass parameters are subject to an exponential suppression on the TeV brane. Thus, if the value of the bare Higgs mass m_H^0 is of the order of the Planck scale, the physics Higgs mass could be warped down to the weak scale, where it is expected to be. This is the reason why the first brane at $y = 0$ is called the Planck brane, whereas the second brane is called the TeV brane. The consequence is that in a theory where the values of all the bare parameters are determined by the Planck scale, hierarchy can be naturally generated between the weak and the gravity scales (Fig. 1.4).

The key feature of this model is that KK gravitons corresponding to small fluctuations around the metric (1.24), have a mass $\sim$ TeV and are localized near the TeV brane so that KK graviton coupling to the entire SM is only $\sim$ TeV suppressed. Furthermore, as a result of the warped nature of the extra dimension, the massless graviton acquires an entire tower of excited states, called Kaluza-Klein modes, that result from the necessary periodic boundary conditions of the field in the extra dimension, and can be produced and searched for at the LHC.
In order to specify the phenomenology it is necessary to fix some features of the model. One is

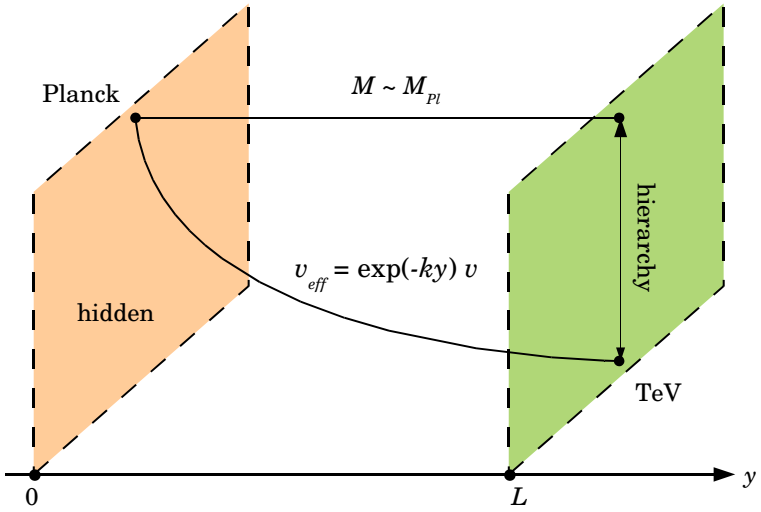


Figure 1.4: The generation of an exponential hierarchy [18].

the scale of the extra dimension, which requires specifying values for k and r_c. In the benchmark model considered, a value of $kr_c \sim 11$ is used. The scale k is therefore the only free parameter of the model, which is fixed to the Planck scale by requiring $k/\bar{M}_{Pl} = 1.0$, where $\bar{M}_{Pl}$ is the reduced Planck mass: $\frac{M_{Pl}}{\sqrt{8\pi}}$. This choice affects the phenomenology of the graviton signature and the way in which the wave functions of the SM populate the extra dimension, influencing their respective couplings. The massless graviton is typically localized near the Planck brane and the Higgs field, and the KK graviton excitations are taken to be localized near the TeV brane. If the quark flavours are localized at different points of the bulk space between the branes, then the different overlaps with the Higgs wave function can explain the hierarchical nature of the masses of the three families. In this model, the coupling to the graviton and its massive Kaluza-Klein modes are small. But their overlap with the wave functions of the longitudinal modes of the gauge boson wave functions (W_L/Z_L) can be significant, and provide a striking signature through their final state decay products. There are two such channels, the $G_{RS} \to W^+W^-$ and the $G_{RS} \to ZZ$, whose couplings are specified by the interaction of the graviton field ($h_{\alpha\beta}$) with the energy momentum tensor of the longitudinal components of the W and Z bosons (${T^{W_L,Z_L}}_{\mu\nu}$) as

$$\mathcal{L}_G \sim \frac{e^{kr_c\pi}}{\bar{M}_{Pl}} \eta^{\mu\alpha}\eta^{\nu\beta} h_{\alpha\beta} {T_{\mu\nu}}^{W_L,Z_L} \tag{1.25}$$

Both of these decays are searched for in this thesis.

1.3.2 The Extended Gauge Model W'

The other new-physics model that can give rise to di-boson signatures is a result of the Sequential Standard Model (SSM) [19]. The SSM is not the consequence of a general solution to a specific "problem" of the SM, but a reference model that predicts new phenomenology at the LHC. It is obtained by taking for the heavy $V^{\pm}$, the so called W', and V^0 the same coupling for the ordinary $W^{\pm}$ and Z in the SM. This not only refers to the coupling to fermions $Vq\bar{q}$ and $Vl\bar{l}$,

but also includes the $V^0W^+W^-$ or $V^\pm W^\mp Z$ vertices. The structure of the couplings to the SM particles are identical to those of the standard W and Z, with possible modifications to the strength of these couplings. For masses below the WW and the WZ thresholds, fermion couplings dominate, while when this threshold is passed, as at the LHC, the decay to WZ and WW dominates and increases like m_V^5 (with m_V the mass of the new heavy boson). This large increase of the coupling is corrected by allowing for mixing between the SM and the SSM gauge bosons. A mixing factor $c\frac{M_W^2}{M_V^2}$, where c is a tunable coupling strength, taken to be $c = 1$ in this search, defines a scaling of the VWW and VWZ couplings. The $c = 1$ choice results in what is commonly referred to as the Extended Gauge Model (EGM). The interaction term of the Lagrangian takes the same form as the triple gauge coupling of the SM with one leg being the new W' boson as

$$\mathcal{L}_{EGM,W'} \sim \frac{M_W^2}{M_V^2} W'WZ \tag{1.26}$$

The specific signature examined in this thesis probes the existence of the $W' \to WZ$ process.

Chapter 2

The LHC and the ATLAS Experiment

This chapter describes the LHC accelerator system and the ATLAS Experiment. The main characteristics of the LHC will be highlighted and the performance achieved during the Run 1 data taking will be reported. A section of this chapter is dedicated to the description of the main sub-detectors of the ATLAS Experiment, focusing in particular on the Electromagnetic and Hadronic calorimeter systems. These two sub-detectors are crucial for the identification and reconstruction of hadronic final states.

2.1 The LHC accelerator

The Large Hadron Collider (LHC) [20] is a two ring superconducting hadron accelerator and collider. It is installed in the 26.7 km tunnel, that was constructed between 1984 and 1989 for the CERN LEP electron-positron accelerator. The tunnel is composed of eight straight sections and eight arcs and lies between 45 m and 170 m below the surface on a plane inclined at 1.4 % sloping towards the lake Geneva. There are two transfer tunnels, each of approximately 2.5 km in length, linking the LHC to the CERN accelerator complex. Figure 2.1 shows the CERN accelerator complex layout. The acceleration chain begins with a LINAC, where protons obtained from a hydrogen gas are accelerated to 50 MeV and injected into the Proton Synchrotron Booster (PSB). This increases the proton energy up to 1.8 GeV, feeding them into the Proton Synchrotron (PS), reaching 26 GeV. The last ring of the injection chain is the Super Proton Synchrotron (SPS), which provides acceleration for protons up to 450 GeV and injects them into the LHC.

The LHC is designed to accelerate two beams of protons (or heavy ions) traveling in opposite directions up to 7 TeV (2.76 TeV per nucleon) in two separated vacuum chambers at a distance of 194 mm. The acceleration is provided by 8 resonant radio frequency cavities (RF), where the electrical field oscillates at 400.8 MHz to achieve a 0.5 MeV/turn energy kick. At the design energy the field strength reaches about 5.5 MV/m. The two proton beams are steered by 1232 cryogenic dipole magnets with a magnetic field of $B = 8.33$ T that allows bending of ~ 0.6 mm per m. These superconducting magnets are immersed in a pressurized bath of superfluid helium at about 0.13 MPa at 1.9 K to maintain a superconducting state, storing an energy of

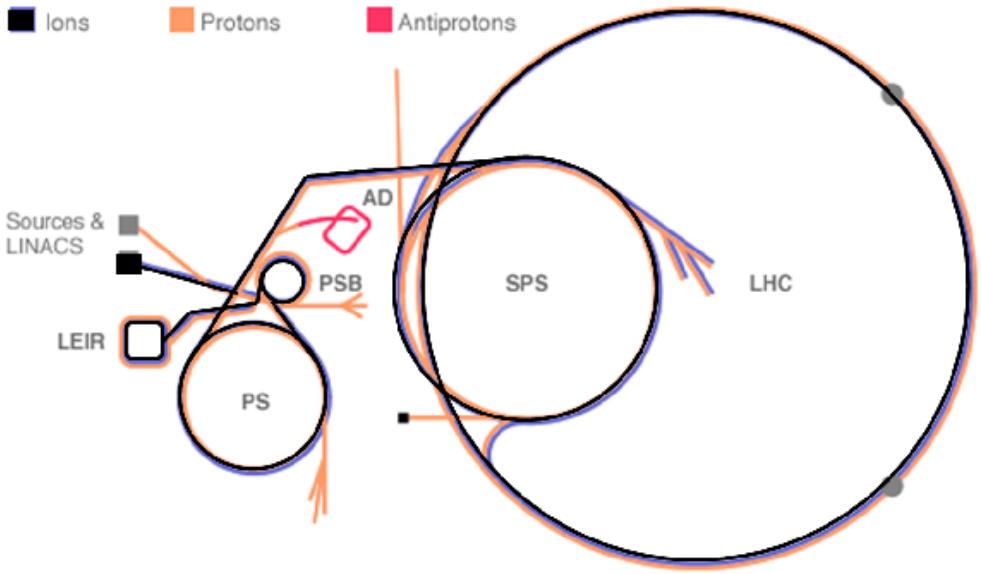


Figure 2.1: The CERN accelerator complex layout [21].

about 600 MJ.
Given the space limitations in the LEP tunnel, and to keep the costs down, a two-in-one, or twin-bore, design has been adopted for almost all of the LHC superconducting magnets. With this particular design the two beam channels use a common cryostat, with opposite magnetic field for the two channels. This makes the magnet structure complicated, especially for dipoles, where the two beam channels are coupled both magnetically and mechanically due to the small distance separation.

The LHC beams are composed of bunches of protons. The number of bunches can vary according to the data taking plans of the LHC, but it has to comply the LHC 40 MHz clock. The machine clock creates 3564 slots for bunches, repeated every 25 ns. Not all slots are used, leaving some collision-free latencies for experiment activities, such as online calibration, and sub-detector synchronisation. A high luminosity bunch train thus corresponds to 2808 bunches for each beam with 25 ns separation and with $\sim 1.15 \times 10^{11}$ protons per bunch.

Integrated luminosity

The integrated luminosity delivered by a particle collider is defined as the ratio between the observed number of produced events, N, and the production cross section of a given process, σ: $L = N/\sigma$. This quantity corresponds to the integral over a time period T of the instantaneous luminosity $\mathcal{L}$, indicating the rate of the collisions per unit of area.

$$L = \int_0^T \mathcal{L} \mathrm{d}t \ . \tag{2.1}$$

It can also be expressed as a function of the characteristic parameters of the collider considered:

$$L = \frac{F n_b N_b^2 f_{\mathrm{rev}} \gamma_r}{4\pi \beta^* \epsilon_n} \ . \tag{2.2}$$

Here, F is a geometric factor determined by the crossing angle between the two beams, n_b corresponds to the number of bunches circulating in the machine, f_{rev} is the frequency of revolution, N_b the number of protons in a bunch, γ_r the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, and β^* the beta function at the collision point. The above expression assumes round beams with equal beam parameters for both beams. The running parameters of the LHC at the design center-of-mass energy $\sqrt{s} = 14$ TeV and the values for the 8 TeV data taking at peak luminosity are shown in table 2.1.

Parameter	Unit	Design value	$\sqrt{s} = 8$ TeV value
Per-beam energy	[TeV]	7	4
β^*	[m]	0.55	0.6
Bunch spacing ($1/f_{crossing}$)	[ns]	25	50
Max number of bunches		2808	1380
Max number of protons	[10^{11}/bunch]	1.15	1.7
Peak luminosity	[$\text{cm}^{-2}\text{s}^{-1}$]	1×10^{34}	7.7×10^{33}
Max $\langle\mu\rangle$ (int/crossing)		19	19

Table 2.1: LHC parameters at design center-of-mass energy $\sqrt{s} = 14$ TeV and at 8 TeV [22,23].

The design instantaneous luminosity peak is 10^{34} $\text{cm}^{-2}\text{s}^{-1}$. The LHC has two high luminosity experiments, ATLAS [24] and CMS [25], both aiming at the design instantaneous luminosity. Two low luminosity experiments, LHCB [26] for B-physics, aiming at a peak luminosity of $\mathcal{L} = 10^{32}$ $\text{cm}^{-2}\text{s}^{-1}$. Installed on the LHC there is a dedicated ion experiment, ALICE [27], aiming for a peak luminosity of $\mathcal{L} = 10^{27}$ $\text{cm}^{-2}\text{s}^{-1}$ for nominal lead-lead ion operation.

Between November 10^{th} and December 16^{th} 2009, the first data at 900 GeV have been collected and analysed by the LHC experiments. During the Run 1 data taking period (from November 2009 to February 2013), LHC has accelerated the proton beams to 1.18 TeV, 3.5 TeV and 4 TeV. After the winter shutdown, on the 31^{st} of March 2010, the LHC began *proton* $-$ *proton* collisions at $\sqrt{s}$=7 TeV, which lasted until the end of 2011. The total delivered integrated luminosity to ATLAS in the 7 TeV data taking period is 5.6 fb^{-1}, as shown in Fig. 2.2a, with a peak instantaneous luminosity of 3.65×10^{33} $\text{cm}^{-2}\text{s}^{-1}$, illustrated in Fig. 2.2c. During 2012, LHC collided protons at $\sqrt{s}$=8 TeV, delivering to ATLAS an integrated luminosity of 23.3 fb^{-1}, shown in Fig. 2.2b. The maximum instantaneous luminosity reached was 7.73×10^{33} $\text{cm}^{-2}\text{s}^{-1}$, as can be seen in Fig. 2.2d.

Since the 3rd of June 2015, the Run 2 data taking at a centre of mass energy is 13 TeV has restarted, opening a new era for the high energy physics.

2.2 The ATLAS Experiment

2.2.1 Detecting particles

Particle physics detectors are devoted to particle identification, often referred as *tagging*, and to the precise measurement of the particle four-momentum.
The different particle types interact differently within the various detector volumes (Fig. 2.3). To detect and measure the electrical charge and momentum of electrically charged particles, the Lorentz force is used, and magnetic fields are employed to bend their trajectories. These

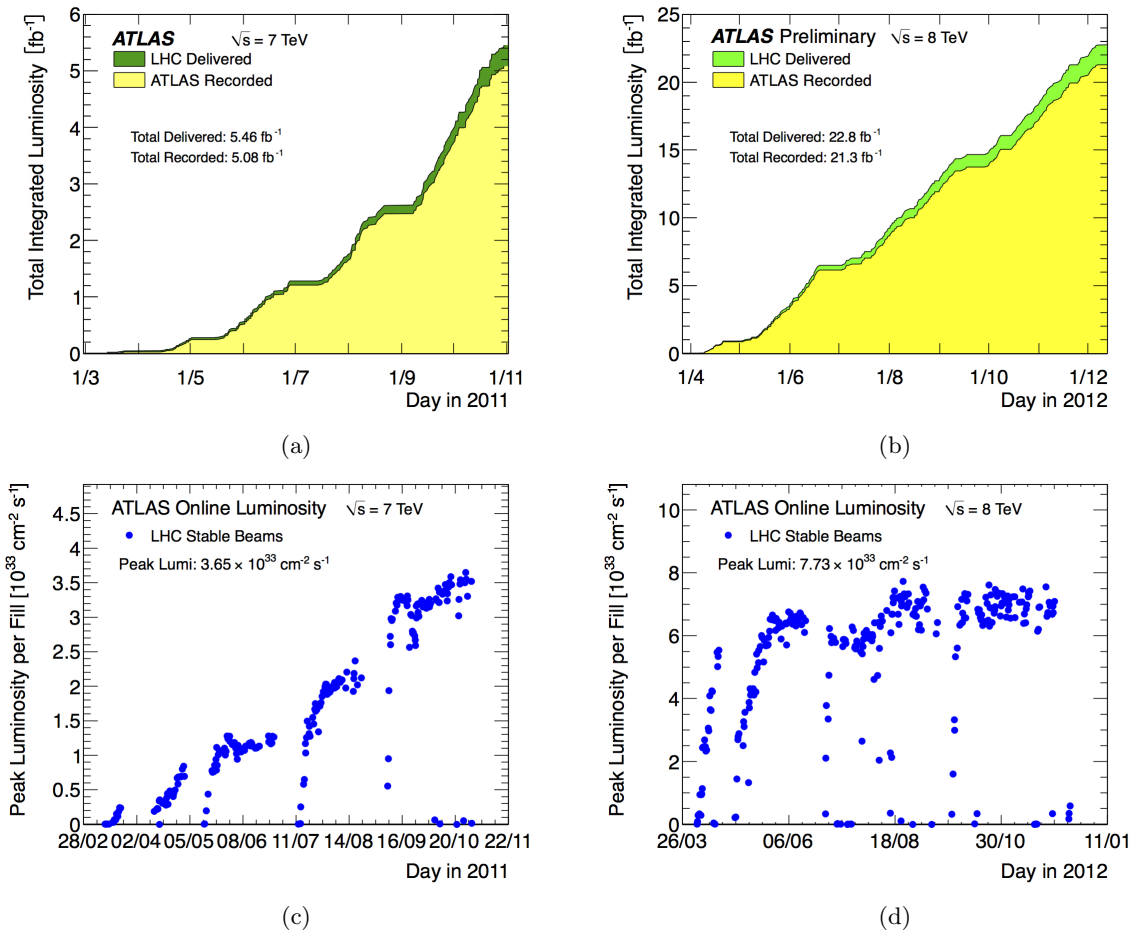


Figure 2.2: Integrated (a,b) and instantaneous (c,d) luminosity for the 7 TeV and 8 TeV ATLAS data taking period, as a function of time.

particles, travelling through the active material of silicon or gaseous detectors, ionize the media and the movement of the electric charges within the high intensity electric field produces electric signals. This signal produces what is generally referred as a *hit*, which indicates the spatial coordinate of the particle interaction. The trajectory of a charged particle is reconstructed interpolating the several *hits* which are left in the different layers of the tracking detectors.

The high luminosity at the LHC implies the production of many particles coming from many collision vertices, therefore a high spatial resolution for hits is needed for the tracking detector, which are generally placed close to the interaction point. The tracking detectors are also employed in the reconstruction of the collision vertex, or primary vertex (PV), and of possible secondary interactions (via the secondary vertices, SV) coming either from long-lived particle decays, like the case of heavy flavour hadrons or τ leptons, or from multiple interactions taking place between the proton bunches (pile-up). This phenomenon can take two forms: in-time and out-of-time pile-up. In-time pile-up is due to multiple collisions happening in a single bunch crossing. This is normally due to high proton multiplicity in each LHC bunch. In this situation the detector needs to be able to distinguish particles coming from multiple interactions.
Out-of-time pileup is due to residual energy from a previous bunch crossing or energy from a future bunch crossing which is deposited before the read-out process is completed. It is primarily

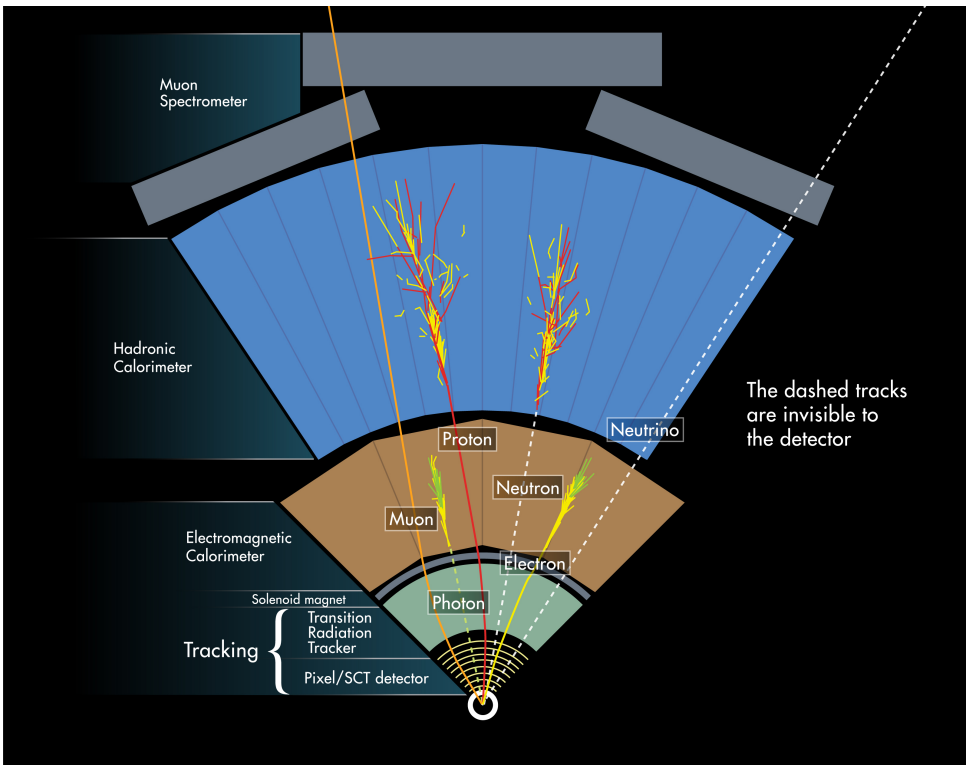


Figure 2.3: Interaction of the different type of particles with the ATLAS detector active volumes.

dependent on the amount of time occurring between bunch crossing, and it can contribute to the electrical signals associated to a given bunch crossing. Pile-up can have different impact on the reconstruction physics objects. In the case of jet reconstruction (discussed in Chapter 4), pile-up might produce additional energy in jets and some of the techniques developed to reduce this effects will be discussed in Chapter 5. Figure 2.4 shows the luminosity-weighted distribution of the mean number of interactions per crossing for the 2011 and 2012 data.

Charged particles also undergo electromagnetic (EM) interactions and can radiate photons (bremsstrahlung). This process has a high rate in the case of low mass charged particles, and photons above the energy threshold $2m_e$ can produce electron pairs, giving rise to electromagnetic showers. Electromagnetic calorimeters are designed to detect these showers. The electromagnetic shower longitudinal development is described in units of the radiation length (X_0), and the calorimeters are designed to contain showers from highly energetic photons and electrons.

Strongly interacting particles, hadrons, undergo inelastic nuclear processes when encountering heavy nuclei, producing hadronic showers in heavy materials. These showers are characterised in units of the interaction length (λ). The hadronic calorimeters are designed to contain highly energetic hadron showers, not fully contained in the electromagnetic calorimeter. All the energy deposits in the calorimeters are clustered and reconstructed as *jets*.

Muons only interact electromagnetically and weakly with matter, releasing minimum ionisation energy in the calorimeters. Therefore, they can penetrate the calorimeters, retaining most of their energy. They are measured by muon detectors, placed outside the calorimeters. These are

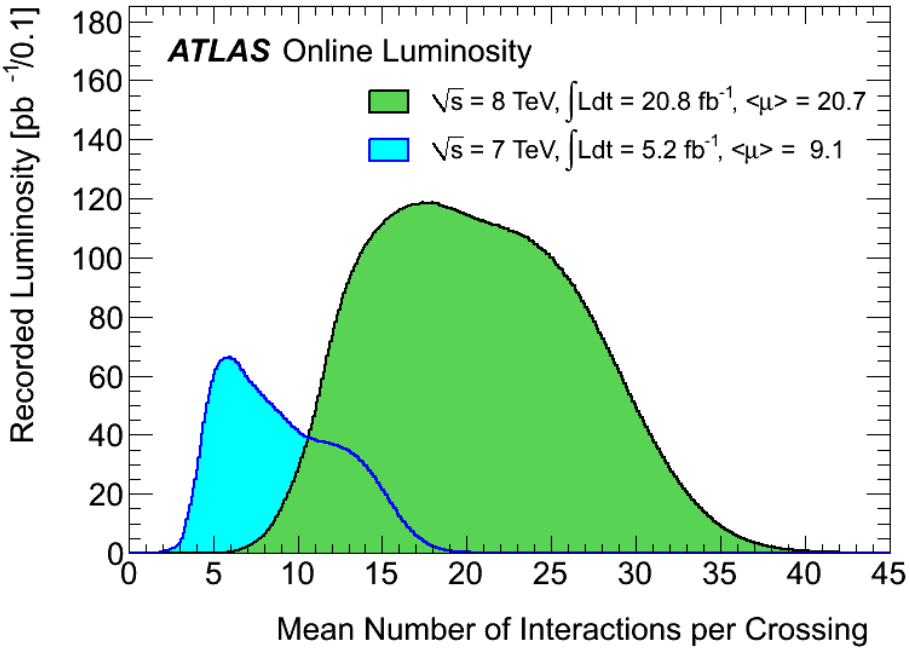


Figure 2.4: Luminosity-weighted distribution of the mean number of interactions per bunch crossing for the 2011 and 2012 data. The mean number of interactions per crossing corresponds to the mean of the poisson distribution of the number of interactions per bunch crossing calculated for each bunch. It is calculated from the instantaneous per bunch luminosity as $\mu = L_{bunch} \times \sigma_{inel}/f_r$ where L_{bunch} is the per bunch instantaneous luminosity, σ_{inel} is the inelastic cross section which we take to be 71.5 mb for 7 TeV collisions and 73.0 mb for 8 TeV collisions, n_{bunch} is the number of colliding bunches and f_r is the LHC revolution frequency.

precise tracking chambers, and the information they provide is often combined with that from the inner tracker to obtain a more precise measurement of the muon track parameters.

In a collider experiment, the neutrinos interact only weakly, with extremely low cross section with the material of a detector. Therefore, neutrinos are generally identified exploiting the momentum conservation in the transverse plane: $p_T^\nu = -\sum_i \overrightarrow{p_T}^i$, where i runs over all the reconstructed particles in the event. p_T^ν is generally referred to as p_{Tmiss}, the missing transverse momentum.

2.2.2 The ATLAS detector

The ATLAS (A Toroidal LHC ApparatuS) detector [24] is designed to exploit the full discovery potential of the LHC accelerator. The ATLAS subdetectors are arranged radially (tracker, calorimeters, muon chambers) from the interaction point up to a radius of about 25 m[1]. The

[1]ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam direction. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being

ATLAS detector is 44 m long and weighs 7 kt, and it is illustrated in fig. 2.5.

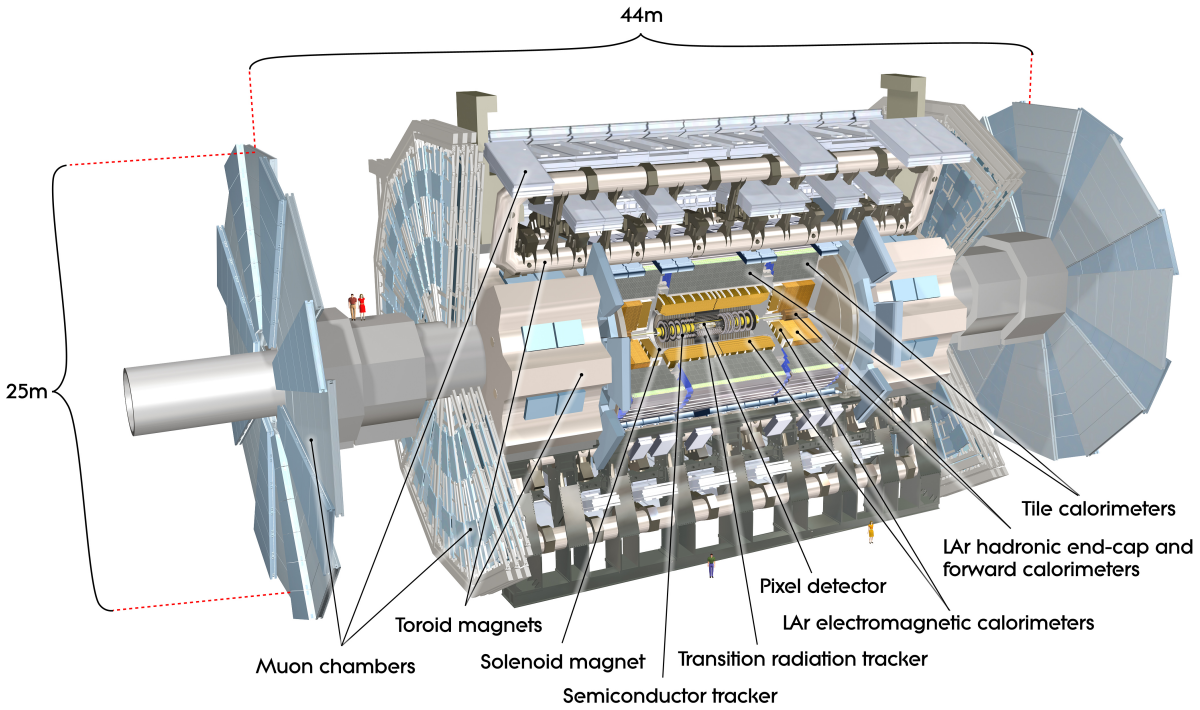


Figure 2.5: Schematic view of the ATLAS detector [24].

As already mentioned, the LHC provides a rich set of phenomena that can be observed. The LHC physics program goes from precise measurements of SM parameters to the search for new physics phenomena. Furthermore, high energy nucleus-nucleus collisions at the LHC give the opportunity to study the properties of strongly interacting matter at extreme energy density, including colour-deconfined[2] states like the *quark-gluon plasma*. A set of processes covering much of the new physics phenomena at the TeV scale have been used to define the detector system requirements.

The high luminosity and the increased cross-sections at the LHC center of mass energies enable high precision tests of QCD, electroweak interactions, flavor physics and top physics. The Higgs boson can be studied in several production and decay channels, and its Standard Model nature can be tested against new physics hypotheses. A wide spectrum of new physics phenomena, from SUSY to heavy gauge bosons, to quark compositeness, can be tested. All these phenomena can be identified via their final state products (signatures), and the subdetectors are designed to resolve them as well as possible.

Thus, the different final state topologies produced by the interesting processes at the LHC, and the methods used for particle identification drive the main requirements for the ATLAS

the azimuthal angle around the beam pipe. The polar angle is generally substituted by the pseudo rapidity, η, defined as $\eta = -\ln(\tan(\theta/2))$, for which a convenient property holds. The $\Delta\eta = \eta_1 - \eta_2$ for two detected objects 1 and 2, is invariant under Lorentz boost along the z axis. Transverse momentum p_T, transverse energy E_T, and the missing transverse energy E_T^{miss} are defined in the $x - y$ plane. The distance in the $\eta - \phi$ plane is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

[2]The colour confinement will be discussed in the context of the QCD phenomenology in sec. 3.1.1

detector. These are:

- fast, radiation-hard electronics and sensors, and high granularity detectors to handle the large occupancy produced by the pile-up events;
- need for the largest possible pseudorapidity acceptance and full azimuthal coverage;
- good resolution and reconstruction efficiency for charged particles in the tracker;
- precise impact parameter resolution in the high radiation dose environment close to the interaction point for secondary vertex reconstruction, aimed at identifying b-jets and τ leptons;
- efficient identification and measurement of photons and electrons in the electromagnetic calorimeter;
- maximal hadronic calorimeter coverage for precise measurements of jets and missing transverse energy;
- good muon identification and momentum resolution for the widest possible range of momenta;
- fast and flexible trigger, to identify interesting events in wide transverse momentum ranges.

The expected performance parameters for each subdetector are listed in Table 2.2. In the following sections, the detector components will be described in detail.

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\%\ p_T \oplus 1\%$	± 2.5	
EM calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimeter			
barrel and endcap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma p_T/p_T = 10\%$ at $p_T = 1$ TeV	± 2.7	± 2.4

Table 2.2: Design performance parameters of the ATLAS detector. E and p_T are expressed in units of GeV.

2.2.3 Magnet system

The ATLAS detector is characterized by a unique hybrid system of four large superconducting magnets. The general layout of the system is shown in Fig. 2.6.

It is 22 m in diameter and 26 m in length, with a stored energy of 1.6 GJ. The ATLAS magnet system consists of:

- a central superconducting solenoid, which is aligned on the beam axis, z, and provides a 2 T axial magnetic field for the inner detector (Sec. 2.2.4);

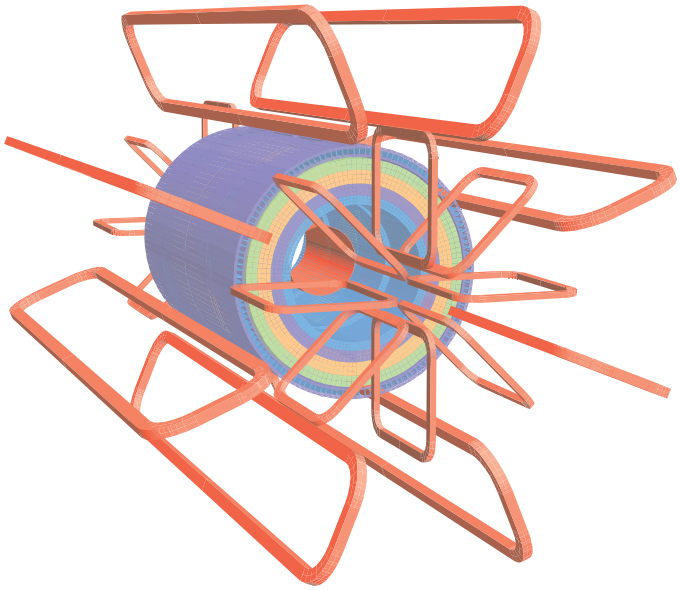


Figure 2.6: The geometry of the magnet system and the hadronic calorimeter steel [24]. The eight barrel toroid coils, with the end-cap coils interleaved are visible. The solenoid lies inside the calorimeter volume. The tile calorimeter is modelled by four layers with different magnetic properties, plus an outside return yoke.

- a barrel toroid and two end-cap toroids, which produce a toroidal and air-core magnetic field for the muon detectors (Sec. 2.2.6) with average values of 0.5 T and 1 T in the barrel and in the end-cap regions, respectively.

The central superconducting solenoid is designed to minimise the amount of material in front of the calorimeter, with $\sim 0.66 X_0$ for perpendicularly incident particles. This is realised by housing the solenoid and the cryostat in the same vacuum vessel of the electromagnetic calorimeter.
The toroidal magnetic field is produced by eight barrel coils, built on different cryostats, and two end-cap cryostats which host eight coils each. The end-cap toroidal magnetic system is rotated by 22.5°, providing a radial overlap and optimizing the bending power in the intersection regions of the two parts of the external magnet system. The choice of toroidal magnets has been made in order to have a magnetic field that is automatically confined without need for additional material, thus reducing the effect of multiple scattering on momentum resolution.
The complex behavior of the ATLAS magnetic field is shown in Fig. 2.7 in terms of its bending power $\int d\ell B$.

2.2.4 The Inner Detector

The ATLAS Inner Detector (ID) consists of three independent but complementary sub-detectors: the PIXEL, the SCT, and the TRT. The ID is designed to provide a robust pattern recognition, to measure precisely the track momentum and to reconstruct both primary and secondary vertex for charged tracks with $p_T > 0.5$ GeV and within the pseudorapidity range $|\eta| < 2.5$. This detector is also devoted to the electron identification over $|\eta| < 2.0$ and for a wide range of energies (between 0.5 GeV and 150 GeV).

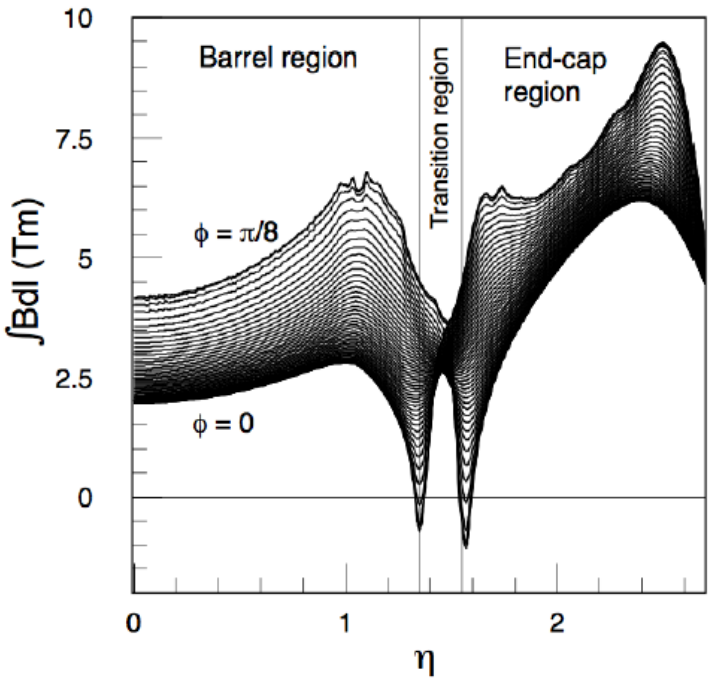


Figure 2.7: Behaviour of the bending power $\int d\ell B$ as a function of pseudorapidity η, for different values of the azimuthal angle ϕ [24].

It is contained inside a cylindrical envelope of length ±3512 mm and radius 1150 mm immersed in the surronding superconducting solenoid described in Sec. 2.2.3 and hence provides a momentum measurement of the particles from their curvature. A view of the ID is given in Fig. 2.8, and its geometrical parameters are listed in Table 2.3.

Item		Radial extension (mm)	Length (mm)
Overall ID envelope		$0 < R < 1150$	$0 < \|z\| < 3512$
Beam pipe		$29 < R < 36$	
Pixel	Overall envelope	$45.5 < R < 242$	$0 < \|z\| < 3092$
3 cylindrical layers	Sensitive barrel	$50.5 < R < 122.5$	$0 < \|z\| < 400.5$
2×3 disks	Sensitive endcap	$88.8 < R < 149.6$	$495 < \|z\| < 650$
SCT	Overall envelope	$255 < R < 549$ (barrel)	$0 < \|z\| < 805$
		$251 < R < 610$ (endcap)	$810 < \|z\| < 2797$
4 cylindrical layers	Sensitive barrel	$299 < R < 514$	$0 < \|z\| < 749$
2×9 disks	Sensitive endcap	$275 < R < 560$	$839 < \|z\| < 2735$
TRT	Overall envelope	$554 < R < 1082$ (barrel)	$0 < \|z\| < 780$
		$617 < R < 1106$ (endcap)	$827 < \|z\| < 2744$
73 straw planes	Sensitive barrel	$563 < R < 1066$	$0 < \|z\| < 712$
160 straw planes	Sensitive endcap	$644 < R < 1004$	$848 < \|z\| < 2710$

Table 2.3: Geometry parameters of the Inner Detector [24].

At inner radii, high-resolution pattern recognition capabilities are available using discrete space-points from silicon pixel layers and stereo pairs of silicon microstrip layers. At larger radii, the transition radiation tracker comprises many layers of gaseous straw tube elements interleaved

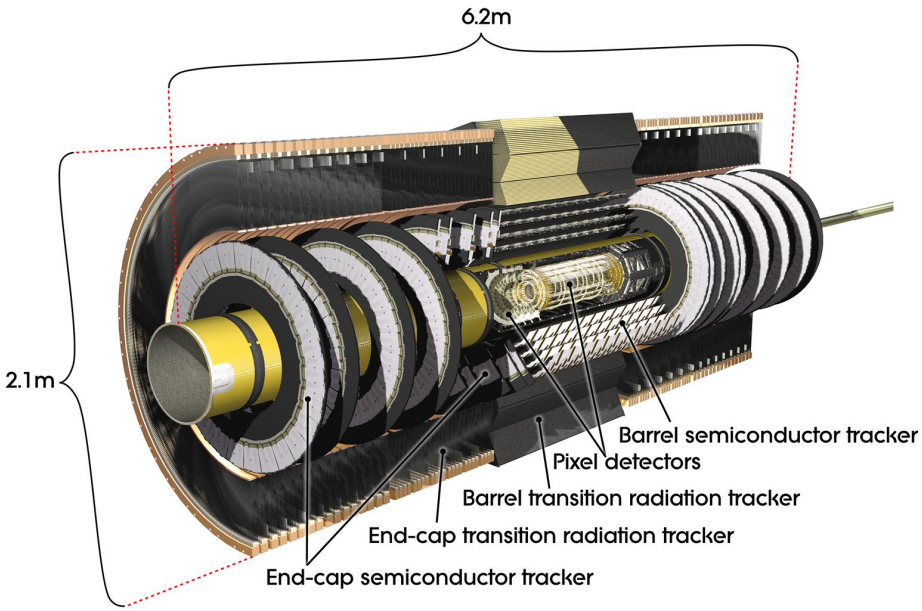


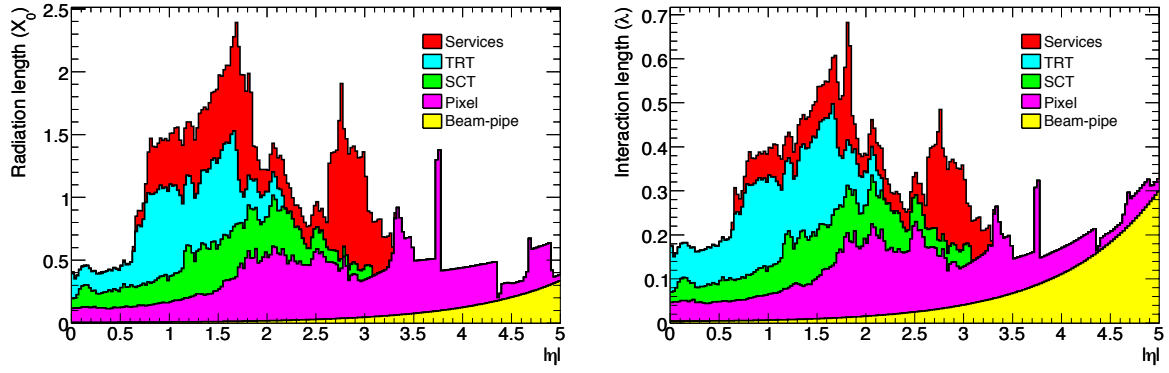
Figure 2.8: Overview of the ATLAS Inner Detector [24].

with transition radiation material.
The ID provides an average of 36 *hits* per track. This enhances the pattern recognition and improves the momentum resolution over $|\eta| < 2.0$. The ID provides also the electron identification, making this detector complementary to the calorimeter over a wide range of energies.
The parameters that can be measured by the ID, which characterise the helicoidal trajectory of a charged particle in a uniform field, are:

- the transverse momentum of the particle, p_{T};
- the track azimuthal angle, ϕ, related to the particle momentum by the relation $\tan\phi = p_y/p_x$;
- the polar angle θ, related to the particle momentum by the relation $\cot\theta = p_z/p_{\mathrm{T}}$;
- the transverse impact parameter d_0, the distance in the xy plane between the track closest point to the z axis and the z axis itself;
- the longitudinal impact parameter z_0, describing the z coordinate of the closest approach point to the z axis.

The location of the ID in the high-radiation environment, close to the beam pipe, imposes stringent radiation-hardness conditions on the sensors. In the course of the life time of the experiment, the pixel innermost layer must be replaced after approximately three years of full operation at design luminosity. In order to keep an acceptable noise after radiation damage, the silicon sensors must be kept at low temperature (approximately $-5°$ to $-10°$C) implying coolant temperature of $\sim$ -25°C.
The high granularity necessary for precision measurements, together with the readout and

cooling, results in an amount of inactive material in the inner detector volume up to a few % of X_0 and less than one λ. This is shown in Fig. 2.9.



(a) (b)

Figure 2.9: The material distribution of the inner detector [24], including services and beampipe, expressed in terms of radiation length X_0 (a) and interaction length λ (b).

As already mentioned, the Silicon Pixel Detector is the closest subdetector to the collision axis. It consists of three cylindrical pixel layers, plus five end-cap disks. The resolution for the three cylindrical layers and of the five end-cap disks is 10 µm in the $r\phi$ plane and 115 µm in z [24].

The Semiconductor tracker consists of eight silicon micro-strip detectors, with single point resolution of 17 µm in the $r\phi$ plane and of 580 µm on z [24]. On average this sub-detector provides eight precision position measurements for a track crossing the barrel region, and it allows to discriminate tracks originating from different particles, if they have a spatial separation of more than 200 µm.

The Transition Radiation Tracker consists of straw tubes, of 4 mm diameter. It can provide up to 36 hits per track in the region $|\eta| < 2$, and its transition radiation foils and fibers allow also discrimination between pions and electrons. The single-hit resolution of the TRT is of the order of 130 µm [24].

Overall, the combined momentum measurement resolution provided by the ID system is

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% \cdot p_T \oplus 1\% \tag{2.3}$$

where $\oplus$ indicates that the two contributions are summed in quadrature.

2.2.5 Calorimetry

The aim of the ATLAS calorimetric system is to measure energy and position of particles through the measurement of their energy deposits in the detector material. The ATLAS calorimetric system consists of two separate subsystems: the electromagnetic and the hadronic

calorimeters. It is designed to provide optimal resolution for high energy photons, electrons, jets and missing transverse energy. For a precise and reliable measurement of the missing transverse energy, the calorimeter is required to be hermetic.

The interaction of energetic particles with matter produces cascades (showers) of particles. Calorimeters must provide a good containment for electromagnetic and hadronic showers, and must limit punch-through into the muon system. These characteristics are obtained with thick calorimeters, located within the radius of the toroidal magnet muon system. The total thickness of the electromagnetic calorimetry is larger than $22X_0$ in the barrel region and larger than $24X_0$ in the endcap region. The total interaction length of the entire system is approximately 10λ, with a peak value of 11λ at $\eta = 0$.
ATLAS calorimeters are non compensating. This means that the hadron response is lower than the response to the electromagnetically interacting particles. The energy of an electromagnetic shower is usually fully detected in the electromagnetic calorimeter, while not all the energy of hadronic shower can be detected in the calorimeters. Invisible energy from nuclear breakdowns and excitation, and partially from invisible particles escaping detection (μ,ν) needs to be taken into account, restoring the correct energy scale for hadronic particles at the offline analysis stage.
The ATLAS calorimetric system employs a sampling technique, separating the absorption from the energy measurement with a *sandwich* structure of active and passive material. This allows for a more compact design with good containment of the shower.
In the next sections, a brief description of the ATLAS electromagnetic and hadronic calorimeter will be provided. The overall scheme of the calorimetry system is shown in Fig. 2.10.

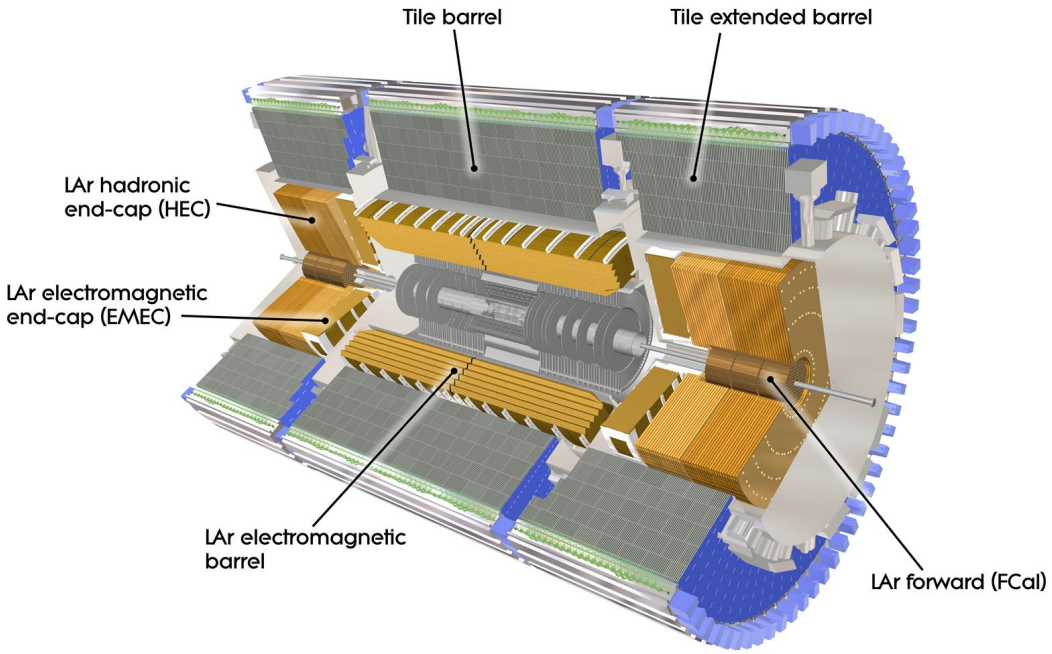


Figure 2.10: Overall view of the calorimetric system [24].

Electromagnetic Calorimeter

The Electromagnetic Calorimeter (EM) is divided into a barrel $|\eta| < 1.475$ and an end cap region ($1.375 < |\eta| < 3.2$). The measurement in the region $1.375 \leqslant |\eta| < 1.52$ is affected by additional material needed to instrument and cool the inner detector. This region is normally excluded from analyses that require precise measurement of electrons. EM uses liquid argon as the active material and lead as absorber, for this reason it is also called *LAr* calorimeter. An additional pre-sampler detector, consisting of an active layer of LAr of 1.1 cm thickness, is installed in front of the cryostat wall. It provides a correction to the energy loss caused by the material in front of the calorimeter in the $|\eta| < 1.8$ region.
When a charged particle travels through the calorimeter, it ionises the liquid argon, and the resulting electrons drift towards the copper electrodes in the read out cells thanks to the presence of an electric field.

The LAr calorimeter is caracterised by an *accordion structure* (Fig. 2.11). This design guarantees complete azimuthal coverage, provides a full ϕ symmetry for traversing particles and allows for a fast readout. The liquid argon is located in the gaps between the 1024 accordion-shaped lead absorbers, leading to a drift time of approximately 450 ns whithin a 2 kV electric field. Parameters such as high voltage, liquid argon temperature and density can influence the drift velocity and hence the energy measurement. Therefore, they are continuously kept under control to reduce the fluctuations in the energy measurement.

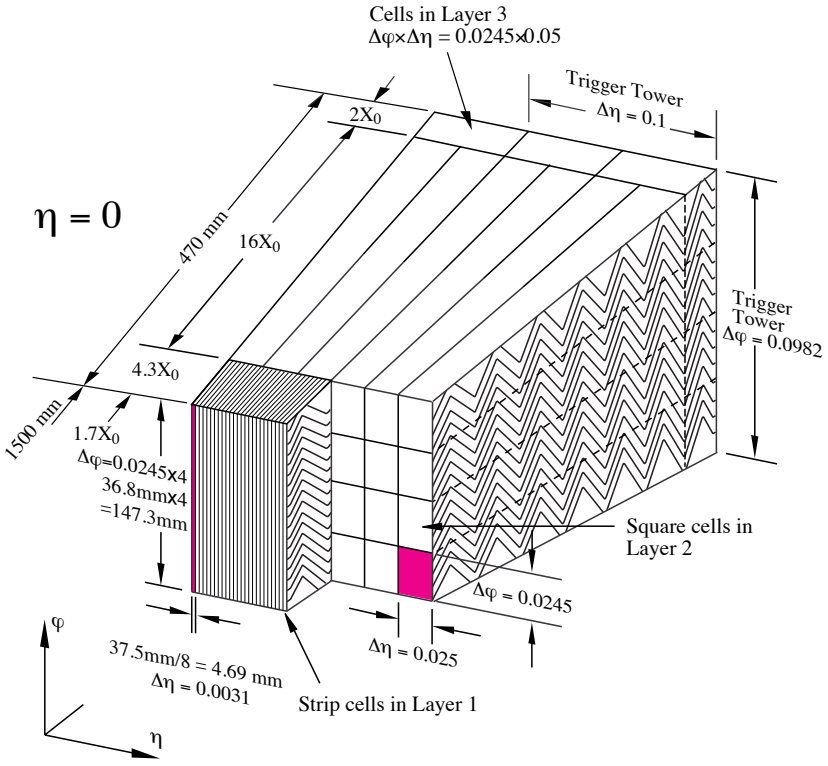


Figure 2.11: Sketch of the EM calorimeter structure [24].

The calorimeter sampling is fine and subdivided in three layers, as can be seen in Fig. 2.11:

- the first layer, $4.3X_0$ thick, consists of strips with $\Delta\eta = 0.0031$, allowing for charged and neutral pion separation;
- the second layer, $16X_0$ thick, is segmented in squared towers of $\Delta\eta \times \Delta\varphi = 0.025 \times 0.025$;
- the third layer has $\Delta\eta \times \Delta\phi = 0.050 \times 0.025$ towers, specifically for electrons and photons with $E > 50$ GeV producing larger showers.

