TEL AVIV UNIVERSITY
RAYMOND AND BEVERLY SACKLER
FACULTY OF EXACT SCIENCES
SCHOOL OF PHYSICS & ASTRONOMY
TEL AVIV UNIVERSITY
אוניברסיטת תל אביב
אוניברסיטת תל-אביב
הפקולטה למדעים מדוייקים
ע"ש ריימונד ובברלי סאקלר
בית הספר לפיסיקה ואסטרונומיה

Search for Dijet Mass Resonance Using Trigger-Level-Analysis in Proton Proton Collision at $\sqrt{s} = 13$ TeV with the ATLAS Detector

Guy Koren

CERN-THESIS-2016-224
14/12/2016

Dissertation submitted in partial fulfillment of the requirements of the degree of M.Sc. in the Particle Physics Department of the Tel Aviv University under the supervision of Prof. **Erez Etzion**

September 2016

guy.koren@cern.ch

Abstract

This work presents a search for a dijet mass resonance using Trigger-Level-Analysis in proton-proton collisions at $\sqrt{13}$ TeV with the ATLAS detector. Dijet searches in ATLAS with sub-TeV masses are statistically limited by the available bandwidth for inclusive jet triggers. The high Standard Model production rate of jets require ATLAS to apply prescale factors on these triggers in order to record full events in a manageable rate. The Trigger-Level-Analysis introduces a new strategy to overcome these limitations; By recording events in a reduced data format, which include only the subset of information necessary for a resonance search, it's possible to record events at much higher rates. The search described in this work considers dijet masses between 394 GeV - 1.24 TeV and uses 3.4 fb^{-1} of data collected during 2015. The data shows agreement with the prediction from the Standard Model, and exclusion limits on possible models were set.

Acknowledgements

The completion of this thesis marks the end of a period in which I have learned a lot and developed a great deal. I would like to take this opportunity and thank the many individuals who provided help and support along the way.

I would first like to express my deep gratitude to Prof. Erez Etzion for his guidance and advice, and for introducing me to the fascinating field of experimental particle physics, first as a teacher and later as a supervisor. I would like to thank my colleagues in Tel Aviv university; Adi Ashkenazi and Merlin Davies, to whom I looked up to as I took my first steps in ATLAS, and to Gideon Bella, Hadar Cohen, Meny Ben-Moshe, Jorge Duarte-Campderros, Nimrod Taiblum, Rotem Gorodeisky, Yan Benhammou, Abner Soffer, Moran Netzer, Noa Feldman, Amir Levi, Sharon Elias, Hadar Greener, and Shaked Zychlinski. I have shared work space with many of them and they all made the last few years enjoyable.

I would like to thank the wonderful people I worked with on this analysis, and especially to Caterina Doglioni, who guided me from day one, and provided advice and support on a daily basis, and to Antonio Boveia and Kate Pachal, for patiently sharing with me from their vast knowledge and experience. I cherish the opportunity I had in working with the three of them.

Though the following dear people did not contribute directly to the content of this thesis, their impact on my life during this time was deeply profound.

I want to thank my friends. I consider myself lucky to be surrounded by such a group of loyal, funny, and *bright* individuals. A special thank you to Hagar Erez, for her love, her compassion, and the countless virtues she brings into my life, and to my family for their constant love and support, not only in the past few years.

Last, but very far from least, to my parents, Mira and Zvi. Thank you. Thank you for everything you did for me throughout the years, for encouraging and believing in me, and for the huge part you have in the person I developed to be. I love you two very much.

Contents

Chapter 1

Introduction

The notion that the world is made up from discrete units can be traced back to the time of the ancient Greeks. The name *"Atoms"* was given then to pieces of matter that can not be further split into smaller portions. However, more than 2000 years had passed before these ideas could be held to experimental tests. The Standard Model of particle physics was developed as the theory that describes the fundamental particles in the universe, and how they interact among themselves. Ever since it's establishment during the mid 1960's and early 1970's, the Standard Model surpassed rigorous testing and it's predictions were able to account for almost all experimental results. Despite it's success, the theory is known to be incomplete; Observations such as the presence of Dark Matter in the universe, asymmetry between the abundances of matter and anti-matter, and neutrino oscillations are examples for phenomena that can not be explained in the framework of the Standard Model. Attempts to resolve these conflicts were made by developing theories that extend the Standard Model, and seeking for their evidence in major experimental operations.

The European Organization for Nuclear Research, CERN[1], located outside of Geneva, Switzerland, is the site of an international laboratory for research of fundamental physics. Since it's establishment in 1954, scientists in CERN operated particle accelerators and detectors with the goal of witnessing evidence of new physics processes, and measuring the current theories to higher precision. The recent generation of experimental machinery

[1]From French: *Conseil Européen pour la Recherche Nucléaire.*

includes the Large Hadron Collider (LHC), a ring shaped particle accelerator used to conduct high energy collision experiments. The LHC started operating in 2010 and in July of 2012 the discovery of the Higgs Boson was announced by the ATLAS and CMS experiments. At the time of writing this work, no sign of physics beyond the Standard Model was found.

The most abundant physics objects produced in proton-proton collisions at the LHC are jets, collimated sprays of particles resulting from the production of high-energy quarks and gluons. Their high production rate exceeds the data-storing capabilities of ATLAS computers. The majority of jets in ATLAS are produced with low energies, whereas beyond the Standard Model signals may come associated with high energy events. To allow more bandwidth for high energy jets events, ATLAS applies prescale factors on jet triggers which reduce the recording rates of low energy jets events. This limits the statistics available in jet related searches below a mass scale of $\sim$1 TeV.
The Trigger-Level-Analysis (TLA) introduces a new strategy for collecting data, applied initially in a dijet mass search; By recording only a subset of the event information, that which is necessary to perform a resonance search, it's possible to record much higher event rates and gain sensitivity to physics in the sub-TeV mass scale.

The work presented here describes the search for a resonance in the dijet mass distribution from proton-proton collision at $\sqrt{s} = 13$ TeV with the ATLAS detector using a Trigger-Level-Analysis.
Chapter 2 gives an introductory theoretical background to the Standard Model, as well as to the relevant physics of proton-proton (pp) collisions, and to one extension model of physics beyond the Standard Model - the Z' dark matter mediator. Chapter 3 describes the experimental setup of the Large Hadron Collider and the ATLAS detector. The implementation of the Trigger-Level-Analysis in ATLAS and the validation of different parts in the analysis chain are reviewed in Chapter 4, and Chapter 5 presents the data and Monte Carlo (MC) samples used in the analysis. Chapter 6 discusses the strategy for estimating the dijet background, and the specific complication relevant in the Trigger-Level-Analysis case. Finally Chapter 7 and Chapter 8 show the results of the search and the corresponding exclusion boundaries respectively.

Chapter 2

Theoretical Background

While this thesis describe an experimental measurement, this opening chapter consists of an introductory theoretical background. A brief introduction to the Standard Model is presented with emphasis on QCD in section 2.1. In section 2.2 phyiscs beyond the Standard Model (BSM) is discussed. The Z' model of a dark matter mediator is presented as an example for an extension to the Standard Model which has a decay mode to dijets. Aspects of physics relevant to a collider experimets are discussed in section 2.3.

2.1 The Standard Model

Most information reviewd in this section had been drawn from [1, 2]. Other sources are cited in the text.

2.1.1 Particle Content

In the current picture of particle physics the universe is composed of a few "building-blocks", called elementary particles. The elementary particles are subject to four fundamental forces: Strong force, Weak force, Electromagnetic (EM) force, and Gravity. The mathematical framework describing how these particles and three of the forces (Gravity excluded) interact is called the Standard Model (SM). Since its development in the mid 1960's - early 1970's the SM has successfully predicted a variety of phenomena and has been able to explain almost all experimental results.

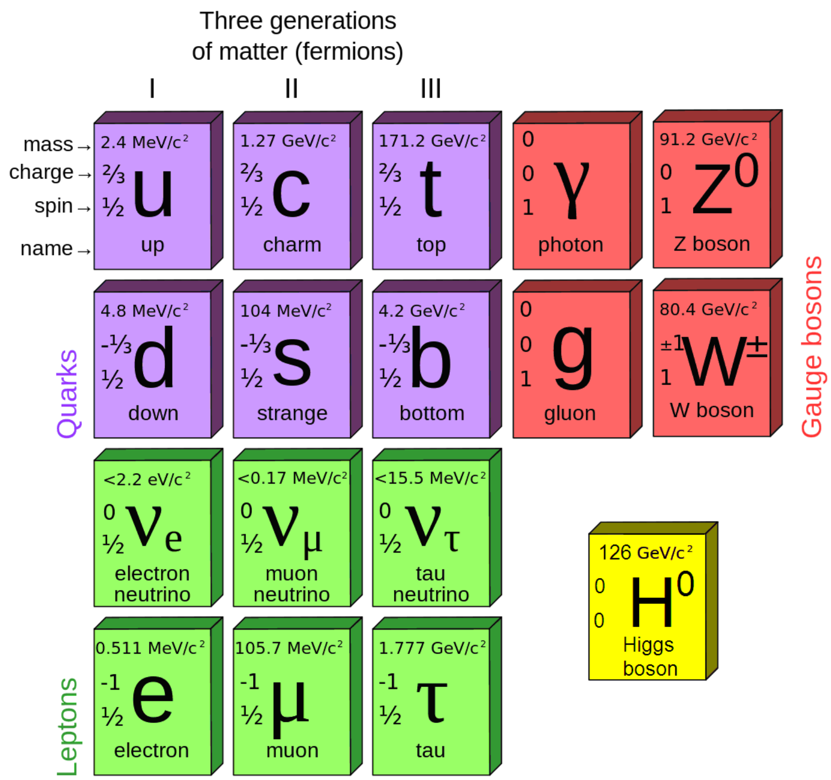


Figure 2.1: The Standard Model particles divided to their respective sub groups and generation.

Elementary particles are often divided into two groups: fermions and bosons.

Fermions - "matter particles" Fermions are the constituents of matter. They carry a half-integer spin and can be further divided into leptons and quarks. Leptons take part in the Electro-Weak interaction only while quarks interact via all SM interactions (Strong and EW). Each group consists of three pairs of particles. For the leptons they are the electron (e) and it's "partner" the electron-neutrino (ν_e) , the muon (μ) and muon-neutrino (ν_μ), and the tau (τ) and tau-neutriono (ν_τ). e, μ, and τ carry a -1 EM charge whereas the three ν's are neutral, i.e carry no EM charge. The three pairs of quarks are the up (u) and down (d) quarks, the charm (c) and strange (s), and top (t) and bottom (b).

Both quarks and leptons have spin $\frac{1}{2}$.

Bosons - Gauge and Higgs Bosons too are split into to two different types. Gauge boson are responsible for mediating the different forces. When two particles interact, they are said to be "exchanging gauge bosons" between them. For this reason gauge boson are often referred to as "force-carrying particles". The gauge bosons consists of the gluon g - responsible for the strong interaction, W^+, W^-, Z^0 - mediate weak interactions, and the photon γ - EM interaction. All gauge bosons have spin 1. The second type of boson is the Higgs boson. It is a scalar particle, meaning its spin is 0. Theorized in 1964 and first observed at the LHC in 2012[3], the Higgs boson is in the heart of the mechanism which allows other elementary particles acquire mass[4].

2.1.2 QCD

Quantum Chromodynamics (QCD) is the gauge field theory describing strong interactions. The strong interaction couples to quarks and is mediated by gluons. The relevant gauge symmetry for QCD is non-abelian $\mathrm{SU}(3)_C$. The charge associated with this symmetry, and hence with strong interactions is the *color* charge. Quarks can be in a state of either "red", "green", or "blue" (r, g, b), and anti-quarks in "anti-red", "anti-green", or "anti-blue" $(\bar{r}, \bar{g}, \bar{b})$. The theory's non-abelian structure dictates that gluons also carry color charge. To enforce the conservation of color in a vertex gluons are required to carry one unit of color and one unit of anti color. A color+anti-color combination can result in eight colored states (Octet) and one colorless state (Singlet), in accordance with the group algebra: $3 \times \bar{3} = 8 \oplus 1$. The Octet states are compatible with SU(3)'s eight required generators and correspond to eight physical gluons. The singlet state has no realization as a particle in nature. Quarks and gluons - *partons* - can not be found as free particles on their own. This is due to a feature of QCD called *confinement*. Confinement determines that only neutral states of color are observed in nature. Until recently such particles were observed in two forms: a combination of a quark and an anti-quark (*meson*) carrying opposite color charges, or three quarks each carrying a different charge of color (*baryon*). In recent years experimental observations confirmed that color neutral states can be constructed from combinations of four (*tetraquark*) and five (*pentaquark*) quarks as well [5, 6]. Another feature of QCD is *asymptotic freedom*. Assymptotic freedom determines that the interaction between two particles grows stronger as the distance between them increases (or their momentum transfer decreases). In terms of the running coupling constant $\alpha_{\mathrm{s}} \propto \frac{1}{log(Q/\Lambda_{\mathrm{QCD}})}$, where Q is the momentum transfer and $\Lambda_{\mathrm{QCD}} \sim 100$ MeV is the QCD scale.

The QCD Langrangian is given by:

$$\mathcal{L}_{QCD} \equiv \sum_i \bar{q}_{i,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} G^C_\mu - m_i \delta_{ab}) q_{i,b} - \frac{1}{4} F^A_{\mu\nu} F^{\mu\nu,A} \tag{2.1}$$

Where $q_{i,a}$ is a quark spinor of flavour i, mass m_i, and color $a = 1 \rightarrow 3$. t^A_{ab} $(A = 1 \rightarrow 8)$ are the $SU(3)$ generators also known as the Gell-Mann matrices. These are 3×3 matrices that satisfy:

$$[t^A, t^B] = i f_{ABC} t^C \tag{2.2}$$

f_{ABC} are called structure constants. G^A_μ is the gluon field associated with generator t^A_{ab}, g_s is the gauge coupling, and $F^A_{\mu\nu}$ is the gluon field tensor given by:

$$F^A_{\mu\nu} \equiv \partial_\mu G^A_\nu - \partial_\nu G^A_\mu - g_s f_{ABC} G^B_\mu G^C_\nu \tag{2.3}$$

2.2 Beyond the Standard Model

The Standard Model has been probed and tested in various ways. The experimental signatures it predict had been seen and measured to great precision. Yet, there is still empirical evidence that it does not tell the complete story. Phenomena such as neutrino oscillations, baryon asymmetry, existence of dark matter, and the mass hierarchy problem are examples to such evidences. A popular way to address these inconsistencies is by introducing complementary models - *extensions* - to the Standard Model. The Exotics, SUSY, and the Higgs BSM groups in ATLAS search the recorded data for detectable signatures of such models. The SUSY group focuses on Supersymmetry, a theory that predicts a ”partner” particle for each of the SM particles. The Exotics group is seeking all other BSM phenomena. The analysis presented here searches for BSM signals in events with two jets in the final states (dijets). A jet is the experimental signature of a parton, and it will be disscused in more detail later in this capter. One model with the potenial of producing dijets signatures, Z’, is reviewd in this section, but a variety of models such as excited quarks (q^*) [7], quantum black holes (QBH) [8], and color octet scalars ($s8$) [9] can also produce detectable dijets resonances [10].

2.2.1 Z' Dark Matter mediator

Dark matter particles are postulated particles that do not interact directly with Standard Model particles. They are motivated by a number of gravitational observations pointing to missing mass in the universe. Dark matter has not yet been detected directly and its existence is established through indirect observations such as the one mentioned above. Z' is a theorized massive particle, arising from a broken $\mathrm{U}(1)_X$ gauge symmetry, with the role of mediating between dark matter particles and SM states. In other words, dark matter can interact with standard model particles via exchange of Z'. Since dark matter particles do not leave a mark in the detector, they are commonly linked to events with missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). But if two SM particles can annihilate to create a Z' particle, the reverse process can occur, and the Z' can decay back into two SM particles. This sort of decay will produce a resonant signal in the mass of the mediator, which can be detected when looking for instance at pairs of jets, pairs of leptons, etc. In the specific scenario outlined in Ref.[11], dijet searches complement $E_{\mathrm{T}}^{\mathrm{miss}}$ searches in regions of parameter space where the decay of the Z' to invisible particles is kinetically suppressed. This model consider a DM fermion χ with mass m_χ, a spin-1 Z' mediator with mass M_{med}, and vector or axial-vector couplings to SM and DM fields. The corresponding interaction Lagrangians:

$$
\begin{aligned}
\mathcal{L}_{\mathrm{vector}} &= g_{\mathrm{q}} \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q}\gamma^\mu q + g_\chi Z'_\mu \bar{\chi}\gamma^\mu \chi \\
\mathcal{L}_{\mathrm{axial-vector}} &= g_{\mathrm{q}} \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q}\gamma^\mu \gamma^5 q + g_\chi Z'_\mu \bar{\chi}\gamma^\mu \gamma^5 \chi.
\end{aligned}
\tag{2.4}
$$

The parameters in the model are thus g_{q} - the coupling of Z' to quarks, g_χ- its coupling to the dark matter fermion, m_χ- the Dark Matter fermion mass, and M_{med} is the Z' mass. For the analysis in this work, the two most relevant parameters are g_{q}, as it drives the resonance height and its width, and M_{med}.

Figure 2.2 shows some examples of the relevant Feynman diagrams for processes involving the DM particles and SM fermions. Processes 2.2(a) - 2.2(c) can be witnessed in a collider. In 2.2(a) the Z' decays back into two quarks, which will create resonant signal in the dijets invariant mass spectrum. In 2.2(b) there are two additional leptons in the final state, as a result of interaction with a Z boson. 2.2(c) is a mono-jet signature where only the jet induced by the gluon is seen in the detector, and the pair of χ's are inferred from missing energy. The last two diagrams are complementary scenarios that can not be produced in a collider experiment, but may be looked for in dedicated DM searches. 2.2(d) shows a process of DM annihilation in which the resulting SM particles are quarks. Signatures from processes like these are mostly sought by indirect DM searches [12]. Direct searches hope to observe a scattering process between DM and SM particles[13], as suggested by the diagram in 2.2(e).

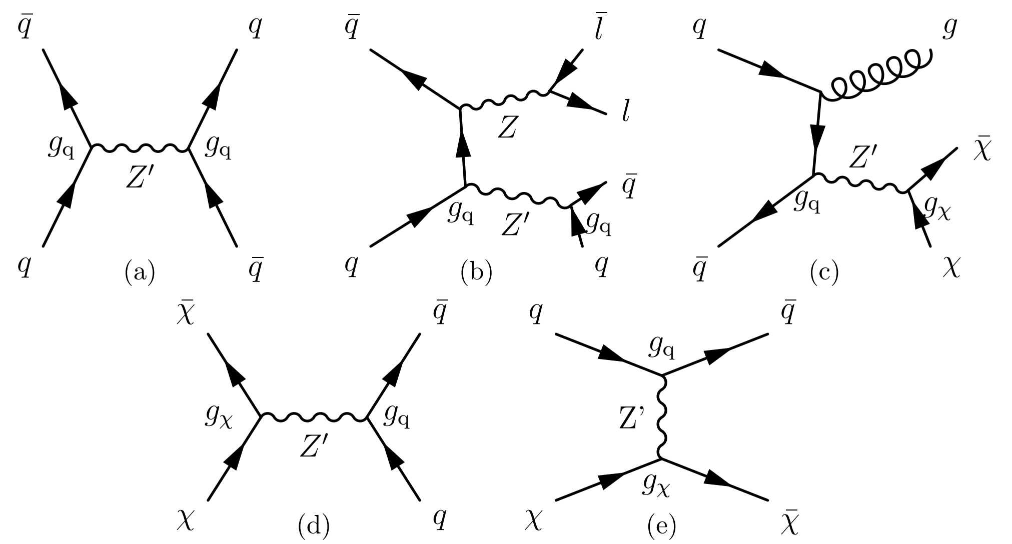


Figure 2.2: Processes involving interaction between SM particles, Z', and DM particles.

2.3 Collider Physics

2.3.1 Jets

As a result of the confinement, states with a color charge different from zero are not observed in nature[14]. This means that a single parton can not

be detected on its own. Instead the experimental signature of a parton is an object called a *jet* - a highly collimated group of energy depositions in calorimeter cells. Each parton has a finite probability of splitting into two partons, which are emitted in a small angle relative to the direction of the original parton. The splitting probability depends on the type of the original parton and the resulting partons ($g \rightarrow gg, g \rightarrow q\bar{q}, q \rightarrow qg$). These splitting are a perturbative effect which is referred to as *parton branching*. After many such splitting, the energy of the resulting partons decreases and it is followed by the step of *hadronization*. At this stage the partons are combined into states of color singlets thus forming *hadrons*. Hadronization is a process that can not be described perturbatively [15].
In ATLAS, jets are reconstructed from topological clusters (*topoclusters*). A topocluster is a group of neighbouring calorimeter cells, which exhibits significant energy depositions. The calorimeter cells are designed to monitor the shower of jets and measure their kinematic properties, which at first approximation matches those of the original parton [16].
Jets are objects that can be defined in several of ways. Each *clustering algorithm* is a well defined mathematical set of rules, meant to combine near by topological clusters to jets. The jets in this analysis were defined using the anti-k_{t} algorithm [17]. Anti-k_{t} belongs to a broader class of algorithms named *sequential recombination algorithms*. In these algorithms the distance between jet constituents is quantified using:

$$\begin{aligned} d_{ij} &= \min({p_{\mathrm{T}}}_i^{2p}, {p_{\mathrm{T}}}_j^{2p})\frac{\Delta R_{ij}^2}{\Delta R^2} \qquad , \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \\ d_{iB} &= {p_{\mathrm{T}}}_i^{2p} \end{aligned} \tag{2.5}$$

With $p_{\mathrm{T}i}$, y_i, and ϕ_i being the transverse momenta, rapidity, and azimuth of particle i. For the anti-k_{t} algorithm the parameter p=-1. The algorithm steps are as follows:

1. Compute all the d_{ij} and d_{iB} and find the minimum.
2. If the minimum is a d_{ij} merge i and j into a single element and return to stage 1.
3. Otherwise, if it's a d_{iB}, declare element i as a final jet, and remove it from list of remaining items. Return to step 1.
4. Repeat until no particles are left.

The fact that p has a negative sign means that the jets are formed around high momenta, *hard* particles, as they are classified first.

