THE UNIVERSITY OF CHICAGO

SEARCHES FOR SUB-TEV HADRONIC RESONANCES WITH THE ATLAS DETECTOR

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
KAROL KRIZKA

CHICAGO, ILLINOIS
MARCH 2017

To my parents.

The lesson of the random walk is that one will always return to the pub.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to acknowledge

Young-Kee Kim for her endless guidance and being a great role model,

Lauren Tompkins for teaching me what a healthy work/life balance is,

Yasuyuki Okamura for teaching me that details matter,

John Alison for teaching me that not all details matter,

Alberto Annovi for trusting me to lead the FTK integration effort,

Gabriel Facini for prioritizing results over bureaucracy,

David, Jim, Mark, Mel and Young-Kee for building one of the best groups in the collaboration,

Jeff and Patrick for being my compatriots at CERN,

the entire UC ATLAS group for many stimulating discussions,

Antonio, Carlo, Caterina, Elisa, Fabrizio, Laurie, Lydia, Nataliia, Prim and Rui as co-analyzers,

Team Waseda (Naoki, Tomoya and Takashi) for being great drinking buddies,

the CERN security for not shutting down the Chicago barbecues,

the XeSMP working group (Ben, Bobak, John, Shannon and Vince) for lessons in statistics,

the LHC accelerator team and the ATLAS collaboration for making my work possible,

my graduate cohort for being an exemplar group of scientist to be associated with,

with special thanks to Alex, Jake, Nate, Ravi, Rhys and Sim.

I would like to apologize to the staff at Restaurant La Meyrinoise and thank them for the endless supply of grappa and limoncello.

ABSTRACT

The Large Hadron Collider has recently began colliding proton beams at a record center-of-mass energy $\sqrt{s} = 13$ TeV. This allows the LHC experiments to drastically improve searches for heavy new particles. However at the same time, the more complicated environment means an increase in the trigger thresholds. Consequently a viable set of theories at lower energies remains unexplored. This thesis describes three ATLAS analyses that used different techniques to target new particles with sub-TeV mass. The first analysis searches for the Higgs boson produced via vector-boson fusion and decaying to two bottom quarks in in $\sqrt{s} = 8$ TeV dataset. The presence of the extra jets from vector-boson fusion and b-tagging is used to reduce the event rate to a manageable level. The second analysis searches for generic resonances decaying to two b-quarks with masses between 600 GeV to 1.2 TeV. The use of b-tagging in the trigger allows it to set unique limits in the sub-TeV mass range. The third analysis searches for hadronic resonances produced in associtaion with energetic initial state radiation (ISR). The use of the ISR as a trigger decorrelates the resonance mass from the trigger efficiency. The result is world-class limits on new particles with masses between 200 GeV to 500 GeV. This space is interesting for particles mediating between the Standard Model and the dark sector. The dark matter relic density prefers $\mathcal{O}(100\text{ GeV})$ mediators, yet the most stringent constraints in this region previously came from the pre-LHC era.

CHAPTER 1

INTRODUCTION

The ultimate goal of physics is to discover the basic principles that govern our world. The key to this is understanding the elementary constituents of our universe and how they interact with each other. The approach of reductionalism was started in ancient Greece, with Democritus claiming that everyting was made up of invidisible partiles called atoms. The first scientific confirmation of the atom came with John Dalton's discovery that elements always react in integer ratios. However Dalton's atom was further broken down into a positive core (the nucleus) surounded by electrons at the turn of the 20^{th} century. Then special relativity and quantum mechanics unleached a whole new era of particle physics.

The marriage of special relativity and quantum mechanics means that energy and wavelength are the same property. And like a microscope needs a short waveleght to resolve small objects, very high energies are needed to resolve the smallest particles possible. A simple source of energetic particles are the violent processes happening in the cosmos. Their products arrive to us in the form of cosmic rays. Through the use of cloud chambers and placing emulsion films at high altitudes, physicists were able to discover and measure particles with energies much higher than available on earth. This lead to the discovery of the positron[14], the muon[82] and the pion[83]. However the field exploded with the invention of particle accelerators. The accelerators can create almost the same energies as cosmic rays, but in a controlled and reproducible environment. Using them, physicists were able to establish our best understanding of elementary particles known as the Standard Model (SM).

The content of the Standard Model is divided into two categories; fermions and bosons. The fermions are spin-1/2 particles and make up the matter. The bosons have integer spin and are the result of enforcing the $SU(3)_C \times SU(2)_W \times U(1)_Y$ symmetry. To satisfy the symmetry, new fields and interactions have to be introduced into the SM lagrangian.

The interactions due to the new fields can be intepreted as forces at macroscopic scales. Furthermore the symmetry is only satisfied if all particles are massless. But our observations tells us that this is not true. The solution is the addition of a spin-0 doublet called the Higgs field. The Higgs field has a potential with a non-zero vaccum expectation value (vev). The ground state of the field then spontaneously breaks the symmetry of the model and particles coupled to the Higgs doublet obtain mass.

The different spin-1 bosons correspond to different parts of the overall symmetry. The Standard Model symmetry can be broken down into electroweak (QED) and colour (QCD) theories. The $SU(3)_C$ corresponds to the QCD theory and any particles charged under it must couple to the coloured massless boson called the gluon. The gluon is known as the carrier of the strong force. The QED theory, corresponding to the $SU(2)_W \times U(1)_Y$ symmetry, is spontagnously broken by the Higgs mechanism. The resulting mass-eigenstates are the Z^0, $W^{\pm}$ bosons and the photon. The first three bosons are massive and correspond to the weak force. The photon is massless and is the cause of the electromagnetic force. The Higgs mechanism is said to have unified these two forces into a single electroweak force.

The matter component of the Standard Model consists of fermions divided into quark and lepton families. Each family has been shown experimentally to consists of three generations categorized by mass. The quarks are charged under both QCD and QED. For each generation, there is a quark with electric charge of $+2/3$ ("up"-type) and charge of $-1/3$ ("down"-type). In total there are six quarks called *up*, *down*, *strange*, *charm*, *bottom* and *top*. They are often referred to by the first letter their name. The leptons interact only via the electroweak force. Like quarks, they are grouped into three generations based on increasing mass and each generation consists of two leptons. The two leptons in each generation are a particle with an electric charge of -1 and a neutrical particle with couplings only to the massive gauge bosons. The generations are known by the name of the charged particle:

	Particle	Mass (MeV)	Electric Charge	Colour Charge
Quarks	up	$2.2^{+0.6}_{-0.4}$	$+2/3$	Y
	down	$4.7^{+0.5}_{-0.4}$	$-1/3$	Y
	charm	127 ± 30	$+2/3$	Y
	stange	96^{+8}_{-4}	$-1/3$	Y
	top	$(173.21^{+0.51}_{-0.71}) \times 10^3$	$-1/3$	Y
	bottom	$(4.8^{+0.04}_{-0.03}) \times 10^3$	$+2/3$	Y
Leptons	electron (e^-)	$0.5109989461 \pm 0.0000000031$	-1	-
	muon (μ^-)	$105.6583715 \pm 0.0000035$	-1	-
	tau (τ^-)	1776.86 ± 0.12	-1	-
	ν_e,ν_μ,ν_τ	0	0	-
Bosons	gluon	0	-	Y
	photon (γ)	0	0	-
	Z^0	$(91.1876 \pm 0.0021) \times 10^3$	0	-
	$W^\pm$	$(80.385 \pm 0.015) \times 10^3$	0	-
	Higgs	$(125.09 \pm 0.21 \pm 0.11) \times 10^3$	-	-

Table 1.1: The particle content of the Standard Model, not including the anti-particles of the fermions. The masses of are taken from the Particle Data Group[85]. The $\overline{MS}$ defintion is used for quark mass.

electron, *muon* and *tau*. The neutral particle is called the neutrino of the given flavour. The masses of the fermions, with the exception of the neutrinos, are also generated by the Higgs mechanism and are free parameters of the model. The neutralino is taken to be massless. Furthermore all of the fermions, again with the possible exception of the neutralinos, are of Dirac-type. That is they have a corresponding anti-particle partner. Whether a separate anti-neutrino also exists or the neutrino is its own anti-particle has not yet been determined.

The full particle content of the Standard Model is summarized in table 1.1.

The Standard Model has been tested over the last few decades by many experiments and shown to be very robust. The last missing piece, the Higgs boson, has been discovered in 2012[21][59]. The prediction for cross-section of common processes at the Large Hadron Collider is compared with the values measured by the ATLAS detector in figure 1.1. There are no significant differences. Yet we know that it still not a complete model. The neutrialinos are treated as massless, even though oscillation experiments prove that they are not[63]. There is not enough asymmetry between matter and anti-matter to be consistent with our

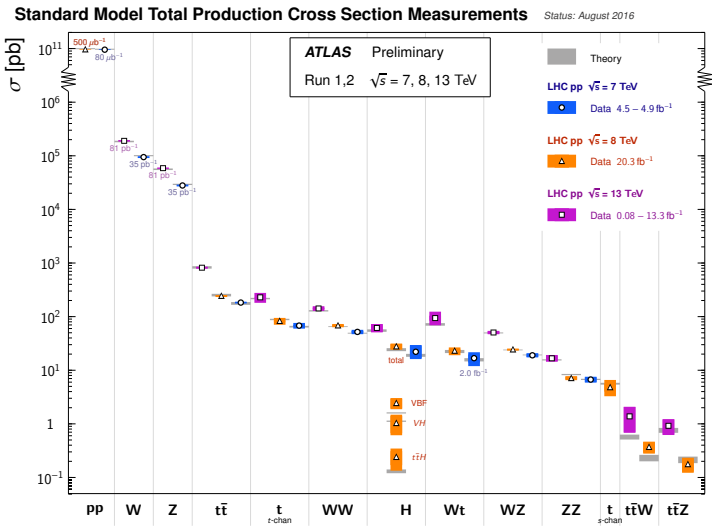


Figure 1.1: List of the common processes at the Large Hadron Collider. The cross-section as measured by ATLAS (points) and their theoretical prediction (gray band) are shown.

universe[9]. Corrections to the Higgs mass have a big effect, unless the model parameters are fine-tuned[69]. There are no particles to account for the invisible mass (dark matter) seen in galactic measurements[87][62][80], no force carrier for gravity or an explanation for dark energy[84] exist.

Many Beyond the Standard Model (BSM, a generic term for new physics) theories have been proposed. They result in a variety of physical signatures. Thus the ability to cast the widest net possible is needed to determine the correct theory. Hadron collider experiments meet this requirement very well. The generic design of the detectors allows one to probe a multitude of final states. For example, many precision tests of the Standard Model can be performed and used to hint at new physics. However the main advantage comes from the ability to directly produce new particles. The only limitation of the direct searches is that the collider center-of-mass energy ($\sqrt{s}$) needs to be larger than the energy scale of new physics. The Large Hadron Collider (LHC)[81] currently provides the most energetic collisions at $\sqrt{s} = 13$ TeV. The results of the collisions are being studied by four detectors, two of which target generic physics. The generic detectors are called CMS[56] and ATLAS[16].

The energy and luminosity of the Large Hadron Collider has been increasing over the last few years. During 2011, the LHC delivered 5.61 fb^{-1} of collisions with $\sqrt{s} = 7$ TeV to the experiments. Then in 2012, 23.3 fb^{-1} of $\sqrt{s} = 8$ TeV collisions was delivered. This concluded the time period known as Run 1 and the accelerator complex experienced a scheduled long shutdown (LS1) to allow for upgrades. Run 2 started in the later half of 2015, with the LHC operating at $\sqrt{s} = 13$ TeV and the design instantaneous lumonisty. By the end of 2016, the collider delivered 43.1 fb^{-1} worth of collisions to the experiments and more is planned for the subsequent two years.

One side effect of all the increases is the need to raise the energy thresholds in the ATLAS detector's trigger system. While this does not affect searches for new heavy particles, it does result in an unexplored gap at lower energies. For example, popular simplified models of dark matter contain a mediator to the dark sector. If such a mediator has couplings to quarks, then it will decay to them. The dijet searches can probe this final state and set limits independent of the dark matter properties. However the standard dijet search is trigger-limited to masses above 1 TeV and there is still a possibility of a sub-TeV mediator. In fact, dark matter relic density prefers lighter mediators with masses at the order of a few 100 GeV.

Similarly while the Higgs boson has been discovered, its most dominant decay mode into two bottom quarks has not been observed yet. The b-quark decay channel is also one of two channels that can be used to demonstrate the fermionic couplings of the Higgs boson at the LHC. The relatively light mass of the boson ($m_H = 125$ GeV) makes the resulting b-jets very soft and hard to distinguish from the inclusive QCD background. This document explores several techniques that can access this phase-space in the current environment.

The rest of the document is divided into seven sections. Section 2 describes the ATLAS

detector and section 3 describes the algorithms used to reconstruct and calibrate the outgoing particles. Section 4 describes the VBF $H \to b\bar{b}$ analysis done at the end of Run 1. It uses the distinct topology of the jets produced by the Vector Boson Fusion process to reduce the background. The Run 2 analyses start with section 5, which describes a standardized procedure for all hadronic resonance searches from the ATLAS Exotics group. In sections 6 and 7, the method is applied to a search for sub-TeV hadronic resonances in the diBjet and dijet+ISR channels correspondingly. The diBjet resonance search targets new particles in the mass range 600 GeV-1.2 TeV with a preferential decay into bottom quarks. A new scalar particle motivated by the diphoton resonance search using the 2015 dataset falls into this category. The dijet+ISR search uses the presence of energetic initial state radiation to trigger recording of events with a sub-TeV hadronic resonance. As mentioned, this is an important final state for dark matter mediators with couplings to quarks. Section 8 concludes.

CHAPTER 2

DETECTOR

The ATLAS detector follows the design of a typical collider experiment by layering different sub-detectors around each other. The sub-detectors, providing near-to 4π coverage, allow one to measure the energy four-vectors of almost all outgoing particles and provide handles on their identification. Starting from the middle, there are the silicon trackers (pixel and strip), the Transition Radiation Tracker (TRT), hadronic calorimeter (Tile), electromagnetic calorimeter (LAr) and a muon spectrometer. The muon spectrometer also contains a set of large toroidal magnets providing a 8 T field and giving ATLAS its name.

The result is a tracker covering the region $|\eta| < 2.5$, electromagnetic and hadronic calorimeters covering $|\eta| < 4.9$, and a muon spectrometer for muons with $|\eta| < 2.7$. The detector also provides a symmetric azimuthal coverage[1].

The ATLAS detector has undergone significant improvements during the long shutdown 1 (LS1) after Run 1. Non-significant changes have occurred to the trigger system and the pixel detector. The following subsections provide more details on the state of the sub-detectors during Run 2. Only differences in Run 1 relevant to the Run I VBF $H \to b\bar{b}$ analysis will be pointed out.

2.1 Inner Detector

The ATLAS Inner Detector (ID) records passing charged particles at discrete points. The ID is surrounded by a solenoid magnet, providing a uniform 2 T magnetic field and giving

1. The ATLAS experiment uses a right-handed Certerian coordinate system with the origin located at the nominal collision point. The $\hat{z}$ axis points along the anti-clockwise beam and the $\hat{x}$ axis points towards the ring center. The polar (θ) angle is defined with respect to the beam axis and the azimuthal (ϕ) angle is measured around the beam axis. Pseudo-rapidity is defined as $\eta = -\ln[\tan(\frac{\theta}{2})]$. Angular distances are defined by $\Delta R^2 = \Delta\phi^2 + \Delta\eta^2$.

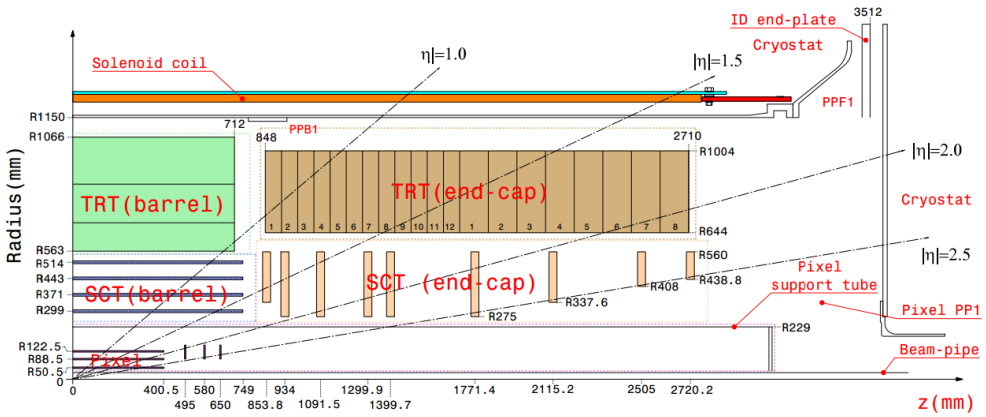


Figure 2.1: Schematic showing the layout of the different components of the inner detector. The newly installed inner b-layer is now shown.

curvature to the charged particles. The measurements, referred to as hits, can be reconstructed as tracks left by the particles' trajectories. The curvature provides a measurement of the particles energy, with the direction corresponding to the sign of the charge.

The Inner Detector consists of three parts; the Pixel tracker (Pixel), the Semi-Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT). It forms a cylinder with radius of 1.2 m and length of 6 m. A diagram showing the complete layout can be seen in figure 2.1. The Pixel and SCT are both silicon-based ionizing detectors. Charged particles from the collisions ionize the active medium, creating electron-hole pairs. The holes and electrons are collected via a 150 V bias voltage. The small area of the active medium allows for precise position measurement of charged particles. The main difference between the Pixel and SCT detectors is that the Pixel detector consists of 2D arrays, while the SCT is made up of strips parallel to the beam-line. The two-dimensional position measurements of the Pixel detector located closest to the interaction beam help to improve the vertex resolution. The one-dimensional SCT is cheaper to construct and allows for a wider coverage at large radius. The TRT consists of straw-tubes and material creating transition radiation.

A single Pixel module is 2x6 cm^2, containing 46080 pixels of 50x400 μm^2 dimension.

The long part of each pixel is aligned parallel to the beam. The sensors are bump-bonded to 16 front-end (FE) chips. To allow for 40 MHz read-out rate, each pixel must have a dedicated connection to a FE chip. But the FE chip, along with a Module Control Chip, can aggregate the data and send it to off-detector electronics using a single optical cable. The data consist of a timestamp and time-over-threshold for all pixels passing a programmable threshold value.

There are 1774 identical Pixel modules in ATLAS. They are organized into three concentric cylinders forming the barrel and three disks in both forward regions. The cylinders, made up of staves containing 13 modules each, are located at radii of 50.5 mm, 88.5 mm and 122.5 mm. The staves are inclined at 20° to remove deadspace by overlapping the modules. The disks contain 8 sectors, with 6 modules each. The support structure contains an integrated cooling mechanism to remove the 15 kW dissipated by the FE chips.

During LS1, a fourth layer has been added closer to the beamline at 3.3 cm radius. It is called the Insertable b-Layer (IBL) and allows for further improvements in vertex resolution. In addition to new read-out electronics, it contains a mix of silicon sensor technologies called planar and 3D. The planar sensors cover most of the active area in the center and are an incremental improvement of the existing Pixel detector. The 3D sensors are a new technology where electrodes pass through the sensor's bulk, allowing for lower bias voltages. This is a new technology for the LHC experiments. The length of the pixels has also shrunk to 50x250 μm^2.

Each Semi-Conductor Tracker module consists of four strip sensors, two at the top and two at the bottom. The sensors themselves consists of p-type silicon strips implanted into n-type active material. The pitch of the strips is 80 μm. The sensors on a single side are daisy-chained to provide 768 strips in 12.8 cm of length (6.4 cm per sensor). The top and bottom pairs of sensors have a stereo angle of $\pm$20 mrad to the module axis. Since each

strip only reports a single dimension, the intersection of the two strips on alternate sides of the module is needed to infer the full 2D coordinate. Following an L1-trigger accept, binary information from the sensors is read-out using 12 custom ASIC read-out chips per module.

In total, the SCT is made up of 4088 modules. They are installed in four concentric cylinders (barrel) and nine disks (endcap) on either side of the detector. Deadspace is reduced by tilting the modules at 11° to allow for overlap. The barrel covers radii between 275 mm and 560 mm. The number of modules per stave varies between 348 to 672, with more modules needed to cover the area at larger radius. The nine disks area placed between 853.8 mm and 2720.2 mm on either side from the detector center. Each disk, with some variation, contains 132 modules.

The Transition Radiation Tracker (TRT) makes up the outermost part of the inner tracker, covering the radius between 0.5 m and 1.5 m. It consists of drift tubes (straws) with a 4 mm diameter outer polymide shell and a 31 μm diameter gold-plated tungsten central wire. The polymide wall is coated with a 0.2 μm aluminum layer on one side and a 5 μm polyurethane layer on the other. The aluminum layer is also protected by a $5-6$ μm layer of graphite-polymide. Carbon fiber is used for further stabilization. Each tube is filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. Any charged particle transversing the TRT will ionize the gas. Freed electrons drift towards grounded central wire due to a -1.5 kV potential. The signal arrival time translates to a 140 μm accuracy on the track's closest approach. An alternate (cheapter) mixture of 70% Ar and 30% CO_2 with degrated electron identification can used during the cosmic runs and comissioning.

The TRT barrel consists of three concentric 144 cm long rings containing a total of 52544 straw tubes divided into 73 layers. The end-caps contain 160 planes of 122880 straws 37 cm long. The barrel straws are parallel to the beam-line while the endcap straws are perpendicular. A track within $|\eta| < 2.0$ is expected to cross at least 36 straws. In the

barrel-endcap transition region ($0.8 < |\eta| < 1.0$), the number of straws hit can be as low as 22. All modules are surrounded by CO_2 to prevent high voltage discharge, Xe leakage and allow for cooling of the drift tubes.

The space between the drift tubes is filled with polypropylene fibers (barrel) and foils (endcap). The polymer serves as the transition radiation material. It causes relativistic particles to emit low energy photons as they transverse its boundary. The amount of radiation depends on the relativistic $\gamma = E/m$ factor and can be used to separate electrons from hadrons. The photons produced by the transition radiation result in much larger signals than minimum ionizing particles. The front-end electronics can distinguish between them by using high and low thresholds.

2.2 Liquid Argon Calorimeter

The closest calorimeter to the interaction point is the liquid argon (LAr) calorimeter. It is a sampling calorimeter that uses liquid argon as the active material and lead as the sampling material. The lead plates promote the creation of electromagnetic showers. The corresponding electrons and photons ionize the argon atoms, which drift to readout electrodes using an applied voltage. The readout electrodes consist of three copper layers separated by polymide sheets. The two outer layers provide the high voltage and the inner layer reads out the signal using capacitive coupling. The electrodes are segmented in η and can be ganged in ϕ to provide a varying resolution.

The components are designed to follow an accordion geometry, shown in figure 2.2. This allows the calorimeter to provide a uniform response over the entire ϕ range without any deadspace. The angles of the accordion change as a function of radius to maintain a constant drift gap. This creates a linear response of the calorimeter as a function of depth.

Three cryostats are used to house the calorimeter. They are separated into a barrel

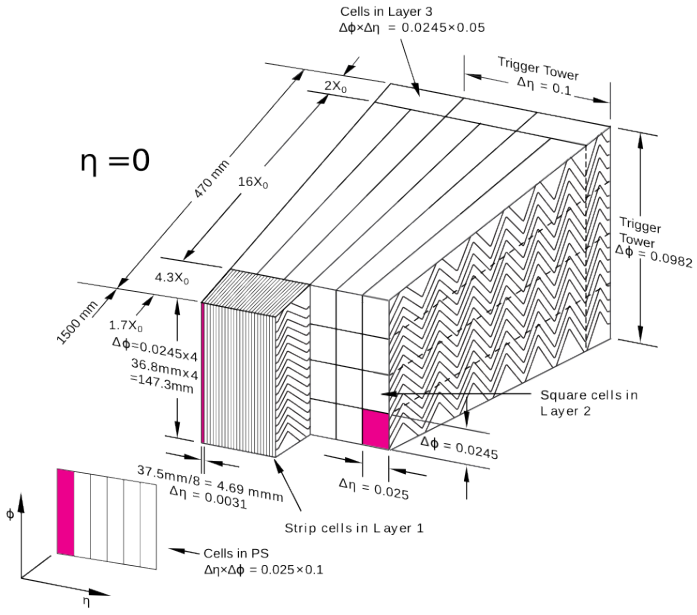


Figure 2.2: Diagram demonstrating the accordion design of a LAr calorimeter barrel module[15]. The granularity of the three layers in ϕ and η is also shown.