The above segmentation allows for high precision spatial measurements, providing a pointing geometry that helps the identification of photons coming from a primary vertex, which is crucial for example in the $H \to \gamma\gamma$ measurement and for jet substructure that will be introduced in Chapter 5.

The lead thickness in the barrel is optimised as a function of η, providing the best possible resolution in energy measurements. The design energy resolution, summarised in Table 2.2, is:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E[\text{GeV}]}} \oplus 0.3\% \ , \tag{2.4}$$

and the pseudorapidity resolution is:

$$\Delta\eta = \frac{40 \text{ mrad}}{\sqrt{E[\text{GeV}]}} \ . \tag{2.5}$$

Hadronic Calorimeter

The ATLAS hadronic calorimeter system is placed directly outside the EM calorimeter envelope. The hadronic calorimeter makes use of plastic scintillator tiles as active material and steel as absorber. The barrel covers the region $|\eta| < 1$ and the two extended barrels $1 < |\eta| < 1.7$. It is also normally referred to as *Tile* calorimeter.
The assembled module forms an almost-periodic steel-scintillator structure with a ratio of volume of approximately 4.7:1. The 3 mm thick scintillator tiles are placed perpendicularly to the colliding beams. The geometry is sketched in Fig. 2.12.

When a particle interacts with the active medium, it produces scintillation light proportional to the deposited energy. This is collected with wavelength-shifting fibres on the tile edges. The readout cells are built by grouping fibers together in projective towers in η, and the scintillation light is collected by photomultiplier tubes (PMT) at each end of the tiles. Its granularity is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$.

The Tile calorimeter is equipped with three calibration systems: charge injection, laser and a ^{137}Cs radioactive source. These systems test the optical and digitised signals at various stages and are used to set the PMT gains uniform within $\pm 3\%$.

In the forward region ($1.5 < |\eta| < 3.2$) the detector experiences a larger particle flux and high energies. The design of the calorimeter system in the forward region is driven by the following requirements: many radiation lengths to fully contain hadronic showers, and radiation-hard technology, due to the large amount of radiation it is exposed to. Therefore the active material employed is liquid argon, while the absorbing material is copper.

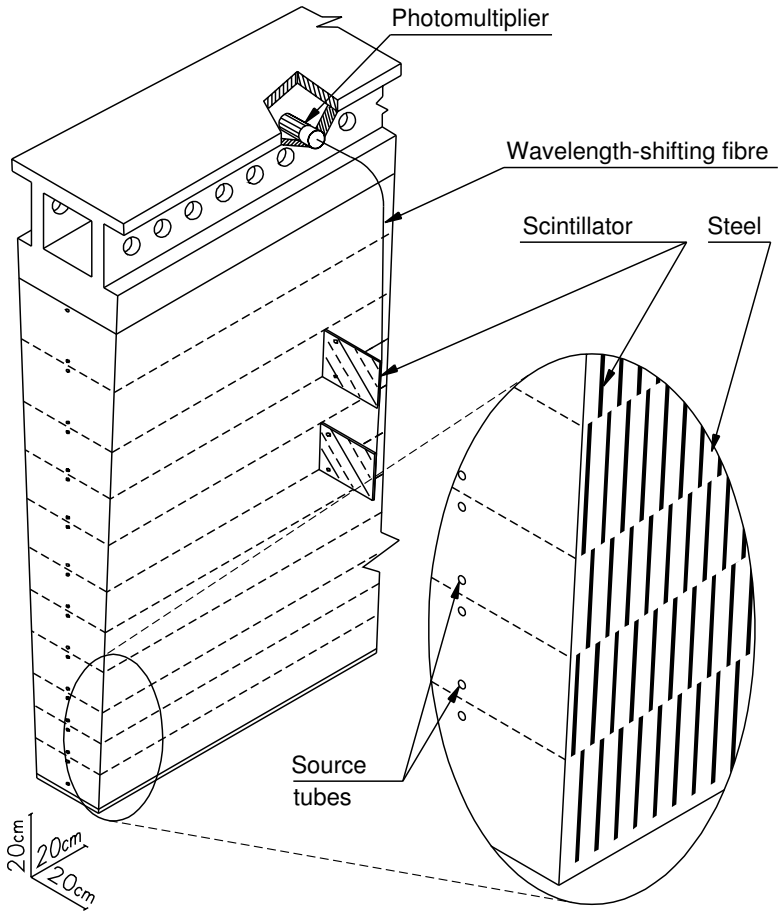


Figure 2.12: Sketch of the Tile calorimeter structure [24].

The hadronic endcap calorimeter (HEC) is placed at 2.03 m from the interaction point. In the HEC, 8.5 mm active gaps are sandwiched between copper plates. The system is composed of two wheels, with different widths for the absorber plates (25 mm and 50 mm for inner and outer wheel, respectively). In order to have a low-HV configuration, the active gaps are split in four drift spaces of $\sim$ 1.8 mm by the presence of three electrodes. The readout cells are pseudo-projective in η and projective in ϕ.

The overall design energy resolution of the TileCal and HEC systems is:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E[\text{GeV}]}} \oplus 3\% \ . \tag{2.6}$$

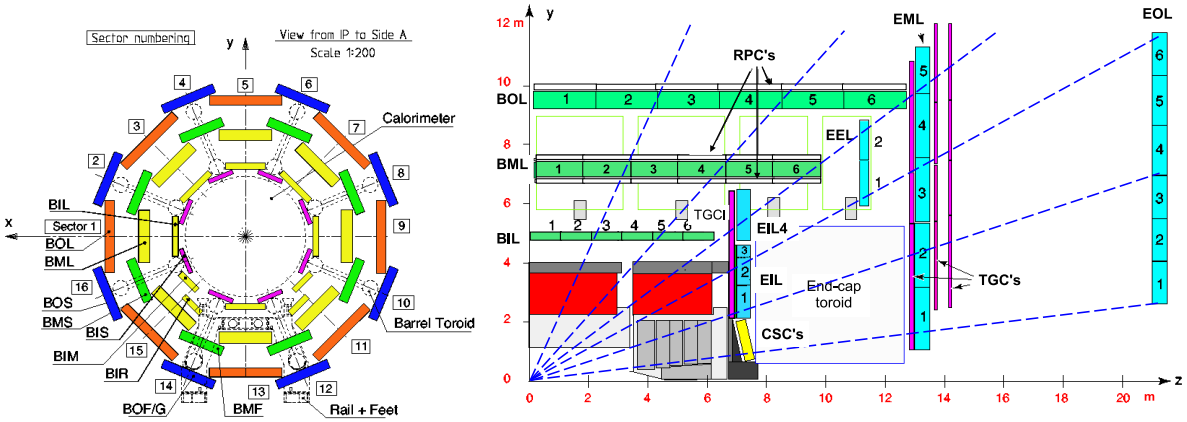
The front face of the forward calorimeter (FCAL) is placed at 4.7 m from the interaction point. The FCAL is placed in the same cryostat if the HEC. The liquid argon thin gaps (2 mm) are centered in tubes parallel to the z axis. The electromagnetic and hadronic parts use different absorbers, copper and tungsten respectively. The FCAL has a high-density design, because of the reduced available longitudinal space and the need to avoid energy leakage. Its design energy

resolution is:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \ . \tag{2.7}$$

2.2.6 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost part of the ATLAS detector. It uses the toroidal air-core ATLAS magnetic system to identify and measure high momentum muons (Fig. 2.13). The direction of the magnetic field causes the muons to bend in the $r - z$ plane, orthogonal to the solenoidal field in the inner detector. It is designed as an independent subdetector,



(a) (b)

Figure 2.13: (a) section of the barrel muon system perpendicular to the beam axis, showing the three concentric cylindrical layers with eight large and eight small chambers. (b) a section of the muon system along the beam axis [24].

exploiting the clean signature of muons, the only charged particles that are not stopped by the calorimeters.

The MS is composed of four subdetectors that make use of different technologies: Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC).
The MDT chambers perform a precise coordinate measurement in the bending direction of the air-core toroidal field, and therefore provide the muon momentum measurement for $|\eta| < 2.7$. This detector is composed of cylindrical aluminum drift tubes of 3 cm diameter with a anodic central wire at a potential of 3080 kV. When a muon crosses the tubes, it produces ionization electrons that are collected on the wire. Using the reconstructed drift time, a precise measurement of the minimum distance between the muon and the wire is provided (with a typical resolution of 80 µm), and this is used to reconstruct the muon trajectory.
The CSC are multi-wire proportional chambers with cathodes segmented into strips, and are used at large pseudo-rapidities ($2.0 < |\eta| < 2.7$) in order to cope with the higher muon rate and background conditions. The RPC and TGC are used for the trigger (next section) and measure also the muon ϕ coordinate.

2.2.7 ATLAS trigger and data acquisition

The highest LHC bunch crossing rate delivered to the ATLAS experiment for the 2012 dataset was 50 ns, providing a collision frequency of 20MHz. As already mentioned in Sec. 2.1, a high luminosity bunch train in LHC corresponds to 2808 proton bunches for each beam with a 25 ns separation and with $\sim 1.15 \times 10^{11}$ protons per bunch. The resulting event rate is about 10^9 Hz, determined by the $proton-proton$ cross section. This needs to be reduced to ~ 100 Hz, for the limited speed of the permanent storage system. Therefore strict requirements of rejection against minimum bias events (a factor of 10^7) are necessary to be fully efficient for the rare physics processes investigated by ATLAS.

The ATLAS trigger and data-acquisition system (DAQ) consists of three levels of online event selection. Each level refines the selection made by the previous one by adding additional information and selection criteria.

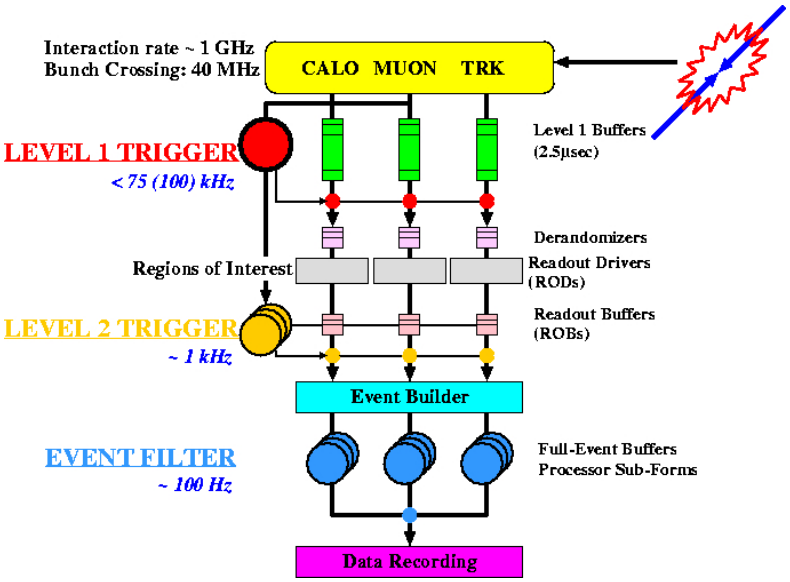


Figure 2.14: Scheme of the ATLAS trigger system [24].

The main requirement for level-1 (LVL1) trigger is to coarsely identify the interesting events, strongly reducing the total rate. The initial selection made by the LVL1 trigger is based on reduced granularity information. The calorimeter trigger identifies energy deposits from: high p_T electrons and photons, jets, hadronically decaying τ leptons and large missing transverse energy. The muon trigger identifies high and low p_T muons. The EM and hadronic calorimeters contribute separately to e/γ and hadron/τ triggers, with the possibility to apply isolation cuts too. Missing and total scalar transverse energy are calculated by summing over trigger towers objects. Due to the need for the LVL1 to be very fast and the lower resolution at high p_T, it does not use information coming from the inner detector. The expected final LVL1 rate is < 100 kHz.

An essential requirement for the LVL1 is to identify with no ambiguity the interesting bunch-crossings, given the very short 25 ns interval. In the muon trigger case, the time-of-flight in the spectrometer is of the same magnitude of the time interval between two bunch-crossing.

On the other hand, for the calorimeter trigger case, the calorimeter signal pulse shape extends over many bunch-crossings. During this latency time (~ 2.5 μs), the information coming from the detector is stored into pipeline memories and then passed to the trigger decision system, consisting of purpose-built hardware processors. The LVL1 system has a very important feature, which is specific to ATLAS: it allows for the identification of the Regions of Interest (ROIs), which represent the position of the trigger objects in the $\eta - \phi$ plane. ROI information is used to reduce the computation time at the next trigger level and the size of the data to be recorded.

The Level-2 trigger (LVL2) runs offline-like algorithms optimised for the on-line processing, using information also from the inner detector. It processes only data coming from the ROIs selected by LVL1[3]. Therefore, it works only on roughly $2 - 5\%$ of the whole ATLAS data size. After the maximum latency time of 10 ms, an event can be either selected and passed to the next trigger level, or discarded and removed from the data flow chain. The expected final rate for LVL2 is about $1 - 2$ kHz.

The last trigger stage in ATLAS is represented by the Event Filter (EF). It performs a full event reconstruction using offline selection algorithms adapted to online decisions and the most up-to-date calibrations for the calorimeters. The EF performs the final physics selection to be written to mass storage for the offline analysis. The expected time available for this last level to make a decision on the event is 1 s. It is expected to reduce the output from LVL2 by an order of magnitude, getting a final output of ~ 600 Hz. The expected final event size is 1 Mbyte. Hence, the output data rate is expected to be of about 600 Mbytes/s, with $O(10^{15})$ bytes of data collected per year.

2.2.8 ATLAS simulation

The ATLAS simulation infrastructure provides detailed simulation of SM and new physics processes, which are propagated through the full ATLAS detector volumes. An overview of the ATLAS simulation data flow is shown in Figure 2.15.

The chain starts with an event generator that produces events in a standard HepMC format [28]. At this stage, the events can be filtered requiring only particular properties. The generator is responsible of any prompt decays, but only "stable" (not decaying while traveling through the detector) particles are stored. Stable particles are then propagated through the ATLAS detector simulation produced by GEANT4 [29]. The energy deposits in the sensitive parts of the simulated detector are recorded as "hits". The coordinates, the energy deposition and the time of the hits are written into an output file that is called "hit file". During both event generation and the simulation of the interaction with the detector, the history of the interaction from the generator, the "truth", including incoming and outgoing particles is recorded. The truth particles are recorded also if they are not passed through the detector simulation.
The next stage is represented by the digitization, where Simulated Data Objects (SDOs) are created from the truth. The detector noise, and the trigger simulation are added at this stage. In particular, the trigger simulation is performed in "pass" mode, meaning that no events are discarded but each trigger hypothesis is evaluated. The output of the digitization are "digits" that are used as inputs for the read out drivers (RODs) in the detector electronics.
Also the ROD functionality is emulated and it produces as output Raw Data Objects (RDO)

[3]The LVL2 can also be used in *full scan* mode. In this case the LVL2 requirements are applied to the full detector.

files. The output produced by the ATLAS detector is in the same RDO (or RAW) file format. The files produced from the simulation and the one from the actual data-taking are interchangeable, with the only exception of the truth information, only present in simulation.

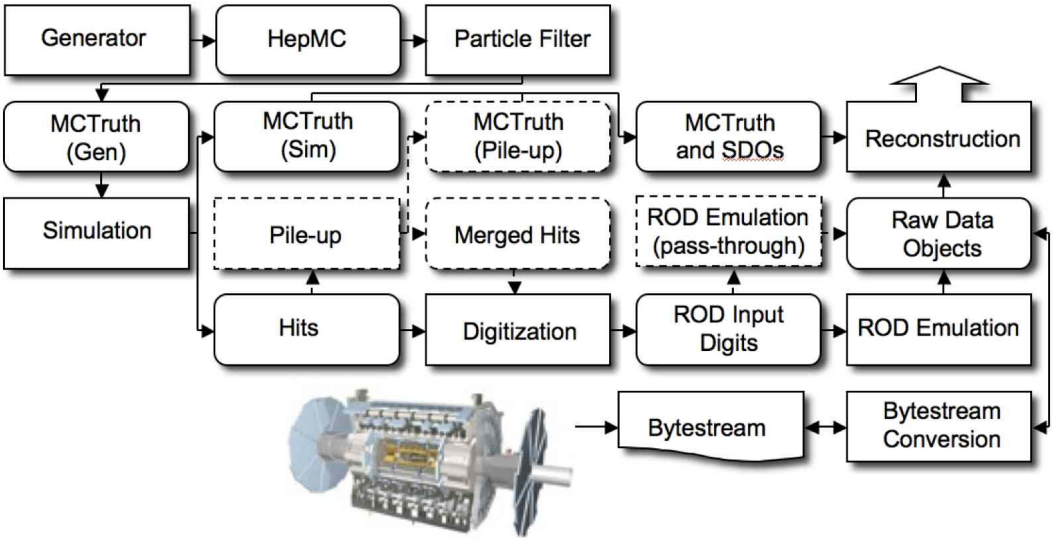


Figure 2.15: The flow of the ATLAS simulation software, from event generators (top left) through reconstruction (top right). Algorithms are placed in square-cornered boxes and persistent data objects are placed in rounded boxes. Generators are used to produce data in HepMC format. Monte Carlo truth is saved in addition to energy depositions in the detector (hits). This truth is merged into Simulated Data Objects (SDOs) during the digitization. Also, during the digitization stage, Read Out Driver (ROD) electronics are simulated [28].

2.2.9 ATLAS offline software

Athena [30] is the ATLAS offline software framework. It implements objects and algorithms in C++, which are configured and loaded by PYTHON scripts, called "job options". The framework inherited the core structure from the Gaudi framework [31, 32], originally developed by LHCb collaboration and written in C++. The basic idea is to have a flexible framework that supports a large variety of applications through base classes and basic functionalities. The infrastructure relies on the CLHEP common libraries [33]. These include utility classes designed for use in high-energy physics software.
The Athena releases are divided in major projects, depending on their functionality and they also contain the detector conditions and geometry definitions. The ATLAS simulation software, the reconstruction, and the physics analysis are some of the areas in which the framework is partitioned. The modular structure allows high flexibility in performing analysis on simulated or collision data using same objects and tools. The framework is designed to run on both local, and distributed computing facilities, like the Worldwide LHC Computing Grig (WLCG), that will be described in the next section.
A generic Athena job consists of three distinct steps. In the initialization step, all the services (accessed many times during a single event) and algorithms (called once per event) are loaded. This initial step is configured within a common PYTHON infrastructure. This avoids recompil-

ing the entire C++ code every time a single configuration parameter is changed, and it allows the configuration of many tools from a common interface. Once the initialization is completed, the event loop begins. For each event all the algorithms are run sequentially. In the final step, all the algorithms are terminated and all the objects are deleted.
The Athena workflow provides maximal flexibility for the user and allows to load only the modules requested, in order to keep each step as light as possible in memory and as fast as possible in processing a single event.

2.2.10 Data formats and computing facilities

The enormous amount of data produced either from LHC collisions or MC simulations requires a high-performance flexible distributed computing system. This requirement is met within the WLCG. The WLCG uses a multi-Tier schema, very well-suited to distribute both storage and computing among the institutes participating in the ATLAS Collaboration.

The Tiers are computing nodes organised in different levels. The EF output data (RAW) are initially recorded and reconstructed at the CERN Tier 0 (T0) data centre, which contains approximately 20% of the computing power of the full WLCG. After copying the RAW data, event reconstruction is performed, leading to the production of ESD (Event Summary Data) and AOD (Analysis Object Data). The ESDs contain the reconstructed quantities measured in the detector (hits), together with the reconstructed physics objects. The ESD event size is about 500 kbytes. The AOD is a smaller size format, with only 100 kbytes per event, in which is stored only reconstructed physics objects. The D3PDs (Derived Physics Data) represent a reduced format of them, which are the data format more flexible and generally used for the final physics analyses[4]. In D3PDs, rather than storing the data in objects, every variable is stored as a simple type, such as an integer, vector (list) of integers, and so on. This format typically is created at a per-analysis or per-working group level, and applies specific selection cuts, triggers, and so on to further reduce the size of the file to a manageable level. Even with this amount of reduction, a typical analysis will work with many TBytes of D3PD data.

Data are distributed to the Tier-1 centres, in which RAW data, AODs, ESDs and D3PDs are copied. Tier-1 centres are distributed in different countries and have the requirement to be able to reprocess RAW data, in order to get more accurate updated versions of smaller-sized data formats.

The physics analyses are mainly performed at the Tier-2 and Tier-3 centres. On average, there are five Tier-2 connected to a Tier-1. Tier-1 and Tier-2 perform the Monte Carlo production and the simulated data are stored in the Tier-1. The multi-Tier structure relies on the presence of GRID technology and middle-ware software, hiding the more complex structure to the user.

[4]D3PDs are produced in standard *Root-ntuple* [34] format.

Chapter 3

QCD and Monte Carlo generators

In the first part of this chapter, the key concepts of QCD will be discussed. The Monte Carlo generators used to simulate the processes discussed in this thesis will be introduced at the end of the chapter.

3.1 Key concepts of QCD

As already mentioned in Sec. 1.2, Quantum Chromodynamics (QCD) is the theory that describes the strong force in the SM. This interaction binds quarks into hadrons, and arises from the requirement of a local non-Abelian gauge invariance under rotation in the color space, $SU(3)_{col}$. Thus all the particles interacting via the strong force carry color charge. The QCD is described by the Lagrangian density:

$$\mathcal{L}_{QCD} = -\frac{1}{4}F_{\alpha}^{\mu\nu}F_{\mu\nu\alpha} + \sum_{f=1}^{n_f}\bar{q}_f(i\gamma^{\mu}D_{\mu} - m_f)q_f \tag{3.1}$$

where the covariant derivative is defined as

$$D_{\mu}q_f = \partial_{\mu}q_f + ig_sG_{\mu\alpha}\frac{1}{2}\lambda_{\alpha}q_f \tag{3.2}$$

and the field strength tensor for the gluon field, $G_{\mu\alpha}$, is

$$F_{\mu\nu\alpha} = \partial_{\mu}G_{\nu\alpha} - \partial_{\nu}G_{\mu\alpha} - g_sf^{abc}G_{\mu b}G_{\nu c} \tag{3.3}$$

In equation (3.1), q_f are the quarks with the sum over the 6 flavors with their corresponding masses represented by m_f. The λ_{α} are the Gell-Mann matrices which allow for the rotation in the color space. g_s is related to the strong coupling constant α_s by the relation $\alpha_s = g_s^2/4\pi^2$. Lastly, f^{abc} are the structure functions of the $SU(3)_{col}$ group.

The QCD Lagrangian contains terms in which the gluon couples to quark-antiquark pairs, in an analogous way to the photon couplings in QED. But, due to the non abelian nature of the

interaction, the third term in the field strength tensor allows also triplet and quartic gluon-gluon interactions. This crucial difference from the QED, where the photons have no charge and hence do not self-interact, leads to many more possible diagrams in QCD (Fig. 3.1) and to the concepts of color confinement and asymptotic freedom.

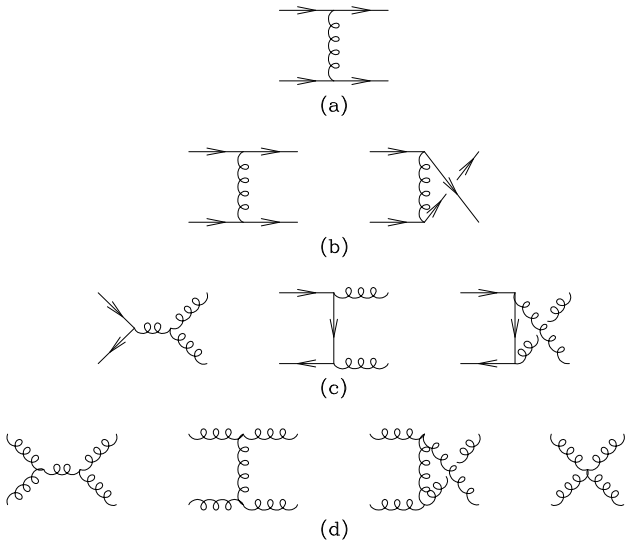


Figure 3.1: Diagrams for the $2 \rightarrow 2$ processes. The diagrams are divided in different sub-processes depending on the initial and final state [35].

3.1.1 Renormalisation, colour confinement and asymptotic freedom

As already mentioned in Sec. 1.2.1, the interaction process can be treated to its lowest order in the coupling constant considered. This is an approximation and it is interpreted as one-quantum exchange mechanism. In order to calculate the complete particle interaction, all the perturbative orders must be summed over. This results in a perturbation series where also loop corrections are added.
In QCD the lowest order corrections to the quark-gluon interaction and gluon-gluon interactions can be obtained through the diagrams shown in Figure 3.2, where an additional gluon is exchanged between the two initial legs of the diagrams.

The loop corrections introduced in the perturbation series for these two interaction processes produce infinite integrals. The series diverge and this effect is known as ultraviolet divergence, due to the fact that the divergence appears because the momentum in the loop is not constrained. To cope with this and to produce physically meaningful (non-infinite) results, an ultraviolet cutoff is introduced, μ_R. This quantity represents the scale at which the QCD is no longer considered to be reliable. The renormalisation procedure consists in effectively subtracting these infinities through counterterms embedded in so-called *bare* parameters that are not measurable. Corrections are introduced to the "renormalised parameters" depending on the renormalisation scale μ_R and on the physics scale at which the experiments are made, taken

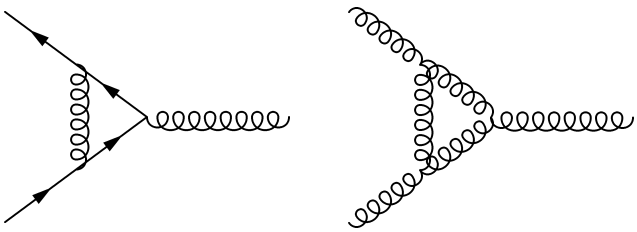


Figure 3.2: Loop diagrams showing the modification to the gluon-gluon and quark-gluon interactions from higher order QCD terms.

as the squared momentum transfer Q^2. Imposing the independence of the final result of the particular choice of renormalisation scale allows to derive an explicit form for the renormalised parameters.
An example of renormalised parameter is represented by the strong coupling constant:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 + \alpha_s(\mu_R^2)\beta_0 ln(Q^2/\mu_R^2)}. \tag{3.4}$$

The coupling constant, initially scale-invariant, *runs* with the energy: becomes a function of the scale of the process. Theoretical and experimental results for the running coupling constant α_s are shown in Figure 3.3.

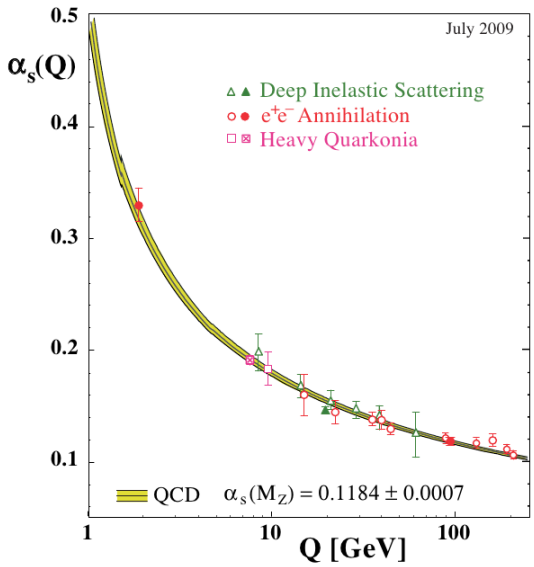


Figure 3.3: 2009 world average of the running coupling constant α_s [36].

As opposed to QED, α_s decreases with increasing Q^2. This result is produced by the β_0 factor

$$\beta_0 = \frac{11N_c - 2N_f}{12\pi} \tag{3.5}$$

where the gluon self interactions lead the factor to be negative. This results in two fundamental

effects: the quark confinement at low Q^2, where the coupling constant is sizable, and the asymptotic freedom due to the decrease of the coupling with increasing momentum transfer.

At low energies, the increase of the strong coupling constant means that the quarks and gluons cannot be observed in isolation, rather they combine to form colourless hadrons. The gluon self interaction explains this effect. In QCD, virtual gluons can self-interact between two separating quarks, and it forces the field lines closer together. This means that the cross sectional area containing the connecting field lines remains roughly constant as the distance increases, as does the number of field lines. Thus the energy increases with the volume between the initial quarks and hence an infinite amount of energy would be required to separate them. When the distance is extended too much, the binding energy stored is large enough to create new hadrons. This consequence of the confinement is called hadronisation.

The asymptotic freedom affects very large momentum transfers, or equivalently, very short distances, where the strength of the colour field decreases and the quarks behave like free fields. Analogously in QED a similar scenario is present at low energies. Hence QCD perturbation theory can be used to describe hard scattering processes, where there are large momentum transfers. This is normally used to test the theoretical predictions.

3.1.2 Parton Distibution Functions and factorisation

In the quark model the nucleons are described as bound states of three quarks, called "valence quarks". The probability for finding a parton[1] of type i carrying a fraction x of the nucleon's momentum between x and $x + dx$ is $f_i(x)dx$, the following equation has to be satisfied for a hadron:

$$\int_0^1 x \sum_i f_i(x)dx = 1 \tag{3.6}$$

where the functions $f_i(x)$ are called Parton Density Functions (PDFs). They are not process dependent and they cannot be predicted in perturbative QCD, hence they need to be measured by experiments. In general, their evolution with x and Q^2 can be determined using splitting functions, formally written down in the DGLAP-equations [38, 39]. This means that the PDFs can be measured in Deep Inelastic Scattering experiments and their subsequent evolution in (x, Q^2) can be measured and compared to the theory.
Figure 3.4 shows the PDFs of the valence quarks of the proton, the gluon, and the sea quarks[2] for two values of the momentum transfer Q^2 at which the proton is probed. The valence quarks carry the largest fraction of the proton momentum, while gluons dominate at low x.

As already mentioned, the particular behaviour of the strong coupling constant allows a perturbative approach for QCD only in the regime where the energy scale of a process is much larger

[1] In the Parton Model [37], the hadrons are considered to be composed by a number of point-like constituents, called *partons*. This model provides a very simple framework for calculating scattering cross sections as well as structure functions for the nucleons. Sea-quark pairs are continuously produced and destroyed in gluon-splitting processes and annihilations, respectively. This constant flux of gluon splits and creations is usually referred as "the sea".

[2] In the Quark Model [40], the hadrons are composed of valence quarks, carrying most of their momentum and contributing to their quantum numbers, and virtual quark-antiquark pairs known as sea quarks.

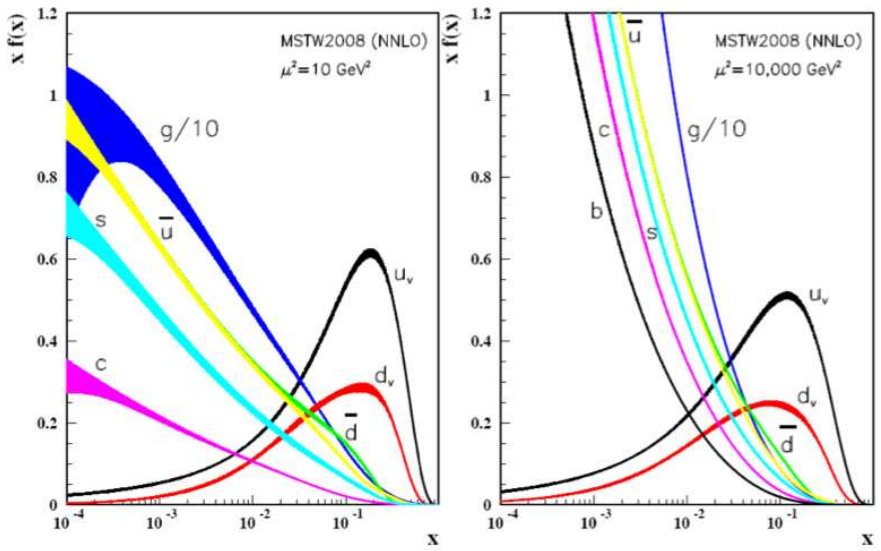


Figure 3.4: Proton PDF functions at transfer momentum $Q^2 = 10$ GeV^2 on the left-hand side, and $Q^2 = 10^4$ GeV^2 on the right-hand side. In the figures $\mu^2 \equiv Q^2$.

than the scale of non-perturbative physics, μ_R^2 (~ 0.3 GeV^2). In order to perform analytic calculations of hadronic processes, the factorisation theorem is needed. This is a crucial concept of QCD which states that a cross section of inclusive processes can be separated into a hard scattering cross section (short distance or large momentum transfer) component and a long-distance component, where the perturbative theory is not applicable. For instance, according to this theorem, the cross section for an *electron-proton* deep inelastic scattering process can be expressed mathematically by

$$\sigma(ep \to eX) = \sum_i C_i^P(x, \alpha_s(Q^2)) \otimes f_i(x, Q^2, \alpha_s(Q^2)) \tag{3.7}$$

where $C_i^P(x, \alpha_s(Q^2))$ represents the perturbative, calculable coefficient function, $f_i(x, Q^2, \alpha_s(Q^2))$ are the PDFs and the sum is over the parton falvors in the interacting hadron. The coefficient functions are characteristic of a given process and are calculable as a series in terms of the coupling constant.

A factorisation scale μ_f must be chosen, such that below it the emissions are included as part of the PDFs and above it the emissions are included in the coefficient function calculation. This is often chosen to be the same as the scale of the physics which is used as the renormalisation scale i.e. $\mu_f^2 = \mu_R^2$.

3.2 Monte Carlo simulations

Monte Carlo (MC) methods simulate physical processes using computer algorithms based on random sampling. These techniques are widely used in a large number of fields.
In particle physics, they are used to simulate particle collisions based on theory, in order to

compare the results obtained from the data to test a given model. The event simulation is composed of two main steps:

- even generation based on prediction from theory;
- simulation of the interactions of the generated particles with the detector volumes.

In particular, the event generation is performed by first calculating the hard process, based on theoretical matrix elements. Then, final state (FSR) and initial state radiation (ISR) is added, and parton radiation produces the parton shower process[3]. Finally the partons are hadronised, their decays are simulated, and the remnants of the proton and the underlying event (UE) is added.
All the different phases can be simulated using different codes, making use of a different matrix element, PDF set, parton shower model or hadronization model. Usually data are compared to a variety of different MC simulations, to better understand and test various models and parametrizations.

As described in sec. 2.2.8, the particle propagation through the detector active and passive volumes is done with GEANT4 [29], which simulates the different types of interactions with the detectors, such as ionisation in the trackers, energy deposition in the calorimeters and decays and radiation along the particle trajectory.
The simulation samples used for this thesis have been produced with a mean number of pile-up interactions per bunch crossing of 20.
The same event reconstruction that is run on the data from the real detector, is then used with this simulated data.

The MC event generators used in this thesis are Pythia, HERWIG++ and POWHEG. Their main characteristics and the differences between them will be briefly discussed in the following.

3.2.1 PYTHIA

PYTHIA version 8 [41] uses perturbative QCD at leading order to compute matrix elements for $2 \to 2$ processes. The parton shower is p_T ordered and interleaved with multiple interactions. Finally the hadronisation is done with the Lund string model [42]. The generator is developed in the C++ programming language, which is now most commonly used for analysis. In PYTHIA 8, multiple interactions are interleaved with both initial-state and final-state radiation, while in PYTHIA 6 (the previous version of the generator) the final-state radiation was not included.
The non-perturbative physics processes, such as ISR and FSR, hadronization, the UE and colour connections, can be tuned to different data.

3.2.2 HERWIG++

HERWIG++ [43] also uses leading order matrix elements in perturbative QCD, but the parton shower in this case is angular ordered and then cluster hadronisation is used. It does not

[3]This process happens when two coloured objects separate during hard scattering processes. The force between the two partons increases together with the probability to radiate and lose some energy. Thus single quarks produced by the hard scatter will result in a whole shower of partons and this process is referred to as "parton showering".

interleave the multiple parton interactions with the parton showers as PYTHIA does. As HERWIG and PYTHIA use very different parton shower and hadronisation models it is useful to compare the two, and their differences are also useful in assessing systematic uncertainties.

3.2.3 POWHEG

POWHEG [44] interfaces next-to-leading-order (NLO) calculations with a parton shower generator, such as PYTHIA or HERWIG. This essentially means that the emission of an additional hard parton is included in the matrix element calculation, rather than only occurring as part of the parton shower. As a result, the modeling of a potential third jet in an event should be more accurate while keeping the flexibility to have more than three jets in the event, which would not be possible with just an NLO calculation.

Chapter 4

Jet reconstruction

Unlike the physical particles that are reconstructed by HEP detectors, jets are not formed by a single particle. They can be considered as tools designed to provide a representation of underlying QCD processes.
In the first part of this chapter a description of the most important jet clustering techniques will be provided, while at the end of the chapter, the jet calibration will be introduced.

4.1 Jet reconstruction

As described in Chapter 3, colour confinement implies that coloured quarks cannot be observed as free particles. Only colourless hadrons can be observed. This means that although coloured quarks can be extracted from a hadron through a hard scattering process, they will recombine quickly into courless hadrons. In particular, a single quark produced in the hard scatter will result in a whole shower of partons (parton showering). When the energy of the shower decreases, coloured partons group themselves into hadrons (hadronisation). In practice, the scattering of a coloured parton in the hard interaction will be observed as a shower of hadrons. The resulting hadrons are detected by the experiments and their combination ideally reflects the physics of the originating parton. Several clustering algorithms are used to collect all the hadrons into jets, which are described in the following.

4.1.1 Inputs to jet reconstruction

The inputs to jet reconstruction are either stable particles with a lifetime of at least 10 ps (excluding muons and neutrinos) in the case of Monte Carlo (MC) generator-level jets (also referred to as particle-jets or truth-jets), charged particle tracks in the case of so-called track-jets [45], or three-dimensional topological clusters (topo-clusters) in the case of fully reconstructed calorimeter-jets.

When particle-jets are reconstructed, the different masses of stable particles used as input of the jet clustering algorithm, such as pions or protons in the simulation, are taken into account. Instead, for track-jet reconstruction, given the very high probability of producing pions in the hadronisation process, tracks are assigned the pion mass when used as input for jet reconstruction. The following quality selection criteria are applied to ensure that only good

quality tracks coming from the hard scattering vertex[1] are used to build track-jets:

- transverse momentum: $p_T^{track} > 0.5$ GeV;
- transverse impact parameter: $|d_0| < 1.0$ mm;
- longitudinal impact parameter: $|z_0| \times \sin(\theta) < 1.0$ mm;
- silicon detector hits on tracks: hits in pixel detector $\geqslant 1$ and in the silicon strip detector $\geqslant 6$;

where the impact parameters are computed with respect to the hard scattering vertex, and θ is the angle between the track and the beam.
In order to reconstruct calorimeter jets, calorimeter cells are clustered together using a three dimensional topological clustering that includes noise suppression. Topo-clusters are considered as massless four-momenta, such that $E = |\overrightarrow{p}|$. Depending on their shape, depth and energy density, they are classified as either electromagnetic of hadronic. In the calibration procedure, corrections are applied to the energy in order to calibrate the clusters to the hadronic scale [46].

Figure 4.1 shows a schematic overview of the ATLAS jet reconstruction.

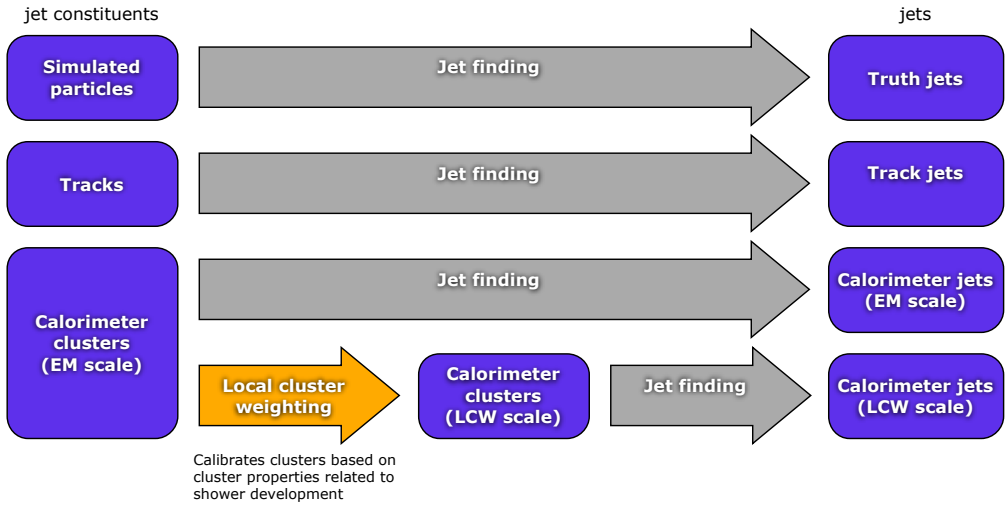


Figure 4.1: Overview of the ATLAS jet reconstruction. After the jet finding, the jet four-momentum is defined as the four-momentum sum of its constituents. LCW is the local cluster calibration [47]. LCW is described in Sec. 4.1.3.

4.1.2 Topo-Clustering technique

The *topological cell clustering*, or topo-clustering, is a technique by which the individual calorimeter cells are topologically combined to follow the shower development produced by the hadronisation of hard partons. The topo-cluster forming procedure is designed to adapt to the

[1]The hard scattering vertex is selected as the primary vertex that has the largest $\sum(p_T^{track})^2$ in the events and that contains at least two tracks.

stochastic processes of electromagnetic and hadronic showers, and thus does not require the energy deposits to be distributed in a regular shape. The basic idea of topological clustering is to group together neighboring cells that have significant energies compared to the expected noise. In order to do this, an estimate of the average amount of noise in a typical calorimeter cell, σ_{noise}, is needed. At the beginning of the LHC data taking, σ_{noise} was considered as the amount of electronic noise in a given cell of the calorimeter. During the LHC Run 1 data taking, the increased level of pile-up activity in the calorimeters dominates the level of the electronic noise, and $\sigma_{noise} \approx \sigma_{noise}^{pile-up}$. The average expected noise allows for the definition of a cell significance, $\zeta = E_{cell}/\sigma_{noise}$. It defines the probability that an observed cell energy comes from noise, and thus provides the signal-to-noise ratio for each calorimeter cell.

The algorithm to form topological clusters from a list of calorimeter cells[2] consists in the following steps:

1. **Find seeds** : Identify cells with $|\zeta| > 4$ and put them into a seed list. Each seed cell forms a "proto-cluster".

2. **Find neighbors** : All the cells identified as seeds are ordered decreasingly in $|\zeta|$. For each seed, its neighboring cells are considered. The cells not included in the seed list, and satisfying the requirement $|\zeta| > 2$, are added to a neighbor seed list and included in the adjacent proto-cluster. If a cell is adjacent to more than one proto-cluster, these proto-clusters are merged. If the signal to noise ratio is above the cell threshold and is between 0 and 2, the cell is included only in the proto-cluster identified as the more significant neighbour to this cell. After the processing of all the seed cells, the original seed list is discarded and the neighbour seed list is used at its place. This procedure is repeated until the seed list is empty.

   The typical definition of neighboring cells includes the eight cells surrounding the seed within the same calorimeter layer. In ATLAS, this number is often larger as the granularity varies between different calorimeter layers and regions. By default, the expanded definition of neighbouring cells is used.

3. **Finalize**: The remaining proto-clusters after the merging are sorted in descending order in E_{T} and converted to clusters. Those with E_{T} (optionally $|E_{\mathrm{T}}|$) less than a given threshold are removed at this step.

In the standard ATLAS reconstruction, two types of topological clusters are built: the electromagnetic "633" clusters and the combined "420" clusters, where the three digits represent the ζ threshold for seed cells, neighbor cells, and minimal signal-to-noise ratio cell, respectively.

Once the raw topo-clusters have been built, the next step is to search for local maxima within each topo-cluster. This is fundamental for the event topologies studied in this thesis, such as collimated Z or W boson hadronic decays, where two energetic quarks can be produced in close proximity. These two quarks can then form a single raw topo-cluster with local maximum energy peaks representing the leading and sub-leading quarks. This ability is crucial for both a proper reconstruction of typical small jets as well as to be able to perform advanced jet substructure techniques. In particular, in order to avoid fluctuations, local maxima are required to have at least $E_{cell} > 500$ MeV, at least four neighbors, and no neighbors with a larger energy signal. The local maximum is also required to occur within the middle layer of the electromagnetic or

[2]Usually all the cells are considered, but they may also be a subset of cells defined by a "region of interest"

hadronic calorimeters, which is much deeper than the first and last layers, and which can thus provide a more reliable energy measurement.

After topo-clusters are constructed, the energies of all of the constituent cells are summed to get the total energy of the topo-cluster. The angles are then set to match the energy-weighted centroid of the topo-cluster, and the mass is set to zero. This then sets the full four-momentum of the topo-cluster.

4.1.3 Local cluster calibration

In order to further refine the performance of topo-clusters and to restore the hadronic energy scale (see Sec. 2.2.5), a Local Calibration Weighting (LCW or LC) [48] scheme has been constructed. This correction is based on MC simulations for single charged and neutral pions, keeping the detector geometry and the topo-cluster reconstruction the same as in the nominal MC simulation samples. The truth particles are generated with a flat distribution in the logarithm of pion energies from 200 MeV to 2 TeV. The corrections are derived with respect to the simulated true deposited energy in the active and inactive detector region, usually referred as "calibration hits". The ATLAS software classifies the true energy depositions in three categories:

- *visible energy*, like the energy deposited by ionization;
- *invisible energy*, the energy absorbed in nuclear reactions;
- *escaped energy* carried away by neutrinos, but recorded in the place where the neutrino is produced.

The local calibration process starts by classifying clusters as mainly electromagnetic or hadronic depending on cluster shape variables. According to the MC predictions, these variables characterize the topology of the energy deposits of electromagnetic and hadronic showers and are defined as observables constructed from the positive cell contents of the cluster and the cluster energy. In particular, the observables considered for the classification are: cluster energy, cluster depth in the calorimeter, and average cell energy density.
Each cluster receives a classification weight that is calculated for small $|\eta|$ regions by mixing neutral and charged pions with a ratio of 1:2. This assumes that 2/3 of the pions should be charged. Clusters are classified as mostly electromagnetic or mostly hadronic and the calculated weight denotes the probability p for a cluster of originating from a hadronic interaction.
From the true energy deposits of simulated pions, the hadronic energy correction weights, w_{HAD}, are calculated and the applied cluster weight is

$$w_{HAD} \cdot p + w_{EM} \cdot (1 - p), \tag{4.1}$$

where $w_{EM} = 1$. For each calorimeter layer dedicated correction weight tables are used. These are binned logarithmically in topo-cluster energy and cell energy density ($\rho_{Cell} = E_{Cell}/V_{Cell}$). The hadronic weighting process is completed by the *out-of-cluster* (OOC) and the *dead material corrections* (DM). OOC corrections are applied to correct for energy deposits inside the calorimeter but outside calorimeter clusters due to the noise thresholds applied during cluster

making. These corrections depend on $|\eta|$, cluster energy and the cluster depth in the calorimeter. DM corrections are applied to compensate for energy deposits in materials outside of the calorimeters. In particular, for energy deposited in the outer cryostat wall between the electromagnetic and hadronic barrel calorimeters a correction is derived performing the geometrical mean of the energies in the samplings just before and just beyond the cryostat wall. When there is not a clear correlation between energy deposits and cluster observables, the corrections are obtained from lookup tables binned in cluster energy, pseudorapidity, and shower depth. Two sets of DM weights for hadronic and electromagnetic showers are used. The weights are applied according to the classification probability p defined above. All corrections are defined with respect to the electromagnetic scale energy of the topo-cluster.
Since only calorimetric information is used, the LCW scheme does not account for low-energy particles which do not create a topo-cluster in the calorimeter. This is, for instance, the case when the energy is absorbed entirely in inactive detector material or particles are bent outside of the calorimeter acceptance.

4.1.4 Jet clustering algorithms

Jet algorithms provide a set of rules for grouping jet constituents into jets. They usually involve one or more parameters that indicate how close two particles, or two jet constituents must be for them to belong to the same jet. The association is performed with a recombination scheme, which indicates which momentum to assign to the combination of two particles. The simplest scheme is the 4-vector sum. Taken together these two ingredients, a jet algorithm and a recombination scheme, form a "jet definition".

In 1990, an agreement was set out by a group of theorists and experimentalists, the "Snowmass accord" [49], in order to identify some general properties of jet definitions. The main properties that should be met by a jet definition are [50]:

1. simple to implement in an experimental analysis;
2. simple to implement in the theoretical calculation;
3. defined at any order of perturbation theory;
4. yields finite cross sections at any order of perturbation theory;
5. yields a cross section that is relatively insensitive to hadronisation.

The first two points are purely practical, as an algorithm is not useful if it is not straightforward to use to calculate theoretical results and make experimental measurements. The remaining points deal with the requirement that the jet be both Infra-Red (IR) and collinear safe, meaning that the addition of a soft particle or the collinear splitting of a hard particle should not change the number of reconstructed hard jets in the event. A description of these two requirements will be provided. Further explanations are provided in Section 2.1.4 of Reference [51].

IR safety requires hard jets not to change if an additional soft emission occurs. This is fundamental given the frequency with which soft gluons are produced in hadronic showers. This effect is schematically represented in Fig. 4.2, where a soft emission between the two jets from

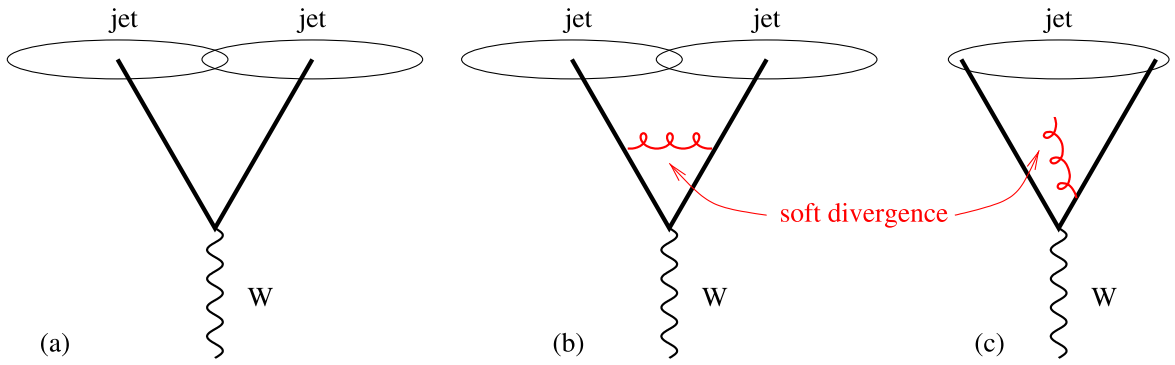


Figure 4.2: The hadronic decay of a W boson into two hard quarks is reconstructed in two separate jets (a). After the emission of a soft gluon, an IR-safe jet algorithm still observes two separate jets (b), while IR-unsafe algorithms merge the two jets into one (c) [51].

a W boson hadronic decay merges everything into a single jet, and which would not have occurred if the soft emission was not present. More generally, the addition of a soft radiation, which contributes with a negligible amount of energy to the jet, can cause the observed hard jet multiplicity to change by modifying the way in which the hard energy deposits are grouped. For instance, a simple example of a jet algorithm where this would happen would start the clustering from the softest partons and then combine objects based purely on angular distance, with a fixed maximum distance cutoff.

Collinear safety represents stability against the single parton splitting process, such as a quark radiating a gluon, a gluon fragmenting into two quarks, or other similar processes. It is crucial for building the jets, because such splittings occur very often in the single hadronic shower produced by hard-scatter partons. Fig. 4.3 shows the way in which this phenomenon can change the number of reconstructed hard jets for algorithms that are not stable against the collinear soft parton emission. In general, the splitting of a hard object into multiple softer objects may cause the observed hard jet multiplicity to change. This occurs due to a change in the way in which the hard energy deposits are clustered together by the algorithms. A simple example of an algorithm where this occurs is the cone algorithm: it starts with the hardest partons and then combines objects based purely on the angular distance, with a fixed maximum distance cutoff. The splitting of the hardest object into two softer objects thus changes the starting point (the seed), and the new hardest parton becomes the geometric centre of the jet. Then, the two parts of the previously hardest object are combined into this jet as they are within the distance cutoff, while the residual hard partons are too far away from the seed to be included. This remaining energy is then collected into a second jet.