2.3.2 Kinematics of Proton Proton Scattering

The main physical process relevant to this analysis is $pp \to pp$ collisions. For these collisions, since the mass and kinetic energy of the incoming protons is equal, the centre-of-mass frame and the laboratory (or detector) frame coincide in the interaction point. However, since the proton is a composite particle, the actual interaction occurs between the parton (quark and gluon) constituents, which means that the colliding centre-of-mass frame differs from the laboratory frame. Regardless of the details of the interaction, some kinematic relation still hold [14]. In the scheme of $1 + 2 \to 3 + 4$, one can define the Mandelstam variables:

$$\hat{s} = (p_1 + p_2)^2 \quad , \quad \hat{t} = (p_1 - p_3)^2 \quad , \quad \hat{u} = (p_2 - p_3)^2 \tag{2.6}$$

where p_i is the four-momenta of parton i. For massless partons the Mandelstam variables satisfy: $\hat{s} + \hat{t} + \hat{u} = 0$ and two of those can be expressed as a function of the third one and the scattering angle θ^* in the centre-of-mass frame:

$$\hat{t} = -\frac{1}{2}\hat{s}(1 - \cos\theta^*) \quad , \quad \hat{u} = -\frac{1}{2}\hat{s}(1 + \cos\theta^*) \tag{2.7}$$

The rapidity of a parton is defined as $y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z})$, with E being the parton energy and p_z its momenta along the collision axis. This variable is invariant up to a constant under a boost in the z-direction. In the centre-of-mass frame, the rapidities of the outcoming partons, y^*, are opposite, and can be related to the scattering angle by:

$$\cos\theta^* = \tanh y^* \tag{2.8}$$

The Mandelstam variable $\hat{s}$ can be expressed using the transverse momentum p_{T} and rapidity in the centre-of-mass frame:

$$\hat{s} = 4p_{\mathrm{T}}^2 \cosh^2 y^* \tag{2.9}$$

One can also express y^* and the centre-of-mass rapidity $\bar{y}$ using the rapidities of the outcoming partons in the laboratory frame $y_{3,4}$.

$$\bar{y} = \frac{y_3 + y_4}{2} \quad , \quad y^* = \frac{y_3 - y_4}{2} \tag{2.10}$$

From the above relation one can derive that the scattering angle θ^* in terms of $y_{3,4}$:

$$\cos\theta^* = \tanh(\frac{y_3 - y_4}{2}) \tag{2.11}$$

2.3.3 Cross Section and Luminosity

In particle physics, the probability of different processes is often referred to in terms of cross section, σ. It is measured in units of *barn*, with 1 b $= 10^{-28} m^2$. But for a given process or phenomena, the cross section alone is not enough to determine the probability of witnessing it in a colliding experiment. For that the concept of luminosity, L, is introduced. L is defined through the rate R of events of some type and the cross section for them to happen.

$$R = L \cdot \sigma \tag{2.12}$$

One can deduce the total number of events in a given time interval δt simply by integrating the above:

$$N = \sigma \int_t^{t+\delta t} L dt. \tag{2.13}$$

which introduces the concept of integrated luminosity. For a given cross section for a new phenomenon, a higher integrated luminosity will thus increase the chances for discovery. Conversely, if the integrated luminosity is known, information on the cross section can be deduced from the number of events observed. N and L are experimental quantities, while σ is a theory parameter, containing the information about the modelling of the physics process — the information we are really interested in. Knowing the integrated luminosity is thus the key to make theory interpretations of an experimental event count, and it is given in units of fb^{-1} or pb^{-1}.

Chapter 3

LHC and the ATLAS detector

3.1 LHC

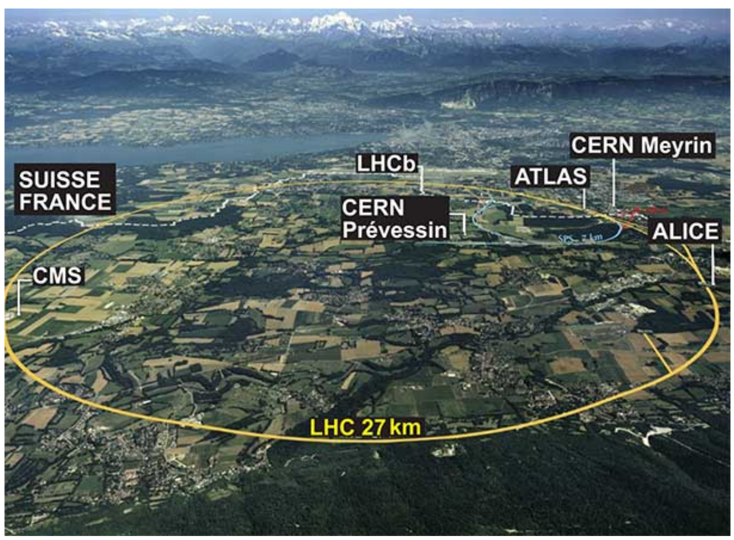


Figure 3.1: Aerial view of CERN with an illustration of the LHC ring.

The Large Hadron Collider (LHC) is a ring collider, designated to collide protons (and heavy ions) with high center-of-mass energy, and located in CERN, on the border of Switzerland and France. The LHC ring is 27 km, inside an underground tunnel of varying depth between 45 - 170 meters. An illustration of LHC with respect to the surroundings is given in Figure 3.1.

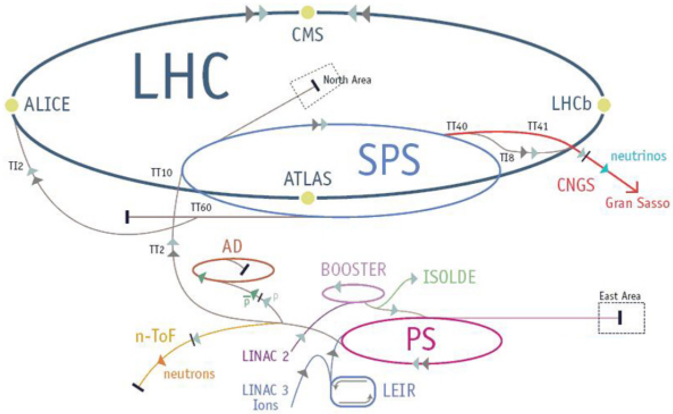


Figure 3.2: An illustration of the steps taken in accelerating the protons to their final energy and injecting them to the LHC.

The process of accelerating and colliding protons is done in a number of steps and is demonstrated in Figure 3.2: Hydrogen atoms are stripped from their electrons, thus becoming protons. They then go through a series of accelerators, from each they emerge more energetic. First is Linac2, a linear accelerator that brings the protons to ∼50 MeV. They are then injected into the Proton Synchrotron Booster (PSB) where they reach an energy of 1.4 GeV. From there the protons reach the Proton Synchrotron (PS), bringing them to 26 GeV, then to the Super Proton Synchrotron, which increases their energy to 450 GeV. Finally the protons are injected in bunches to the LHC ring, where their energy is further increased to 6.5 TeV. In this last step the bunches of protons are injected into the LHC in both directions, clockwise and anticlockwise. Their paths meet in four different points along the ring, where the collisions occur. Around each such point is a detector, set to measure and analyse the products of the collision.

The four detectors are ALICE[18], ATLAS[19], CMS[20], and LHCb[21]. ALICE experiment is set to probe *heavy-ion* physics, LHCb focuses on *b-physics*, and both ATLAS and CMS are multipurpose detectors designed to find the *Higgs boson* and to provide high precision measurements of proton-proton, heavy ions, and proton-heavy ions collisions.

LHC started operating in 2010 with a centre-of-mass energy of 7 TeV. During Run-1 in 2012, the proton beams had energy of 4 TeV (centre-of-mass energy of 8 TeV) and bunch spacing of 50 ns. Run-1 was followed by a two years shut down in which the LHC underwent an upgrade, and when Run-2 started in May 2015, it provided with beams of 6.5 TeV at 25 ns bunch spacing.

3.2 The ATLAS Detector

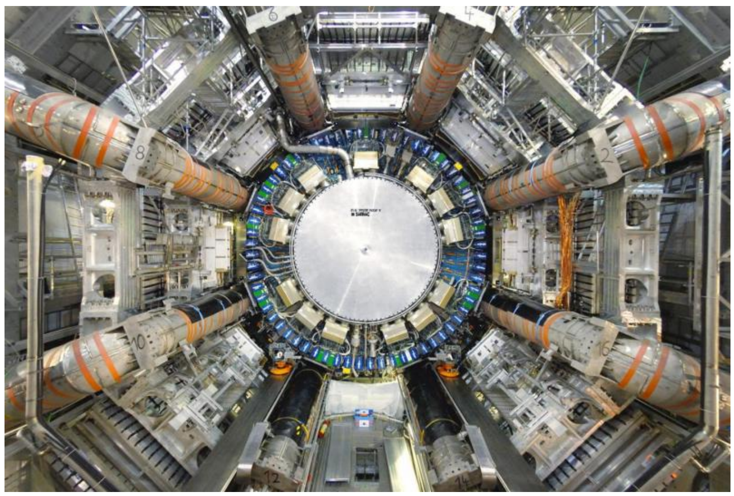


Figure 3.3: The ATLAS detector at CERN.

This section outlines a reviews of the ATLAS detector, mainly based on information drawn from [19]. Other sources are cited in the text.

The ATLAS detector, shown in Fig. 3.3, is a 25 meter high and 45 meter long symmetric cylinder centred around the interaction point. It is comprised of several sub-detector systems, each using different techniques to detect and measure the properties of particles moving through it. The four major sub detector systems in ATLAS are the Inner Detector (ID), the calorimeter system, the muon spectrometer, and the magnet system. They are schematically depicted in Fig. 3.4. Each sub detector system is sensitive to different types of particles, and the role each of the sub detector system plays in the measurement is illustrated in Figure 3.5; Charged particles, like electrons and protons, interact with material in the ID, while neutral particles pass through it leaving no detectable signs. The calorimeter system uses two techniques to measure the energy deposited by the particles passing through it; Photons

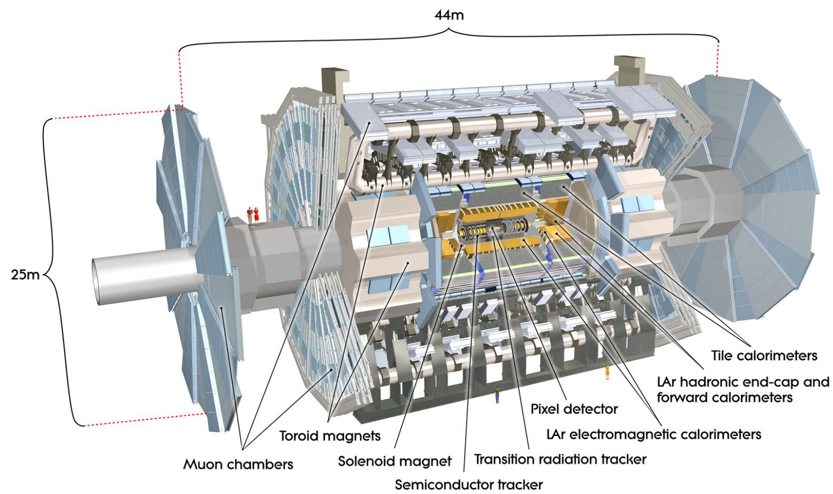


Figure 3.4: A schematic view of ATLAS and it's sub systems.

and electrons give birth to EM showers in the EM calorimeter, and Hadrons are detected in the Hadronic calorimeter. Muons barely interact with any of the material in the ID and calorimeters. They make it all the way out to the muon spectrometer where they can be detected. Neutrinos escape undetected as they almost do not interact with the material. Their presence can be deduced from non-conservation of the total transverse momentum in a collision. Hence the energy carried by neutrinos in a collision is denoted by $E_{\mathrm{T}}^{\mathrm{miss}}$.

Before looking with more detail into the different sub detectors, let us review the coordinate system used in the measurements of all properties.

3.2.1 Coordinate System

ATLAS uses a right handed coordinate system with its origin in the middle of the detector at the interaction point, and its z-axis along the beam pipe. The x-axis is directed towards the centre of the LHC ring, and the y-axis is pointing upwards. It is possible to describe a point in space using a cylindrical (r, ϕ, z) set of coordinates, with $r = \sqrt{x^2 + y^2}$ and $\phi = \arctan(y/x)$. More often the pseudorapidity η is being used instead of the polar angle θ.

$$\eta = -\ln(\tan(\theta/2)) \tag{3.1}$$

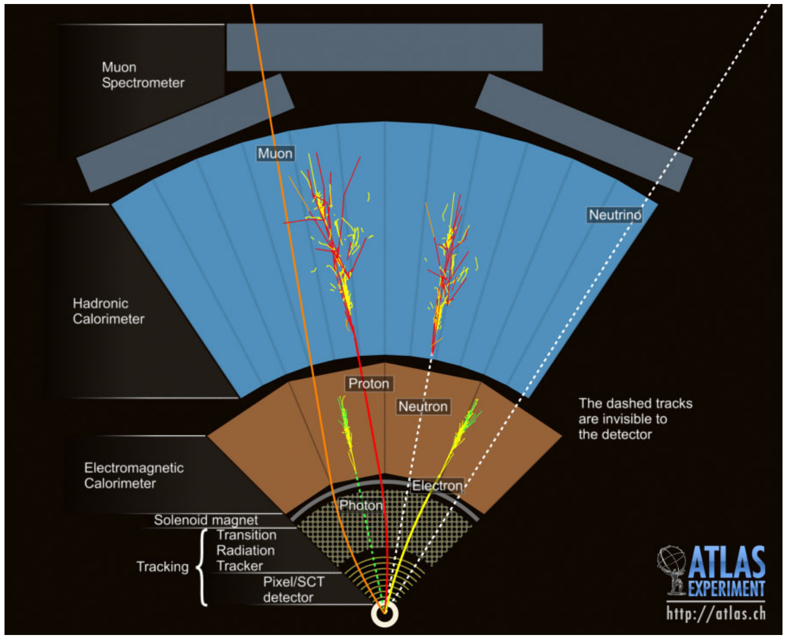


Figure 3.5: The principals of each sub system in ATLAS.

Where $\theta = \arccos(\frac{z}{\sqrt{r^2+z^2}})$. For relativistic particles the pseudorapidity is a good approximation for the rapidity of the particle y, defined in Sec. 2.3.2.

3.2.2 Inner Detector

The inner detector[22] is the closest sub-detector to the interaction point. It consists of three sub-detectors exploiting different techniques for particle registration: the pixel detector, the semi-conductor tracker (SCT) and the transition radiation tracker (TRT). They are used for the reconstruction of trajectories (tracks) of charged particles, as well as the position of an interaction (vertex), and for electron identification. To measure the particle momenta based on the curvature of the reconstructed tracks, the ID tracking system is embedded in a 2 T solenoidal magnetic field. A more detailed description of the magnets system in ATLAS will be given later in this section.

Pixel Detector The pixel detector comprises four layers in the barrel and three disks in the end-cap region. In total, there are 1744 sensors with 46080 read-out pixels each, yielding approximately 80 million read-out

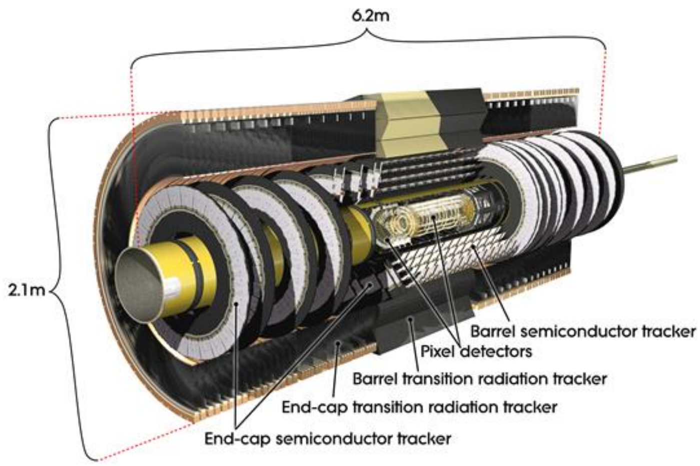


Figure 3.6: The ATLAS Inner Detector.

channels. The pixels have a minimum size of $50 \times 400\ \mu\mathrm{m}^2$. When charged particles traverse the silicon sensors, many electron-hole pairs are created. Due to an electric field which is applied across the sensor, electrons and holes drift in opposite directions. The signal is read out individually for every pixel. The pixel detector was originally designed with only three layers in the barrel. The fourth layer, named Insertable B-Layer (IBL), was installed as part of an upgrade to the detector in the long shutdown during 2013/2104 [23]. It is the innermost layer at a distance of 3.2 cm and it provides improved tracking performance and improved vertecing resolution near the interaction point.

SCT - Semi Conductor Tracker The SCT is a silicon strip detector and thus based on a similar principle of operation as the pixel detector. It also consists of cylindrical layers in the barrel region. There are four of these layers and nine additional disks in the end cap regions on both sides. One SCT module consists of 768 strips which are 6.4 cm long and separated by 80 μm from each other. The strip detector can only deliver a precise measurement in one dimension. This is extended to two dimensions by gluing two modules backwards to each other where one module is rotated by 40 mrad. The resolution which can be reached is 17 μm in the R-ϕ direction and 580 μm in z-direction. In total the SCT has 6.3 million read-out channels.

TRT - Transition Radiation Tracker The TRT is the most outer part of the ID. It provides a large number of tracking measurements (typically > 30 hits per track), good pattern recognition, and it contributes to identify electrons and distinguish them from other particles. The detector consists of straw tubes filled with a gas mixture based on Xe, CO_2, and O_2. In the middle of the straw there is a tungsten wire which is coated with gold. When particles move through the TRT they are ionizing the gas and the freed electrons move to the wire of a high potential. The measured time it takes to reach the wire, determines the distance of the track to the wire to a precision of 130 μm. The TRT identifies electrons by measuring the amount of current collected by the wires in the tubes. As all charged particles, electrons emit radiation when they cross a boundary between two media of different refractive index. This radiation is called *transition radiation*, and is proportional to $\gamma = E/m$, which is much higher for electrons compared with hadrons, due to their small mass.

3.2.3 Calorimeters

The goal of a calorimetr is to measure the energy deposited by particles passing through it. ATLAS calorimeter system has two major components: the electromagnetic (EM) calorimeter, and the hadronic calorimeter. Both parts use the concept of a sampling calorimeter, in which active and passive detector material alternate. The passive materials (absorber) are interacting heavily with the passing particles which starts the showering process. The active layers on the other hand are detecting the produced particles and measure their energy. For an accurate energy measurement, it's important that all interacting particles will be fully absorbed in the calorimeter material, and deposit their entire energy. The resolution of the EM calorimeter can be parameterized:

$$\frac{\sigma(E)}{E} = \sqrt{(\frac{a}{\sqrt{E}})^2 + (\frac{b}{E})^2 + c^2} \tag{3.2}$$

The first term is a stochastic term, dependent on the choice of absorber and active materials and the thickness of the sampling layers, the second term is due to electronic noise, and the third term originates from local non-uniformities in the response of the calorimeter. The incident particle energy resolution of the ATLAS calorimeter, neglecting the noise term which is subject to large variations depending on the LHC pile-up conditions, is shown in Table 3.1. The coverage of the ATLAS calorimeters spans to the region of $|\eta|$ <4.9.

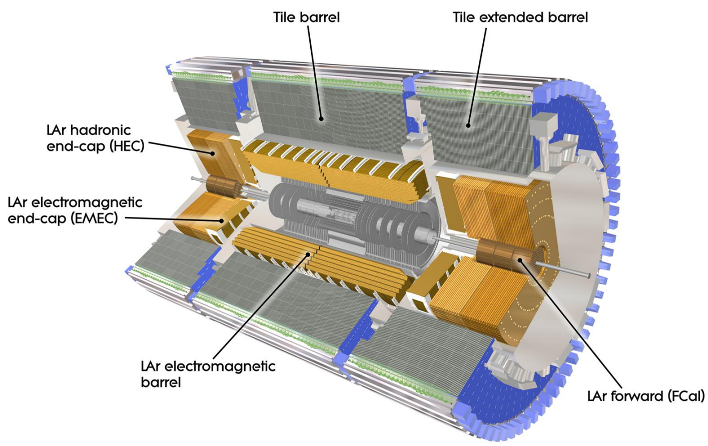


Figure 3.7: The ATLAS Calorimeter.

EM Calorimeter The EM calorimeter uses lead as the absorber and liquid argon (LAr), cooled to 87 K, as the detecting material. In the barrel region the coverage of the EM calorimeter is $|\eta| < 1.475$. It has two half-barrels that extend in radial direction from 2.8 m to 4 m. Each half-barrel consists of 16 modules, with each module covering an angle of $\Delta\phi = 22.5°$. The end-cap region covers $1.375 < |\eta| < 3.2$ and is made out of two coaxial wheels which are divided into eight modules. The EM calorimeter uses two main processes to measure particles energy: For photons it is e^+e^- production, and for electrons and positrons it is s emission of extra photons through bremsstrahlung radiation. One can describe these showers using the *radiation length* X_o. One X_o is the mean path an electron has travelled before losing all but $\frac{1}{e}$ of its initial energy to bremsstrahlung. It is equivalent to ~7/9 of the mean free path an average photon above threshold travels, before it undergoes pair production. The thickness of the electromagnetic calorimeter can be given in units of X_o. The barrel region has a thickness of 22 X_o, which corresponds to an inner radius of 1.15 m and an outer radius of 2.25 m, while the end-cap region is 24 X_o thick (6.65 m in each direction along the beam axis).

Hadronic Calorimeter The hadronic calorimeter surrounds the EM calorimeter. It is designed to measure the energy of hadrons, in particular the neutral ones which do not leave any track in the inner detector. The hadronic calorimeter also prevents hadrons from penetrating through and arriving to the muon spectrometer. The physics of the hadronic shower makes hadronic calorimetry much more complicated than electromagnetic calorimetry. The lower momentum transfer, soft QCD interactions between hadrons and nucleons are very difficult to model, meaning the simulations of hadronic calorimeter are typically less accurate. The presence of missing energy also add to the difficulty of modelling hadronic calorimetry. Either the energy scale of the hadronic calorimeter must be independently determined, or the calorimeter must be built specially to compensate for this lost signal. The ATLAS hadronic calorimeter is non-compensating. Hadronic energy loss is typically quantified by the *absorption length*, λ_0, defined as the mean distance required to reduce by a factor $\frac{1}{e}$ the number of incoming relativistic hadrons travelling through a medium. This is typically $\mathcal{O}(10)$ times longer than the radiation length in a medium, requiring hadronic calorimeters to be very thick. ATLAS features three hadronic calorimeters sub systems, that use different techniques and/or materials depending on the respective detector region.

The **Tile Calorimeter** uses steel absorbers and scintillating polystyrene tiles as active material. A charged particle will excite some of the atoms in the scintillator, these excited atoms will emit photons of particular wavelengths. The scintillation radiation is the signal that is measured. The tile calorimeter consists one central barrel, covering $|\eta| < 1.0$ and on each side, an extended barrel, covering $0.8 < |\eta| < 1.7$. The average radial depth is $7.4\lambda_0$, and its main purpose is to measure the energy and directions of hadronic jets.

The **Hadronic End-Cap (HEC)** calorimeter is built out of copper as passive material and LAr as active material. It is located inside the hadronic tile calorimeter extended barrels, has a depth of $\sim 12\lambda_0$, and covers $1.5 < |\eta| < 3.2$.