($|\eta| < 1.475$) and two endcap ($1.375 < |\eta| < 4.9$) regions. There is also a thin presampler over the entire η range. The barrel calorimeter is divided into two 3.2 m long half-barrels spanning the radius from 2.8 m to 4 m. It can contain radiation lengths ranging from 22 X_0 in the central region to 33 X_0 at $|\eta| = 1.3$. Each half-barrel is divided into 16 modules along ϕ for mechanical reasons. Each module is further divided in the radial direction into three layers with decreasing granularity. High granularity in the first two layers is important for determining the direction of photons. The third layer only receives the tail end of the electromagnetic shower, so a high angular resolution is not necessary.

The forward cryostats house three calorimeters; the electromagnetic endcap calorimeter (EMEC, $1.375 < |\eta| < 3.2$), the hadronic endcap calorimeter (HEC, $1.5 < |\eta| < 3.2$) and a forward calorimeter (FCal, $3.1 < |\eta| < 4.9$). All three are based around liquid argon sampling material and use the same accordion design as the barrel calorimeter with the waves perpendicular to the beam. The HEC and FCal use a copper and copper-tungsten absorbers correspondingly. The EMEC has an active thickness between $24X_0$ to $38X_0$ and is divided

into three layers with decreasing granularity.

The electrodes are kept at an operating voltage of 2000 V. For the constant 2.1 mm gap between the absorber and electrode, the drift time of 450 ns is larger than the 25 ns bunch spacing of the accelerator. This can cause the signal to be present in several collisions, creating out-of-time pile-up. Signal shaping is necessary to extract precise timing information to distinguish energy deposits between adjacent events.

2.3 Hadronic Calorimeter

The ATLAS experiment contains several calorimeters aimed at measuring the energy of strongly interacting particles. The Tile calorimeter sits in the barrel and covers the region $|\eta| < 1.7$. The η range is further extended by the hadronic end-cap calorimeter (HEC) and the forward calorimeter (FCal). The HEC and FCal are described in section 2.2.

The Tile calorimeter is a sampling detector made up of repeating steel and scintillator tiles. The steel promotes hadronization of strongly interacting particles. The size of calorimeter, spanning the radius from 2.8 m to 4.25 m, provides radial interaction depth of 7.4λ. The tiles are mounted perpendicular to the beam and have a 4.7:1 steel-scintillator ratio by volume. Longitudinally the Tile calorimeter is subdivided into three parts. A 5.8 m long central barrel is surrounded by the 2.6 m long extended barrels on each side. The three barrels are further subdivided into 64 modules along the azimuth.

The parton shower induces ultraviolet light in the scintillator, which is then shifted by the read-out fibers into the visible spectrum. The read-out fibers bring the scintillation light to photomultiplier tubes (PMT) located in an aluminum support structure (drawer) located radially outside. The fibers are grouped to provide cells 0.1×0.1 in $\Delta\eta \times \Delta\phi$ for the first two radial layers and 0.2×0.1 for the third. The two sides of the scintillator are readout by independent PMTs to provide redundancy and partially equalize signals from particles

entering at different angles. The front-end electronics, located in the drawer, also calculate analogue sums of the cells to form trigger towers for the L1 trigger.

2.4 Trigger and Data Acquisition System

The ATLAS trigger and data acquisition (TDAQ) system is required to handle input event rates of up to 40 MHz, corresponding to the 25 ns bunch spacing of the LHC. This task has become more complicated in Run 2 with the increases in the center-of-mass energy, instantaneous luminosity and the number of proton-proton interactions per collision. To maintain the physics goals of the experiment, the trigger system has been upgraded during the LS1 period between Run 1 and Run 2[43]. This section describes the TDAQ system in Run 2 after the upgrades, with notable differences mentioned at the end.

The current trigger system consists of two levels. The level 1 (L1) trigger is hardware based, using a multitude of custom electronic boards to reduce the detector rate from 40 MHz to 100 kHz. The High Level Trigger (HLT) employs a computing farm made up of commercial hardware to run offline-like reconstruction on the events passing L1 and reduce the rate further down to 1 kHz. The output of the HLT is then saved to disk.

The L1 trigger decision is based on the partial outputs of the muon spectrometer and the calorimeters. The multiplicity of objects above a certain E_T threshold identified by the L1 subsystems are passed to the Central Trigger Processor (CTP). The output of the Minimum Bias Trigger Scintillators (MBTS), LUCID Chernekov counters and Zero-Degree Calorimeter (ZDC) is also used as CTP input. The CTP holds a programmable look-up table defining up to 256 combinations of input signals to trigger an L1 accept. The L1 accept causes all the Read-Out Drivers (RODs) to send full detector data into the Read-Out System (ROS) buffers that feed the HLT. The objects identified by the Level 1 trigger are called Regions of Interest (RoI) and can be used to seed the HLT trigger algorithms.

The Level-1 muon trigger looks for coincidences in different layers of the muon chambers. The strip within which muon hits are considered is called a *road*. The width of the road is correlated with the p_T of the muon. Extra logic exists to reject muons not originating from the primary vertex.

The logic of the Level-1 calorimeter trigger (L1Calo) is more complicated as it tries to identify jets, photons, electrons and calculate transverse missing energy. Its input is coarse calorimeter clusters called trigger towers. The trigger towers are 0.1×0.1 in $\Delta\eta \times \Delta\phi$ (larger at higher $|\eta|$) and are provided by the front-end electronics. A sliding window algorithm is run to identify clusters corresponding to physics objects. The electron/photon/tau identification is performed by the Cluster Processor (CP). The CP looks for 2×2 clusters in the electromagnetic calorimeter only. It can also apply energy isolation by considering the 12 surrounding towers or veto on large hadronic activity in the corresponding hadron calorimeter 2×2 window. The Jet/Energy-sum Processor (JEP) is responsible for using jet elements (2×2 trigger tower sums) to identify jets and calculate scalar jet E_T sums and missing transverse energy. The jet RoI's are identified by searching for 4×4 or 8×8 windows with the sum of the hadronic and electromagnetic calorimeter energy exceeding a given threshold and being maximized inside the 2×2 central core.

The High Level Trigger uses commercial computer servers running Scientific Linux CERN to perform event reconstruction. The use of the lower input rate, full detector information and software allows for more flexibility in the trigger decisions. The same algorithms as used for the offline reconstruction and object identification are employed. The trigger decisions are organized into trigger chains and use the information about calibrated object multiplicity and advanced photon/electron/b-jet identification.

Due to computational complexity of track reconstruction, it is only performed for a subset of trigger chains (ie: b-jet triggers) and in Regions of Interest only. This prevents the

full calibration chain and advanced pile-up suppression to be run, since some of the inputs depend on track information. Full tracking information will be available in 2017 via the FastTracKer upgrade[89].

The output of the HLT is sent to the Tier-0 computing farm located at CERN to run the full offline reconstruction software suite.

The Run 1 TDAQ architecture followed a similar design, but operated at a lower rate. The main difference was the High Level Trigger being divided into the Level 2 (L2) and the Event Filter (EF) triggers. The L2 trigger used the 75 kHz output rate of the L1 trigger and reduced it to 3.5 kHz by running partial event reconstruction inside RoIs. The EF then based its decision on full event reconstruction to save events to disk at 400-600 Hz. The overall L1 design stayed the same, but the hardware was upgraded to allow for a higher bandwidth and more advanced corrections. The scope of the L1 algorithms benefited from the increased flexbility in the switch from Application-Specific Integrated Circuits (ASIC) to generic Field Programmable Gate Arrays (FPGA).

CHAPTER 3

OBJECT RECONSTRUCTION

The ATLAS event reconstruction employs many algorithms to identify and calibrate the physics objects in the LHC collisions. To account for any differences between data and simulation, the algorithms need to be calibrated and consequent uncertainties assigned. This chapter describes the algorithms relevant to the analyses described in this document. Section 3.1 describes photon identification, section 3.2 describes jet reconstruction and calibration, section 3.3 describes identifications of jets produced by bottom quarks and section 3.4 describes quark-like vs gluon-like jet discrimination.

3.1 Photons

The photon identification in the ATLAS software[42][44] deals with prompt photons only. Prompt photons are produced from the hard scatter and deposit all of their energy in narrow clusters in the electromagnetic calorimeter. The search for the clusters is seeded via a sliding window ($\Delta\eta \times \Delta\phi = 0.075 \times 0.123$) algorithm. Photons can be either converted or unconverted. Converted photons pair produce electron and a positron before reaching the calorimeter. The presence of two tracks, without any hits in the first pixel layer, roughly matched to the EM cluster is the fingerprint of a converted photon.

The background to prompt photons are photons originating from decays of hadrons and misidentified jets (fake photons). The fine segmentation in the first two layers of LAr calorimeter can be used to distinguish the former. Furthermore photons originating form hadronic decays tend to come in pairs and are accompanied by large activity in the hadronic calorimeter. The fake photons can be reduced using the shape of the electromagnetic showers. The complete list of discriminating variables related to the shower shape is given in

Category	Description	loose	tight
Hadronic leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over range $\|\eta\| < 0.8$ or $\|\eta\| > 1.37$)	×	×
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over range $0.8 < \|\eta\| < 1.37$)	×	×
EM Middle layer	Ratio of 3×7 $\eta \times \phi$ to 7×7 cell energies	×	×
	Lateral width of the shower	×	×
	Ratio of 3×3 $\eta \times \phi$ to 3×7 cell energies		×
EM Strip layer	Shower width calculated from three strips around the strip with maximum energy deposit		×
	Total lateral shower width		×
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips		×
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima		×
	Ratio of the energy difference associated with the largest and second largest energy deposits to the sum of these energies		×

Table 3.1: Discriminating variables[44] used as input to the photon identification algorithm. The right columns indicate their presence in the *loose* and *tight* definition.

table 3.1. The photon ID algorithm comes in two flavours. The *loose* selection only uses the hadronic calorimeter and the second LAr layer. The *tight* selection also uses the first LAr layer and applies tighter cuts.

Prompt photons are also isolated with little activity around them. The isolation energy can be defined as the sum of transverse energy in close-by clusters not including the photon cluster (E_T^{iso}) or the scalar sum of the transverse momenta of close-by tracks not including converted electrons (p_T^{iso}). Corrections for pile-up or any underlying event activity must be applied. The recommendation for Run 2 is $E_T^{iso} < 0.065 \cdot E_T$ and $p_T^{iso} < 0.05 \cdot p_T$. The close-by definition uses a cone of $\Delta R < 0.2$.

The efficiency of the photon identification is measured using three data-driven methods. They include using the radiative Z boson decays ($Z \to l^+l^-\gamma$) to select a purified photon sample, extrapolation of shower shapes left by electrons in $Z \to e^+e^-$ and the matrix method. The matrix method studies the effect of the isolation cut on the photon identifica-

tion. The ABCD method described in section 3.4 is based on it. The photons produced by the simulation are smeared such that the identification efficiency matches that seen in data. The measured efficiency for photons, including identification and isolation cuts, is around 80%.

3.2 Jets

Jets play a crucial role in the ATLAS experiment. Many of processes in the proton-proton collisions contain quarks and gluons in the final state. As they try to fly-off in different direction,, colour confinement causes them to hadronize into a multitude of colliminated mesons, baryons and other particles. The energy left in the calorimeters by the final stable particles is clustered into blobs of energy called jets. The jets serve as a proxy for the original partons produced in the collision. They have roughly the same energy and direction as the original parton.

There are several properties that good clustering algorithms should satisfy. First of all, the reconstructed jets should not be affected by the parton splitting into two particles or soft radiation. This is called collinear and infrared safety correspondingly. Also jets should have a predictable shape and area. This is advantageous for experiments, where corrections (ie: pile-up removal) depend on the area of the jet. The anti-k$_t$ algorithm produces conical jets of radius R that satisfy all of these properties.

The anti-k$_t$ algorithm works by calculating the distance between any two clusters i and j as

$$d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{3.1}$$

and the distance between a cluster i and the beam as

$$d_{iB} = k_{t,i}^{2p}. \tag{3.2}$$

The momentum of cluster i is defined by the transverse momentum, rapidity and azimuthal angle $(k_{t,i}, y_i, \phi_i)$. The angular distance between two clusters is given by $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. The calculated distance parameters are sorted. If the smallest parameter comes from d_{ij}, then the two clusters are merged. But if it is d_{iB}, then cluster i is defined as a jet and removed from the list. This is repeated until the list of clusters is empty.

The algorithm is parametrized by p and R. If $p = -1$, then the algorithm is called anti-k$_t$. Alternative versions called k_t ($p = 1$)[65] and Cambridge/Aachen ($p = 0$)[94] also exist. The parameter R is the radius parameter. For anti-k$_t$ jets it corresponds to radius of the circular boundary within which clusters are considered for merging.

The ATLAS experiment uses the the anti-k$_t$ algorithm[7] with $R = 0.4$ on topological clusters[24] (topoclusters) calibrated to the electromagnetic scale to reconstruct jets. A topocluster is a topologically connected set of calorimeter cells with significant energy deposits. Splitting is performed such that each topocluster only contains one energy maximum. The anti-k$_t$ algorithm can also be run on stable MC particles to form truth jets or on tracks to form track jets.

The reconstructed jets need to be calibrated to account for detector response and the effects of pile-up need to be subtracted. A generic calibration procedure was developed by the ATLAS collaboration[34].

The first step is to fix the jet direction to point towards the reconstructed hard scatter vertex.

Next the contribution from the underlying pile-up events is estimated by using the average energy density (ρ) and the area of the jet. The area of a jet is defined by the ghost association of infinitesimally soft particles. The anti-k$_t$ algorithm is rerun with a multitude of fake low energy depositions, called ghosts, added uniformly over the detector. Being very low energy, the particles do not change the jets formed. The fraction of the ghost particles associated

with a jet is taken to indicate the area. To calculate the event energy density, the k_t algorithm is run over the central ($|\eta| < 2.0$) region of the detector. The median energy density (p_T/A) of the k_t jets is used is define ρ as an event-level variable. To account for residual effects of pile-up, further corrections parametrized in the number of reconstructed vertices (N_{PV}) and the average interactions per bunch ($< \mu >$) are applied. The complete pile-up correction can be summarized by

$$p_T^{\text{corr}} = p_T^{\text{EM}} - \rho \times A - \alpha \times (N_{PV} - 1) - \beta \times < \mu > . \tag{3.3}$$

The jet energy scale (JES) correction is applied by studying dijet events simulated using Pythia8. Each reconstructed jet is matched to a truth jet using $\Delta R < 0.3$ criterion. The average response is parametrized as a function of η and follows the structure of the calorimeter. Only jets that contain no other jets of $p_T > 5$ within $\Delta R < 1.0$ are used to derive this calibration. Such jets are called isolated and are used to remove ambiguities from matching reconstructed jets to truth jets.

The next step is to apply JES corrections based on non-kinematic features of the jet. This step is called the Global Sequential Correction (GSC)[35] and accounts for differences between quark-initiated and gluon-initiated jets. The extra jet properties used in the GSC correction, in order, are

1. the fraction of energy deposited in the first layer of the Tile calorimeter,
2. the fraction of energy deposited in the third layer of the electromagnetic calorimeter,
3. the associated number of tracks,
4. the p_T-weighted width of the tracks,
5. the number of associated muon segments.

The last step is the insitu correction to account for differences between the data and simulation. This can be anything ranging from misestimation of the pileup to mismodelling of the detector. It is derived from events where the jet balances against a well-calibrated object in the transverse plane. The reference object is either be a photon or a leptonically decaying Z boson. High-p_T jets can also be balanced against multiple lower energy jets. The double-ratio,

$$f = \frac{p_{T,\mathrm{MC}}^{\mathrm{jet}} / p_{T,\mathrm{MC}}^{\mathrm{ref}}}{p_{T,\mathrm{data}}^{\mathrm{jet}} / p_{T,\mathrm{data}}^{\mathrm{ref}}}, \tag{3.4}$$

is used to bring the energy scale of the jets reconstructed in data to those in simulation.

The uncertainty on the calibration factors is around 6% for jets with $p_T < 100$ GeV, above which it drops to a constant 1%. It is dominated by the uncertainty on flavour composition at low p_T and turns over to the absolute insitu JES for $\mathcal{O}(100\text{ GeV})$ jets. Needles to say, the JES uncertainty plays a negligible role in all analyses considered.

3.3 Bottom Quark Tagging

The identification of jets initiated by b-quarks (b-tagging) plays an important part at hadron colliders. It is useful for reducing the dominant QCD background that most analyses face. This is especially true if a particle decays preferentially to b-quarks, like the Higgs boson with BR($H \to b\bar{b}$)=60%. The decay into b-quarks means the creation of b-hadrons. The unique lifetime of b-hadrons allows them to travel a few μm's before decaying. The tracker system in a modern detector is designed to have a fine enough resolution to detect its decay as a secondary vertex.

The ability to identify the secondary vertex is vital for b-tagging algorithms to discriminate against other jets. The other jets consist of charm-initiated jets (c-jets), jets initiated by gluons and quarks lighter than the charm (light-jets) and taus. The charm quark has a

mass close to the bottom quark, giving it a similar behaviour. This makes c-jets a unique background that is hard to discriminate against.

The ATLAS experiment employs the output of three separate b-tagging algorithms called IP3D, SV and JetFitter[33]. All of them try to use the tracks associated with a jet to form continuous discriminating variables related to the secondary vertex. The track association is done by angular matching.

The IP3D algorithm is based on the signed impact parameter significance. The impact parameter is defined as the closest approach of a track to the primary vertex. The sign is assigned with respect to the direction of the jet under consideration. Tracks associated with a b-jet are expected to have large impact parameters due to the presence of the b-hadron. IP3D uses probability distributions to calculate the likelihoods of the tracks under b-jet and light-jet hypotheses. The likelihood ratio forms the final discriminant. The IP3D algorithm uses both the transverse and longitudinal impact parameters. An alternative algorithm that only uses the transverse impact parameter is called IP2D.

The Secondary Vertex Finding Algorithm (SV) tries to explicitly reconstructs the secondary vertex and use its properties to identify b-jets. Tracks are first preselected to remove the presence of other long lived particles (ie: K_S or Λ) by testing all possible two-track combinations. The remaining tracks, further filtered to remove outliers, are then used to form a single vertex. The probability to reconstruct the secondary vertex of a b-jet depends on p_T and is between 70% to 80%. The properties of the secondary vertex that can be used for b-jet identification include the invariant mass, the number of tracks, the number of two-track vertices, transverse and 3D decay lengths, and the energy fractions of the SV tracks to all of the jet's tracks.

The JetFitter algorithm tries to reconstruct the full decay chain (PV $\rightarrow b \rightarrow c$) of a b-quark. It uses a Kalman filter to define the flight path that the three decay vertices lie

on. This method is accurate at identifying the secondary vertex even when it has only one track associated to it. Variables similar to the output of the SV algorithm are calculated by JetFitter.

The three algorithms are combined into the final product, called the MV2 algorithm, using a Boosted Decision Tree. The BDT is trained using a large simulated $t\bar{t}$ sample with jets labelled as signal or background using truth-tagging. The truth-tagging algorithm identifies b-jets by looking for a b-hadron inside the jet cone. If no b-hadrons are found, then the procedure is repeated with c-hadrons. Jets not truth-tagged as b-jets or c-jets fall into the light-jet category. The BDT can be trained with varying proportions of charm jets. The default training used by ATLAS is called MV2c20, where the background contains 20% c-jets. At 70% b-jet acceptance, it rejects a factor of 440 light jets and a factor of 4.5 c-jets. Different operating points are supported by the ATLAS software. They are calibrated to account for any differences between simulation and data[32].

The b-tagging performance has significantly improved from Run 1 to Run 2. The light-jet rejection has improved by a factor of ≈ 4 for the same b-jet efficiency. The gains are a result of the new pixel layer (IBL), improved tracking in high-density environments[27] present in high p_T jets and reworking of the algorithms themselves. The Run 1 algorithm[29], called MV1, fed the inputs of the three sub-algorithms into a neutral network to form the discriminant. The separation of the algorithm into secondary vertex finding and multivariate discrimination simplified the process and improved the performance.

3.4 Quark/Gluon Tagging

Quarks and gluons, albeit being very different types of particles, look almost the same in the ATLAS detector. They are reconstrcuted as the jets described in section 3.2. The ability to separate quark-initiated (quark-like) jets from gluon-initiated (gluon-like) jets would be

beneficitial to many analyses. Many signals produce quarks at tree-level, while the QCD background is dominated by gluon-like jets. For example, the VBF-jets in the VBF $H \to b\bar{b}$ analysis are quarks. This section summarizes a study on developing a generic quark/gluon tagging algorithm[30][40]. The study explores several jet moments, properties defined from the jet constituents, to determine their power in differentiating quark-like and gluon-like jets. In addition to gauging the performance, the amount of mismodelling of the jet moments by the Monte Carlo generators is evaluated.

An important process in the parton shower of a hadron is the emission of a gluon[92], with a probability depeneding on the colour factor. Quarks carry a colour factor of $C_F = 4/3$, while gluons carry $C_A = 3$. Thus gluons will create more particles in the final state. Dividing a fixed jet energy among a higher particle multiplicity means individual particles in gluon-initiated jets carry less energy and hence are less collimated. This results in "wider" jets. The following jet moments are used to exploit the colour charge[68][57][86][78]:

- Number of tracks in jet:

$$n_{\text{trk}} = \sum_{\text{trk}\in\text{jet}} \tag{3.5}$$

- p_T weighted width of the jet from tracks:

$$w_{\text{trk}} = \frac{\sum_{\text{trk}\in\text{jet}} p_{T,\text{trk}} \Delta R_{\text{trk,jet}}}{\sum_{\text{trk}\in\text{jet}} p_{T,\text{trk}}} \tag{3.6}$$

- E_T weighted width of the jet:

$$w_{\text{calo}} = \frac{\sum_{\text{const}\in\text{jet}} p_{T,\text{const}} \Delta R_{\text{const,jet}}}{\sum_{\text{const}\in\text{jet}} p_{T,\text{const}}} \tag{3.7}$$

- Fraction of energy carried by the largest energy constituent:

$$f^{\text{largest}} = \frac{E_{\text{largest const}}}{E_{\text{jet}}} \tag{3.8}$$

- Two point energy correlation function:

$$C_\beta = \frac{\sum_{i,j\in\text{jet}} E_{T,i}E_{T,j}\left(\Delta R_{i,j}\right)^\beta}{\left(\sum_{i\in\text{jet}} E_{T,i}\right)^2} \tag{3.9}$$

Another difference between quarks and gluons is electric charge, with the former having fractional charge ($\pm\frac{2}{3}$ for up-like and $\pm\frac{1}{3}$ for down-like quarks) and the latter having no charge at all. While the charge of the particles associated to a jet will not equal the intial hadron's charge, it should be correlated with it. The jet charge[75] is defined as

$$Q^\kappa = \frac{1}{(p_T)^\kappa}\sum_{\text{trk}\in\text{jet}} q_i \times \left(p_{T,trk}\right)^\kappa \tag{3.10}$$

A dedicated study and unfolding of the jet charge was done in [41].