As shown, a jet algorithm which is either IR or collinear unsafe can produce a variation of jet multiplicity and therefore a change in the observed physics. There are currently many jet algorithms which are both IR and collinear safe, including *cone-type* algorithms and *successive recombination* algorithms, and the majority of them are implemented in FastJet [52]. The ATLAS jet reconstruction software makes extensive use of the FastJet software library due to its simple and uniform interface, as well as the large selection of functionality and the continuos development.

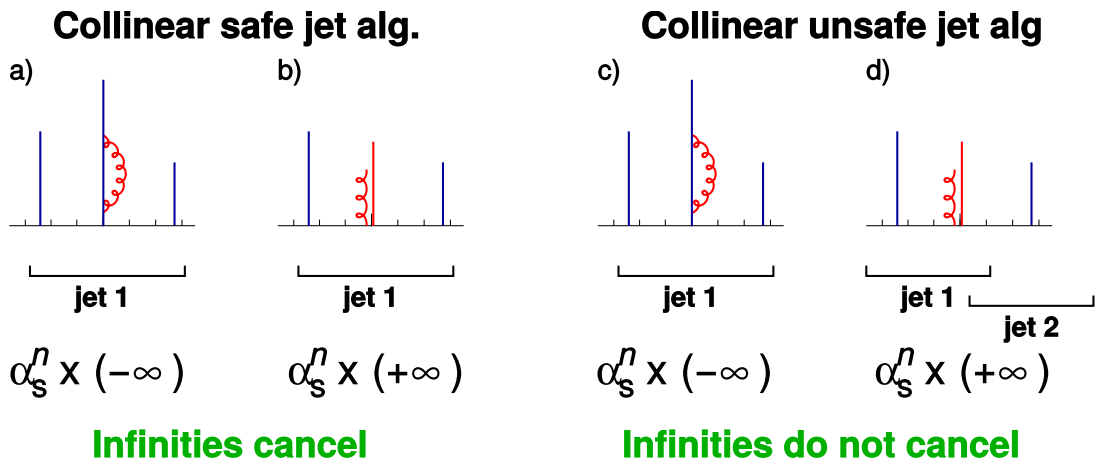


Figure 4.3: The comparison of two processes shows the difference between collinear safe and unsafe algorithms. Parton p_T is represented by the height of the vertical lines, while the horizontal axis represents the parton rapidity. In the presence of a gluon correction to the quark propagator (a and c), since no particle is emitted, the jet clustering algorithms are stable. In perturbation theory, the gluon correction should be cancelled by gluon radiation, where the gluon and parton have the same p_T. For collinear safe algorithms (b) this feature is provided, while for collinear unsafe algorithms (d) the perturbative approach breaks down. The observed hard jet multiplicity, for a collinear unsafe algorithm, is influenced by the presence of an extra parton, leading to the reconstruction of two separate jets, rather than one [51].

Cone-type algorithms

A cone algorithm is designed following the simple idea of taking a cone with a fixed radius, $R = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$, around a particle. The size of the cone is chosen such that the energy contained in the cone is maximised. This radius optimisation can be done iteratively, by taking a seed particle, combining the momenta of all the particles in a cone around it to define a new axis and repeating until the axis remains stable. This makes the resulting jets dependent on the choice of seed. Another problem of this family of algorithms arises when two cones overlap and a decision must be taken whether to merge the jets or which particles should be assigned to each one. As a result, cone algorithms are generally affected by additional soft emissions and hence are not IR safe (with the exception of SIScone [53]). For this reason, cone algorithms are currently not often used in physics analyses.

Successive recombination algorithms: the k_T-family

Successive recombination algorithms are the current the most used method for forming jets. These algorithms attempt to work backwards through the branchings that occur in the parton shower by merging two particles together at each step. A distance d_{ij} is introduced between entities (particles, pseudojets) i and j and a distance d_{iB} between entity i and the beam (B). The (inclusive) clustering proceeds by identifying the smallest of the distances. If it is a d_{ij}, entities i and j are recombined, while if it is d_{iB}, i is called a jet and it is removed from the list of entities. The distances are recalculated and the procedure repeated until no entities are left.

The algorithms included in the "k_T-family" lie on the following definition of distance measures:

$$d_{ij}^2 = min(p_{T_i}^{2p}, p_{T_j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.2}$$

$$d_{iB} = p_{T_i}^{2p} \tag{4.3}$$

where $\Delta R^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and p_{Ti}, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i. In addition to the usual radius parameter R, the parameter p defines the contribution of the constituents p_T in the definition of the jet distance parameter d_{ij}. Different values of p can be used, resulting in different algorithms. Anti-k_T algorithm [54] for $p = -1$, Cambridge/Aachen (C/A) algorithm [55] for $p = 0$, and k_T algorithm for $p = 1$ represent the most widely used infrared and collinear-safe jet algorithms available for hadron-hadron collider physics today.

A full understanding of the algorithm properties is crucial for the implementation of jet substructure techniques that will be discussed in the next chapter. Fig. 4.4 shows the same event clustered in the $\eta - \phi$ space using different jet clustering algorithms. It can be seen that the

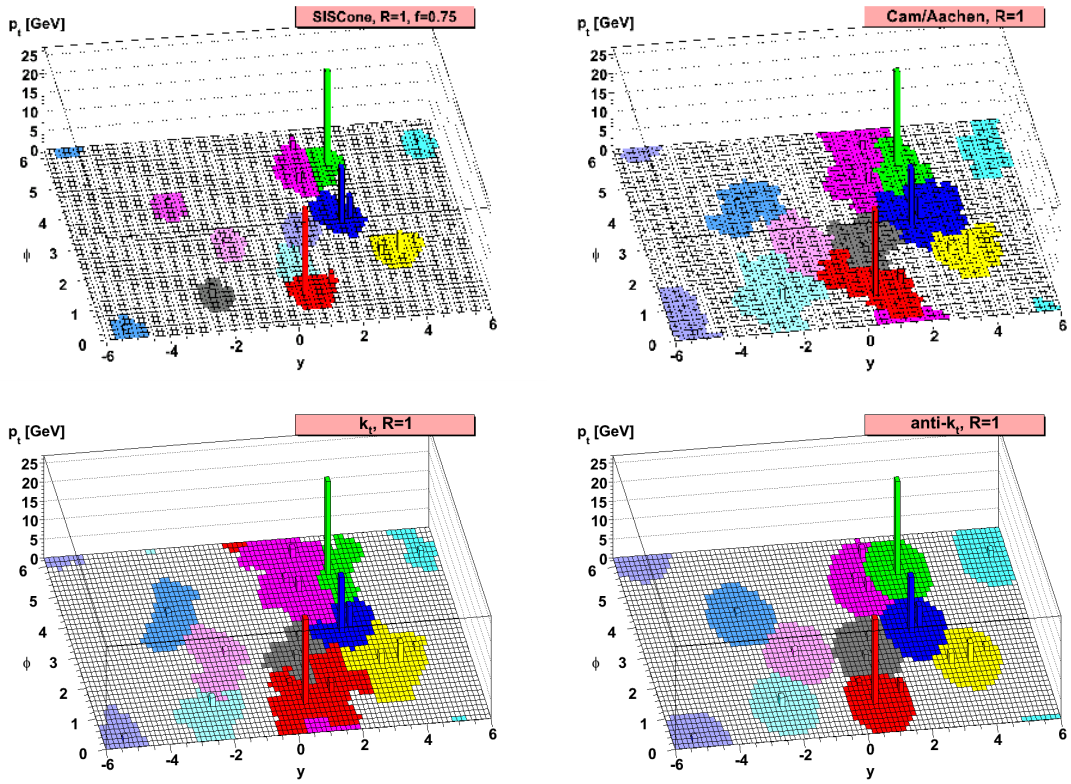


Figure 4.4: The same event clustered using different jet clustering algorithms, SiSCone (top left), Cambridge-Aachen (top right), k_T (bottom left) and anti-k_T (bottom right) in $y - \phi$ space [54].

four algorithms considered (SiSCone, C/A, k_T, and anti-k_T) generally agree about the hardest

jets, although combine different soft constituents into them.

4.2 Particle level jet calibration

The calorimeter jets are built from the topo-clusters with positive energy entering as massless particles in the jet algorithm. As discussed in Sec. 4.1.2, the topo-clusters are initially reconstructed at the EM scale, which correctly measures the energy deposited in the calorimeter by particles produced in electromagnetic showers. The LCW calibration (see Sec. 4.1.3) not only restores the correct jet energy scale, but aims also at an improved resolution compared to the EM scale by correcting the signal from hadronic deposits, and thus reduces fluctuation due to the non-compensating nature of the ATLAS calorimeter [47].
However, the LCW is not sufficient to fully correct all effects and ATLAS implemented a full calibration scheme for calorimeter jets, which restores the jet energy scale to that of jets reconstructed from stable simulated particles (truth particle level). This procedure consists of the following four steps:

- **Pile-up correction**
  Pile-up interactions cause an energy offset in the jet formed from topo-clusters at the EM and LCW scale. A correction is defined to account for this offset produced by both additional proton collision in a recorded event (in-time pile-up) and by past and future collisions influencing the energy deposited in the current bunch crossing (out-of-time pile-up). This correction is derived from MC simulation as a function of the number of primary vertices in the event and the expected number of interactions (μ, sensitive to out-of-time pile-up) in bins of jet η and p_T.

- **Origin correction** This correction is applied on the jet direction to make the jet pointing to the primary vertex instead of the nominal centre of the ATLAS detector.

- **Jet calibration based on MC simulation** The calibration of energy and η of a reconstructed jet is derived comparing these quantities to the corresponding ones of the matching truth jet in MC simulations. In particular, the ratio

$$\mathrm{R}^{EM(LCW)} = E_{jet}^{EM(LCW)} / E_{jet}^{truth}, \tag{4.4}$$

  calculated for various jet energies as a function of the jet pseudorapidity η_{det} measured in the detector reference frame, is the average energy response and represents the inverse of the energy calibration function. It can be applied to jets formed from topo-clusters at EM or at LCW scale with the resulting jets being referred to as calibrated with EM+JES or with the LCW+JES scheme.

- **Residual in situ corrections** The remaining data-to-MC calibration differences are assessed using in situ techniques exploiting the transverse momentum balance between the jet and a well-measured reference object.
  For instance, jets in the central region are calibrated using photons or Z bosons as reference objects up to 800 GeV, while jets with higher p_T are calibrated using systems of low-p_T jets recoiling against a high-p_T jet.

Figure 4.5 shows an overview of the ATLAS calibration scheme for calorimeter jets used for the 2011 dataset.

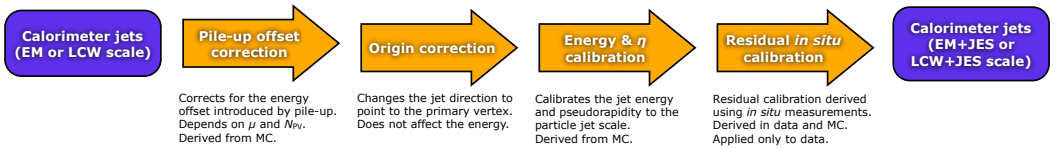


Figure 4.5: Overview of the ATLAS jet calibration scheme used for the 2011 dataset. The pile-up, absolute JES and the residual in situ correction calibrate the scale of the jet, while the origin and the η corrections affect the direction of the jet [47].

Chapter 5

Jet Substructure

The new energy regime explored at the LHC enables the production of Lorentz-boosted heavy particles, whose hadronic decay products can be reconstructed as one large-area jet. Studying the internal structure of jets allows to go beyond the four-momentum description of a single parton and yields new approaches to test QCD and to search for new physics in hadronic final states. In this chapter the main techniques of jet grooming and tagging will be described and particular emphasis will be given to the tagging of hadronically decaying boosted vector bosons, as they are crucial for the analysis described in this thesis.

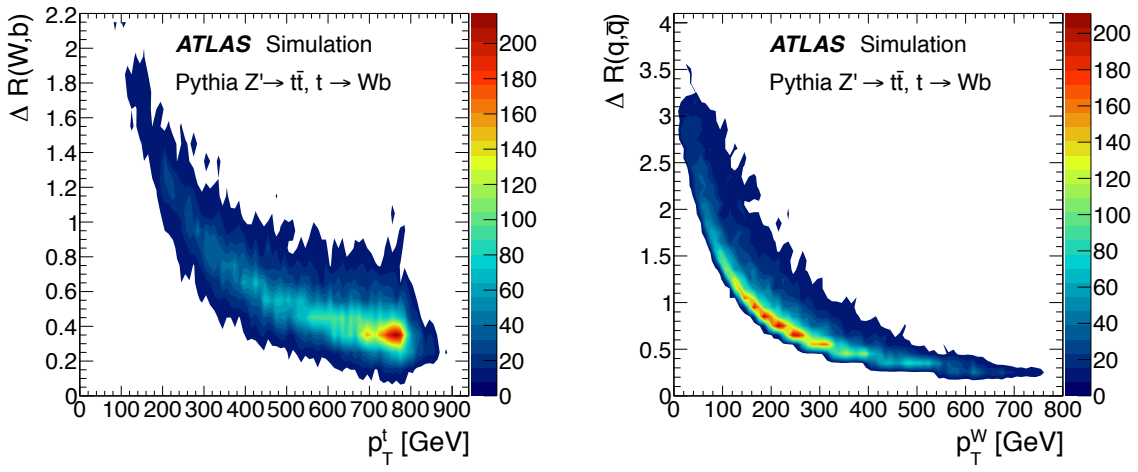
5.1 Motivations to investigate the jet substructure

The center-of-mass energy of the LHC opens a new kinematic regime to experimental study. SM particles can be produced in a new phase space, and new massive particles could decay to highly boosted SM particles. These signatures necessitate the development of new techniques to conduct measurements in boosted final state topologies. As shown in fig. 5.1, the decay products of W bosons, top quarks, and Higgs bosons, if sufficiently boosted, can become collimated to the point that standard reconstruction techniques begin to fail. For example, if a W boson has a high transverse momentum, the two quarks from its decay product can be at a distance in the $y - \phi$ plane smaller than 0.4. This case would be impossible to resolve using the standard ATLAS jet reconstruction based on anti-k_{T} jets with a distance parameter of 0.4.

In the case of hadronic decays, the decay products of a significantly Lorentz-boosted object tend to merge into a single jet. From basic kinematics, the angular separation ΔR between two massless objects, namely quarks, from the decay of a boosted massive object X can be expressed approximately as

$$\Delta R(q\bar{q}) \approx \frac{2m(X)}{p_T(X)} \tag{5.1}$$

where $\Delta R = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$, and p_T and m are the transverse momentum and the mass, respectively, of the decaying particle. Moreover, the high luminosity conditions at the LHC often produce multiple *proton-proton* interactions per bunch crossing that add soft particles unrelated to the hard scattering. These particles can contaminate jets, increasing the quantity



(a) $t \to Wb$ (b) $W \to q\bar{q}$

Figure 5.1: (a) The angular separation between the W boson and b-quark in top decays, $t \to Wb$, as a function of the top-quark transverse momentum (p_T^t) in simulated PYTHIA [41] $Z' \to t\bar{t}$ ($m'_Z = 1.6$ TeV) events. (b) The angular distance between the light quark and anti-quark from $t \to Wb$ decays as a function of the p_T of the W boson (p_T^W). Both distributions are at the generator level and do not include effects due to initial and final-state radiation, or to the underlying event.

of energy clustered and considerably degrading the jet mass resolution, weakening dramatically the sensitivity to new physics processes. This effect increases in the case of large-R jets, as it is proportional to the jet geometrical area.

As discussed in the previous chapter, the aim of the jet reconstruction is to identify and describe partons from hard scattering processes with single four-momenta. This approach simplifies a very complicated picture, given the presence of a very high particle multiplicity in the *proton-proton* collision, into a small set of four-vectors. The goal of the jet substructure technique is to go beyond the simple four-momentum description, in order to exploit the full potential of the high spatial granularity of electromagnetic and hadronic calorimeters, and identify decay products of new or existing particles within the jet.
Investigating the jet substructure, a new set of observables describing the energy distribution into the jet can be defined and used to discriminate between signal and background events. In the next section, a description of the most important *jet filtering* techniques (grooming) and of the substructure observables will be provided, with a particular focus to the BDRS-A split filtering technique, which is the key point of the ATLAS heavy diboson resonance search event selection.
A complete description of the state-of-the-art of the jet substructure techniques can be found in References [45, 51, 56].

5.2 Algorithms, grooming and tagging variables

The implementation and development of the jet substructure is focused on the successive recombination algorithms, the k_T-family, described in Sec. 4.1.4.
In the case of the k_T and C/A algorithms, given the p_T dependence on the distance measures, the clustering history of the algorithm (the ordering and structure of the pair-wise subjets recombinations made during jet reconstruction) provides spatial and kinematic information about the substructure of that jet.
As already mentioned, for jets reconstructed with the anti-k_T algorithm the highest-p_T constituent are clustered first. This yields very stable and circular jets, making this algorithm the most commonly used in ATLAS and CMS analyses with hadronic final states. On the other hand, the structure of the jet as defined by successive recombinations carried out by the anti-k_T, provides little or no information about the p_T ordering of the shower or wide angular-scale structure. In order to take advantage of the anti-k_T jet stability for analyses searching for highly boosted topologies, these jets are used to select events based on jet kinematics (η and p_T), and then the jet constituents are reclustered with the k_T algorithm to make use of the k_T-ordered splitting scales and of advanced substructure selections.
The four-momentum recombination scheme is used in all cases and the jet finding is performed in rapidity-azimuthal angle (y-ϕ) coordinates.

5.2.1 Jet grooming algorithms

In Sec. 4.1, the basic idea of jet reconstruction has been discussed. While the jets are reconstructed, other particles not coming from the process of interest might be clustered inside the jets. In particular, there might be contributions from initial state radiation, pile-up, and multi parton interactions that lead to a degradation of the resolution of the properties of the jet.
Jet grooming is the application of an algorithm intended to recover the resolution of the jet and better define the constituents as those from the interesting boosted physics object. Three main types of jet grooming have been implemented and are presently in use in ATLAS:

- splitting and filtering (split-filtering)
- trimming
- pruning.

In the next sections, a description of these grooming strategies is provided.

Split-filtering

Split-filtering is usually referred as *mass-drop filtering* or "BDRS" [57], representing the initials of the authors of the paper where this technique was first described. The BDRS algorithm was developed for the search for associated production of a Higgs boson with a W or Z boson with the subsequent decay $H \to b\bar{b}$. Therefore, the technique is focused on fat jets with a two-body-like substructure and is useful for W/Z tagging as well.

This procedure is aimed at finding isolated energy clusters in a jet, through the identification of subjets which have to be symmetric, and with a significantly smaller mass than the jet. This

latter characteristic is called mass-drop. It was developed and optimized using a C/A jets, because this algorithm provides information on the angular ordering of the shower constituents. The C/A algorithm, indeed, clusters showering products together based on their distance in the $y - \phi$ plane. Because of this, the presence of potentially large splittings within the final jet can be easily identified.
The requirement placed on the fat-jet internal structure is able to retain its hard components without applying explicit selection to reject soft-p_T or wide-angle jet constituents.
In particular, the split-filtering procedure has two stages:

- **Mass-drop and symmetry** The last stage of the C/A is undone and the jet "splits" into two sub-jets, j_1 and j_2, ordered according to their mass: $m^{j_1} > m^{j_2}$. A significant mass-drop is required such that:

$$m^{j_1}/m^{jet} < \mu_{frac}, \tag{5.2}$$

where μ_{frac} is a parameter of the algorithm. A relatively symmetric splitting is also required through the cut on the momentum balance

$$\frac{min[(p_T^{j_1})^2, (p_T^{j_2})^2]}{(m^{jet})^2} \times \Delta R^2_{j_1,j_2} > y_{cut}, \tag{5.3}$$

where $\Delta R_{j_1,j_2}$ is the j_1, j_2 angular distance, and y_{cut} defines the energy sharing between the two sub-jets in the original fat-jet. This last requirement forces both the sub-jets to carry some significant fraction of the momentum of the original jet. Figure 5.2 represents a schematic view of this procedure. If the mass-drop and symmetry criteria are not satisfied, the jet is discarded.

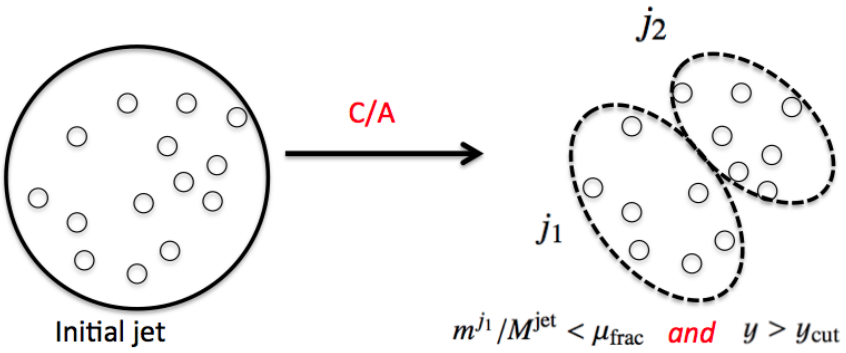


Figure 5.2: The mass-drop and symmetric splitting criteria [45].

- **Filtering** After identifying the two hardest proto-jets matching the mass-drop and symmetry criteria, j_1 and j_2 are reclustered using C/A algorithm with distance measure $R_{filt} = min[0.3, \Delta R_{j_1,j_2}/2]$. All the constituents outside the three hardest reclustered sub-jets are discarded: the jet is then filtered. The choice of three sub-jets allows the capture of one additional deviation from a two-body decay due to the presence of radiation. By dynamically choosing the angular scale of the sub-jet reclustering, the ability to resolve the substructure and the sensitivity to highly collimated decays is maximized. The filtering mechanism is illustrated in figure 5.3

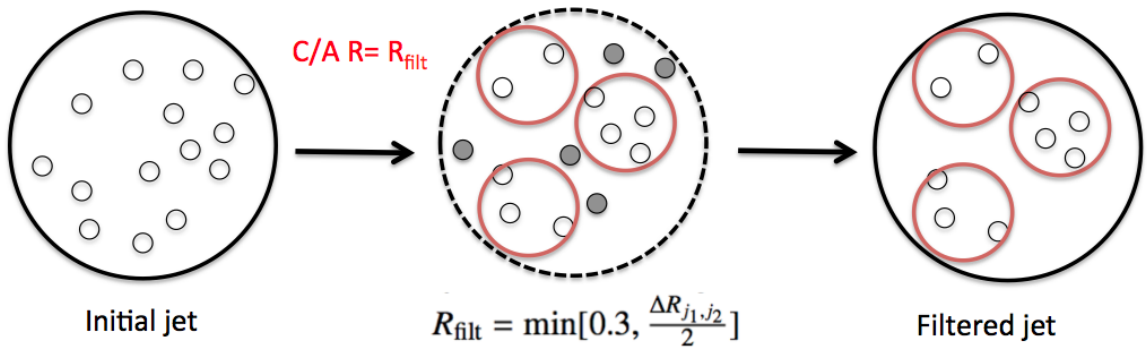


Figure 5.3: Schematic view of the filtering stage for the BDRS procedute [45].

Trimming

The initial motivation to study trimming [58] came from the necessity to improve the performance of jet algorithms in identifying and reconstructing hard partons. In the ideal case of no pileup or multi-parton interactions (MPI), and no initial state radiation (ISR), the best jet algorithm is the anti-k_{T}. Moreover, if a distance parameter R$\sim$1 for such algorithm is used, the bulk of the final state radiation (FSR) coming from the parton shower, can be captured. As expected, this ideal situation does not occur when pile-up, MPI and ISR are included in the event generation.

In order to mitigate this effect, the trimming algorithm takes advantage of the fact that all the sources of contamination in the reconstructed jets mentioned above are often much softer than the outgoing partons associated with the hard-scatter and their FSR, which represent the real target of jet reconstruction.
The removal of this sort of contribution is achieved by reclustering the constituents of the original anti-k_T fat-jet into a set of k_T small-R jets. The distance measure of the reclustered sub-jets is a parameter of the algorithm and it is usually referred to as $R_{sub-jet}$. Any sub-jets are removed if

$$p_{T_i}/p_T^{jet} < f_{cut} \tag{5.4}$$

where p_{Ti} is the transverse momentum of the i^{th} sub-jet, and f_{cut} is the second parameter of the method, which is typically a few percents.
The trimming procedure is illustrated in figure 5.4.

Many combinations of the trimming parameters have been studied and optimized in the context of analyses using jet substructure techniques to reconstruct boosted heavy objects. Some of these results are summarize in Reference [45]. For instance, low-mass jets ($m^{jet} < 100$ GeV) from light-quarks or gluons lose typically $30-50$ % of their mass in the trimming procedure, while this reduction for jets containing the decay products of a boosted object is much smaller and mainly due to the removal of pile-up or underlying event (this effect increases with the number of PV in the event). The effect of trimming and BDRS are shown for both QCD dijet and $Z \to q\bar{q}$ simulated events in Fig. 5.5. Moreover, figure 5.5 shows also a clear improvement in jet mass resolution due to the two algorithms.

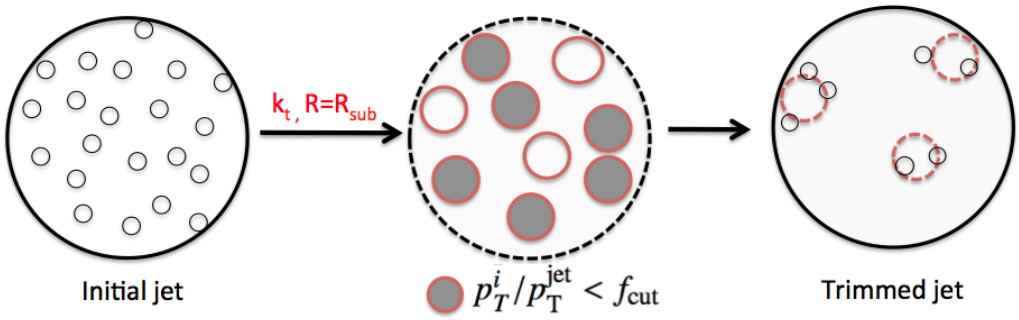
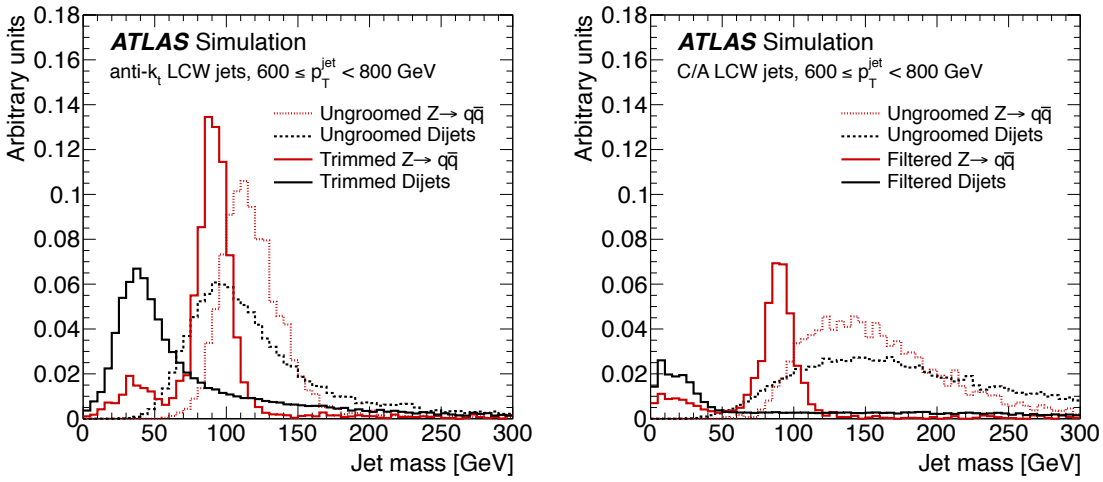


Figure 5.4: Diagram depicting the jet trimming procedure [45].



(a) anti-k_t, $R = 1.0$ (b) C/A, $R = 1.2$

Figure 5.5: Leading-p_T^{jet} jet mass for simulated HERWIG $Z \to q\bar{q}$ signal events (red) compared to POWHEG+PYTHIA dijet background events (black) for jets in the range 600 GeV $\leqslant p^{jet_T} <$ 800 GeV. The Monte Carlo generators used for the simulations are described in detail in sec. 3.2. The dotted lines show the ungroomed jet distributions, whereas the solid lines show the (a) trimmed and (b) mass-drop filtered jet distributions. The trimming parameters are $f_{cut} = 0.05$ and $R_{sub-jet} = 0.3$ and the mass-drop filtering parameter is $\mu_{frac} = 0.67$ [45].

Pruning

Similarly to trimming, the pruning algorithm removes constituents with a small relative p_T, but also adds a veto on wide-angle radiation. The procedure starts after finding a large-R jet with the anti-k_T or C/A algorithm. The jet constituents are then reclustered using either a C/A or a k_T algorithm in order to reverse the clustering stages and identify the angular scale and energy of the originating proto-jets.
Pruning does not require the reconstruction of sub-jets. The procedure can be summarized as follows:

- The C/A or k_T algorithm is run on the original jet constituents.

- At each recombination step of constituents j_1 and j_2 (where the two are p_T ordered), either $p_T^{j_2}/p_T^{j_1+j_2} > z_{cut}$ or $\Delta R_{j_1,j_2} < R_{cut} \times (2m^{jet}/p_T^{jet})$ must be satisfied. Therefore, the two parameters of the algorithm to be tuned in order to optimize the sensitivity in analyses involving boosted objects are z_{cut} and ΔR_{cut}.
- j_1 and j_2 are merged if one or both of the above criteria are met, otherwise the j_2 is discarded and the algorithm continues.

Figure 5.6 illustrates the pruning procedure.

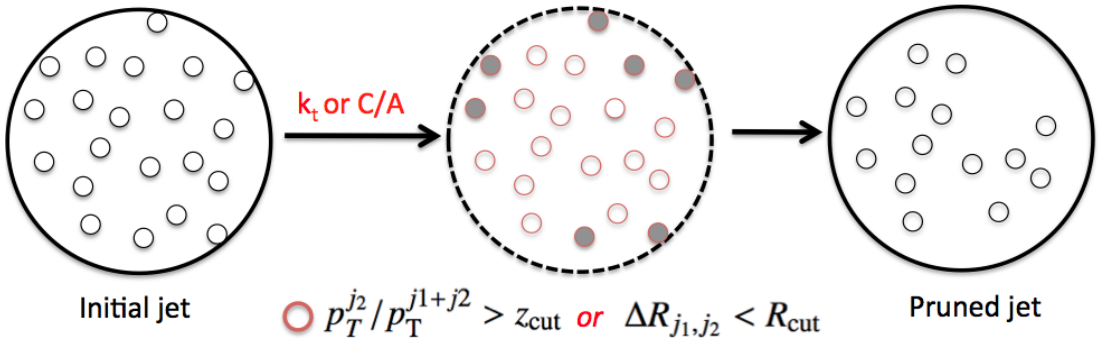


Figure 5.6: Diagram illustrating the jet pruning procedure [45].

5.2.2 Jet properties and substructure observables

In order to discriminate between jets from boson decay and QCD jets a number of jet moments, functions of the jet constituent four momenta, exists that parameterizes the information contained in the full set of jet constituents. For the studies presented here, a subset of four jet substructure variables is considered.

Jet mass

The jet mass is defined as the mass calculated from the four-momentum sum of all jet constituents. Depending on the input to the jet algorithm (see sec. 4.1.1), the constituents may be considered as either massive or massless four-momenta.
It is one obvious discriminant, because jets from W and Z bosons are expected to have mass peaks near 80 and 91 GeV, respectively, while those from QCD sources are expected to have a smoother distribution.

Mass-drop: μ_{12}

The mass-drop fraction [57] μ_{12} is also defined at the last step of recombination where two *proto-jets* are combined to form one jet. μ_{12} is the fraction of mass carried by the most massive proto-jet:

$$\mu_{12} = \frac{max(m_1, m_2)}{m_0} \tag{5.5}$$

where m_1 and m_2 are the masses of the two proto-jets and m_0 is the mass of the final clustering step. In the case of boson jets, where a two-prong structure is expected, the mass-drop tends to have small values. On the other hand, for QCD jets the mass of the final jets is expected to be very close to the mass of the leading proto-jet and then the mass-drop results in a value of $\mu_{12} \approx 1$.

Momentum balance: y_f

Another powerful discriminant is the momentum balance y_{f} [57] defined as

$$y_{\mathrm{f}} = \frac{\min(p_{\mathrm{T1}}^2, p_{\mathrm{T2}}^2)}{m_0^2} \Delta R_{12}^2 \approx \frac{\min(p_{\mathrm{T1}}, p_{\mathrm{T2}})}{\max(p_{\mathrm{T1}}, p_{\mathrm{T2}})} \tag{5.6}$$

where p_{T1} and p_{T2} respectively denote the transverse momenta of the two subjets, ΔR_{12} is the distance between them, and m_0 is the mass of the parent jet. This is approximately the ratio of the lower and higher p_T values for the two subjets. It is expected to peak sharply near zero for the background coming from gluon radiation, while the signal distribution is flatter, extending up to one for longitudinal bosons, i.e. for which quarks are emitted perpendicular to the boson's direction of motion in the rest frame. To emphasize the difference in shapes between the two distributions, $\sqrt{y_{\mathrm{f}}}$ will often be used.

N-subjettiness

The N-subjettiness variables τ_N [59] are observables related to the subjet multiplicity. In particular, they describe to what degree the substructure of a given jet agrees with N or fewer subjets. The τ_N variable is calculated by clustering the constituents of the jet with the k_T algorithm and requiring that exactly N subjets be found. This is done using the exclusive version of the k_T algorithm [60] and it is based on reconstructing clusters of particles in the jet using all of the jet constituents. These N final subjets define axes within the jet. The variables τ_N are then defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{Tk} \times min(\delta R_{1k}, \delta R_{2k}, ..., \delta R_{Nk}), \ with\ d_0 \equiv \sum_k p_{Tk} \times R \tag{5.7}$$

where R is the jet radius parameter in the jet algorithm, p_{Tk} is the p_T of constituent k and δR_{ik} is the distance from subjet i to constituent k. According to this definition, τ_N defines how well jets can be described as containing N or less k_T subjets. This is done by taking the hypothesis of N sub-jets, and quantifying the alignment of the jet constituents with the axes of such sub-jets. The most powerful variables in discriminating the number of hard sub-jets are not the τ_N, but the ratios τ_N/τ_{N-1}. For example, the ratios τ_2/τ_1 can be used to discriminate between jets from parton shower of initial-state gluons or light-quarks, and jets from two hadronic prong decays (from vector or scalar bosons). This ratio is usually referred to as τ_{21}. The value $\tau_{21} \simeq 1$ corresponds to the case where a jet contains only one sub-jet, whereas a lower value implies a jet better described by two subjets.

5.3 A modified BDRS algorithm for the high-p_T boson tagging: the BDRS-A

The BDRS split-filtering procedure has been studied in the energy regime interesting for the associated production of a Higgs boson with a W or Z boson with the subsequent decay $H \to b\bar{b}$. Searching for resonances with masses larger than 1 TeV require a different tuning of the split-filtering algorithm in order to have excellent grooming and tagging performance in the very high-p_T regime.

BDRS-A algorithm is a modified version of the canonical BDRS and it was studied to a greater extent in [61]. It follows the same procedure as the original BDRS, initially finding jets with the C/A algorithm with a distance measure of $R = 1.2$. However, the following modifications have been applied to the splitting and filtering processes:

- The mass drop is required to be $\mu_{filt} = 1$, because at very high momentum, the momentum-balance requirement at the splitting stage is enough to reject all the jets from QCD.
- The momentum balance cut is loosened to the value of $\sqrt{y_f} > 0.2$. This ensures that a larger fraction of jets from boosted-boson decay at high values of $\sqrt{y_f}$ pass the requirement. Furthermore, QCD jets tend to accumulate along the boundary of this requirement and so moving it to lower values generates a larger separation between signal and background, which can be used to make a post-selection requirement.
- The subjet radius used in the filtering stage is fixed to be $R = 0.3$ to ensure that very high-p_T, collimated decays will not lead to reclustering with a distance measure finer than the granularity of the calorimeter. This guarantees good mass resolution performance.

The parameters used for the BDRS-A algorithm are listed in Table 5.1.

Table 5.1: Parameters for the BDRS-A filter.

Name	Value
$\sqrt{y_{\text{fmin}}}$	0.20
μ_{fmax}	1.00
R_{r}	0.30
n_{r}	3

Although the jet constituents are topi-clusters locally calibrated (see Sec. 4.1.3), a Monte Carlo based particle-level calibration has been derived and applied to the jets used in this thesis to restore the hadronic scale. Energy, mass and η are calibrated according to the calibration procedure documented in detail in Appendix A.

5.3.1 Angular scale

The particular choice of the BDRS-A filtering parameters allows to discover during the grooming the angular scale of the jets (the angle between the two subjets under the assumption that the

jet has a two-prong structure) without pre-defining this parameter in the algorithm. In the case of jets produced by a boson hadronic decay, $V \to qq$, the scale is set by ΔR_{qq}, the distance in the $y - \phi$ plane between the two quarks, dependent on the mass and the momentum of the decaying particle. For the BDRS-A algorithm, the measured value of the angular scale ΔR_f has been found to be very close to the truth angular scale ΔR_{qq}. Loosening the parameters of the filtering tends to produce higher values of ΔR_f, because of wide-angle gluon radiation, contribution from pile-up, and underlying event. On the other hand, if the parameters are too tight, ΔR_f tends to have smaller values and the filtering procedure would continue past the correct scale and discard one of the subjets of interest. Figure 5.7 shows the difference between

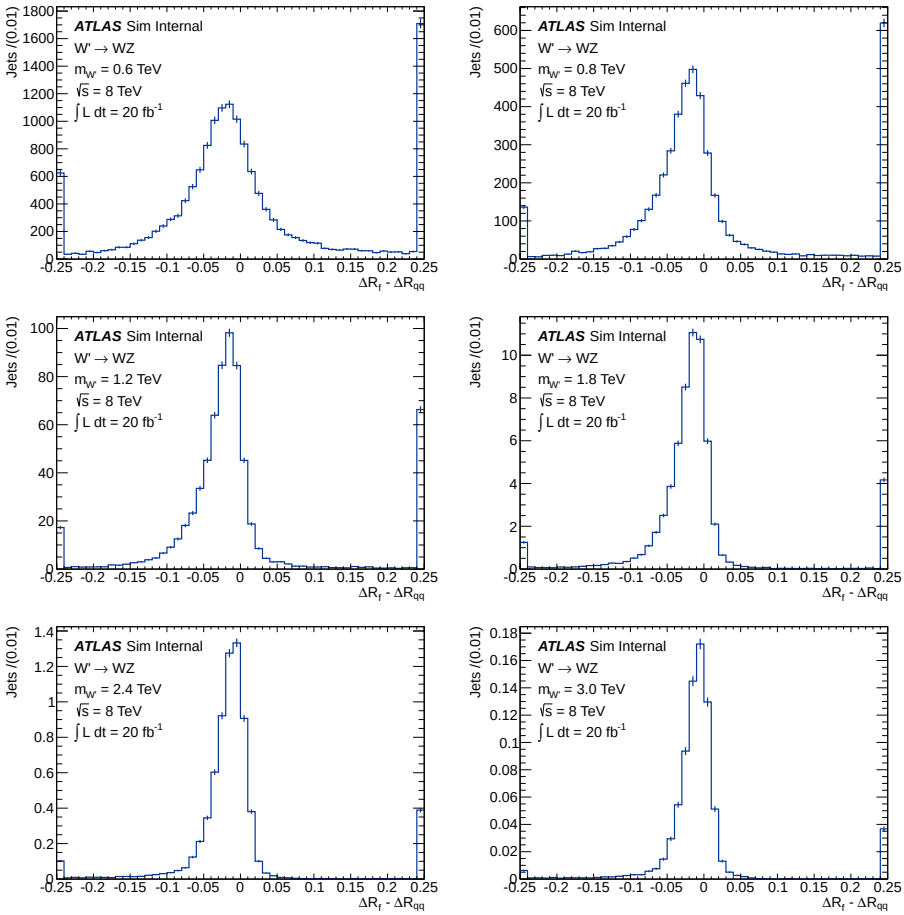


Figure 5.7: Difference between reconstructed subjet opening angle and boson diquark opening angle distributions for the $W' \to WZ$ samples. The first bin includes underflows and the last bin includes overflows.

ΔR_f and the true diquark angle ΔR_{qq}. Table 5.2 gives the fractions of jets for which this difference is smaller than 0.1 and 0.2. For high resonance masses, $m_{W'} > 1$ TeV, the filter finds the right stopping point about 90% of the time. The filter stops early ($\Delta R_f - \Delta R_{qq} \gg 0$) more often than continuing past the correct point ($\Delta R_{qq} - \Delta R_f \gg 0$), and so a larger value of y_{fmin} might be justified.

These features of the BRDS-A split-filtering have a big impact on the implementation of a boson tagging procedure for the heavy diboson resonances search. These aspects will be discussed in the next chapter.

Table 5.2: Fraction of reconstructed jets matched to bosons for generator-level cuts and with $|\Delta R_f - \Delta R_{qq}| < \Delta_{\max}$, i.e. corresponding to the distributions in Fig. 5.7.

$m_{W'}$	$\Delta_{\max}$	
[TeV]	0.10	0.20
0.6	0.70	0.82
0.8	0.75	0.83
1.2	0.82	0.87
1.8	0.88	0.91
2.4	0.90	0.93
3.0	0.92	0.94

Chapter 6

Search for heavy diboson resonances decaying in di-jet final state

As already mentioned in the Chapter 1, many extensions of the SM, such as Randall-Sundrum models with warped extra dimensions, Technicolor, and Grand Unified Theories are predicted to manifest themselves as heavy resonances decaying to vector boson pairs. The opportunity of searching for fully hadronic signatures, if in control of the large QCD background, can enhance the sensitivity of these searches in the very high energy regime. The analysis presented is focused on searching for narrow diboson resonances that should appear as a narrow peak in the dijet invariant mass spectrum. The dominant background for this search is due to dijet events from QCD processes. The application of jet substructure for jet grooming and boson tagging on the two highest p_T identified jets in the events is able to significantly suppress the large QCD dijet background and to increase the sensitivity of this search. In this chapter, a detailed description of the search for fully hadronic diboson resonances [62] is provided. In particular, the chapter is organized as follows:

- collision data and data-quality requirements;
- simulated data samples;
- boson-jet-identification implementation and optimization;
- full event selection definition;
- data-driven background parametrization;
- systematic uncertainties.

The results will be discussed in Chapter 7 and 8.

6.1 Collision data

This analysis has been performed with 8 TeV $proton - proton$ collisions recorded by the ATLAS detector in 2012. Events have been required to be selected by a large-radius jet trigger,

EF_j360_a10tcem, which is intended to keep all events with at least one jet above the p_{T} threshold of 360 GeV, reconstructed by clustering topo-clusters at EM scale with the anti-k_t $R = 1.0$ algorithm. Before an event is considered for the analysis, it must satisfy the Good Runs List (GRL) requirements, which ensure that all sub-detectors operated correctly during data-taking. Further data quality requirements are applied beyond the GRL; they are discussed in the following section.

6.1.1 Data quality requirements beyond the GRL

After passing the trigger and the GRL selection, events are excluded according to additional data quality guidelines.*Incomplete events*, events with noise bursts in the LAr calorimeters, or events with data corruption in either the Tile or LAr calorimeters, must be rejected.

During the 2012 data-taking, a data acquisition (DAQ) system restart, usually referred as "TTC restart", was developed to recover possible detector busy conditions without restarting the full ATLAS DAQ system. This situation may happen when a particular sub-detector is causing issues to the data taking operations and it needs to be restarted. In the luminosity block after a TTC restart, there may be events where some detector information is missing: they are usually referred as *incomplete events*. These events have been checked and removed from the analysis.

Events with noise bursts in the LAr calorimeter, as well as events with data corruption in either the tile or LAr calorimeters are automatically flagged during data taking and are rejected.

The tile calorimeter may have module trips (mainly due to problem in the power-supply system), which are generally corrected for and thus do not affect the data for analyses. However, in some 2012 data taking periods (G to J), there were occasional trips in one module which were not identified and masked. The events affected by this problem have been flagged as bad and removed from the dataset.

Table 6.1 gives the numbers of excluded and retained events by run period. Additional checks were also made for duplicate events and for events that may be copied twice in the datasets. None were found.

6.1.2 Integrated luminosity

The total integrated luminosity achieved selecting events with EF_j360_a10tcem trigger and GRL, was found to be 20.3 fb^{-1}, with a $\pm 2.8\%$ relative uncertainty. The integrated luminosity calculation has been derived following a methodology described in Ref. [63]. It excludes events marked with the errors in the LAr calorimeter, while events with errors in the Tile are included in the luminosity calculation because these can be corrected at the reconstruction level. The integrated luminosity obtained per data-taking period is shown in Tab. 6.12.

6.2 Simulated data samples

Several Monte Carlo simulated samples have been used in order to study the signal and background topologies, and to optimise the sensitivity of this search to diboson resonances.

In particular, the leading-order Monte Carlo generator PYTHIA 8.170 [41] has been used to

Table 6.1: Counts of retained and excluded events for the data quality checks by run period. The numbers of processed events and the numbers passing the GRL are also shown. Events may be excluded for multiple reasons. The retained column refers to all events passing the data quality selection in the skimmed dataset we use. The"Trigger 360 + DQ" column refers to all events accepted by the EF_j360_a10tcem trigger, passing the GRL, and data quality selection.

Run Period	*Raw*	*GRL*	*Trigger 360 + GRL*	*Incomplete*	*LAr error*	*Tile error*	*Tile trip*	*Retained all triggers*	*Trigger 360 + DQ*	*Efficiency [%]*
A	4434830	4149358	338797	0	43012	0	0	4144404	336956	99.5
B	8216277	7924172	2211847	0	51446	61096	0	7881514	2181444	98.6
C	2313602	2244672	616537	0	6029	0	0	2238757	613349	99.5
D	5318256	5195697	1457897	0	26031	2	0	5174725	1446480	99.2
E	4113414	3935790	1144622	0	33802	1	0	3923078	1137624	99.4
G	2047083	1996423	574522	0	4089	0	88	1992395	572265	99.6
H	2260633	2155846	649266	0	6569	6293	103	2149820	645818	99.5
I	1660485	1535970	464488	0	72002	0	51	1528992	460677	99.2
J	4193485	3995555	1157365	0	17460	0	105	3979106	1148831	99.3
L	1377921	1298283	375625	0	4868	0	0	1293784	373601	99.5
All	35935986	34431766	8990966					34306575	8917045	99.2

Table 6.2: Integrated Luminosity per run period.

Run Period	$\int \mathcal{L}dt$ [fb^{-1}]	Events / fb^{-1} 10^5
A	0.8	4.21
B	5.0	4.36
C	1.4	4.38
D	3.3	4.38
E	2.6	4.38
G	1.2	4.77
H	1.5	4.31
I	1.0	4.61
J	2.6	4.42
L	0.9	4.15
All	20.3	4.39

model $W' \to WZ$ events. The W' boson samples are generated assuming EGM couplings [19] for the W'. It is required to decay to a W and a Z boson, which are both forced to decay hadronically. The samples generated for this analysis use MSTW2008 [64] PDFs, with parton shower parameters tuned to ATLAS underlying-event data [65]. An additional set of W' samples generated with PYTHIA for the hard scattering interaction and HERWIG++ [66] for parton showering and hadronisation is used to assess systematic uncertainties on the signal efficiency due to uncertainties on the parton shower and hadronisation model. The partial width of $W' \to WZ$ decays thus scales linearly with $m_{W'}$, leading to a narrow width over the entire accessible mass range. Because of the anti-quark parton distribution functions involved in the production, a significant part of the W' cross section for large W' masses is due to off-shell interactions which produce a low-mass tail in the W' mass spectrum. The relative size of the low-mass tail increases with the W' mass: the fraction of events with a diboson mass below 20% of the pole mass of the W' increases from 10% for $m_{W'} = 1.3$ TeV to 22% for $m_{W'} = 2.0$ TeV and to 65% for $m_{W'} = 3.0$ TeV. As described in the first chapter, an extended RS model with a warped extra dimension is used for the excited graviton benchmarks [67]. The RS excited graviton samples are generated with CalcHEP 3.4 [68] setting the dimensionless coupling constant $k/\bar{M}_{\rm Pl} = 1$, where the k parameter is the curvature of the warped extra dimension and $\bar{M}_{\rm Pl}$ is the reduced Planck mass. The graviton resonance is decayed to WW or ZZ, and the resulting W or Z bosons are forced to decay hadronically. The cross section times branching ratio as well as the resonance width are provided by CalcHEP for the RS. Events are generated using CTEQ6L1 [69] PDFs, and use PYTHIA 8 for the parton shower and hadronisation. The cross section times branching ratio as well as the resonance width calculated by CalcHEP for the EGM W' and RS model exited gravitons are listed in Table 6.3 for 5 mass values.

To characterise the expected dijet invariant mass spectrum in the mass range spanning from 1.3 to 3.0 TeV, simulated QCD dijet events, diboson events, and single W or Z bosons produced with jets are used. Contributions from SM diboson events are expected to account for approximately 6% of the selected sample, and single boson production is expected to contribute less than 2%. Contributions from $t\bar{t}$ production, studied using MC@NLO [70] and HERWIG++ [66]

showering, were found to be negligible and are not considered further. Both QCD dijet events and W/Z+jets samples are produced with PYTHIA 8 using CT10 [71] and CTEQ6L1 PDFs, respectively. Diboson events are produced at the generator level with POWHEG [44], using PYTHIA for the soft parton shower.

For all the generated samples, the contribution from additional minimum-bias pp interactions is accounted for by overlaying additional minimum-bias events generated with PYTHIA 8, matching the distribution of the number of interactions per bunch crossing observed in collision data. In the 8 TeV data taking, the average number of pp interactions per bunch crossing was approximately 20.

Table 6.3: The resonance width (Γ) and the product of cross sections and branching ratios ($\sigma \times$BR) to four-quark final states used in modelling $W' \to WZ$, $G_{RS} \to WW$, and $G_{RS} \to ZZ$, for several values of resonance pole masses (m). The fraction of events in which the invariant mass of the W' or G_{RS} decay products lies within 10% of the nominal resonance mass ($f_{10\%}$) is also displayed.

			$W' \to WZ$		$G_{RS} \to WW$		$G_{RS} \to ZZ$	
m [TeV]	$\Gamma_{W'}$ [GeV]	$\Gamma_{G_{RS}}$ [GeV]	$\sigma \times$BR [fb]	$f_{10\%}$	$\sigma \times$BR [fb]	$f_{10\%}$	$\sigma \times$BR [fb]	$f_{10\%}$
1.3	47	76	19.1	0.83	0.73	0.85	0.37	0.84
1.6	58	96	6.04	0.79	0.14	0.83	0.071	0.84
2.0	72	123	1.50	0.72	0.022	0.83	0.010	0.82
2.5	91	155	0.31	0.54	0.0025	0.78	0.0011	0.78
3.0	109	187	0.088	0.31	0.00034	0.72	0.00017	0.71

6.3 Boson Identification

The jet filtering algorithm introduced in Section 5.3 has limited discriminating power for jets from the decay of boosted bosons and QCD jets. Several physics observables have been studied in order to enhance this separation [72] distinguishing between *signal* and *background* jets. This has been achieved applying to each jet additional cuts on three discriminating variables:

Filtered jet mass: the mass calculated from the filtered jets is expected to be small for QCD jets but should reflect the boson mass for jets from boosted hadronic boson decays.

Subjet momentum balance $\sqrt{y_f}$: by requiring a relatively symmetric momentum sharing between the two C/A subjets that form the jet, the dominant component of the QCD background involving soft gluon radiation can be rejected[1].

Number of tracks associated to the unfiltered jet: because energetic gluon emission tends to create jets with large hadron multiplicity, this cut can reject background events with hard gluons that may pass the $\sqrt{y_f}$ cut.

[1] Unlike the filtering cut $\sqrt{y_{\text{fmin}}}$, this cut does not affect which C/A subjets are identified as the boson daughter candidates through the iterative procedure.

In this section, the most important aspects of the boson identification cuts will be summarized, beginning with the baseline event-selection requirements imposed on all signal and background MC samples used in this procedure. A complete description of the optimization of the boson tagging selection can be found in Ref. [72].