The **Forward Calorimeter (FCal)** also uses LAr as active material. It consists of three layer, and uses copper for an absorber in the first layer, and tungsten in the outer two layers. The FCal is located inside the endcap cryostat, in a region of high radiation level. It covers $3.1 < |\eta| < 4.9$ and is

Calorimeter Segment	Energy resolution σ_E/E
EM Barrel	$\frac{10\%}{\sqrt{E}} \oplus 0.7\%$
Tile Calorimeter	$\frac{50\%}{\sqrt{E}} \oplus 3\%$
HEC	$\frac{50\%}{\sqrt{E}} \oplus 3\%$
FCal	$\frac{100\%}{\sqrt{E}} \oplus 10\%$

Table 3.1: Energy resolution in the ATLAS calorimeter systems

$\sim 10\lambda_0$ long.

3.2.4 Muon Spectrometer

The outermost detector system of the big ATLAS cylinder is the muon system. The only hint that signal in these chambers comes from a muon is the simple fact that other charged particles are rarely expected to make it as far out in the ATLAS detector system. Aside from the measurement of the muon momentum the muon spectrometer is participating in the trigger system. For that two different types of detectors are utilized: fast detectors for the trigger system and precise detectors to measure the muon spatial position accurately. The barrel part consists out of 3 layers at 5 m, 7.5 m and 10 m from the beam line, and is divided into 16 azimuthal sections. The two end-caps are built out of three wheels on each side at 7.4 m, 14 m and 21.5 m from the interaction point. The muon spectrometer has a number of different types of sensors in order to measure muon tracks, and magnetic field is required for momentum measurements and is delivered by the toroid magnets.

The **Monitored Drift Tubes (MDTs)** consists of three to eight layers of aluminium tubes, 30 mm in diameter, filled with 93% Ar and 7% CO_2 gas mixture at a pressure of 3 bar. It provides precision tracking measurements in region of $|\eta| < 2.7$. A single MDT has a resolution of 80 μm, but a chamber with few layers of tubes reaches a resolution of 35 μm.

The **Cathode-Strip chambers (CSC)** are used in the inner-layer in the forward region ($2.0 < |\eta| < 2.7$) due to their higher rate capability and time resolution. They as well provide precision tracking measurements. The

CSC are multi-wire proportional chambers, composed out of radial anode wires, and cathode strips perpendicular to the wires. The CSCs have a precision of 40 μm in the bending plane and 5 mm in the non-bending plane. The precision-tracking chambers are complemented by a system of fast triggers chambers capable of delivering track information within a few tens of nanoseconds.

The **Resistive Plate Chambers (RPC)** provide triggers in the barrel region, where $|\eta| < 1.05$. The RPC chambers, arranged in towers in three concentric barrels, are composed of units filled with a gas mixture of $C_2H_2F_4$, isoC_4H_{10} and SF_6. The plates are coated from the inside with graphite and connected to the high voltage system.

The **Thin Gap Chambers (TGC)** provide triggering in the endcaps region, where $1.05 < |\eta| < 2.4$. These are multi-wire proportional chambers. The chambers are filled with a gas mixture of CO_2 and nC_5H_{12}. Inside each chamber there are anode wires, made of gold plated tungsten, and cathode strips which are perpendicular to the wires. When a charged particle will pass through the gas, it will ionize it and create a negative (positive) signal in the wires (strips) which is measured.

3.2.5 Magnet System

ATLAS uses two magnet systems, designed to generate fields which curve the tracks of EM charged particles moving in the detector. The bending radius of a charged particle trajectory is proportional to its velocity in the direction perpendicular to the magnetic field. By considering the mass of the particles, the momentum in that direction can be determined.
Fig. 3.8 shows a sketch of the magnet system geometry; Surrounding the inner detector is a superconducting solenoid magnet, aligned with the beam axis and provides a 2 Tesla axial magnetic field for the inner detector. The second magnet system consists of three air-core toroids, one in the barrel region and two in the end-cap region. These magnets produce a toroidal magnetic field of approximately 0.5 Tesla and 1 Tesla for the muon detectors in the central and end-cap regions respectively.

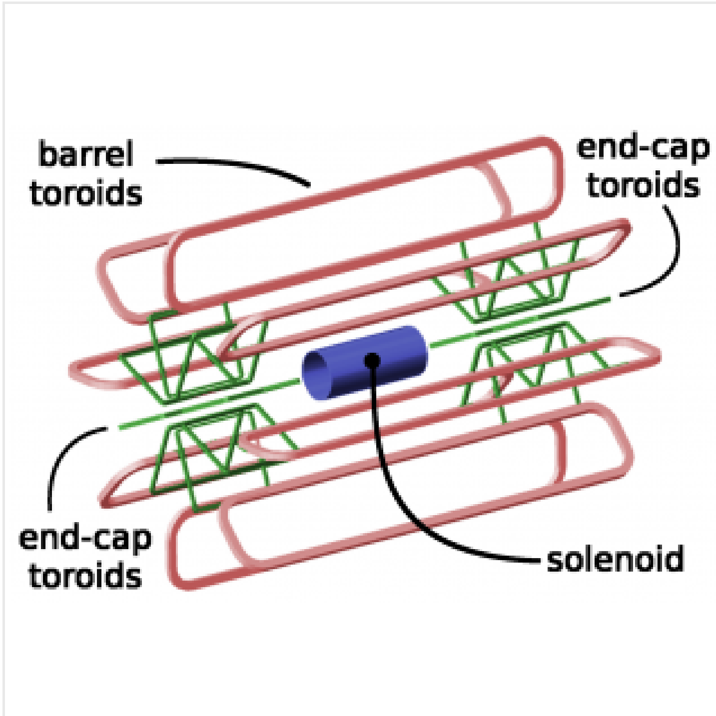


Figure 3.8: The ATLAS Magnets System.

3.3 The ATLAS Trigger System

At the high luminosity intensity of the LHC the trigger plays an important role. With a bunch spacing of 25 ns, ATLAS generates data at a rate of 40 MHz, corresponding to ∼40 GB/sec [24]. This is vastly more data that can be written to disk and the majority of the data will contain information of no interest. The trigger system decides, in real-time, whether to record or reject the data of a collision. The triggered and recorded information is called an event. By selecting events passing predefined hardware and software criteria, the recording rate is reduced to ∼1 kHz. There are several channels of potential trigger chains which run in parallel. For instance events with deposited energy above predefined energy threshold may be saved for further analysis. In a specific data analysis, the triggers are assigned in order to select the events that are interesting with respect to the phenomenon one is interested in.

For RUN-1 the ATLAS Trigger was composed of three parts: the Level-1 trigger (L1) which is hardware-based, and the Level-2 (L2) and Event Filter (EF) which are implemented in software. The two software triggers were replaced in Run-2 by a single High-Level-Trigger (HLT).

Level-1 Trigger L1 is built with fast custom electronics in order to get a latency of less than 2.5 μs, reducing the rate to a maximum of 100 kHz. From the information of the calorimeters and muon tracks, the L1 trigger identifies physics objects and classifies them to: jets, electrons/photons, muons, and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). Geometrical areas in the detector with the presence or multiplicity of certain signals are flagged as Regions of Interest (ROI) and are processed in the next trigger level.

High-Level-Trigger The HLT is a computing system connected by fast dedicated networks. Sophisticated selection algorithms are applied to the ROIs of L1, or to the full-event information, within an average processing time of 200 ms. The output rate is reduced in the HLT from the 100 kHz of L1, to an output rate of approximately 1kHz. All events passing the HLT are written to disk.

Chapter 4

Trigger Level Analysis

The p_{T} distribution of jets produced at the LHC pp collisions is steeply falling. This means that high-energy jet events are very rare compared to low-energy ones, and jet triggers at low energy threshold fire much more often than the ones at higher threshold. But the high energy events are often of much more interest for many BSM searches. In order to reduce the relative triggering rate of less interesting events, allowing more bandwidth for the ones considered more interesting, the ATLAS trigger strategy is to apply prescale factors on triggers with low energy threshold. A prescale factor p means that only one in every p triggered events is actually recorded. A trigger with prescale factor $p = 1$ is called un-prescaled. To properly correct for the trigger prescale, the corresponding number of events is multiplied by p. The statistical precision is however smaller than that achieved by recording the full set of events. This means a loss of sensitivity to new phenomena at lower p_{T}, implying that it will take more data (longer time) to discover them. Thus far dijets searches in ATLAS have been focusing on events with high p_{T} jets and large prescale factors had been applied for triggers in the low p_{T} regime. The result is a region of low masses which is not being probed for new phenomena. For spin-dependent interactions between quarks and dark matter, light masses are favored due to thermal relic density constraints [11]. For instance, Figure 4.1 shows parameter region excluded from theory, compared to constraints from different searches, for the Z' model discussed in Sec. 2.2. The constraints are presented as a function of the DM fermion mass and the mediator mass, given as m_χ and M_{R} respectively, and considering axial-vector couplings to both quarks and χ with $g_{\mathrm{q}} = \frac{1}{6}$ and $g_\chi = \frac{3}{2}$. The plot shows that for $M_{\mathrm{R}} < 1$ TeV there is still an unexplored parameter space for detecting such a model. This mass region is not covered by the 2015 dijets search, in whcih the invariant mass distribution starts from 1.1 TeV [25]. This demonstrates the importance of reaching lower dijet masses, and leaving no stone unturned.

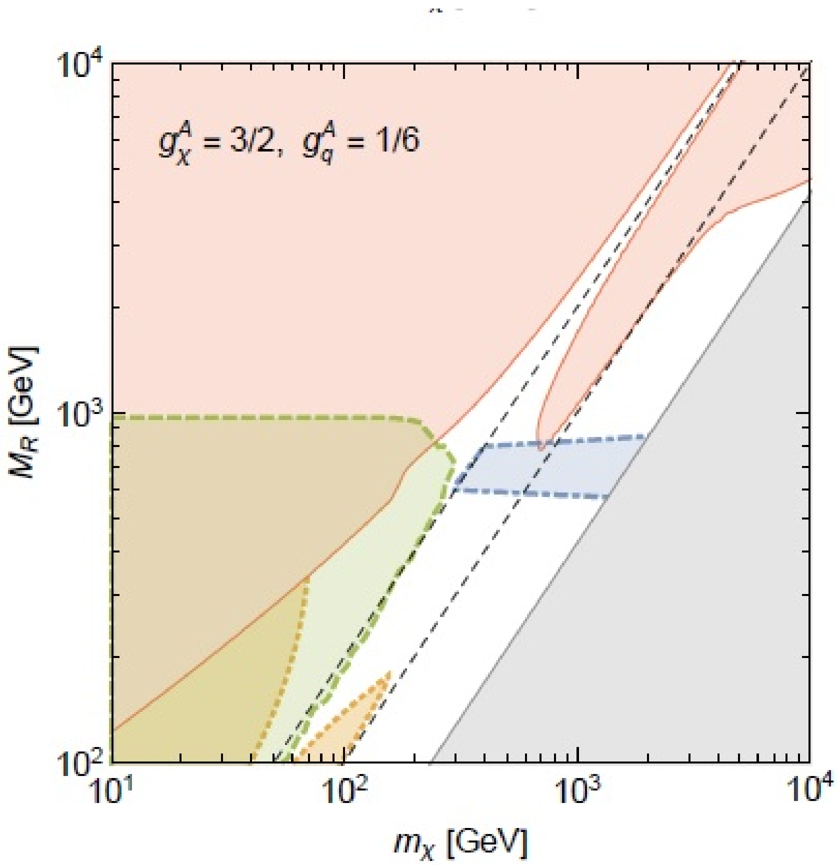


Figure 4.1: Combined constraints from direct detection experiments (orange, dotted), searches for monojets (green, dashed) and dijets (blue, dot-dashed), as well as constraints based on theoretical arguments (red, grey). This figure is taken from ref. [11].

The Trigger-Level-Analysis was formed to address this very subject. It gains sensitivity at lower masses by using un-prescaled low p_{T} triggers. In order to record these events within the available bandwidth, only partial events, containing the small subset of information needed for a dijet resonance search, are saved to disk.

This chapter contains a detailed overview of the Trigger-Level-Analysis, with the necessary steps that were needed to implement this method in Run-2 of the LHC.

4.1 The Data Scouting Stream

The trigger bandwidth limitation described above can be avoided by recording only partial event information from the HLT. For 2015 the **Data Scouting (DS)** stream [26] has been integrated into the ATLAS TDAQ system. The following chapters uses the terminology:

- **Data Scouting (DS):** The technology used to stream HLT reconstructed objects via partially built events.
- **Trigger Level Analysis (TLA):** The full analysis strategy and chain, i.e. the usage of HLT objects (recorded via Data Scouting) in physics analysis.

The DS stream records the event header and all HLT jet objects above a p_{T} threshold[1] for all events firing the L1_J75[2] trigger. For each HLT jet, the stream records the jet four-momentum and a set of variables characterizing the detector information within the jet (such as the jet timing, quality variables, and limited information about the structure of the jet). The stream does not record any readout of individual calorimeter elements, nor does it record information from the tracking or muon detectors. For 2016 it is planned to have trigger L1_J100 seeding the DS stream, as the instantaneous luminosity is expected to be higher. The 2015 average DS event size, only containing the jet and minimal event information as detailed above, was 30 kB/event. It can be reduced to 10 kB/event when placing a cut of 20 GeV on the trigger jets that are saved to storage, as planned for 2016.

Figure 4.2 depicts a comparison between the rate of the Data Scouting trigger to the combined rate of single jet triggers, demonstrated here during typical two hours of data taking in standard conditions. This show the

[1]In 2015 lower-than-expected luminosity allowed operation with the lowest possible threshold of 4 GeV, while in 2016 the jet threshold will be 20 GeV.

[2]The naming convention of ATLAS trigger is *TriggerLevel_TriggerObjects_*p_T *Thresold*. For example, in this particular trigger, the first part of the name suggests it is a L1 trigger, the 'J' and '75' indicates that it is triggered by events with jets which have p_T of at least 75 GeV.

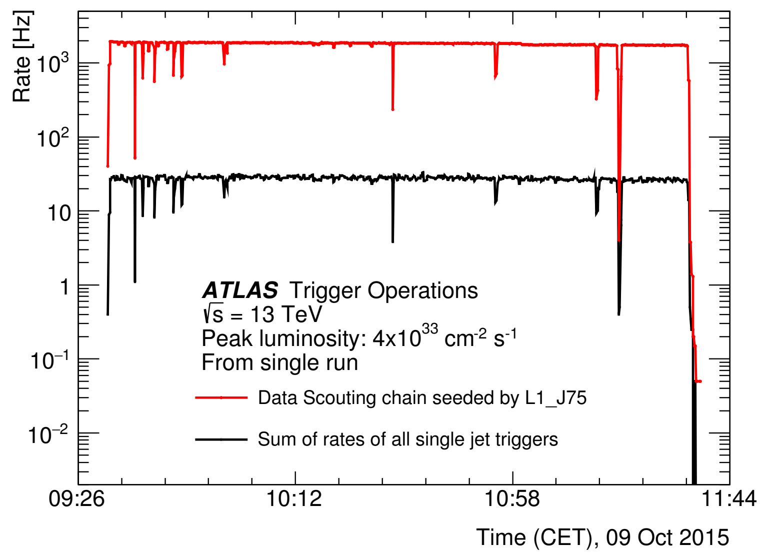


Figure 4.2: Trigger rate for the data scouting chain seeded by the Level-1 trigger J75 compared to the sum of the rates of all prescaled and unprescaled central single jet triggers, during a time range of a single run, taken on Oct 9, 2015. Overlaps in the rate of the single jet triggers are considered negligible.

increase in statistics achievable using this data taking strategy. The single jet rate is much lower than the Data Scouting rate due to the prescales applied to remain within the overall bandwidth budget if recording the full event.

4.2 Performance of Trigger Jets

To make use of the DS data, the performance of the objects reconstructed at the trigger level need to be comparable to that of objects reconstructed offline. The definition of performance includes reconstruction efficiency, jet identification (cleaning), calibration and corresponding estimates of fake rates, energy scale resolution and uncertainties.

This work does not go into full detail in describing the performance studies of the analysis. Broader discussion and explanation is given in [27].

Reconstruction of Trigger Jets In Run-2, trigger jets and offline jets share the same calorimeter inputs (topoclusters), and the jet reconstruction software is the same in both cases. Still, there are differences in the calorimeter reconstruction that mostly affect low-p_T jets:

- A **Bunch Crossing IDentification (BCID) offset correction** is applied to calorimeter cells offline, accounting for out-of-time pile-up effects.

- A difference in the **energy and timing calculation for the Tile Calorimeter** is present due to a different algorithm for online and offline calculation.

Trigger Jet Identification and Cleaning Jet-like signature in the detector can arise from processes other than the *pp* collision. These processes can be sporadic noise bursts, beam induced background or cosmic rays. The DS stream contain calorimeter information from the events, which helps in identifying $fake$ jets. This is known as the *jet identification (cleaning)* procedure. The cleaning procedure in the TLA is closely related to the offline procedure, outlined in [25], with the exception that the track-based **charged fraction** (CHF) variable is not available for events in the DS stream. This is not a concern as the effects of jet cleaning are minimal for this analysis. Noisy cells, beam background and cosmic rays usually give rise to single-jet events. The dijet topology is already effective in rejecting most of these events, as they will not contain two central jets with a small rapidity separation. In any case, the impact of not having the CHF cut is cross checked in the analysis and found to be negligible [27].

Trigger Jet Calibration Jets are reconstructed from multiple signals in the detector, combined into a single object by a set of mathematical rules. Energies of reconstructed jets therefore do not necessarily reflect the true energy of the original partons. Differences can arise from the details of the clustering algorithm, non-responsive parts of the detector, energy that falls outside of the calorimeter or cells with energy below threshold, and discrepancies between the energy measurement of electromagnetic and hadroinc showers, where the hadronic calorimeter does not properly account for the latter[19]. The jet calibration is applied to bring the jets as close as possible to the energy of the original partons.

Trigger jets can not pass the same calibration chain as offline jets since some of the steps rely on detector information unavailable in the DS stream. Whenever available, calibration constants are derived using *trigger* jets quantities. The missing items in the calibration of trigger jets are:

- **Residual pile up correction**, which is based on the number of primary vertices in an event (NPV), and the expected number of additional collisions in the same crossing (μ). Both variables are not recorded in the DS stream.

- **Jet origin correction** which accounts for the jet position, not it's energy.

- **Global Sequential Correction (GSC)** which improves the jet resolution.

To compromise for missing the above steps, an additional data-driven correction is applied to trigger jets, bringing them to the offline scale. This correction is applied prior to the final in-situ correction.

After calibration, the offline Jet Energy Scale (JES) uncertainties are applied to trigger jets, with additional corrections to account for the differences mentioned above.

Figure 4.3 show the invariant mass distribution of trigger jets compared to offline jets after applying the jet calibration procedure. Events in the plot are selected by the HLT_j110 single jet trigger and are required to pass the TLA selection cuts (with $|y^*| < 0.6$). The average m_{jj} response of trigger jets compared to offline jets is shown in black and is within 1% of unity.

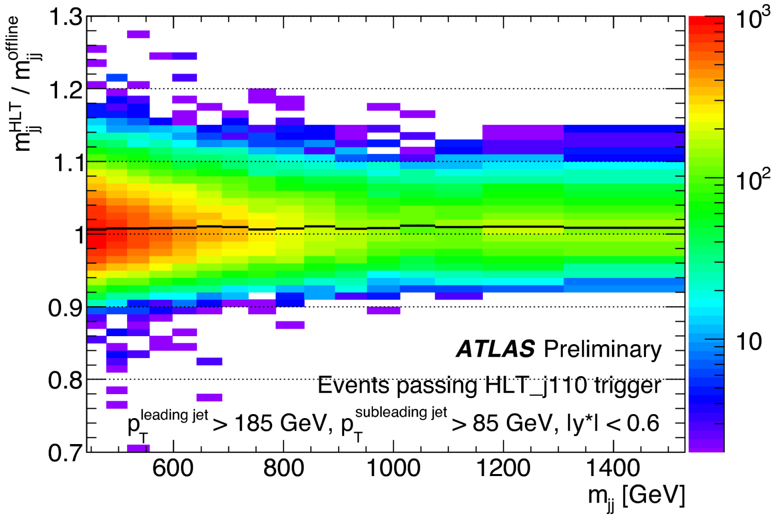


Figure 4.3: m_{jj} of trigger jets compared to offline jets, after applying jet calibration. Events are selected by the HLT_j110 single jet trigger and are required to pass the TLA selection cuts (with $|y^*| < 0.6$).

Chapter 5

Data and Background Simulation

5.1 Data

This analysis uses two separate data stream. The Data Scouting stream mentioned above is used for the search, and an additional Physics Main stream is used for calibration and cross checks of the trigger jet performance.

Physics Main Stream This stream is used by all ATLAS analysis. It is seeded from several triggers in the physics menu, and contains full event information, including jets reconstructed both online at the HLT level (*trigger jets*), as well as offline (*offline jets*) within the same event. A subset of these events selected by a single jet trigger are those used to cross-check the performance of trigger jets against offline jets.

Data Scouting Stream As mentioned in section 4.1, the DS stream contains trigger jets seeded from HLT, which are used to create the m_{jj} distribution used in the search phase of the analysis. The DS stream recorded an integrated luminosity of 3.4 fb^{-1} during 2015, with the LHC operating at 25 ns bunch spacing.

5.2 MC simulation

5.2.1 QCD Background

This analysis uses simulation of QCD processes at leading order (LO) in perturbative QCD. These are used for comparisons of jet quantities and for tests of the techniques applied to the data to obtain the final background estimate. The LO samples, from the Pythia8 [28] event generator, use the A14 tune [29] and NNPDF2.3 LO PDF set [30], along with simulated pile-up events corresponding to the 25 ns bunch spacing.

The *un-prescaled* nature of this analysis yields very high statistics, especially in low mass. This require generating equivalent statistics in MC, which using the Pythia8 standard generation scheme is beyond the ATLAS resources. Pythia8 provides a *Power Law* option to make event-level generation of a large number of events more efficient. This is done through oversampling of $2 \rightarrow 2$ QCD events by a factor $\hat{p_{\mathrm{T}}}^{POW}$, where $\hat{p_{\mathrm{T}}}$ is the transverse momentum in the rest frame of the hard subprocess and POW is an input parameter. Events are than assigned a compensating weight, the inverse of this, which is used when plotting histograms of event properties. In the generation of events for this analysis the power was set to $POW = 5$. 2.07 billion events were generated, and the gain in CPU time achieved with this method is by a factor of approximately 100.

5.2.2 Statistical Power of Samples

In order to derive the statistics of a background sample, we define the effective number of entries, N_{eff}, as the number of events that would produce the observed (i.e. *Poisson*) uncertainty in a bin assuming all events have identical weight w_{eff}. If N is the bin content and δN is it's error, then:

$$N_{eff} = ((N)/(\delta N))^2 \tag{5.1}$$

$$w_{eff} = N/N_{eff} \tag{5.2}$$

N_{eff} is the figure of merit for deciding whether a background sample is sufficient to reach the statistical power expected in data. One can also represent this in terms of equivalent precision for unweighted events. N_{eff} is the number of unweighted events needed to have the same precision as the weighted

Monte Carlo sample. If $N_{eff} > N$, where N is the number of events scaled to a given luminosity, then the precision is sufficient to properly simulate the desired luminosity.