A truth flavour tagging algorithm is used to study the discriminating variables in Monte Carlo. It labels a jet by looking for the highest energy parton inside the $\Delta R < 0.4$ jet cone as either quark-like or gluon-like. To remove dependance on heavy flavour, for which dedicated tagging algorithms already exist, quark-like jets are further separated into b-jet, c-jet and light-jet groups. The truth-tagging algorithm is applied to dijet, γ+jet and Z+jet samples to generate truth-tagged distributions, referred to as truth templates, in the different flavour-tagged categories. The three samples are used to understand the sample dependance of the discriminating variables. To understand the accuracy of different generators, both Pythia8 and Herwig++ are used to generate the samples and compared against templates extracted

using a data-driven method.

Since it is difficult to obtain a pure samples of quark or gluon jets, a procedure previously developed by the ATLAS collaboration is used to extract the distributions from a collection of inclusive samples. The procedure assumes that an inclusive distribution is a linear combination of pure templates with little sample dependance. The binned distributions of the discriminating variables in the three samples can be described by

$$\begin{pmatrix} p_{\text{dijet},i} \\ p_{\gamma+\text{jet},i} \\ p_{\text{Z+jet},i} \end{pmatrix} = \begin{pmatrix} f_{\text{dijet},q} & f_{\text{dijet},g} \\ f_{\gamma+\text{jet},q} & f_{\gamma+\text{jet},g} \\ f_{\text{Z+jet},q} & f_{\text{Z+jet},g} \end{pmatrix} \begin{pmatrix} p_{q,i} \\ p_{g,i} \end{pmatrix} + \begin{pmatrix} f_{\text{dijet},hf}\, p_{\text{dijet},hf} \\ f_{\gamma+\text{jet},hf}\, p_{\gamma+\text{jet},hf} \\ f_{\text{Z+jet},hf}\, p_{\text{Z+jet},hf} \end{pmatrix} + \begin{pmatrix} 0 \\ f_{\text{impure}}\, p_{\text{impure},i} \\ 0 \end{pmatrix}, \quad (3.11)$$

where $p_{\text{sample},i}$ is the content in bin i for *sample* (inclusive sample or template) and $f_{\text{sample},x}$ is the fraction of jet type x in *sample*. The histograms are normalized to unit area. The left side of the equation corresponds to the inclusive distributions obtained from data. On the right hand side, the first two terms correspond to the templates of a given jet type weighted by their amounts. The fractions for quarks, gluons and the heavy flavour contributions can be obtained from Monte Carlo. The last term corresponds to contaminations from other samples and is only relevant for γ+jet selection where the photon can be faked by a jet. Its value is estimated using the ABCD method. The details on how the individual values obtained are provided later. Since a lot of the values depend on the kinematics of a jet, the procedure is repeated in p_T and $|\eta|$ categories listed in table 3.2. The two unknown quantities of interest are the bin contents of the templates, $p_{q,i}$ and $p_{g,i}$, and can be obtained by solving

Lower p_T bin boundary (GeV)							
25	40	90	120	150	190	210	400
dijet, Z+jet	dijet, Z+jet, γ+jet	dijet, γ+jet					

Upper $\|\eta\|$ bin boundary			
0.8	1.2	2.1	2.5

Table 3.2: The p_T and $|\eta|$ bin boundaries used to study the properties of the quark- and gluon-like jets. The p_T bin table also lists the samples used as input into the extraction procedure.

the linear system of equations. Since the system in equation 3.11 is over-constrained and not exact (due to detector effects and imperfect assumptions), it cannot be solved exactly. This is bypassed by finding the least-squares solution or removing one of the samples. The latter is used for high p_T jets, where there are not many Z+jet events. The p_T bins and list of samples used in each are also shown in table 3.2.

To estimate systematic effects, the extraction procedure is repeated several times with each input quantity being varied within its uncertainty. The gluon-like jet rejection for a fixed quark-like acceptance is then evaluated for each varied template and compared against the rejection from the nominal template. This allows an envelope to be built around the nominal tagger to serve as the systematic uncertainty for the performance of the quark/gluon tagger. The uncertainties stemming from each of the variations are treated as uncorrelated and added in quadrature. The paragraphs in this section describing the source of each input also describe their definition of uncertainty.

The only exception is the sample dependence systematic. The sample dependence systematic accounts for any dependence the jet moments have on the underlaying hard-scatter process. To evaluate its effect, a modified version of the template variation approach is used. First the nominal templates are extracted from the inclusive Monte Carlo templates. The

truth-tagged dijet, γ+jet and Z+jet templates are used as the variations. Even though the Monte Carlo is not guaranteed to give a good description of the discriminating variables, the relative difference between the samples should at least be well modelled.

The inclusive distributions consist of dijet, γ+jet and Z+jet samples. Separate versions are created using simulated events and events recorded by the ATLAS detector with $\sqrt{s} =$ 8 TeV collisions. The events are required to pass the standard event cleaning cuts. The sample-specific selection is listed below. In all cases, only the properties of the leading jet in each event are studied and the events are reweighted to match the jet p_T spectrum in the dijet dataset. The latter reduces the dependence of the variables on the jet p_T. Furthermore the leading jet is required to be at least $\Delta R > 0.8$ (twice the jet radius) away from the closest jet with $p_T > 15$ GeV and originating from any vertex in the event. The jet isolation requirement reduces the effect of near-by jets seen in figure 3.1.

The dijet sample is sourced using the single jet triggers. For each template bin, the least prescaled trigger that is fully efficient in leading jet p_T for the corresponding p_T bin is chosen. Then following event selection is applied:

- subleading jet $p_T > 20$ GeV
- $\Delta\phi_{\text{leading, subleading jet}} > 2.5$
- $p_{T,3^{\text{rd}}\text{ jet}} < \max\left\{12\text{ GeV}, 0.25 p_{T,\text{avg}}\right\}$, with $p_{T,\text{avg}}$ being the average p_T of the two leading jets
- $\eta_{1^{\text{st}}\text{ jet}} \cdot \eta_{2^{\text{nd}}\text{ jet}} > 0$ and $\eta_{1^{\text{st}}\text{ jet}} < \eta_{2^{\text{nd}}\text{ jet}}$

The first three cuts ensure that the sample is dominated by 2-to-2 parton scattering, since the simulated dijet events are only generated at leading order by Pythia8 and Herwig++. The last cut is to have more events with the leading jet being gluon-like. This is needed to make the linear system in equation 3.11 non-degenerate. Having the two jets be in the same

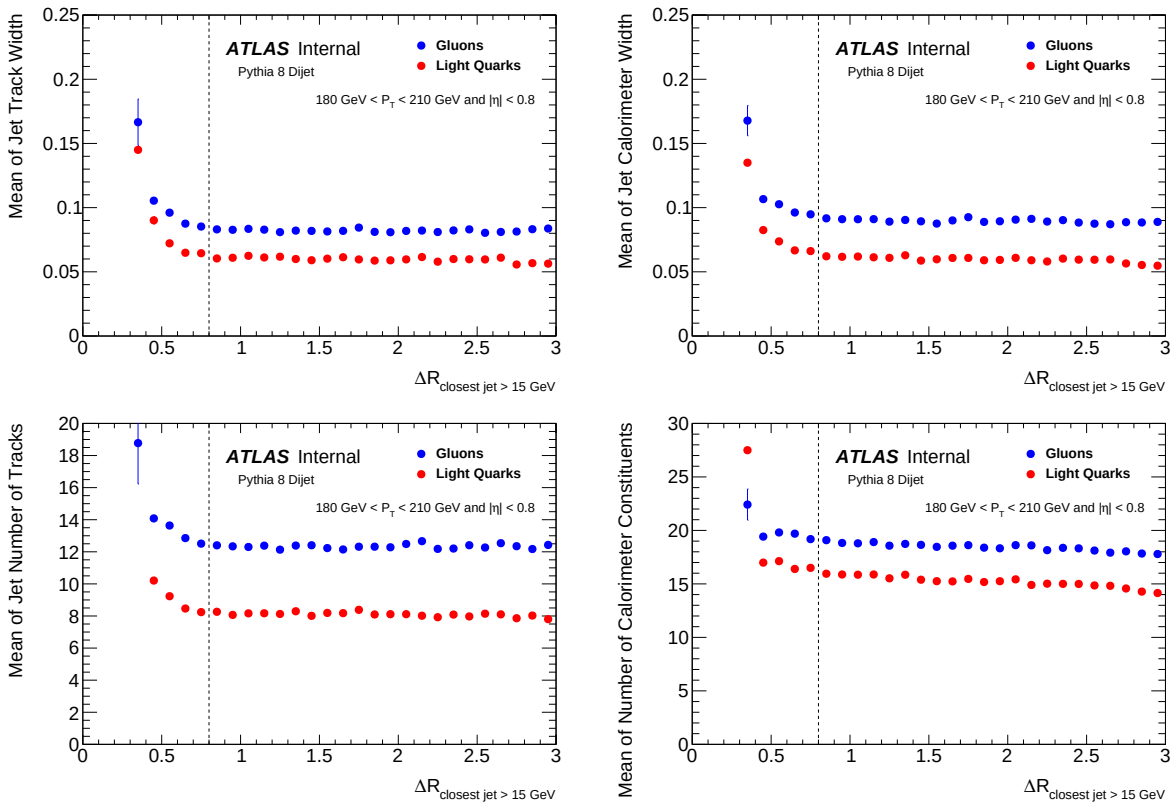


Figure 3.1: Mean of the w_{trk} (top-left), w_{calo} (top-right), n_{trk} (bottom-left) and number of constituents (bottom-right) as a function of the distance to the nearest jet above 15 GeV in truth-tagged Pythia 8 dijet events. The template bin for 180 GeV $< p_T <$ 210 GeV and $|\eta| < 0.8$ is chosen. The $\Delta R = 0.8$ selection used to separate isolated jets from non-isolated jets is shown as a dashed line.

hemisphere along the z-direction and forward causes collisions with a high Q to be selected. Such events tend to happen between a valance quark and a sea gluon. By colour charge conservation, having a single gluon in the initial state results results in a gluon in the final state.

The γ+jet selection provides a sample that is rich in quark-like jets. It uses the single photon triggers, also chosen such that each template bin uses the least prescaled fully-efficient trigger in the corresponding jet p_T bin. The rest of the selection is as follows:

- leading photon passes tight ID
- photon isolation energy less than 3 GeV in a $\Delta R < 0.4$ cone
- $\Delta\phi_{\text{leading photon,jet}} > 2.9$
- $p_{T,\text{subleading jet}} < \max\left\{40 \text{ GeV}, 0.3 p_{T,\gamma}\right\}$

The jets used in the selection are required to not overlap with the leading photon within $\Delta R < 0.2$. The first two cuts are used to reject photons faked by jets. However even after their application, the data still contains a large background from dijet events with a jet faking a photon. It is corrected for by the "impurity" term in equation 3.11. The last two cuts applied make the data sample compatible with the leading-order γ+jet generators (Pythia8 and Herwig++).

A dedicated $Z \to e\bar{e}$ trigger is used to select the Z+jet sample. The sample is further purified via the following cuts:

- two electrons passing the medium++ criteria
- 66 GeV $< m_{ee} <$ 116 GeV
- leading jet $p_T > 25$ GeV
- $\Delta\phi_{\text{leading jet,Z}} > 2.0$

- $p_{T,\text{subleading jet}} < \max\left\{40 \text{ GeV}, 0.3 p_{T,Z}\right\}$

To select the jets, the same overlap criteria as with the γ+jet sample is used after identifying good electrons. Two reference Monte Carlo datasets are generated then showered, one using Alpgen/Herwig++ and second with Powheg/Pythia8.

The fractions of quark and gluon jets in each sample are taken from the truth-tagged jets in Monte Carlo. The Pythia8 generator is used to provide the nominal values, while an alternate set for systematics is provided by Herwig++. The values from both generators are shown in figure 3.2. The dijet sample is evenly split between quarks and gluons, with gluons dominating at lower energies. A similar trend is true at high $|\eta|$, but with the fraction of quarks being slightly higher. This fraction would have been even higher if the "high-Q" cut is not present in the dijet selection. On the other hand, the γ+jet and Z+jet events are dominated by quark-like jets. The quark fraction in both samples is above 80% for all η. Both photons and Z's couple to quarks and not gluons, resulting in a larger number of available Feynman diagrams containing both particles.

Having one sample be dominated by gluons (dijet) and another by quarks (γ+jet, Z+jet) is a requirement for the procedure to work. It causes the fraction matrix in equation 3.11 to not have a determinant close to zero. This means that the solution is stable under perturbations (ie: statistical, detector resolution) of the input distributions, reducing the uncertainty in the extracted templates.

The contribution from heavy flavour to each sample is also taken from simulated events showered by Pythia8. Since the contributions are small, mismodelling should not have big effect on the extracted templates. This is quantified by calculating the systematical uncertainty using the template variations. Template variations are generated by varying the c/b-jet fractions by 20% in the dijet sample and 50% in the γ+jet sample. For the uncer-

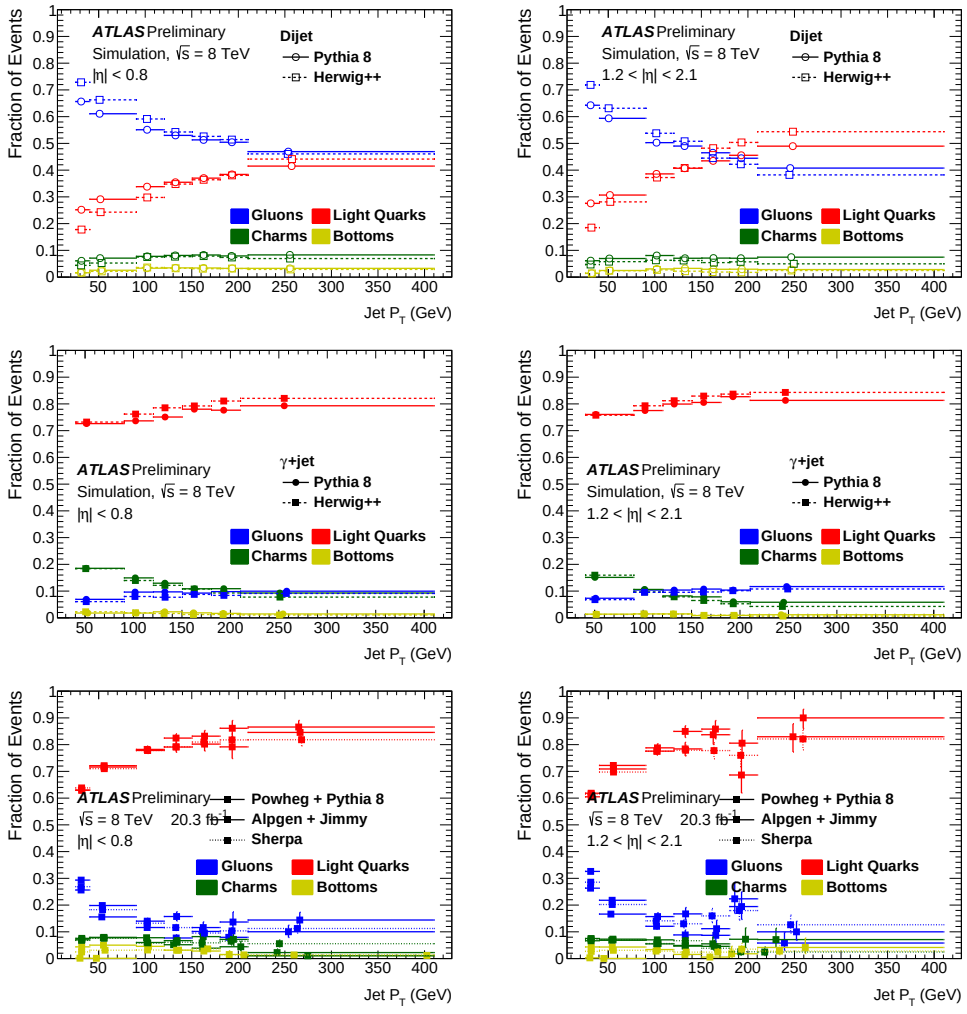


Figure 3.2: The fractions of quark-like (light and heavy flavour) and gluon-like jets in the dijet (top), γ+jet (middle) and Z+jet (bottom) samples used in the extraction procedure. Values are shown for both central (left) and most forward (right) template bins.

tainty on the c-jet shape, the dominant decay modes of the c-mesons are varied by 10%. The largest deviation from the nominal c-shape is taken as the template variation. For the b-jet shape, the alternate distribution is obtained from the HERWIG++ simulation.

The last piece is the impurity term. The dijet and Z+jet samples have a high purity, but the γ+jet sample does not. A photon can be faked by a jet, resulting in some dijet events sneaking into the γ+jet sample. Since the effect of the fake γ+jet events on the fanal sample is hard to predict, a data-driven "ABCD" method[18] is used to derive their contributions. The ABCD method divides the events into four regions (A, B, C and D) based on two uncorrelated variables; isolation energy and photon ID. Region C corresponds to the signal region, where the photon passes tight ID and has isolation energy < 3 GeV. Region A has the same isolation energy cut, but a loosened "non-tight" photon identification cut. Regions B and D have the same photon selection, but require isolation energy greater than 5 GeV. The definition of the four regions is illustrated in figure 3.3. The simplest version of the ABCD method estimates the contamination in region C (signal region) by using region D as a reference and applying a transfer factor calculated between regions B and A as

$$N^C_{\text{bkg}} = N^D_{\text{bkg}} \frac{N^A_{\text{bkg}}}{N^D_{\text{bkg}}}. \tag{3.12}$$

This assumes that all of the signal is contained in region C, which is not true as seen in figure 3.3. The correction for the signal leakage uses γ+jet Monte Carlo (PYTHIA8) to estimate the effect by defining the amount of signal in region X by $c_X = N^X_{\text{sig}}/N^C_{\text{sig}}$ and subtracting it from the number of observed events (ie: $N^X_{\text{bkg}} = N_X - c_X N^C_{\text{sig}}$). This requires equation 3.12 to be solved using the quadratic formula, with the smaller solution used to define the purity as $f_{\text{impure}} = 1 - N^C_{sig}/N_C$. The purity is calculated independently for all templates bins as summarized in figure 3.4. Events with jet $p_T > 150$ GeV tend to be almost 100% pure, while lower energy events are at least 60% pure. The shape of the contaminiation, p_{impure},

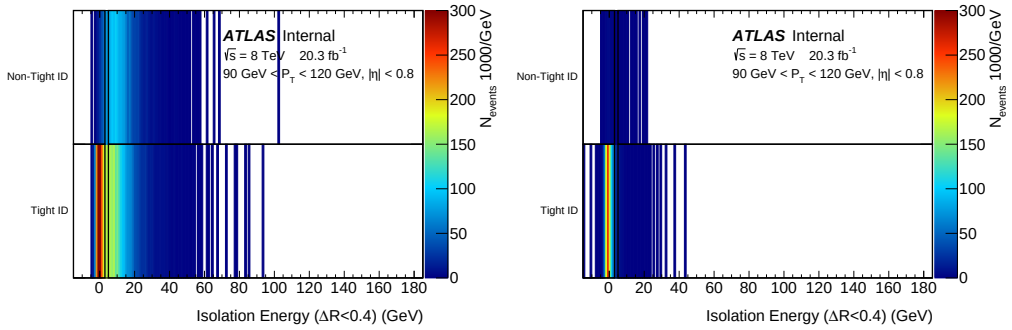


Figure 3.3: Example of the ABCD method used to estimate the purity of the γ+jet sample used to extract the quark/gluon tagger templates. The template bin with jet 90 GeV $< p_T <$ 120 GeV and $|\eta| < 0.8$ is shown. The black lines represend the definition of the ABCD regions, reading from left to right. The event count in data (left) and the signal leakage estimated by PYTHIA8 (right) are shown.

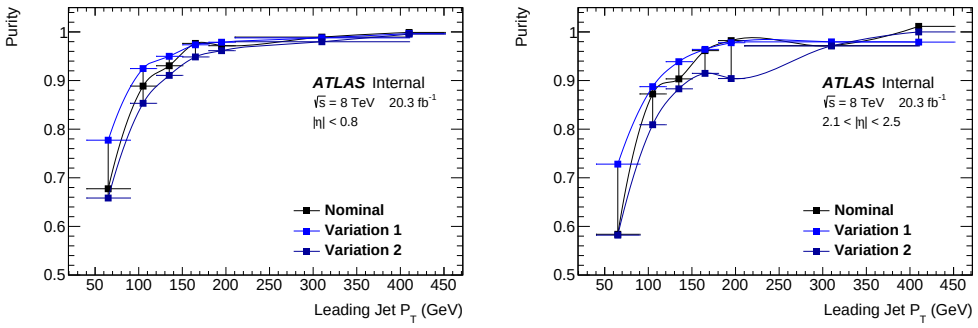


Figure 3.4: The purity of the γ+jet sample used to extract the quark/gluon tagger templates as a function of the jet p_T. The template bin with jet 90 GeV $< p_T <$ 120 GeV and $|\eta| < 0.8$ is shown. The nominal value is shown in black. The other curves correspond to two variations of the non-tight photon identifcation to estimate the uncertainty.

is taken from region A.

The uncertainty in the purity calculation is estimated by varying the definition of the "non-tight" photon ID. Both the shape and normalization are varied at the same time in the extraction procedure.

The result of running the template extraction procedure on data and Monte Carlo are shown in figure 3.5 for calorimeter-based variables and figure 3.6 for track-based variables. The truth-tagged templates are not used in the comparisons to cancel any bias from the

extraction procedure. Both central and forward distributions are shown for a single p_T template bin for illustration. The p_T dependence is demonstrated in figure 3.7 by plotting the mean for example variables. There are a few important conclusions that can be made from the results.

The first observation is simple. For jet p_T less than 100 GeV, there is a strong dependance on the momentum. This motivates the reweighting the leading jet p_T spectra to a common distribution. The p_T dependence is one of the largest contributions to the sample dependance of the tagger. This also demonstrates a problem with building a generic quark/gluon tagger. The efficiency, and any related scale factors, will be heavily dependant on the selection under study.

Second is the accuracy of Monte Carlo generators in modelling the distributions. For quark-like jets, both generators provide reasonable predictions. The means of the distributions are similar within 10% and the tails match up reasonably. On the other hand, the modelling of gluon-like jets is generator dependant. The PYTHIA8 generator performs poorly, with a 10% sytematic shift in all variables. The HERWIG++ generator does a better job, showing agreement at the level of quark-like jets. The consequence is that any attempt at using the variables in an analysis selection also requires generator-dependant scale factors.

The extracted templates can be used to evaluate the rankings of the discriminating variables based on their usefulness in a generic quark/gluon tagger. Figure 3.8 shows the gluon-like jet rejection for a fixed quark-like jet efficiency of 70%. The efficiency is defined as the fraction of quark-like jets left after cutting on a single variable. The rejection is defined as one minus the corresponding gluon-like efficiency using the same cut value. The performance separates the variables into two categories; those based on colour charge reject 50% of gluon-like jets and those based on electric charge reject 40%.