6.3.1 Event pre-selection criteria for boson-identification studies

In order to optimize the boson identification cuts, the following criteria have been used to select events in the most relevant region of the phase-space for each of the signal models considered for this search. This is needed since the p_{T} spectra of the jets are rapidly falling and the optimization procedure may be dominated by the low energy region if all the event are taken inclusively. The criteria require that both jets pass the BDRS-A filter (described in Sec. 5.3), and additional constraints on the dijet invariant mass and rapidity difference, as listed in Table 6.4 . The fraction of signal events that survives these requirements is presented in Tables 6.5, 6.6, and 6.7.
The dijet rapidity difference requirement, $|\Delta y_{12}| < 1.2$, favors central events produced by an s-channel mechanism. This cut is included in the full event selection of this search and it will be discussed in detail more in the following sections.
The dijet mass requirement selects events for which the invariant mass of the two leading jets system is included in a window of 10 % around the generated W' mass. This allows to study the boson tagging performance in the energy regime of each generated signal sample, avoiding contributions from off-shell production tails, particularly pronounced for test masses above 2.5 TeV and due to the LHC center-of-mass energy.
After applying these cuts, the signal shows narrow peaks in the energy and p_{T} distributions. The mean and RMS for those distributions are presented in Tables 6.8 and 6.9. In order to have a similar kinematic regime for QCD background jets, and to study the performance of the cuts on single jet samples rather than dijet events, the jets are selected into a window of 2 standard deviations about the mean of the jet p_{T} distribution for a given signal hypothesis.

Table 6.4: Selection applied to reconstructed jets for boson identification studies. Here R stands for the heavy resonance that decays to bosons.

Both leading jets pass jet filter
$\frac{\lvert m_{jj} - m_R \rvert}{m_R} < 10\%$
$\lvert \Delta y_{12} \rvert < 1.2$

6.3.2 Optimization procedure

The optimization procedure uses Monte Carlo samples to identify optimal boson tagging cuts for the jet mass and subjet momentum balance variables. The power of each discriminant was assessed by evaluating $S/\sqrt{B}$ as a function of a threshold on that variable. S is the number of signal and B the number of background events that survives a given cut on the discriminant.

Table 6.5: Fraction of $W' \to WZ$ signal events used in the boson identification studies: ε_m is the fraction falling in the ±10% dijet mass window, $\varepsilon_{\Delta y}$ is the fraction of those surviving the Δy_{12} cut, and $\varepsilon_{\text{disc}}$ is the total fraction accepted.

$m_{W'}$ [TeV]	ε_m	$\varepsilon_{\Delta y}$	$\varepsilon_{\text{disc}}$
1.2	0.72	0.70	0.52
1.8	0.66	0.71	0.47
2.4	0.50	0.71	0.36
3.0	0.27	0.69	0.19

Table 6.6: Fraction of $G^* \to WW$ signal events used in the boson identification studies: ε_m is the fraction falling in the ±10% dijet mass window, $\varepsilon_{\Delta y}$ is the fraction of those surviving the Δy_{12} cut, and $\varepsilon_{\text{disc}}$ is the total fraction accepted.

m_{G^*} [TeV]	ε_m	$\varepsilon_{\Delta y}$	$\varepsilon_{\text{disc}}$
1.2	0.67	0.82	0.56
1.8	0.65	0.83	0.55
2.4	0.61	0.83	0.51
3.0	0.53	0.82	0.45

Table 6.7: Fraction of $G^* \to ZZ$ signal events used in the boson identification studies: ε_m is the fraction falling in the ±10% dijet mass window, $\varepsilon_{\Delta y}$ is the fraction of those surviving the Δy_{12} cut, and $\varepsilon_{\text{disc}}$ is the total fraction accepted.

m_{G^*} [TeV]	ε_m	$\varepsilon_{\Delta y}$	$\varepsilon_{\text{disc}}$
1.2	0.64	0.81	0.52
1.8	0.63	0.82	0.52
2.4	0.59	0.82	0.49
3.0	0.51	0.82	0.42

Table 6.8: Mean and RMS of the reconstructed jet energy distributions of leading and subleading jets in signal and Pythia multijet background events passing the event selection in Table 6.4.

	E [GeV]							
Mass	$W' \to WZ$		$G^* \to WW$		$G^* \to ZZ$		Pythia QCD	
[TeV]	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS
1.2	737	217	698	172	697	173	719	227
1.8	1015	218	985	176	984	181	989	225
2.4	1293	212	1269	182	1266	185	1257	228
3.0	1567	211	1553	188	1549	191	1528	226

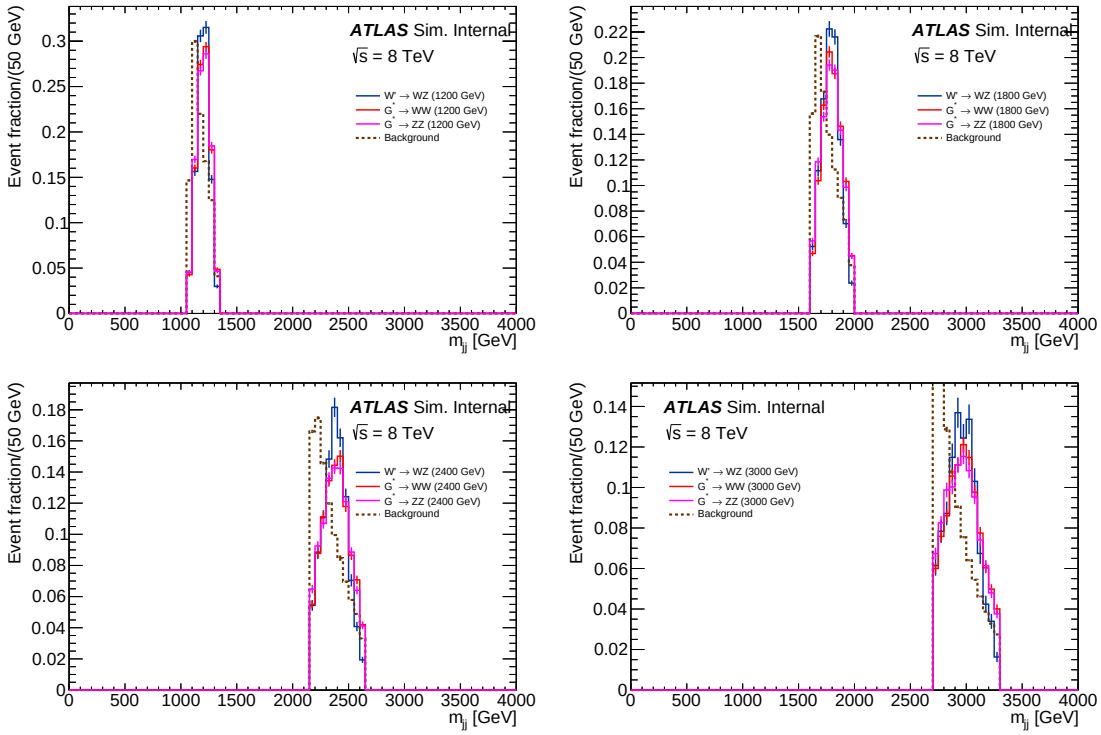


Figure 6.1: Reconstructed dijet mass for the benchmark signals and Pythia multijet background for the event selection in Table 6.4. All spectra are normalized to unity.

Table 6.9: Mean and RMS of the reconstructed p_{T} distributions of leading and subleading jets in signal and Pythia multijet background events passing the event selection in Table 6.4.

	p_{T} [GeV]							
Mass	$W' \to WZ$		$G^* \to WW$		$G^* \to ZZ$		Pythia QCD	
[TeV]	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS
1.2	563	59	573	65	571	67	534	61
1.8	845	90	860	88	857	92	799	92
2.4	1122	135	1144	110	1140	112	1060	120
3.0	1382	216	1422	131	1419	130	1319	144

6.3.3 Jet mass selection

The jet mass distributions from W and Z bosons are expected to peak near 80 and 91 GeV respectively, unlike jets from QCD processes where the mean of the squared mass distribution scales with $p_T^2 R^2$ [73]. Figure 6.2 shows the signal and background jet mass distributions obtained applying the baseline selection described in Section 6.3.1 for each of the W' mass hypotheses. The signal shows the expected peaks near the W and Z boson masses. Due to the poor mass resolution of the fully hadronic boson decay, the W and Z peaks for the W' signal samples are merged into a single peak. The background shows a broad peak just above 20% of the jet p_{T}, as well as some structure below 50 GeV mostly arising from jets with a very small

number of constituent clusters (i.e. two or three).
Efficiency and sensitivity curves have been obtained from these distributions for the W' signals, and are shown in Figs. 6.3 and 6.4. These are used to determine the mass window cut:

$$60 \leqslant m_{\text{jet}} < 110 \text{ GeV} \tag{6.1}$$

as the first selection requirement for boson jet identification. The jet mass selection defined above represents a looser selection. It has been extensively used to derive systematic uncertainties for jet kinematic and substructure properties, and to perform studies on the dijet background shape that require a larger acceptance for statistical precision.
In order to maximize sensitivity for the W' search, the jet mass selection described above has been tightened with a data driven approach, taking into account jet mass scale and jet mass resolution uncertainties for BDRS-A jets. Since the peaks of the signal jet mass distributions are different for the two bosons in a $W' \to WZ$ decay, the optimized mass cuts for the WZ final state are different for the two jets. In the optimized $W' \to WZ$ selection, the mass of one jet (the more massive of the two leading jets) must fall in a 26 GeV symmetric window about the Z reconstructed mass, and the mass of the other jet must fall in a 26 GeV symmetric window about the W boson reconstructed mass [74]:

$$\begin{aligned} &69.4 \leqslant m_{\text{jet 1}} < 95.4 \text{ GeV} \quad \text{and} \quad 79.8 \leqslant m_{\text{jet 2}} < 105.8 \text{ GeV}, \text{ or} \\ &69.4 \leqslant m_{\text{jet 2}} < 95.4 \text{ GeV} \quad \text{and} \quad 79.8 \leqslant m_{\text{jet 1}} < 105.8 \text{ GeV}. \end{aligned} \tag{6.2}$$

For the graviton signals, both mass cuts use the same 26 GeV window around either the W or Z boson reconstructed mass.

6.3.4 Sub-jet momentum balance $\sqrt{y_{\text{f}}}$

The other discriminant considered is the sub-jet momentum balance, y_{f}, defined in Equation 5.6. It corresponds approximately to the ratio between the lower and higher p_{T} values for the two subjets. Therefore it is expected to peak sharply near zero for the background coming from gluon radiation, while the signal is wider and extends up to one for longitudinal bosons, for which the rest-frame quarks are emitted perpendicular to the boson direction of motion. The difference in shape between the two distributions is larger when $\sqrt{y_{\text{f}}}$ is used.

Figure 6.5 shows the signal and background distributions for this variable and Fig. 6.6 shows the corresponding efficiencies and sensitivity for the W' signal. The distributions are similar for the different signal mass points and a common requirement $\sqrt{y_{\text{f}}} \geqslant 0.45$ has been used for all the jets considered.

6.3.5 Track multiplicity

The radiation of a hard gluon can allow QCD jets to mimic a two-pronged structure and pass the tagging cuts described above. Gluon-induced jets contain on average a larger number of charged hadrons than quark-induced jets of corresponding energy, and the multiplicity of charged particles for both quarks and gluons increases with the energy scale [75]. Since the relevant scale for a boosted-boson decay is the boson mass, much lower than the jet's momentum which sets the scale for the QCD dijet background, further discrimination between boson jets and QCD background can be attained by cutting on the charged hadron multiplicity. The

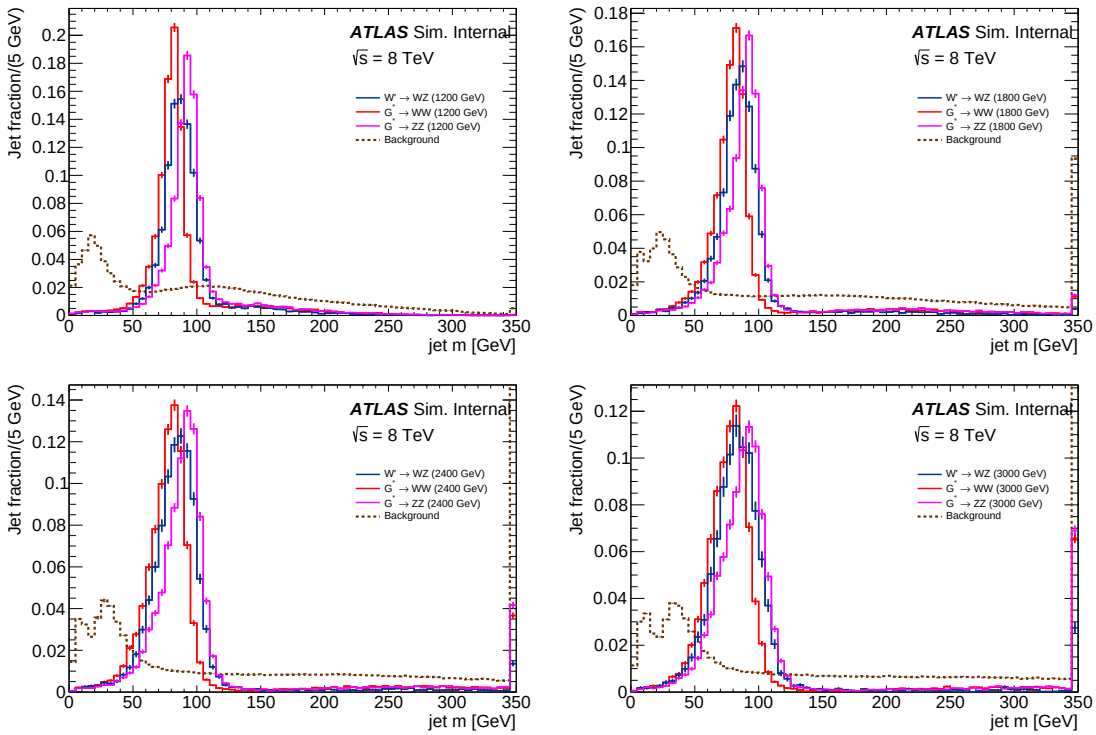


Figure 6.2: Jet mass for the benchmark signals and Pythia multijet background for the event selection in Table 6.4. All spectra are normalized to unity. Both leading and subleading jets are included.

charged hadron multiplicity has been estimated with the multiplicity of tracks reconstructed by the ID associated to the jet, n_{trk}.

Tracks with p_{T} greater than 500 MeV and $|\eta| < 2.5$, which satisfy standard impact parameter and quality cuts for tracks in jets and are associated with the event primary vertex, are associated to jets using the ghost-association procedure [76]. In the ghost-association procedure, additional constituents for the jet finding algorithm are formed from each track, using a negligible momentum magnitude and the direction of the track. To reduce the bias in the track multiplicity distribution due to the tracking detector acceptance ($|\eta_{\text{trk}}| < 2.5$) the jets themselves are required to satisfy $|\eta_{\text{jet}}| < 2.0$.

Figure 6.7 shows the distribution of the number of ghost-associated tracks for the C/A ungroomed jets, for the benchmark signals and background Monte Carlo events passing the selection requirements in Table 6.4 and the cuts introduced in the previous two sections ($60 \leqslant m < 110$ GeV and $\sqrt{y_{\text{f}}} \geqslant 0.45$). Both Pythia and Herwig Monte Carlo background samples are shown. A data-driven study has been used to determine the optimal threshold. In this study, the relative efficiency and background rejection of a n_{trk} cut are estimated by its effect on the boson mass peak in a sample of V+jets events. Efficiency correction scale factors for this cut applied to MC signal events are also determined using the same approach [74].

The efficiency measurements in the V+jets samples indicated that a cut of $n_{\text{trk}} < 30$ provides a sensitivity increase of 10-15%, so this threshold has been adopted as a boson identification requirement.

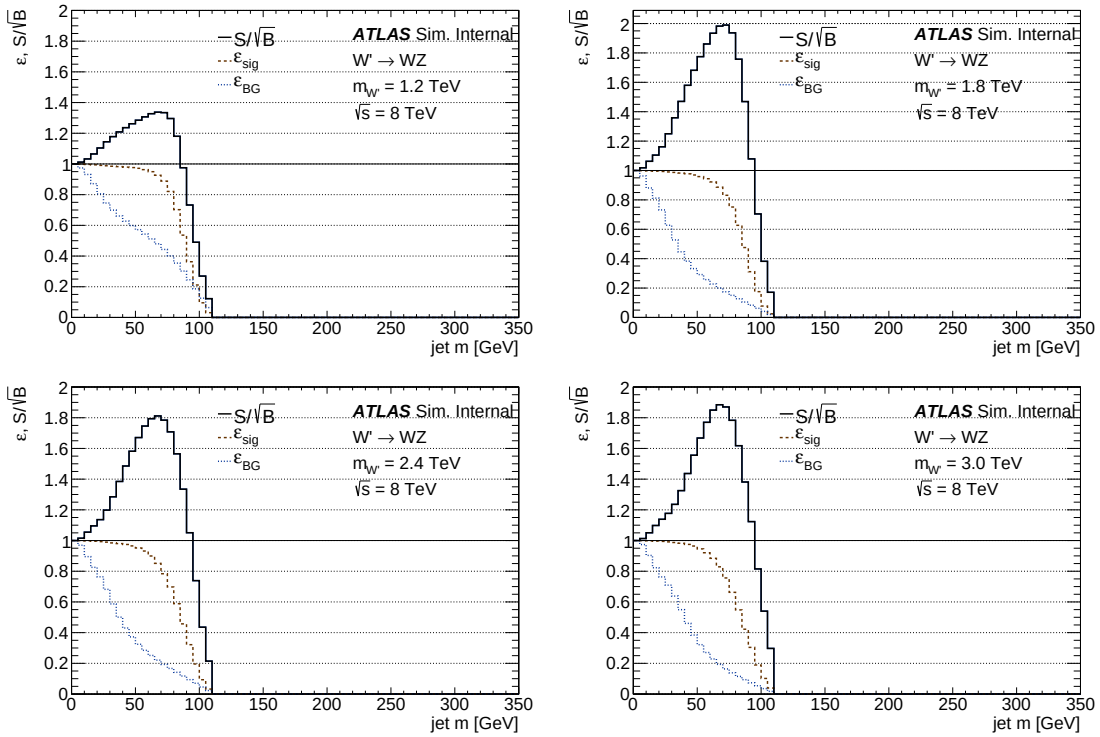


Figure 6.3: Efficiencies as a function of lower jet-mass threshold for the $W' \to WZ$ signal and Pythia multijet background using the event selection in Table 6.4 plus an upper mass threshold $m < 110$ GeV. Both leading and subleading jets are included. The $S/\sqrt{B}$ sensitivity based on these efficiency curves is also plotted as a function of the same variable.

6.3.6 Boson-tagging selection summary and expected performance

Jets are initially formed with the C/A algorithm with distance parameter $R = 1.2$. These jets are then filtered with the mass drop filter described in Sec. 5.3, with the parameters $\mu_{\rm f} = 1.0$, $\sqrt{y_{\rm f}} = 0.2$, $R_r = 0.3$ and $n_r = 3$. After the filtering, jets are selected by requiring the jet mass to fall within a window of full width 26 GeV around the reconstructed boson mass, the sub-jet momentum balance to pass the cut $\sqrt{y_{\rm f}} \geqslant 0.45$ and the number of tracks associated to the ungroomed jet to satisfy $n_{trk} < 30$. Jets passing these requirements are referred to as *tagged* in the following sections. As already mentioned, to enhance the statistical sample in deriving systematic uncertainties on kinematic and substructure observables studies and in studing the dijet background shape, a wider mass window of $60 < j_m \leqslant 110$ GeV has been used.

The expected tagging performance metrics per-jet are presented in Table 6.10. The first three columns give the expected cumulative signal selection efficiencies, and the last three columns give the expected cumulative background rejection. The expected efficiencies have been measured in a sample of events that were pre-selected by the cuts in Table 6.4, as well as requiring that $|\eta_1| < 2.0$ and $|\eta_2| < 2.0$. The efficiencies and rejections are the average of the values for the leading and sub-leading jets. Pythia Monte Carlo is used both for signal and background, and no scale factor is applied for the track multiplicity selection. The signal efficiencies are calculated on the graviton samples, selected within a window round the W or the Z mass as in the previous section. Consistent values have been found between W' and graviton samples.

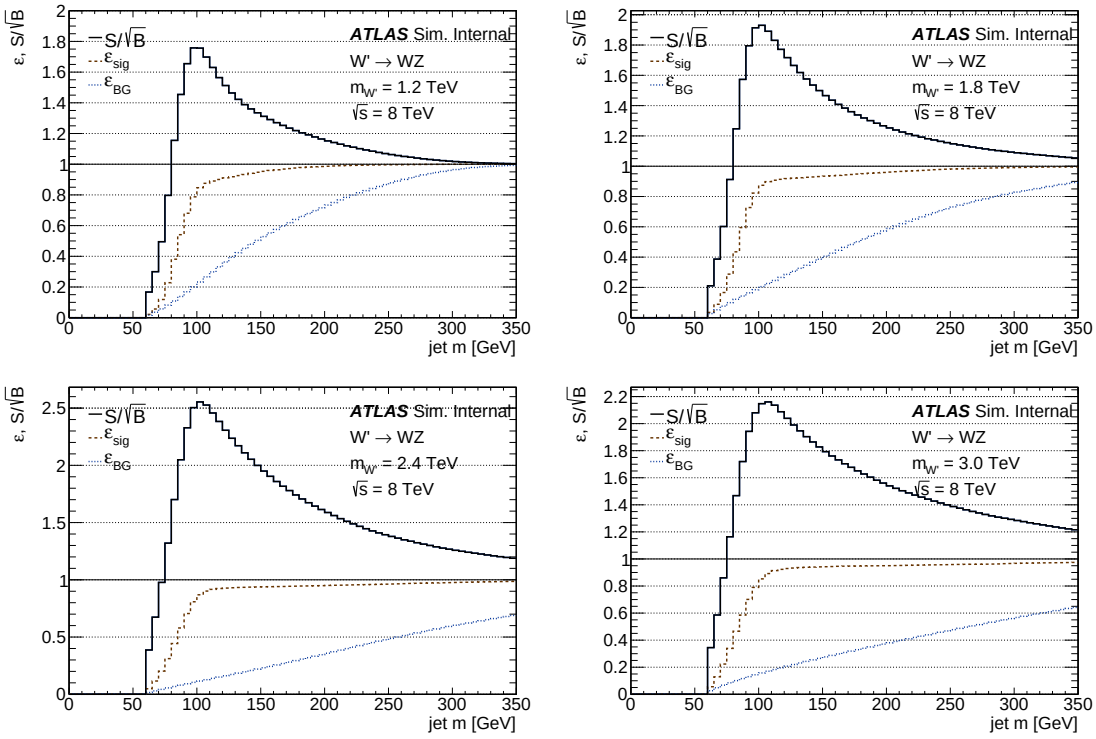


Figure 6.4: Efficiencies as a function of upper jet-mass threshold for the $W' \to WZ$ signal and Pythia multijet background using the event selection in Table 6.4 plus a lower mass threshold $m \geqslant 60$ GeV. Both leading and subleading jets are included. The $\sqrt{S}/B$ sensitivity based on these efficiency curves is also plotted as function of the same variable.

Table 6.10: Expected tagging performance per-jet in graviton samples (numbers are consistent for the W' samples). Efficiencies are given relative to a sample pre-selected by the criteria described in Section 6.3.6, and an additional requirement $|\eta_1| < 2.0$ and $|\eta_2| < 2.0$. The first three columns give the cumulative selection efficiencies per jet. The last three columns give the cumulative background rejections. The first rejections of the background correspond to Pythia and the second to Herwig.

m_{G^*} [TeV]	ε^S_m [%]	$\varepsilon^S_{m,\sqrt{y_f}}$ [%]	$\varepsilon^S_{m,\sqrt{y_f},\, n_{trk}}$ [%]	$1-\varepsilon^B_m$ [%]	$1-\varepsilon^B_{m,\sqrt{y_f}}$ [%]	$1-\varepsilon^B_{m,\sqrt{y_f},\, n_{trk}}$ [%]
1.2	67.5	49.5	44.0	89.6 / 94.5	97.5 / 97.3	98.8 / 98.1
1.8	61.5	45.3	40.0	94.0 / 93.8	98.0 / 97.9	99.0 / 98.6
2.4	56.8	40.4	36.0	95.0 / 94.8	98.4 / 98.3	99.2 / 98.8
3.0	50.8	38.6	33.9	95.5 / 95.2	98.6 / 98.4	99.4 / 99.0

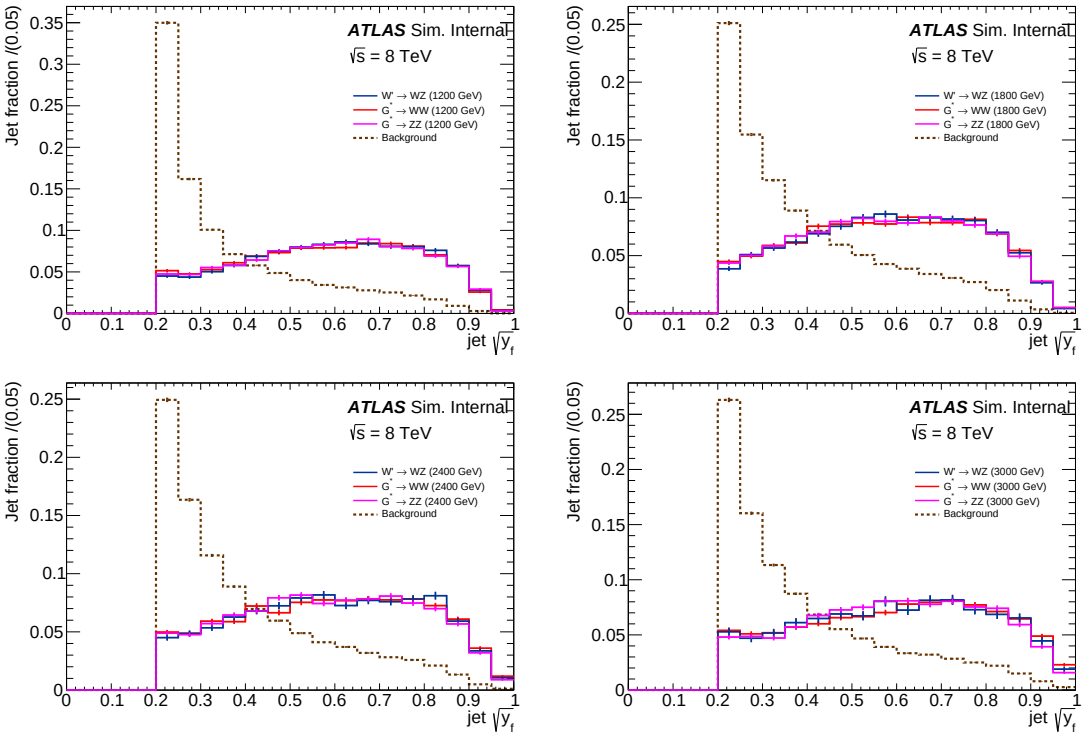


Figure 6.5: Momentum balance for the benchmark signals and Pythia multijet background for the event selection in Table 6.4 plus the requirement on the jet mass, $60 < j_m < 110$ GeV. Spectra are normalized to unity. Both leading and subleading jets are included.

6.4 Event selection

The final statistical analysis is performed searching for an excess in the dijet invariant mass spectrum of the two leading jets for selected events in data. The goal of the event selection is to maximize the sensitivity for the observation of a $W' \to WZ$ or other narrow diboson resonances. In order to accomplish this, the selection must reject most of the QCD background while maintaining high efficiency for signal events.

The following selection criteria have been defined, that will be described in the next sections:

- The trigger is the lowest unprescaled[2] large-radius jet trigger for 2012, EF_j360_a10tcem.
- The two C/A 1.2 leading jets in the event are required to pass the mass-drop filter described in Sec. 5.3.
- The mass of the dijet-system is required to be above 1.05 TeV, to avoid regions where the trigger is not fully efficient.
- The rapidity gap between the two leading jets is required to be $|\Delta y_{12}| < 1.2$ to reject QCD t-channel dijet production.

[2]During the data taking, the bandwidth of each trigger item is limited. If the rate of events selected exceeds this bandwidth limit, the trigger item is prescaled. An unprescaled trigger has prescale factor equal to 1, in order to be ready to select events at every bunch crossing.

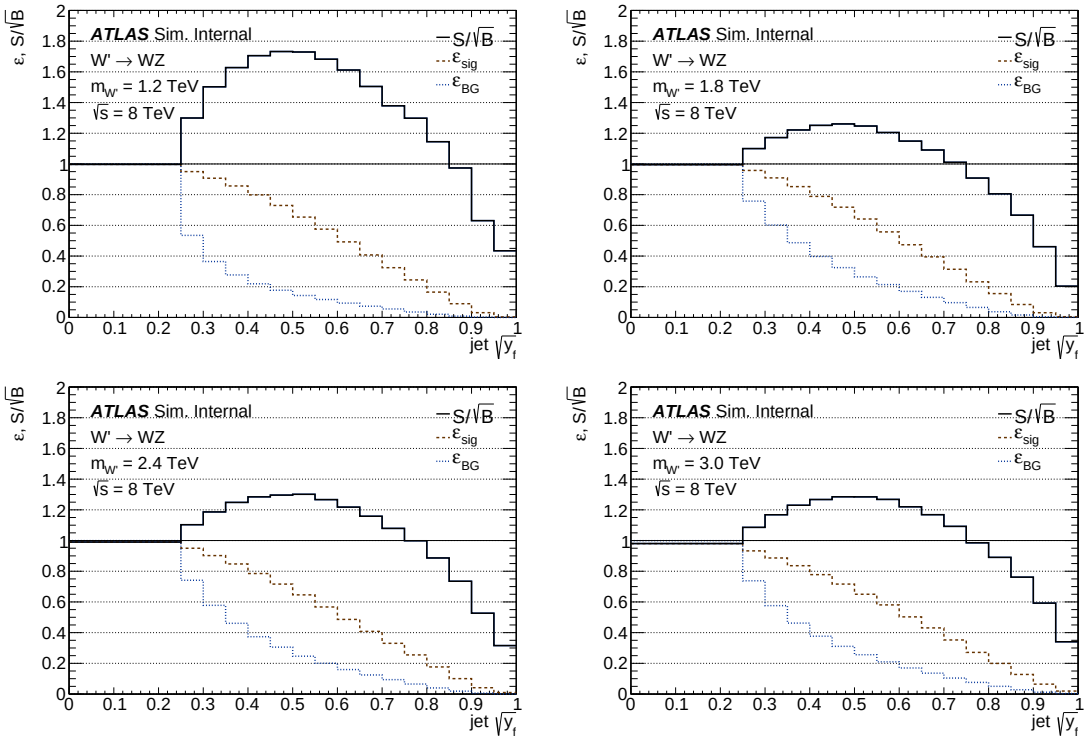


Figure 6.6: Efficiencies as a function of momentum-balance threshold for the $W' \to WZ$ signal and Pythia multijet background using the event selection in Table 6.4 plus the window on jet mass $60 \leqslant m < 110$ GeV. Both leading and subleading jets are included. The $\sqrt{S}/B$ sensitivity based on these efficiency curves is also plotted as function of the same variable.

- The η of the two leading jets is required to be within $|\eta| < 2.0$ to ensure good overlap with the inner detector so the ghost-associated tracks are usable as a discriminant.
- The p_{T} asymmetry, A, between the two leading jets, as defined in Equation 6.4 is required to be $A < 0.15$ to avoid jets at the tails of the p_{T} resolution and with issues in the recontruction.
- Events where one of the two leading jets has all three leading clusters in a *bad* [3] calorimeter region are rejected to avoid badly-calibrated jets.
- The two leading jets are required to pass the boson identification criteria described in the previous section.

6.4.1 Trigger selection and dijet mass threshold

As previously mentioned, events are required to be selected by the 360 GeV large-radius jet trigger, EF_j360_a10tcem, which is intended to keep all events with at least one jet above the p_{T} threshold of 360 GeV at EM scale. This trigger was unprescaled through all the 2012 data-taking. It uses as inputs for the anti-k_t algorithm with distance parameter $R = 1.0$, topo-clusters at the electromagnetic scale. The use of large-radius jet trigger is crucial to have similar

[3] In a masked calorimeter cell or in an adjacent cell.

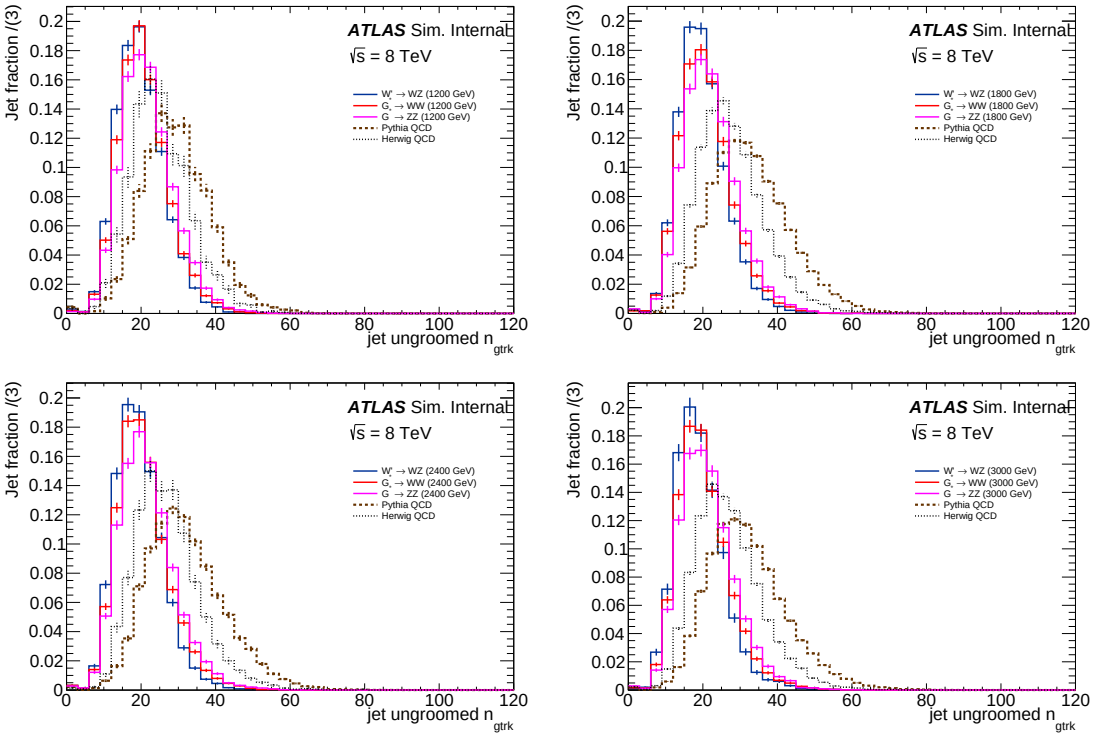


Figure 6.7: Track multiplicity for the benchmark signals and for the Pythia and Herwig multijet backgrounds for the event selection in Table 6.4 plus the window on jet mass $60 \leqslant m < 110$ GeV and momentum balance threshold $\sqrt{y_{\mathrm{f}}} \geqslant 0.45$. All spectra are normalized to unity. Both leading and subleading jets are included.

object definitions at the trigger level and in the analysis, in order to guarantee a high selection efficiency for jet masses up to 200 GeV and for dijet masses larger than 1 TeV. The use of a small-R jet trigger would add inefficiencies for objects with large jet masses and low p_{T}, because the jet algorithm would not be able to contain the full extent of the boson-decay products.
Possible differences in the trigger and offline jet algorithms, as well as differences in the calibration and detector condition databases used in jet reconstruction, may lead to a slow turn-on of the trigger efficiency as a function of jet p_{T} and therefore to distortion in the low dijet mass region. In order not to bias the trigger efficiency measurement, the turn-on region has been studied with respect to the leading jet p_{T} in data control samples selected from the muon triggers. In particular, events passing either the EF_mu25i or EF_mu36 triggers (events with at least one isolated muon with p_{T} above 25 GeV and with p_{T} above 36 GeV, respectively) have been used. This represents an unprescaled trigger selection that has negligible correlation with EF_j360_a10tcem, and therefore it allows the study of the plateau efficiency as well as the turn-on shape. The trigger has been tested among events with a muon corresponding to the trigger and a C/A R=1.2 jet with $p_T > 250$ GeV. Figure 6.8 shows the relative trigger efficiency measured in this sample as a function of the leading C/A R=1.2 jet's p_{T} before boson tagging, and in the subsample of events in which the leading jet satisfies boson tagging requirements. The efficiency for C/A R=1.2 jets is found to be greater than 50% for jets with ungroomed p_{T} above 461.6 ± 0.5 GeV, and shows a plateau at 99.6 ± 0.2 %. For tagged C/A R=1.2 jets, the efficiency is greater than 50% for ungroomed p_{T} above 468 ± 02 GeV and has a plateau at 99.8 ± 0.4 %.

The trigger efficiency curve as a function of the jet p_{T} is parametrized as:

$$\epsilon(p_{\mathrm{T}}) = \varepsilon \left(1 - \frac{1}{\sqrt{2\pi}} \int_{p_{\mathrm{T}}}^{\infty} \exp - \left[\frac{(x-T)}{\sqrt{2}\sigma} \right]^2 dx \right) \tag{6.3}$$

where ε is the efficiency-plateau value, T is the threshold and σ represents the threshold rise. The fitted parameter's uncertainties and their correlations are used to derive the trigger efficiency and its uncertainty as a function of the jet p_{T}, as summarised in Table 6.11.

6.4.2 Central events

The s-channel production mechanism of W' and G_{RS} resonances tends to create more central events than the t-channel mechanism, which is the dominant one for the QCD background. In order to distinguish these two categories of events, the difference in rapidity between the two leading jets is considered, usually referred to as Δy_{12}. The signal is expected to peak near $\Delta y_{12} = 0$, while the background peaks at much larger values of $|\Delta y_{12}|$, as shown in Figure 6.9 for reference masses $m_{W'} = 1.2, 1.8, 2.4$ and 3.0 TeV. The W' signal is multiplied by the indicated scale factor to show a number of events comparable to the background for visualisation purposes. Only events with their two leading jets having $|\eta| < 2.8$, passing the BDRS-A filter and with a

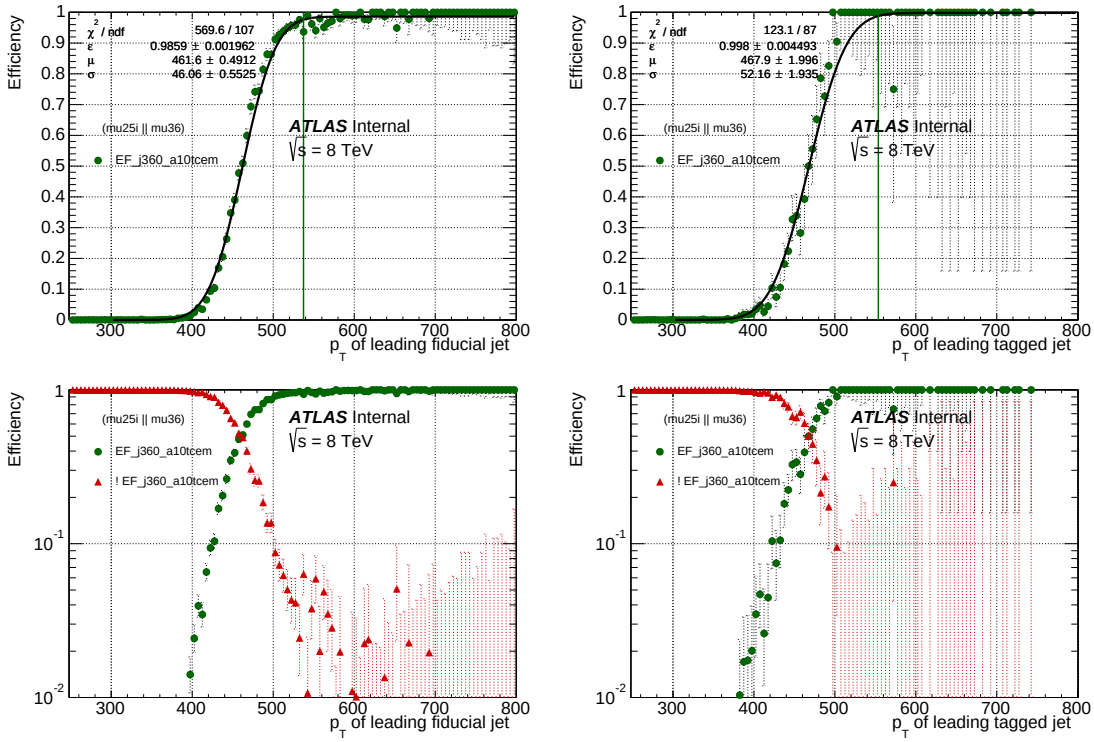


Figure 6.8: Relative trigger efficiency as a function of leading jet p_{T} before boson tagging (left) and after (right). The efficiency is shown in the upper plots in linear scale; efficiency and inefficiency are shown in the lower plots in log scale. The point at which the trigger reaches 99% of the plateau efficiency is indicated with a green line on the upper plots. The jet momenta used for these figures is the ungroomed jet momentum, and jets are not calibrated to the particle level.

Table 6.11: Trigger turn-on curve parameters, obtained by fitting the jet trigger efficiency distribution in muon-triggered events. The leading C/A 1.2 jet in the event is required to be within $|\eta| < 2.0$, to be filtered as described in 5.3 and then required to pass the tagging criteria $\sqrt{y_f} \geqslant 0.45$, and $60 < j_m < 110$. The jet momenta used for this table is the ungroomed jet momentum, and jets are not calibrated to the particle level.

Parameter	Value
Efficiency	0.998 ± 0.004
Threshold	468 ± 2 GeV
Rise	52 ± 2 GeV

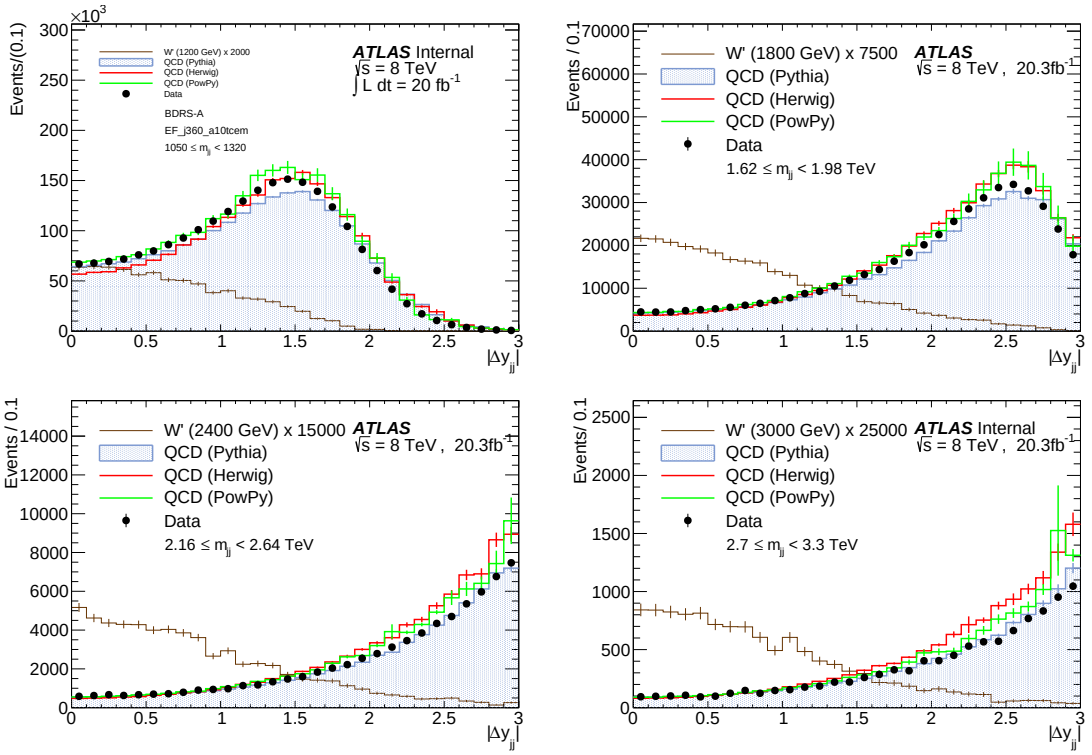


Figure 6.9: Rapidity gap between the two leading BDRS-A jets in QCD MC and W' MC after the particle-level calibration has been applied. The QCD MC is normalized to the integrated luminosity of the data and has been reweighted to the p_{T} of the leading jet. The signal MC is normalized to the integrated luminosity times a scale factor defined in the legend. Events in the histograms have been selected after all pre-selection requirements described in the previous sections, the EF_j360_a10tcem trigger is applied. Each panel displays a different dijet mass range corresponding to the relevant phase space for the benchmark signal indicated.

dijet mass, m_{12}, such that $|m_{12} - m_{W'}|/m_{W'} < 0.1$ are included. Based on these figures, and considering the ratio of signal to background, the event selection requirement $|\Delta y_{12}| < 1.2$ has been defined and used in the analysis.

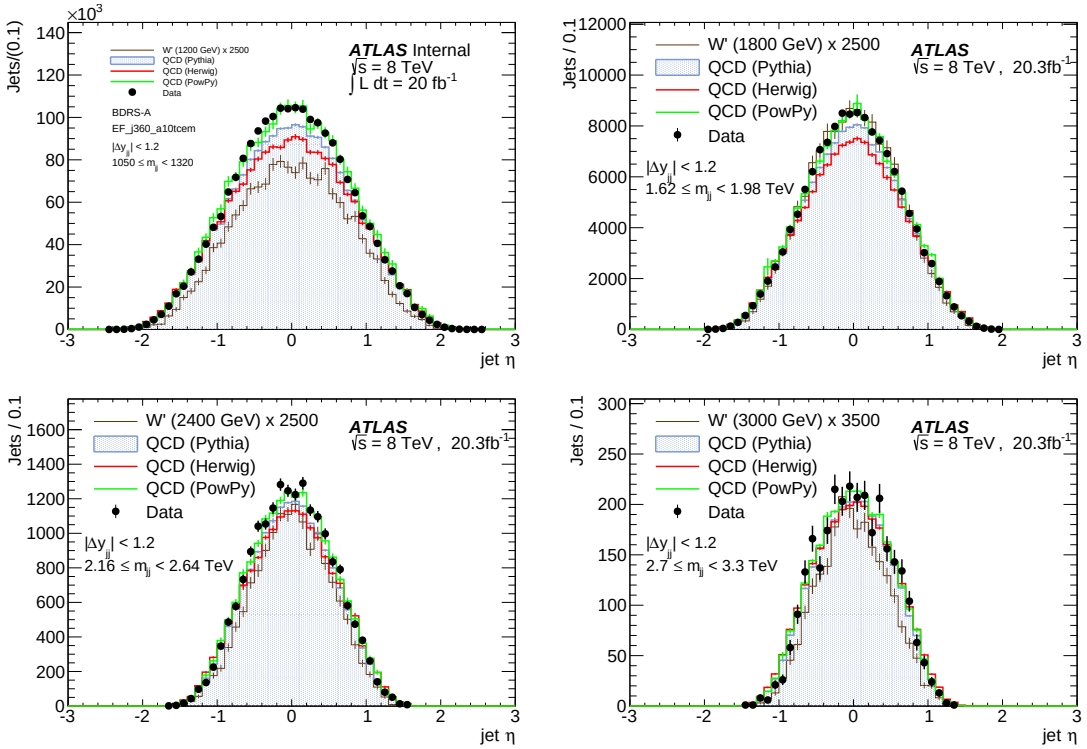


Figure 6.10: Pseudo-rapidity distribution for leading and sub-leading BDRS-A jets QCD MC and W' MC after the particle-level calibration has been applied. The QCD MC is normalized to the integrated luminosity of the data and has been reweighted according to the p_{T} of the leading jet. The signal MC is normalized according to the integrated luminosity times a scale factor as defined in the legend. Events in the histograms have been selected after all pre-selection requirements described in the previous sections and the requirements given by each figure. Each panel displays a different dijet mass range corresponding to the relevant phase space for the benchmark signal indicated.

6.4.3 Jet η

Since a cut on track multiplicity is used in the boson-tagging (see Sec. 6.3.5), the two leading jets are required to have $|\eta| < 2.0$. This requirement avoids the forward calorimeter region and ensures good overlap with the inner tracker. Figure 6.10 shows the distributions of signal and background for reference masses $m_{W'} = 1.2, 1.8, 2.4$ and 3.0 TeV.

6.4.4 Dijet p_{T} Asymmetry

The final-state topology for the signal models considered for this analysis is characterised by the presence of only two very energetic jets coming from the decay of a very massive object. The two jets are required to form a balanced system in the transverse plane, therefore jets are expected to have a small difference in p_{T}. A measure of this difference is the dijet p_{T} asymmetry, defined by

$$A = \frac{p_{\mathrm{T1}} - p_{\mathrm{T2}}}{p_{\mathrm{T1}} + p_{\mathrm{T2}}}. \tag{6.4}$$

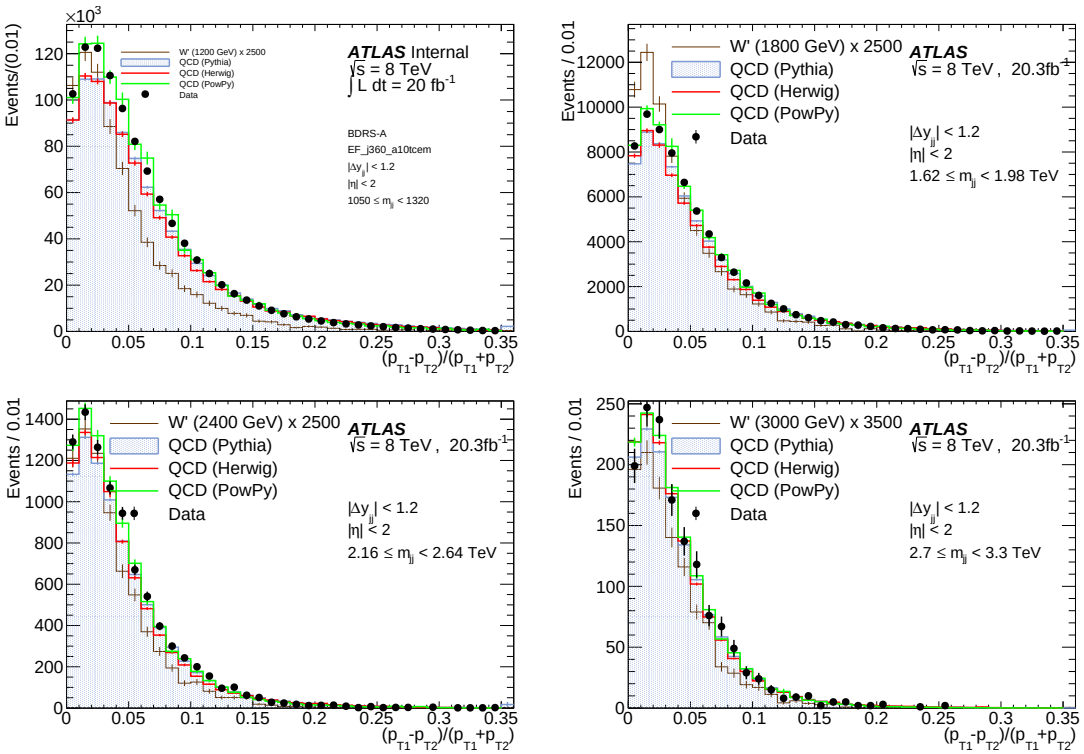


Figure 6.11: Transverse momentum asymmetry distribution between leading and sub-leading BDRS-A jets in QCD MC and W' MC after the particle level calibration. The QCD MC is normalized to the integrated luminosity of the data and has been reweighted according to the p_{T} of the leading jet. The signal MC is normalized to the integrated luminosity times a scale factor as defined in each legend. Events in the histograms have been selected after all pre-selection requirements described in the previous sections and the requirements in each legend. Each subfigure displays a different dijet mass range corresponding to the relevant phase space for the benchmark signal indicated.

The p_{T} asymmetry would be zero for the $W' \to WZ$ signal (and the dijet background) in the absence of radiation and other reconstruction effects which give finite jet p_{T} resolution. Here p_{T1} and p_{T2} respectively denote the p_{T} of the leading and subleading jets. Figure 6.11 shows this distribution for signal, QCD background and data for the reference masses.

Events are required to have $A < 0.15$. This cut rejects events in data where one jet p_{T} is not properly reconstructed due to detector problems or reconstruction anisotropies. The signal event selection efficiency of this requirement is greater than 98%.