The 2.07 billion of truth-level events generated for testing the background estimation corresponds to effective statistics for 3.4 fb^{-1} starting from dijet invariant masses of 478 GeV. To better represent QCD nature, *data-like* m$_{jj}$ spectra can be made by drawing numbers from a Poisson distribution around bin values of a weighted MC background. Fig. 5.1 shows the number of effective entries, the data-like distribution (in the region where the effective statistics is sufficient), and weighted background, expected at truth level for 3.5 fb^{-1}. The lack of effective events in certain bins is due to events with very large weights. They could be removed, but there is no reason to do so as they are outside the region where the background estimation is tested using this sample.

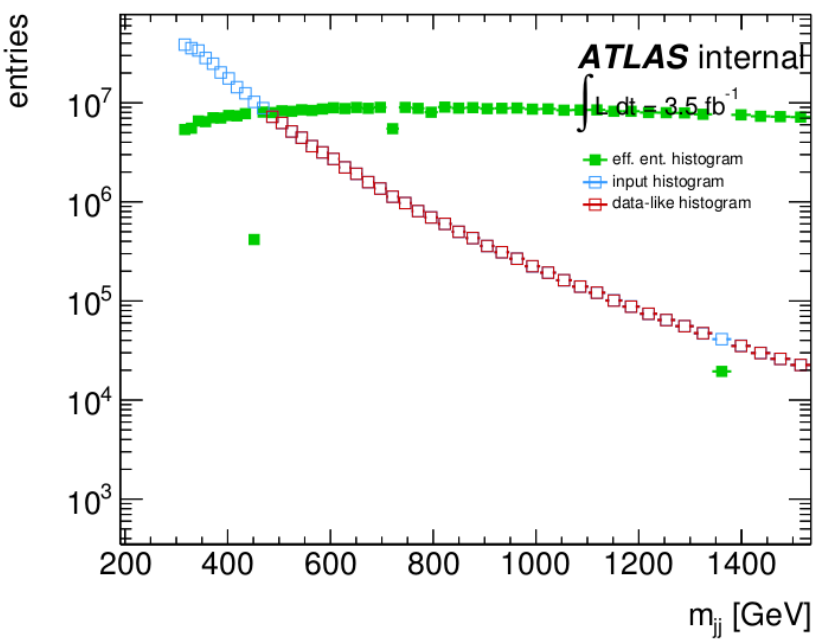


Figure 5.1: Effective number of entries, data-like and scaled distribution for 2.07B truth-level Pythia QCD events.

5.2.3 MC Signal

The benchmark model for this analysis is a leptophobic Z' DM mediator with axial-vector couplings, described in section 2.2. Signal distributions with varying M_{med} and g_{q} values are generated with Pythia8 event generator. It is necessary however to choose parameter values that correlate to the sensitivity reach of the search. The strategy to determine the signal grid is by performing a search on QCD distributions with signal at different cross sections. This can shed light on the sensitivity of the analysis. For this analysis however, at the time of the signal preparation, the statistical precision of the background MC was not sufficient to conduct this study. Instead the sensitivity had to be estimated by comparing spectra with and without injected signal, using the same statistical tools employed in the search phase, and explained in detail in the following chapters. The final signal grid is shown in Table 5.1, the DM mass $m_\chi$was fixed at 10 TeV and g_χ was set to 1.5.[1]

M_{med}/GeV	g_q				
450	0.1	0.15	0.2	0.3	
550	0.1	0.15	0.2	0.3	
650	0.1	0.15	0.2	0.3	0.4
750	0.1	0.15	0.2	0.3	0.4
850	0.1	0.15	0.2	0.3	0.4
950		0.15	0.2	0.3	0.4

Table 5.1: Parameter grid in g_q and M_{med} requested for full simulation with 25 ns configuration.

The signal samples cross-sections, acceptances and Gaussian core widths are shown in Table 5.2. The width of the Gaussian core is calculated by fitting the signal distribution using a Crystal ball function. The fitted parameters of the distribution are used to define the core of the signal. The Gaussian core width of the signal templates is roughly 6-8% relative to the resonance mass. The acceptance depends on the kinematic cuts specified in the next section, and on the invariant mass range of the search.

[1]For $m_\chi$values far above the kinematic limit, the values of $m_\chi$and g_χ are not expected to have any effect on the dijet resonance production and decay.

Sample	cross-section (fb)	Acceptance (m_{jj}>450 GeV)	Width of Gaussian core	Cross-section within core (fb)
$M_{Z'}$=550, m_{DM}=1000, g_q=0.1, g_{DM}=1.5	22303	0.354	0.071	7703
$M_{Z'}$=550, m_{DM}=1000, g_q=0.15, g_{DM}=1.5	50223	0.344	0.078	16975
$M_{Z'}$=550, m_{DM}=1000, g_q=0.2, g_{DM}=1.5	89393	0.351	0.079	30722
$M_{Z'}$=550, m_{DM}=1000, g_q=0.3, g_{DM}=1.5	202950	0.329	0.091	64456
$M_{Z'}$=650, m_{DM}=1000, g_q=0.1, g_{DM}=1.5	11838	0.409	0.064	4176
$M_{Z'}$=650, m_{DM}=1000, g_q=0.15, g_{DM}=1.5	26645	0.415	0.069	9949
$M_{Z'}$=650, m_{DM}=1000, g_q=0.2, g_{DM}=1.5	47453	0.409	0.070	17372
$M_{Z'}$=650, m_{DM}=1000, g_q=0.3, g_{DM}=1.5	107930	0.401	0.077	40122
$M_{Z'}$=650, m_{DM}=1000, g_q=0.4, g_{DM}=1.5	196600	0.385	0.094	74098
$M_{Z'}$=750, m_{DM}=1000, g_q=0.1, g_{DM}=1.5	6808	0.453	0.061	2482
$M_{Z'}$=750, m_{DM}=1000, g_q=0.15, g_{DM}=1.5	15330	0.446	0.131	6801
$M_{Z'}$=750, m_{DM}=1000, g_q=0.2, g_{DM}=1.5	27305	0.442	0.067	9616
$M_{Z'}$=750, m_{DM}=1000, g_q=0.3, g_{DM}=1.5	62203	0.431	0.076	22861
$M_{Z'}$=750, m_{DM}=1000, g_q=0.4, g_{DM}=1.5	113750	0.412	0.086	40847
$M_{Z'}$=850, m_{DM}=1000, g_q=0.1, g_{DM}=1.5	4155	0.459	0.060	1451
$M_{Z'}$=850, m_{DM}=1000, g_q=0.15, g_{DM}=1.5	9356	0.458	0.061	3284
$M_{Z'}$=850, m_{DM}=1000, g_q=0.2, g_{DM}=1.5	16670	0.463	0.064	6198
$M_{Z'}$=850, m_{DM}=1000, g_q=0.3, g_{DM}=1.5	38040	0.454	0.079	14485
$M_{Z'}$=850, m_{DM}=1000, g_q=0.4, g_{DM}=1.5	69833	0.433	0.089	25433
$M_{Z'}$=950, m_{DM}=1000, g_q=0.15, g_{DM}=1.5	5974	0.464	0.061	2142
$M_{Z'}$=950, m_{DM}=1000, g_q=0.2, g_{DM}=1.5	10643	0.462	0.065	3745
$M_{Z'}$=950, m_{DM}=1000, g_q=0.3, g_{DM}=1.5	24338	0.454	0.068	8448
$M_{Z'}$=950, m_{DM}=1000, g_q=0.4, g_{DM}=1.5	44878	0.424	0.085	15395

Table 5.2: Characteristics of the Z' samples generated with the PYTHIA 8 event generator.

5.3 Event Selection

The physics objects used in this analysis are trigger jets, reconstructed with the anti-k_t algorithm [17], and a distance parameter of $\Delta\mathrm{R} = 0.4$.

Dijets events selected for the analysis must have at least two jets with $p_\mathrm{T} > 50$ GeV, and within $|\eta| < 2.8$. The rapidity separation of the two jets must also fulfil $|y^*| < 0.6$. Additional kinematic cuts on the highest p_T (*leading*) and second highest p_T (*sub-leading*) jets in an event require $p_\mathrm{T} > 180$ GeV and $p_\mathrm{T} > 75$ GeV respectively, as well as for both leading, sub-leading, and third jet to pass the cleaning criteria of Sec. 4.2. This selection leads to an unbiased mass distribution for $m_{jj} > 455$ GeV in Pythia QCD MC when using a $|y^*| < 0.6$ cut. The equivalent value with a $|y^*| < 0.3$ cut is 410 GeV. However, this does not determine the starting point of the mass region considered for the search. The lowest mass used is determined during the background estimation procedure.

Chapter 6

Background Estimation and Search Phase

This section describes the background estimation and search phase strategy used in this analysis. Section 6.1 outlines the basic principles and challenges of obtaining a background estimation representing QCD. Section 6.2 discusses an alternative fitting method for estimating the dijet background. In Section 6.3 the strategy of the search, and the statistical framework to interpret the data are explained in detail. The tests conducted to check the robustness of the background estimate and its response to the existence of signal are presented in Section 6.4.

6.1 Estimating QCD Background

Searches in ATLAS for a resonance in dijet invariant mass distributions are generally done using a data-driven background estimate (e.g. [10]). The distribution is fitted using a function which falls smoothly with increasing mass. The fitted function yields the baseline background estimation. It is the hypothesis of a QCD-only interpretation of the results. A data-driven method is preferred because a fully MC-based background prediction introduces theoretical and experimental uncertainties (e.g., the uncertainty on the jet energy scale) which would reduce sensitivity to signals. The challenges of this approach is the lack of a (physically motivated) functional form describing the behaviour of the data, and the statistical precision - A consequence of the (very) large amount of statistics recorded by the DS stream. The m_{jj}

Functional form	Citation	Label
$f(x) = p_1(1-x)^{p_2}x^{p_3}$	[25]	`our3param`
$f(x) = p_1(1-x)^{p_2}x^{p_3+p_4\ln x}$	[31]	`our4param`
$f(x) = \frac{p_1}{x^{p_2}}e^{-p_3x-p_4x^2}$	[32]	`UA2`
$f(x) = p_1(1-x^{1/3})^{p_2}x^{-p_3}$	[33]	`Multijet9`
$f(x) = (1-x^{1/3})^{p_1}x^{-p_2-p_3\ln(x)}$	[34]	`GammaGamma3ParThird`
$f(x) = (1-x^{1/3})^{p_1} * x^{-p_2-p_3\ln(x)^2}$	[34]	`GammaGamma4ParThird`

Table 6.1: Functional forms explored for fitting the QCD background, with citation to previous results in which they have been used to fit similar distributions and label for reference in the rest of the text.

distribution predicted by QCD is a complicated non-analytic convolution of matrix elements, quark and gluon PDFs, detector resolution, and kinematic selections. However, one can draw conclusions from the prediction without knowing it's true functional form. The SM predicts an m_{jj} distribution which is smoothly falling. We expect BSM phenomena at a specific mass to produce an excess in the observed m_{jj} distribution, localized around that mass. If some smooth functional form can describe the overall shape of the m_{jj} spectrum, such localized signal can be hopefully detected. A necessary feature is that this background parametrization allows for a degree of smooth variation of the QCD mass distribution, but does not accommodate localized excesses. The overall shape can be determined, independent of the detector response, by a fit of the function to the data. A list of functions considered by this analysis is presented in Table 6.1, along with references to the analyses in which they have previously been employed. ATLAS has used the form of `our4param` for all of its prior dijet searches. This ansatz was found to provide a satisfactory fit to the QCD prediction of dijet production [25, 31].

There is no *a priori* reason to expect that a simple analytic function will remain able to accurately describe the Standard Model contribution to the data. Trigger Level Analysis takes the technique to a much larger dataset than previously studied. As the number of events increases, the statistical uncertainty associated with it's measurement decreases, as dictated by Poisson statistics. One should question whether with increased precision the new data will finally resolve small differences between simple analytic parametrizations and the true complexity of the underlying mass distribution. Indeed, the ultimate goal of the TLA strategy is to take the search for lower-rate signals into a statistics-rich regime. This implies that the precision with which the background distribution shape is modeled, is a key to the ultimate sensitivity of the search.

This method must be thoroughly tested in MC simulation, prior to applying it on data. Due to the above mentioned uncertainties in the MC simulation, one cannot draw rigid conclusions about the performance of the method on data. Nevertheless, the simulation is not random; it is our best prior estimate of the shape of the m_{jj} distribution of SM background events. In addition, the presence of a small and broad signal in the data cannot be ruled out. Such signal may not be detected as a localized excess, but rather bias the fit and produce an inaccurate background parametrization.
The above discussion highlights the importance of having a characterization of the method performance on QCD simulation. This is the reason the data is *blinded* during the validation of the background estimation. That is, only after the background estimation technique is found to work well on simulation, it is considered ready to be applied, and the data is *unblinded*.

6.2 Sideband Fit Method

The search region for the analysis extends a lower bound in dijet mass, revealing an unprecedented amount of statistics. Prior to the derivation of the MC samples, the analysis team was concerned that a single, *global*, fit to the entire spectrum will not accommodate the high precision dictated by the data, as explained in Section 6.1. As an alternative, the *sideband fit* method of estimating the background was developed, suggesting to divide the *global* fit into multiple fits on multiple subregions of the spectrum. The diminished mass range should help the fit to accommodate for the precision implied by the high statistics. When the MC distributions were prepared and stood to a fit test, both fitting methods fairly accomplished in fitting the data. Since the sideband fit was not yet fully understood (in particular the relevant systematics) it was decided to discard this method for the moment and continue the studies using the full-spectrum fit. For this reason its performance was not tested to the extent shown in section 6.4. It is however very likely that with the coming 2016 data, this method, or with some variation, will be used for the background estimation.

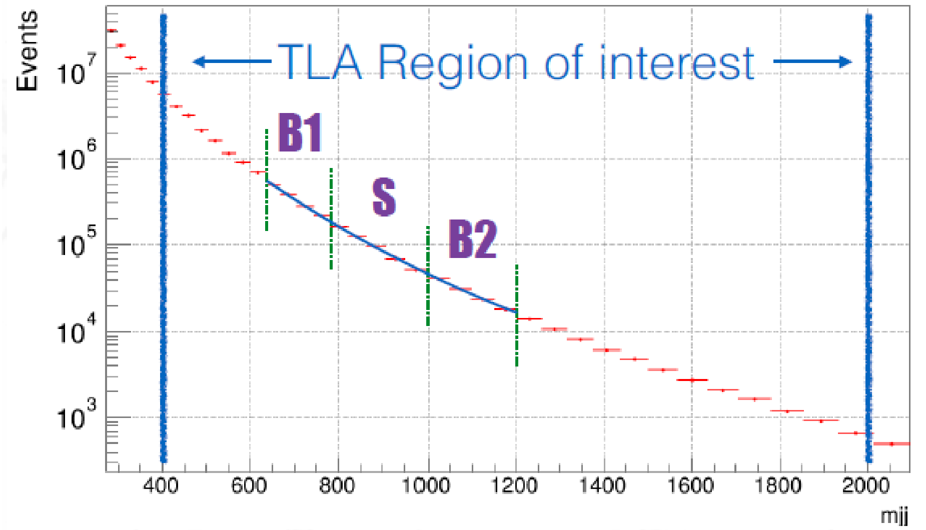


Figure 6.1: The basic scheme of the sidebands method fits. S is excluded from the fit range, which consists of $B1$, $B2$ only. All three regions are iteratively "sliding" through the TLA region of interest.

6.2.1 Overall Description

The sidebands method is an iterative background estimation technique which fits consecutive windows using any fit function ansatz from Table 6.1. The technique starts from a desired mass point, and defines three region in each iteration. $B1$, $B2$ are the sidebands regions and S is the signal region. For the fit, all bins that fall outside regions $B1$ & $B2$ are excluded, thus the retrieved fitted parameters correspond to these regions only. This is of course to avoid a bias in the estimation if a signal exists in data. The fit is used to determine the background only in the central bin of the S region. Next, $B1$, $B2$, and S "slide" by 1 bin, and the adjacent bin is estimated by another fit, with a range set by the new sidebands regions. This is repeated in iterations until the full desired mass range is covered. An illustration of the sidebands fit principles is demonstrated in Figure 6.1.

This method is more suitable for low mass regions where distributions are rich in statistics. In regions with less events, the reduced fit range makes the impact of random fluctuation much more crucial. Since each bin in the final background is estimated using a different subset of data bins, the resulting high mass spectrum is often characterized by a non-smooth form.

This is demonstrated in Figure 6.2, where the sidebands method is applied to a data-like MC mass distribution featuring poor statistics at high masses. This simulation was produced for the 2015 dijet analysis[25]. The simulated event counts are marked in black circles and the predicted background is shown by the red line. In the mass region where event counts are less than $\sim$50 the background exhibits fluctuations and do not maintain the expected smooth form. The bottom panel shows the significance of the background compared to the simulated data. In producing Figure 6.2 S was set to 3 bins and $B1 = B2 = 4$ bins.

A by-product feature of this method is that the final background estimation begins from a higher mass than the full-spectrum fit. This is because in the first iterations the low-edge bins of the data distribution are used to calculate the fit (i.e. included in $B1$), but are not being estimated (not included in S in any part of the calculation). One might think that this can be accounted for by simply extending the edge of the fit by a few bins to the left. However, the starting mass for the background estimation is determined to avoid the kinematic bias of the trigger, and moving to a lower mass edge degrades the reliability of the data. The shift in the starting mass between the two methods depends on the size of $B1$ and S. It is however possible to minimize this bias by varying the amount of bins in $B1$ in the first iterations, while keeping the total d.o.f of the fit fixed. This idea can be illustrated with an example: Assuming sideband sizes ($B1 = B2 = 5$ bins) for the main part of the procedure, one can set $B1$ to 2 bins, and $B2$ to 8 bins for the first iteration. In the next steps, instead of sliding the three windows, bins will be migrated from one sideband to the other, until the regions are in the desired setup. From there the procedure continues like initially described. In principle, the same effect occurs in the high mass end of the spectrum. This however is less a concern as in this case the end point for the estimation is not constrained by the performance of the experimental system.

6.2.2 Current Status

The sidebands method introduces extra free parameters to the stage of the background estimation, which are the number of bins chosen for each window. This choice has implication both on the resulting background, the strategy of the search, and the systematic uncertainties of the final background estimation.

The number of bins in the S region can be motivated by the typical width

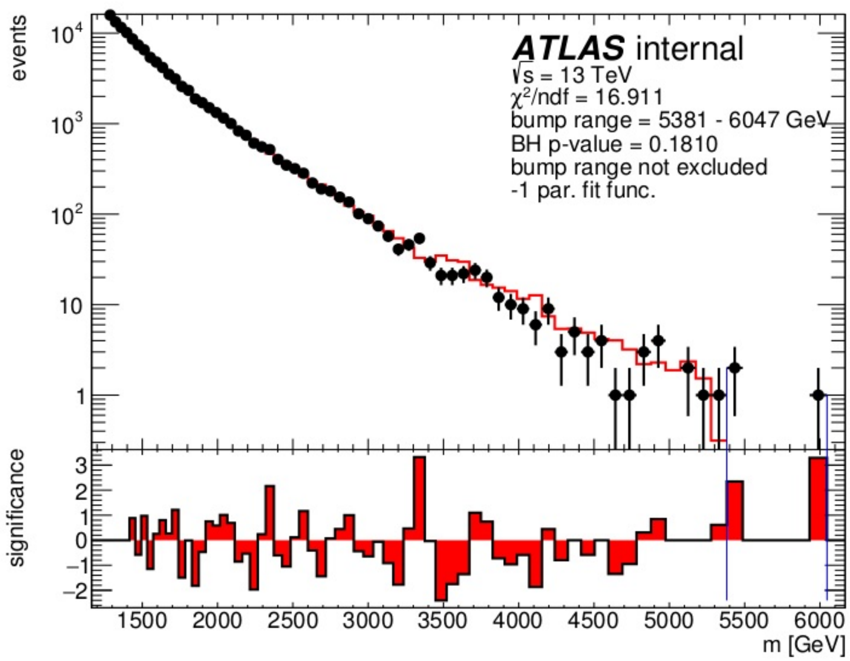


Figure 6.2: Background prediction using the sidebands method on a MC simulated dijets mass distribution. In the higher mass region where the simulated counts are low, the prediction exhibits fluctuations.

of signals relevant to an analysis. Choosing a too narrow signal window will bias the spectrum and make it harder to distinguish an excess in the data, as it is not all excluded from the underlying fits. A wide S means that the background in S is estimated by more remote regions of the spectrum. For the sidebands, using narrow windows may help the fit converge more easily. On the other hand it is required to have a fair amount of bins ($d.o.f$) to better describe the shape of the distribution, and decrease the impact of random fluctuations. In any case, the windows sizes should be picked by a dedicated study, whether on MC or data. At the time this work was written, no such study was conducted mainly because of time and man-power constraints. It is of high priority to conduct such study, and minimize the arbitrary choices made in the analysis. Any choice made in the analysis, is accompanied by some systematic uncertainty regarding that choice. This case is not different. The implications of using a different setup for $B1$, $B2$, S is yet to be quantified, and the extent of the affect this choice has on the resulting background must be understood. Once these systematics will be defined, it will be combined with the fit uncertainty, derived from the fit covariance matrix, to create the full systematics for this method.

If the data holds an excess around some mass, part of the background will be biased by it even if S agrees with the width of the signal. The estimation of bins on the left of the excess, is done when $B2$ overlaps with signal bins. As a result bins in just left of the signal will be overestimated, as the fit is biased by the signal in $B2$. A similar thing occurs when the signal overlaps with $B1$ to the bins just right of the signal. This however, has implications on the limit setting phase only, as it does not affect the ability to discover a signal, only on what the background looks on each side of it.

The next steps in the development of the sidebands fit should focus:

- Developing a robust procedure for selecting sidebands and signal regions width.
- Understanding the systematics from choosing different values for the sidebands and signal regions widths.
- Conducting the same studies elaborated in the following sub-sections, specifically: it's performance with injected signal, and resistivity against spurious signals.