It should be noted that a random tagger would reject 30% of the gluon-like jets at the

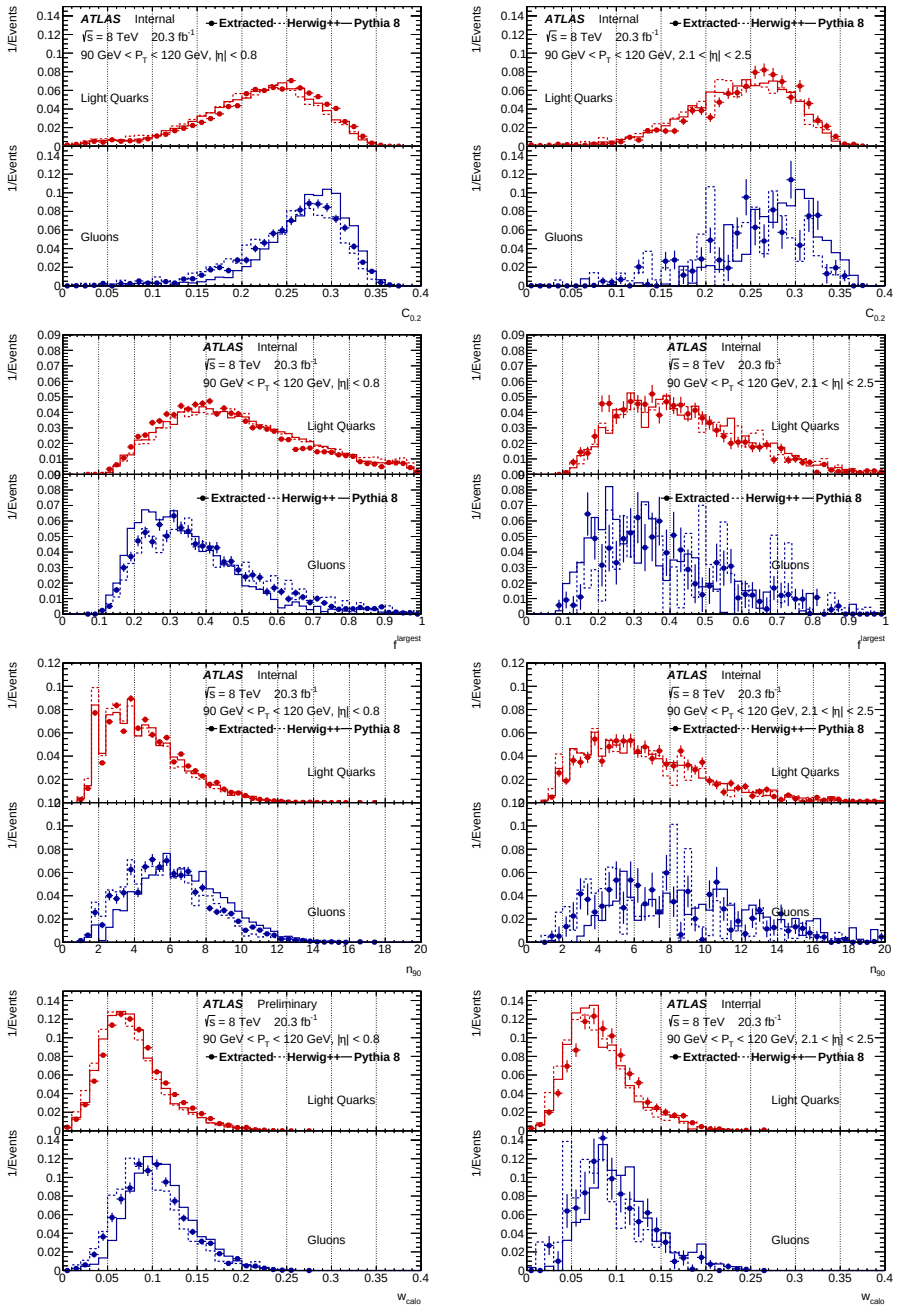


Figure 3.5: Example distributions of calorimeter-based variables for 90 GeV $< p_T <$ 120 GeV with $|\eta| < 0.8$ (left) and $2.1 < |\eta| < 2.5$. The templates extracted from data (points) are compared with templates extracted from PYTHIA8 (solid) and HERWIG++ (dashed) simulated events.

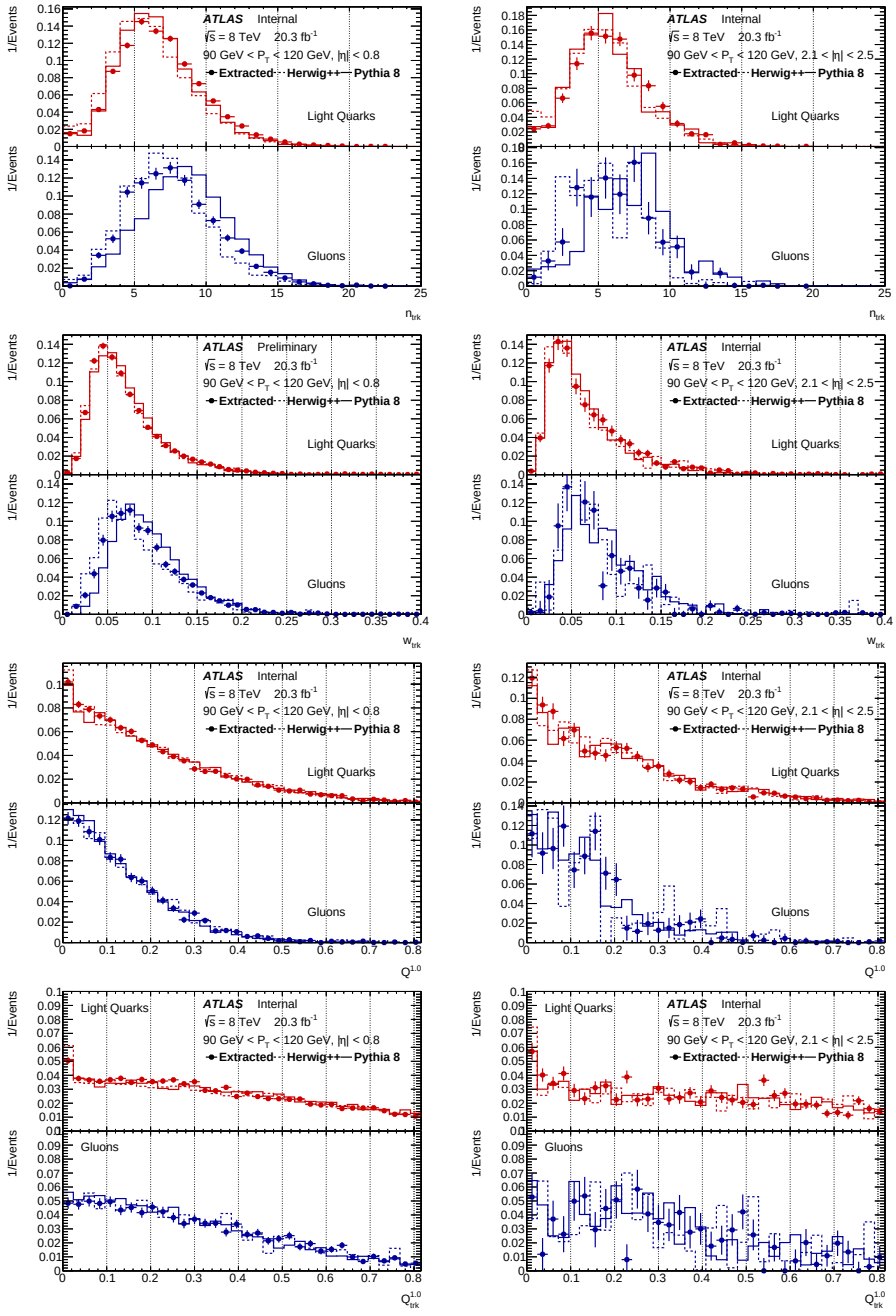


Figure 3.6: Example distributions of track-based variables for 90 GeV $< p_T <$ 120 GeV with $|\eta| < 0.8$ (left) and $2.1 < |\eta| < 2.5$. The templates extracted from data (points) are compared with templates extracted from PYTHIA8 (solid) and HERWIG++ (dashed) simulated events.

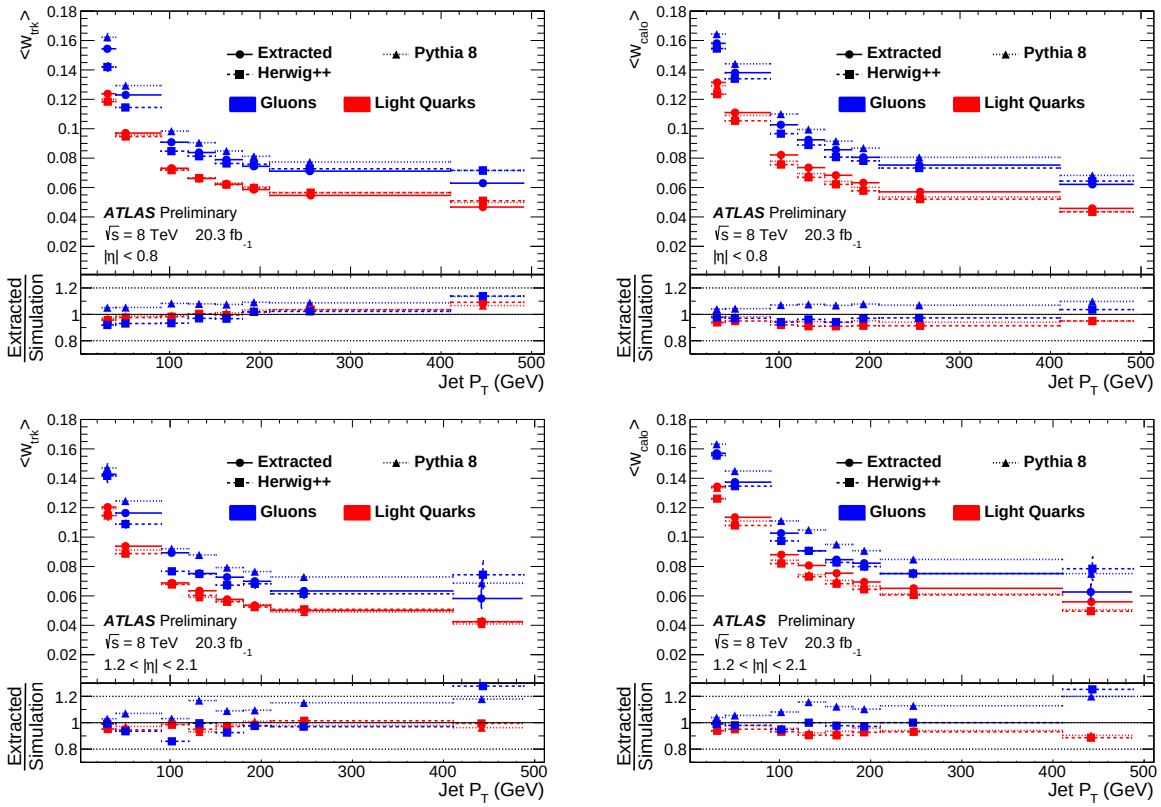


Figure 3.7: Arithmetic mean of discriminating variables as a function of the jet p_T for $|\eta| < 0.8$ (top) and $1.2 < |\eta| < 2.1$ (bottom). The track width (left) and calorimeter width (right) as used as examples.

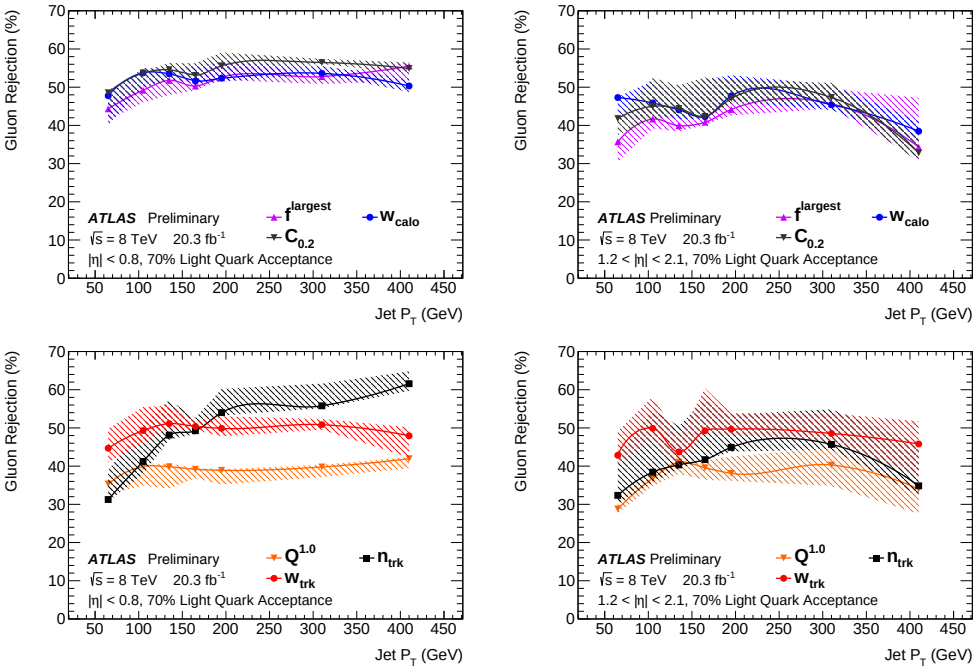


Figure 3.8: The gluon rejection for 70% quark acceptance of the different disciminating variables tested for the quark/gluon tagger. Both calorimeter (top) and track (bottom) based variables are shown for central (left) and forward (right) $|\eta|$ range. The hashed area correspond to the total (statistical+systeamtic) uncertainty.

same working point, so the observed performance is not very substantial. Combined with the large uncertainty, especially at large $|\eta|$, a generic data-driven quark/gluon tagger is not the optimal way to proceed. Instead it might be more fruitful to provide unfolded distributions for the authors of the generators to tune their parton shower algorithm. With a better description in Monte Carlo, analyzers can then use the disciminating variables directly as inputs into their optimization procedure without creating new systematics. A similar approach is taken for the VBF $H \to b\bar{b}$ mutlivariate analysis described in section 4.3. Since the modelling of the quark-like jets in the signal is reasonable and the estimation for background gluon-like jets is data-driven, no correction is applied to the jet moments of the events used to train the Boosted Decision Tree.

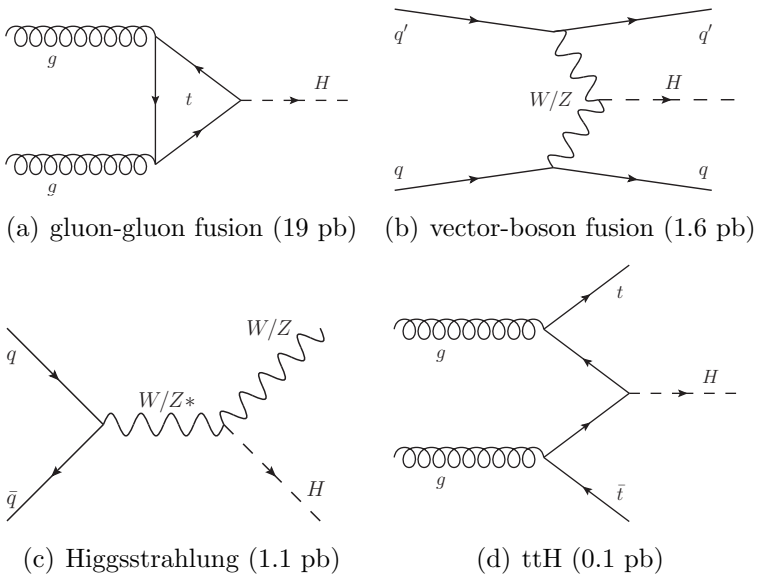
CHAPTER 4

VECTOR-BOSON FUSION HIGGS $\rightarrow$ BB

On July 4, 2012 the ATLAS and CMS collaborations announced the discovery of a new resonance in different channels at mass of 125 GeV[21][59]. Studies of its couplings[25] and spin[22] since have indicated that it behaves like the Higgs boson predicted by the Standard Model. However the most favoured decay mode of a Higgs boson to two b-quarks, with a branching ratio of 60%, still remains elusive. In the most dominant production diagram, where two incoming gluons couple to the Higgs via a top loop (gluon-gluon fusion or ggF), the final state of two b-jets looks very similar to the overwhelming background from the QCD dijet production. There are very few handles that can be used to design an efficient trigger to record the Higgs signal events and discriminate the background from the signal offline. Combined with a low production cross-section, it is very difficult to obtain a statistically significant sample.

The situation can be improved by producing the Higgs boson in associating with extra particles. The presence of the extra particles provides extra handles on the signal that can be used as a source for the trigger or to lower the background by cutting on their kinematics. There are three experimentally favoured mechanisms that can accomplish this. In order of decreasing cross-section, they are vector-boson fusion (VBF), Higgsstrahlung and ttH. The corresponding Feynman diagrams and thei cross-sections are shown in figure 4.1. All three are actively being pursued by the main LHC experiments[54][61][39][60][53][74]. This section reports on search for the Higgs boson decaying to two b-quarks in the VBF production channel at 8 TeV with 21.2 fb^{-1} of data.

The VBF $H \rightarrow b\bar{b}$ signal consists of four jets in the final state. Two are the b-jets from the decay of the Higgs boson. The other two are light quarks jets that radiated the vector bosons used to produce the Higgs. They are often referred to as the "VBF jets". There are



(a) gluon-gluon fusion (19 pb) (b) vector-boson fusion (1.6 pb)

(c) Higgsstrahlung (1.1 pb) (d) ttH (0.1 pb)

Figure 4.1: The Feynman diagrams, along with the corresponding cross-sections at 8 TeV[79], showing the common production methods of the Higgs boson at the LHC.

two main backgrounds with similar signatures: multi-jet QCD and Z+jets.

The VBF jets have two features that can be used to differentiate them from the background. First of all, there is no colour charge connection between the two. As far as they know, the two quarks have just scattered off each other. This means that the rapidity separation between the two, $\Delta\eta_{jj}$, tends to be quite large. The other is that they are quark-initiated, while a lot of the background has gluon-initiated jets. The use of quark/gluon tagging can be used to further suppress the background.

The two Higgs jets can be identified as b-jets through the use of b-tagging algorithms. Furthermore their invariant mass should reconstruct the Higgs mass at 125 GeV, while for the background a combination of two random b-jets leads to a smoothly falling spectrum. This allows a fairly straight-forward search for a resonant peak over a smooth background. One can also define a control region in the m_{bb} spectrum outside a window centred at 125 GeV and use it for a data-driven selection optimization and background estimation. This removes the need for simulated QCD samples, which is both computationally intensive to produce in the required phase-space and not very trustworthy physically. The Z+jets background also

appears as a resonant peak, but is centred at a mass of 90 GeV. The overlap with the Higgs signal region is not large.

It should be noted that ggF $H \to b\bar{b}$ can have the same final state, if one considers extra jets induced by next^n-leading order effects. It is complicated to differentiate such events from the VBF produced Higgs and doing so would detract from the main goal of observing the Higgs coupling to b-quarks. Thus it is best to consider it as part of the signal.

Two analyses, with some commonalities, were used to search for VBF $H \to b\bar{b}$ in 2012 ATLAS data. The main analysis used a multivariate technique employing a Boosted Decision Tree (BDT) and four different signal regions to increase the sensitivity. Kinematics of the VBF jets and jet shapes useful for quark/gluon tagging were used as inputs to the BDT. As a cross-check, a cut-based analysis is performed. It only had a fixed cut on number of jets, $\Delta\eta_{jj}$ and Δm_{jj}.

The rest of the chapter is organized into four sections. Section 4.1 describes the common samples, trigger strategy, pre-selection and definition of systematics for the two analyses. Section 4.2 describes the cut-based selection and results. Sub-section 4.3 touches on the multivariate analysis.

4.1 Pre-Selection

The multivariate and cut-based analyses share the same Monte Carlo sample, trigger strategy, common pre-selection and list of systematics. Using the same triggers is unavoidable, since the trigger menu is designed around simple selections using jet kinematics, multiplicity and b-tagging. Sharing the pre-selection allows to define a baseline for comparison between the two analyses. Comparisons include checking for errors in calibration procedures and improvements in sensitivity through further optimization. The standardization of systematics is another point of the cross-check, as they be comparable between the two analyses.

Simulated samples are generated for the $Z \to b\bar{b}$ background and the $H \to b\bar{b}$ signal. The NLO CT10 parton distribution function (PDF) is used for both samples. Also PYTHIA8 [90] with the AU2 parameter tune[19][17] is used to model the hadronization and parton shower. No samples are generated for the non-resonant background. Obtaining a statistically significant simulated sample in the required phase-space needs the use of complex filters andits usefulness in accurately modelling the QCD background is questionable.

The ggF and VBF components of the $H \to b\bar{b}$ signal are generated using POWHEG. The normalization for the VBF component contains full QCD and EW next-to-leading-order (NLO) corrections. Approximate next-to-next-to-leading-order (NNLO) QCD corrections are also applied. A reweighting factor is calculated by comparing the HAWK [12][13] prediction with and without EW NLO corrections applied. The reweighting is used to correct the shape of the Higgs p_T distribution in the signal sample. For the ggF component, the normalization is taken at NNLO in QCD and includes a soft-gluon resummation up to next-to-next-to-leading logarithm terms (NNLL). Reweighting is also used to correct the Higgs p_T shape. This time the HRES [64][72] program is used for the inclusive dataset and MINLO HJJ [8] for the events with at least two jets.

The $Z \to b\bar{b}$ events are generated in MADGRAPH5 using diagrams containing extra jets from EW and QCD at tree-level. Diagrams containing up to 3 partons at tree-level, including the VBF process, are merged using the CKKW-L procedure in PYTHIA8. The decay of the Z boson into two b-quarks is also done in PYTHIA8.

With the exception of calorimeters, the detector simulation of the samples is done using methods summarized in section 3. For the calorimeter response, a fast technique using parametrized models is applied.

The relevant data is recorded due to an OR of three b-jet triggers listed in table 4.1.

	Trigger	Integrated Luminosity [fb^{-1}]
A	EF_2b35_loose_4j35_a4tchad	20.2
B	EF_b35_medium_j35_a4tchad_vbf_2L1FJ15	4.41
C	EF_b35_medium_j35_a4tchad_vbf_3L1J15_FJ15	4.43

Table 4.1: List of triggers used by the VBF $H \to b\bar{b}$ analysis.

All three triggers require at least one 15 GeV jet at Level-1 and at least one b-tagged 35 GeV jet in the Event Filter. The differences are η range, the number of required extra jets and the b-tagging strategy. The main trigger (EF_2b35_loose_4j35_a4tchad, trigger A) requires four central jets ($|\eta| < 3.2$) above 35 GeV, with at least two of them passing the "loose" b-tagging operating point. This trigger operated unprescaled for the entirety of the 8 TeV run of the ATLAS detector, gathering 20.2 fb^{-1} of data. The other two triggers were added near the end of the data-taking period, gathering only 4.4 fb^{-1} of data each. They are designed around the topology of the VBF signal and focus on the presence of forward jets. At Level-1, they either require two forward ($|\eta| > 3.2$) jets with $p_T > 15$ GeV (EF_b35_medium_j35_a4tchad_vbf_3L1J15_FJ15, trigger B) or three jets with $p_T > 15$ GeV and one of them forward (EF_b35_medium_j35_a4tchad_vbf_3L1J15_FJ15, trigger C). The criteria is further supplemented by requiring at least one jet above 35 GeV passing the "medium" b-tagging operating point in the Event Filter. The three triggers can be considered as orthogonal, since the overlap between them is found to be negligible. Furthermore, the events passing triggers B and C are weighted to the full 20.2 fb^{-1} luminosity.

Figure 4.2 shows the trigger efficiency for the signal corresponding to the three triggers. An efficiency of almost 80% can be reached using a combinations of triggers A and C, with trigger C being more important for lower 4^{th} jet p_T. However due to the low trigger C luminosity, only 40 to 50% of the signal is recorded. Trigger B does not contribute a significant amount to the signal yield.