6.4.5 Jets in problematic regions of the calorimeter

When the hadronic decays of the boosted bosons sought in this analysis are reconstructed as jets, it is characteristic of these jets to have a two-prong structure following the pattern of quarks from the boosted boson decay. Most of the energy deposited in the calorimeter is very collimated so the resulting clustering profile is a few highly energetic clusters surrounded by less energetic clusters. This pattern is exploited by the filtering and tagging algorithms.

Tile calorimeter modules can suffer low voltage trips. When a trip occurs, the region corre-

sponding to the affected module ($\approx 0.8 \times 0.1$ in the $(\eta - \phi)$ plane) is masked. An offline correction has been defined to account for the energy in the cells of the masked modules. The correction assigns an energy estimate to masked cells based on the average energy density of the neighbouring cells. This correction is not optimal for the narrow high-p_{T} energy deposits that form the jets used in this analysis, because these deposits tend to be contained in a small number of high-p_{T} cells surrounded by very low-p_{T} deposits. When the correction is applied to jets with narrow high-p_{T} deposits in cells neighbouring a masked module, the jets are over-corrected. Conversely, when the correction is applied to jets with narrow high-p_{T} deposits in cells within a masked module, the jets are under-corrected. A detailed study of the effects of this correction on the high-p_{T} jets used in this analysis can be found in Appendix B.

Events with over-corrected jets are rejected to avoid potentially harmful distortions to the di-jet mass spectrum. The variable LC3 has been defined for both the leading jets in the event. LC3 is defined as the number of the three most energetic topoclusters inside a jet falling within a masked module in a module neighbouring a masked module. An event is rejected if $LC3 < 3$.

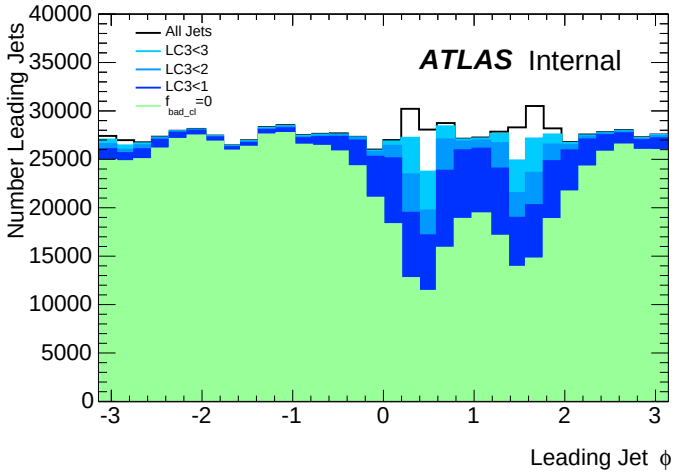


Figure 6.12: ϕ distribution for untagged groomed jets in the data for different $LC3$ cut values. $\mathrm{f}_{bad_{cl}}$ distribution corresponds to events where no problematic clusters have been used while reconstructing the two leading-jets.

Figure 6.12 shows that this prescription is fairly effective at eliminating both under-corrected and over-corrected jets. This figure shows all untagged groomed jets in data versus ϕ. The effects of two dead modules are clearly apparent near $\phi \sim 0.5$ and $\phi \sim 1.5$. The distribution for all jets shows over-corrections near these two modules. The distribution for $LC3 < 3$ is used in this analysis and is sufficient to avoid distortions in the dijet mass spectrum. See Appendix B for further details and a demonstration of this statement.

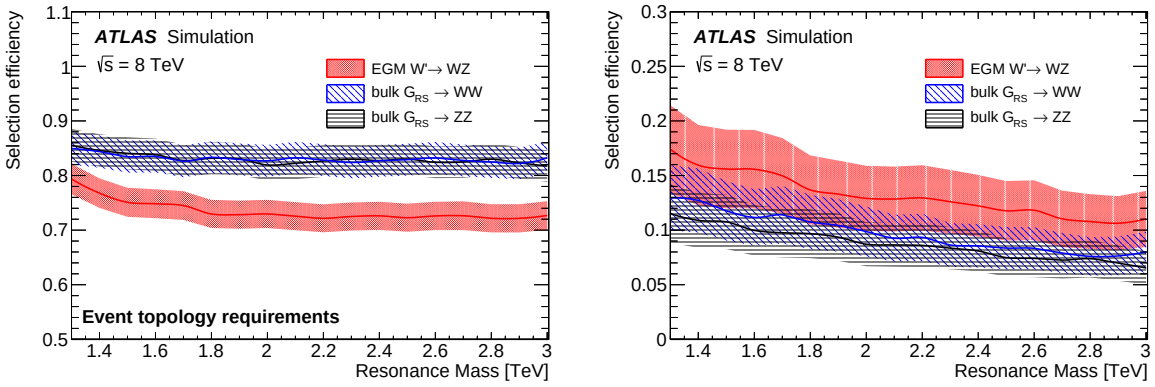
The loss of event efficiency due to the $LC3 < 3$ cut is approximately 5%. This efficiency loss for bad (masked or neighboring-masked) regions is measured in Monte Carlo and verified by data. The resulting per-jet efficiency scale factor is applied to Monte Carlo, as shown in Appendix B.

Table 6.12: Counts and efficiencies for the special jet cleaning.

Run Period	DQ + Trig + JQ	Retained	Efficiency [%]
A	334733	323646	96.7
B	2171478	2104749	96.9
C	610988	589378	96.5
D	1441530	1390444	96.5
E	1133357	1092670	96.4
G	569491	542428	95.2
H	643584	606329	94.2
I	459249	432361	94.1
J	1145081	1078045	94.1
L	372273	350401	94.1
All	8881764	8510451	95.8

6.4.6 Event selection summary

The event selection requirements are listed in Table 6.13. Table 6.14 shows the relative efficiencies of the event selection cuts after the particle level calibration has been applied. Figure 6.13a shows the selection efficiency of the event topology requirements for signal events with resonance mass within 10% of the nominal signal mass for the $W' \to WZ$, bulk $G_{RS} \to WW$ and bulk $G_{RS} \to ZZ$ benchmark models, with statistical and systematic uncertainties indicated by the width of the bands in the figure.



(a) (b)

Figure 6.13: Event selection efficiencies as a function of the resonance masses for EGM $W' \to WZ$ and bulk $G_{RS} \to WW$ and ZZ for simulated events with resonance mass within 10% of the nominal signal mass. In (a), the event topology requirements are applied to EGM $W' \to WZ$, $G_{RS} \to WW$ and $G_{RS} \to ZZ$ samples, while in (b), the WZ, WW and ZZ boson tagging selections are also applied in the EGM $W' \to WZ$, $G_{RS} \to WW$ and $G_{RS} \to ZZ$ samples respectively and the efficiencies shown are corrected by the simulation–to–data scale factor. The width of the bands in each figure indicates both the statistical and systematic uncertainties.

Table 6.13: Event selection requirements.

Selection requirement
Event passes EF_j360_a10tcem trigger
Data quality requirements
Jet cleaning (bad-looser, ugly, jets in problematic region of the calorimeter)
Event has at least two ungroomed C/A 1.2 jets
Leading two jets survive filter
$\|\Delta y_{12}\| < 1.2$
$\|\eta\| < 2.0$
$A < 0.15$
Leading and subleading jets tagged as bosons with two mass windows
$m_{jj} > 1.05$ TeV

6.5 Background parametrization

The search for diboson resonances is performed by looking for statistically significant narrow[4] peaks on a smoothly falling dijet mass spectrum. In the mass range considered, between 1.3-3.0 TeV, the dominant background component are dijet events from QCD processes. Contributions from SM diboson events are expected to account for approximately 6 % of the selected sample, and single boson production is expected to contribute less than 2 %. The contribution of $t\bar{t}$ production has been studied using $MC@NLO$ [70] and $HERWIG++$ [66] samples, and was found to be negligible. The contamination from W-bosons produced by QCD final-state radiation has been studied and it shows a falling dijet mass spectrum, without any turn-on effects at high dijet masses that could mimic a signal.
Although the available theoretical prediction for the QCD background is not expected to accurately describe the data in the very high mass region, the dijet mass spectrum is expected to be smoothly falling. The background estimation is data driven and it is obtained fitting the observed dijet mass spectrum with a parametrized shape. The background model has been empirically characterized by the function:

$$\frac{\mathrm{d}n}{\mathrm{d}x} = p_1(1-x)^{p_2+\xi p_3}x^{p_3}, \tag{6.5}$$

where $x = m_{jj}/\sqrt{s}$, and m_{jj} is the dijet invariant mass, p_1 is a normalisation factor, p_2 and p_3 are dimensionless shape parameters, and ξ is a dimensionless constant chosen after fitting to minimise the correlations between p_2 and p_3[5]. A maximum-likelihood fit, with parameters p_1, p_2 and p_3 free to float, is performed in the range 1.05 TeV $< m_{jj} <$ 3.55 TeV, where the lower limit is dictated by the point where the trigger is fully efficient for tagged jets and the upper limit is set to be in a region where the data and the background estimated by the fit are well below one event per bin for the tagged distributions. The likelihood is defined in terms of

[4]The peak width is expected to be dominated by detector resolution for the EGM $W\prime$, while wider for RS gravitons.

[5]The fit is performed iteratively changing the value of ξ in order to obtain the minimal correlation between the two parameters p_2 and p_3.

Table 6.14: Summary of event selection and cumulative efficiencies for signal samples.

		Efficiency [%]											
		1.2 TeV			1.8 TeV			2.4 TeV			3.0 TeV		
	Data	W'(WZ)	G(WW)	G(ZZ)	W'(WZ)	G(WW)	G(ZZ)	W'(WZ)	G(WW)	G(ZZ)	W'(WZ)	G(WW)	G(ZZ)
Pre-selection	8510451	76.9	90.0	86.7	81.9	93.2	92.9	72.2	91.6	90.7	52.6	90	85.8
2 filtered jets	8387735	76.8	90.0	86.7	81.8	93.2	92.9	72.1	91.6	90.7	52.6	89.9	85.7
$\lvert\Delta y_{12}\rvert < 1.2$	5023199	63.1	77.2	74.2	60.4	77.4	77.0	53.6	75.7	74.6	39.6	74.9	70.2
$\lvert\eta\rvert < 2$	4981723	62.9	77.2	74.2	60.4	77.4	77.0	53.6	75.7	74.6	39.5	74.9	70.2
$A < 0.15$	4069018	59.1	69.4	64.3	57.3	71.5	69.0	51.6	71.6	68.7	37.9	71.8	65.6
Tag (WZ)	2082	15.0	11.2	11.4	12.8	9.4	11.1	10.4	8.2	10.2	7.9	8.3	9.3
$m_{jj} > 1.05$ TeV	602	14.7	10.9	11.0	12.6	9.3	11.0	10.1	8.1	10.1	7.3	8.2	9.2
Tag (WW)	1467	10.5	12.4	6.3	9.4	10.6	6.8	7.2	8.8	5.9	5.6	8.7	5.4
$m_{jj} > 1.05$ TeV	423	10.2	12.1	6.0	9.3	10.5	6.8	7.1	8.7	5.9	5.3	8.5	5.4
Tag (ZZ)	1205	9.1	5.1	10.2	7.5	4.2	9.1	6.2	3.5	8.1	4.6	3.5	6.7
$m_{jj} > 1.05$ TeV	336	8.9	4.9	9.9	7.4	4.1	9.1	5.9	3.5	8.0	4.2	3.4	6.6

events binned in 100-GeV-wide bins in m_{jj} as

$$\mathcal{L} = \prod_i \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}, \tag{6.6}$$

where n_i is the number of events observed in the i^{th} m_{jj} bin and λ_i is the background expectation for the same bin. The selected value of bin width has been chosen based on the expected width of signal models and on the measured dijet mass resolution is 5 % for $m_{jj} > 1$ TeV.

The functional form in eq. 6.5 has also been tested in order to test the flexibility of the background fit model and the compatibility of the expected background with dijet mass distributions obtained using MC simulations and several sidebands in the data[6]. Figure 6.14 shows fits to the HERWIG++ and PYTHIA simulated dijet events that pass the full event selection and tagging requirements on both jets, where the predictions from these leading-order generators are corrected by reweighting the untagged leading-jet p_{T} distributions to match the untagged distribution in data. Figure 6.15 shows the results of fitting the dijet mass spectrum before tagging, and in otherwise tagged events where both the leading and subleading jet have masses falling below the boson-tagging mass windows, in the range $40 < m_j < 60$ GeV. The trigger efficiency as a function of the dijet mass has been taken into account in the fit for the data selected before boson tagging. In particular, the first dijet mass bin has been removed from the fit range, because in this jet sample the trigger becomes fully efficient for higher values of m_{jj}. The fitted background functions in Figures 6.14 and 6.15 are integrated over the same bins used to display the data, and labelled "background model" in the figures. The size of the shaded band reflects the uncertainties on the fit parameters. The lower panels in the figures show the significance, defined as the signed z-value of the difference between the distribution being modelled and the background model's prediction [77]. The significance with respect to the maximum-likelihood expectation is displayed in red, and the significance when taking the uncertainties on the fit parameters into account is shown in blue.

The results of these fits are summarized in Table 6.15, as well as fits to data where the mass of one jet falls in the low-mass sideband ($40 < m_j \leqslant 60$ GeV) and the other falls in a high-mass sideband from $110 < m_j < 140$ GeV, and where both jet masses fall in the high-mass sideband. The dijet mass distribution of the simulated background and of each of these background-dominated selections are well described by the functional form in eq. (6.5).

6.6 Systematic uncertainties

6.6.1 Uncertainties on background parametrization

The uncertainty on the background expectation is determined by the fitting procedure, which assumes a smoothly falling m_{jj} distribution. Several alternative families of parametrisations have been studied in order to assess systematic uncertainties due to the particular choice of background model. Signal plus background fits have also been performed to simulation of the dominant background and control regions of data in which a signal contribution is expected to be negligible. These effects were estimated to be smaller than 25% of the statistical uncertainty at any mass in the search region. The effect of the uncertainty on the trigger efficiency, the

[6] The data sidebands have been obtained varying the tagging requirements on the two leading jets.

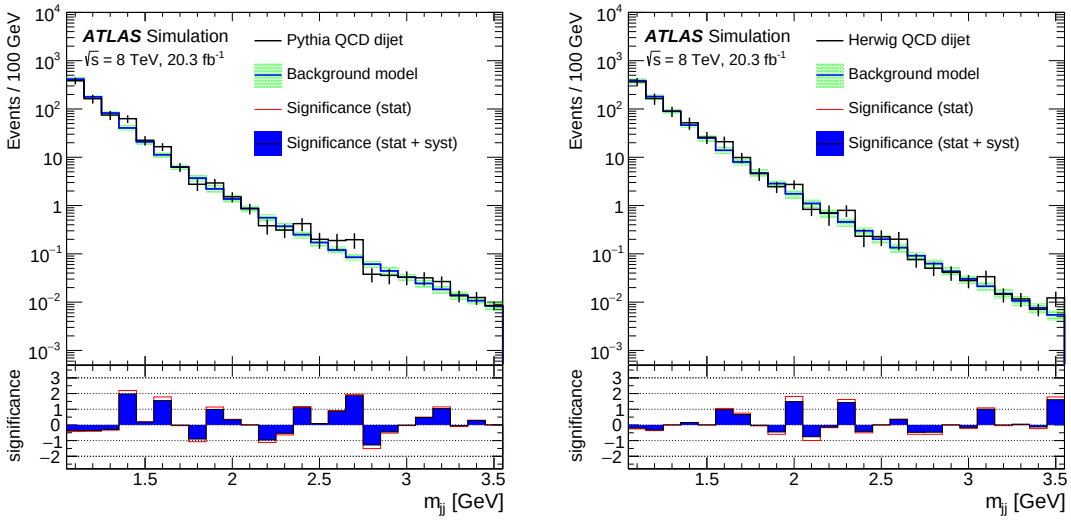


Figure 6.14: Fits of the background model to the dijet mass (m_{jj}) distributions in (left) PYTHIA 8 and (right) HERWIG++ simulated background events that have passed all event selection and tagging requirements. The events are reweighted in both cases to correctly reproduce the leading-jet p_{T} distribution for untagged events, and the simulated data samples were scaled to correspond to a luminosity of 20.3 fb^{-1}. The significance shown in the bottom panels for each bin is calculated using the statistical errors of the simulated data.

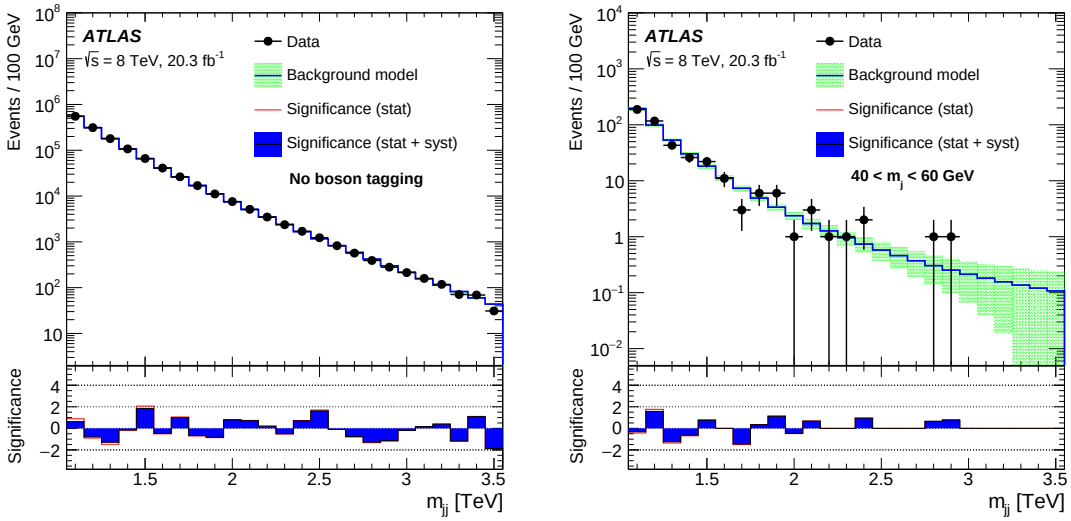


Figure 6.15: Fits of the background model to the dijet mass (m_{jj}) distributions in data events (left) before boson tagging, and (right) where both jets pass all tagging requirements except for the m_j requirement, and instead satisfy $40 < m_j \leqslant 60$ GeV.

Table 6.15: Goodness-of-fit for maximum-likelihood fits of the background model to the dijet mass distribution in simulated events, and in selected mass sidebands from data events where at least one of the leading and subleading jet fails the jet mass selection. One-sided χ^2 probabilities are displayed; for the three data sideband fits, these probabilities were calibrated using pseudo-experiments to avoid biases due to empty bins.

Sample	χ^2/nDOF	Probability
Pythia dijet events	24.6/22	0.31
Herwig++ dijet events	15.9/22	0.82
Data with $110 < m_{j1} \leqslant 140$ GeV and $40 < m_{j2} \leqslant 60$ GeV	12.1/11	0.79
Data with $40 < m_j \leqslant 60$ GeV for both jets	19.8/13	0.56
Data with $110 < m_j \leqslant 140$ GeV for both jets	5.0/6	0.91

variations of the selection efficiencies as a function of the kinematic properties of the background and the composition of the background were also studied and were found to be well-covered by the uncertainties from the fit. Therefore, the only systematic uncertainties related to the background expectation have been evaluated directly by the background estimation procedure using the fit errors as uncertainties.

6.6.2 Uncertainties on signal expectation and shape

The systematic uncertainties associated with the signal expectation and shape are summarised in the following section. These uncertainties are assessed and expressed in terms of nuisance parameters with specific probability density functions (pdfs). The input signal is considered as a function of the nuisance parameters. Where systematic effects only affect the normalisation, the input histogram is simply normalised correspondingly. The overall normalisation is a product of scale factors, each corresponding to an identified nuisance parameter. If the shape is also affected by the systematic uncertainties, the systematic variation is included when the signal histogram is generated.

Both detector effects and MC modeling can be source of systematic uncertainties. They are considered and described in the following sections.

Luminosity scale

The uncertainty in the measurement of the integrated luminosity has been modeled assuming a Gaussian pdf. The discussion in Section 6.1.2 implies that the integrated luminosity scale factor

$$S_L = L_{\text{int}}/(20\ \text{fb}^{-1}) \tag{6.7}$$

has the pdf:

$$P(S_L) = 1.015\ G(S_L|1, 0.028) \tag{6.8}$$

where the coefficient is the ratio of the measured integrated luminosity to the value assumed in the simulation. Here $G(x|\,\overline{x},\sigma)$ denotes a Gaussian pdf:

$$G(x|\,\overline{x},\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\overline{x})^2}{2\sigma^2}} \tag{6.9}$$

Jet energy scale

The systematic uncertainty relative to the jet energy scale (JES) moves the signal peak on a background which is rapidly falling with dijet mass. It reduces the expected sensitivity to the signal when the energy scale decreases. The jet p_{T} (or energy) scale α is defined by

$$p_{\mathrm{T}} = \alpha p_{\mathrm{T}}^{\mathrm{MC}} \tag{6.10}$$

where $p_{\mathrm{T}}^{\mathrm{MC}}$ denotes the value of p_{T} reconstructed in the simulation and p_{T} the value expected in data. This translates directly to a shift in the scale of the dijet mass

$$m_{jj} = \alpha m_{jj}^{\mathrm{MC}} \tag{6.11}$$

assuming that the uncertainty in jet direction is negligible. Furthermore, at very high p_{T} also the uncertainty on the jet mass scale has a negligible effect on the dijet mass scale with respect to the p_{T} scale.

This systematic uncertainty, and also the one on jet mass scale and momentum balance scale have been calculated using track-to-calo double ratios between data and simulation. This technique has been used and described in Ref. [45]. The systematic uncertainty on a particular observable is assessed by applying the jet reconstruction and filtering algorithms to inner-detector track constituents, which are treated as massless, and matching these track jets to the calorimeter jets. The comparison of two measurements of the same object, performed with two separate subsystem of ATLAS, is intended to cancel out theoretical uncertainties, while pointing out any systematic mismodelling in the detector simulation.

This ratio for the jet p_{T} is shown in Figure 6.16 as a function of the p_T of the jet for jets passing the tagging selection requirements $\sqrt{y_f} \geqslant 0.45$ and $60 < j_m < 110$ GeV, in events where $|\Delta y_{jj}| < 1.2$ and $|\eta| < 2.0$. Additionally, the same ratio for a track collection taking into account the tracking inefficiency is shown. There is no evidence of significant dependence on p_{T} and this has been also checked for other kinematic variables such as jet eta and jet mass.

A systematic uncertainty of 1 ± 0.02 is assigned for the p_{T} scale, based on these studies. Treating this as a Gaussian uncertainty, the pdf for the m_{jj} scale is

$$P(\alpha) = G(\alpha|1, 0.02) \tag{6.12}$$

This momentum scale affects only the shape of the signal distribution and is included following Eqn. 6.11 when filling the signal histogram.

Jet energy resolution

The jet energy resolution (JER) changes the reconstructed width of the diboson resonance. A value of 20 % has been considered as systematic uncertainty on JER [47]. A nominal 5%

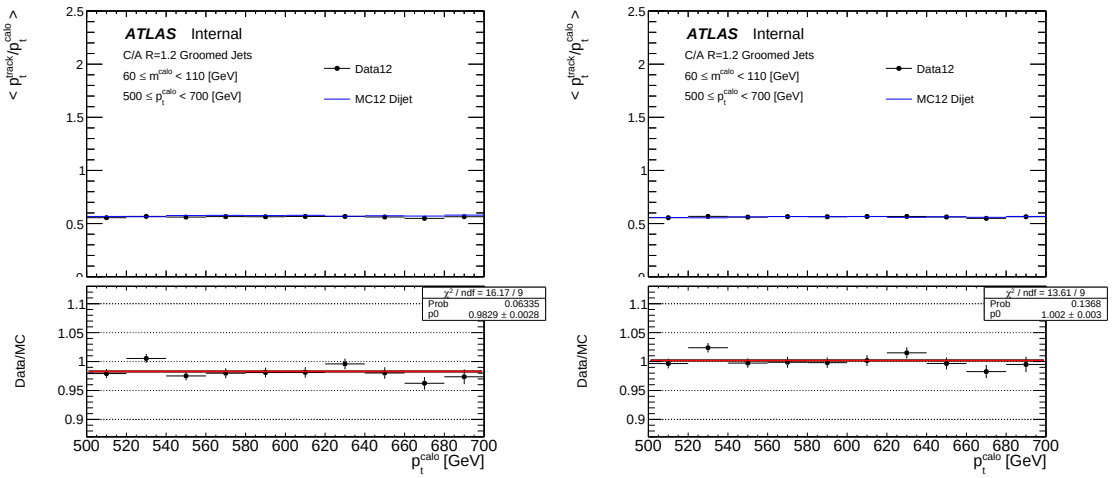


Figure 6.16: Data-to-simulation track-to-calorimeter mean p_{T} double ratio as a function of calorimeter jet p_{T} for the standard simulation (left) and one where the tracking efficiency has been decreased by its uncertainty (right).

JER is derived based on the width of the energy response for the MC signal jets after applying the tagging selection. The effect of a degradation of the jet energy resolution with respect to the nominal MC is modeled by applying a Gaussian smearing to the p_{T} of the reconstructed jets. In order to take into account the degradation of the jet energy resolution, each jet p_{T} is multiplied by a random number generated according to a Gaussian distribution. The Gaussian mean is one and a width corresponding to the nominal JER, σ_{r_E}. The uncertainty on the JER is modeled with a Gaussian pdf for σ_{r_E}

$$P(\sigma_{r_E}) = G(0, 0.05 \times \sqrt{1.2^2 - 1^2}), \tag{6.13}$$

Jet mass scale

The jet mass scale is relevant for the boson-identification procedure. The scale uncertainty on this variable directly produces a change in the normalization of the signal and on its shape. As for the jet p_{T}, the jet mass scale uncertainty has been evaluated using the track-to-calo double ratio technique mentioned in sec. 6.6.2. Figure 6.17 shows these double ratios for the jet mass for jets passing the tagging selection requirements $\sqrt{y_f} \geqslant 0.45$ and $60 < j_m < 110$ GeV, in events where $|\Delta y_{jj}| < 1.2$ and $|\eta| < 2.0$ as a function of the mass of the jet. A 3% uncertainty in the JMS is extracted from these studies.

This result has been cross checked in data and simulation by fitting the V+jets spectra shown in Figure 6.18. The position of the peaks agree and the fits give a mass scale uncertainty of 3.0%.

Based on these considerations, the mass scale α_m:

$$m = \alpha_m m^{\mathrm{MC}} \tag{6.14}$$

is assigned an uncertainty of 3%, i.e. it is described with the pdf

$$P(\alpha_m) = G(\alpha_m | 1, 0.03) \tag{6.15}$$

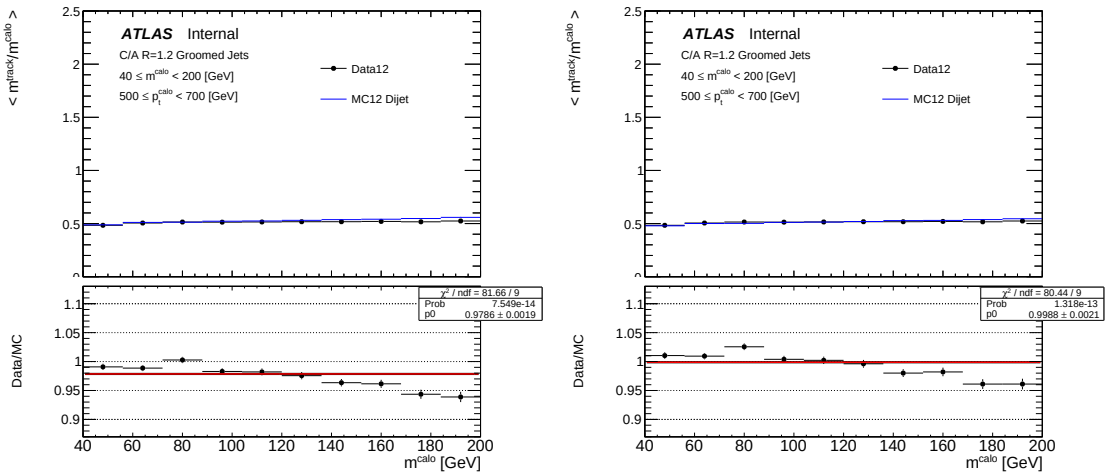


Figure 6.17: Data-to-simulation track-to-calorimeter mean mass double ratio as a function of calorimeter jet mass for the standard simulation (left) and one where the tracking efficiency has been increased by its uncertainty (right).

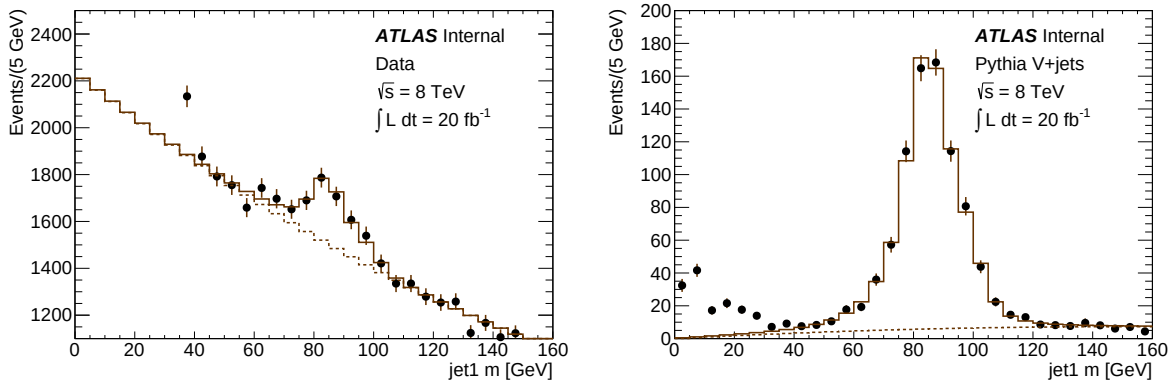


Figure 6.18: Fits to the jet mass spectrum in V+jets events for data (left, showing considerable QCD background) and for simulation (right). Events are required to be accepted by the trigger EF_j360_a10tcem, pass the $|\Delta y_j j| < 1.2$ requirement and their two leading jets are required to have $|\eta| < 2.0$. One jet in the event is required to pass the tagging selections, $\sqrt{y_f} \geqslant 0.45$, and $n_{\text{trk}} < 30$. The p_T of the jets is required to be 0.56 TeV $< p_T <$ 0.65 TeV.

Jet mass resolution

A 20% jet mass resolution uncertainty over the nominal mass resolution has been considered [47].The nominal mass resolution is extracted from the width of the W/Z mass shape in a control sample. The width is obtained by fitting the V+jets spectra shown in Figure 6.18. The widths for data and MC are in good agreement and a nominal resolution of 7.5% has been obtained.

The effect of a degradation of the jet mass resolution with respect to the nominal MC is modeled by applying a gaussian smearing to the m of the reconstructed jets. To degrade the resolution of the jets, each jet mass is multiplied by random number generated according to a Gaussian distribution. The gaussian mean is one and a width σ_{r_m}.The uncertainty on the JMR is modeled

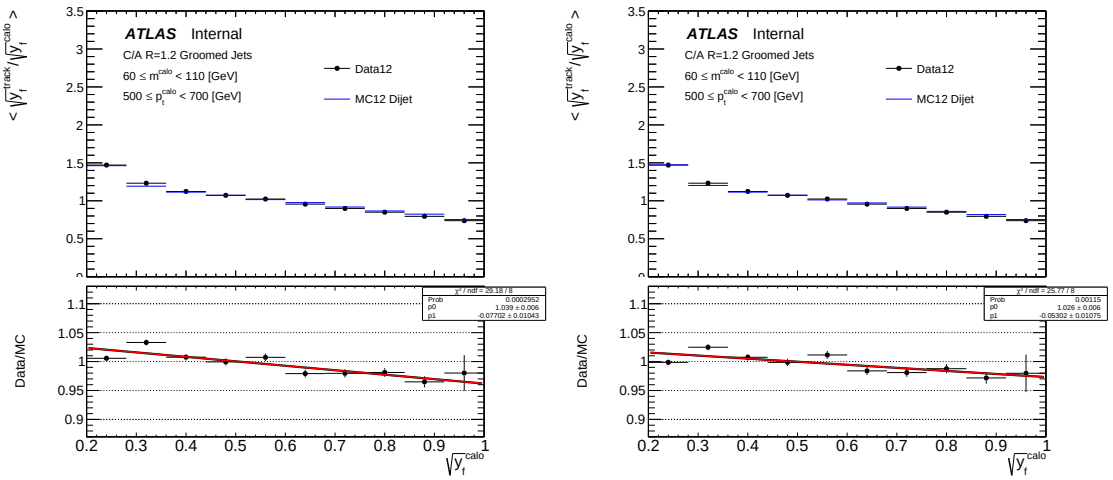


Figure 6.19: Data-to-simulation track-to-calorimeter mean $\sqrt{y_f}$ double ratio as a function of calorimeter jet $\sqrt{y_f}$ for the standard simulation (left) and one where the tracking efficiency has been increased by its uncertainty (right). Both plots represent the data after boson tagging.

with a gaussian pdf for σ_{r_m}

$$P(\sigma_{r_m}) = G(0, 0.075 \times \sqrt{1.2^2 - 1^2}) \tag{6.16}$$

Momentum balance scale

The subjet momentum balance $\sqrt{y_f}$ is an important discriminant for boson jet tagging. Figure 6.19 shows the ratio of the mean track-to-calo ratio for this variable as a function of the variable itself. There are variations over the range of the variable, but for the most important region is the one near the threshold that is applied in the tagging, i.e. near $\sqrt{y_f} = 0.45$. By fitting for an in-situ scale shift:

$$\sqrt{y_f} = \alpha_{\sqrt{y_f}} \sqrt{y_f}^{\mathrm{MC}} \tag{6.17}$$

the scale is off by only 1% at this threshold. Figure 6.19 also shows the same double ratio with the tracking efficiency reduced by its uncertainty, resulting in a 2% shift in the momentum balance scale. The scale is assigned the pdf

$$P(\alpha_y) = G(\alpha_y | 1, 0.02) \tag{6.18}$$

and treated as a signal normalization systematic uncertainty, due to its effect on the selection efficiency.

Momentum balance resolution

A 20% momentum balance resolution uncertainty over the nominal momentum balance resolution has been considered considered [47]. A nominal momentum balance resolution of 16% is extracted from the response of the momentum balance in MC for signal jets.The effect of a degradation of the jet momentum balance resolution with respect to the nominal MC is modeled by applying a gaussian smearing to the $\sqrt{y_f}$ of the reconstructed jets. To degrade the resolution

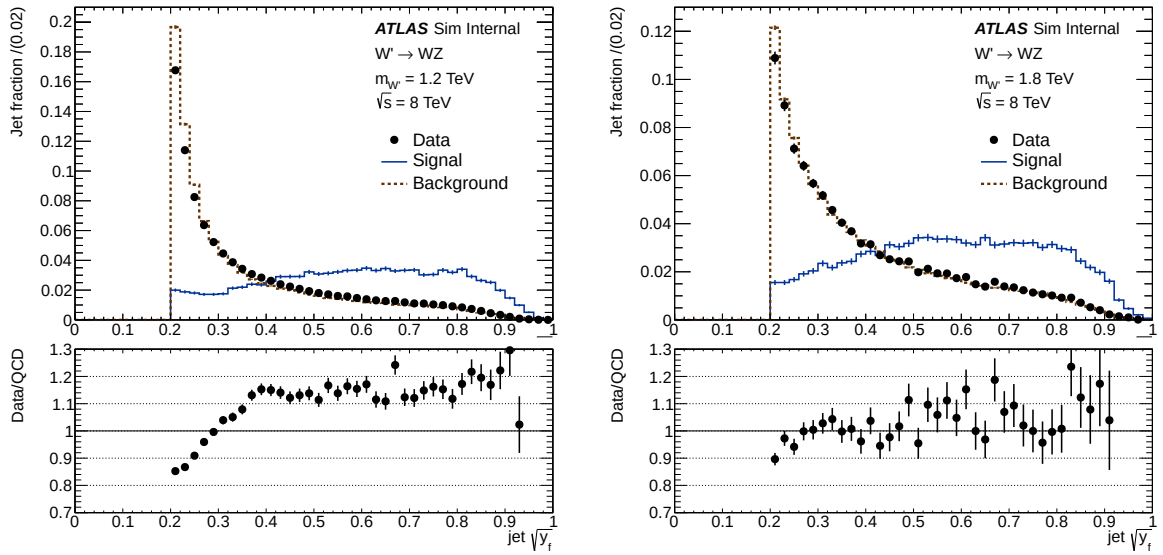


Figure 6.20: Data and simulation distributions of $\sqrt{y_{\mathrm{f}}}$ before before boson tagging but including the jet mass requirement $60 < m < 110$ GeV. Events are required to have dijet mass within 10% of the W' mass indicated on each plot.

of the jets, each jet $\sqrt{y_{\mathrm{f}}}$ is multiplied by random number generated according to a Gaussian distribution. The gaussian mean is one and a width σ_{r_y}.

The uncertainty on the $\sqrt{y_f}$ resolution is modeled with a gaussian pdf for σ_{r_y}

$$P(\sigma_{r_y}) = G(0, 0.16 \times \sqrt{1.2^2 - 1^2}) \tag{6.19}$$

Track-multiplicity efficiency

The n_{trk} variable is not modelled sufficiently well in simulation [78], so it is necessary to apply a scale factor to the simulated signal to correct the selection efficiency of the n_{trk} requirement to match the data. A scale factor of 0.90 ± 0.08 is derived from the ratio of the selection efficiency of this cut in a data control region enriched with W/Z + jets events, where a high-p_{T} W or Z boson decays hadronically, to the selection efficiency of this cut in simulation. The data control region is defined by selecting events in the kinematic range where the jet trigger used in the search is fully efficient, and where only the leading-p_{T} jet passes the tagging requirement on $\sqrt{y}$. Fits to the jet mass spectrum of the leading-p_{T} jet determine the number of hadronically decaying W and Z bosons reconstructed as a single jet that pass the n_{trk} requirement as a function of the selection criteria on n_{trk}. Figure 6.21 shows the jet mass spectra used in this evaluation. The dominant uncertainty on these yields is the mismodelling of the jet mass spectrum for non-W or non-Z jets, and is evaluated by comparing the yields obtained when using two different background models in the fit. The resulting scale factor is 0.90 ± 0.08. Since the n_{trk} requirement is applied twice per event in the selections used in the search, a scale factor of 0.8 is applied per selected signal event, with an associated uncertainty of 20%

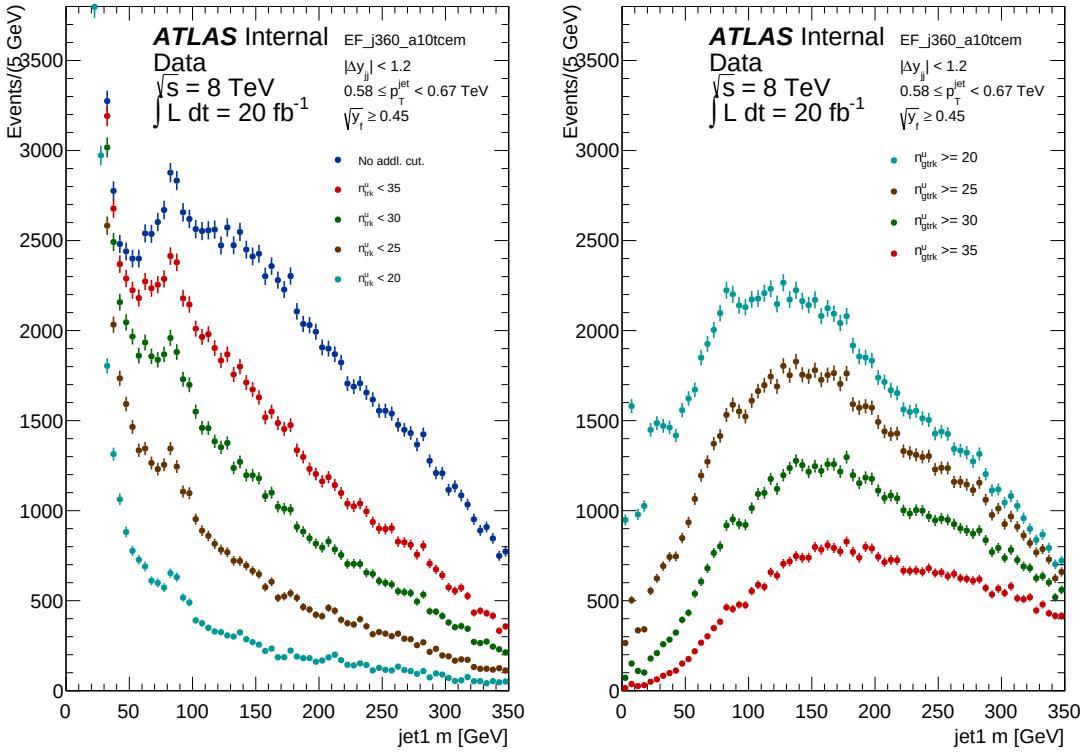


Figure 6.21: Leading jet mass spectra used to evaluate the efficiency of the track-multiplicity cut. The W-boson mass peak is clearly evident at $m_{\text{jet}} = 80$ GeV. The spectrum is shown for different cut thresholds indicated on the legends. On the left there are the spectra with the cut applied as an upper threshold to enhance the boson signal and on the right there are the spectra with the same selection requirements reversed. The efficiency is evaluated for each threshold by comparing the number of events in the two peaks obtained in fits of the two spectra. Both leading and sub-leading jets distributions have been used to evaluate the efficieny of the n_{trk} cut.

The scale factor pdf is found to be well-described by a truncated Gaussian distribution:

$$G_{\text{t}}(\varepsilon \,|\, \varepsilon_0, \sigma_\varepsilon, \varepsilon_{\max}) = \begin{cases} 0 & \varepsilon < 0 \\ Ke^{-\frac{(\varepsilon-\varepsilon_0)^2}{2\sigma_\varepsilon^2}} & 0 \leqslant \varepsilon \leqslant \varepsilon_{\max} \\ 0 & \varepsilon > \max \end{cases} \tag{6.20}$$

with the following parameters:

$$P(S_{\text{L}}) = G_{\text{t}}(S_{\text{t}} | \, 0.893, 0.08, 1.056). \tag{6.21}$$

Other event selection uncertainties

There are systematic uncertainties associated to each of the event selection requirements summarized in Table 6.13, but the dominant uncertainties are those associated with the boson tagging and described above. Given that the trigger, data quality, filtering and dijet p_{T}balance requirements remove less than 2% of the signal events, the associated uncertainties are neglected.

The cut on Δy_{12} removes about 30% of the events but the uncertainty on the jet angles is negligible compared to the width of that distribution. All but the effects of the three boson tagging selection requirements are neglected in the treatment of systematic uncertainties.

PDF, α_S, NLO and beam energy uncertainties on the signal

Uncertainties in the PDFs and α_S used in the signal generation (due to higher order corrections than LO generation) and in the beam energy contribute to the uncertainty in the expected signal shape. At high W' mass, there is a significant off-shell (low-mass) contribution to the cross section, but almost all of the sensitivity of this search comes from events in the on-shell peak. Here the on-shell and off-shell parts are discussed separately in the following.

Neglecting its small contribution to the sensitivity, the off-shell contribution is relevant only when the limits (or more generally the posterior distribution) are converted from signal strength (i.e. relative to the baseline signal histogram) to total cross section. The effect of the off-shell contribution can be eliminated by expressing the limit on a fiducial cross section which excludes the off-shell part. Alternatively, the off-shell fraction can be treated as nuisance parameters with pdf evaluated using PDF error sets and possibly including NLO corrections or uncertainty. Here a mixed approach is used: the signal is normalized using the total cross section, but the uncertainty from the off-shell contribution is neglected.

The normalization of the on-shell part of the cross section is not relevant because the same cross section used to convert limits from signal strength to cross section (see Section 7.3.1) is used to normalize the expected signal histograms. Doubling the cross sections in Table 6.3 would have no effect on the final cross-section limits. It would affect limits on signal strength or coupling strengths, and cross-section normalization uncertainties should be included if those are the final product of a search. That is not the case here.

This leaves only the shape of the on-shell peak. Although the sources of uncertainty considered in this section can all affect this shape, their dependence on m_{jj} is weak. A variation smaller than 1% is expected over the narrow window in m_{jj} where the search is sensitive to the signal. Thus, corrections and uncertainties arising from PDFs, α_S, higher-order corrections to the cross section and beam energy are all neglected.

Central normalization

Figure 6.22 shows the probability density function for the full-scale factors produced by combining the pdfs for the luminosity, parton shower, and track multiplicity:

$$S = S_L \, S_{ps} \, S_t^2 \tag{6.22}$$

where S_L is the scale factor for the integrated luminosity and S_t is that for the track multiplicity cut in the boson jet tagging. The latter enters squared because both jets are required to be boson tagged in this analysis. The mean of the pdf is 0.794 and the RMS is 0.17 (21.4%).

Dependence on m_{jj}

Most of the systematic uncertainties discussed here have been evaluated at the lower end of the dijet mass search spectrum, and there is the problem of assessing them at higher masses.

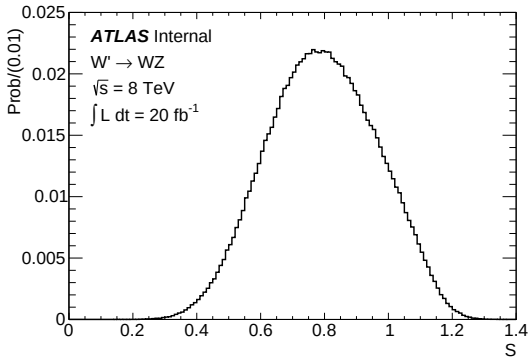


Figure 6.22: Scale factor pdf.

This is a common problem in new physics searches at energies. With the assumption that any additional m_{jj}-dependent in-situ corrections to the selection efficiency are small compared to the variation in efficiency, such additional corrections are found to be negligible. No additional uncertainly is assigned to account for the extrapolation to high-m_{jj}.

6.6.3 Summary of systematic uncertainties

All the systematic uncertainties used and their corresponding models are summarised in Table 6.16.

Table 6.16: Summary of the systematic uncertainties and their corresponding models.

Systematic uncertainty	pdf
Luminosity	$G(S_{\mathrm{L}} \,\vert\, 1, 0.028)$
Jet energy scale	$G(\alpha \,\vert\, 1, 0.02)$
Jet energy resolution (additional smearing)	$G(\sigma_E \vert\, 0, 0.05 \times \sqrt{1.2^2 - 1^2})$
Jet mass scale	$G(\alpha_m \,\vert\, 1, 0.03)$
Jet mass resolution (additional smearing)	$G(\sigma_m \vert\, 0, 0.075 \times \sqrt{1.2^2 - 1^2})$
Momentum balance scale	$G(\alpha_y \,\vert\, 1, 0.02)$
Momentum balance resolution (additional smearing)	$G(\sigma_y, \vert\, 0, 0.16 \times \sqrt{1.2^2 - 1^2})$
Track-multiplicity efficiency	$G_{\mathrm{t}}(S_{\mathrm{t}} \vert\, 0.89, 0.095, 1.07)$
Parton shower	$G(1.0, 0.05)$

Chapter 7

Analysis results

The first section of this chapter reports on the background fit to boson-tagged data. After applying the full selection to the data, an excess of events with respect to the background expectation with $\sim 3\sigma$ significance has been found in the dijet invariant mass spectrum at approximately 2 TeV. For this reason, dedicated studies of the impact of the boson tagging and the event selection have been performed on the data and are described in the second part of the chapter, togheter with additional studies of possible sources of systematics effects coming from detector anisotropies and data-taking instabilities. In the last part of the chapter, the statistical interpretation of the results is discussed and in particular exclusion limits are extracted for RS graviton and EGM W' hypotheses.

7.1 Background fit to data

As described in Sec. 6.5, the background is estimated by fitting the data with an empirical distribution. The results of the background fit to the data are shown in Figure 7.1. In this figure, the fitted background functions are integrated over the same bins used to display the data. Each fitted smoothly falling function is labeled "expectation" and the size of shaded band line reflects the uncertainties of the fitting parameters, which are propagated to show the uncertainty on the expectation from the fit. The lower panels in the figure show the significances, defined as the signed z-value of the difference between the observed data and the expectation [77]. The significance with respect to the maximum likelihood expectation is displayed in red, and the significance when taking the uncertainties on the fit parameters into account is shown in blue. Table 7.1 gives the fitted values of the parameters for the data selected before tagging and after the WZ, WW and ZZ selections, as well as the number of events observed.

7.2 Background shape cross checks

In Figures 7.1a, 7.1b, and 7.1c, obtained after applying the full tagging selection, the dijet invariant mass spectrum shows a local significant deviation from the background only hypothesis in the 2 TeV region. In order to check the assumption of a steeply falling QCD background shape, all the possible systematics effects that could sculpt this spectrum have been investigated.

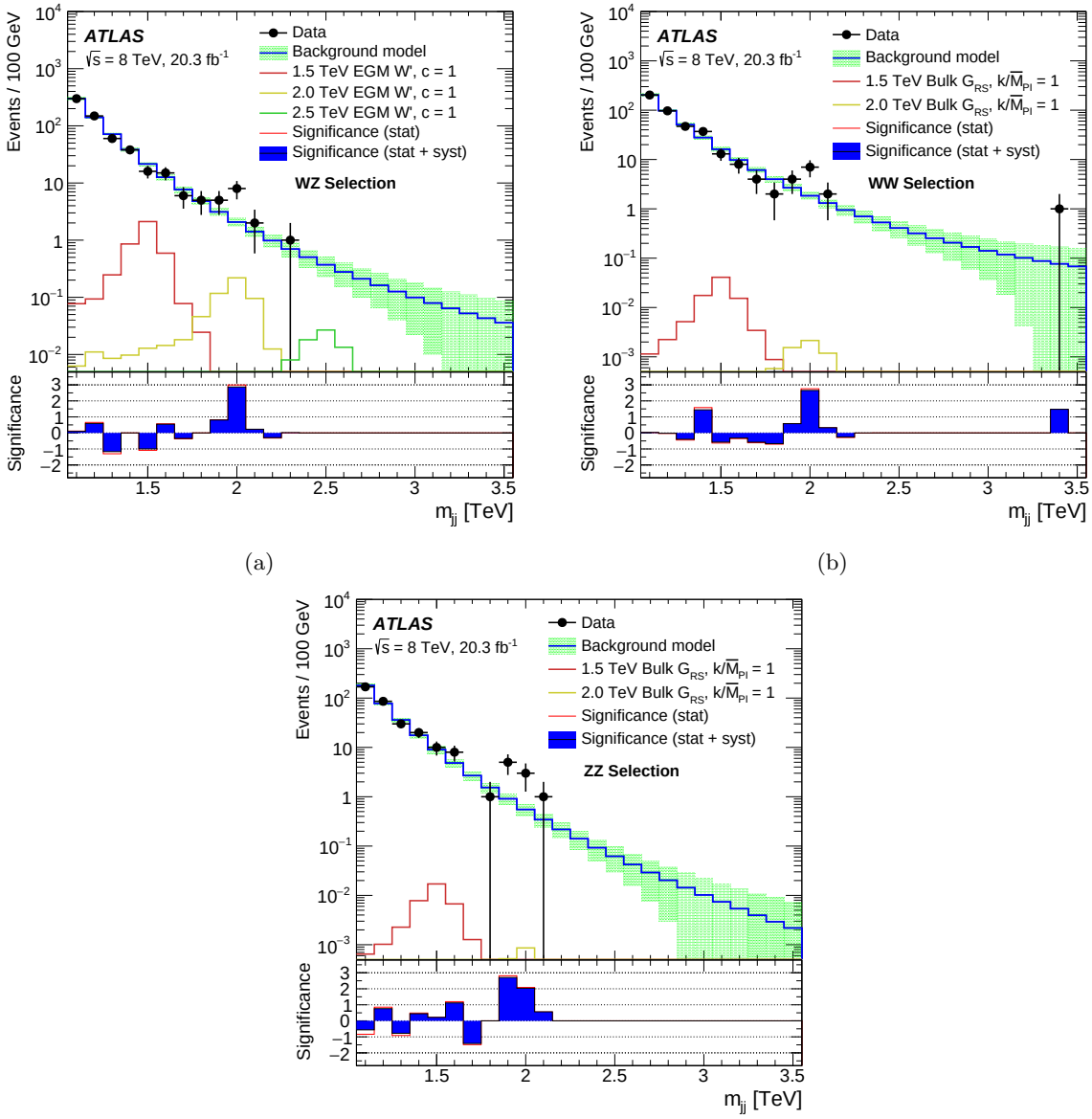


Figure 7.1: Background-only fits to the dijet mass (m_{jj}) distributions in data before tagging (a), after tagging with the WZ selection (b), after tagging with the WW selection (c) and after tagging with the ZZ selection (d). The significance shown in the inset for each bin is the difference between the data and the fit in units of the uncertainty on this difference. The spectra are compared to the signals expected for an EGM W' with $m_{\mathrm{W}'} = 2.0$ TeV or to an RS graviton with $m_{G_{RS}} = 2.0$ TeV.