Figure 6.3 shows background predictions using the sidebands method (green) and global fit (red). A calculation at next-to-leading order (NLO)

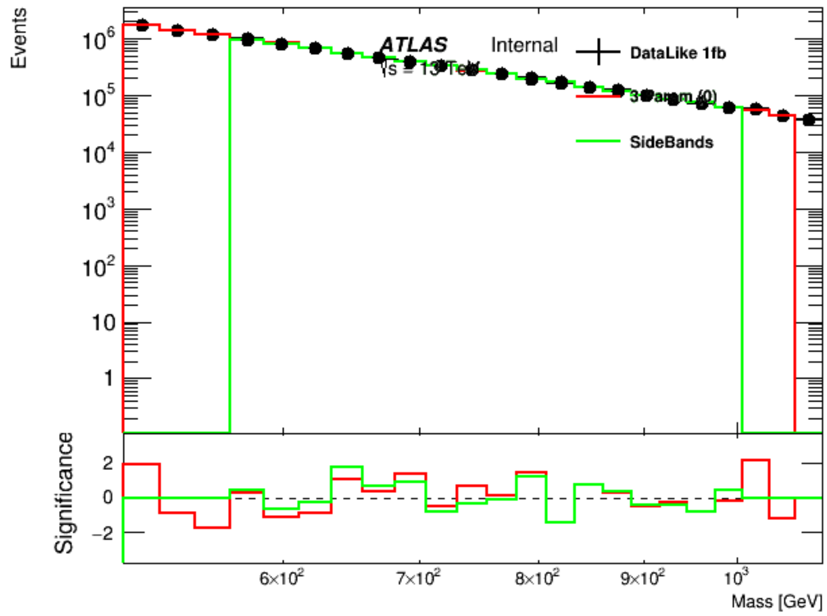


Figure 6.3: Background prediction using the sidebands (green) and global fit (red) methods. In the region where both predictions overlap, they show fair agreement in their shape.

was used to derive the mass distribution, which was scaled to simulate 1 fb^{-1} of data. The bottom panel, which again gives the significance of the background with respect to the input histogram, shows that the shape of the two background estimations fairly agree. The plot also demonstrates the different threshold mass of the two background, with the sidebands estimation starting at a higher mass and ending in a lower one. The configuration of the sidebands and signal region in the figure is $S = 3$ bins, $B1 = B2 = 5$ bins.

Figure 6.4 show the result of applying the sidebands method on the MC simulation used to test the full background estimation procedure, and described in chapter 5.2.

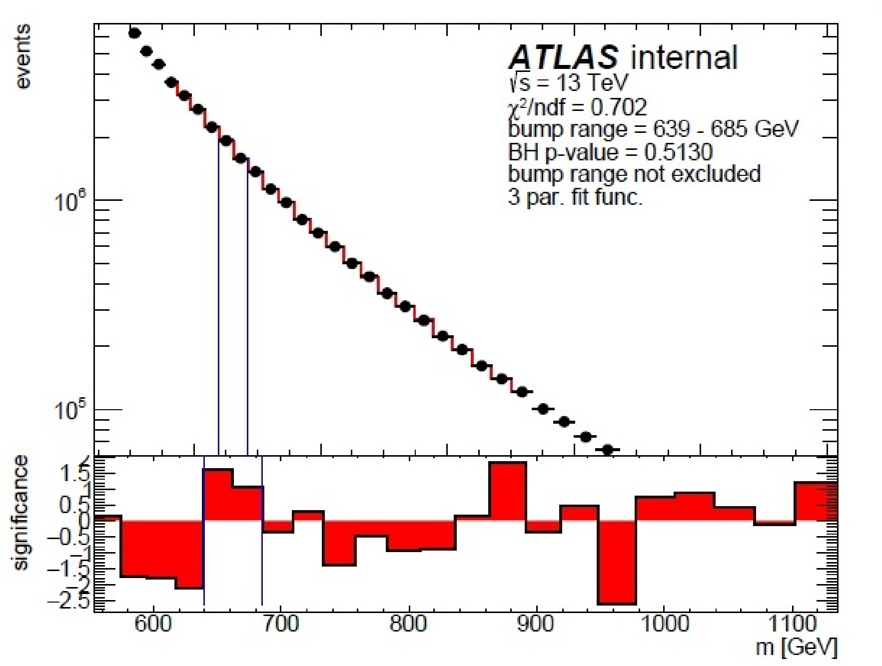


Figure 6.4: Background prediction using the sidebands method with a data-like MC mass distribution scaled to 3.5 fb^{-1}.

6.3 Search Strategy

6.3.1 Search Phase

The term *search phase* refers here to the procedure done in the analysis to determine the presence of a *new physics* signal. This procedure is based on the BumpHunter (BH) algorithm [35], which scans the statistical significance of bins in the data. This procedure is the same as the one used in ATLAS high mass dijet analysis [25].

6.3.1.1 Statistical Significance of Data/Expectation

The significance of the difference between the data and the QCD hypothesis H_0 (which is the fit described in section 6.1) is represented by a p-value. It is the probability of finding a deviation at least as big as the one observed in the data, under the assumption that H_0 describes nature. The number of entries in each bin is assumed to follow a Poisson distribution. In this case, if B events are expected in a bin, and D events are observed in that bin (D is an integer and B is a real number), the p-value for that bin is:

$$p\text{-value} = \begin{cases} \sum\limits_{n=D}^{\inf} \frac{B^n}{n!} e^{-B} & \text{if} \quad D > B \\ \sum\limits_{n=0}^{D} \frac{B^n}{n!} e^{-B} & \text{if} \quad D \leq B \end{cases} \tag{6.1}$$

The p-value range usually spans over many orders of magnitude. Therefore, it is possible to express it in terms of the z-value, which is the number of standard deviations to the right of the mean of the normal distribution. The mapping between a p-value to a corresponding z-value is given by:

$$p\text{-value} = \int_{z\text{-value}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \tag{6.2}$$

A z-value ≥ 0 is equivalent to a p-value ≤ 0.5 and negative z-values correspond to p-value ≥ 0.5. Significant deviations are characterized by small p-values and z-value $\sim$3 and above. Common statistical fluctuations in the data usually yield z-values around 1-2. In this case the deviation is considered not interesting, as these fluctuations are expected. The deviations are considered even less significant when the z-value is negative. It is worth noting that the z-values do not tell us whether the observed number of events is greater than the expected or not. It merely reflects how probable such difference is. For this reason, the *residual* is preferred when presenting the results. For the residual, negative z-values are given the value 0, as they are least interesting. For the remaining (positive) z-value, in bins in which $D < B$ the z-value is assigned a negative sign to indicate that the data falls below the hypothesis.

6.3.1.2 The BumpHunter Algorithm

In cases where a signal is searched over a broad mass range and not in an *a priori* defined mass location, the significance calculation must take into account the fact that an excess in events anywhere within the range can be considered a signal. This is known as the *look-elsewhere-effect* [36]. In addition, a potential signal is expected to be distributed over a number of bins rather than a single bin. The BumpHunter (BH) test can be used to determine whether a significant excess is present in the spectrum. It examines every set of adjacent bins from a minimum width of two bins to (in this analysis) a maximum relative width of 20%. The statistical significance and z-value are calculated for each such window by counting the total number of events in the window. The region with the lowest probability is the most prominent "bump."

A frequentist p-value for the data is defined as the probability of observing a bump at least as prominent in a large number of repetitions of the experiment. It is important to distinguish between this frequentist p-value and the one given by Equation 6.1. In Equation 6.1, the p-value is the probability of the observed data/hypothesis difference in a **single bin**, assuming Poisson statistics. The BH p-value outlined in this paragraph is not calculated analytically from some mathematical expression. Instead, a large number of pseudo-experiments (10000) are generated by random Poisson fluctuations around H_0. The BH test is used to scan each set of pseudo-data and find the most significant excess (or deficit), anywhere in the spectrum. The p-value is directly estimated from the results as the portion of pseudo-experiments which yielded a more significant deviation than that observed in the data. The "*Look-elsewhere-effect*" is inherently taken into account in the algorithm, as the entire spectrum of each set of pseudo-data is sought for deviations.

If the $p-$value obtained from the BumpHunter for the data is larger than 0.01, one may conclude that no statistically significant excess has been observed and the analysis proceeds to the limit-setting stage. If, however, a

$p-$value smaller than that is observed, it indicates that an equally bumpy spectrum would only occur by chance in fewer than 1% of cases if the data is genuinely distributed according to to the background-only hypothesis given by the fit. In this case it becomes necessary to ensure that this background estimate is not itself biased by the presence of signal. The fit is then discarded and a new fit is calculated without the bins corresponding to the most discrepant window. The $p-$value is then recalculated by running the BH only on the remaining regions of the spectrum, and by using an identically restricted range for the fit and BH in all pseudo-experiments. If this $p-$value is found to be above 0.01, the process stops and the last fit is used as the final estimate. Otherwise, an additional minimal bin or bins are removed, the fit is re-run, and the procedure continues as above until the $p-$value is acceptable. It is desirable to remove the minimal number of bins so as to minimise the chance of excluding genuine background fluctuations. After refitting, if the largest remaining bump found is adjacent to the currently excluded window, it is assumed to represent the tail of a signal bump, and a single bin is added to the exclusion window on that side. If, however, the p-value remains low and the largest remaining bump is not next to the excluded window, it is not clear which side to extend, so one bin is added to each end of the excluded window.

6.3.2 Search strategy for the TLA

The background estimation strategy for the TLA consists of fitting an ansatz function to the entire dijet mass distribution (*global fit*), as explained earlier in this section. This method is essentially identical to that applied in the high-mass dijet analysis.

The fit is restricted to a range of m_{jj}. The lowest dijet invariant mass for this search is determined from two inputs:

- The mass cut required to remove the bias of the m_{jj} distribution from the trigger turn-on curve, as mentioned in Sec. 5.3.
- Tests of the fit on MC, varying the endpoint below and above the point determined by the trigger bias study, to understand how well the fit can accommodate bias from the trigger efficiency curve that is larger than the 0.5% level assumed in the trigger turn-on study.

These studies, described in the following, will not conclusively determine a reliable starting point as the Monte Carlo distribution does not necessarily represent the data distribution, but indicate that it may be as low as 443 GeV and likely no higher than 515 GeV. The upper edge of the mass range is less critical and has less impact on the success or failure of various fits. A value of 1.236 TeV was found to work in all MC tests and was selected for the data. The exact start of the fit range is selected alongside the best and alternate fit functions by the algorithm described below.

More than a single fit ansatz was considered for the search. An algorithm is used to determine the fit function that is best suited for the dataset involved by ranking the functions of Table 6.1 using the χ^2 p-value of each function to the data as a figure of merit. The function with the highest p-value is selected as the nominal function, with the second highest selecting an alternate parametrisation which is used to define an uncertainty on the function choice. All functions considered for the background estimation are tested in Monte Carlo, as described in Section 6.4.

6.3.2.1 Algorithm to be followed in each individual search to select a fit function and a fit range

The high precision of the TLA data requires functions to accurately describe the shape of the data to a better degree than in previous dijet searches. In addition, tests on MC are not sufficient to determine the exact impact of the trigger turn-on on the shape of the m_{jj} distribution. For these reasons an algorithm was established for selecting the start point and the functions used to fit a spectrum by iteratively testing various options on the spectrum itself.

The algorithm takes into account that discrepancies between a fit and the data can come from either a bump (signal), or a function that fails to describe the data. Several observations were used in the formulation of the algorithm to assist in this distinction:

- Lowering the start location for the fit generally degrades the fit quality because:
  - The spectrum begins to be affected by the trigger turn-on, which changes the shape away from what is expected
  - Rapidly increasing statistics in the lowest mass bins means that fundamental disagreements between the true shape of QCD and an *ad hoc* parametrisation will become more significant

- Significant disagreements due to the above factors are decreased by raising the start location of the fit.
- When a disagreement arises because of a true bump (such as one created by injecting signal events into MC), the bump remains detectable even with a small shift in the fit start location.
- Allowing a window exclusion during a fit can do one of the two:
  - Remove the impact of a signal to obtain an unbiased background-only estimate (its designed purpose)
  - Allow for better agreement between data and a fit which does not describe it well, by eliminating a region of significant discrepancy.

  Because of the tendency to eliminate a window when no signal is present but a poor function is chosen, any fit which does not require a window exclusion should be prioritised over any fit which does, regardless of which of their p-values is higher.

The algorithm determined by the analysis team is the following:

1. Select a cutoff below which a fit result is considered unacceptable. We selected (χ^2 p-val < 0.05||BH p-val < 0.01).
2. Begin fitting from the lowest possible mass bin as determined in MC (443 for $|y*| < 0.6$, 394 GeV for $|y^*| < 0.3$).
3. Test fitting the set of possible functions and do not permit the exclusion of a window.
4. If two functions provide fits with p-values above the cutoff, the process stops.
   - Take the function with the best χ^2 p-value to be the nominal and that with the second best to be the alternate fit for defining the fit choice uncertainty
   - Re-fit with the selected functions, allowing for a window exclusion. This is the nominal background prediction.
   - If the resulting BH p-value is > 0.01, set limits. Otherwise, a bump has been found. Add systematics to the BumpHunter and follow the procedure in the case of an excess.

5. If fewer than two functions provide successful fits, refit with all functions allowing for a window exclusion of up to 20% window width to mass ratio.

    - If the p-values of the portion of the spectrum outside the window now pass the cutoff criteria, this is a candidate bump. Test if it is real by looking at start locations one bin and two bins higher up. If a successful fit without a window exclusion is possible in either of these cases, this is not a true bump. The successful fit will be taken as the nominal background description.
    - If the p-values of the portion of the spectrum outside the window are still below the cutoff, move the start value to the next highest bin and repeat from step 3.

6. If the preceding step is followed, moving the start value upwards a bin at a time, until only 10 bins remain, and still no good fit is found, we accept that the *global* fit can not describe the data, and focus on the development of the sidebands method.

This is summarised in the flowchart shown in Figure 6.5.

The algorithm was tested extensively in data-like MC with and without injected signal, as described below in Section 6.4.1. The results showed stable performance in accordance with the expected outcomes, so the method was adopted for use in data.

Choice of signal region In addition to the standard search spectrum with a cut of $|y^{*}| < 0.6$, an additional signal region with $|y^{*}| < 0.3$ is also considered in the analysis. This extra signal region is considered to access a lower unbiased dijet invariant mass starting point, given the minimum $\mathrm{p}_T >$ 185 GeV imposed by the trigger turn-on. This region allows an unbiased dijet mass distribution above 410 GeV. The sensitivity of the two signal regions was tested using expected limits with MC backgrounds. For a signal mass of 550 GeV or below, the $|y^{*}| < 0.3$ was found to be more sensitive for all couplings. For this reason, the search phase for $|y^{*}| < 0.3$ is performed only if no signal has been found in the $|y^{*}| < 0.6$ signal region. The results from the $|y^{*}| < 0.3$ signal region are used for the 450 mass points and for limits on Gaussian signals where 2σ of the Gaussian template is not contained in the fit region.

Figure 6.5: Flowchart illustrating the algorithm for selection of a background function and fit range.

6.3.2.2 Systematic uncertainties of the background estimation

There are two systematic uncertainties associated with the global fit method: the uncertainty associated with the choice of fit function, and the uncertainty associated with the fit quality.

Uncertainty due to choice of fit function This uncertainty is related to the differences between the selected nominal and alternate fit functions. Both functional forms are fitted over a range of pseudo-data. The size of the uncertainty in each bin is the RMS of the differences in that bin, and the directionality is chosen to be from the nominal to the alternate.

Uncertainty due to fit quality The second uncertainty is that associated with the quality of the fit itself. Under ideal circumstances this would be derived as a confidence band on the function determined by the covariance matrix of the fitted parameters. However, in cases where the likelihood function has a badly-behaved maximum, it is not possible to accurately compute this covariance matrix using the minimization algorithms of this analysis. In this analysis, fits including a parameter for normalization of a signal template introduce this problematic behaviour by using a parameter which peaks at very small values but has a hard physical cutoff at zero. A different method for estimating the error band is thus required. Since the confidence interval on a function is meant to represent the 1σ region within which the fit would fall in the large-number limit of repeated trials, it can also be found by throwing pseudo-experiments and fitting each. This method does not need accurate estimation of the parameter errors; instead, the pseudo-experiments are generated using Poisson statistics based on H_0 so only the fit itself is required. Here, as before, H_0 is defined from the nominal fit; that is, the fit to the real data. A large number of these pseudo-datasets are generated, each is fit using the same starting conditions as the observed data, and the error on the fit in each bin is defined to be the RMS of the function value in that bin for all the pseudo-experiments.

6.4 Robustness of the Background Estimation

This section describes tests done on truth-level MC to ensure an acceptable background estimation, robust against a number of possible scenarios. The requirements of robustness for the background estimation technique are:

- The background estimation must not exhibit presence of spurious signals, defined as significant localized excesses or deficiencies due to differences between the QCD shape and background estimation;
- If the background estimation is significantly affected by the presence of a signal, the signal must be identified and the signal window excluded;
- If the signal is not identified, the predicted background values must not be significantly affected, neither in the regions neighbouring the signal nor globally.

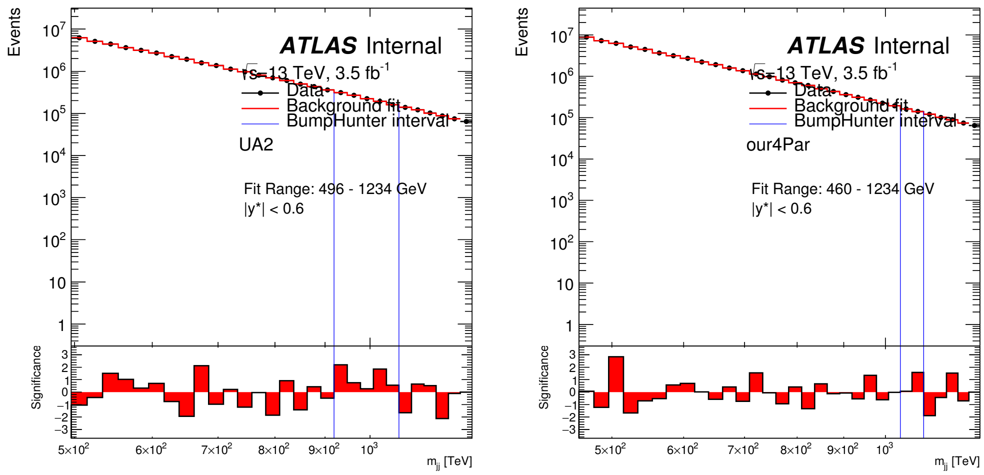
First the algorithm is applied to data-like spectra and is required to give the expected outcome with and without the presence of signal. It is than tested for robustness against spurious signal, ability to distinguish a real signal, and the bias to background from the presence of a signal.

6.4.1 Validation of search algorithm

99 different data-like pseudo-spectra were generated from the truth MC sample to test the algorithm of 6.3.2.1. The pseudo-spectra were created using a different random number seed, and hence each represents one possible outcome of QCD. The algorithm was followed to determine a background estimate for each, consisting of a nominal function, an alternate function, and a start location for the fit.

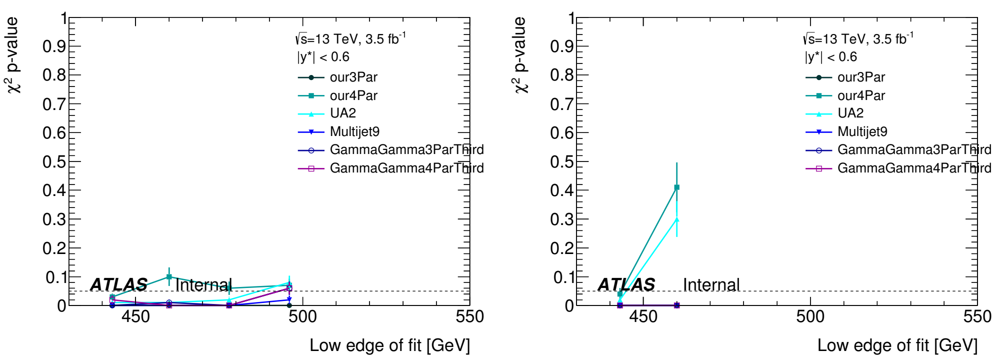
QCD-only MC Figure 6.6 shows the selected nominal distributions for two random seeds. The evolution of the χ^2 p-value as a function of the m_{jj} starting point for those random seeds, is shown in Fig. 6.7 for all functions of 6.1.

In approximately 75% of the cases, the nominal and alternate functions selected were the UA2 function and the dijet four-parameter function, in some order. Successful tests were possible for 95 out of the 99 seeds (where



(a) (b)

Figure 6.6: Nominal fits selected by the algorithm for pseudo-data drawn from two different random seeds. Fit ranges are marked on the plot, along with the name of the nominal function selected.



(a) (b)

Figure 6.7: χ^2 p-value as a function of lowest mass of fit region, for two random seeds. The results show the value of all functions considered for the search.

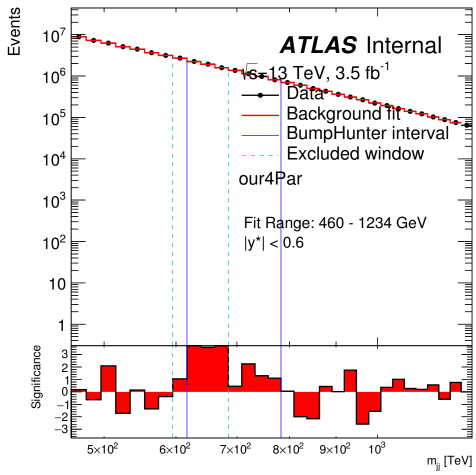
the other 4 suffered from some missing files). Of these 95 cases, all arrived at a successful background estimation without a window exclusion beginning from a mass point no higher than 515 GeV. The fact that the algorithm produced no cases of a window exclusion/discovery in background-only MC provides reassurance that any bump-like feature found in the data will be the product of some actual feature rather than of a poor background parametrisation. This resistance to false discoveries is the most important trait of the background estimate algorithm.

Tests in the presence of a signal For five of the 99 seeds used in the background-only test, a set of new spectra were created, each with one of the Z' signals added to the data-like MC at its nominal cross section. Thus each mass point and width/coupling was tested in 5 different scenarios of statistical noise. Examples of four different mass points in a single background pseduo-spectrum are shown in Figure 6.8.

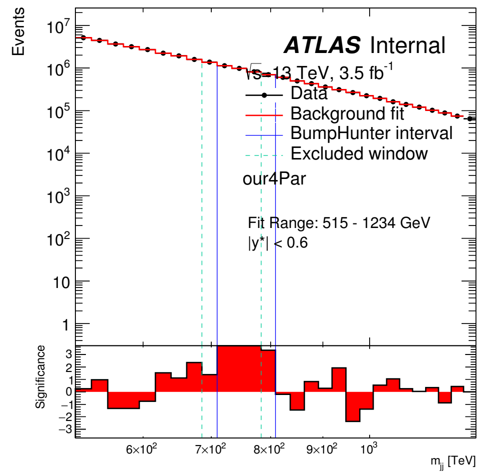
For samples with signals present, 84% of cases resulted in the UA2 and dijet four-parameter functions being selected as nominal and alternate functions, in some order. The rates of discovery for the various points are given in Table 6.2, for the $|y^*| < 0.6$ signal region. Results for the $|y^*| < 0.3$ signal region were not available in time as they would have needed the processing of at least another 2B MC events to have sufficient data-like statistics below 450 GeV.