The two analyses share a common pre-selection. The first step is a standard event clean-

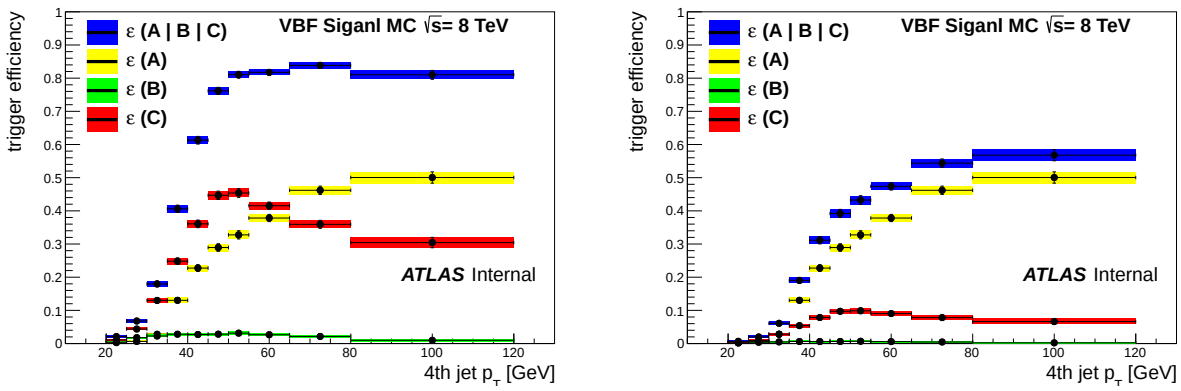


Figure 4.2: The trigger efficiency for the VBF $H \to b\bar{b}$ signal in terms of the 4^{th} jet p_T. Both efficiencies without (left) and with (right) luminosity weight for the VBF triggers are shown.

	VBF Higgs			ggF Higgs		
Cut	N_{sig}	Cut efficiency	Total efficiency	N_{sig}	Cut efficiency	Total efficiency
generated	18535	-	-	45322	-	-
$n^{PV}_{tracks} > 3$	18528	1.00	0.999	45315	1.00	0.9998
LAr noise removal	18528	1.00	0.999	45315	0.04	0.9998
trigger	1031	0.06	0.056	2037	1.00	0.0449
4 jets with $p_T > 50$ GeV	393	0.38	0.021	697	0.34	0.0154
truth matching	343	0.87	0.018	578	0.32	0.0127
η-sorting	267	0.78	0.014	183	0.83	0.0040
online ROIs b-jet matching	215	0.81	0.012	145	0.79	0.0032
centrality of b-jets	214	0.99	0.012	145	1.00	0.0032
offline b-tagging of b-jets	174	0.81	0.009	120	0.83	0.0027
$p_T(b\bar{b}) > 100$ GeV	129	0.74	0.007	95	0.79	0.0021

Table 4.2: The expected signal event yield for different cuts at the pre-selection stage.

	QCD Z Production			EW Z Production		
Cut	N_{sig}	Cut efficiency	Total efficiency	N_{sig}	Cut efficiency	Total efficiency
generated	88003186	-	-	46593	-	-
$n^{PV}_{tracks} > 3$	87999994	0.99996	0.99996	46470	0.997	0.997
LAr noise removal	87999994	1.00000	0.99996	46470	1.000	0.997
trigger	125474	0.00143	0.00143	2775	0.060	0.060
4 jets with $p_T > 50$ GeV	40046	0.31916	0.00046	1134	0.409	0.024
truth matching	28772	0.71847	0.00033	804	0.709	0.017
η-sorting	7145	0.24834	0.00008	285	0.354	0.006
online ROIs b-jet matching	5428	0.75972	0.00006	221	0.774	0.005
centrality of b-jets	5408	0.99632	0.00006	220	0.998	0.005
offline b-tagging of b-jets	4481	0.82843	0.00005	182	0.826	0.004
$p_T(b\bar{b}) > 100$ GeV	3748	0.83661	0.00004	160	0.881	0.003

Table 4.3: The expected Z boson event yield, separated into QCD and electroweak production, for different cuts at the pre-selection stage.

Cut	N_{sig}	Cut efficiency	Total efficiency
any trigger and at least 4 jets	119881920	-	-
good runs list	114367080	0.95	0.954
$n_{tracks}^{PV} > 3$	114367080	1.00	0.954
LAr noise removal	114093210	1.00	0.952
trigger	46949274	0.41	0.392
4 jets with $p_T > 50$ GeV	11877433	0.25	0.099
online ROIs b-jet matching	2857351	0.24	0.024
centrality of b-jets	2816464	0.99	0.023
offline b-tagging of b-jets	1417385	0.50	0.012
$p_T(b\bar{b}) > 100$ GeV	874539	0.62	0.007

Table 4.4: The number of events found after different cuts at the pre-selection stage.

ing to remove problems with noisy components and the requirement of a well-reconstructed primary vertex. For data, additional selection is applied to only select events recorded with a fully operational detector (Good Run List). Next exactly four jets with $p_T > 50$ GeV and $|\eta| < 4.5$ are required. The four jets are then ordered in η such that $\eta_1 < \eta_2 < \eta_3 < \eta_4$. The two outside ($\eta_1$ and η_4) jets are defined as VBF jets and the two middle (η_2 and η_3) jets are defined as Higgs jets. The motivation for such ordering comes from the forward property of the VBF jets. This toplogy is demonstrated by the event displays seen in figure 4.3. The two Higgs jets are required to be within the tracker acceptance ($|\eta| < 2.5$) and be identified as b-jets using the 70% operating point. When applied to a $t\bar{t}$ sample, this operating point has a b-jet (light-jet) mis-tag rate of less than 20% (1%). The HLT b-jets identified by the trigger must match to one of the Higgs b-jets using ΔR matching. A scale factor is applied to account for mismodelling of online and offline b-tagging performance in simulation. Finally the reconstructed Higgs p_T is required to be above 100 GeV. It removes sculpting of the non-resonant background from the jet p_T cuts. Tables 4.2-4.4 show the effect of the selection on the event count for the signal, Z boson background and data. Additional selection aimed at reducing the background is performed. The MVA analysis used several input variables to train a Boosted Decision Tree. The cut-based analysis cut on jet multiplicity, m_{jj} and $\Delta\eta_{jj}$ to maximize signal significance. More details are provided in the corresponding sections.

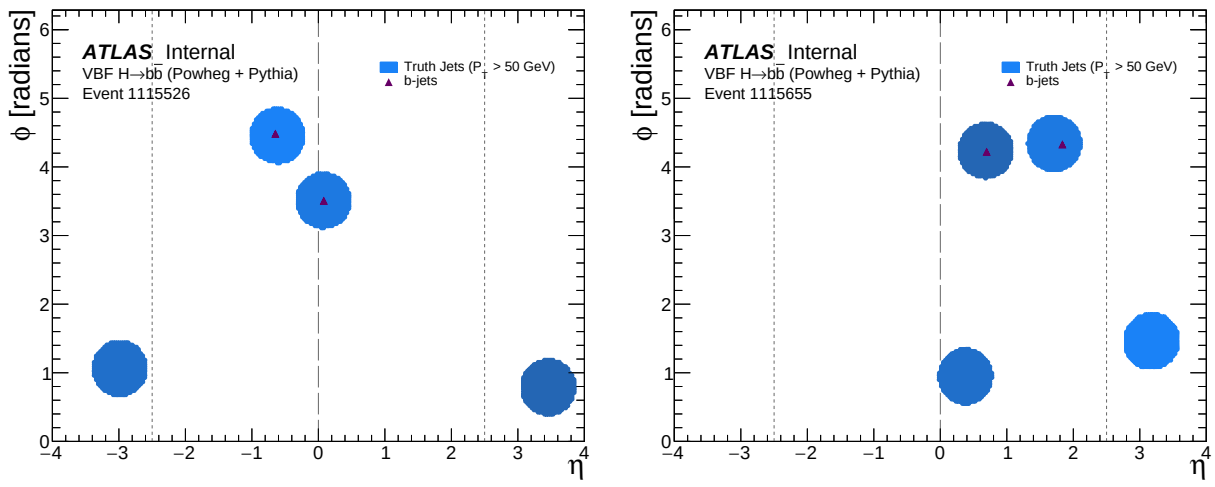


Figure 4.3: Event displays showing the topology of a typical $H \to b\bar{b}$ event. The b-jets stemming from the Higgs decays are identified at truth-level using a triangle.

To improve the m_{bb} resolution, additoinal corrections are applied to the energy of the Higgs jets for events passing the pre-selection. The b-jets containing a muon with $p_T > 4$ GeV and $|\eta| < 2.5$ have their four-vector modified by the muon's momentum. If there is more than one muon inside a b-jet's cone, the one closest to the jet axis is used. The added muon energy is corrected for any depositions already in the calorimeter. A second, p_T-dependant, correction of up to 5% is applied to account for biases in the Jet Energy Scale (JES). The correction is based on VH samples[39]. The two corrections map to 1.5% and 0.7% improvements in the m_{bb} resolution correpondingly.

The following is the list of the considered systematics. They are grouped into several categories.

- **Detector-related:** Jet energy scale and resolution, online and offline b-tagging scale factor, modelling of jet width, luminosity uncertainty
- **MC statistics:** Statistical uncertainty on the expected number of signal events.
- **MC signal modelling:** Signal acceptance uncertainty due to parton shower, QCD

scale and parton density function.

- **Z yield:** Uncertainty on the prediction of the peaking Z background.
- **Choice of function:** Mismodelling of the non-resonant background due to the fit function choice.
- **Statistical:** Statistical uncertainty.

The systematics effecting the shape of the m_{bb} background are calculated using pseudo-experiments. They are found to be compatible with the observed values from the final fit of the unblinded dataset.

Other commonalities include the definition of background, the signal/control regions and the fit range. Both analyses assume that the m_{bb} spectrum is made up of three components; the resonant $H \to b\bar{b}$ signal, resonant $Z \to b\bar{b}$ and non-resonant QCD background. The amount of signal is extracted via a fit to the m_{bb} spectrum in the range of 70 GeV to 300 GeV. The fitting function is a linear sum of the three components. The $H \to b\bar{b}$ and $Z \to b\bar{b}$ peaks are modelled using templates from fully-simulated Monte Carlo. The position and width of the peak templates are fixed in the fitting procedure. The QCD background is modelled using a polynomial, with the degree depending on the selection. The coefficients correspond to the number of events in each component. The extracted Higgs signal yield is cited as the signal strength

$$\mu = \frac{N_{\text{obs}}}{N_{\text{exp}}}. \tag{4.1}$$

In the case of the $Z \to b\bar{b}$ background, the number of events is fixed to the Monte Carlo prediction, equivalent to subtracting its contribution from the histogram. There are some differences in the fitting procedure between cut-based and multivariates analyses that will be discussed in their corresponding sections.

The signal region is defined as 100 GeV $< m_{bb} <$ 140 GeV and is blinded during the development of the selection. The behaviour of variables is studied via control regions consisting of 20 GeV side-bands on either side of the SR. Their dependence on m_{bb} was checked by comparing the upper (80 GeV $< m_{bb} <$ 100 GeV) and lower (140 GeV $< m_{bb} <$ 160 GeV) sidebands. For optimization, the amount of non-resonant background in the SR is estimated using an extrapolation via a polynomial fit to the blinded data.

4.2 Cut-based Analysis

The cut-based analysis uses simple cuts on variables to define the final selection. The choice of cut values is done by optimizing $N_{\text{sig}}/\sqrt{N_{\text{bkg}}}$ using the pre-selection as a baseline. Figure 4.4 shows the variables that are found to be m_{bb} independent and provide a good separation. The m_{bb}-independence is checked by comparing the variable distributions in the upper and lower sidebands. The optimal selection, along with the corresponding cuts, is:

- no fifth jet with $p_T > 25$ GeV and $|\eta| < 2.4$
- $\left|\eta_{jj}\right| > 3.0$
- $m_{jj} > 650$ GeV

The extra jet cut is motivated by the presence of only 4 jets in the tree-level VBF $H \to b\bar{b}$ diagram. The last two cuts are motivated by the forward nature of the two VBF jets. This directly leads to a large $\left|\Delta\eta_{jj}\right|$ and indirectly to a large m_{jj}. The correlation between $\left|\Delta\eta_{jj}\right|$ and m_{jj} is checked and found to be negligible for the choice of cut-value. Table 4.5 shows the cutflow for the signal and background following this selection. There are 69 signal events (15% ggF) expected in the final sample.

The fit of the m_{bb} distribution uses an unbinned likelihood fit implemented by the RooFit[93] statistical framework. A binned fit would result in a biased result, because

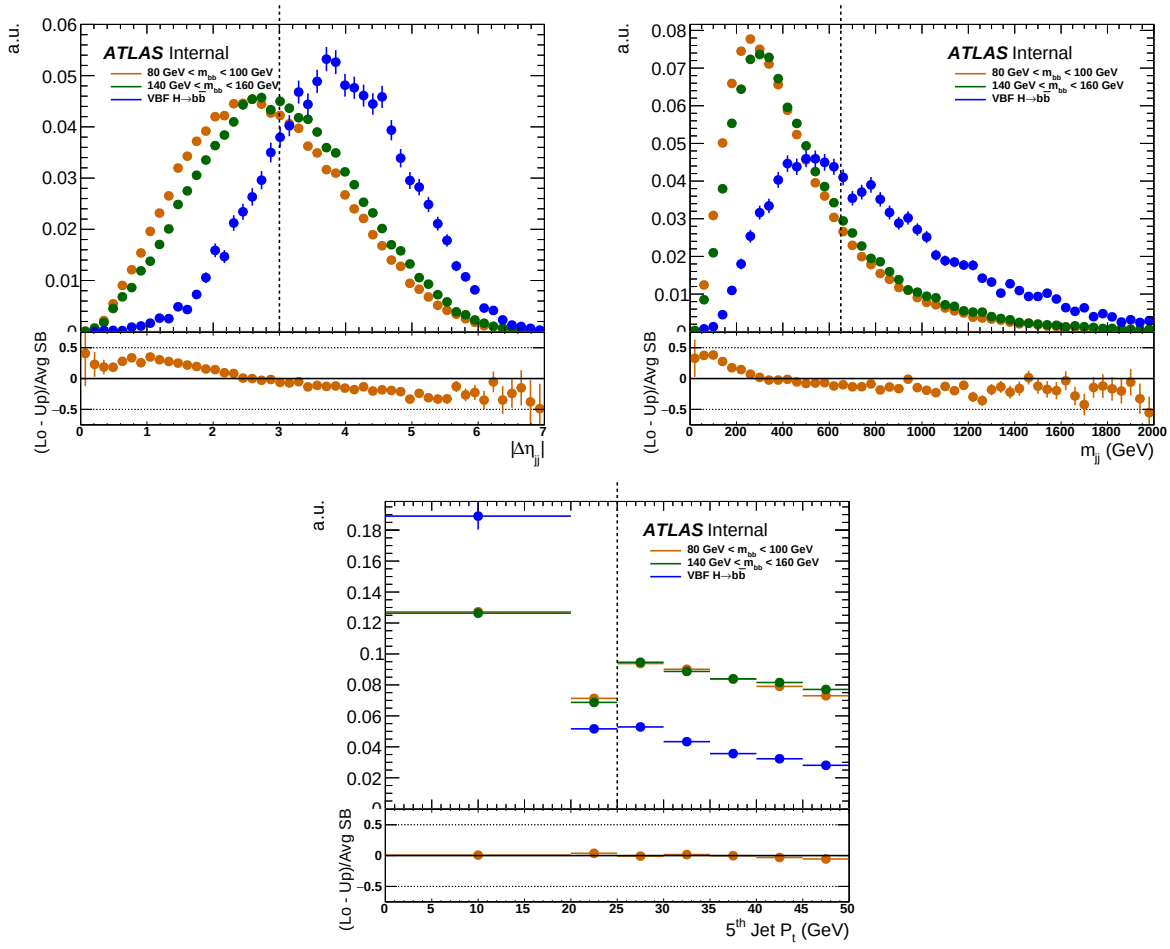


Figure 4.4: Distributions of the variables used for optimization of the analysis. They are normalized to unit area to allow for a comparison of the upper (green) and lower (orange) sidebands with the signal. The ratio plot compares the two sidebands to check their dependence on m_{bb}. The dashed line corresponds to the cut value in the cut-based selection.

Process	Selection criterion	N_{Events}	Efficiency	Total efficiency
VBF Higgs	pre-selection	129	—	6.97×10^{-3}
	no extra jets with $p_T > 25$ GeV and $\|\eta\| < 2.4$	104	0.808	5.97×10^{-3}
	$\|\Delta\eta_{jj}\| > 3.0$	89	0.797	4.49×10^{-3}
	$m_{jj} > 650$ GeV	59	0.710	3.18×10^{-3}
ggF Higgs	pre-selection	95	—	2.27×10^{-3}
	no extra jets with $p_T > 25$ GeV and $\|\eta\| < 2.4$	51	0.541	1.23×10^{-3}
	$\|\Delta\eta_{jj}\| > 3.0$	21	0.402	4.93×10^{-4}
	$m_{jj} > 650$ GeV	10	0.491	2.42×10^{-4}
Z	pre-selection	3909	—	6.05×10^{-6}
	no extra jets with $p_T > 25$ GeV and $\|\eta\| < 2.4$	2299	0.588	3.56×10^{-6}
	$\|\Delta\eta_{jj}\| > 3.0$	555	0.242	8.59×10^{-7}
	$m_{jj} > 650$ GeV	271	0.488	4.20×10^{-7}
Data	pre-selection	874539	—	5.39×10^{-10}
	data-only: blind 100 GeV $< m_{bb} <$ 140 GeV	730013	0.835	4.50×10^{-10}
	no extra jets with $p_T > 25$ GeV and $\|\eta\| < 2.4$	423222	0.580	2.61×10^{-10}
	$\|\Delta\eta_{jj}\| > 3.0$	193022	0.456	1.19×10^{-10}
	$m_{jj} > 650$ GeV	87895	0.455	5.41×10^{-11}

Table 4.5: Cutflow for additional selection in the cut-based analysis.

RooFit speeds up the NLL minimization algorithm by using the bin center to define the x-value. This assumption breaks down for steeply-falling distributions, which is a common feature of most combanitoric backgrounds. Using the average x-value instead would alleviate this problem. The fit model used is described in section 4.1. In addition, the peaking components are modelled by a bukin function[6] fitted to the Monte Carlo prediction. This provides a smooth model and is consistent with the VH analysis. An example of a bukin-based template is shown in figure 4.5. A fourth-order Bernstein polynomial is used for the non-resonant background, with all parameters left to float during the fit. The polynomial order is increased until the minimized negative log likelihood (NLL) no longer changed. The next order is also chosen as the alternate function to evaluate the function choice systematic.

The validity of the fit is validated using a linearity test. A background-only fit is done to the blinded dataset to estimate the background contribution. It is then combined with the signal template corresponding to different μ values to generate pseudo-experiments. Each pseudo-experiment is used as a dress-rehearsal to determine an extracted signal strength.

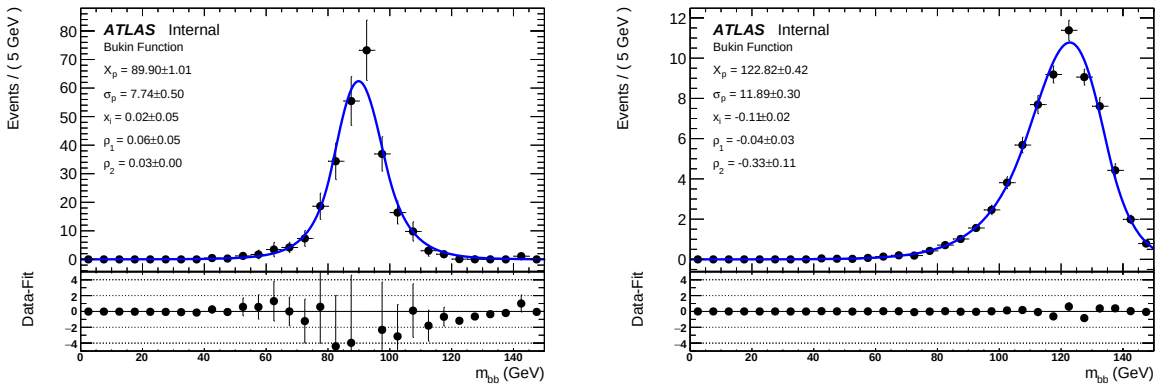


Figure 4.5: Example fit of the bukin function used to model the peaking components in the data. The nominal distributions for the Z background (left) and Higgs signal (right) are shown.

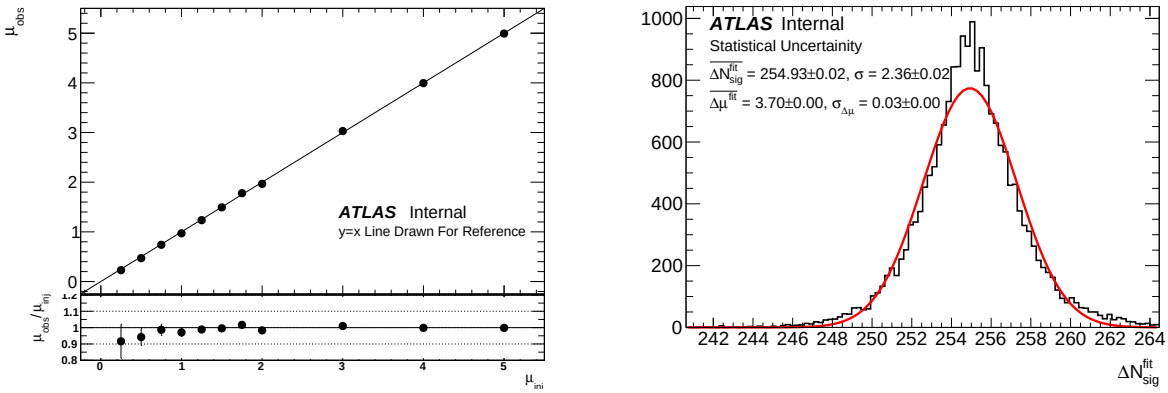


Figure 4.6: Results of the linearity test using pseudo-experiments. Left: Extracted signal strength vs injected signal strength. Right: Distribution of extracted signal strength for $\mu_{\text{inj}} = 1$.

The mean and standard deviation of the extracted signal strengths as a function of injected signal strength is shown in figure 4.6. It demonstrates that the fitting procedure can extract the correct signal strength. The standard deviation also gives the expected statistical uncertainty.

A similar procedure is used to evaluate the expected systematics affecting the m_{bb} distribution. The same pseudo-experiments as above are used. However the fitting templates are created using Monte Carlo smeared by $\pm\sigma$ for the systematic under study. The average

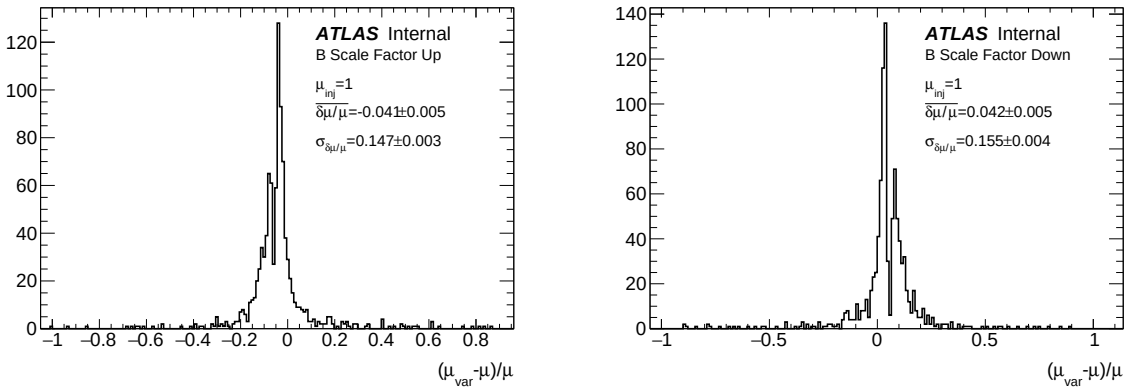


Figure 4.7: Example of the pseudo-experiments used to calculate the expected systematic variations. The pseudo-data is fit with the b-tagging scale factor up (left) and down (right) variations applied to the peak models.

difference from the nominal μ,

$$\delta\mu = \frac{\mu_{var} - \mu}{\mu}, \tag{4.2}$$

is used to define the final expected uncertainty. Figure 4.7 shows an example distribution for the offline b-tagging scale factor uncertainty. The bimodal distribution is expected from the leakage of the Z peak into the SR. Floating the Z peak during the fit would have removed this feature. This procedure is repeated for different signal strengths, demonstrating the constant relative systematic seen in figure 4.8. Table 4.6 lists the expected uncertainties.