Table 7.1: Number of observed events and parameters from the background-only fits to the dijet mass spectrum for each tagging selection. The parameter p_1 is translated to n_B, defined as the total background in the range $1.05 < m_{jj} < 3.55$ TeV.

Parameter	before tagging	WZ	WW	ZZ
n_B	$(1.336 \pm 0.001) \times 10^6$	640 ± 25.3	425 ± 20.6	333 ± 18.2
ξ	4.3	3.8	4.2	4.5
p_2	30.95 ± 0.03	31.0 ± 1.4	32.5 ± 1.5	39.5 ± 2.0
p_3	-5.54 ± 0.03	-9.1 ± 1.5	-9.4 ± 1.6	-9.5 ± 2.3
Observed events	1335762	604	425	333

	WZ selection (Events)				WW selection (Events)				ZZ selection (Events)			
m_{jj} bin [GeV]	obs	exp	+1 σ	-1 σ	obs	exp	+1 σ	-1 σ	obs	exp	+1 σ	-1 σ
1050–1150	299	298	313	281	203	202	215	189	169	176	189	163
1150–1250	149	143	147	135	97	98.0	103	93.0	86	77.5	82.9	72.9
1250–1350	60	71.4	75.3	67.2	47	50.8	54.1	47.4	30	36.7	40.0	33.5
1350–1450	38	38.2	41.0	35.4	37	27.9	30.3	25.5	20	18.5	20.8	16.4
1450–1550	16	21.5	23.4	19.6	13	16.1	17.8	14.5	10	9.86	11.3	8.52
1550–1650	15	12.6	13.9	11.4	8	9.75	10.9	8.62	8	5.50	6.39	4.66
1650–1750	6	7.67	8.56	6.86	4	6.12	6.90	5.35	0	3.19	3.77	2.67
1750–1850	5	4.82	5.47	4.25	2	3.98	4.54	3.43	1	1.92	2.32	1.58
1850–1950	5	3.12	3.62	2.70	4	2.67	3.09	2.27	5	1.19	1.48	0.96
1950–2050	8	2.08	2.47	1.75	7	1.84	2.18	1.53	3	0.76	0.98	0.59
2050–2150	2	1.41	1.74	1.15	2	1.30	1.58	1.06	1	0.50	0.67	0.37
2150–2250	0	0.98	1.26	0.77	0	0.95	1.18	0.74	0	0.34	0.48	0.23
2250–2350	1	0.70	0.94	0.52	0	0.70	0.91	0.52	0	0.23	0.35	0.15
2350–2450	0	0.50	0.71	0.35	0	0.53	0.71	0.37	0	0.16	0.26	0.09
2450–2550	0	0.37	0.55	0.24	0	0.41	0.58	0.27	0	0.12	0.20	0.06
2550–2650	0	0.28	0.44	0.17	0	0.32	0.48	0.20	0	0.08	0.16	0.04
2650–2750	0	0.21	0.36	0.12	0	0.25	0.40	0.15	0	0.06	0.13	0.02
2750–2850	0	0.16	0.29	0.08	0	0.21	0.35	0.11	0	0.05	0.11	0.02
2850–2950	0	0.13	0.25	0.06	0	0.17	0.30	0.08	0	0.04	0.09	0.01
2850–2950	0	0.13	0.25	0.06	0	0.17	0.30	0.08	0	0.04	0.09	0.01
2950–3050	0	0.10	0.21	0.04	0	0.14	0.27	0.06	0	0.03	0.08	0.01
3050–3150	0	0.08	0.18	0.03	0	0.12	0.25	0.05	0	0.02	0.07	0.00
3150–3250	0	0.06	0.16	0.02	0	0.10	0.23	0.03	0	0.02	0.06	0.00
3250–3350	0	0.05	0.15	0.02	0	0.09	0.22	0.03	0	0.01	0.06	0.00
3350–3450	0	0.04	0.13	0.01	1	0.08	0.21	0.02	0	0.01	0.05	0.00
3450–3550	0	0.04	0.12	0.01	0	0.07	0.20	0.01	0	0.01	0.05	0.00

Table 7.2: Number of events observed in the WZ, WW, and ZZ selected samples in each dijet mass bin used in the analysis, compared to the prediction of the background-only fit.

In particular, checks on selection effects, jet reconstruction, and detector issues are summarized in the following sections.

7.2.1 Kinematic cross checks

It is crucial that the cuts that are applied to the jets do not affect the shape of the background, so that the dijet mass function used to model it is valid. In order to check these aspects, various distributions of the jet kinematic variables have been produced and compared in jet mass signal and control regions (sidebands). The dijet mass distribution has been also produced by applying the efficiency of each of the boson tagging cuts to a data driven background distribution, to check for any distortion of the spectrum introduced by the tagging.

Sideband Comparisons

The distributions of various kinematic quantities have been compared for jets in the several jet mass control regions.The variables considered are the tagging variables, n_{trk} and $\sqrt{y_f}$, as well as η, ϕ, p_T for both the jet and dijet masses. Both jet leading and and sub-leading jet were considered and the distributions for the two jets have been compared and plotted against the dijet mass to look for any correlations. The correlation of the jet η with the dijet mass is shown in Figure 7.2 for the leading jet and Figure 7.2 for the sub-leading jet. No significant correlation is observed between η and the dijet mass. The full set of plots and a detailed description of the tests are reported in Reference [74] and they show no significant correlations between tagging and kinematic variables with the dijet mass.

Figure 7.4 shows fits of the background model to the dijet-mass distributions in data events where both jets pass all tagging requirements except for the m_j requirement. No significant excesses are observed in dijet-mass spectra obtained selecting jets in jet-mass sidebands.

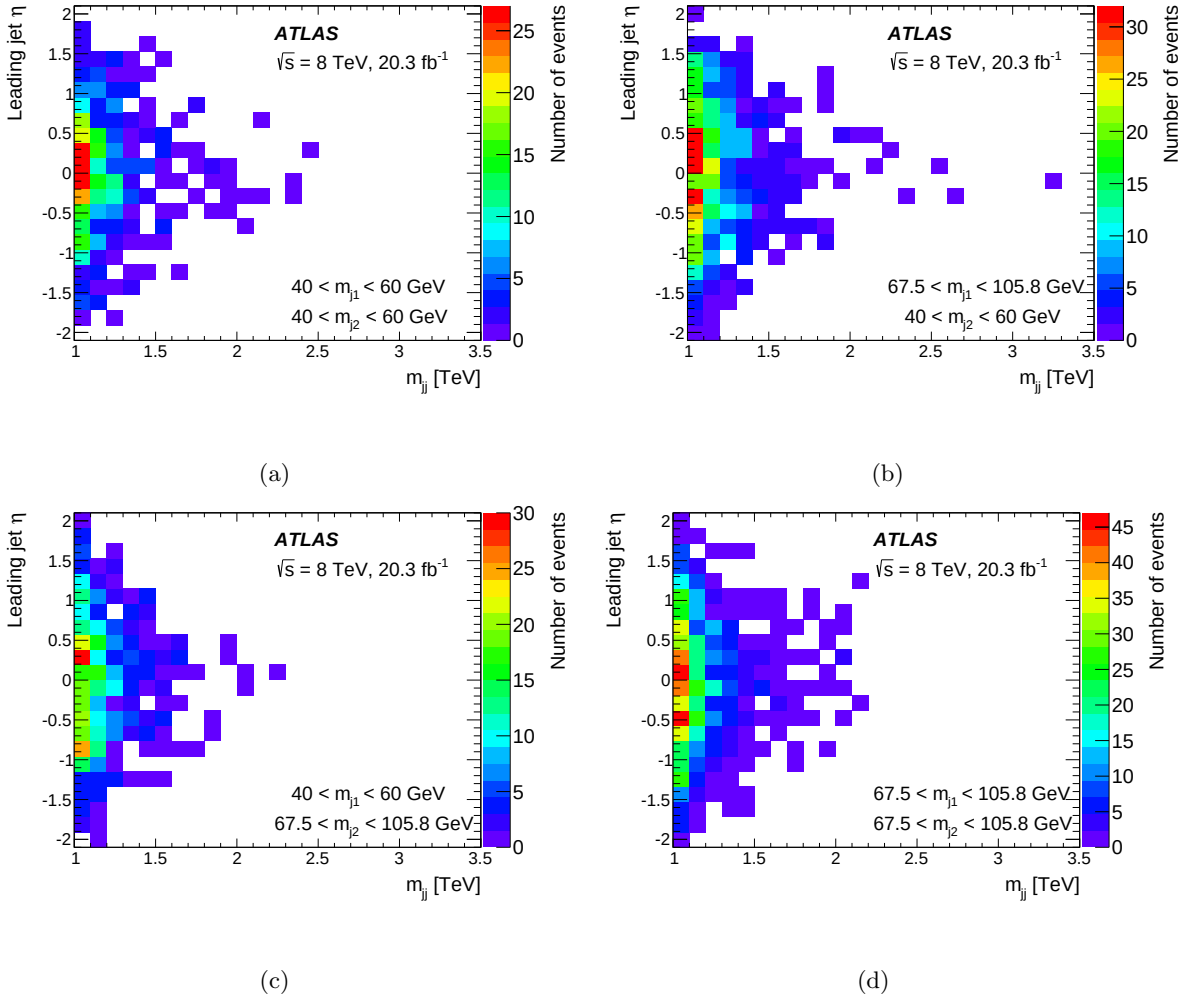
Cut Efficiencies

An excess of events in the dijet mass distribution might be produced by a rapid change in the shape of the efficiency for one of the tagging cuts. The efficiency for n_{trk} and cuts $\sqrt{y_f}$ was calculated for both the leading and the sub-leading jet as a function of the jet p_T. To preserve the statistics of the sample, the efficiency was calculated applying the boson tagging only on one of the two leading jets at a time. For example to calculate the efficiency for n_{trk} cut for the sub-leading jet the formula is

$$\mathrm{Eff}(n_{trk}^{jet2}) = \frac{n_{evt}(60 < m_{jet2} < 110 \text{ GeV}, n_{trk}^{jet2} < 30)}{n_{evt}(60 < m_{jet2} < 110 \text{ GeV})} \tag{7.1}$$

For both numerator and denominator the events are required to survive all the cuts that are included in the event selection but not in the boson tagging. For these tests only the wide mass window requirement is considered. The plots of these efficiencies for both the leading and sub-leading jet, for the n_{trk} and $\sqrt{y_f}$ cuts are shown in Figure 7.5.

The efficiency was also calculated for two out of the three boson tagging cuts after the other has already been applied. This is necessary to check if the two tagging cuts might cooperate



(a) (b)

(c) (d)

Figure 7.2: The distributions of the leading jet η and the dijet mass in several control regions (a-c), and in a region containing the three signal regions (d).

in producing "sculpting" effects on observed the dijet mass spectra. For example the efficiency of the cut on n_{trk} after the cut on $\sqrt{y_f}$ has already been applied to the jet, again for both the leading and sub-leading jets as a function of p_T. This was done using a similar formula as above:

$$\text{Eff}(n_{trk}^{jet2}|\sqrt{y_f^{jet2}} \geqslant 0.45) = \frac{n_{evt}(60 < m_{jet2} < 110GeV, \sqrt{y_f^{jet2}} \geqslant 0.45, n_{gtrk}^{jet2} < 30)}{n_{evt}(60 < m_{jet2} < 110GeV, \sqrt{y_f^{jet2}} \geqslant 0.45)} \tag{7.2}$$

giving the efficiency for sub-leading jet for the cut on n_{trk} after applying the cut on $\sqrt{y_f}$. The full set of plots and a detailed description of the tests are reported in Reference [74] and they exclude systematic effects due to the shape of the efficiency for the tagging cuts.

Furthermore, dijet mass distributions depleted of signal have been produced by requiring that one jet passes the boson tagging selection, and the other is in the correct mass window but fails one of the tagging cuts. The efficiencies evaluated above for the different tagging selections have been applied to the partially-tagged jets, in order to emulate the effect of the selection that has

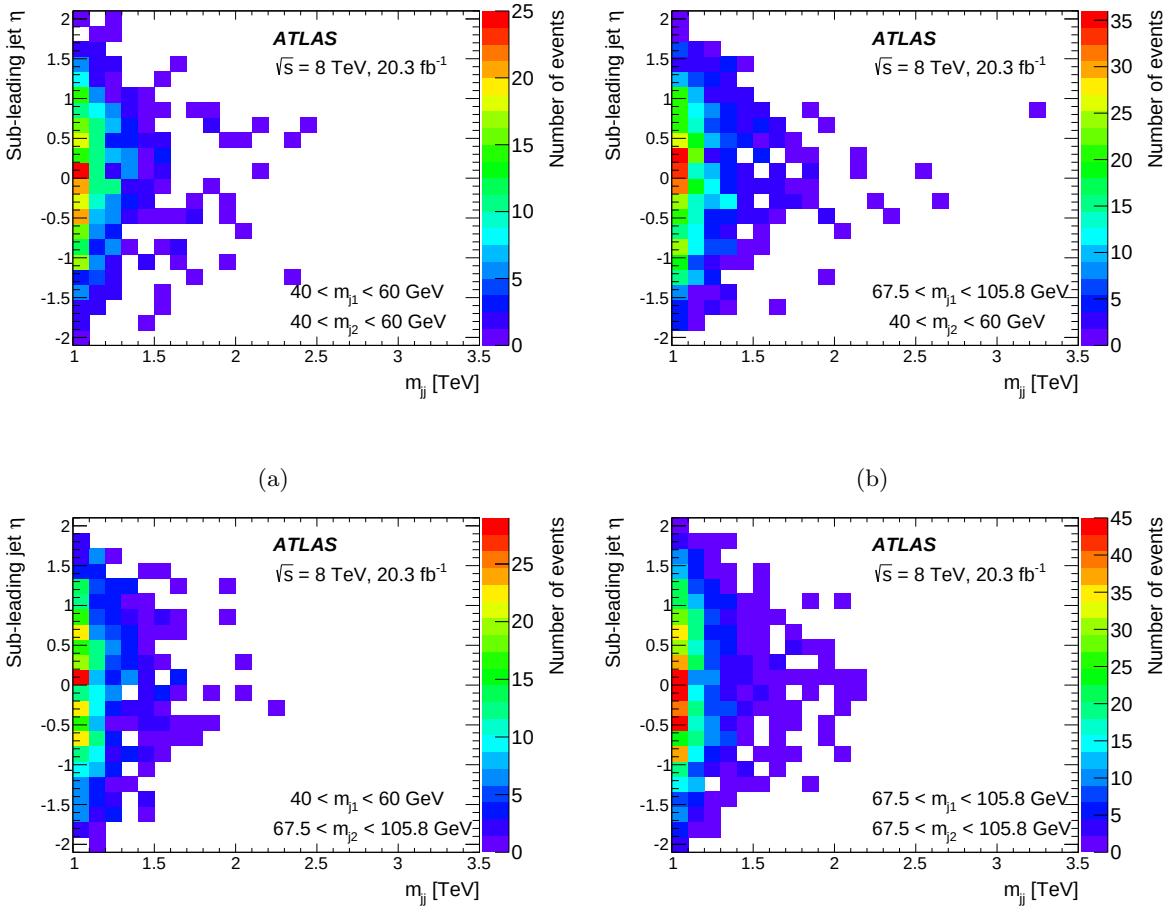


Figure 7.3: The distributions of the subleading jet η and the dijet mass in several control regions (a-c), and in a region containing the three signal regions (d).

been removed. Also these tests are reported in Reference [74] and they result in smooth dijet mass distributions, providing an additional argument in excluding systematic effect due to the shape of the tagging efficiency.

7.2.2 Effect of the tagging cuts

To check that the distortion in the dijet mass spectrum was not caused by one particular cut in the boson-tagging selection, the dijet-mass distribution was plotted varying the set of tagging cuts applied. First the dijet-mass spectrum was made applying only one of the tagging cuts at a time. The result is shown in Figure 7.6, where applying only one tagging cut the dijet mass spectrum remains smoothly falling. None of these cuts alone produces a distortion in the spectrum.

A second cross check was done by looking at the dijet mass spectrum removing only one tagging

cut from the full event selection. Figure 7.7 shows a smooth spectrum when the $\sqrt{y_f}$ cut is removed, while the excess can is still visible when either the jet mass or n_{trk} cuts are removed. This is effect is expected, given the background efficiency 6.3.2 of each of the boson-tagging cuts.

In Figures 7.8 and 7.9 the bakground fit is performed on the obtained m_{jj} distributions and also an EGM W' signal with $m_{W'} = 2$ TeV (passing the same event selection) is added to the plots. The spectra obtained applying n_{trk} and $\sqrt{y}$ selections are shown in Figure 7.8a and Figure 7.8b. These selections are relatively weak compared to the WZ jet mass selection shown in Figure 7.8c.

Effect of changing the $\sqrt{y_f}$ cut

Two cross checks have been made to study the $\sqrt{y_f}$ cut. The first has been done by producing the dijet mass spectrum for different values of the cut and refitting the resulting spectrum. The distributions obtained are shown in Figure 7.10 for values varying between 0.4 and 0.6. These plots show that the fit performs well in all cases and the distortion gradually appears as the cut is increased, although tightening the $\sqrt{y_f}$ cut the statistics becomes very limited. This also shows that the events in the excess at 2 TeV are not all from near the threshold of the cut.

The dijet mass distribution was also produced with one jet in the wide mass window, $60 < m < 110$ GeV and one jet in the mass sideband, $40 < m < 60$ GeV, when the standard $\sqrt{y_f} < 0.45$ cut is applied and for a tighter cut of $\sqrt{y_f} < 0.55$. Additionally the jet in the sideband is required to have $0.06 < m/p_T < 0.12$. The results are in Figure 7.11, where no structure can be seen in the distribution.

A similar study has been performed also for the n_{trk} selection, and no evidence of bias or other problems has been observed [74].

7.2.3 Comparison between calorimeter and track reconstructed jets

Some of the events with $m_{jj} > 1.8$ TeV have one or both the jets falling in the crack region of the Tile calorimeter, or close to this region. The calorimeter compensation for an energy deposit in the crack region could shift up or down the energy response and then shift the angular coordinates of the reconstructed jets. To exclude angular mismeasurements, the reconstructed calorimeter BDRS-A jets have been matched to the closest BDRS-A track-jet. The matching criterion is $\Delta R_{Calo-Track} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$.

The scatter plots in Figure 7.12 show the comparison between the calorimeter-measured and track-measured η coordinates for both leading and subleading jets. The colors of the markers are related to the different m_{jj} bins. For the highest mass bin (in red) the η measurements for calorimeter and track jets are in good agreement.

An important ingredient in the m_{jj} measurement is the opening angle between the two leading jets. Figure 7.13 shows the measurements of the opening angle and the difference in rapidity between the two leading calorimeter jets and matched leading track jets. Also in this case, for the events in the highest mass bin (red) the measurement of calorimeter and tracker are in good agreement.

Figure 7.14 shows another comparison between calorimeter and track jets. For each calorimeter jet matched with a track jet, a "mixed" four-vector is made. The energy variables (p_T and mass) are taken from the calorimeter jet and the angular variables (η and ϕ) from the track jet. The calculation of the m_{jj} is then performed with the mixed 4-vectors and compared with the one obtained with the calorimeter jets. Possible mismeasurements of the angular coordinates of the jets would appear as outliers in the scatter plot in Figure 7.14, where the mixed invariant mass (Y axis) is plotted as a function of the calorimeter invariant mass (X axis). The plot in Figure 7.14 also shows that the ordering of the tagged di-jet events in m_{jj} is preserved using the mixed four-vectors and this excludes migration effects due to problems in the measurement of the angular coordinates of the jet four-vector.

7.2.4 Data taking stability

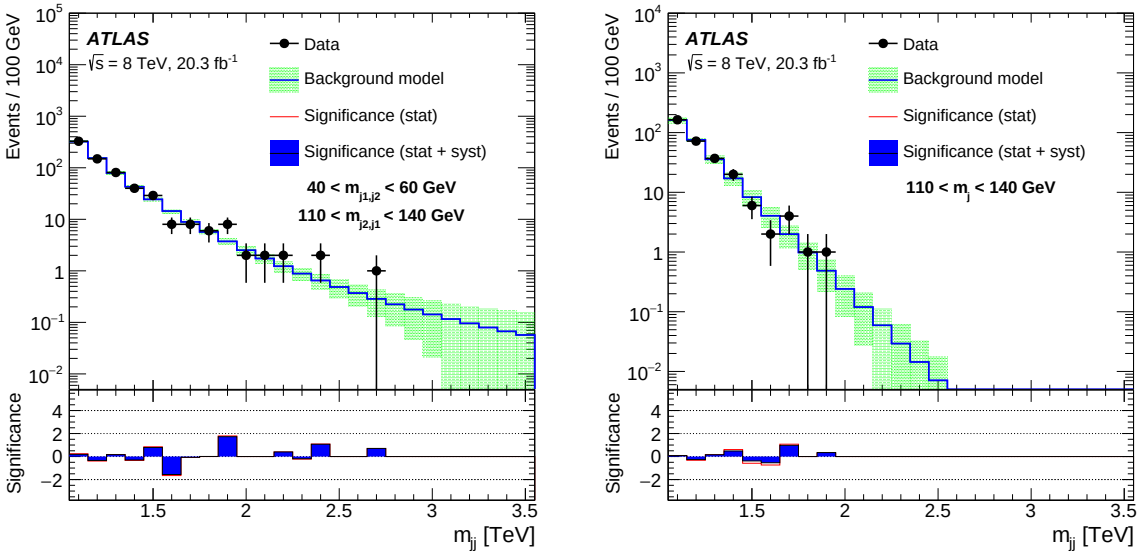
Figure 7.15 shows the m_{jj} for selected boson-tagged (with a jet mass window of 60-110 GeV) events as a function of the data taking run number.

Figure 7.16 shows the cut-flow histograms for the different data taking periods. The entries of the histograms are normalized to the number of events passing the trigger selection. The last three bins show the effect of the application of each of the 3 tagging cuts (jet mass cut, $\sqrt{y_f}$ and number of ghost associated tracks), after applying the full event selection. The ratio plot shows a spread after the LC3 cut. This effect is due to the sensitivity of this cut, implemented to remove jets in problematic region of the calorimeter, which vary over the different data taking periods.

The same normalization procedure has been applied in Figure 7.17, where the number of selected and boson-tagged events in 5 m_{jj} bins are shown as function of the data taking period. The different cut-flow curves are consistent in the statistical uncertainties, and do not show any suspicious effect in the data taking conditions that could be responsible for the observed excess of events in the high dijet mass region.

7.2.5 Stability of the trigger selection

As described in the previous chapter, events in this search must satisfy the EF_j360_akt10em trigger. Here, the stability of the trigger efficiency turn-on curve as a function of jet p_{T} over the 2012 data-taking period is evaluated. The study uses data collected with an uncorrelated muon trigger selection and requires at least one C/A 1.2 jet within $|\eta| < 2.0$. The trigger efficiency is plotted and fit as a function of the p_{T} of the leading jet satisfying these requirements; since the number of tagged jets in the muon sample is small, fitting the trigger efficiency curve as a function of tagged-jet p_{T} for each run period does not produce conclusive results. Table 7.3 shows that the plateau efficiency is consistent across all run periods. The difference among the thresholds and widths in different run periods is larger compared to the fit statistical errors, but no clear trend is observed.



(a) (b)

Figure 7.4: Fits of the background model to the dijet mass distributions in data events where both jets pass all tagging requirements except for the jet mass requirement. In (a) one jet satisfies $40 < m_j \leqslant 60$ GeV, and the other satisfies $110 < m_j \leqslant 140$ GeV, and in (b) both jets satisfy $110 < m_j \leqslant 140$ GeV. The significance shown in the bottom panel for each bin is the difference between the data and the fit in units of the uncertainty on this difference.

Table 7.3: Fitted trigger-efficiency parameters for untagged C/A 1.2 jets measured from muon stream data, for the run periods of the 2012 dataset.

Run Period	Plateau efficiency	Threshold	Width
A	0.988 ± 0.009	464 ± 3	52 ± 3
B	0.987 ± 0.004	462 ± 1	46 ± 1
C	0.985 ± 0.008	457 ± 2	43 ± 2
D	0.989 ± 0.004	459 ± 1	45 ± 1
E	0.979 ± 0.006	459 ± 1	42 ± 2
G	0.987 ± 0.008	463 ± 2	45 ± 2
H	0.992 ± 0.006	465 ± 2	46 ± 2
I	0.988 ± 0.008	467 ± 2	56 ± 3
J	0.977 ± 0.007	462 ± 1	43 ± 2
K	0.998 ± 0.007	468 ± 3	52 ± 3

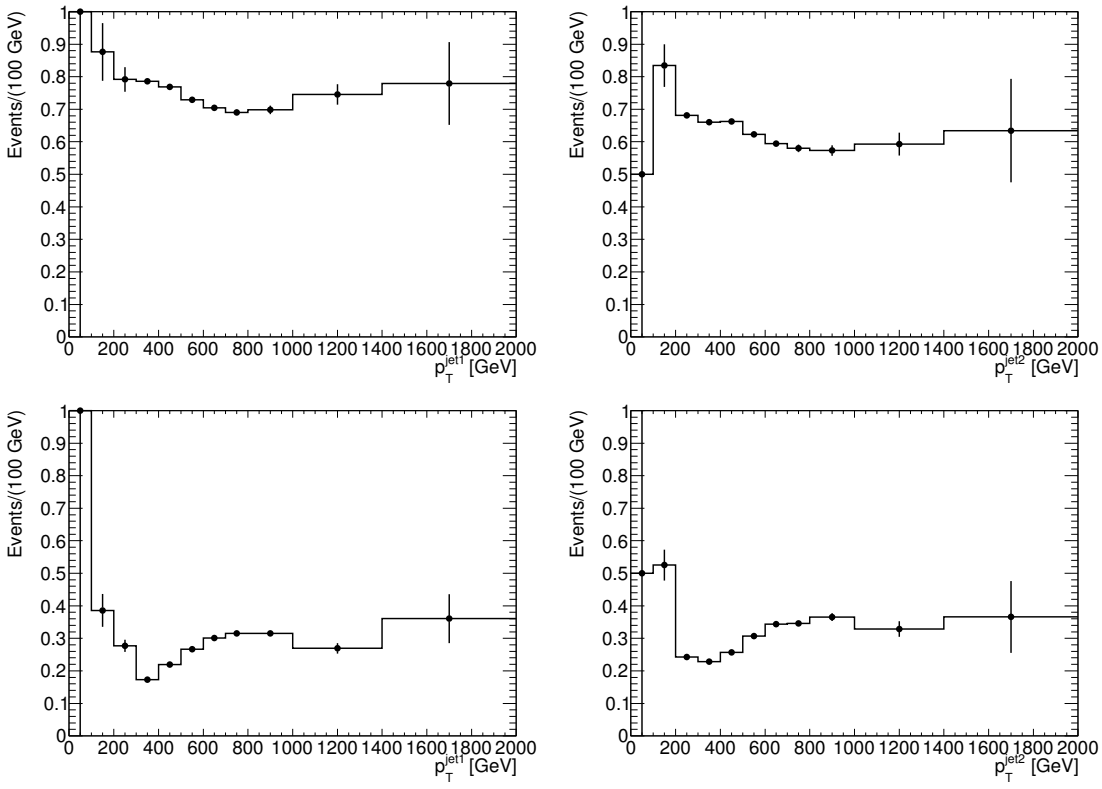


Figure 7.5: The cut efficiencies for the cut on n_{trk} (top) and $\sqrt{y_f}$ (bottom) for the leading (left) and subleading (right) jets.

7.3 Statistical analysis

A frequentist statistical analysis is used to interpret the search results. For each of the benchmark new physics models tested, the parameter of interest in the statistical analysis is the signal strength, μ, defined as a scale factor on the total number of signal events predicted by the model. Thus, the background only hypothesis corresponds to $\mu = 0$, and the hypothesis of a signal plus background model corresponds to $\mu = 1$. A test statistics, $\lambda(\mu)$, based on the profile-likelihood ratio [79] is used to test the new physics models. The test statistics is designed to extract the information on μ from a maximum likelihood fit of the signal plus background model to the data. The significance of an excess observed in the data with respect to the background only expectation is quantified in terms of the local p_0, defined as the probability of the background only model to produce an excess at least as large as the one observed. The global probability of an excess with local p_0 as the most significant excess to be observed anywhere in the search region is quantified with pseudo experiments taking into account the overlap of events in the different signal regions. Exclusion limits of 95% confidence level are set in the absence of a significant excess over the background only prediction following the CL_s prescription [80].

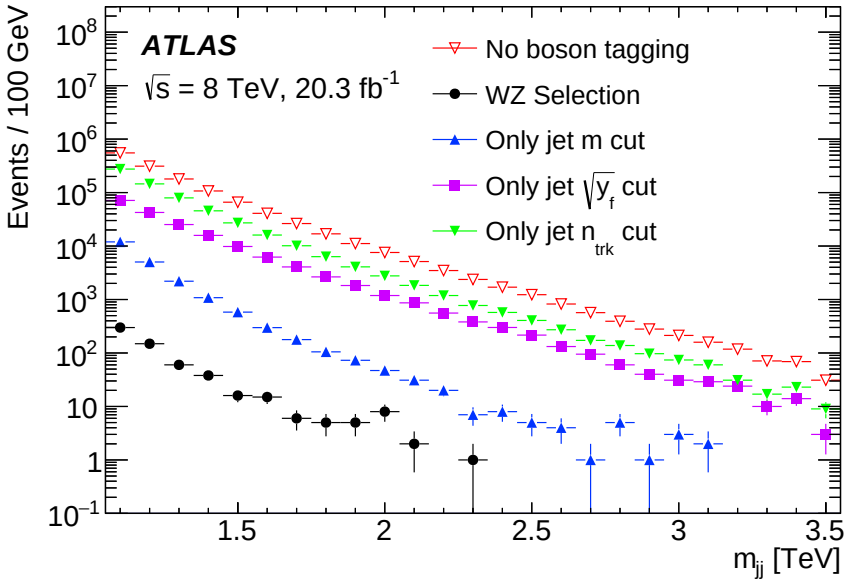


Figure 7.6: The dijet mass distribution formed when only one of the tagging cuts is applied. The fully tagged spectrum (black dots) is shown for comparison

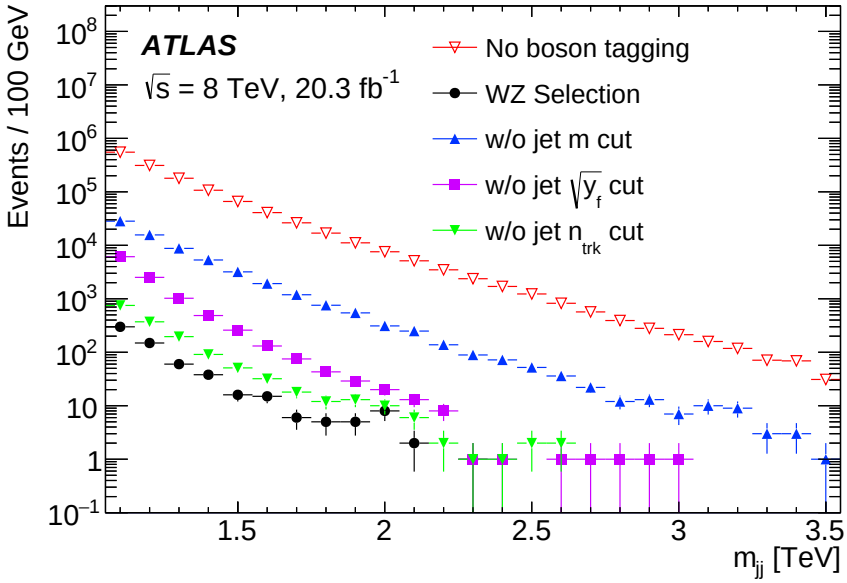
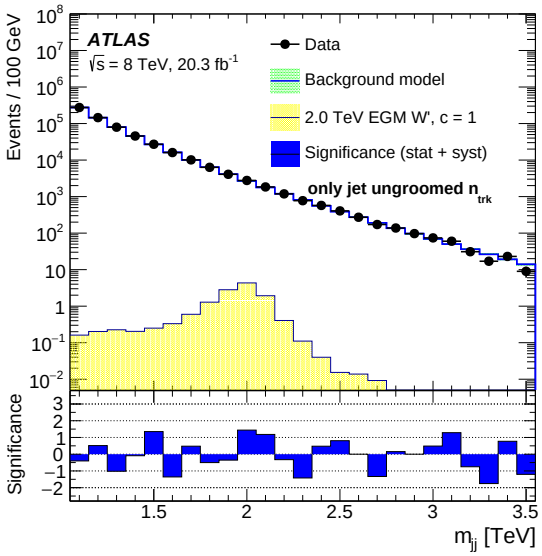
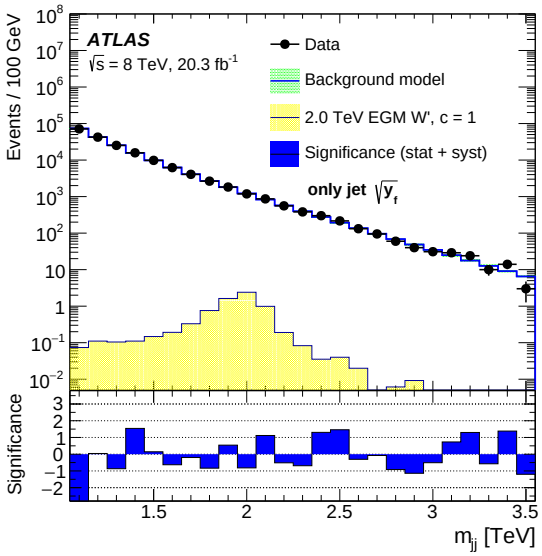


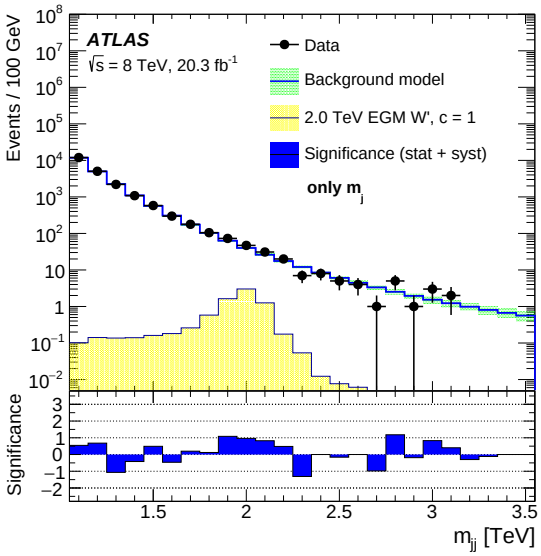
Figure 7.7: The dijet mass distribution formed when one of the tagging cuts is removed, along with the fully tagged spectrum for comparison



(a)

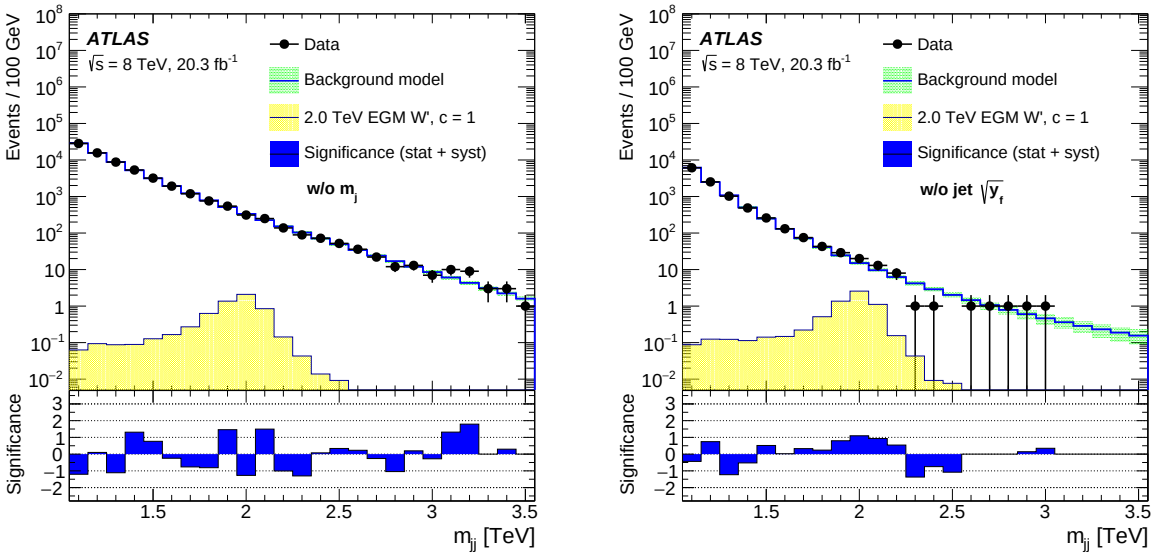


(b)



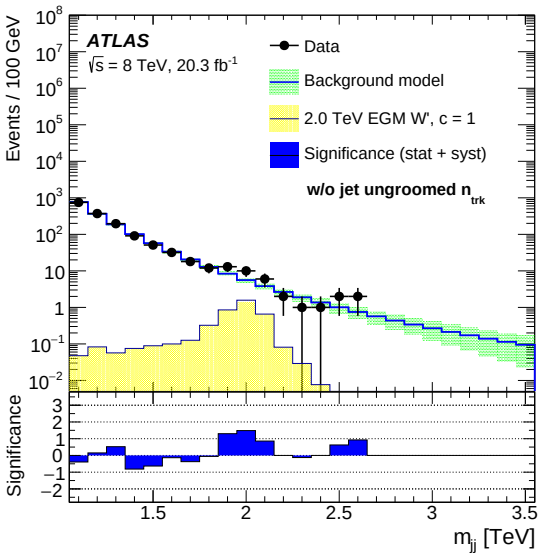
(c)

Figure 7.8: Comparison of the observed m_{jj} spectrum and the expected spectrum plus a 2 TeV W' signal, when applying a single boson tagging requirement. In (a) only the n_{trk} requirement is applied to each jet, in (b) only the $\sqrt{y}$ requirement is applied to each jet, and in (c) only the jet mass requirements corresponding to the WZ selection are applied. In each figure, the significance shown in the bottom panel for each bin is the difference between the data and the fit in units of the uncertainty on this difference.

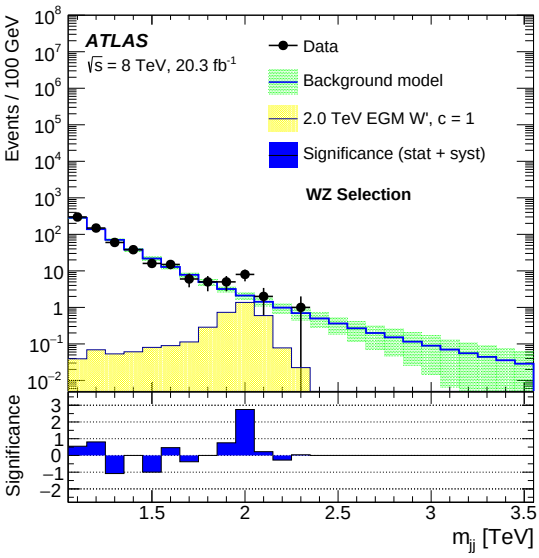


(a)

(b)



(c)



(d)

Figure 7.9: Variation of the observed m_{jj} spectrum selected with the WZ selection, and the expected spectrum with a 2.0 TeV W' signal when applying all tagging requirements except one. In (a), the WZ mass selection is not applied, in (b) the $\sqrt{y}$ requirement is not applied, in (c) n_{trk} requirement is not applied. In (d) all requirements for the WZ selection are applied. In each figure, the significance shown in the bottom panel for each bin is the difference between the data and the fit in units of the uncertainty on this difference.

7.3.1 Likelihood model

The starting point for the statistical analysis is the likelihood of the observation (data) given the expectation (background). The likelihood model for the observation is

$$\mathcal{L} = \prod_i P_{\text{pois}}(n^i_{\text{obs}}|n^i_{\text{exp}}) \times G(\alpha_{\text{p}_\text{T}}) \times G(\alpha_m) \times G(\sigma_{\text{r}_\text{E}}) \times \mathcal{N}(\theta) \tag{7.3}$$

where $P_{\text{pois}}(n^i_{\text{obs}}|n^i_{\text{exp}})$ is the Poisson probability to observe n^i_{obs} events if n^i_{exp} events are expected, $G(\alpha_{\text{p}_\text{T}})$, $G(\alpha_m)$, and $G(\sigma_{\text{r}_\text{E}})$, are the pdfs of the nuisance parameters modeling the systematic uncertainties related to the shape of the signal, and $\mathcal{N}$ is a log-normal distribution for the nuisance parameters, θ, modeling the systematic uncertainty on the signal normalisation. The expected number of events is the bin-wise sum of those expected for the signal and background: $\mathbf{n}_{\text{exp}} = \mathbf{n}_{\text{sig}} + \mathbf{n}_{\text{bg}}$. The number of expected background events in each dijet mass bin i, n^i_{bg}, is obtained by integrating dn/dx obtained from Eqn. 6.5 within that bin. Thus $\mathbf{n}_{\text{bg}}$ is a function of the dijet background parameters p_1, p_2, p_3. The number of expected signal events, $\mathbf{n}_{\text{sig}}$, is evaluated based on MC simulation assuming the cross section of the model under test multiplied by the signal strength, including the effects of the systematic uncertainties described in Section 6.6. Thus the expected number of signal events is a function of the signal strength μ and the nuisance parameters modeling the systematic uncertainties on the signal.

7.3.2 Search for a new diboson resonance

The local significance of the excesses observed in data with respect to the background expectation is quantified by estimating p_0 using a large ensemble of 500,000 background only pseudo-experiments[1]. Figure 7.18 shows the local p_0 as a function of the test model mass in the WZ, WW and ZZ selections. Since WZ, WW and ZZ selections are not orthogonal so the observed p_0 are not independent between the three selections. The observed excess of events in the region around 2.0 TeV in Figures 7.1a, 7.1b and 7.1c leads to small p_0 values in the region around 2.0 TeV. The most significant excess with respect to the background expectation is seen in the WZ selection with a p_0 value of 2.9×10^{-4}, equivalent to a local significance of 3.4 σ. The global p_0 is computed by obtaining the probability for a background-only experiment to observe a test statistics as significant as the most significant excess in the search. The computation uses 100,000 pseudo experiments accounting for events selected in more than one signal region and considering the full mass range of the search (1.3-3.0 TeV). In these pseudo experiments the events are generated according to the background fit for the inclusive case without discriminating between the signal selections (Fig. 7.19). For each generated event, the invariant mass of the two leading jets is generated according to the expected jet mass distribution in a particular range on m_{jj}. Then the two leading jet masses are used to classify the events as WZ, WW, or ZZ. Finally, the probability to get locally significant excesses in any of the three signal regions is evaluated and it leads to a global significance of 2.5 σ. Since this level of global significance is not large enough to claim a discovery, exclusion limits have been set on the models discussed in this thesis. The results are presented in the next section.

[1]The pseudo-experiments are performed generating a large set of background-only distributions from the observed background fit, in order to check what is the probability to get a statistical fluctuation of the background similar to the excess observed in data

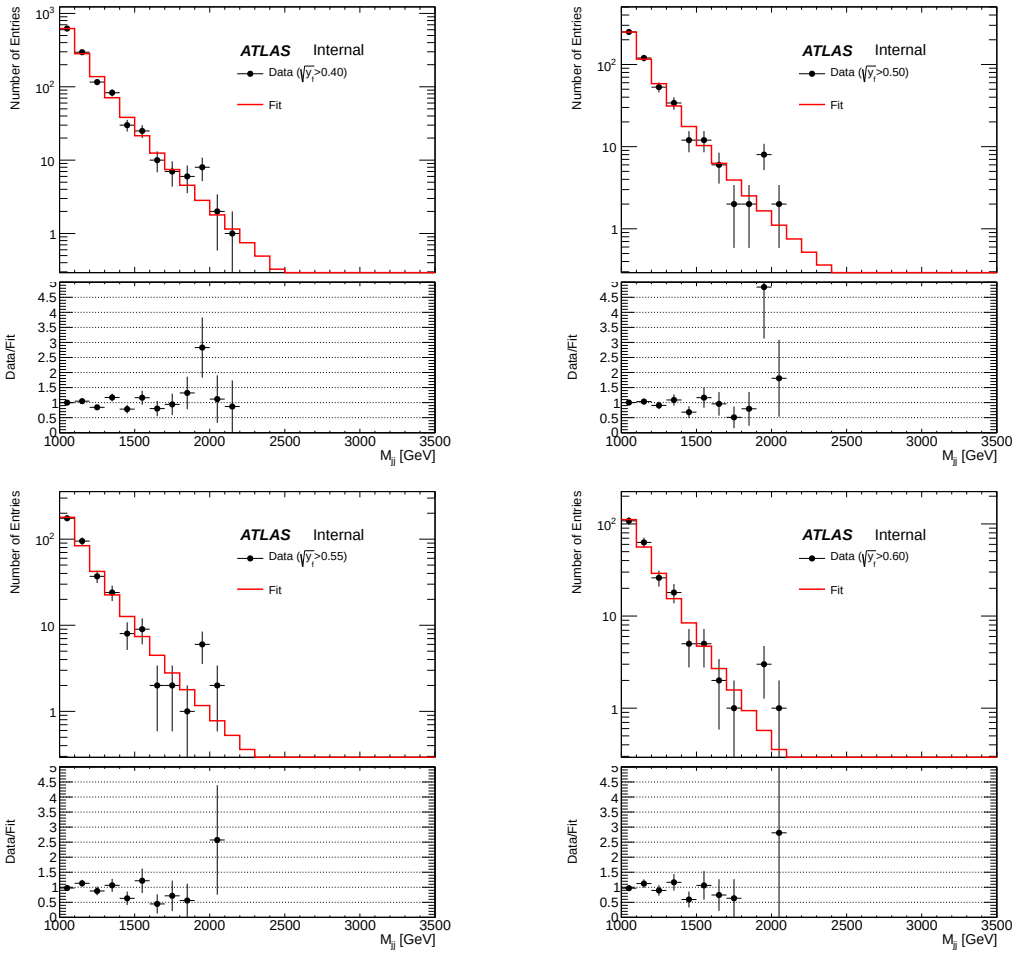


Figure 7.10: The dijet mass distribution formed with after applying different $\sqrt{y_f}$ cuts from 0.40 to 0.6.

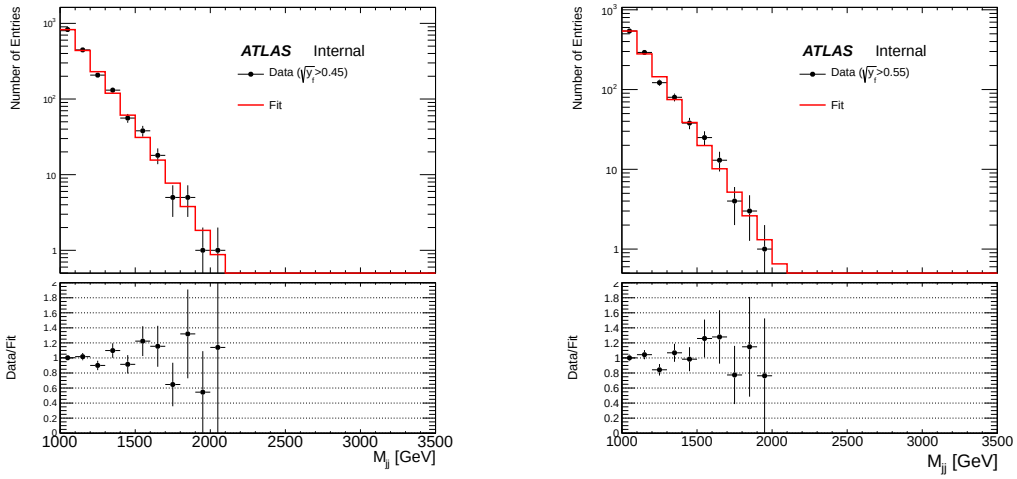


Figure 7.11: The dijet mass distribution formed with one jet in the wide mass window, $60 < m < 110$ GeV and one jet in the mass sideband, $40 < m < 60$ GeV, when the standard $\sqrt{y_f} < 0.45$ cut is applied and for a harsher cut of $\sqrt{y_f} < 0.55$

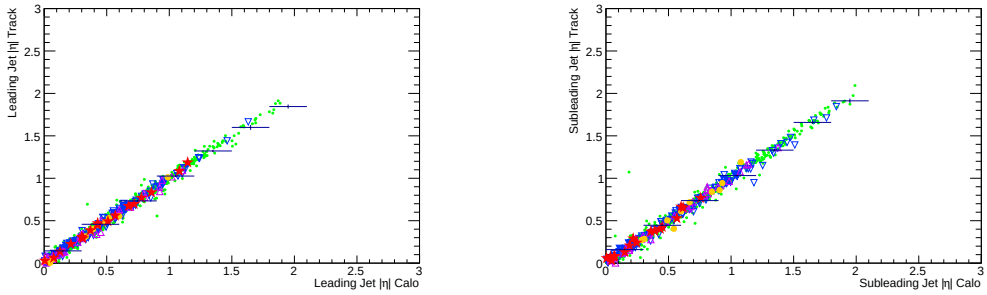


Figure 7.12: Comparison of calorimeter and tracker measurements of η coordinate for both leading (left) and subleading (right) jets. The following color convention has been used: $1.0 \leqslant m_{jj} < 1.2$ TeV (green), $1.2 \leqslant m_{jj} < 1.4$ TeV (blue), $1.4 \leqslant m_{jj} < 1.6$ TeV (violet), $1.6 \leqslant m_{jj} < 1.8$ TeV (orange), $m_{jj} \geqslant 1.8$ TeV (red).

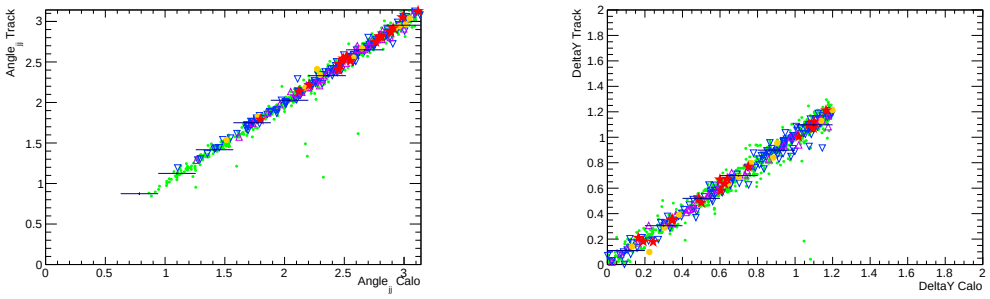


Figure 7.13: Comparison of calorimeter (X axis) and tracker measurements (Y axis) of the opening angle (left) and the difference in rapidity (right) of the two leading jets. The following color convention has been used: $1.0 \leqslant m_{jj} < 1.2$ TeV (green), $1.2 \leqslant m_{jj} < 1.4$ TeV (blue), $1.4 \leqslant m_{jj} < 1.6$ TeV (violet), $1.6 \leqslant m_{jj} < 1.8$ TeV (orange), $m_{jj} \geqslant 1.8$ TeV (red).

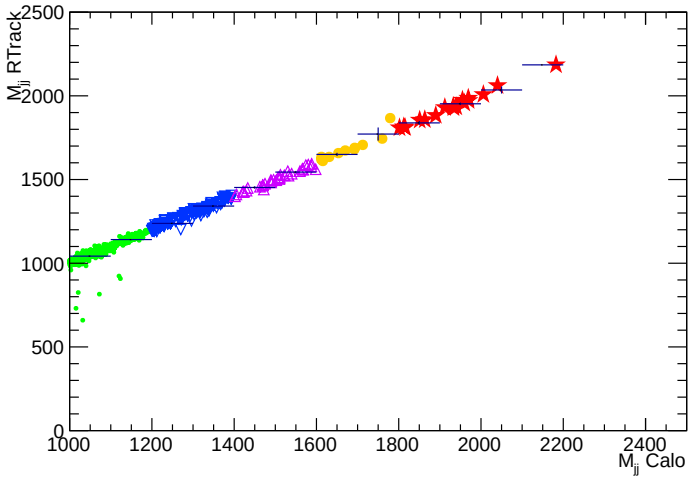


Figure 7.14: Comparison of calorimeter (X axis) and mixed four-vector measurements (Y axis) of m_{jj}. The following color convention has been used: $1.0 \leqslant m_{jj} < 1.2$ TeV (green), $1.2 \leqslant m_{jj} < 1.4$ TeV (blue), $1.4 \leqslant m_{jj} < 1.6$ TeV (violet), $1.6 \leqslant m_{jj} < 1.8$ TeV (orange), $m_{jj} \geqslant 1.8$ TeV (red).