Mass [GeV]	g_q=0.1	g_q=0.15	g_q=0.2	g_q=0.3	g_q=0.4
450	0.0	0.0	-	0.0	-
550	0.2	0.6	0.4	0.2	-
650	0.0	0.4	0.6	1.0	1.0
750	0.0	0.6	1.0	1.0	1.0
850	0.0	0.0	0.6	1.0	1.0
950	-	0.4	0.4	1.0	1.0
1050	-	-	-	1.0	0.8

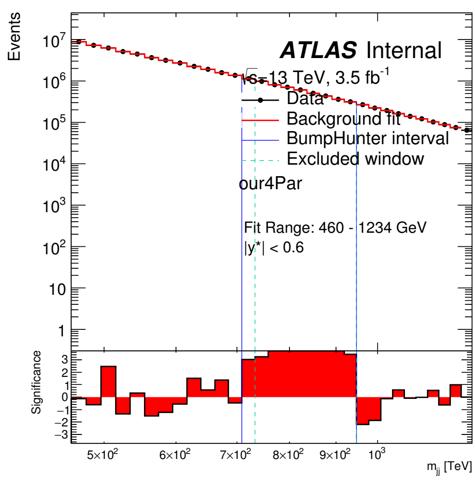
Table 6.2: Fraction of cases, out of 5 tests in unique background spectra, in which each signal mass point was discovered.



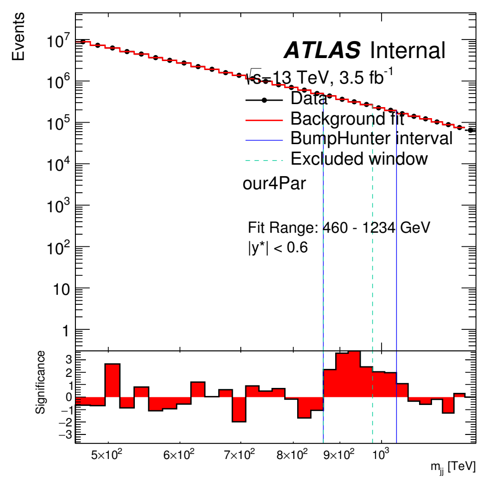
(a) M_{med} = 650 GeV, g_{q}= 0.15



(b) M_{med} = 750 GeV, g_{q}= 0.20



(c) M_{med} = 850 GeV, g_{q}= 0.30



(d) M_{med} = 950 GeV, g_{q}= 0.30

Figure 6.8: Nominal fits selected by the algorithm for pseudo-data when fitting a spectrum with four different injected Z' signals. A window was excluded in all four cases, as denoted by the dashed lines. In each case the excluded region corresponded to the location of the injected signal. Fit ranges are marked on the plot, along with the name of the nominal function selected.

6.4.2 Spurious signal tests

Deviations between the QCD background estimation and the actual QCD shape can result in spurious signal or deficiencies. To evaluate such discrepancies the background estimation is done on a Monte Carlo distribution which is scaled to the expected luminosity. The distribution is not introduced Poisson noise as the fluctuation can overshadow any trends in the fit deviation. Thus, this is a worst case scenario where the fit differences are correlated bin-by-bin.

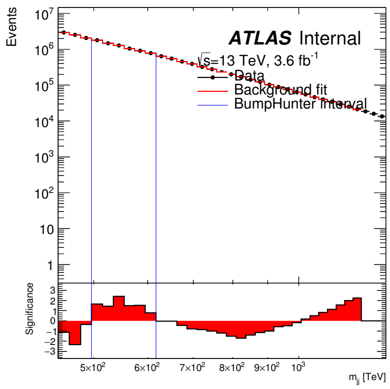
The spurious signal is defined as the difference in events within the most discrepant region chosen by the BumpHunter algorithm. This number is compared against benchmark signal yield and uncertainty on the background estimation.

Figure 6.9 shows the search phase results on the scaled MC distribution with all functions of Table 6.1. For both UA2 and the dijets four-parameters function, the number of spurious signal events is under 20% of the signal yield and no larger than 20% of the background uncertainty, which are the requirements set by the high mass dijets analysis [25]. The other four functions considered for the search are exhibiting visible trends in deviations, as can be seen in their respective residuals pads.

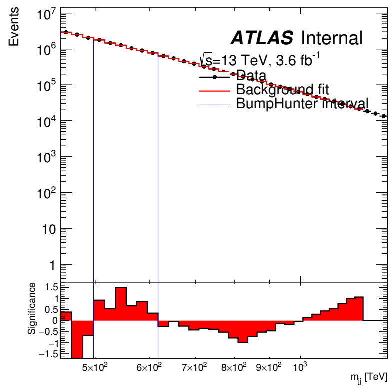
The spurious signal in each bin is compared to the fit function uncertainty in Figure 6.10, for the case of the UA2 and dijets four-parameters function.

6.4.3 Gaussian signals injection studies

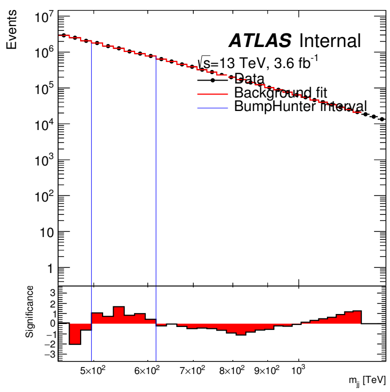
To ensure that the fitting procedure is robust and stable in the presence of signal, pseudo-data distributions containing both background QCD and possible signal events are fitted using the four parameter dijets function, `our4par` of Table 6.1. The initial background distribution is obtained by generating random Poisson numbers around a fit to the scaled QCD mass distribution from Monte Carlo. The signal events are drawn from Gaussian distributions with varying mean and width. The Gaussian relative widths tested are between 3 - 15 % and their mean mass values between 550 - 950 GeV using a $|y^*| < 0.6$ cut. An additional 450 GeV mass point was tested using a $|y^*| < 0.3$ cut with a 15% width excluded. Applying a harder cut makes it possible to have effective statistics from a lower mass threshold, whereas with $|y^*| < 0.6$ a signal at 450 GeV will coincide with the low edge of the fit. The signal normalization is iteratively chosen to be close to the



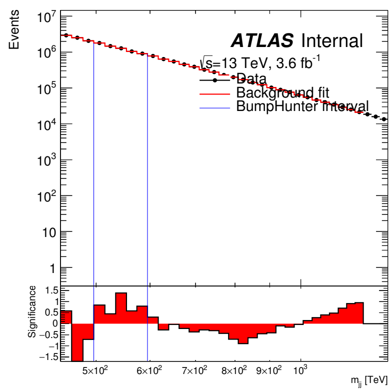
(a) Dijet 3-parameter function



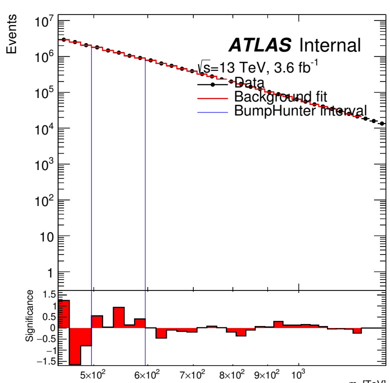
(b) Diphoton 3-parameter function



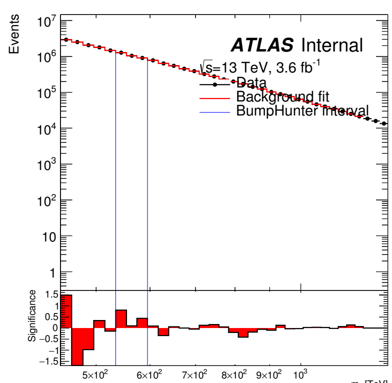
(c) Ninth multijet function



(d) Diphoton 4-parameter function

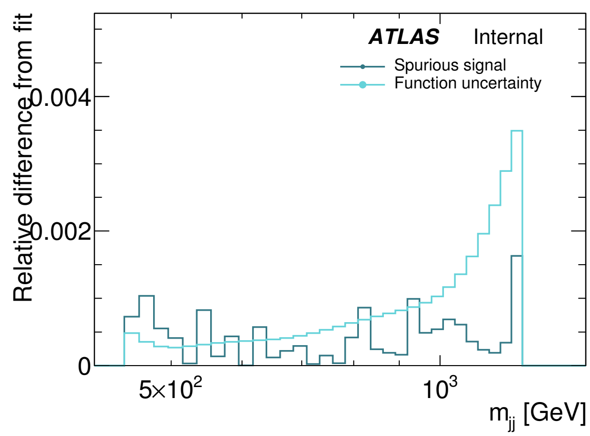


(e) UA2 function

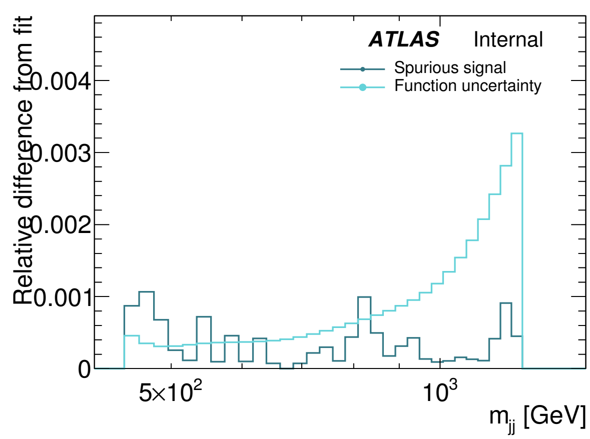


(f) Dijet 4-parameter function

Figure 6.9: Spurious signal found by fitting the full scaled MC distribution from 443 GeV to 1236 GeV with each of the six best-performing fit functions.



(a) UA2 function

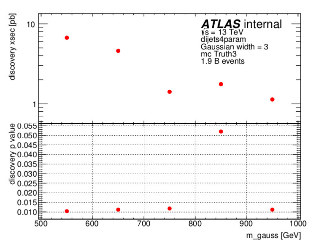


(b) Dijet 4-parameter function

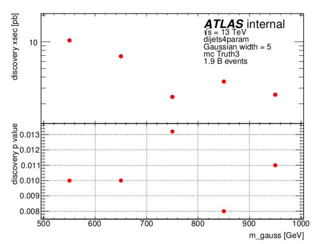
Figure 6.10: Size of spurious signal compared to fit uncertainty for the two best-performing fit functions.

threshold p-value for the BumpHunter algorithm (~ 0.01) where the signal window is removed from the fit.

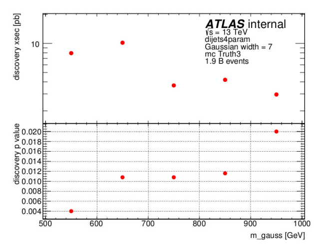
The top panels of Fig. 6.11 show the number of signal events needed for a detection, divided by the background luminosity (3.5 fb^{-1}), labeled as discovery cross-section in pb. The bottom panels of the same figure show the $p-$value closest to the discovery threshold. The individual cases where the $p-$value were not close to 0.01 have been investigated: the cause of the difference is a background fluctuation in the Poisson draw of the events interfering with the algorithm used to increase the normalization. Forcing a higher cross-section by hand, or changing initial data-like distribution, resolves this issue. Algorithmically, this is not optimal. A possible solution would be to draw more pseudo-experiments where different background distributions are used and averaging the results. However, this study as is already gives a rough idea of the cross-sections that the search phase is sensitive to, and what kind of background biases they can induce.



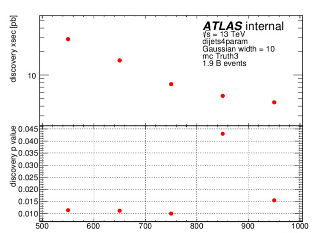
(a) signal width = 3%



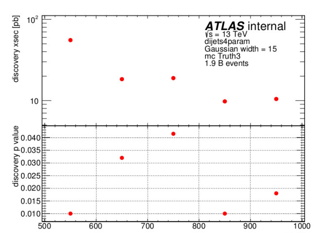
(b) signal width = 5%



(c) signal width = 7%



(d) signal width = 10%



(e) signal width = 15%

Figure 6.11: discovery cross sections for five different Gaussian widths. The cross section here is the number of entries in the signal histogram divided by the luminosity

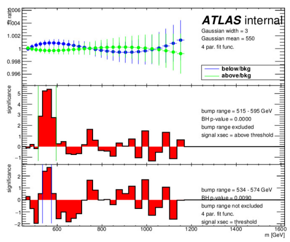
What can be extracted from the results of Fig. 6.11 is that the search phase is not sensitive to resonances as wide as 15% of their mass. In a search with a restricted fit range, and with low masses, 15% means a significant number of bins.

Mass distributions containing signal right above and below the threshold were also tested. Example figures for the 550, 750, and 950 GeV mass points, Figs. 6.12, 6.13 and 6.14, show comparisons between the fit in cases of inclusion/exclusion of the bump region. The 450 GeV mass point with the $|y^*| < 0.3$ cut is shown as well in Fig. 6.15. The top panel in the plots show the ratio of the fit with and without the injected signal. The blue line represents the ratio of the fit without any signal injected to the fit with injected signal that is just under the threshold, thus indicating the bias from having signal but not being able to discover it. The green line shows the ratio of the fit without any signal injected to the fit where the signal normalization slightly above threshold: in this case, the signal window is removed by the BumpHunter algorithm and the background is refitted without the signal window. The bottom two panels show the significance of the fits residuals between pseudo-data (with signal) and the fit with signal for the two cases.

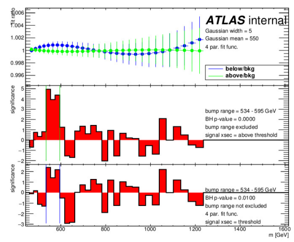
As expected, when the signal window is excluded, the difference between the fit with and without signal is generally below 0.1%. This gives confidence in the background estimation procedure in combination with the search phase. Whenever the signal is not discovered however, the background estimation with injected signal shows a bias that is larger for wider signals and for signals that are closer to the edges of the fit. This bias can be compared to the fit function uncertainty and to the background uncertainty of the fit, shown in Fig. 6.10 when including statistical uncertainties only. For higher-mass, wider signals (> 10%), the bias is considerably larger than the current uncertainty. From these studies, testing widths that are above 10% in the Gaussian limit setting procedure is avoided.

6.4.4 Sanity-check for data-derived JES correction

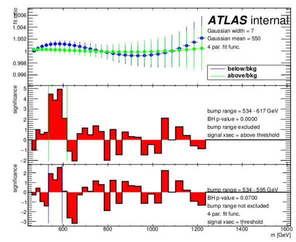
An additional sanity-check is done testing the data-derived correction applied in the calibration stage. The test is done by setting the correction factors to the high-statistics MC truth sample. The MC simulation is already in a fairly good agreement between trigger jets and offline jets response, and therefore the jets would not need any further correction. Applying the scale factors on the MC simulation results in a biased m_{jj} distribution, which can be used as a "worst case scenario" of applying a wrong scale on the trigger



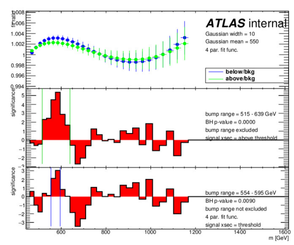
(a) signal width = 3%



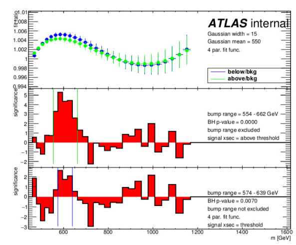
(b) signal width = 5%



(c) signal width = 7%



(d) signal width = 10%

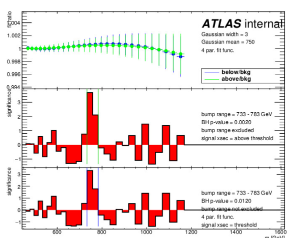


(e) signal width = 15%

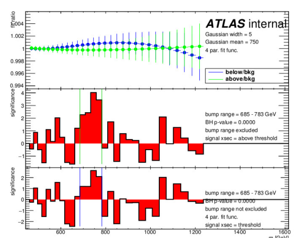
Figure 6.12: Ratio of fits with injected signal to pure background distributions, for Gaussian mean of 550 GeV and various widths for signals right above and right below the discovery threshold.

jet kinematics in data. This check is done to ensure that a wrong scale would not create an artificial bump.

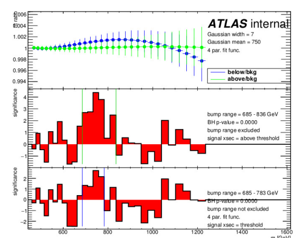
Data-like spectra are derived from the MC m_{jj} both before and after applying the scale factors. These spectra are than fitted using the dijets four parameter function, and the fits results are compared. Generally it is not expected that this procedure should work on MC simulation. However, Figure 6.16(a) shows that the fit remains stable on MC simulation even when the scale factors are applied. In comparison with the fit on the original MC distribution (Figure 6.16(b)) the fit results are worsen: At the high edge of the mass range the residuals show no difference. In lower mass however, there are differences in the residuals shape creating a localized 3-bins excess. This is seen also as the BumpHunter p-value drops from 0.64 to 0.11 in the scale factors case. This however is not significant enough to exclude a region from the fit.



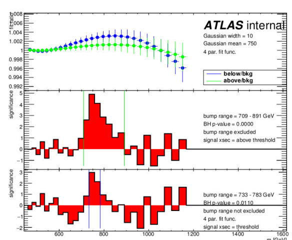
(a) signal width = 3%



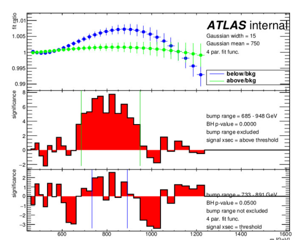
(b) signal width = 5%



(c) signal width = 7%

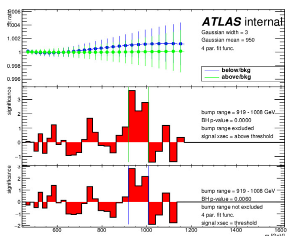


(d) signal width = 10%

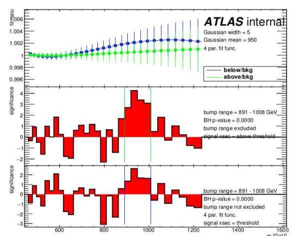


(e) signal width = 15%

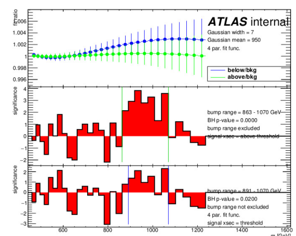
Figure 6.13: Ratio of fits with injected signal to pure background distributions, for Gaussian mean of 750 GeV and various widths.



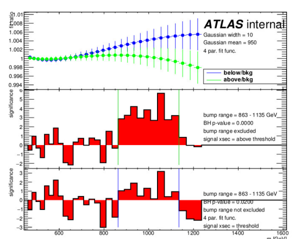
(a) signal width = 3%



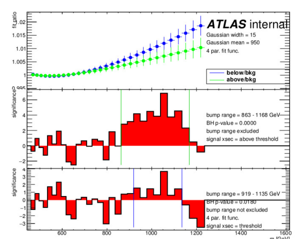
(b) signal width = 5%



(c) signal width = 7%

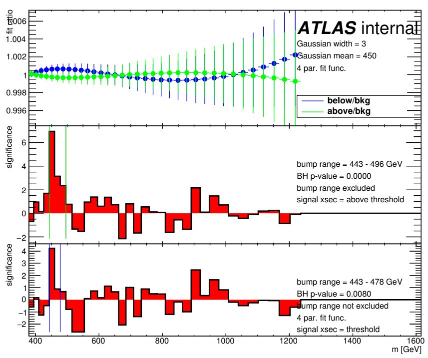


(d) signal width = 10%

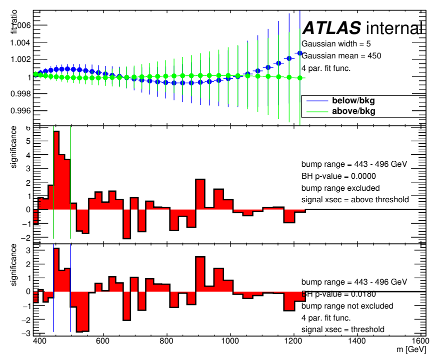


(e) signal width = 15%

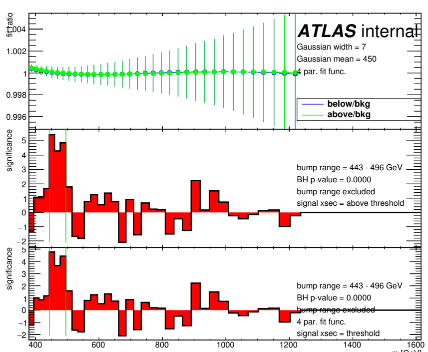
Figure 6.14: Ratio of fits with injected signal to pure background distributions, for Gaussian mean of 950 GeV and various widths.



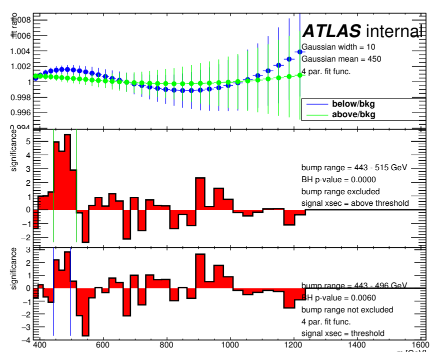
(a) signal width = 3%



(b) signal width = 5%

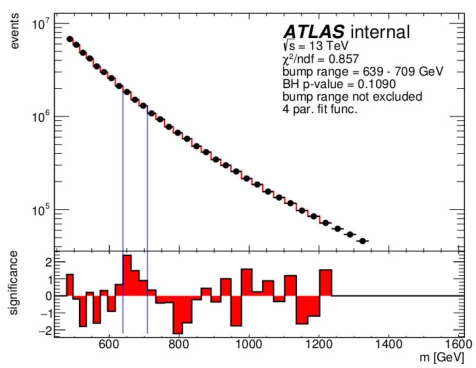


(c) signal width = 7%

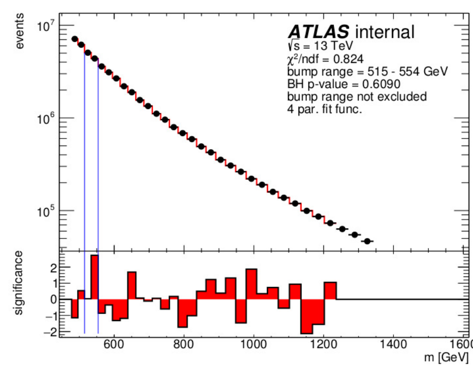


(d) signal width = 10%

Figure 6.15: Ratio of fits with injected signal to pure background distributions, for Gaussian mean of 450 GeV and various widths for signals right above and right below the discovery threshold, using a $|y^*| < 0.3$ cut.