The uncertainty on the choice of the background function is evaluated by scaling the pre-selected m_{bb} distribution to the final selection. Figure 4.9 shows the ratio of the two distributions, with the linear fit used as the scale factor. The resulting distribution has a shape similar to the final selection, but is background dominated. The scaled distribution is fit using the main three-component model described earlier. Because the Z contribution is not negligible, the bukin function used for the $Z \to b\bar{b}$ prediction is derived from a similarly scaled $Z \to b\bar{b}$ Monte Carlo. For the non-resonant background, different functional forms are tested.

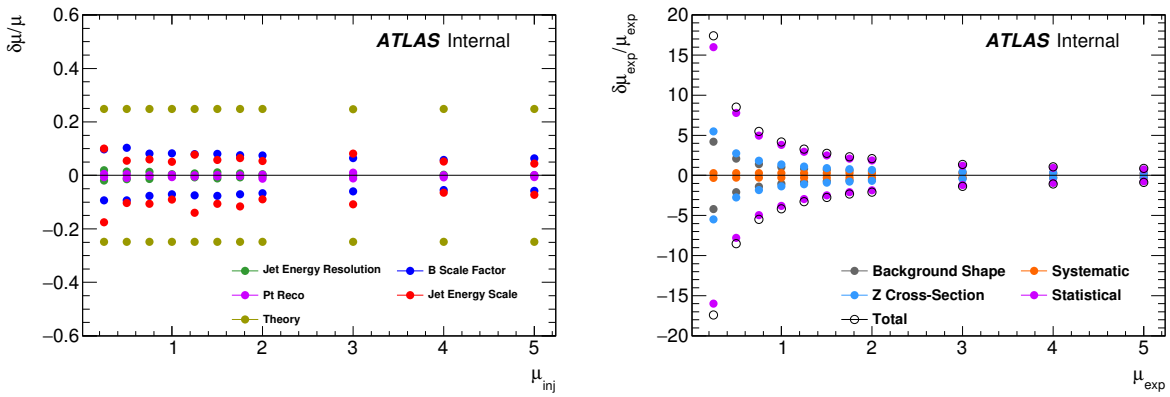


Figure 4.8: The relative uncertainty on the signal strength due to different components Left: Systematic uncertainty broken up into different components. Right: Total, statistical and systematic uncertainties.

Source of uncertainty	$\delta\mu_{\text{exp}}$
Statistics	
	± 3.7
Background modelling	
Multi-jet background	± 1.05
Z cross-section theory	± 1.37

Source of uncertainty	Uncertainty on the acceptance and cross section			Fit	$\delta\mu_{\text{exp}}/\mu_{\text{exp}}$
	$\delta N_{\text{VBF}}/N_{\text{VBF}}$	$\delta N_{ggF}/N_{ggF}$	$\delta N_{\text{total}}/N_{\text{total}}$		
Experimental uncertainty					
B-tag efficiency	$\pm 5.3\%$	$\pm 5.7\%$	$\pm 5.3\%$	x	$+4.2\%/-4.1\%$
Trigger (online b-tagging) efficiency	$\pm 7.9\%$	$\pm 8.1\%$	$\pm 8.0\%$	x	$+7.1\%/-5.8\%$
p_T-reco	—	—	—	x	$+0.3\%/-0.6\%$
Jet energy scale	$\pm 11.9\%$	$\pm 14.8\%$	$\pm 12.1\%$	x	$+5.1\%/-9.0\%$
Jet energy resolution	$\pm 2.1\%$	$\pm 7.3\%$	$\pm 0.8\%$	x	$\pm 0.4\%$
Luminosity	$\pm 2.8\%$	$\pm 2.8\%$	$\pm 2.8\%$		$\pm 2.8\%$
MC Statistics	$\pm 1.4\%$	$\pm 9.1\%$	$\pm 1.8\%$		$\pm 1.8\%$
Theoretical uncertainty					
QCD Scale (jet multiplicity)	$\pm 4.5\%$	$\pm 116.8\%$	$\pm 17.6\%$		$\pm 17.6\%$
QCD Scale (kinematic variables)	$\pm 2.0\%$	$\pm 86.2\%$	$\pm 12.8\%$		$\pm 12.8\%$
Parton shower	$\pm 10.5\%$	$\pm 10.8\%$	$\pm 10.5\%$		$\pm 10.5\%$
PDFs	$\pm 3.2\%$	$\pm 7.3\%$	$\pm 2.9\%$		$\pm 2.9\%$
QCD shape corrections for ggF	$\pm 0.0\%$	$\pm 25.7\%$	$\pm 3.8\%$	x	$+3.6\%/-3.5\%$
EWK shape corrections for VBF	$\pm 2.0\%$	$\pm 0.0\%$	$\pm 1.7\%$	x	$+1.8\%/-1.7\%$
$H \to b\bar{b}$ Branching Fraction		$+3.2\%/-3.3\%$			$+3.2\%/-3.3\%$
Total systematic uncertainty excluding the background modelling					
	$\pm 25.2\%$	$\pm 153.0\%$	$\pm 33.4\%$		$\pm 28\%$

Table 4.6: Summary of the expected uncertainties divided by source.

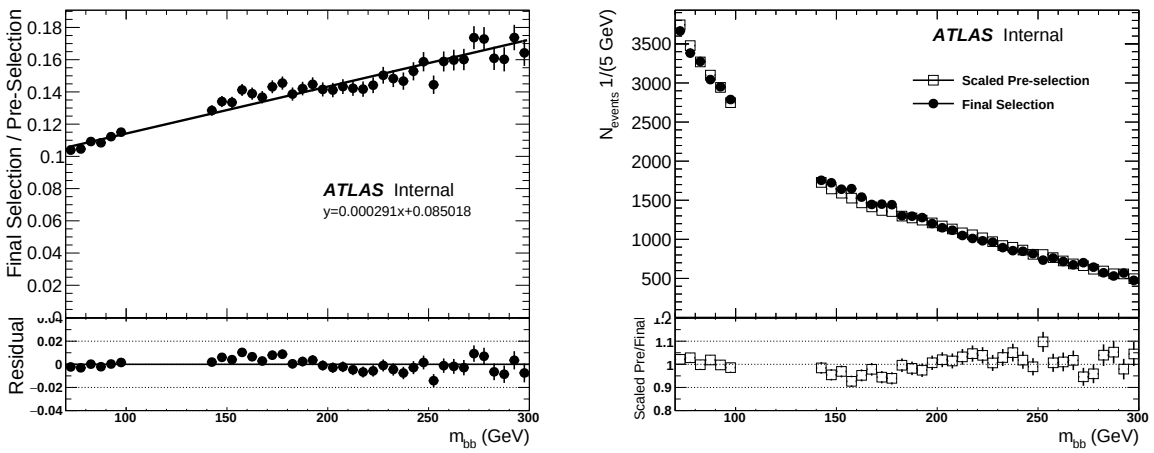


Figure 4.9: Construction of the control region used for testing the non-resonant background function choices. Left: Ratio of the pre-selection to the final selection. The fitted line is used as a scale factor to construct the control region. Right: Sideband comparison of the final selection with the control region constructed by scaling pre-selection.

One class of functions are the Bernstein polynomials,

$$f_{\mathrm{polN}}(x) = \sum_{k=0}^{k=N} \left(\frac{x-70}{230}\right)^{k} \left(\frac{300-x}{230}\right)^{N-k}, \tag{4.3}$$

with N corresponding to the polynomial order. The second class is a sum of exponentials,

$$f_{\mathrm{expN}}(x) = \sum_{k=0}^{k=N} \exp\left(\frac{-x}{\tau_N}\right), \tag{4.4}$$

with N up to 2. In case of the double exponential, τ_1 is restricted from 50 GeV to 200 GeV and τ_2 to the range spanning 1000 GeV and 100000 GeV. The last class is a polynomial times an exponential,

$$f_{\mathrm{polN-exp}}(x) = \left(\sum_{k=0}^{k=N} \left(\frac{x-70}{230}\right)^{k} \left(\frac{300-x}{230}\right)^{N-k}\right) \exp\left(\frac{-x}{\tau}\right). \tag{4.5}$$

In all definitions, x stands for m_{bb}. They also contain an extra factor (not shown) for

Function	$\mu_{\text{alternative}} - \mu_{\text{pol4}}$
pol5	1.05
pol3-exp	11.60
pol4-exp	0.90

Table 4.7: The difference between the fitted signal strength in the scaled pre-selection sample when different functions are used for the non-resonant background.

Process	Selection criterion	N_{Events} (entire m_{bb} region)	N_{Events} (signal region)
Data	pre-selection	874539	144526
	no extra jets with $p_T > 25$ GeV and $\|\eta\| < 2.4$	506718	83496
	$\|\Delta\eta_{jj}\| > 3.0$	231991	38969
	$m_{jj} > 650$ GeV	105510	17615

Table 4.8: Cutflow of the data in the cut-based analysis after unblinding.

normalization.

The number of signal events in the scaled validation sample is expected to be zero, with any deviations either statistical or due to mismodelling. To correct for statistical fluctuations, the N_{sig} from the nominal fourth order polynomial is used as a reference. Table 4.7 shows the deviations of N_{sig} when the fifth order polynomial and fourth-order polynomial times exponential are used, with corresponding fits shown in figure 4.10. Both give a good description of the non-resonant background in background-only fits of the scaled validation sample. The value obtained from the fifth order polynomial is used to define the function choice systematic, with the latter acting as validation. As a result, an absolute value of $\Delta\mu$=1.05 is used for the function choice systematic.

The cutflow inside the SR after unblinding is shown in table 4.8. Figure 4.11 shows the fit of the unblinded m_{bb} distribution. The observed $H \to b\bar{b}$ signal strength is -5.2 ± 3.7(stat), compatible with SM and no signal scenarios at 2σ level. The fit is also repeated with varied templates from systematics, with the results summarized in table 4.6. The combined systematical uncertainty is $^{+2.5}_{-2.2}$. The lack of a signal leads to limit setting using the CL_s method as the next step.

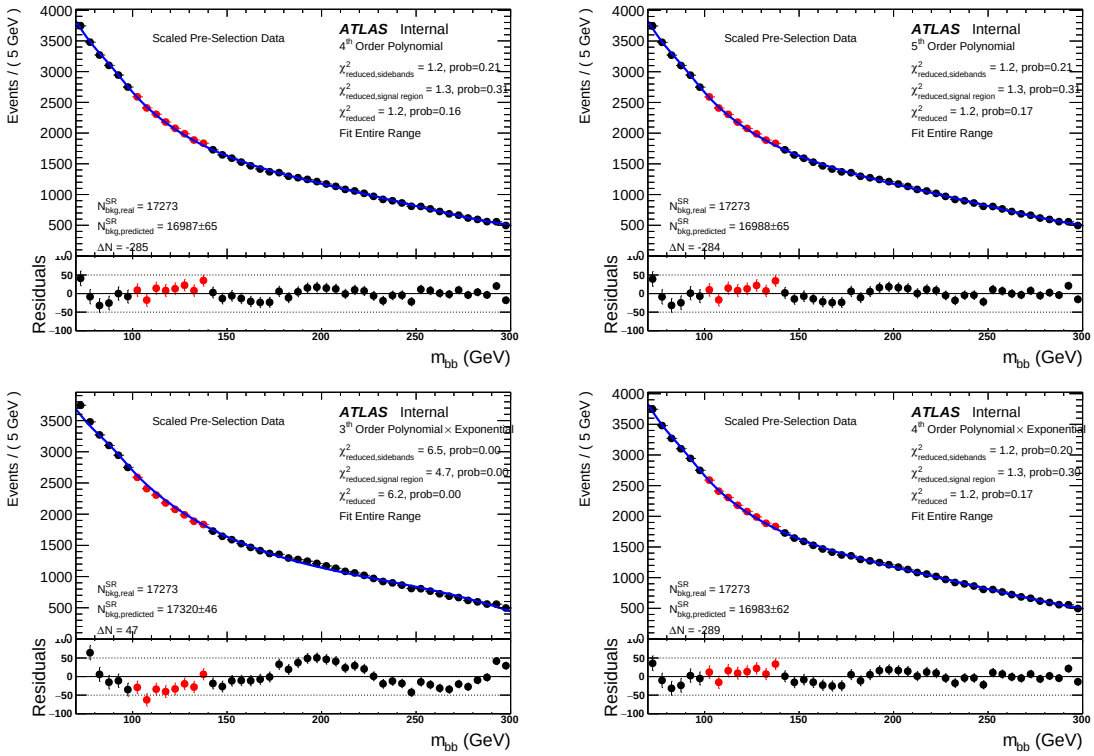


Figure 4.10: The result of fitting a fourth-order polynomial (top-left), fifth-order polynomial (top-right), third-order polynomial times exponential (bottom-left) and fourth-order polynomial times exponential (bottom-right) to the scaled pre-selection sample.

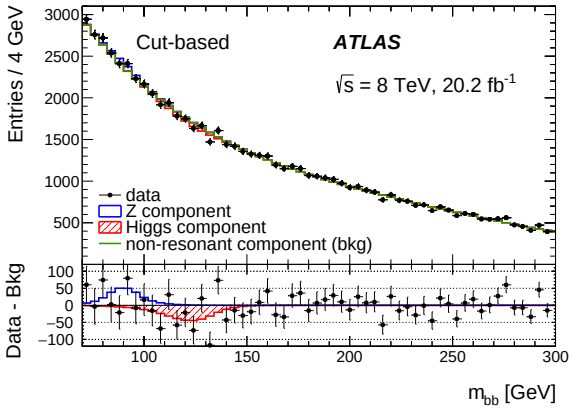


Figure 4.11: The fit of the unblinded m_{bb} distribution.

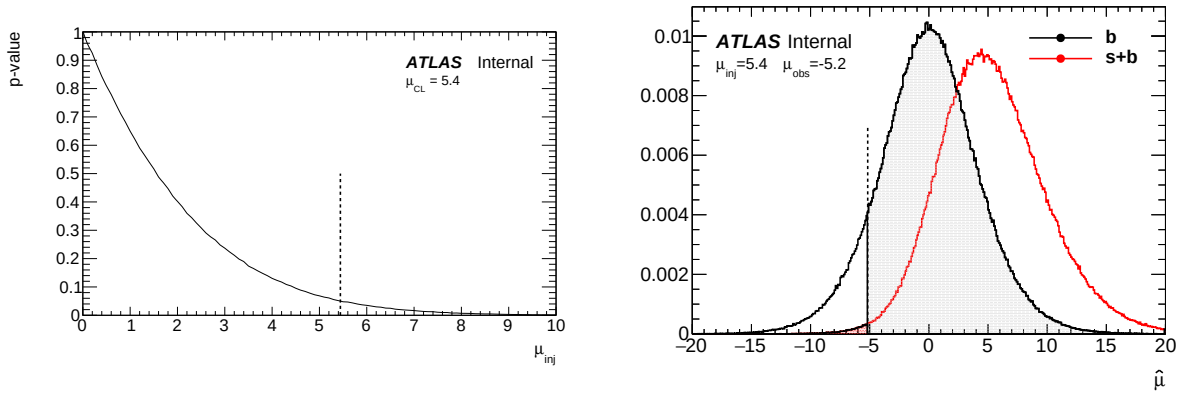


Figure 4.12: Left: The p-value as a function of the signal strength hypothesis, with the dashed line indicating the 95% CL_s exclusion limit at $\mu = 5.4$. Right: The corresponding probability density functions for signal+background and background-only hypothesis.

The limit setting is done by comparing the background-only and signal + background hypotheses. The comparison is done using the CL_s test statistic, defined by

$$CL_s = \frac{P(N_{\text{sig}} \leq N_{\text{obs}}|S+B)}{1 - P(N_{\text{sig}} \geq N_{\text{obs}}|B)}, \tag{4.6}$$

where numerator is the probability to observe less than than N_{obs} events under the signal+background hypothesis and the denominator is to observe less than N_{obs} signal events under the background-only hypothesis. A Gaussian shape with a relative width is assumed for all systematic uncertainties. The 95% CL_s exclusion limit is found at $\mu = 5.4$. Figure 4.12 shows the probablilty density distributions without and with signal at this value. The expected limit from a null-signal extraction is $\mu < 8.5^{+3.0}_{-2.5}$.

4.3 Multivariate Analysis

The mutlivariate analysis employs the TMVA[73] implementation of a Boosted Decision Tree (BDT) to separate the Higgs signal from the non-resonant background. The input variables are required to be uncorrelated with m_{bb}, allowing the events in the sideband regions (lower: 70 GeV $< m_{bb} <$ 90 GeV, upper: 150 GeV $< m_{bb} <$ 190 GeV) to be used for training

and validation. Although no attempt is made to discriminate between the signal and the Z background, there small amounts of the Z resonance present in the lower sideband. Only the VBF sample is included as signal during training. The variables, ordered in decreasing discriminating power, are

- Jet width of the two VBF jets with $|\eta| < 2.1$:

$$w_{\text{calo}} = \sum_{c \in \text{cluster}} \frac{E_{T,c} \Delta R_{\text{jet},c}}{E_{T,c}} \tag{4.7}$$

- Scalar sum of the p_T's of additional jets with $p_T > 20$ GeV:
- Invariant mass of the two VBF jets (m_{jj})
- η separation between the two VBF jets ($\left|\Delta\eta_{jj}\right|$)
- The maximum $|\eta|$ of the two VBF jets
- Separation between the $|\eta|$ averages of the two VBF jets and two Higgs jets:

$$\frac{|\eta_{j1}| + |\eta_{j2}|}{2} - \frac{|\eta_{b1}| + |\eta_{b2}|}{2} \tag{4.8}$$

- Cosine of the polar angle of the cross product of the VBF jets momenta in the rest frame of the Higgs boson candidate

They exploit the forward nature of the VBF jets and the fact that they are initiated by light quarks. The jet width, discussed in section 3.4, serves as a proxy for quark/gluon tagging. The last variable is sensitive to the production mechanism. The comparison of the variable shapes compared between side-bands is shown in figure 4.13 and indicates no dependence on m_{jj}.

Figure 4.14 compares the BDT response of the entire pre-selected dataset with the $H \to b\bar{b}$ signal and Z boson background. The ggF $H \to b\bar{b}$ component of the signal looks similar to

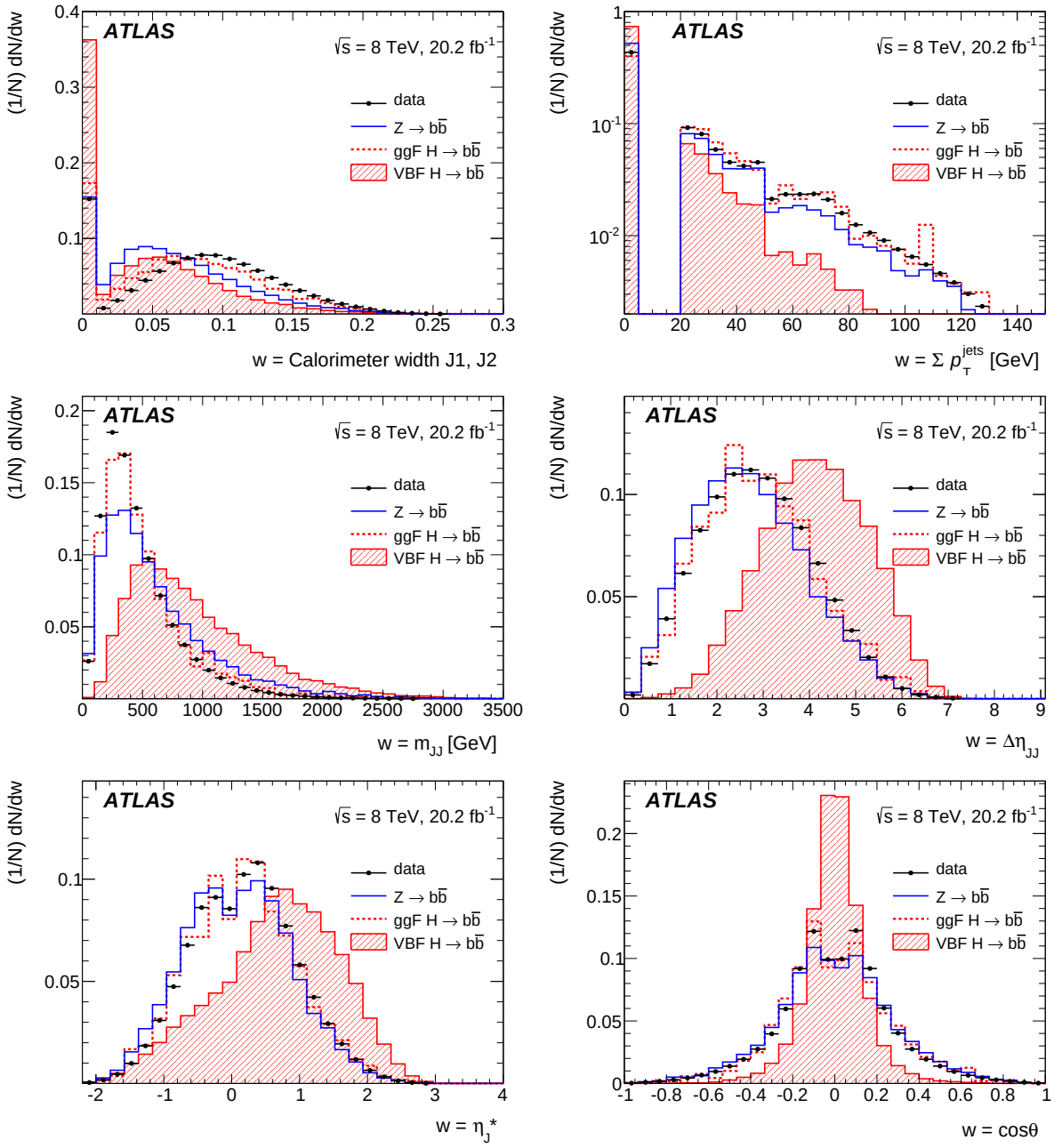


Figure 4.13: The input variables to the BDT used for disciminating the $H \to b\bar{b}$ signal from the non-resonant background. Distributions for the data in the lower and upper sidebands (black points), $Z \to b\bar{b}$ background, ggF and VBF $H \to b\bar{b}$ signal are shown. The variables are ordered left-to-right in descreasing separation power.

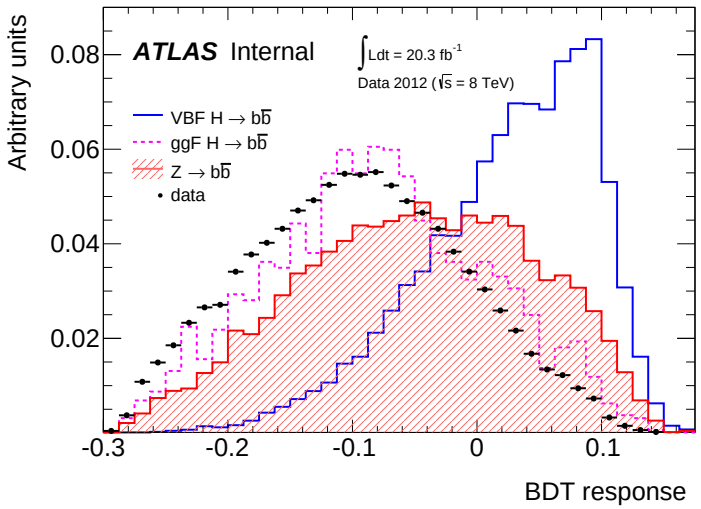


Figure 4.14: The response of the trained BDT used for the multivariate $H \to b\bar{b}$ analysis. Distributions for the data in the lower and upper sidebands (black points), $Z \to b\bar{b}$ background, ggF and VBF $H \to b\bar{b}$ signal are shown.