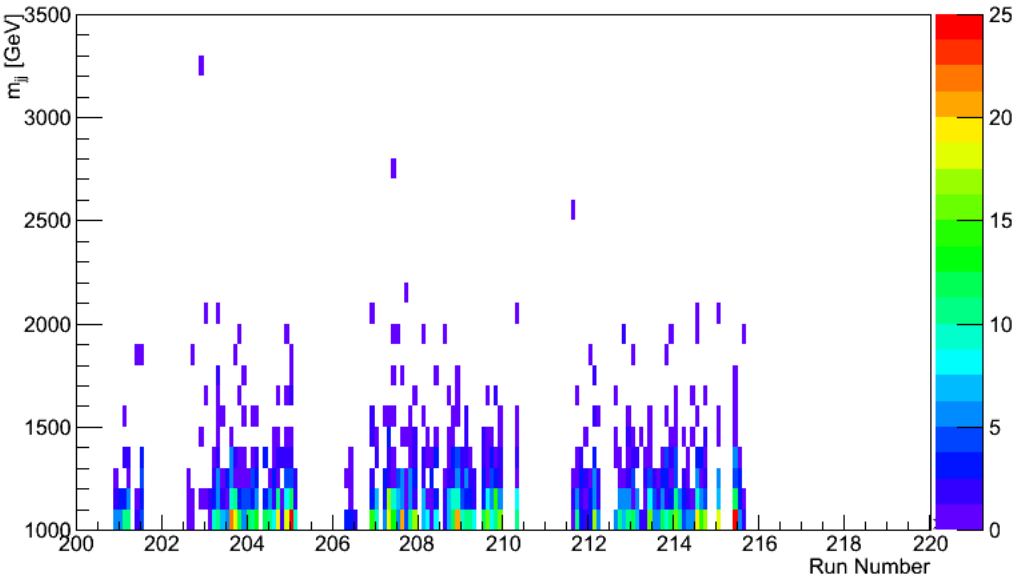


Figure 7.15: Di-jet invariant mass for selected boson-tagged (jet mass window 60-110 GeV) events as a function of the data taking run number.

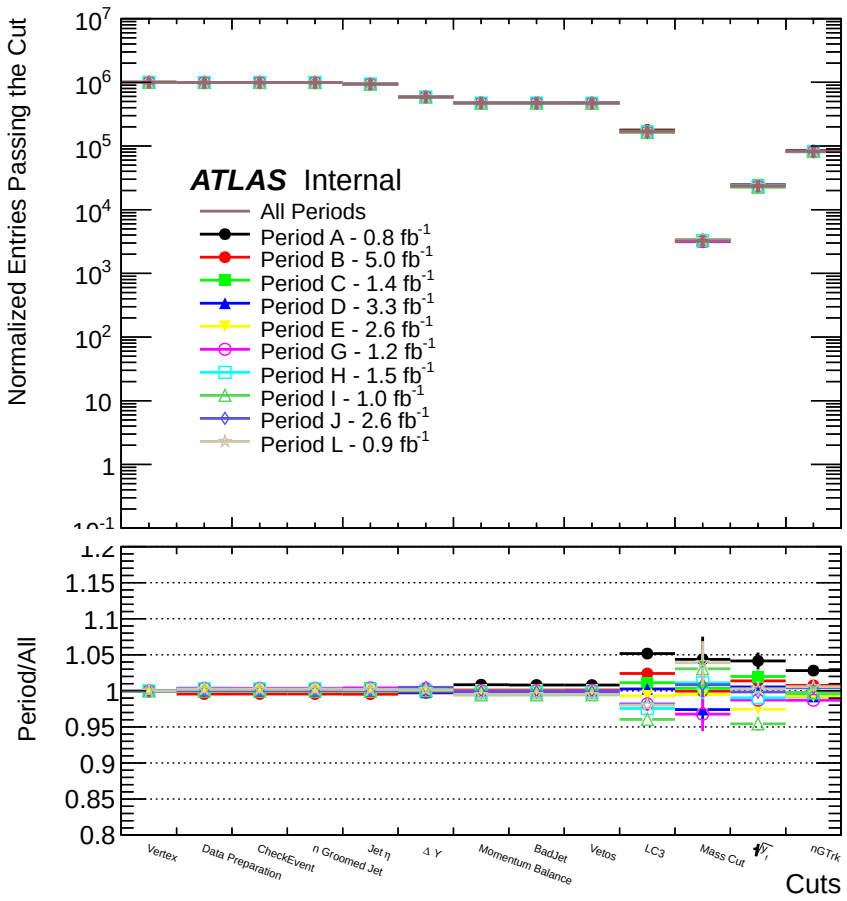


Figure 7.16: Cut-flow histograms for different data taking periods (top). The number of entries are normalized to the number of events passing the trigger selection for each period. The bottom plot shows the ratios between the cut-flow histogram for the different periods and the full data one.

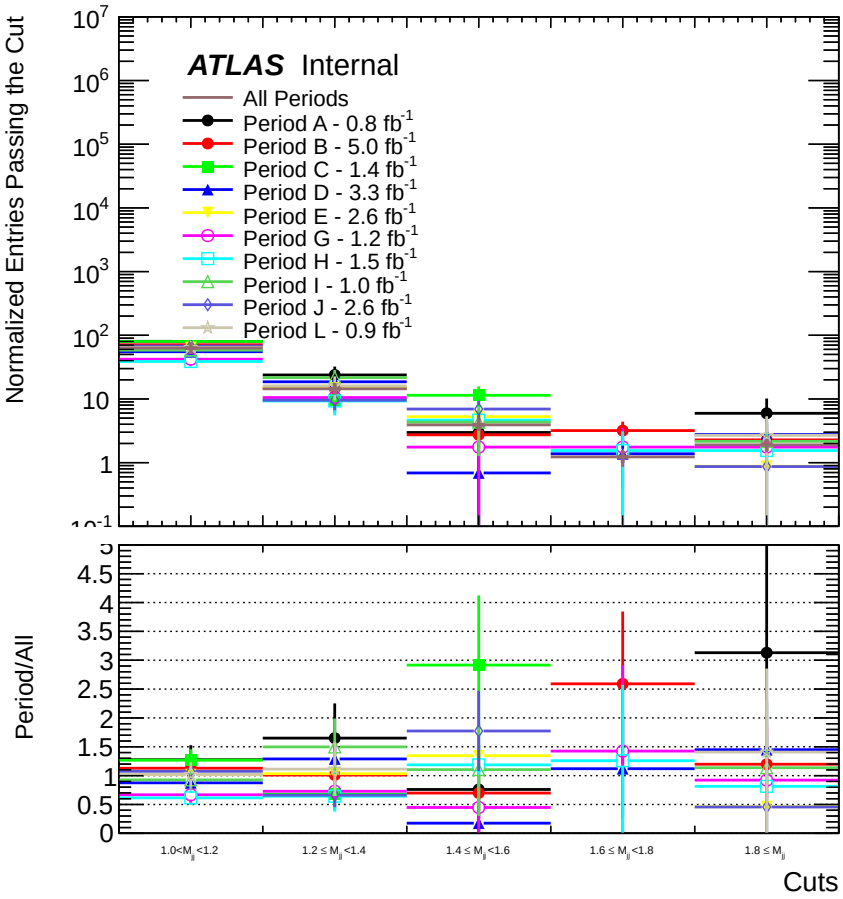


Figure 7.17: Number of selected and boson-tagged events in 5 m_{jj} bins as function of the data taking period. The number of entries are normalized to the number of events passing the trigger selection for each period. The bottom plot shows the ratios between the histogram for the different periods and the full data one.

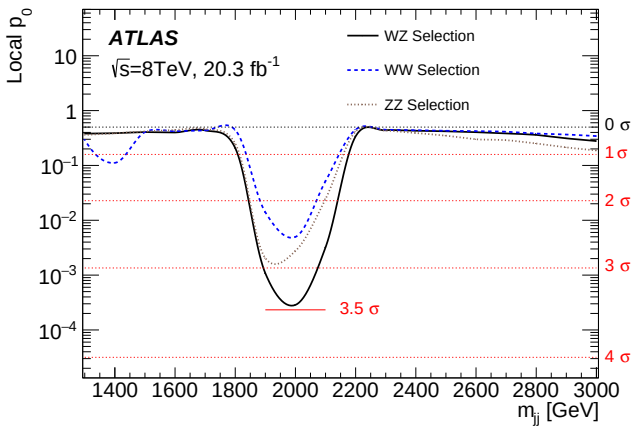


Figure 7.18: Local p_0 values for the WZ, WW and ZZ mass-window selections as a function of dijet mass. Since the WZ, WW and ZZ selections are not orthogonal, the observed p_0 are not independent between the three selections.

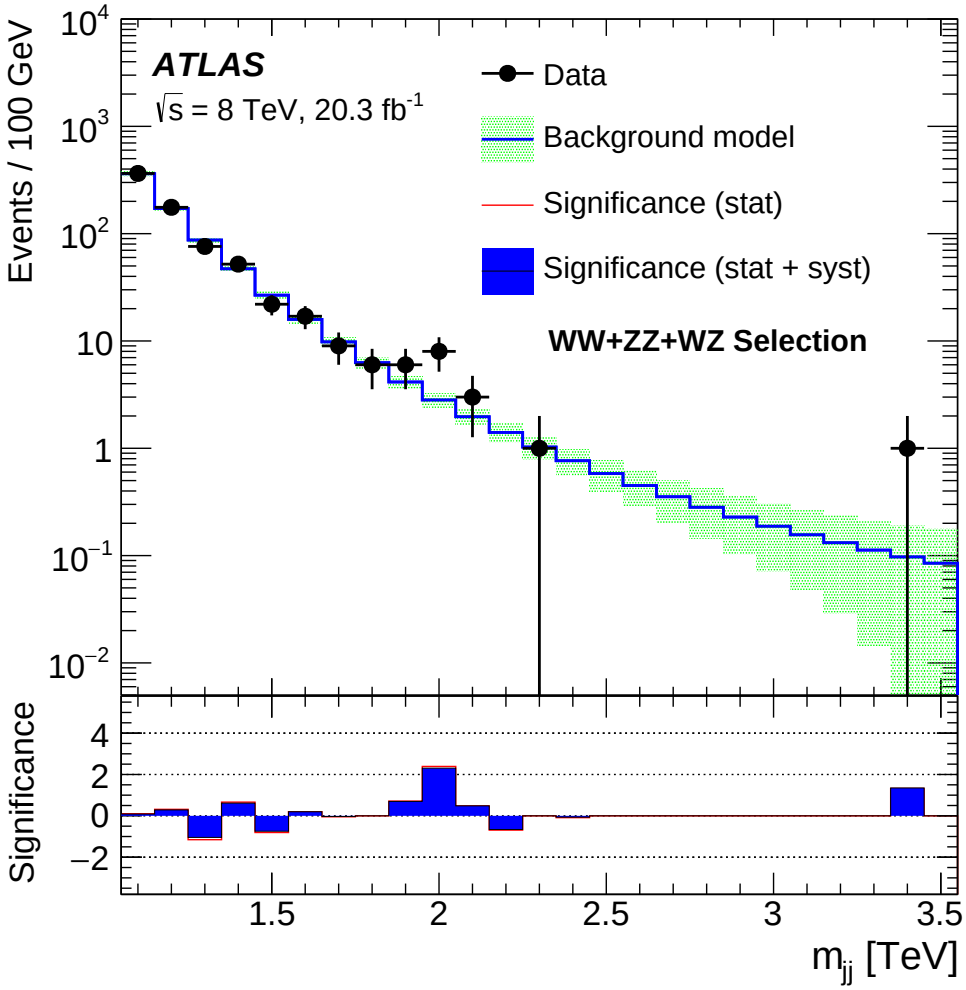


Figure 7.19: Background fit to the dijet mass distributions passing cuts any of the three analysis selections (WZ, WW, and ZZ).

7.3.3 Exclusion limits on new diboson resonances

Limits on the production cross section times branching ratio of massive resonances are set in each diboson channel as a function of the resonance mass, using the EGM W' as a benchmark for the WZ selection, and the bulk G_{RS} for the WW and ZZ selections. The observed excess of events in Figures 7.1a, 7.1b and 7.1c leads to observed limits which are weaker than expected for masses near 2 TeV. Figure 7.20a, shows the observed 95% CL upper limits on the cross section times branching ratio on the EGM $W' \to WZ$ hypotheses as function of the W' mass. EGM $W' \to WZ$ are excluded masses between 1.3 and 1.5 TeV with 95% CL. Figures 7.20b and 7.20c show the observed 95% CL upper limits on the cross section times branching ratio for the bulk $G_{RS} \to WW$ and ZZ, respectively. The cross section times branching ratio for excited graviton production with the model parameters described in Section 6.2 is too low to be excluded with the sensitivity of the dataset analyzed in this thesis.

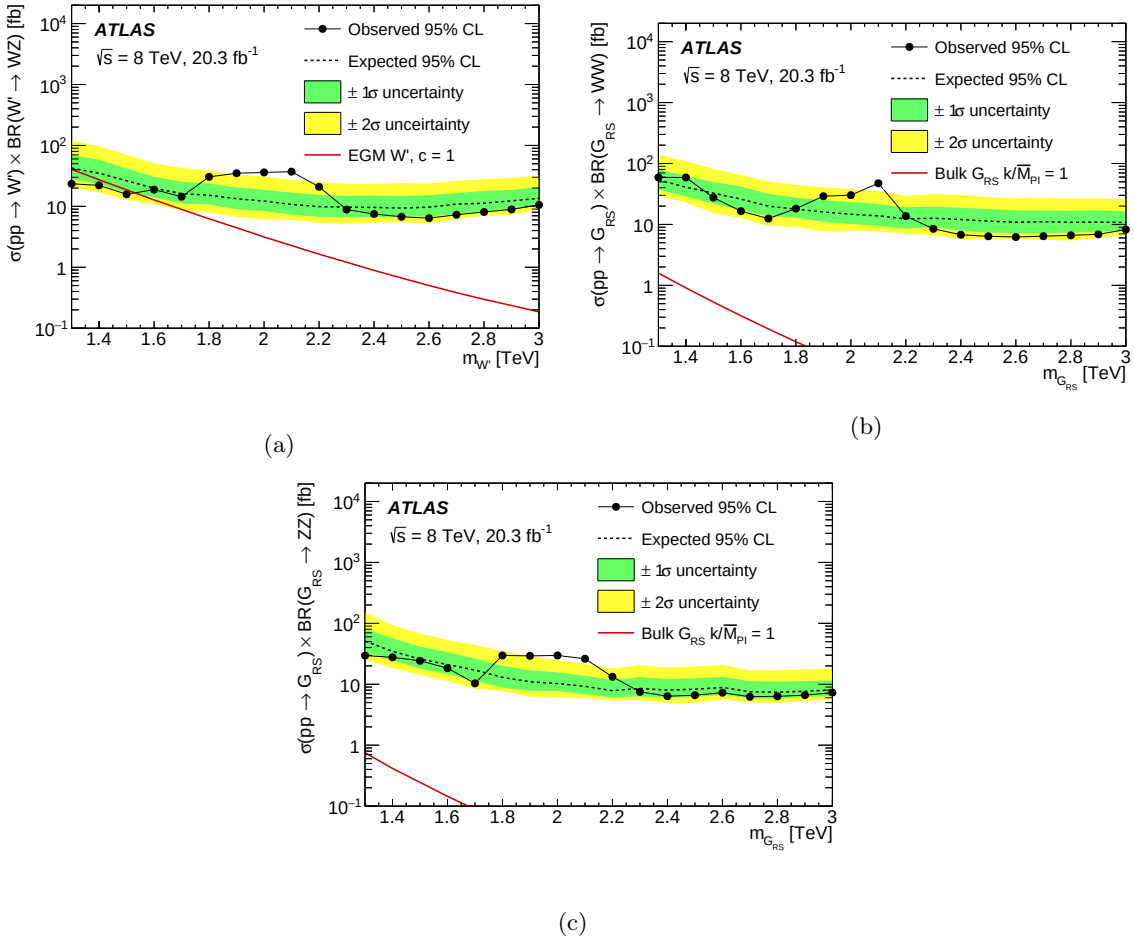


Figure 7.20: Upper limits, at 95% C.L., on the section times branching ratio limits for (a) the WZ window selection as a function of $m_{W'}$, and for (b) the WW window selection and (c) the ZZ window selections as a function of $m_{G_{RS}}$. The solid red line in each figure displays the predicted cross section for the W' or G_{RS} model as a function of the resonance mass.

W'(WZ)			95% CL Limits [fb]					
m [GeV]	σB_{qqqq} [fb]	Γ [GeV]	obs	exp	-2 σ	-1 σ	+1 σ	+2 σ
1300	40.5059	46.66	23.47	42.43	20.51	27.61	69.49	120.84
1400	27.2790	50.30	22.08	35.01	16.96	23.41	58.15	96.95
1500	18.6721	53.99	15.79	26.28	12.82	16.81	41.96	70.12
1600	12.7655	57.60	18.82	19.84	10.97	13.99	30.62	50.61
1700	8.9308	61.32	14.40	16.19	8.55	11.18	25.36	42.78
1800	6.2346	65.00	30.47	15.22	7.95	10.67	23.64	38.51
1900	4.4633	68.66	34.96	13.35	6.69	8.96	20.40	33.37
2000	3.1448	72.30	36.04	12.18	6.12	8.39	18.71	30.84
2100	2.2750	76.00	37.12	10.74	6.02	7.31	16.77	27.06
2200	1.6403	79.60	20.92	9.88	5.36	6.73	15.40	25.36
2300	1.2126	83.34	8.87	9.72	5.24	6.70	14.89	23.30
2400	0.8888	87.00	7.49	9.61	5.40	6.68	15.05	23.37
2500	0.6664	90.68	6.75	9.37	5.47	6.76	14.83	23.97
2600	0.5018	94.30	6.40	9.69	5.25	6.80	15.22	23.60
2700	0.3833	98.03	7.28	10.88	6.24	7.49	16.53	26.11
2800	0.2973	101.70	8.08	11.26	6.96	8.24	17.88	27.75
2900	0.2358	105.40	8.91	12.32	7.27	8.86	18.61	29.37
3000	0.1856	109.00	10.58	13.54	8.86	10.12	20.71	32.11

Table 7.4: Observed and expected limits on the EGM W' models in the WZ selection.

G(WW)			95% CL Limits [fb]					
m [GeV]	σB_{qqqq} [fb]	Γ [GeV]	obs	exp	-2 σ	-1 σ	+1 σ	+2 σ
1300	1.5887	69.00	59.16	53.46	29.46	37.36	80.36	139.90
1400	0.9081	76.00	59.00	40.90	23.03	28.67	62.92	109.16
1500	0.5318	83.00	27.57	32.60	15.39	21.85	49.69	81.17
1600	0.3173	90.00	16.53	26.04	12.63	17.18	41.93	64.89
1700	0.1924	96.00	12.47	19.92	9.44	13.56	31.84	50.82
1800	0.1188	103.00	18.18	17.83	9.15	12.96	27.57	44.35
1900	0.0744	109.00	29.01	15.99	7.68	10.97	24.93	40.50
2000	0.0470	116.00	30.23	14.68	7.87	10.18	23.25	37.90
2100	0.0300	123.00	47.39	13.83	7.72	9.46	21.67	35.14
2200	0.0194	129.00	13.70	12.34	7.02	8.53	19.19	29.46
2300	0.0126	136.00	8.48	12.63	6.67	9.01	19.53	31.31
2400	0.0083	142.00	6.76	12.06	5.78	7.86	18.78	30.54
2500	0.0055	149.00	6.39	11.32	5.68	7.36	17.92	28.18
2600	0.0036	155.00	6.21	10.87	5.64	7.19	16.99	26.76
2700	0.0024	161.00	6.41	11.07	5.58	7.10	16.94	26.91
2800	0.0016	168.00	6.62	10.84	5.41	7.41	16.72	26.60
2900	0.0011	174.00	6.87	10.93	6.02	7.50	17.07	26.76
3000	0.0008	180.00	8.24	10.76	6.61	8.25	16.39	25.85

Table 7.5: Observed and expected limits on the bulk G_{RS} models in the WW selection.

G(ZZ)			95% CL Limits [fb]					
m [GeV]	σB_{qqqq} [fb]	Γ [GeV]	obs	exp	-2 σ	-1 σ	+1 σ	+2 σ
1300	0.7532	69.00	29.55	50.91	25.19	33.72	81.26	152.30
1400	0.4155	76.00	27.71	34.78	18.59	24.56	56.97	93.29
1500	0.2456	83.00	24.30	25.93	14.29	18.15	41.45	67.91
1600	0.1445	89.00	18.39	21.13	11.35	14.58	33.67	53.01
1700	0.0888	96.00	10.36	17.07	8.46	11.42	26.44	44.35
1800	0.0557	103.00	29.71	12.98	7.74	8.78	20.04	34.98
1900	0.0338	109.00	29.24	10.98	6.28	7.83	16.94	28.79
2000	0.0213	116.00	29.73	10.26	6.07	7.72	15.68	24.91
2100	0.0137	123.00	26.23	9.25	5.93	6.87	13.70	21.35
2200	0.0088	129.00	13.27	7.84	5.41	6.09	11.76	18.20
2300	0.0058	136.00	7.50	8.55	5.51	6.30	13.15	20.57
2400	0.0037	142.00	6.37	8.02	4.91	6.01	12.16	18.93
2500	0.0025	149.00	6.61	8.30	4.96	6.15	12.57	19.61
2600	0.0019	155.00	7.27	8.85	5.64	6.73	13.23	20.72
2700	0.0011	161.00	6.24	7.53	5.00	5.88	11.26	17.05
2800	0.0008	168.00	6.34	7.41	4.96	5.75	11.00	16.88
2900	0.0005	174.00	6.62	7.66	5.28	6.22	11.27	17.50
3000	0.0003	181.00	7.31	8.05	5.69	6.42	11.52	18.21

Table 7.6: Observed and expected limits on the bulk G_{RS} models in the ZZ selection.

Chapter 8

Conclusions

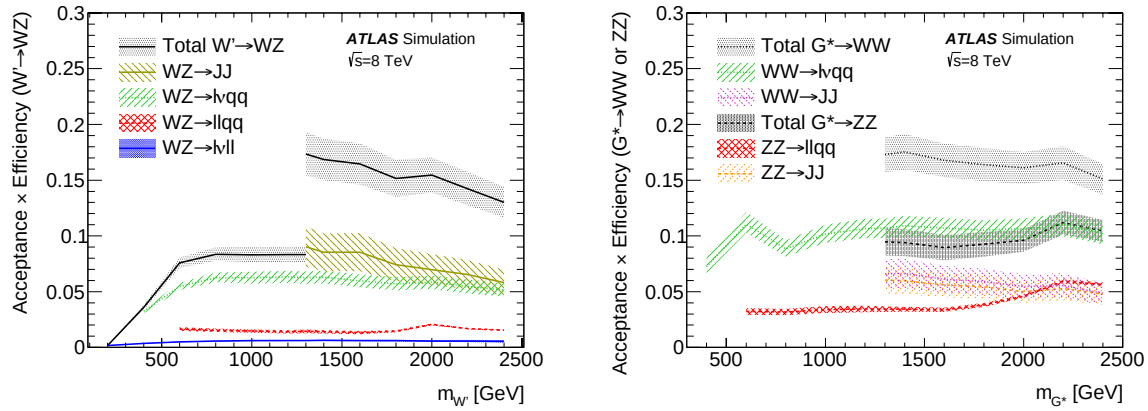
A search for massive particles decaying to WZ, WW, or ZZ has been performed using 20.3 fb^{-1} of $\sqrt{s} = 8$ TeV $proton-proton$ collision data collected at the LHC with the ATLAS detector in 2012. This is the first ATLAS search for resonant di-boson production in a fully hadronic final state. It strongly relies on the suppression of the dijet background with a substructure-based jet grooming and boson-tagging selection. The boson-tagging procedure includes different jet mass criteria to identify W and Z boson candidates and thus produces three overlapping sets of selected events for the searches in the WW, WZ, and ZZ decay channels. The most significant discrepancy with the background-only model occurs around 2 TeV in the WZ channel with a local significance of 3.4 σ and a global significance of 2.5 σ, when the entire mass range of the search in all three channels is taken into account.

Upper limits on the production cross section times branching ratio of massive resonances are set in each diboson channel as a function of the resonance mass, using an EGM $W' \to WZ$ as benchmark model for the WZ decay channel, and an excited bulk graviton G_{RS} to represent resonances decaying to WW and ZZ. A W' with EGM couplings and mass between 1.3 and 1.5 TeV is excluded at 95% CL.

8.1 ATLAS di-boson combination and CMS results

Using the same dataset, the ATLAS experiment has performed searches for heavy bosons decaying to di-boson final states in multiple decay channels. In particular, charged di-boson resonances decaying to WZ have been searched in the $\ell\nu\ell'\ell'$ (with $\ell = \mu, e$), $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ final states, while neutral di-boson resonances decaying to WW and ZZ have been searched in $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ final states. The results of these searches have set exclusion bounds on the production and decay of the EGM $W' \to WZ$ and excited bulk graviton G_{RS} decaying to WW and ZZ final states [81–83]. To improve the sensitivity to new di-boson resonances, the fully hadronic (JJ) analysis results have been combined with the results from the semi-leptonic analyses [84]. The four analyses implemented orthogonal event selections in order to be statistically independent. The correlations between the systematic uncertainties of the different channels have been considered while obtaining the final results. The acceptance times efficiency for the different analyses is shown as a function of W' and G_{RS} mass in Figure 8.1.

In the statistical procedure performed for each tested hypothesis, only the channels sensitive to

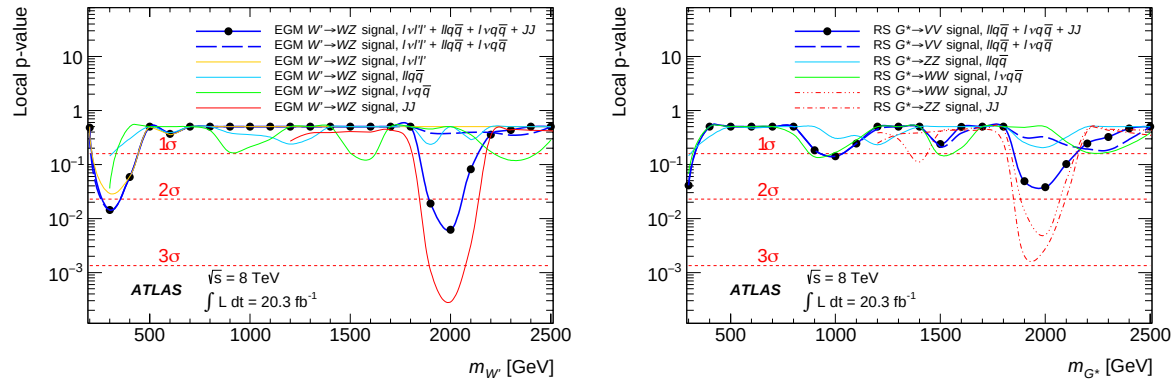


(a) (b)

Figure 8.1: Signal acceptance times efficiency for the different analyses entering the combination for (a) the EGM W' model and (b) for the bulk G_{RS} (G^* is used in place of G_{RS} in the plot) model. The branching ratio of the new resonance to di-bosons is included in the denominator. The error bands represent the combined statistical and systematic uncertainties [84].

that hypothesis are included in the combination. The analysis has been performed using the frequentist approach with a test statistics based on profile-likelihood ratio [79].

The combined p_0-value obtained in the search for EGM W' and G_{RS} as a function of the resonance mass for all the four analyses is shown Figure 8.2. In the same plots also p_0-value for individual channels are shown.



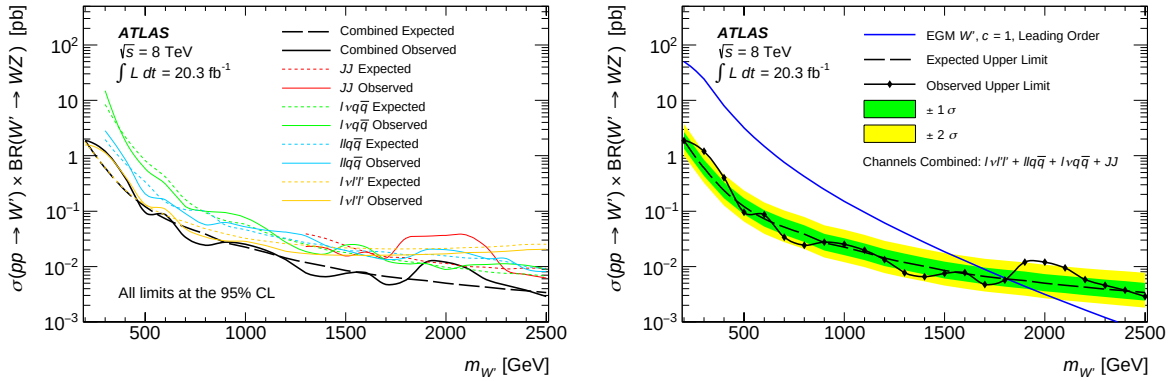
(a) (b)

Figure 8.2: The p_0-value for the individual and combined channels for (a) the EGM W' search in the $\ell\nu\ell'\ell'$ (with $\ell = \mu, e$), $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ and JJ channels and (b) the bulk G_{RS} (G^* is used in place of G_{RS} in the plot) search in the $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ and JJ channels [84].

For the full combination the largest deviation from the background-only expectation is found in the EGM W' search at around 2.0 TeV with a p_0-value corresponding to 2.5 σ. This value is smaller than the p_0-value of 3.4 σ observed in the JJ channel alone because the leptonic and semi-leptonic channels are more consistent with the background-only hypothesis. For the EGM W' benchmark the compatibility of the combined $\ell\nu\ell'\ell'$ (with $\ell = \mu, e$), $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ channels

with the JJ channel has been checked using a profile-likelihood-ratio test [84] and it was found to be at the level of 2.9 σ. When accounting for the probability for any of the four channels to fluctuate the compatibility is found to be at the level of 2.6 σ. The bulk G_{RS} interpretation shows better compatibility with the background-only hypothesis.

Combined upper limit on the EGM W' production cross section times its branching ratio to WZ at the 95% CL in the mass range from 300 GeV to 2.5 TeV is shown in Figure 8.3. In particular, Figure 8.3a shows observed and expected limits of the individual and combined channels, while Figure 8.3b shows combined observed and expected limits compared with the theoretical EGM W'. The resulting combined lower limit on the EGM W' mass is observed to be 1.81 TeV, with an expected limit of 1.81 TeV.

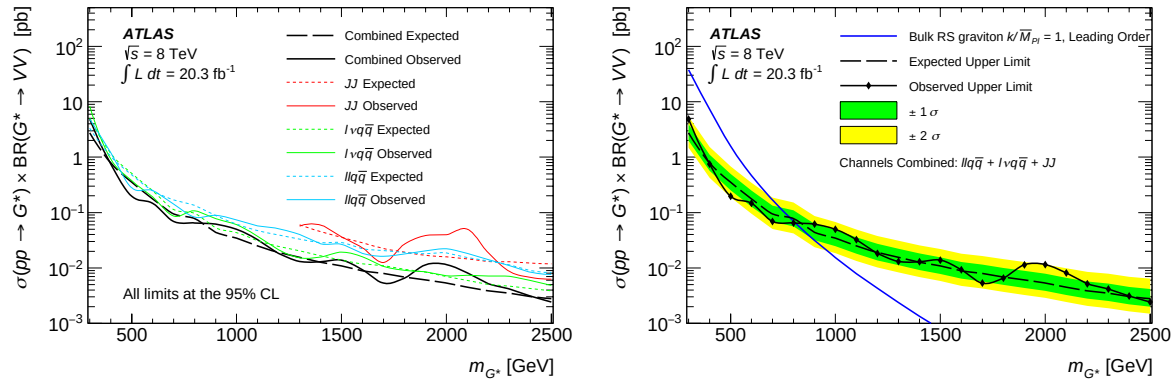


(a) (b)

Figure 8.3: The 95% CL limits on (a) the EGM W' using the $\ell\nu\ell'\ell'$, $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ and JJ channels and their combination, and (b) the combined 95% CL limit with the green (yellow) bands representing the 1 σ (2 σ) intervals of the expected limit including statistical and systematic uncertainties [84].

Figure 8.4 shows observed and expected upper limits at the 95% CL on the bulk G_{RS} production cross section times its branching ratio to WW and ZZ are shown in the mass range from 200 GeV to 2.5 TeV. The combined, lower mass limit for the bulk G_{RS} is 810 GeV, with an expected limit of 790 GeV.

The combination improves both the cross-section limits and the mass limits for EGM W' and bulk G_{RS} production over the most stringent limits of the individual analyses. For both the signal interpretations the most stringent lower mass limit from the individual $\ell\nu\ell'\ell'$, $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ and JJ channels provided from the $\ell\nu q\bar{q}$ channel. $\ell\nu q\bar{q}$ channel has signal efficiency comparable to the JJ analysis in when searching for EGM W', while higher signal efficiency for bulk G_{RS} decaying to WW final state. The background rejection in semi-leptonic channel is higher than in the JJ. Dominant systematics for $\ell\nu q\bar{q}$ analysis are due to the modeling of the Z+jets background in the case of G_{RS} limit. In the intermediate and mass region up to 1.5 TeV the uncertainty on the normalization of the W+jets background is the dominant systematic uncertainty and it produces a 20% to 30% degradation of the EGM W' limit and 25% to 55% degradation of the G_{RS} limit depending on the mass point, while in the high mass region up to 2 TeV the shape uncertainty on the W+jets background dominates with a degradation of around 25% for the EGM W' limit and 35% for the G_{RS} limit.



(a) (b)

Figure 8.4: The 95% CL limits on (a) the bulk G_{RS} using the $\ell\nu\ell'\ell'$, $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$ and JJ channels and their combination, and (b) the combined 95% CL limit with the green (yellow) bands representing the 1 σ (2 σ) intervals of the expected limit including statistical and systematic uncertainties [84].

Using the same integrated luminosity, CMS collaboration has performed searches for diboson resonances in the semi-leptonic channels [85] and the fully hadronic channel [86], requiring that both $q\bar{q}$-pairs from W/Z-boson decays are reconstructed as a single jet and using jet substructure techniques to improve sensitivity. From these, the fully hadronic channel is able to exclude W' boson masses up to 1.7 TeV and bulk Randall-Sundrum (with $k/\bar{M}_{Pl} = 0.1$) graviton masses up to 1.2 TeV.

8.2 The ATLAS fully hadronic di-boson resonance search in LHC Run 2

The increase of center of mass energy and instantaneous luminosity in the next LHC data-taking period, gives rise to higher particle multiplicity than those of the 8 TeV data taking, in which the jet reconstruction and the identification of hadronic boosted topologies is more challenging. The implementation of new jet filtering and boson tagging techniques, exploring the full potential of the detector, is crucial. Multivariate classifiers and machine learning techniques are already showing promising results in tagging jets from boson decay and they increase the sensitivity of this kind of searches, exploiting the full potential of the big set of observables defined when studying the substructure of jets.

These technical improvements, together with increase of the cross-section for the expected signals due to the large increase of parton luminosity at high energies, provide enough sensitivity to understand with less than 10 fb^{-1} if the excess observed in at 2 TeV was just a statistical fluctuation, or if it was produced by a new physics process.

An "early" 13 TeV analysis has been published in December 2015 by the ATLAS collaboration [87]. The analysis has been performed on 3.2 fb^{-1} of $proton - proton$ collision data at 13 TeV and in the low mass region it provided comparable sensitivity with respect to the 8 TeV analysis presented in this thesis. The collected integrated luminosity was not enough to constrain the data-driven background in the high-mass region and the analysis on 2016 dataset

is expected to exceed the 8 TeV sensitivity in the full invariant mass range up to 3 TeV.

Appendix A

Particle-level calibration for BDRS-A jets

In order to scale the measured jet energy (mass) back to the energy (mass) of the particle–level jet, a jet energy (mass) calibration is derived. One of the main aims of this calibration is to correct for the non-uniformity of the jet energy and mass response as a function of the jet direction in the detector. The calibration corrects reconstructed jets based on the numerical inversion technique using MC simulations, with the resulting jets referred to as LCW+JES jets. Reconstructed jets are matched to a particle–level jets within $\Delta R < 0.3$.

The jet response is defined as

$$\mathcal{R}_E^{\text{jet}}(\eta_{\text{det}}) = E_{\text{reco}}^{\text{jet}}/E_{\text{truth}}^{\text{jet}}, \quad \mathcal{R}_m^{\text{jet}}(\eta_{\text{det}}) = m_{\text{reco}}^{\text{jet}}/m_{\text{truth}}^{\text{jet}} \ .$$

The LCW calibrated jet is corrected as follows:

$$E_{\text{LCW+JES}}^{\text{jet}} = \frac{E_{\text{LCW}}^{\text{jet}}}{\mathcal{F}_{\text{calib}}^{E}(E_{\text{LCW}}^{\text{jet}})}, \ m_{\text{LCW+JES}}^{\text{jet}} = \frac{m_{\text{LCW}}^{\text{jet}}}{\mathcal{F}_{\text{calib}}^{m}(E_{\text{LCW}}^{\text{jet}} * \text{JES})} \ ,$$

where $E_{\text{LCW}}^{\text{jet}}$ and $m_{\text{LCW}}^{\text{jet}}$ are the energy and mass of the jet at the LCW scale, respectively and $\mathcal{F}_{\text{calib}}^{E}(E_{\text{LCW}}^{\text{jet}})$ is the calibration function. The calibration function $\mathcal{F}$ is derived in different η_{det} and E_{truth} bins. For each η_{det} and E_{truth}-bin, the mean jet energy (mass) response $\left\langle \mathcal{R}_E^{\text{jet}} \right\rangle$ ($\left\langle \mathcal{R}_m^{\text{jet}} \right\rangle$) is derived as the peak position of the Gaussian fit to the response distribution and the mean reconstructed energy $\left\langle E_{\text{LCW}}^{\text{jet}} \right\rangle$ is calculated. The inverse of the mean jet response as a function of the mean reconstructed energy is then parametrized as:

$$\mathcal{F}_{\text{calib}}(E_{\text{reco}}^{\text{jet}}) = \sum_{i=0}^{N_{\text{max}}} a_i \cdot \left(\ln(E_{\text{LCW}}^{\text{jet}})\right)^i \ ,$$

where a_i are free parameters of the fit and N_{max} is chosen between one and six depending on the goodness of the fit.

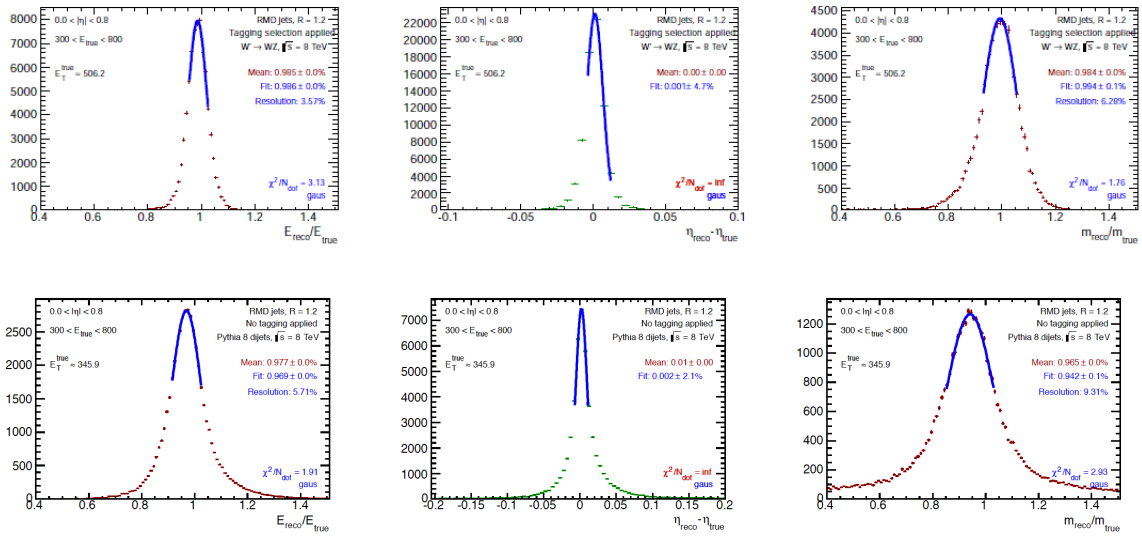


Figure A.1: Jet energy response (left), $\Delta\eta$ (middle) and jet mass response (right) for (top) W/Z - jets and (bottom) QCD jets

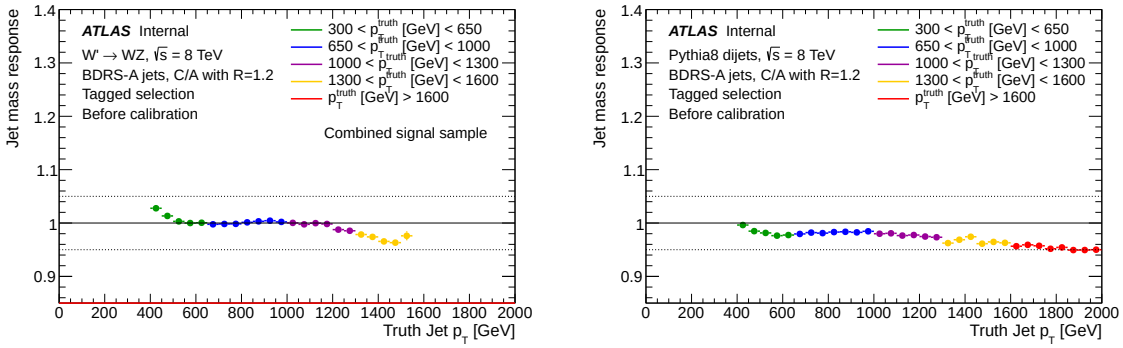


Figure A.2: Jet mass response as a function of the truth jet p_T for (top) W/Z - jets and (bottom) QCD jets

A.0.1 Details of the calibration

The calibration of the jet energy and mass is derived based on QCD MC jets. Although the jet energy and mass scale for QCD and signal jets (jets containing a boson) is significantly different, as shown in Figure A.1, the calibration is derived based on QCD jets which would comprise the majority of jets measured in search for new physics. The overcorrection of the jet mass and energy response after applying this QCD based calibration is fairly constant over all energy and η bins and it can be easily taken into account when optimizing the tagging selection. The comparison of the jet mass response for QCD and W/Z jets as a function of the truth jet p_T and η is shown in Figure A.2 and Figure A.3, respectively. While the mass response as a function of the η is centered around its expected value of 1.0, the QCD mass response is significantly lower.

The jet calibration is derived from the Pythia 8 dijet Monte Carlo samples (Sec. 6.2). Since

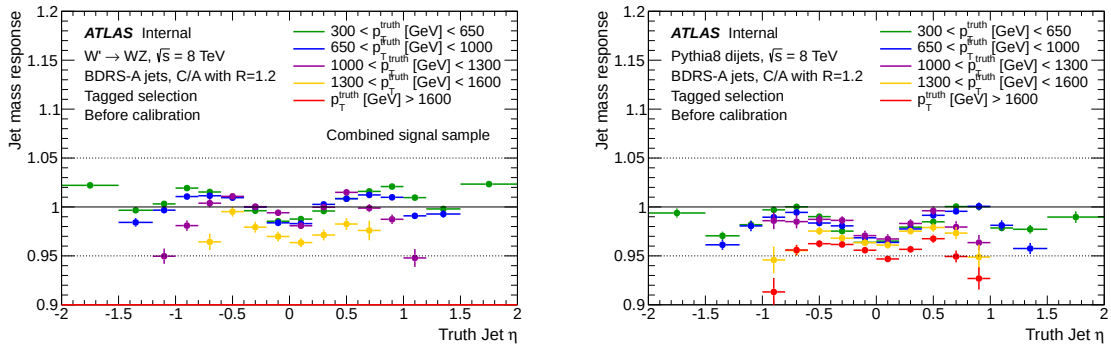


Figure A.3: Jet mass response as a function of the truth jet η in different truth jet p_T bins for (top) W/Z - jets and (bottom) QCD jets

the Pythia MC samples are split in leading truth jet p_T, each sample is reweighed according to the cross-section and the MC event weight in order to obtain a smooth p_T truth distribution. Due to the low statistics of the QCD samples especially at high jet energies, only five different truth energy bins:

- $300 < E_{\text{truth}} < 650$ GeV,
- $650 < E_{\text{truth}} < 1000$ GeV,
- $1000 < E_{\text{truth}} < 1300$ GeV,
- $1300 < E_{\text{truth}} < 1600$ GeV,
- $E_{\text{truth}} > 1600$ GeV

were used for the calibration. In order to reduce the η dependence, see Fig. A.3, the calibration is derived in the following η-bins: -2.0, -1.5, -1.2, -1.0, -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.5, 2.0. The jet energy response as a function of the jet η before and after applying the jet energy correction is shown in Fig. A.4. Before the energy correction, a clear shift of the energy response as well as a detector dependence is visible. After the jet energy correction, the energy response for different energy values is flat in the different η-bins and centered around its expected value. The jet mass response as a function of the jet η before and after applying the jet mass correction is shown in Fig. A.5. Fig. A.6 depicts the $\eta_{\text{reco}} - \eta_{\text{truth}}$ distribution before and after applying the η correction. The p_T closure after applying the jet energy, mass and η correction can be found in Fig. A.7.

The jet energy response as a function of the truth jet p_T and the truth jet η is shown in Fig. A.8. The corresponding plots for the jet mass response are shown in Fig. A.9. The mass response before applying a calibration is in better agreement with the expectation than after applying the correction. Even though 15 η-bins have been chosen, an η-dependence is still visible for signal jets as well as QCD jets. This is caused by then non-Gaussian behavior of the jet mass response which makes a calibration more difficult than for the energy. Overall the correction for the jet mass is of the order of 5% going up to 13% at high energies due to the lack of statistics. At these high energies, only one high energy bin with $E_{\text{truth}} > 1600$ GeV is

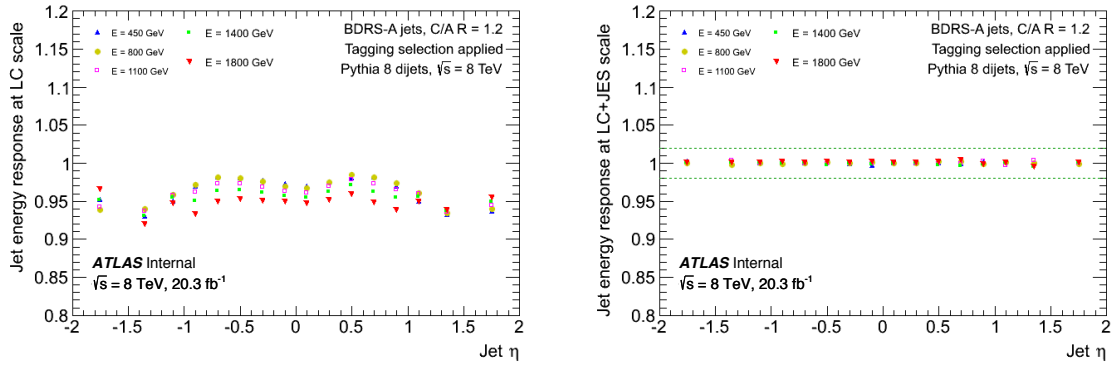


Figure A.4: Jet energy response before (left) and after (right) the jet energy calibration as a function of the jet η for different energies.

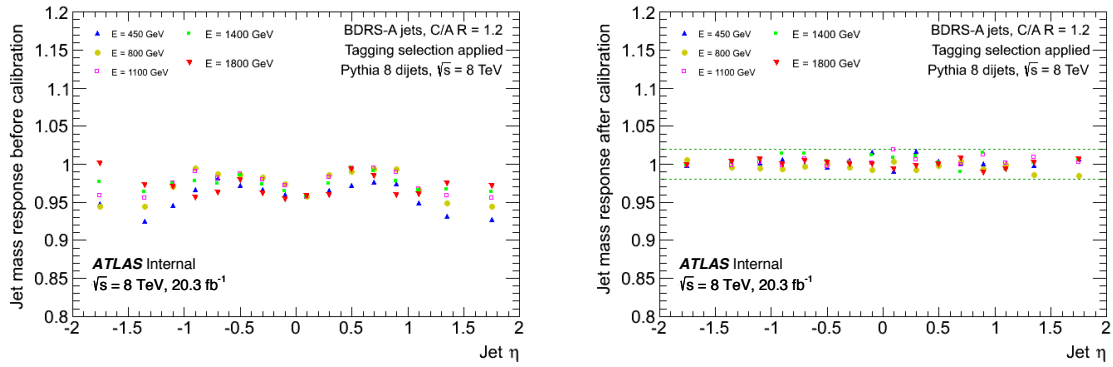


Figure A.5: Jet mass response before (left) and after (right) the jet mass calibration as a function of the jet η for different energies.

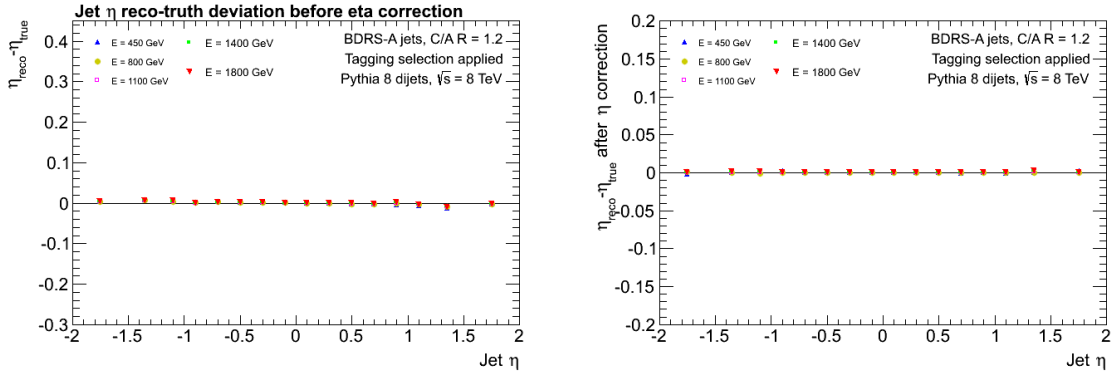


Figure A.6: $\eta_{\text{reco}} - \eta_{\text{truth}}$ before (left) and after (right) the jet mass calibration as a function of the jet η for different energies.

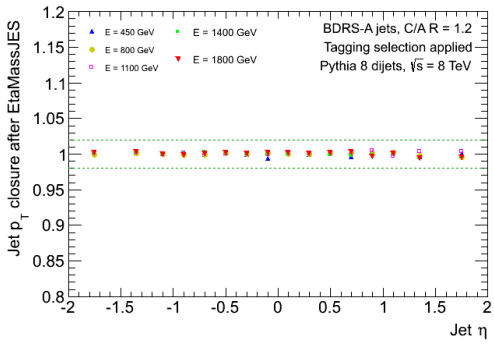


Figure A.7: p_T closure after the jet energy, mass and η calibration as a function of the jet η for different energies.

used but taking into account that the analysis does not select many jets above 2000 GeV, the effect is negligible. The jet energy and mass calibration is symmetric in η and no η and energy bins show an unexpected behavior.