(a)



(b)

Figure 6.16: Fits on data like spectra generated from MC distribution with (a) and without (b) correction factors applied. In the lower panels the significance of the deviations is shown.

Chapter 7

Search Phase results

This section contains the results of the search on data. Section 7.1 presents a preliminary test on a single run (25 pb^{-1}), as well as the results from unblinding a subset of the full data (0.8 fb^{-1}). The results using the full 3.4 fb^{-1} recorded by the Data Scouting stream are presented in Section 7.2.

7.1 Results on partial datasets

A test on data prior to unblinding was done by running the search phase procedure on a single run (281317, 25 pb^{-1}). This corresponds to $\sim 1/100$ of the full Data Scouting stream statistics. The three parameter fit function labeled `our3param` in Table. 6.1 was used with fit range between 460 - 1236 GeV. Results are shown in Figure 7.1. The three parameter function is not the best performing function as predicted by the full background estimation tests. It is used here because of the small statistics in the sample chosen for this test. With such a low luminosity, the three parameter function is sufficient and it does not add unnecessary degrees of freedom to the fit.

The next stage of the unblinding process examined a subset of the data, constituting 800 pb^{-1} of luminosity and equivalent therefore to roughly one quarter of the full dataset. For this partial unblinding the full procedure for selection of a function and a starting point was followed, as detailed in Section 6.3.2.1. The resulting fits are shown in Figure 7.2. For the spectrum with $|y^*| < 0.3$ a p-value of 0.877 is found. The chosen function is the four-parameter dijet function `our4param` with the resulting background starting from 379 GeV. The best function for the spectrum with $|y^*| < 0.6$ is UA2

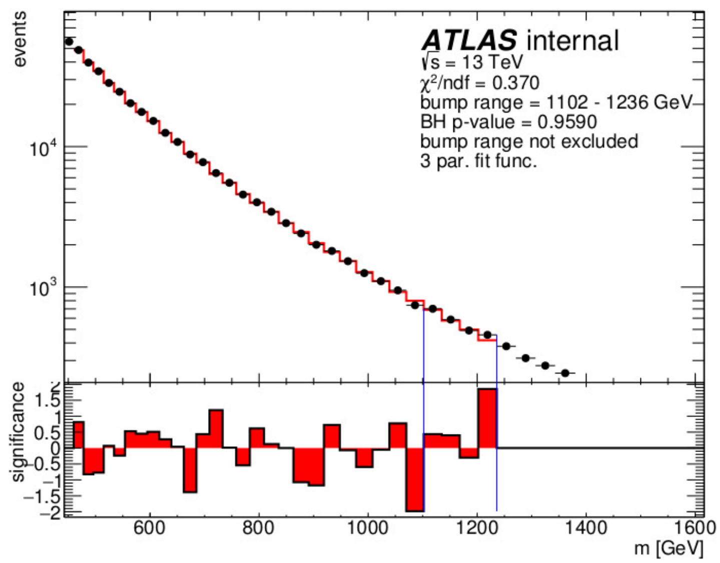
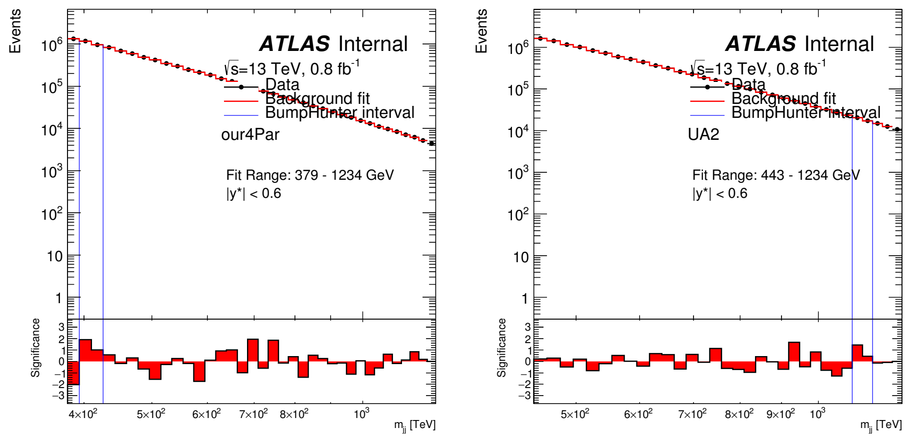


Figure 7.1: Fit of a single run (281317, 25 pb^{-1}) from the data scouting stream. The three parameter function accurately describes the data when the statistics are low.

with a fit range starting from 443 GeV. The p-value for this dataset is 0.921. As the resulting p-values suggests, in neither case a significant discrepancy was observed, and thus the unblinding proceeded to the final stage.

7.2 Results on the full dataset

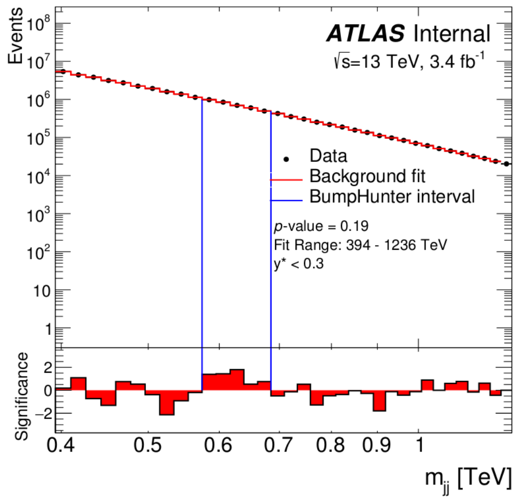
The final unblinded search phase results using the full 3.4 fb^{-1} of 2015 data are shown in Figure 7.3. For both the $|y^*| < 0.3$ and $|y^*| < 0.6$ spectra, the algorithm discussed in Section 6.3.2.1 selected the UA2 function as the nominal background description with the dijet four-parameter function as the alternate parametrisation. The $|y^*| < 0.3$ spectra is fit from $m_{jj} = 394$ GeV, while the $|y^*| < 0.6$ spectra is fit starting from $m_{jj} = 443$ GeV. This is in agreement with the results found in Monte Carlo, where ~75% of fits to background-only spectra exhibited a preference for these two fit functions in one or the other order. The resulting BumpHunter p-values are 0.186 for the $|y^*| < 0.3$ spectrum and 0.439 for the $|y^*| < 0.6$ spectrum. This is in good agreement with the Standard Model and shows no indication of new



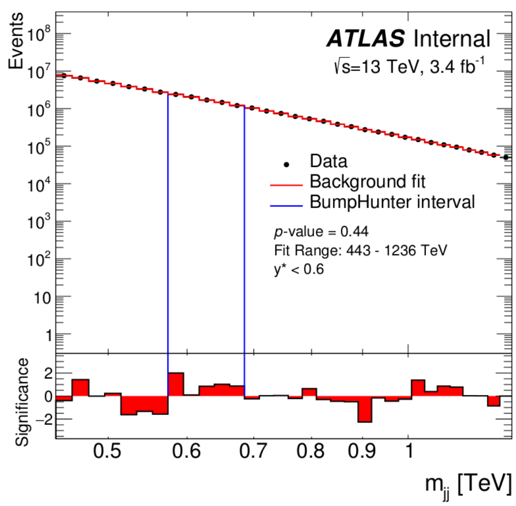
(a) (b)

Figure 7.2: Search phase results in a subset of the data, corresponding to 0.8 fb^{-1}. (a) shows the search phase result using a cut of $|y^*| < 0.3$, and (b) shows the equivalent result using the nominal cut of $|y^*| < 0.6$.

physics. Figure 7.4(a) shows the distribution of BumpHunter test statistics from the pseudo-experiments used to derive the p-value for $|y^*| < 0.6$. The red arrow represents the BH test result for the data themselves. Figure 7.4(c) shows the p-values for all mass windows considered in the search phase using $|y^*| < 0.6$. The analogous plots for $|y^*| < 0.3$ are presented in Figures 7.4(b) and 7.4(d).

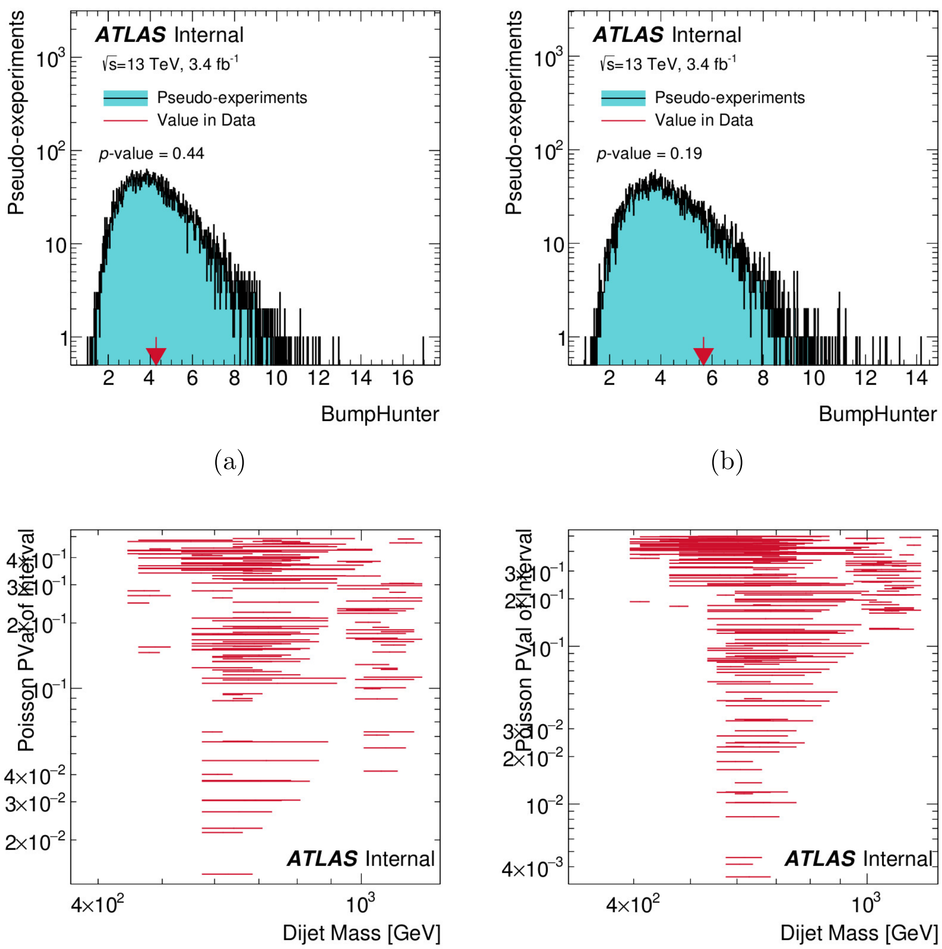


(a)



(b)

Figure 7.3: Search phase results in the full 2015 dataset corresponding to 3.4 fb^{-1} of data. (a) shows the search phase result using a cut of $|y^*| < 0.3$, and (b) shows the equivalent result using the nominal cut of $|y^*| < 0.6$.



(a) (b) (c) (d)

Figure 7.4: Search phase results in the full 2015 dataset corresponding to 3.4 fb^{-1} of data. (a) and (b) show the value of the BumpHunter test statistic in data for $|y^*| < 0.6$ and $|y^*| < 0.3$ respectively, compared to the distribution of test statistics for an ensemble of pseudo-experiments generated from the background fit. (c) and (d) shows the Poisson p-value of each of the mass intervals considered in the search phase for $|y^*| < 0.6$ and $|y^*| < 0.3$ respectively.

Chapter 8

Excluding Bounds

With no observed indication of BSM physics, a limit setting procedure is followed to set constrains on possible new physics models. Section 8.1 describes the method used in the limits setting stage. The resulting 95% credibility-level (CL) upper limits on for Z' models are shown in 8.2. Model-independent limits, assuming a Gaussian shaped signal, are presented in 8.3.

8.1 Limit setting method

The limits setting procedure uses a Bayesian approach, in which the probabilities of new physics models are evaluated from the observations of an experiment, according to Bayes theorem [37]. These probabilities are then used to set 95% CL exclusion limits on the signal strength of such models (given in $\sigma \times \mathcal{A} \times BR$) as a function of the resonance mass. The posterior probability that the sought model $H(\lambda, \theta_i)$ was observed given the data D is:

$$P\big(H(\lambda_i, \theta_i)|D\big) \propto L\big(D|H(\lambda_i, \theta_i)\big)\pi(\lambda_i, \theta_i) \tag{8.1}$$

θ_i are the nuisance parameters, describing our knowledge about the experiment and the systematic uncertainties. λ is a parameter of the model, and is proportional to the signal strength. $\pi(\lambda, \theta_i)$ is called the *prior*, and it gives the probability distribution for the values of λ and θ_i prior to the observation. It is assumed that $\pi(\lambda, \theta_i) = \pi_\lambda(\lambda)\pi_\theta(\theta_i)$ and $\pi_\lambda(\lambda)$ is taken as a constant. This indicates no prior knowledge on the signal strength in the sought model. $L\big(D|H(\lambda, \theta_i)\big)$ is the likelihood function, which gives the probability for D given $H(\lambda, \theta_i)$. To set a limit on the cross section of a new physics process the posterior probability must depend on λ only. The nuisance parameters are integrated out to obtain:

$$P\big(H(\lambda)|D\big) = \int P\big(H(\lambda, \theta_i)|D\big) d\theta_i \tag{8.2}$$

To integrate Eq. 8.2 one must take into the account the systematic uncertainties employed in θ_i. The following systematic uncertainties are considered for the calculation:

- **Fit choice uncertainty** (Section 6.3.2.1)
- **Fit quality uncertainty** (Section 6.3.2.1)
- **Luminosity uncertainty** of 5%
- **JES uncertainty** derived for the calibration constants

The integration is performed numerically using a *Markov Chain Monte Carlo* approach [38]. This is referred to as marginalization.

The 95% CL upper limit on the signal strength, λ_{upper}, is obtained by solving:

$$0.95 = \int_0^{\lambda_{upper}} P\big(H(\lambda)|D\big) d\lambda \tag{8.3}$$

8.2 Limits on Z' models

Limits on specific models are set using signal templates, usually derived in MC. As described in Sec. 5.2.3, Z' distributions were generated at various masses $M_{\rm med}$ and SM couplings $g_{\rm q}$, shown in Table 5.1. The models predict a cross section dependant on $M_{\rm med}$. When the simulated events are applied the reconstruction steps and the same selection cuts as data, the acceptance $\mathcal{A}$, defined as the fraction of events passing the selection cuts, is obtained. Each signal distribution is used in combination with the background estimation, as the hypothesis for a Z' model with the corresponding $M_{\rm med}$ and $g_{\rm q}$. In the language of the previous section, this hypothesis is represented by $H(\lambda, \theta_i)$. The procedure in Sec. 8.1 is than applied, and the result of Eq. 8.3 gives the observed 95% CL upper limit on cross section times acceptance times branching ratio, $\sigma \times \mathcal{A} \times BR$, for that Z' model. Observed limits are set for masses in which a MC signal distribution was generated. They can be compared to the expected limits, which are obtained by treating the background estimation as an observation. The background estimation is fluctuated multiple times with Poisson statistics, and each fluctuated distribution replaces D for the derivation in the previous section. The expected limit is the median of the distribution of 95% CL exclusion limits, obtained from the various fluctuated pseudo-data. A theoretical representation for the Z' model is obtained from a MC calculation of the cross section for the specific model as a function of the mass. The limit on the Z' mass for a specific model occurs when the observed limit is the same as the theoretical prediction.

The observed limits, expected limits, and theoretical prediction can be seen in Figure 8.1 for the $|y^*| < 0.6$ selection, and in Figure 8.2 for $|y^*| < 0.3$. The distribution of pseudo-data used to derive the expected limits is also used to set 1σ and 2σ error bands, indicating the values where 68% and 95% of the distribution is held.

If indeed no new physics process is observed in the data, the expected and observed limits should fairly agree. It can be seen in Figure 8.1 that the limits are weakened with respect to the expectation near the signal points at 650 GeV. This can be explained by noticing that the most discrepant interval found by the BumpHunter algorithm is between 595-685 GeV. In regions where the data is high compared to the fit, it requires more signal events to simulate a significant excess with respect to the data.

Figure 8.3 shows the ratio of observed 95% CL upper limits to the value predicted by theory, as a function of M_{med} and g_{q}. Expected limits are shown in parenthesis. A ratio below unity implies that the model is excluded with 95% credibility. For the 450 GeV mass point, the signal region with $|y^*| < 0.3$ was used to calculate limits. Masses above and including 550 GeV are constrained with $|y^*| < 0.6$. The solid (dashed) red curve represent the observed (expected) limit, derived from extrapolation across the coupling axis. For the $|y^*| < 0.3$ this curve is given as a dot at 450 GeV.

The TLA results can be combined with exclusions from other ATLAS searches to give current bounds for possible Z' models, as shown in Figures 8.4 - 8.5. 95% CL observed and expected upper limits on g_{q} as a function of M_{med} are shown in Fig. 8.4. The darker blue curves show the results

from a "high-mass" search using the traditional triggering strategy for dijet resonances [39]. The red (purple) curves show the results from a search triggering on events containing a photon (gluon), such as those radiated from the initial state partons that would produce the resonance [40]. The lighter blue curves show the results from the Trigger-Level-Analysis.
Figure 8.5 show regions in a DM mass–mediator mass plane excluded at 95% CL by a selection of ATLAS dark matter searches. Two scenarios are considered for the computation of the exclusions: In (a) the Z' couplings to DM and to quarks are taken to be g_χ= 1 and g_q= 0.25, and in (b) g_q= 0.1 and g_χ= 1.5. In both scenarios the coupling of Z' to quarks is assumed to be universal to all flavors.

8.3 Model-independent limits

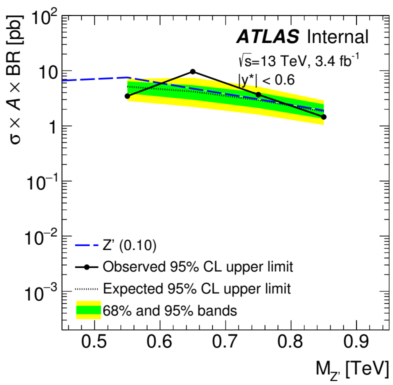
The fact that many models show similar features in data makes it plausible to set limits which are as close as possible to model-independent exclusion limits. Instead of simulated MC events, the signals probed are generated from a Gaussian template. This is a valid approximation, as many signals have a Gaussian or Gaussian-like core. The normalization of the Gaussian plays the role of λ in terms of the nomenclature of Sec. 8.1. When divided by the luminosity $\sigma \times \mathcal{A} \times BR$ is obtained, and the limit can be compared to that of Z'. Gaussian signals with mean values m_G between 425 GeV - 1.1 TeV and widths-to-mass ratios of 10%, 7%, and the detector resolution were probed. Different widths have implications on the affect of statistical fluctuations on the resulting limits. This is explained by the fact that similar numbers of events are spread over a different number of bins. Narrower signal can be aligned with local fluctuations, making it most vulnerable for this effect. The limits on $\sigma \times \mathcal{A} \times BR$ are plotted in Fig. 8.6 and summarized in Tables 8.1 and 8.2. As is the case of the model-specific limits, the generic Gaussian limits are weakened near ∼650 GeV, the region with the largest excess in the data compared to the fit. As for signal region, the $|y^*| < 0.6$ cut was used to set the limit in cases where 2σ of the Gaussian was within the fit range. The two signal regions are separated in Fig. 8.6.

Mass [GeV]	Resolution width	7% width/mass	10% width/mass
500	4.1	-	-
550	2.0	2.5	-
600	4.3	5.4	9.3
650	3.9	5.2	7.2
700	2.5	3.4	4.7
750	1.3	1.9	2.4
800	1.1	1.2	1.5
850	0.73	0.81	0.95
1000	0.81	1.0	1.4
1050	1.0	1.6	-
1100	1.1	-	-

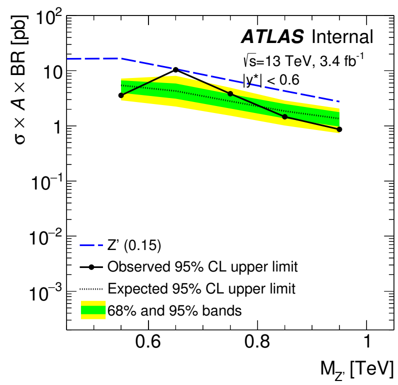
Table 8.1: Limits on cross section times acceptance times branching ratio for generic Gaussian signals using the $|y^*| < 0.6$ cut.

Mass [GeV]	Resolution width	7% width/mass	10% width/mass
425	3.8	4.1	-
450	2.4	2.7	3.5
500	-	1.8	2.2
550	-	-	4.4

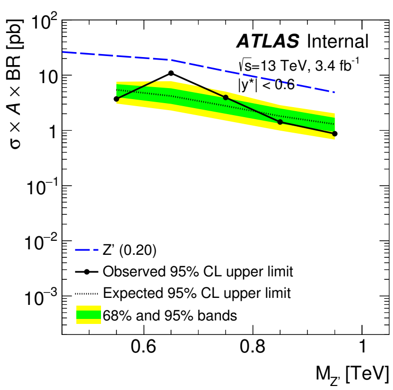
Table 8.2: Limits on cross section times acceptance times branching ratio for generic Gaussian signals using the $|y^*| < 0.3$ cut.



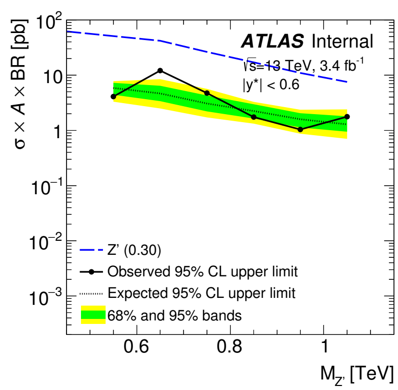
(a) g_{q}=0.1



(b) g_{q}=0.15

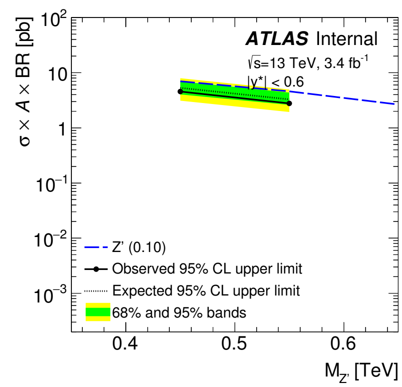


(c) g_{q}=0.2

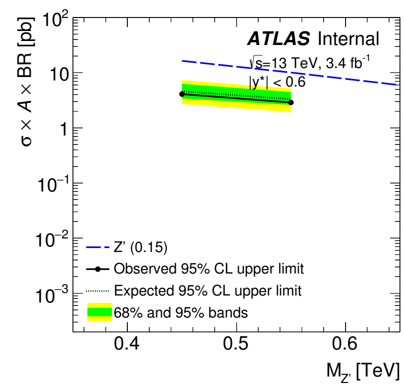


(d) g_{q}=0.3

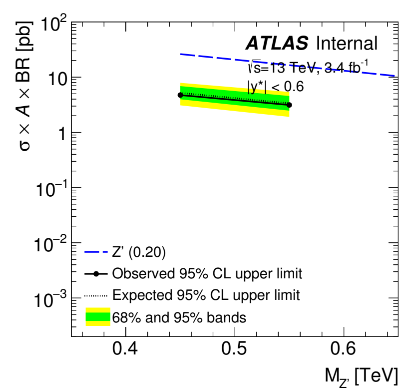
Figure 8.1: Observed and expected limits on Z'models with a range of coupling, using the $|y^*| < 0.6$ selection.