Process	Pre-selection	Category I (-0.08 to 0.01)	Category II (0.01 to 0.06)	Category III (0.06 to 0.09)	Category IV ($>$ 0.09)
VBF $H \to b\bar{b}$	130	39	33	23	19
ggF $H \to b\bar{b}$	94	331	9.5	3.8	1.6
$Z \to b\bar{b}$	3700	1100	350	97	49
Data	554302	176073	46912	15015	6493

Table 4.9: The number of expected events in each BDT category. The regions are ordered by descreasing sensitivity, with the BDT score boundary listead in the table header.

the QCD dominated background. However good discrimination between the VBF $H \to b\bar{b}$ signal, the target of the analysis, is visibile. To maximize the sensitivity of the analysis, events are sorted into four regions based on their BDT score. The boundaries between the regions are chosen to maximize the statistical significance, defined by $N_{\text{sig}}/\sqrt{N_{\text{sig}} + N_{\text{bkg}}}$. Table 4.9 describes the four regions along with the expected number of events in each. The m_{bb} distributions from the four regions are fit simultanously to extract a single μ.

The data is fit with the three-component model described in section 4.1. To remove the assumptions on the functional form of the peaking components, distributions from Monte

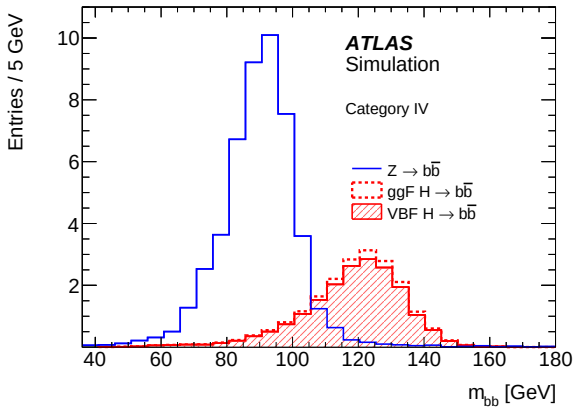


Figure 4.15: The m_{bb} distributions used to model the Z and Higgs components in the MVA $H \to b\bar{b}$ analysis.

Carlo are used directly. Since no dependence of m_{bb} on the BDT category is seen, a single template is constructed from the inclusive sample. In addition, the MC statistical uncertainty is reduced by smoothing the templates with the 353QH algorithm[67]. Figure 4.15 shows the templates normalized to the event yield in the most sensitive category. The shapes of the ggF and VBF signal components are similar.

A smootly falling function is used to model the non-resonant background. The functional forms described in 4.2 are tested by fitting the blinded m_{bb} distributions. The fit quality and number free parameters (less is better) are used to pick the nominal and alterante background models. The chosen functions and their free parmeters depend on the BDT category, as summarized in table 4.10. Polynomials are chosen for all nominal functions. The shape of the non-resonant function is fixed from the sideband fit. A systematic due to a possible bias from the function choice is assigned by comparing the predictions of the nominal and alternate functions.

The four BDT categories are profiled simultaneusly with a likelihood consisting of bin-by-bin poisson probablilities. The signal strength, multiplying the cross-section predictions in each BDT category, is parametrized in terms of several nuisance parameters. The NPs

	Category I	Category II	Category III	Category IV
Nominal	4^{th} pol.	3^{rd} pol.	3^{rd} pol.	3^{rd} pol.
Alternate	2^{nd} pol. $\times$ exponential	3 exponentials	2 exponentials	exponential

Table 4.10: The nominal and alternate functions used to described the non-resonnant background in the MVA VBF $H \to b\bar{b}$ analysis.

NP	Category	Constraint
μ_Z	All	1.00 ± 0.26
$\alpha_{Z,I}$	Category I	1.00 ± 0.30
$\alpha_{Z,II}$	Category II	1.00 ± 0.34
$\alpha_{Z,III}$	Category III	1.00 ± 0.34
$\alpha_{Z,IV}$	Category IV	1.00 ± 0.45

Table 4.11: Constraints applied to the Z yield in MVA profile likelihood.

corresponding to the systematics listed in section 4.1 are constrainted with Gaussian distributions. The Z peak normalization is constrained to the Standard Model prediction, with corresponding nuisance parametres summarized in table 4.11. They consists of uncertainties from next-to-leading order corrections to the total cross-section (μ_Z) and category dependance ($\alpha_{Z,i}$). The normalization and shape of the non-resonant background is floated without any priors. Differences between the normal and alternate function are studied via eigenvalue decomposition of fits preformed on pseudo-data. Nuisance parameters for the dominant components are added to the likelihood. This approximates floating the function parameters without double-counting the statistical uncertainty.

The result of the profile likelihood fit on the unblinded dataset is shown in figure 4.16. It measures a Higgs signal strenght of $\mu = -0.8 \pm 1.3(\text{stat})^{+1.8}_{-1.9}(\text{sys})$. Table 4.12 shows the breakdown of the uncertainties and compares to them with values obtained by the cut-based analysis.

The observed μ is consistent with both the Standard Model prediction and a null-signal hypthesis. To determine the maximum possible μ given the data, a test statistic is used to

Source of Uncertainty		Uncertainty on μ	
		MVA	Cut-based
Experimental Uncertainties	Detector-related	$+0.2/-0.3$	$+1.6/-1.2$
	MC statistics	± 0.4	± 0.1
Theoretical Uncertainties	MC signal modelling	± 0.4	± 1.3
	Z yield	$+0.6/-0.5$	± 1.4
Non-Resonant Background Modelling	Choice of function	± 1.0	± 1.0
	Sideband statistics	± 1.7	± 3.7
Statistical uncertainties		± 1.3	
Total		± 2.3	$+4.6/-4.4$

Table 4.12: The observed uncertainties in the MVA (left column) and cut-based (right column) VBF $H \to b\bar{b}$ analyses.

set a limit via the CL_s procedure. The test statistics is defined as a ratio of likelihoods,

$$q_\mu = 2 \ln \mathcal{L}(\mu, \vec{\theta}_\mu) / \mathcal{L}(\hat{\mu}, \vec{\hat{\theta}}_\mu), \tag{4.9}$$

with $\vec{\theta}$ corresponding to NPs that maximize the likelihood for a fixed μ and the denominator likelihood maximizing both $\hat{\mu}$ and $\vec{\hat{\theta}}_\mu$. The observed 95% confidence level is found to be $\mu < 4.7$. The expected limits in absence of a signal and SM hypothesis are 5.4 and 5.7 correspondigly.

4.4 Summary and Outlook of VBF $H \to b\bar{b}$

Two overlapping analysis are used to search for the decay of the Higgs boson into two bottom quarks using the Run 1 ATLAS data. A multi-variate analysis uses several event-level variables as inputs to a Boosted Decision Tree to separate the signal from the enormous QCD background. As a cross-check, a cut-based analysis uses fixed cuts on similar variables. The multivariate analysis observed a signal strength of

$$\mu_{\text{MVA}} = -0.8 \pm 1.3(\text{stat})^{+1.8}_{-1.9}(\text{sys}) \tag{4.10}$$

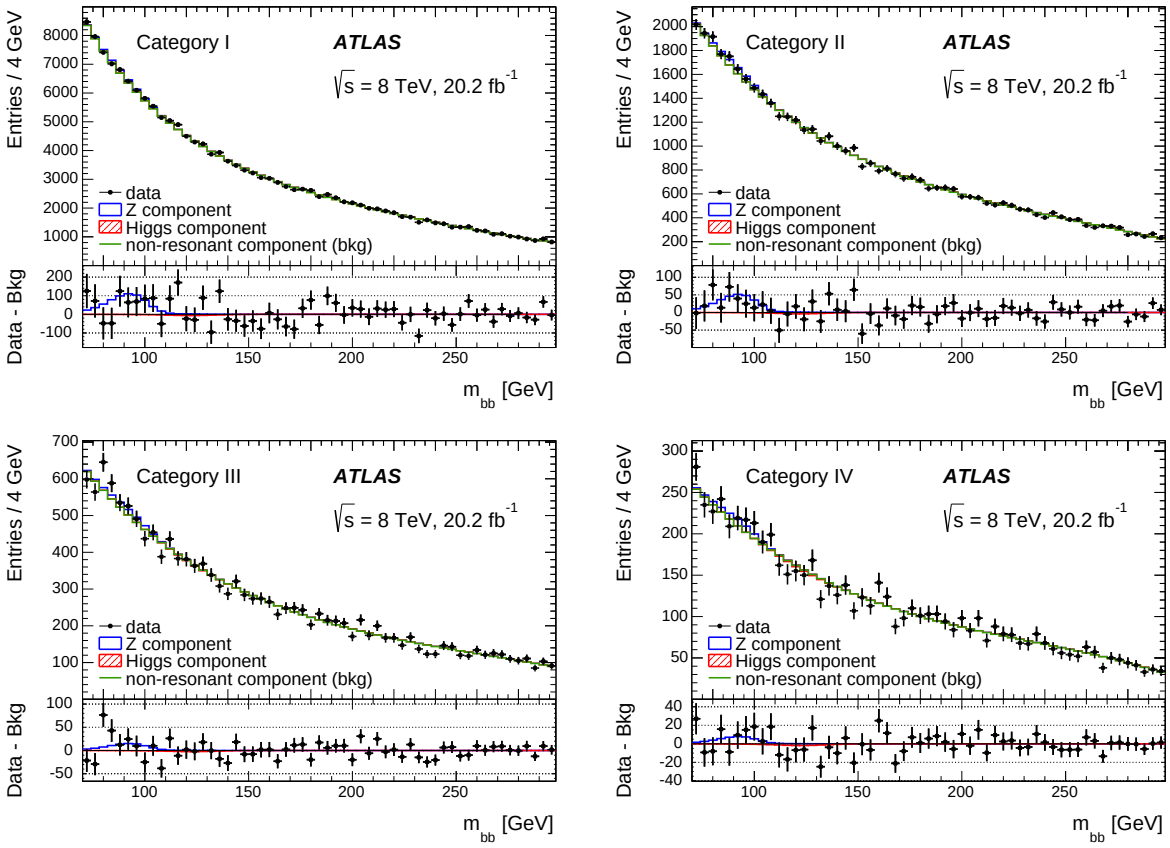


Figure 4.16: The unlinded dataset in the four BDT categories. The Higgs signal (red) and non-resonant (green) and Z (blue background) are shown with the normalization from maximizing the likelihood.

while the cut-based analysis observed

$$\mu_{\text{cut-based}} = -5.2 \pm 3.7(\text{stat})^{+2.7}_{-2.5}(\text{sys}). \tag{4.11}$$

Neighter having seen a significant excess in the m_{bb} spectrum, the CL_s procedure is used to set limits on the $H \to b\bar{b}$ signal. The observed limits are

$$\mu_{\text{MVA}} < 4.7 \qquad \text{and} \qquad \mu_{\text{cut-based}} < 5.4 \tag{4.12}$$

at the 95% confidence level.

For Run 2, the ATLAS experiment and accelarating complex has undergone several improvements that increase the prospects the VBF $H \to b\bar{b}$ search. First the Large Hadron Collider will run at a higher center-of-mass energy and luminosity. The Higgs cross-section increases with energy more than the background, going from 1.6 pb to 3.7 pb. On the other hand, the background only increases by a third from 300 μb to 400 μb. This leads to a factor of 2 increase in sensitivity as measured by $s/\sqrt{b}$. The increase in luminosity will also increase the available statistics. Already at the end of the first year, we have over 30 fb^{-1} of data available. This is a 50% improvement over the dataset used in the Run 1 analysis. Around 300 fb^{-1} is expected by the end of Run 2.

The detector upgrades in the trigger system are also useful for the VBF $H \to b\bar{b}$ search. The addition of the FastTracKer[89] and the L1Topo[26] hardware will allow the design of new menus to improve the trigger efficiency. FTK will provide offline-like track performance for every event passing the L1 trigger. This will be important at high luminosities, where the CPU-based tracking software will not be fast enough. The hardware tracking will allow ATLAS to maintain the current thresholds by having the ability run b-tagging algorithms even for low-p_T jets. But the immediate gain will come from the L1Topo trigger, because it

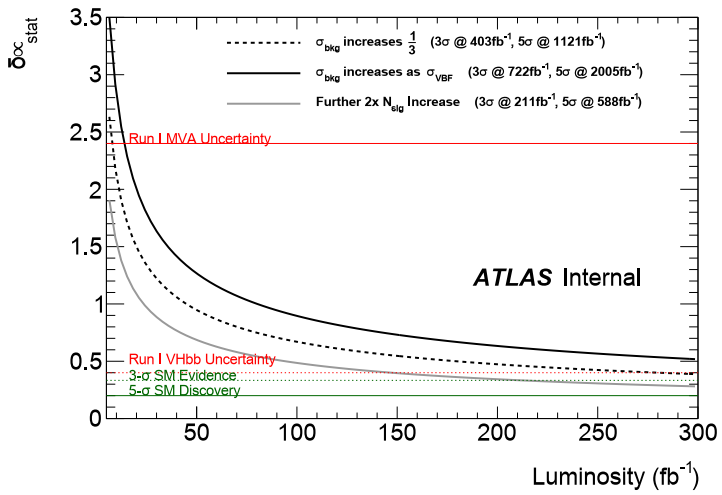


Figure 4.17: Extrapolation of the VBF $H \to b\bar{b}$ limit into the Run 2 dataset. Assumptions that the background increases like QCD (dashed), like VBF signal (solid) and signal increases by an additional factor of two (gray) are tested. The red lines show the current VBF $H \to b\bar{b}$ (solid) and V$H \to b\bar{b}$ (dashed) results.

helps to bypass the L1-limitation of the current analysis. It allows the L1-trigger to calculate correlations between the kinematics of detected particles, like m_{jj} and $|\Delta\eta_{jj}|$, already present in the offline selection. Cutting earlier on these variables will increase the signal acceptance by allowing for lower jet p_T thresholds. The impact of L1Topo-based triggers on the VBF $H \to b\bar{b}$ analysis still need to evaluated.

The sensitivity of the VBF $H \to b\bar{b}$ analysis in Run 2 can be predicted using a simple calculation. Since the search is statistics-limited, the limits scale like $N_{\text{sig}}/\sqrt{N_{\text{bkg}}}$. For example, the improvement due to the increase in the center of mass energy will reduce the statistical uncertainty by a factor of 2 at 20 fb^{-1}. Figure 4.17 shows the significance of the signal as a function of luminosity, using the MVA statistical uncertainty as a baseline. The estimate for the increase of background presented earlier is just a rough estimation, since it applies to inclusive QCD production. For this reasion, estimates for two assumptions are shown. One follows the inclusive cross-section increase and the second assumes that the background scales like the signal. The indication is that the full Run 2 dataset is needed

for an observation. For comparison the analysis requiring the production of the Higgs in association with a vector-boson is close to 3-sigma evidence with only the Run 1 dataset.

CHAPTER 5

STANDARDIZED RUN 2 DIJET RESONANCE PROCEDURE

Starting from this chapter, the content changes from the Run 1 search for the Standard Model Higgs to searches for new physics with data recorded at the start of Run 2. The subsequent two chapters describe the results of the sub-TeV diBjet and the dijet+ISR analyses. Both analysis were done under the management of the ATLAS Exotics group and, like the $H \to b\bar{b}$ search, they involve searching for a new particle decaying into two jets (dijet-like). Under the guise of the the high-mass dijet analysis[20], the Exotics group has developed a framework for doing resonance searches in the dijet invariant mass spectrum. The common framework provides a validated procedure that can be quickly applied to new searches. It was one of the key ingredients to get the two described analyses to progress at a fast pace.

The procedure consists of four stages; selection, background modelling, search phase and limit setting. The selection stage involves optimizing the event selection to maximize the signal sensitivity and comparing the recorded data with Monte Carlo prediction to spot any detector problem. Next a smooth function has to be chosen to model the smooth QCD background. The BumpHunter algorithm is then used in the search phase to identity the region where the data has the largest deviation from the fitted function. If no significant deviations are found, the limit setting procedure is used to eliminate parameter space in a series of benchmark models.

This chapter provides more details on each of the steps. The results of procedure application to the diBjet and dijet+ISR analysis are described in chapters 6 and 7 correspondingly.

5.1 Selection

The selection of all dijet-like analyses has several attributes in common. First is the application of the Good Run List to ensure that the relevant parts of detector were operational when the data was recorded. There are several valid GRLs describing the early Run 2 data. For 2015 running, the newly installed Inner B-Layer pixel detector had to be shut-off for a period that recorded 230 pb^{-1} of data. Similarly, the toroid magnets had problems during the early part of 2016 for 670 pb^{-1} of data. The decision whether to include the corresponding runs is analysis dependant. In addition to GRL, all analyses apply a standard set of event cleaning cuts to reject non-collision background and eliminate jets faked by noisy calorimeter cells. Also pile-up reweighting is applied to any simulated samples to ensure that the number of primary vertex distributions match the accelerator conditions.

The next step is to select the trigger that was used to record the desired events. An important part of the trigger selection is to measure the trigger efficiency and only select events for which the efficiency is constant. This eliminates any bias that the sharp turn-on curve can have on the observable variables. Since the relevant triggers require an object (jet or photon) to pass a p_T threshold, this leads to a requirement on the fully-calibrated p_T of the corresponding object. For example, the dijet+jet trigger required at least one jet with p_T above 380 GeV, as reconstructed by the trigger software. This translate to offline jet p_T of at least 430 GeV, resulting in such cut being applied in the dijet+jet selection.

The final common selection for the dijet-like searches increases the signal sensitivity by cutting on the difference in the pseudo-rapidity, defined as

$$y^* = \frac{|y_1 - y_2|}{2}, \tag{5.1}$$

where y_i is the rapidity of the resonance jet i. It selects jets that are close together in the longitudinal direction and effectively selects central jets. The purpose of the cut is to

reject the t-channel QCD diagrams that have a pole close to the beam. Studies of alternate variables suggest that y^*, by far, is the most powerful variable for signal optimization.

The individual analyses also apply a small set of dedicated cuts. For example, the diBjet analysis runs a b-tagging algorithm on the resonance jets. The extra cuts are be describes in their corresponding analysis section.

After the selection is applied, a comparison between the data and Monte Carlo is used to validate the selection. The Monte Carlo consists of generated events for the dominant background. Although there is no expectation that the simulation captures the physics and detector performance perfectly, simple comparisons are still useful. For example, the SHEPRA γ+jet simulation underestimates the subleading jet p_T spectrum in the dijet+γ analysis by 20% below 100 GeV. There is no expectation of an exact match, since multijet events are difficult to model and the prediction does not include contamination from misidentified photons. However the important part is that there are no sudden spikes or large drop-offs in the data indicating problems with the detector or the data acquisition system.

Finally, the dijet-like analyses emply a very loose blinding strategy. No blinding is done outside of debugging code on simulated samples and not looking at the m_{jj} distribution before other distributions are shown to be reasonable. There are no control or validation regions. The main deterrent for subjective bias is that the full procedure is fixed when the statistical interpretation is performed.

5.2 Background Model

As will be seen in the comparisons of the dijet invariant mass spectra with Monte Carlo, the simulation of QCD is not very accurate. This leads to the need for a data-driven approach to estimate the background. Since the background does not contain any known resonances

and is smoothly falling, the simplest approach is to perform a likelihood fit with a smoothly falling function. The main question is what the function should be. The answer to that depends on the luminosity. A small dataset has many statistical fluctuations, requiring a rigid function to smootly interpolate between the points. On the other hand, a large dataset has a small relative statistical uncertainty meaning that a flexible function is necessary to capture all of its features. To find the optimal compromise, the likelihood ratio test was used to choose from the class of nested functions parametrized by

$$f(x) = c_0(1-x)^{c_1}x^{c_2+c_3\ln x+c_4\ln x^2+c_5\ln x^2}, \quad x = \frac{m_{jj}}{\sqrt{s}}. \tag{5.2}$$

The c_i are parameters that can be floated or fixed to reduce the degrees of freedom, giving the nested property. For example, the three parameter function (f_3) is defined by $c_3 = c_4 = c_5 = 0$. The likelihood ratio test uses the Wilk's Theorem to provide the probability that two nested functions differening by m degrees of freedom give the same description of the data. According to Wilk's Theorem the likelihood ratio,

$$\rho_n = -2\ln\left(\frac{\mathcal{L}_{n\text{ par}}}{\mathcal{L}_{n+m\text{ par}}}\right), \tag{5.3}$$

follows a χ^2 distribution with degrees of freedom equalling m. For the procedure in the dijet-like analyses, only adjacent functions ($m = 1$) are compared.

If the probability is less than 5%, then the extra parameter in the $n+1$ function has a significant effect on the background model. Assuming that the significant effect is due to the need for a more flexible function, then it follows that $n+1$ function should be chosen as the nominal background model. This is best illustrated by following the probability of ρ_n in an accumulating dataset. For example, in the dijet+γ analysis the ρ_3 (3-parameter function is the lowest order considered) probability quickly drops around 1 fb^{-1}. The consequence is that the four-parameter function (ρ_4 remains high for the remainder of the dataset) is used

to model the final m_{jj} distribution.

It should be noted that a better likelihood does not equal goodness of fit. The χ^2 of the model still needs to be considered.

The likelihood ratio test is run on the recorded data to determine the final function choice. However simulated background samples are used at the development stage to validate the procedure. This leads to the need for pseudo-data that accounts for the statistical correlations seen in an accumulating dataset. Thus the following generation algorithm is employed to grow the "data-like" toys bin-by-bin using the background Monte Carlo as a probability density function. First the number of effective MC entries in each bin of the m_{jj} histogram is defined by

$$N_{\text{eff}} = \frac{(\sum w_i)^2}{\sum w_i^2}, \tag{5.4}$$

where w_i is the weight of event i. The number of effective entries is then compared against the number of expected entries (N_{exp}) at a given luminosity to ensure that the statistical uncertainty of the simulated sample is negligible when compared to the expected statistical uncertainty of the pseudo-data. It ensures that any deviations in the data-like sample are a result of mismodelling and not biased by statistical fluctuations. The next step is to generate a set of N_{eff} random numbers in the range $[0, 1]$. If the random number is smaller than $N_{\text{exp}}/N_{\text{eff}}$, then the content of the bin being processed is incremented by 1 event. The result is a bin content with expected mean N_{exp} and variation following binomial statistics. In the limit where $N_{\text{eff}} \gg N_{\text{exp}}$, the variation reduces to Poisson uncertainty.

The pseudo-data generation can be repeated for different values of N_{exp}, corresponding the requested luminosity values. The data-like statistics follow, because for two luminosities $\mathcal{L}_1 < \mathcal{L}_2$ the events passing $N_{\text{exp},1}/N_{\text{eff}}$ are a subset of events passing $N_{\text{exp},2}/N_{\text{eff}}$. This method also allows the luminosity steps to be further refined by rerunning the procedure only at the new values.

5.3 Search Phase

The search phase employs the BumpHunter algorithm[11] to locate the region with the most significant excess and quantify its likeliness as a statistical fluctuation. It works on a binned distribution, with the bin sizes corresponding to the detector resolution of m_{jj}. The detector resolution is determined by comparing truth jets with reconstructed jets after full detector simulation in QCD Pythia 8 dijet Monte Carlo.

The first step of the algorithm to estimate the background by fitting the function chosen in section 5.2. To locate the region with the most significant deviation, all combinations of adjacent bins from two to half the histogram range are tested. Poisson statistics are used to assign a probability to the deviation of the data in each range with respect to the background prediction. The region with the smallest probability, P_{sig}, is defined as the "signal". If $P_{\text{sig}} < 0.05$, then data is refit with the "signal" blinded and the search is repeated over the entire dataset. This ensures that the "signal" does not bias the background-only fit. The plot showing the tested range on the x-axis and its probability on the y-axis is called a tomography plot. It is a useful diagnostics plot, since it can show patterns in fluctuations or detect a second possible excess.