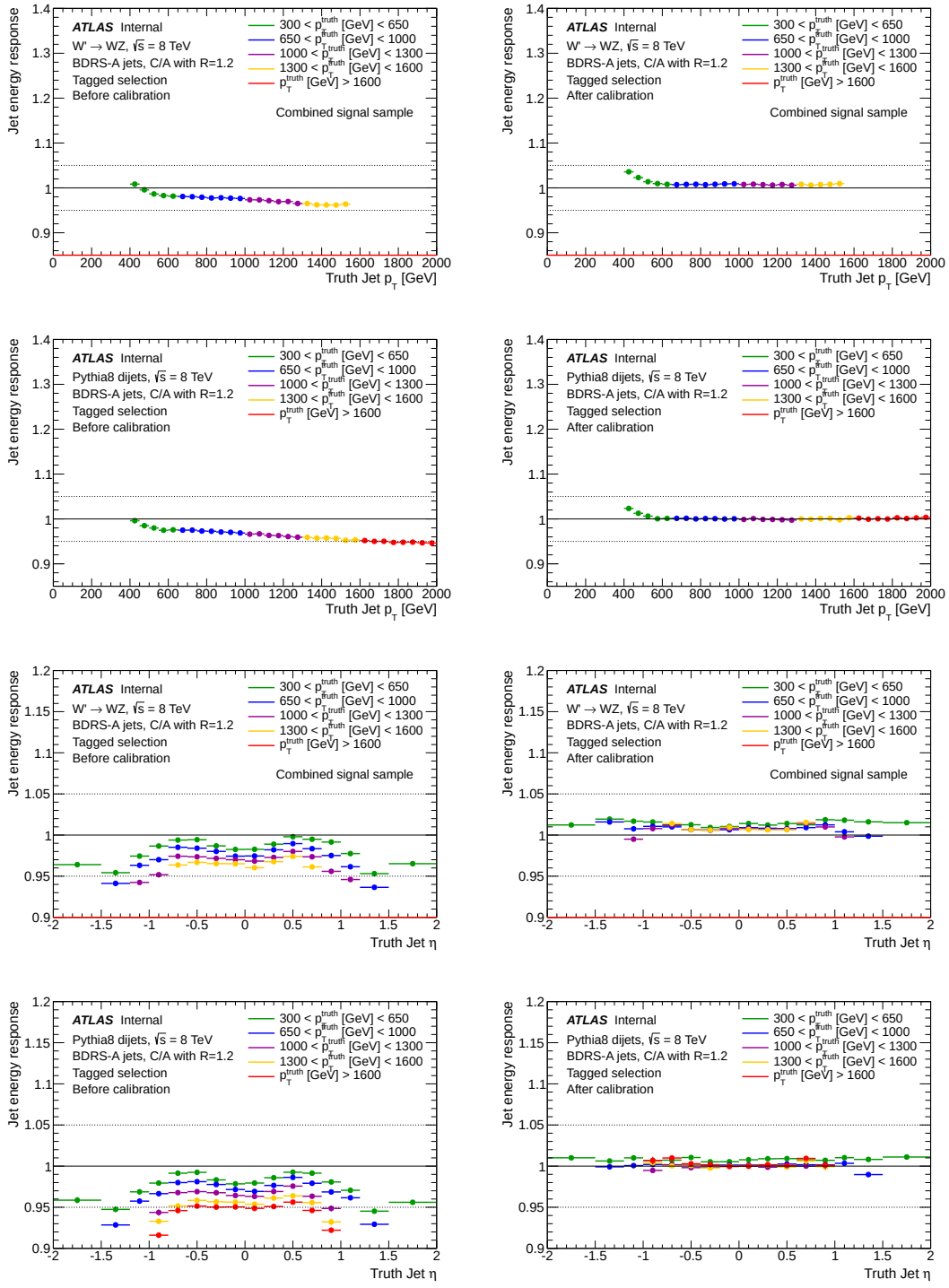


Figure A.8: Jet energy response as a function of the truth jet p_T and as well as the truth jet η for signal and QCD jets without (left) and with (right) the derived calibration applied. The jet energy response is shown in different truth p_T bins.

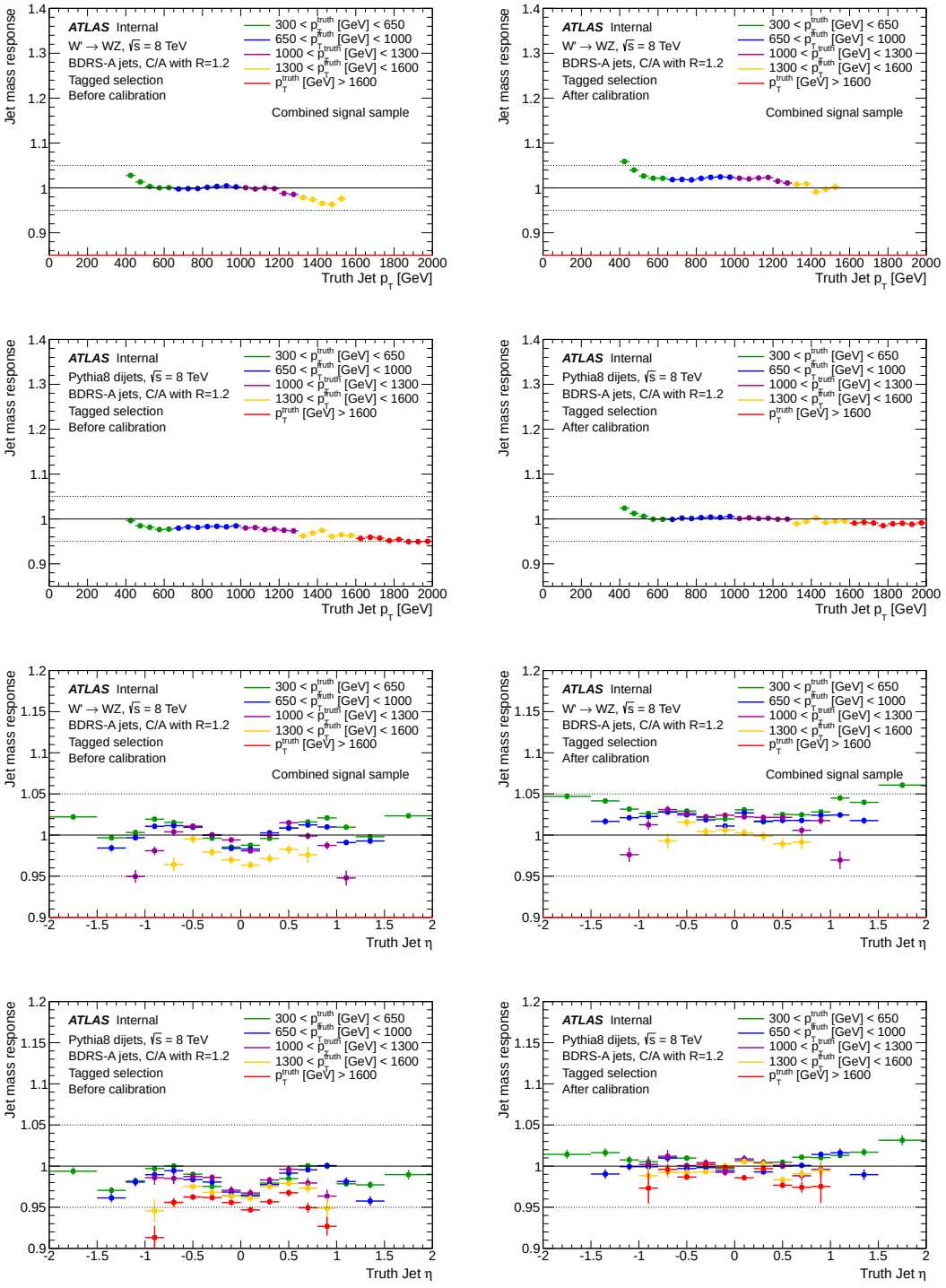


Figure A.9: Jet mass response as a function of the truth jet p_T and as well as the truth jet η for signal and QCD jets without (left) and with (right) the derived calibration applied. The jet mass response is shown in different truth p_T bins.

Appendix B

Jets in problematic regions of the calorimeter

Figure B.1 shows the number of events with leading groomed jets with $p_T > 500$ GeV as a function of the centroid of the jets in data. The location of long-term masked modules is easily distinguished by the deficiency of leading jets in these locations. The over-correction doe to a poor calibration for narrow high-p_{T} jets is also easily seen in the hot spots surrounding the masked regions.

Events with over-corrected jets are rejected to avoid potentially harmful distortions to the di-jet mass spectrum. An event is rejected more than two of the three most energetic topoclusters in either of its two leading jets fall within a masked module or in a module neighbouring a masked module ($LC3 < 3$).

Temporarily masked modules were not included in the detector simulation, therefore, the effects of the masked modules and the cleaning cut on the signal acceptance and efficiency are estimated in the following two steps. In the first step, the signal acceptance loss is estimated by emulating the frequency and location of masked modules and rejecting the events failing the cleaning cut. The signal acceptance loss due to this cleaning cut is approximately 5%. Figure B.2 shows the number of groomed jets as a function of the number of their three leading clusters in masked regions in data and Monte Carlo. In the second step, a scale factor is applied for those jets containing clusters in the problematic regions to account for the mismodeling of the energy in the masked modules. The scale factor is derived by selecting signal truth-jets with their centroids in problematic regions that pass the tagging criteria, and comparing the selection efficiency of the reconstructed jets in the simulated masked module to the selection efficiency in other modules. The resulting per-jet scale factor is $SF_{bch} = 0.66 \pm 0.02$, this scale factor is then used to correct the per-jet efficiency. Figure B.3 shows the boson tagging efficiency as a function of the W' signal mass, red points correspond to jets falling in the simulated masked module, and blue points correspond to jets falling in the other modules.

Since not all the masked modules are simulated in Monte Carlo, it has been decided to remove events containing jets affected by this miscalibration. The removal of groomed jets based on their centroids has been found to be ineffective, as the Cambridge-Aachen algorithm produces jets of irregular shape.

The BDRS-A groomed jets have irregular shapes as well as the ungroomed Cambridge/Aachen

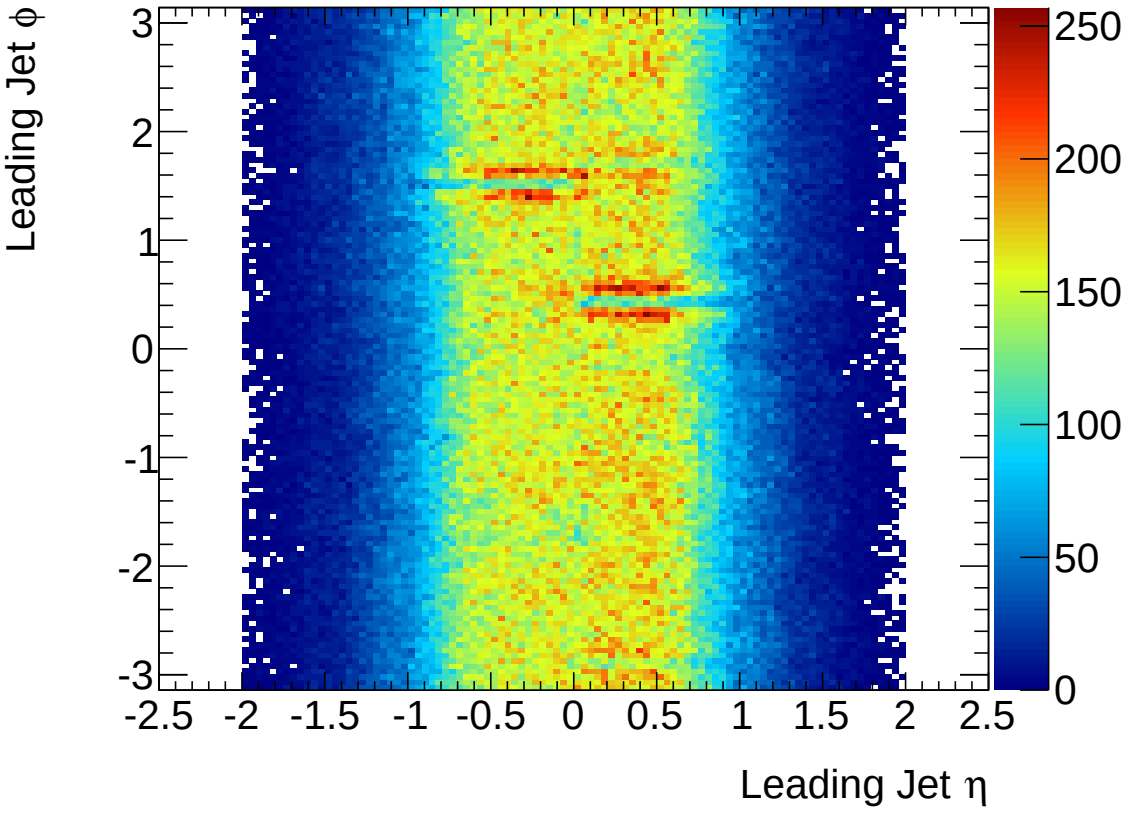


Figure B.1: Number of leading groomed jets with $p_T > 500$ GeV in data as a function of the jet centroids. The location of long-term masked modules is easily distinguished by the deficiency of leading jets in these locations. The over-corrective effect of the average energy density correction is also easily seen in the hot spots surrounding the masked regions.

jets, but with a reduced size, determined by the filtering procedure. Moreover, their energy is concentrated in a small number of clusters, corresponding to the hard W decay products. Therefore, the event removal strategy is based on the number of high-p_{T} topoclusters per jet falling in the masked regions. Only the three highest p_{T} (leading) topoclusters are taken into account to define the variable $LC3$ on which to cut.

This choice is motivated by Figure B.4, showing the untagged jet p_{T} and mass response in QCD Pythia MC as representative of the jet response for the main background for the analysis, considering exclusive $LC3$ bins ($LC3 = 0, 1, 2, 3$). From Figure B.5, where the jet response is evaluated after applying the tagging selection, it is visible that the W-tagging further reduces the p_{T} and mass response discrepancy when compared to the $LC3 < 3$ cut only. The response is defined as $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{true}$, as a function of the jet p_{T}^{true}. It can be seen that the response increases with the number of the jet leading clusters falling in the masked regions increases, as expected from the behaviour of the correction. The cut at $LC3 < 3$ is chosen so that the response of problematic jets is contained within the 2% p_{T} scale systematic uncertainty, in both the inclusive and exclusive cases.

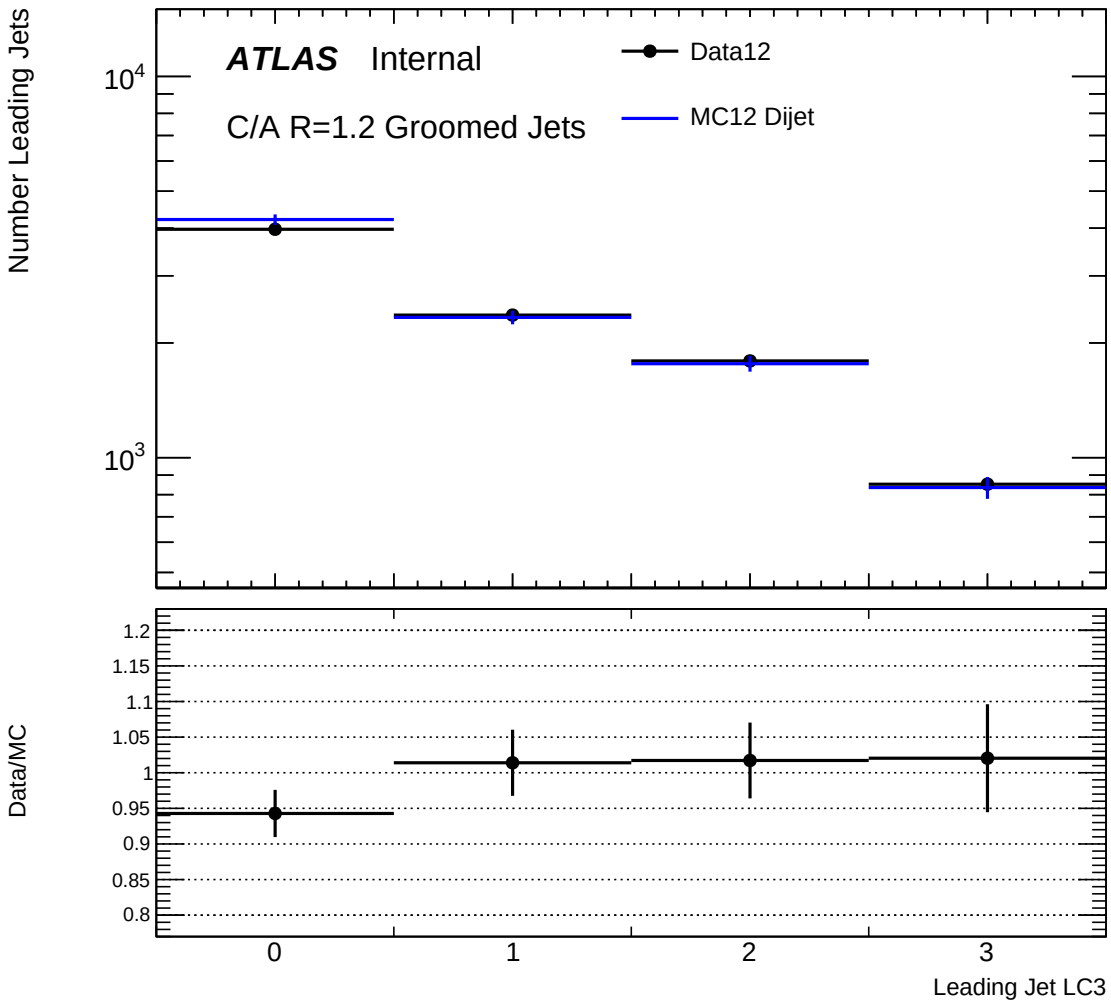


Figure B.2: Number of groomed jets as a function of the number of clusters out of their three leading clusters falling in masked regions in data and Monte Carlo.

B.0.0.1 Effect of event removal on boson tagging variables

The effect of the selection of the $LC3$ cut has been studied on data before applying the boson tagging selection. Figure B.6 shows the effect of the $LC3 < 3$ cut on the leading jet $(\eta - \phi)$ distribution. The regions corresponding to the Tile calorimeter modules LBA05 and LBC16 are still visible, but the hot spots on the edges have been completely removed. Figure 6.12 shows the leading jet ϕ distributions from the data using different values of the $LC3$ cut: it can observed that cutting at 3 removes the excesses on the edges of the two most prominent masked modules.

In order to investigate possible distortions in the spectra of the most sensitive boson tagging variables (i.e. jet mass, $\sqrt{y_f}$) and the jet p_{T}, ratio plots have been produced between the distribution for the sample where no bad clusters (falling within a masked module or neighboring a maksed module) are allowed to enter the jet ($f_{bad\_cl} = 0\%$) and the samples with the $LC3 < 1$, $LC3 < 2$ and $LC3 < 3$ cuts on the fraction of bad clusters. The fraction of bad clusters is

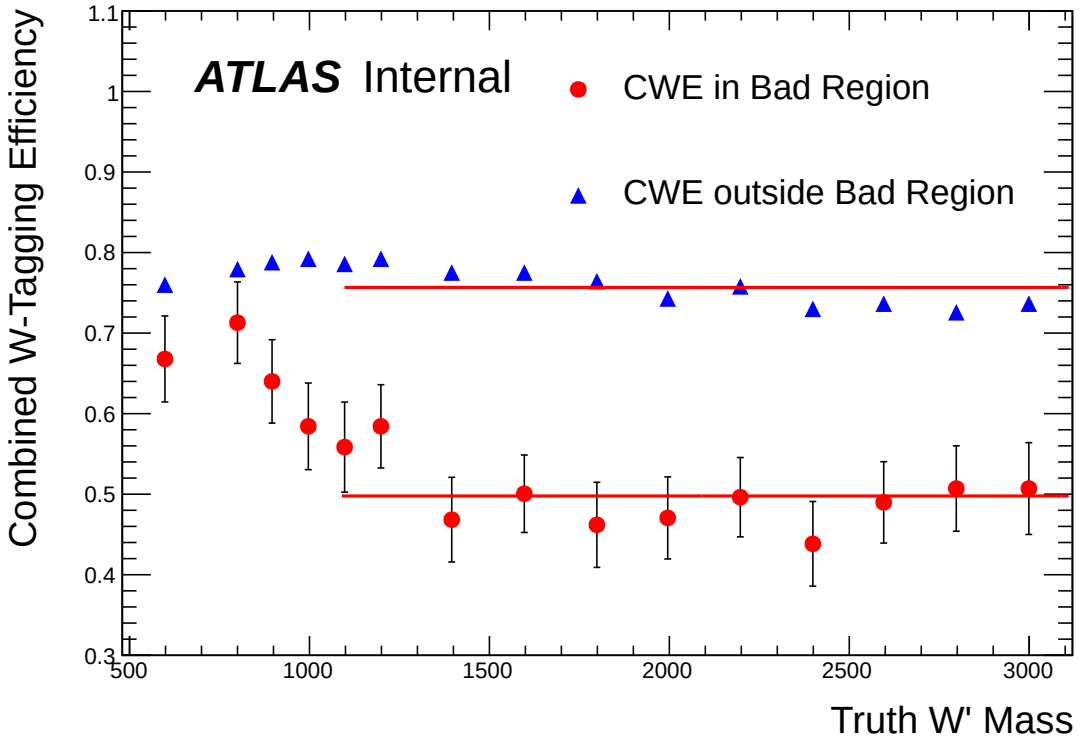


Figure B.3: Boson tagging efficiency as a function of the W' signal mass, red points correspond to jets falling in the simulated masked module, and blue points correspond to jets falling in the other modules.

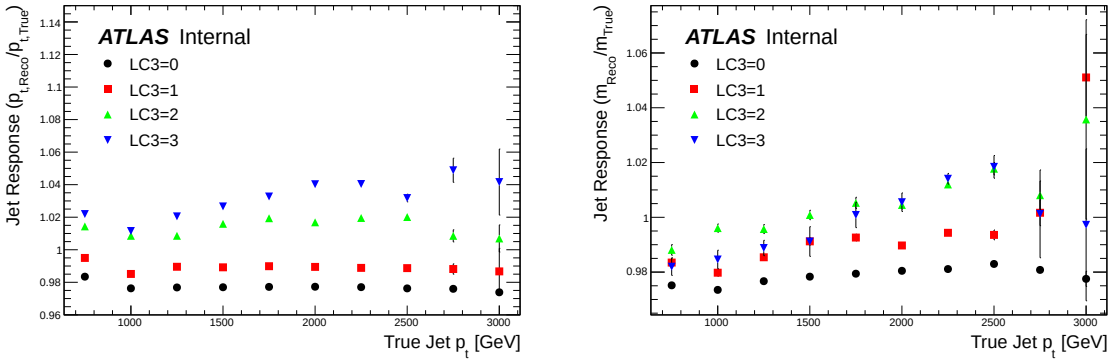


Figure B.4: p_{T} (left) and mass (right) response before applying the W-tagging selection.

defined as follows:

$$f_{bad\_cl} = \frac{N(BadClusters)}{N(TotalClusters)}$$

All the distributions before performing the ratio are normalized to unit area to investigate shape differences. From the ratio plots in Figure B.7, it is possible to observe that the shapes of the distributions for the $LC3 < 1, 2, 3$ cuts are in agreement within statistical uncertainties with the cleanest sample (equal to 1 in the ratio plots). The distribution where no cut has been

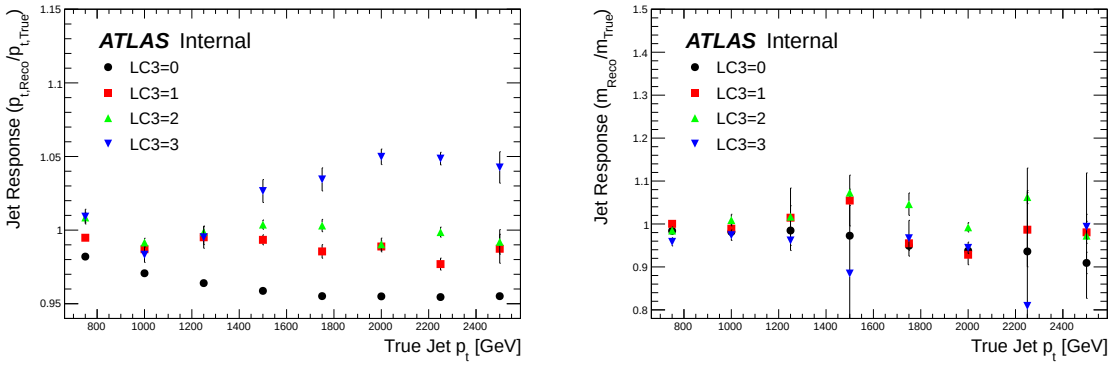


Figure B.5: p_{T} (left) and mass (right) response after applying the W-tagging selection.

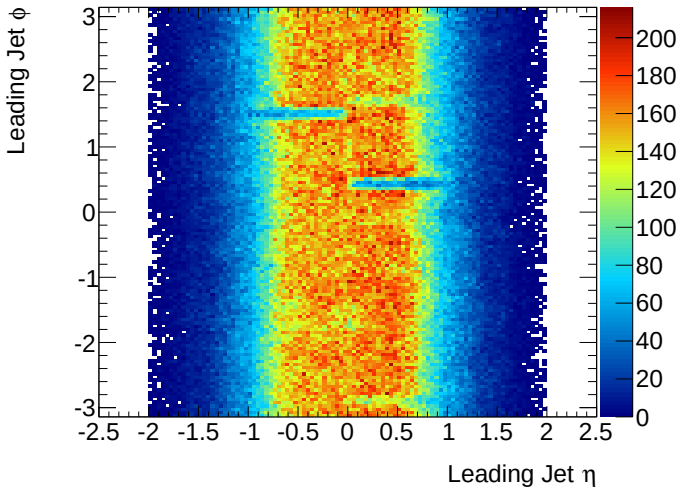


Figure B.6: $(\eta - \phi)$ map for untagged groomed jets after applying the $LC3 < 3$ cut.

applied, instead, diverges in the high p_T region. This is understood as the cut chosen removes the high p_{T} tail produced by the overcorrection in the masked Tile calorimeter regions.

In the ratio plots we observe a distortion only for momentum balance closer to 1, in the case of very balanced subjets. This is predominantly due to the limited statistics in the tail of the sample spectrum, where no bad clusters are allowed. Since the boson tagging cut applied to this variable is 0.45, the distortion of the cleaning selection on the variable is considered as negligible.

The differences between the three cut values are more significant in the case of untagged jet mass, as shown in Figure B.4. In particular, the bigger spectrum distortion takes place in the mass window selected by the boson tagging. The distribution for tighter $LC3$ cuts shows a large reduction of the high mass peak (> 100GeV) approaching the tightest selection, while the second peak is the result of the large spatial extension of the jet.

In Figure B.8 the untagged jet mass distribution is shown for different numbers of topoclusters. Reducing the number of clusters within the jets reduces the second mass peak due to the resulting reduction in jet size.

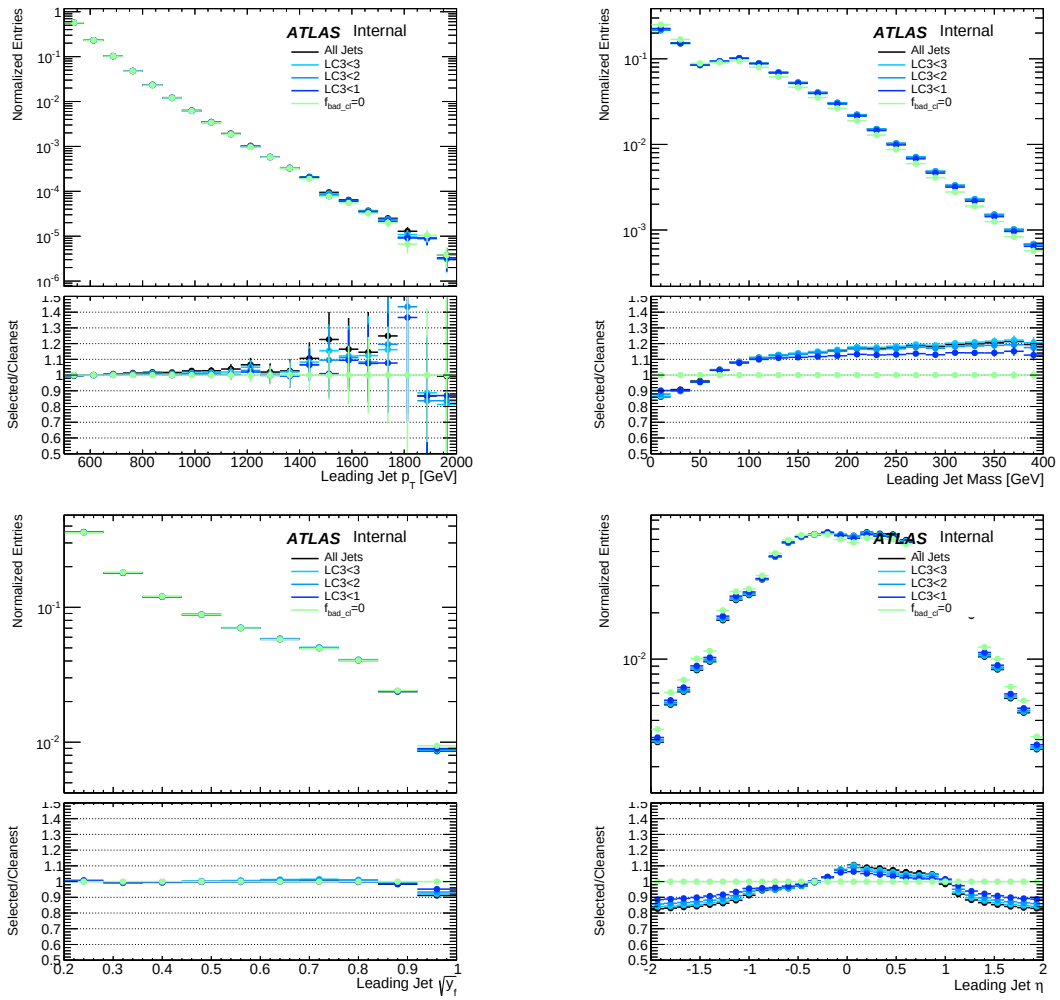


Figure B.7: p_{T}, mass, y_f and η distributions for the different $LC3$ cut values compared with the cleanest sample, where $f_{bad\_cl} = 0$.

The number of hadronic topoclusters passing the f_{bad_cl} selection is shown as a function of different value of the cut in Figure B.9. The jets with a larger spatial extension have a higher probability to contain at least one leading cluster falling in the region of a masked module: this explains the decrease of the distribution of the number of clusters for jets in events surviving the cut and the jet mass distribution distortion.

The $LC3 < 3$ cut has the following consequences:

- removes the hot spots in the leading jets $(\eta - \phi)$ distribution;
- reduces the high p_{T} tail in the leading jet p_{T} distribution;
- removes jets with average p_{T} response outside the 2% p_{T} scale systematic uncertainty;
- preserves the shapes of the boson tagging variables.

After applying the full event selection and the boson tagging, the cut removes $\sim 5\%$ of MC signal events and $\sim 2.5\%$ of background events. This difference between the rejection for signal

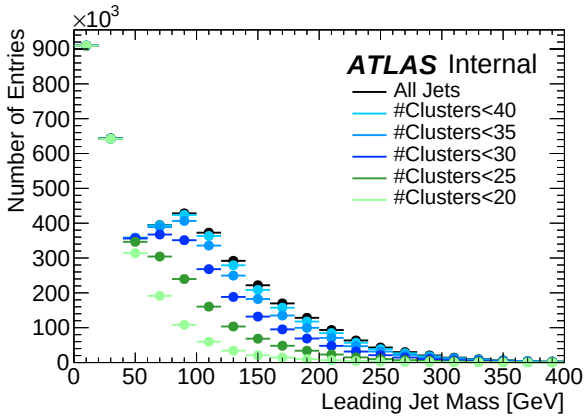


Figure B.8: Mass distribution before W-tagging for different numbers of topoclusters.

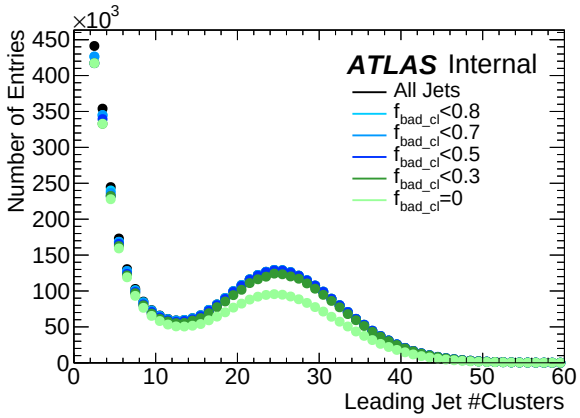


Figure B.9: Number of hadronic topoclusters for different values of f_{bad_cl}.

and background is due to fact that most of the Tile modules trips occur in the barrel region and the signal, characterized by more central topologies and narrower energy distributions, is more affected by the cut.

B.0.0.2 Systematic uncertainties and corrections

Since the $LC3$ method was developed on Monte Carlo simulation, the presence of a systematic uncertainty due to a possible mismodeling of the variable in the simulation was studied. As it can be seen in Figure B.2, no significant effect was found within statistical uncertainties.

As previously mentioned, in the Monte Carlo only the presence of one dead module is fully simulated. Other modules, which were masked just for a period of data taking, can be emulated in the Monte Carlo using the information on their spatial coordinates, available in the detector operation database. Therefore, their corresponding BCH_CORRECTION_CELL effect is not simulated. This affects the reconstruction and W-tagging efficiencies, resulting in an

overestimate of both.

The $LC3$ cut applied to the fully simulated masked regions, removes the hotspot around the dead module, but does not recover the reconstruction inefficiency caused by it. This is not true in the case of the emulated masked regions. The $LC3 < 3$ cut does not completely remove the hotspots, thus it is possible that one of the subjets of the candidate W boson partially falls into it. This causes its energy to be increased dramatically. These events are rejected at the W-tagging stage by requiring momentum balance between the two leading subjets ($\sqrt{y_f}$). This causes an efficiency loss in W-tagging jets falling in the masked regions.

To take into account the effect on both reconstruction and tagging, the CWE (Combined W-tagging and reconstruction Efficiency) variable is defined as a ratio with as numerator the number of groomed truth jets with their centroids in the bad (good) regions and as denominator the number of groomed truth jets satisfying the following conditions:

- their centroids are located in the bad (good) region;
- they are W-tagged;
- they satisfy the $LC3 < 3$ cut.

This variable is calculated both inside at the fully simulated masked region and outside of it. Its distribution is shown in Figure B.3. The ratio between the CWE inside the fully simulated masked region and the CWE outside it is taken as a scale factor and applied to jets falling inside the emulated masked regions, corresponding to 10% of the total jets. The scale factor is found to be 0.66 ± 0.02, and its corresponding systematic uncertainty is negligible.

Bibliography

[1] H. Weyl, *Theory of Groups and Quantum Mechanics*. 1929.

[2] F. Halzen and A. D. Martin, *Quarks & Leptons: An Introduction in Modern Particle Physics*. Wiley, 1984.

[3] I. J. Aitchison and A. J. Hey, *Gauge Theories in Particle Physics: A Practical Introduction, Fourth Edition*. CRC Press, 2012.

[4] S. L. Glashow, *Partial-symmetries of weak interactions*, **22** (Feb., 1961) 579–588.

[5] S. Weinberg, *A Model of Leptons*, Physical Review Letters **19** (Nov., 1967) 1264–1266.

[6] A. Salam, *Weak and electromagnetic interactions*, . Proc. of the 8th Nobel Symposium on 'Elementary particle theory, relativistic groups and analyticity', Stockholm, Sweden, 1968, edited by N. Svartholm, p.367-377.

[7] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys.Rev.Lett. **13** (1964) 321–323.

[8] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys.Lett. **12** (1964) 132–133.

[9] TEVNPH (Tevatron New Phenomina and Higgs Working Group), CDF, D0 Collaboration, *Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data*, 2012. arXiv:1203.3774 [hep-ex]. http://lss.fnal.gov/cgi-bin/find_paper.pl-conf-12-065.

[10] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the {ATLAS} detector at the {LHC}*, Physics Letters B **716** (2012) no. 1, 1 – 29. http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[11] C. Collaboration, *Observation of a new boson at a mass of 125 GeV with the {CMS} experiment at the {LHC}*, Physics Letters B **716** (2012) no. 1, 30 – 61. http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[12] ATLAS Collaboration and CMS Collaboration Collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (May, 2015) 191803. http://link.aps.org/doi/10.1103/PhysRevLett.114.191803.

[13] V. Trimble, *Existence and Nature of Dark Matter in the Universe*, Annual Review of Astronomy and Astrophysics **25** (1987) no. 1, 425–472, http://dx.doi.org/10.1146/annurev.aa.25.090187.002233. http://dx.doi.org/10.1146/annurev.aa.25.090187.002233.

[14] P. J. E. Peebles and B. Ratra, *The cosmological constant and dark energy*, Rev. Mod. Phys. **75** (Apr, 2003) 559–606. http://link.aps.org/doi/10.1103/RevModPhys.75.559.

[15] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, Phys. Rev. Lett. **83** (1999) 3370–3373, arXiv:hep-ph/9905221 [hep-ph].

[16] L. Randall and R. Sundrum, *An Alternative to compactification*, Phys. Rev. Lett. **83** (1999) 4690–4693, arXiv:hep-th/9906064 [hep-th].

[17] H. Davoudiasl, J. Hewett, and T. Rizzo, *Experimental probes of localized gravity: On and off the wall*, Phys. Rev. D **63** (2001) 075004, arXiv:hep-ph/0006041 [hep-ph].

[18] M. Gabella, *The Randall-Sundrum Model*, www-thphys.physics.ox.ac.uk/people/MaximeGabella/rs.pdf **6** (2006) no. 9, .

[19] G. Altarelli, B. Mele, and M. Ruiz-Altaba, *Searching for new heavy vector bosons in p anti-p colliders*, Z. Phys. C **45** (1989) 109.

[20] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) no. 08, S08001. http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[21] T. Nayak and B. Sinhah, *Physics at the Large Hadron Collider*, ch. Search and Study of Quark-gluon Plasma at the CERN-LHC, pp. 131–144. Springer India, New Delhi, 2009. http://dx.doi.org/10.1007/978-81-8489-295-6_9.

[22] ATLAS Collaboration, *Luminosity Public Results*, . https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[23] ATLAS Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C73** (2013) no. 8, 2518, arXiv:1302.4393 [hep-ex].

[24] ATLAS Collaboration, *ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[25] CMS Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[26] LHCb Collaboration, A. A. Alves, Jr. et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[27] ALICE Collaboration, K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[28] ATLAS Collaboration, G. Aad et al., *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det].

[29] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[30] ATLAS Collaboration Collaboration, *ATLAS Computing: technical design report.* Technical Design Report ATLAS. CERN, Geneva, 2005. https://cds.cern.ch/record/837738.

[31] P. Mato, *Status of the GAUDI event-processing framework*, . https://cds.cern.ch/record/1745147.

[32] G. Barrand et al., *GAUDI - The software architecture and framework for building LHCb data processing applications*, in *Proceedings, 11th International Conference on Computing in High-Energy and Nuclear Physics (CHEP 2000)*, pp. 92–95. 2000. http://chep2000.pd.infn.it/abst/abs_a152.htm.

[33] L. Lonnblad, *CLHEP: A project for designing a C++ class library for high-energy physics*, Comput. Phys. Commun. **84** (1994) 307–316.

[34] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81–86.

[35] B. R. W. R. K. Ellis, W. J. Stirling, *QCD and Collider Physicsn.* Cambridge Monographs on Particle Physics Nuclear Physics and Cosmology, 2003.

[36] S. Bethke, *The 2009 World Average of alpha(s)*, Eur. Phys. J. **C64** (2009) 689–703, arXiv:0908.1135 [hep-ph].

[37] R. P. Feynman, *THE BEHAVIOR OF HADRON COLLISIONS AT EXTREME ENERGIES*, .

[38] V. N. Gribov and L. N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) no. 4, 438–450. https://cds.cern.ch/record/427157.

[39] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298.

[40] Wikipedia, *Quark model — Wikipedia, The Free Encyclopedia*, 2016. https://en.wikipedia.org/w/index.php/title=Quark_model&oldid=714288023. [Online; accessed 10-April-2016].

[41] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[42] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31–145.

[43] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. Seymour, and B. Webber, *HERWIG 6.5 Release Note*, arXiv:hep-ph/0210213 .

[44] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].

[45] ATLAS Collaboration, G. Aad et al., *Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **09** (2013) 076, arXiv:1306.4945 [hep-ex].

[46] ATLAS Collaboration, T. Barillari et al., *Local hadronic calibration*, .

[47] ATLAS Collaboration, G. Aad et al., *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 17, arXiv:1406.0076 [hep-ex].

[48] ATLAS Collaboration, G. Aad et al., *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. **C73** (2013) no. 3, 2304, arXiv:1112.6426 [hep-ex].

[49] W. T. Giele and E. W. N. Glover, *Physics with jets in $p\bar{p}$ collisions*, . https://cds.cern.ch/record/283947.

[50] S. D. Ellis, Z. Kunszt, and D. E. Soper, *One-jet inclusive cross section at order α_s^3. Gluons only*, Phys. Rev. D **40** (Oct, 1989) 2188–2222. http://link.aps.org/doi/10.1103/PhysRevD.40.2188.

[51] G. P. Salam, *Towards Jetography*, Eur. Phys. J. **C67** (2010) 637–686, arXiv:0906.1833 [hep-ph].

[52] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[53] G. P. Salam and G. Soyez, *A Practical Seedless Infrared-Safe Cone jet algorithm*, JHEP **05** (2007) 086, arXiv:0704.0292 [hep-ph].

[54] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[55] S. Bentvelsen and I. Meyer, *The Cambridge jet algorithm: features and applications*, The European Physical Journal C - Particles and Fields **4** no. 4, 623–629. http://dx.doi.org/10.1007/s100520050232.

[56] D. Adams et al., *Towards an Understanding of the Correlations in Jet Substructure*, Eur. Phys. J. **C75** (2015) no. 9, 409, arXiv:1504.00679 [hep-ph].

[57] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett. **100** (2008) 242001, arXiv:0802.2470 [hep-ph].

[58] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084, arXiv:0912.1342 [hep-ph].

[59] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, JHEP **03** (2011) 015, arXiv:1011.2268 [hep-ph].

[60] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, Nucl. Phys. **B406** (1993) 187–224.

[61] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57–61, arXiv:hep-ph/0512210 [hep-ph].

[62] ATLAS Collaboration, G. Aad et al., *Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **12** (2015) 055, arXiv:1506.00962 [hep-ex].

[63] ATLAS Collaboration, ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) no. 8, 2518, arXiv:1302.4393 [hep-ex].

[64] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv:hep-ph/0901.0002 [hep-ph].

[65] ATLAS Collaboration, ATLAS Collaboration, *Further ATLAS tunes of PYTHIA6 and Pythia 8*, ATL-PHYS-PUB-2011-014 (2011) . http://cdsweb.cern.ch/record/1400677.

[66] M. Bahr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton, et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639–707, arXiv:0803.0883 [hep-ph].

[67] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Warped Gravitons at the LHC and Beyond*, Phys. Rev. D **76** (2007) 036006, arXiv:hep-ph/0701186 [hep-ph].

[68] A. Belyaev, N. D. Christensen, and A. Pukhov, *CalcHEP 3.4 for collider physics within and beyond the Standard Model*, Comput. Phys. Commun. **184** (2013) 1729–1769, arXiv:1207.6082 [hep-ph].

[69] D. Stump, J. Huston, J. Pumplin, W.-K. Tung, H. Lai, et al., *Inclusive jet production, parton distributions, and the search for new physics*, JHEP **0310** (2003) 046, arXiv:hep-ph/0303013 [hep-ph].

[70] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **0206** (2002) 029, arXiv:hep-ph/0204244 [hep-ph].

[71] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[72] ATLAS Collaboration Collaboration, *High-p_{T} boson jet tagging with the ATLAS detector*, ATLAS-COM-PHYS-2013-899 (2013) .

[73] S. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann, *Jets in hadron-hadron collisions*, Prog.Part.Nucl.Phys. **60** (2008) 484–551, arXiv:0712.2447 [hep-ph].

[74] ATLAS Collaboration Collaboration, *ATLAS search for diboson resonances with jets in 20 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV*, ATLAS-COM-PHYS-2013-900 (2013) .

[75] B. Webber, *Average Multiplicities in Jets*, Phys.Lett. **B143** (1984) 501.

[76] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, JHEP **0804** (2008) 005, arXiv:0802.1188 [hep-ph].

[77] G. Choudalakis and D. Casadei, *Plotting the differences between data and expectation*, Eur. Phys. J. Plus **127** (2012) no. 2, 25.

[78] ATLAS Collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C **74** (2014) no. 8, 3023, arXiv:1405.6583 [hep-ex].

[79] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv:1007.1727 [physics.data-an]. [Erratum-ibid. **C 73** (2013) 2501].

[80] A. L. Read, *Presentation of search results: The CL_s technique*, J. Phys. G **28** (2002) 2693–2704.

[81] ATLAS Collaboration, G. Aad et al., *Search for WZ resonances in the fully leptonic channel using pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Lett. **B737** (2014) 223–243, arXiv:1406.4456 [hep-ex].

[82] ATLAS Collaboration, G. Aad et al., *Search for resonant diboson production in the $\ell\ell q\bar{q}$ final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 69, arXiv:1409.6190 [hep-ex].

[83] ATLAS Collaboration, G. Aad et al., *Search for production of WW/WZ resonances decaying to a lepton, neutrino and jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. **C75** (2015) no. 5, 209, arXiv:1503.04677 [hep-ex]. [Erratum: Eur. Phys. J.C75,370(2015)].

[84] ATLAS Collaboration, G. Aad et al., *Combination of searches for WW, WZ, and ZZ resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Lett. **B755** (2016) 285–305, arXiv:1512.05099 [hep-ex].

[85] CMS Collaboration, V. Khachatryan et al., *Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV*, JHEP **08** (2014) 174, arXiv:1405.3447 [hep-ex].

[86] CMS Collaboration, V. Khachatryan et al., *Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **08** (2014) 173, arXiv:1405.1994 [hep-ex].

[87] *Search for resonances with boson-tagged jets in 3.2 fb^{-1} of p-p collisions at $\sqrt{s}$= 13 TeV collected with the ATLAS detector*, ATLAS-CONF-2015-073, CERN, Geneva, Dec, 2015. http://cds.cern.ch/record/2114845.

Ringraziamenti

Il primo giugno 2011 ha avuto inizio la mia esperienza di dottorato all'Università di Ginevra. In questi 5 anni la mia vita è cambiata molto. Sono cresciuto, maturato, sia dal punto di vista personale, sia per quello che riguarda la carriera di un giovane fisico delle alte energia. Tante persone hanno intersecato le loro vite con la mia e mi hanno arricchito. Nelle prossime righe cercherò di esprimere i miei ringraziamenti alle persone che hanno segnato e reso possibile questo percorso.

Il primo ringraziamento va al prof. Iacobucci. Egli ha innanzitutto creduto in me, dandomi l'opportunità di entrare nel programma il dottorato in fisica delle alte energie con il gruppo ATLAS dell'Università di Ginevra. Dalle nostre discussioni ho imparato quanto per un fisico sperimentale sia importante la cura maniacale dei dettagli e la capacità di presentare in modo chiaro e preciso il proprio lavoro.

Il mio dottorato si è svolto quasi totalmente al CERN. Questo laboratorio è un posto speciale, in cui spesso si fa l'errore di considerare normale l'essere circondati da persone eccezionali. Tra queste mi preme menzionare e ringraziare Caterina Doglioni ed Enrique Kajomovitz Must. Entrambi sono rari esempi di preparazione, dedizione e talento. Caterina è stata la post-doc che mi ha seguito durante il dottorato. Da lei ho imparato tutto quello che conosco sulla fisica dei jets, dalle performance all'uso di questi "oggetti" all'interno di un'analisi. Inoltre, mi ha insegnato come essere parte attiva di un un gruppo di ricerca e quanto sia fondamentale non avere timore nel fare domande. Con Enrique ho condiviso gli ultimi due anni di dottorato. La gran parte delle cose che ho imparato dal lavoro di analisi presentato in questa tesi la devo a lui e ai nostri confronti quotidiani. Anche nella frenesia dei ritmi del CERN, abbiamo sempre cercato il tempo per sederci davanti ad una lavagna o un foglio di carta e ragionare. Per questo motivo il suo contributo alla mia maturazione professionale è stato fondamentale.

Un grande ringraziamento va a tutti i miei colleghi di dottorato, post-doc, professori e studenti del gruppo ATLAS dell'Università di Ginevra. Lavorare con voi è stato un privilegio. Ho davvero apprezzato il clima amichevole e stimolante che si è instaurato tra noi. È stata una bella squadra e vi auguro di continuare ad essere così uniti ed affiatati.

La vita al CERN non sarebbe stata la stessa senza il nostro chiassoso gruppo di amici italiani. Vi ringrazio per aver condiviso con me tanti momenti di questa esperienza. Abbiamo avuto modo di confrontarci quasi quotidianamente esplorando qualsiasi argomento, dalla politica alla fisica, passando per le nostre frustrazioni e soddisfazioni. Ringrazio in particolare Valerio Ippolito, Antonio Bellotta, Dario Berzano e Francesco Guescini, per essere stati e per essere ancora (spero per molto) punti cardine della mia vita a Ginevra e soprattutto per essere delle persone speciali.

Un ringraziamento speciale va alla "mia" persona speciale, Chiara Debenedetti. Ci siamo

conosciuti che eravamo due dottorandi. Abbiamo condiviso (e continuiamo a farlo) frustrazioni, passioni e la vita in questo angolo di mondo. Ci siamo tenuti per mano quotidianamente, ci siamo confrontati, e senza il suo supporto tutto sarebbe stato più difficile. Grazie per essere una parte speciale del mio mondo.

Le ultime parole le dedico alla mia famiglia. Cinque anni sono tanti e la vita ci ha fatto capire quante sono le cose che possono accadere in un tale lasso di tempo. Ringrazio mio padre: il mio modello, il mio sorriso e la mia positività. Vivi con me tutti i giorni e mi spingi ad apprezzare la vita. Ringrazio mia madre, per non essersi mai accontentata dei miei risultati e per aver creduto in me. Ringrazio mia sorella: un misto di passione e talento, che ha paura di crescere.

Infine, dedico questo lavoro di tesi a tutta la gente curiosa.

Acknowledgements

On June 1st, 2011 I started my PhD experience at the University of Geneva. During these 5 years my life changed a lot. I grew up, I matured, both personally and professionally, as a young high-energy physicist. Many people entered my life and enriched it. In the following lines I will try to thank all the people who left a very clear impact on this path and that made it possible.

First, I want to thank Prof. Iacobucci. He believed in me giving me the opportunity to start the PhD in high-energy physics with the ATLAS group at the University of Geneva. From our discussions I learned how important it is for an experimental physicist to focus on all the details and how useful it is to be able to present results in a precise and clear way.

During my PhD I was mostly working at CERN. This laboratory is a special place where you work side by side with, and are surrounded by exceptional people, that too often we "wrongly" consider as normal. Among them, I want to mention and to thank Caterina Doglioni and Enrique Kajomovitz Must. Both of them are rare examples of high level preparation, hard work, and incredible talent. Caterina has been the post-doc I worked with during my PhD. She taught me most of the things I learned about jet physics, spanning from performance to the proper use of these "objects" in physics analyses. Furthermore, she taught me how to become an active member of an analysis team and how fundamental asking "all" the questions is. With Enrique I shared the last two years of my PhD activity. He is responsible for most of the things I learned in the analysis work presented in this thesis. Even during the frenetic life at CERN, we always found the time to sit together in front of a whiteboard or a piece of paper, to discuss and think about physics problems. For this reason his contribution to my career has been fundamental.

A big thanks goes to all my PhD colleagues, post-docs, professors and students of the ATLAS group of the University of Geneva. Working with you has been a privilege. I really enjoyed the nice and exciting environment I was part of. It is a fantastic team and I wish for you all the best.

CERN life would not be the same without our "noisy" group of Italian friends. I really thank you for having shared with me many important moments of this experience. I had the opportunity to speak almost daily about many topics, from physics to politics, passing through our frustrations and satisfactions. In particular, I want to thank Valerio Ippolito, Antonio Bellotta, Dario Berzano e Francesco Guescini, for being the most important people of my life in Geneva and for being special.

A special thanks goes to "my" special person, Chiara Debenedetti. We met each other at beginning of our PhDs. We shared (and we keep sharing) frustrations, passions and everyday

life in this small portion of the world. We help each other and confront ourselves every day, and without her support everything would have been much more difficult. Thanks a lot for being a special part of my world.

The last lines are dedicated to my family. Five years are a lot and life showed us how quickly it can change. I thank my father: my role model, my smile and my positivity. You live with me every day and you push me to enjoy life. I thank my mother, to have trusted me and to have never considered my results enough. I thank my sister: a mixture of passion and talent, who does not want to grow up.

Lastly, I dedicate this thesis work to all the curious people.