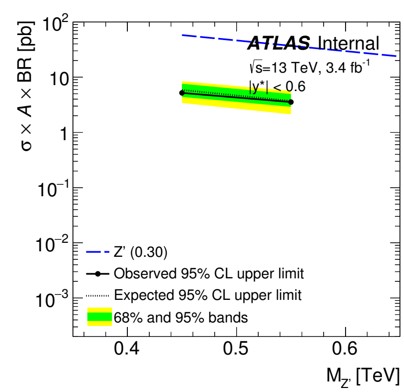
(a) g_{q}=0.1



(b) g_{q}=0.15



(c) g_{q}=0.2



(d) g_{q}=0.3

Figure 8.2: Observed and expected limits on Z'models with a range of coupling, using the $|y^{*}| < 0.3$ selection.

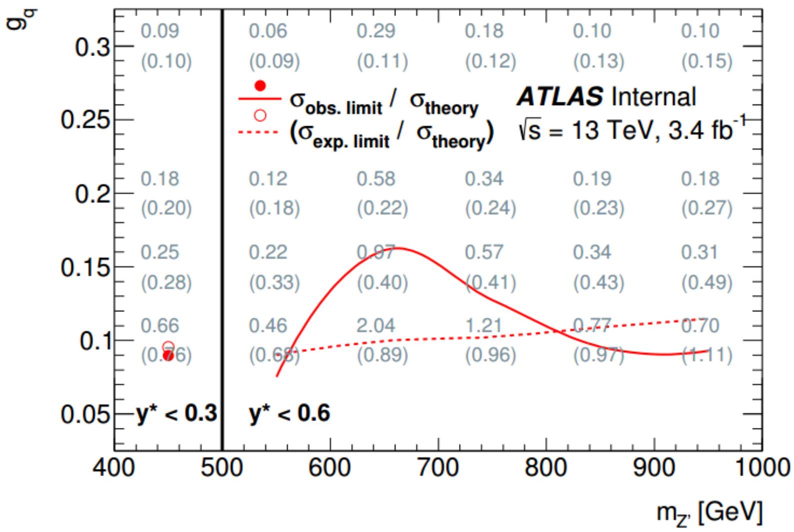


Figure 8.3: The ratio of 95% CL upper limits to cross sections predicted by theory, for Z' models as a function of $M_{\rm med}$ and $g_{\rm q}$. The solid (dashed) red curve, for the $|y^*| < 0.6$ spectra, and filled (empty) marker, for the $|y^*| < 0.3$ spectra, show the observed (expected) excluded limits.

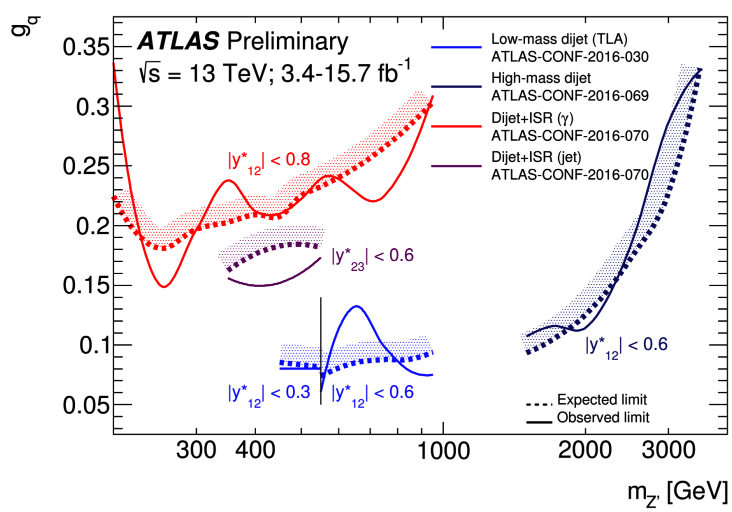
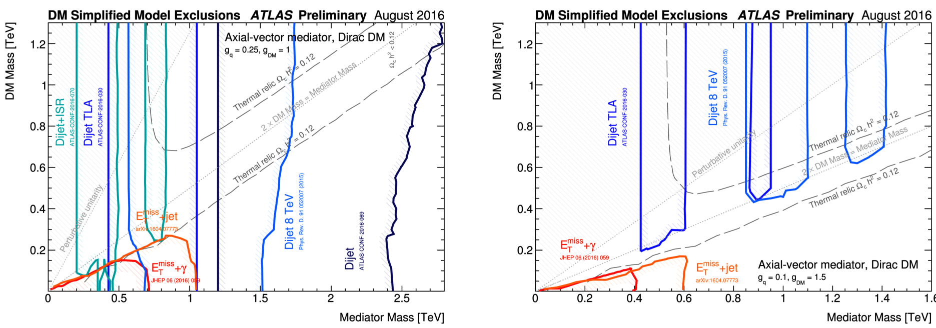


Figure 8.4: The 95% CL upper limits obtained from four ATLAS searches using 2015 and 2016 data on coupling g_{q} as a function of the resonance mass M_{med} for the Z' model. The expected limits from each search are indicated by dotted curves. The selection in y^* of the two jet used to reconstruct the invariant mass are also shown for the four searches. Coupling values above the solid curves are excluded, as long as the signals are narrow enough to be detected using these searches (10% signal width / mass for dijet+ISR and TLA, 15% for high-mass dijets).



(a) g_{q}= 0.25, g_{χ}= 1

(b) g_{q}= 0.1, g_{χ}= 1.5

Figure 8.5: Regions in a dark matter mass–mediator mass plane excluded at 95% CL by a selection of ATLAS dark matter searches. Dashed curves labeled "thermal relic" indicate combinations of dark matter and mediator mass that are consistent with a dark matter density of $\Omega_c = 0.12h^2$ and a standard thermal history, as computed in MadDM[41]. Between the two curves, annihilation processes described by the simplified model deplete Ω_c bleow 0.12 h^2. A dotted curve indicates the kinematic threshold where the mediator can decay on-shell into dark matter. Points in the plane where the model is in tension with the perturbative unitary considerations of [42] are indicated by the shaded triangle at the upper left. The exclusion regions, relic density contours, and unitarity curve are not applicable to other choices of coupling values or model.

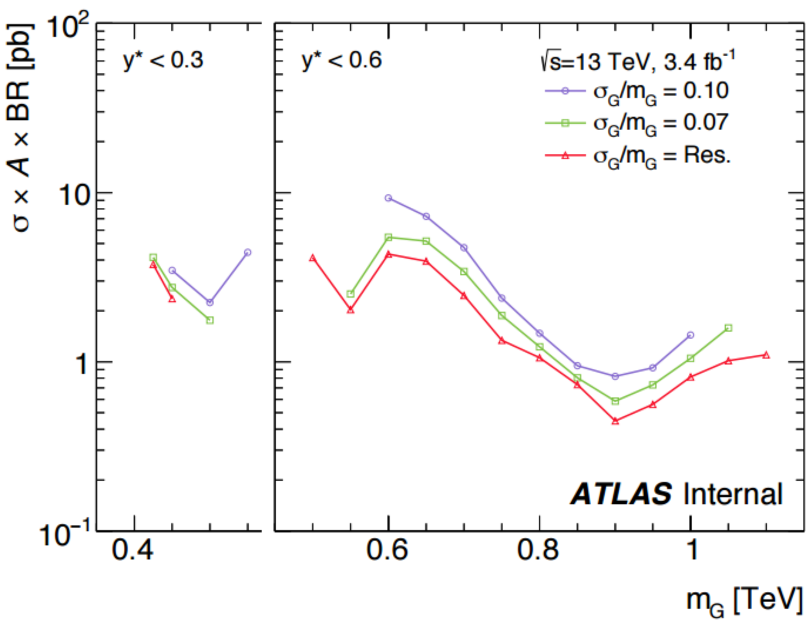


Figure 8.6: 95% CL upper limits on $\sigma \times \mathcal{A} \times BR$ for hypothetical Gaussian shaped signals as a function of the Gaussiam mean m_G. Three scenarios were considered for the width of the signal: a relative width of 10%, 7%, and the detector resolution at that mass.

Chapter 9

Conclusions

This thesis presents a search for a resonance in dijet final states using Trigger-Level-Analysis, which employs a new strategy in ATLAS to increase sensitivity in low mass regions. Masses below 1 TeV are prominent search regions for interactions between quarks and dark matter through the decay of a mediator particle. Currently, The bandwidth allocated to inclusive jet trigger forces ATLAS to apply prescale factors on these triggers in order to record data in a manageable rate. As a consequence, dijet searches in ATLAS have been focusing on larger masses, and in Run-2 dijet masses below 1.1 TeV have not been yet covered.

The jets used in the Trigger-Level-Analysis are triggered by a single unprescaled L1_J75 trigger. They are reconstructed at the HLT level and only a subset of the jet properties is saved to disk. Using this reduced data format lowers the average event size from $\sim$ 1-2 MB to $\sim$ 30 kB. This increases the available statistics in sub-TeV dijet masses, but introduces other complications; Some of the information discarded is generally used to veto events based on jet properties, as well as in the jet calibration scheme. To address this, alternative methods were developed when possible and the performance of the jets was validated through various tests before proceeding to analyze the data.

The large amount of statistics available in the Trigger-Level-Analysis imposes difficulties for a data-driven background estimation; As dictated by Poisson statistics, an increment in event counts decreases the relative statistical uncertainties on the measurement. Fitting a simple analytic function to the data, as generally done in dijet searches, is not guaranteed to work when the precision of the data is very high. In addition, in order to remain sensitive to new physics, the uncertainties on the background estimation are required to remain low. The procedure for modeling the background and searching the data for a resonance is robust against various possible scenarios. It takes into account the possibility that the data is too precise to be modeled using the current methods, and was carefully designed to distinguish this case from that of a potential new physics signal. This procedure was rigorously tested in MC before applying it on data.

During 2015, 3.4 fb^{-1} of data were collected in the Trigger-Level-Analysis. The data shows agreement with the Standard Model as the most prominent resonance corresponds to a p-value of 0.439 (0.186 when using a tighter constrain on the dijet rapidity difference). With no sign of physics beyond the Standard Model, exclusion bounds were set on various Z' models, as well as on generic gaussian-like core models.

This is the first analysis in ATLAS to use jets reconstructed at trigger level for a physics search. This procedure is to be repeated both using the data collected in 2016, and in further analyses.

Bibliography

[1] P. Langacker. *The standard model and beyond*. Series in High Energy Physics, Cosmology and Gravitation. Boca Raton, FL: Taylor and Francis, 2010. URL: https://cds.cern.ch/record/1226768.

[2] C. P. Burgess and G. D. Moore. *The standard model: A primer*. Cambridge University Press, 2006. ISBN: 9780511254857, 9781107404267, 9780521860369.

[3] G. Aad and others (ATLAS Collaboration). 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'. In: *Phys. Lett.* B716 (2012), pp. 10–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[4] V. A. Bednyakov, N. D. Giokaris and A. V. Bednyakov. 'On Higgs mass generation mechanism in the Standard Model'. In: *Phys. Part. Nucl.* 39 (2008), pp. 10–16. DOI: 10.1007/s11496-008-1002-9. arXiv: hep-ph/0703280 [HEP-PH].

[5] R. Aaij et al. 'Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays'. In: (2016). arXiv: 1606.07895 [hep-ex].

[6] R. Aaij et al. 'Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays'. In: *Phys. Rev. Lett.* 115 (2015), p. 072001. DOI: 10.1103/PhysRevLett.115.072001. arXiv: 1507.03414 [hep-ex].

[7] U. Baur, I. Hinchliffe and D. Zeppenfeld. 'Excited Quark Production at Hadron Colliders'. In: *International Journal of Modern Physics A* 02.04 (1987), pp. 1285–1297. DOI: 10.1142/S0217751X87000661. eprint: http://www.worldscientific.com/doi/pdf/10.1142/

S0217751X87000661. URL: http://www.worldscientific.com/doi/abs/10.1142/S0217751X87000661.

[8] D. M. Gingrich. 'Quantum black holes with charge, colour, and spin at the LHC'. In: *J. Phys.* G37 (2010), p. 105008. DOI: 10.1088/0954-3899/37/10/105008. arXiv: 0912.0826 [hep-ph].

[9] T. Han, I. Lewis and Z. Liu. 'Colored Resonant Signals at the LHC: Largest Rate and Simplest Topology'. In: *JHEP* 12 (2010), p. 085. DOI: 10.1007/JHEP12(2010)085. arXiv: 1010.4309 [hep-ph].

[10] T. AAkesson et al. *Search for New Phenomena in Dijet Events with the ATLAS Detector at $\sqrt{s}$=13 TeV with the full 2015 dataset*. Tech. rep. ATL-COM-PHYS-2015-1205. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2055242.

[11] M. Chala et al. 'Constraining Dark Sectors with Monojets and Dijets'. In: *JHEP* 07 (2015), p. 089. DOI: 10.1007/JHEP07(2015)089. arXiv: 1503.05916 [hep-ph].

[12] J. M. Gaskins. 'A review of indirect searches for particle dark matter'. In: (2016). arXiv: 1604.00014 [astro-ph.HE].

[13] T. Marrodán Undagoitia and L. Rauch. 'Dark matter direct-detection experiments'. In: *J. Phys.* G43.1 (2016), p. 013001. DOI: 10.1088/0954-3899/43/1/013001. arXiv: 1509.08767 [physics.ins-det].

[14] R. M. Harris and K. Kousouris. 'Searches for Dijet Resonances at Hadron Colliders'. In: *Int. J. Mod. Phys.* A26 (2011), pp. 5005–5055. DOI: 10.1142/S0217751X11054905. arXiv: 1110.5302 [hep-ex].

[15] G. P. Salam. 'Towards Jetography'. In: *Eur. Phys. J.* C67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].

[16] P. Loch and the Atlas Collaboration. 'Jet measurements in ATLAS'. In: *Journal of Physics: Conference Series* 323.1 (2011), p. 012002. URL: http://stacks.iop.org/1742-6596/323/i=1/a=012002.

[17] M. Cacciari, G. P. Salam and G. Soyez. 'The Anti-k(t) jet clustering algorithm'. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[18] The ALICE Collaboration. 'The ALICE experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08002. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08002.

[19] G. Aad and others (ATLAS Collaboration). 'The ATLAS Experiment at the CERN Large Hadron Collider'. In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[20] S. Chatrchyan et al. 'The CMS experiment at the CERN LHC'. In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[21] A. Augusto Alves Jr. et al. 'The LHCb Detector at the LHC'. In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[22] E. Stanecka. 'The ATLAS Inner Detector operation, data quality and tracking performance'. In: *Proceedings, 32nd International Symposium on Physics in Collision (PIC 2012)*. 2013, pp. 383–388. arXiv: 1303.3630 [physics.ins-det]. URL: https://inspirehep.net/record/1223989/files/arXiv:1303.3630.pdf.

[23] M. Backhaus. 'The upgraded Pixel Detector of the {ATLAS} Experiment for Run 2 at the Large Hadron Collider'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 831 (2016). Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors, pp. 65 –70. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/j.nima.2016.05.018. URL: http://www.sciencedirect.com/science/article/pii/S0168900216303837.

[24] S. Prince. 'The updated ATLAS Jet Trigger for the LHC Run II'. In: *Meeting of the APS Division of Particles and Fields (DPF 2015) Ann Arbor, Michigan, USA, August 4-8, 2015*. 2015. arXiv: 1511.00972 [physics.ins-det]. URL: https://inspirehep.net/record/1402596/files/arXiv:1511.00972.pdf.

[25] G. Aad and others (ATLAS Collaboration). 'Search for new phenomena in dijet mass and angular distributions from pp collisions at with the {ATLAS} detector'. In: *Physics Letters B* (2016), pp. –. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/j.physletb.2016.01.032. URL: http://www.sciencedirect.com/science/article/pii/S0370269316000447.

[26] R. Abreu et al. *Data Scouting in ATLAS: Trigger, Tier0 and Analysis*. Tech. rep. ATL-COM-GEN-2014-010. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1957173.

[27] R. Abreu et al. *Search for low-mass dijet resonances with Trigger-Level Analysis*. Tech. rep. ATL-COM-PHYS-2016-049. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2125945.

[28] T. Sjostrand et al. 'A Brief Introduction to PYTHIA 8.1'. In: *Comput. Phys. Commun.* 178 (11 2008), pp. 852–867. arXiv: 0710.3820 [hep-ph].

[29] "Aad, G. and others (ATLAS Collaboration). *ATLAS Run 1 Pythia 8 tunes*. 2014. URL: http://cds.cern.ch/record/1966419.

[30] S. Carrazza, S. Forte and J. Rojo. 'Parton Distributions and Event Generators'. In: *Proceedings, 43rd International Symposium on Multiparticle Dynamics (ISMD 13)*. 2013, pp. 89–96. arXiv: 1311.5887 [hep-ph]. URL: http://inspirehep.net/record/1266070/files/arXiv:1311.5887.pdf.

[31] G. Aad and others (ATLAS Collaboration). 'Search for new phenomena in the dijet mass distribution using pp collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector'. In: *Phys. Rev. D* 91 (5 2015), p. 052007. DOI: 10.1103/PhysRevD.91.052007. URL: http://link.aps.org/doi/10.1103/PhysRevD.91.052007.

[32] J. Alitti et al. 'A Measurement of two jet decays of the W and Z bosons at the CERN $\bar{p}p$ collider'. In: *Z.Phys.* C49 (1991), pp. 17–28. DOI: 10.1007/BF01570793.

[33] G. Aad and others (ATLAS Collaboration). 'Search for strong gravity in multijet final states produced in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC'. In: (2015). arXiv: 1512.02586 [hep-ex].

[34] G. Aad and others (ATLAS Collaboration). *Search for resonances decaying to photon pairs in 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2015-081. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2114853.

[35] G. Choudalakis. 'On hypothesis testing, trials factor, hypertests and the BumpHunter'. In: *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*. 2011. arXiv: 1101.0390 [physics.data-an]. URL: https://inspirehep.net/record/883244/files/arXiv:1101.0390.pdf.

[36] E. Gross and O. Vitells. 'Trial factors or the look elsewhere effect in high energy physics'. In: *Eur. Phys. J.* C70 (2010), pp. 525–530. DOI: 10.1140/epjc/s10052-010-1470-8. arXiv: 1005.1891 [physics.data-an].

[37] G. Cowan. 'Statistics for Searches at the LHC'. In: *LHC Phenomenology*. Springer, 2015, pp. 321–355.

[38] A. Beskos and A. Stuart. 'Computational complexity of Metropolis-Hastings methods in high dimensions'. In: *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Springer, 2009, pp. 61–71.

[39] G. Aad and others (ATLAS Collaboration). *Search for New Phenomena in Dijet Events with the ATLAS Detector at $\sqrt{s}$=13 TeV with 2015 and 2016 data*. Tech. rep. ATLAS-CONF-2016-069. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2206212.

[40] G. Aad and others (ATLAS Collaboration). *Search for new light resonances decaying to jet pairs and produced in association with a photon or a jet in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2016-070. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2206221.

[41] M. Backovic, K. Kong and M. McCaskey. 'MadDM v.1.0: Computation of Dark Matter Relic Abundance Using MadGraph5'. In: *Physics of the Dark Universe* 5-6 (2014), pp. 18–28. DOI: 10.1016/j.dark.2014.04.001. arXiv: 1308.4955 [hep-ph].

[42] F. Kahlhoefer et al. 'Implications of unitarity and gauge invariance for simplified dark matter models'. In: *JHEP* 02 (2016), p. 016. DOI: 10.1007/JHEP02(2016)016. arXiv: 1510.02110 [hep-ph].

תקציר

מחקר זה עוסק בחיפוש אחר סימנים לפיסיקה חדשה בהתפלגות המסה של צמדי-סילונים מהתנגשויות פרוטון-פרוטון באנרגיית מרכז מסה של 13 TeV בגלאי אטלס שבמאיץ ההדרונים הגדול. חיפושים באטלס המשתמשים באירועים של צמדי-סילונים במסות נמוכות מ- 1 TeV מוגבלים מבחינה סטטיסטית עקב הסינון המוקדם הנעשה עוד בשלב הדרבון (רישום ראשוני של הנתונים לפני עיבודם). קצב היצירה הגבוה של סילונים במודל הסטנדרטי מאלץ את אטלס לברור את האירועים המעובדים בשלב מוקדם זה של רישום הנתונים, ולהשתמש רק בחלקם בשלב הניתוח הסטטיסטי. ע"י עיבוד ברמת הדירבון (Trigger-Level-Analysis) ניתן להתגבר על חלק מן ההגבלות הללו ולהגדיל את הרגישות לזיהוי תהליכים במסות נמוכות. בשיטה זו תוצרי ההתנגשות נשמרים במתכונת מצומצמת אשר כוללת רק את המידע הנחוץ על מנת לבצע חיפוש בהתפלגות המסה של צמדי-הסילונים. השימוש בתצורה זאת מאפשר קצב רישום מהיר הרבה יותר של אירועים מסוג זה. החיפוש נעשה במסות שנעות בין 394 GeV ועד 1.24 TeV וכולל 3.4 fb^{-1} של מידע שנאסף במהלך שנת 2015. תוצאות המחקר הראו התאמה למודל הסטנדרטי ולא נצפו סימנים מצביעים לפיסיקה מעבר למודל הסטנדרטי. בהתאם לכך, נקבעו חסמים על האפשרות לקיומן של תיאוריות המנבאות הרחבה של המודל הסטנדרטי של חלקיקי היסוד.

TEL AVIV UNIVERSITY
RAYMOND AND BEVERLY SACKLER
FACULTY OF EXACT SCIENCES
SCHOOL OF PHYSICS & ASTRONOMY

אוניברסיטת תל-אביב
הפקולטה למדעים מדוייקים
ע"ש ריימונד ובברלי סאקלר
בית הספר לפיסיקה ואסטרונומיה

חפוש סימנים לפיסיקה חדשה בהתפלגויות צמדי-סילונים במאיץ הפרוטונים LHC

גיא קורן

חיבור מדעי המוגש כחלק מדרישות התואר "מוסמך אוניברסיטה" במחלקה לאנרגיות גבוהות באוניברסיטת תל אביב תחת הדרכת פרופסור **ארז עציון**.

ספטמבר 2016

Guy.koren@cern.ch