Since the BumpHunter algorithm searches through many regions, it is very likely that it will find an unlikely fluctuation. The look-elsewhere effect is accounted for by calculating the likelihood of P_{sig} given a background-only dataset. The fitted function is used to generate pseudo-experiments. The search algorithm is then run on each pseudo-experiment and the probability of the most significant deviation is recorded. The result is then used to define the global p-value of the signal as the probability to see an excess of P_{sig} or smaller. If the p-value is greater than 5%, then the signal is not real.

5.4 Limit Setting

In the case where no significant signal excess is observed, the Bayesian method is used to set the 95% credibility level (CL) on the $\sigma \times \mathcal{A}$ of the benchmark models. Signal samples were generated for different values of the resonance mass and a constant prior was assumed for the signal strength. A logarithm interpolation between the mass points provides continuous limits. Nuisance parameters are added to account for systematic uncertainties due to the luminosity, the parton distribution function, the background model, the jet energy scale and resolution. In the case of the diBjet search, the b-tagging algorithm introduces an additional systematic. The posterior probabilities of the systematics are integrated in the marginalization procedure. With the exeption of the background model uncertainty, all of the systematics are only applicable to the signal prediction.

The uncertainty on the luminosity is taken from the x-y beam separation scans[23] performed in August 2015 as $\pm 5\%$.

The LHAPDF [5] software provides an ensembles of parton distribution functions, with ability to reweight between the members of each ensemble. The set created by the NNPDF group was used when generating the signal samples. The standard deviation of the reweighted variations results in a $\pm 1\%$ effect on the signal acceptance.

The effects of the jet energy scale are described by 4 nuisance parameter recommended by the JetETMiss group. Alternate descriptions were also tested as a cross-check. For the energy scale uncertainty, varied signal templates are generated at $\pm 0.5, 1, 1.5, 2, 2.5, 3$ sigma of each nuisance parameter. The several steps in each nuisance parameter are required to not introduce negative values when extrapolating the templates during marginalization. The jet energy resolution uncertainty is taken as a flat $\pm 2\%$ observed of the acceptance variation in the signal window.

The uncertainty on the function used to model the QCD background model is a combination statistical uncertainty and a systematic due to the choice of the parametrization.

The covariance matrix computed by the minimization code cannot be trusted to calculate the statistical uncertainty as the function parameters are heavily correlated. Instead the nominal model is used to generate pseudo-experiments. Each pseudo-experiment is then fit using the same initial parameters. The standard deviation of the function value each bin provides the statistical uncertainty. For the choice of the function, the data itself is used to generate toys. The fits fom the nominal function and the function with an extra degree of freedom are compared. The root-mean-square difference in each bin is computed and used to set the scale of the uncertainty. A single nuisance parameter is then used to linearly interpolate from the nominal fit to the alternate version given by the average deviation.

Generic signals modeled by Gaussian distribution are also used as inputs to the limit setting procedure. The results for different Gaussian widths are reported. They can then be matched up to corresponding resonances in a variety of models. The same systematic uncertainties describes above are considered. For the jet energy scale, the widths are changed by 1% corresponding to the variations seen in simulated signal samples.

CHAPTER 6

DIJET RESONANCE WITH B-TAGGING

The ATLAS search for generic resonances decaying into two bottom quarks was first performed at the beginning of Run 2[51]. Similar searches have previously been done by the CDF experiment[55] at the Tevatron and the CMS experiment in Run 1[58]. The ATLAS analysis used the unprescaled single jet trigger, which is fully efficient for events containing at least one jet with p_T above 440 GeV. This constraints the minimum possible resonance mass to be above 1 TeV. However due to the poor efficiency of b-tagging at high p_T, the resulting limits are very weak. In fact, the inclusive high-mass dijet search[48] can set more stringent limits on the same model. This chapter examines an extension of the diBjet search to lower masses by using dedicated b-jet triggers. The low mass analysis searches for resonances in the 500 GeV $< m_{bb} <$ 1200 GeV range[49].

6.1 Signal Model

The main benchmark model used by the low-mass diBjet analysis is a SSM gauge boson, referred to as a Z'. It arises in extensions of the Standard Model that have a new U(1) symmetry. The Z' has the same couplings to the the fermions as the Standard Model Z boson. The implementation of the Z' in Pythia8, combined with the NNPDF 2.3 PDF set, is used to generate signal samples with masses of 0.6, 0.8, 1 and 1.25 TeV. Only the decay of the Z' into b-quarks is enabled during simulation.

However the real motivation for the diBjet search in this mass range is data-driven. In the 2015 data, the ATLAS experiment has observed a 4-sigma signal (local significance) in the diphoton mass spectrum at 750 GeV[50]. This has resulted in an avalanche of phenomenological papers. From an experimental point of view, the most useful publication was

[66]. It made no theoretical assumptions about the underlaying model. Instead it used the total width (Γ) and $\sigma_{\gamma\gamma}$ measured in the diphoton spectrum to predict the cross-section in an alternate channel, assuming the new resonance only couples to one additional particle. What follows is a summary of the calculation in the scenario where the additional particle is a bottom quark. Even though the excess did not reappear in the 2016 dataset[52], indicating that it was just a statistical fluctuation, the calculation still demonstrates a useful technique for studying new resonances.

The cross-section for the process $pp \to S \to f\bar{f}$, where S is a new colour-less boson with couplings to f, is given by

$$\sigma(pp \to S \to f\bar{f}) = \frac{2J+1}{M\Gamma_S}\left[\sum_{\mathcal{P}} C_{\mathcal{P}\bar{\mathcal{P}}}\Gamma_{\mathcal{P}\bar{\mathcal{P}}}\right]\Gamma_{f\bar{f}}. \tag{6.1}$$

The sum is over all partons ($\mathcal{P} \in \{g, u, d, c, s, b, \gamma\}$), with $C_{\mathcal{P}\bar{\mathcal{P}}}$ corresponding to the dimensionless integral for the parton density of $\mathcal{P}$. Γ_S and $\Gamma_{X\bar{X}}$ correspond to the full and partial width of S. Since the new resonance is seen in the decay to two photons, the spin J can either be 0 or 2. For now, spin 0 is assumed.

Assuming that the new resonance in the diphoton channel is produced from b-quark annihilation, equation 6.1 becomes

$$\sigma(b\bar{b} \to S \to \gamma\gamma) = \frac{1}{M\Gamma_S}C_{b\bar{b}}\Gamma_{b\bar{b}}\Gamma_{\gamma\gamma}. \tag{6.2}$$

The left-hand side and Γ_S can be obtained from the ATLAS result as 10 fb^{-1} with $\Gamma_S/M = 0.06$. The PDF factor is given in reference [66] as 15.3. Plugging the values into equation 6.2, one obtains the equality

$$\frac{\Gamma_{b\bar{b}}}{M}\frac{\Gamma_{\gamma\gamma}}{M} = 2.8 \times 10^{-4}\frac{\Gamma_S}{M} = 1.8 \times 10^{-5}. \tag{6.3}$$

Furthermore $\Gamma_S = \Gamma_{b\bar{b}} + \Gamma_{\gamma\gamma}$ holds, allowing 6.3 to be solved for $\Gamma_{b\bar{b}}$. Feeding this back into equation 6.1, the cross-section for S decaying into $b\bar{b}$ is given by

$$\sigma(b\bar{b} \to S \to b\bar{b}) = 2.1 \text{ pb}. \tag{6.4}$$

A diBjet signal with a cross-section of 2.1 pb would be visible in the 3.2 fb^{-1} of data recorded at the time. It should be noted that this is the most optimistic scenario for the diBjet search, since it is possible (and likely) that the scalar couples to more particles.

6.2 Selection

The diBjet analysis is performed with 3.2 fb^{-1} of data gathered at the end of 2015 operations, excluding runs with the IBL detector in stand-by mode. The key to reaching masses below a TeV is to apply b-tagging algorithms in the trigger. The trigger required two b-jets passing the medium working point, one with $p_T > 175$ GeV and the second with $p_T > 60$ GeV. To be on the kinematic plateau of the trigger, the leading b-jet is required to have $p_T > 230$ GeV and the subleading b-jet to have $p_T > 90$ GeV after the full jet corrections are applied offline. Both jets are need to be in the tracker to make b-tagging possible, resulting in a restriction of $|\eta| < 2.4$. To suppress the t-channel multijet background, a $|y^*| < 0.6$ cut is applied.

The two leading b-jets are identified using the 70% working point of the standard ATLAS b-tagging algorithm. In a t$\bar{t}$ sample, it accepts 70% of b-jets. The efficiency of the Z' signal is summarized in figure 6.1. This working point was shown to give the optimal sensitivity to find a narrow resonance. Due to differences in track reconstruction and discriminating variables, the offline b-tagging is not correlated with the algorithm employed by the trigger. The efficiency of the trigger-level b-tagging with respect to the offline b-tagging algorithm, shown in figure 6.2, is measured as a function of p_T using dilepton $t\bar{t}$ events. For the jet kinematics in the analysis, the efficiency ranges from 87% ($p_T \approx 100$ GeV) to 60% ($p_T \approx 600$ GeV).

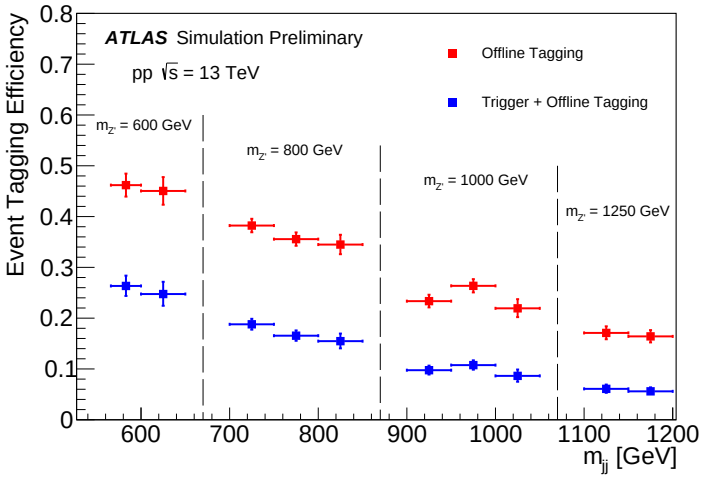


Figure 6.1: The offline (red) and trigger (blue) b-tagging efficiency for different masses of the Z' signal.

It is applied as an extra weight to the simulated events. This is on top of the scale factor applied to correct the offline b-tagging efficiency.

The background-dominated data is compared with multijet samples simulated by PYTHIA8. The QCD Monte Carlo is generated with inclusive jet-flavours, meaning that only a small fraction is the diBjet background. This results in large MC statistical uncertainty and prevents statistical tests on the simulated sample from giving sensible results. However that is not detrimental, since the whole procedure for the background estimation is data-driven. The comparisons only serve as a reference.

The p_T's of the leading and subleading jets are shown in figure 6.3. The m_{bb} and $|y^*|$ distributions are shown in figure 6.4. An example of the Z' signal with mass of 800 GeV is overlayed. The m_{bb} distribution has the region around the 750 GeV blinded. It is kept blinded during the initial sensitivity estimate and selection optimization to prevent a possible bias from the signal seen in the diphoton analysis.

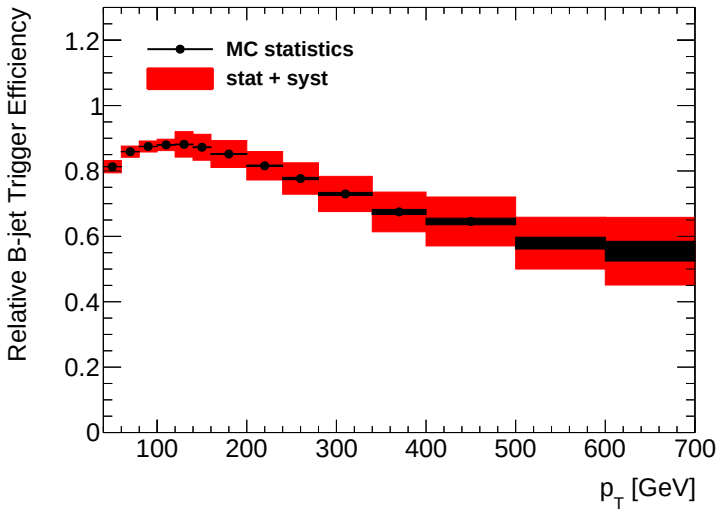


Figure 6.2: The efficiency of the medium working point of the b-jet trigger for jets passing the 70% WP of the offline b-tagging algorithm.

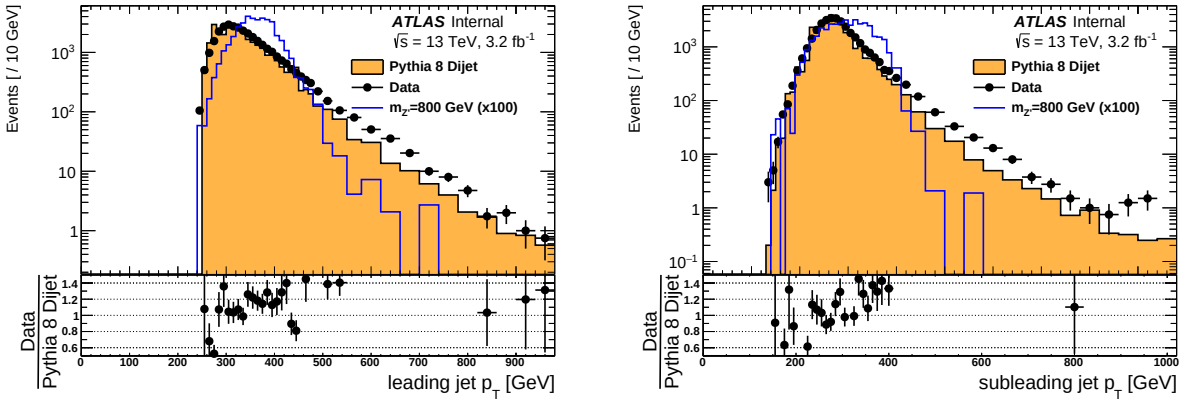


Figure 6.3: The leading (left) and subleading (right) jet p_T spectrum after the diBjet selection for data (black), simulated PYTHIA8 8 multijet (yellow) and Z' events (blue).

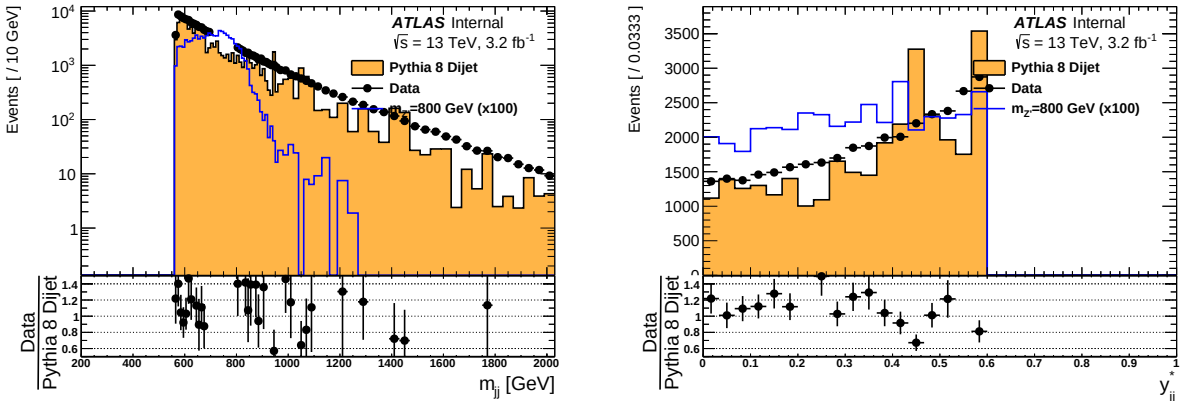


Figure 6.4: The m_{bb} (left) and $|y^*|$ (right) distributions after the diBjet selection for data (black), simulated PYTHIA8 8 multijet (yellow) and Z' events (blue).

6.3 Results

The simulated background samples did not have enough statistics to allow for sensible output from the statistical code. Regardless the whole background model is data-driven. The likelihood ratio test on the 2015 data concluded that adding parameters beyond three does not change the fit to the background. Figure 6.5 shows the result of running BUMPHUNTER. The most significant excess is between 600 GeV and 700 GeV, however it comes with a high p-value of 0.7.

Due to the lack of a signal, the limit setting procedure outlined in section 5.4 is used to set 95% CL exclusions on the Z' and generic Gaussian signals. The b-tagging introduced additional systematic uncertainties that have to be considered in addition to the ones already mentioned. The effect of the offline b-tagging algorithm is evaluated using the method described in [50][32]. It resulted in the largest systematic uncertainty of approximately 30%. An additional term is included for the extrapolation to high p_T jets. It is derived by studying the behaviour of the input variables in simulation. Several sources, including statistical, purity and modelling, are considered for the uncertainty on the trigger efficiency. This uncertainty is found to be p_T dependant and range from 10% to 30%.

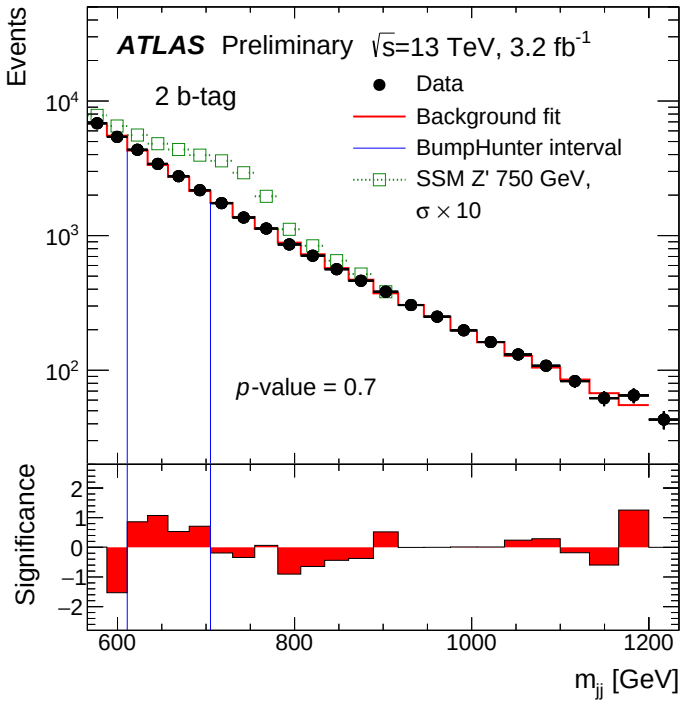


Figure 6.5: The result of the BumpHunter algorithm applied to the diBjet data.

The observed limits are shown in figure 6.6. Around mass of 750 GeV, a resonance with $\sigma \times \mathcal{A} \times \epsilon$ of 0.2 pb is excluded at 95% CL. Taking values of $\epsilon = 0.15$ and $\mathcal{A} = 0.7$ from the Z' signal, this translates to a limit on the cross-section of 1.9 pb. Thus the CL touches the prediction for the optimistic model based on the 750 GeV signal seen in the diphoton invariant mass spectrum.

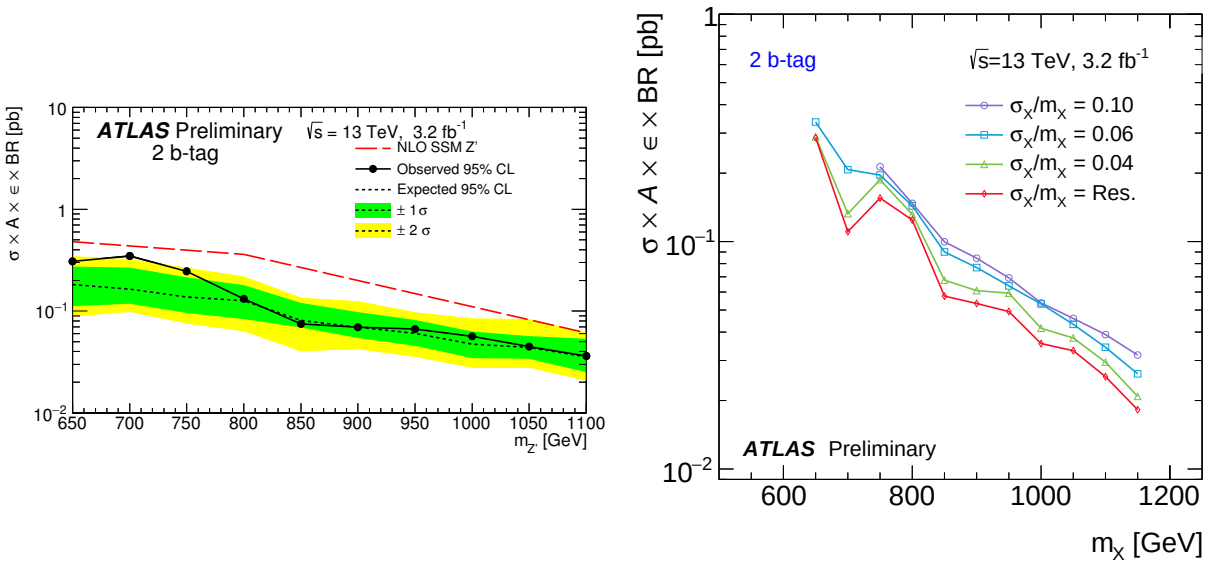


Figure 6.6: Limits on $\sigma \times \mathcal{A} \times \epsilon$ for the diBjet analysis. The 95% CL exclusions for the Z' (left) and generic Gaussian (right) signals are shown.

CHAPTER 7

DIJET RESONANCE WITH INITIAL STATE RADIATION

The searches for hadronic resonances considered so far involve b-tagging. This is advantageous if the target new particle couples preferentially to heavy quarks. However that is not always the case. Even the SSM Z' considered in diBjet search (chapter 6) decays into bottom quarks only a sixth of the time. This chapter looks for new resonances in an inclusive dijet final state, produced with additional initial state radiation[46]. It is complimentary to the high mass dijet search[47] and the dijet Trigger Level Analysis (TLA)[45]. Together they use a combination of techniques to reach resonant masses of a few hundred GeVs. The high mass dijet analysis uses the unprescaled single jet trigger to set limits above 1 TeV. The Trigger Level Analysis forgoes the readout of the entire detector and only saves the physics objects reconstructed by the trigger. The reduction in bandwidth allows it to record unprescaled events with leading jet of only 200 GeV and thus search for new particles with mass of 400 GeV or above. Finally the dijet+ISR class of searches uses initial state ration to trigger the event, allowing the search to go as low as 200 GeV at the cost of signal acceptance. The initial state radiation can either be a photon (dijet+γ channel) or a jet (dijet+jet channel).

7.1 Signal Model

There are several general motivations to search for new resonances in the dijet spectrum. The simplest one being if a new particle is produced at a hadron collider it should also decay to hadrons. However a good model is always helpful, as it guides the optimization of the selection and identifies sensitive regions of the theoretical phase-space. For this reason, the simplified model recommended by the Dark Matter Forum[1] has been adopted as the benchmark model. It is the simplest possible extension of the Standard Model to include dark matter, be valid at the LHC energies and allow the experimental program to perform

fairly generic searches. The simplified model can always be reinterpreted in the context of a more complete theory.

The model used in the dijet-like searches is best summarized by the Feynman diagram in figure 7.1. It adds two new particles to the Standard Model, corresponding to dark matter and a mediator (Z'). The presence of dark matter is driven by cosmological measurements including rotational curves of galaxies[87] and the Cosmic Microwave Background[62]. The mediator is inserted to connect the dark sector to the existing particles and make the model sensible at the LHC. It replaces the effective field theory (EFT) approach taken in Run 1[38][36][37][31][71], where the dark matter coupled to the Standard Model particles via an effective vertex. Only axial-vector couplings to quarks are considered to evade the spin dependent constraints from the direct detection experiments[2]. Furthermore any possible couplings to leptons are ignored, since there are strong constraints from leptonic resonance searches.

There are four new parameters in the simplified model. They are the masses of the dark matter (m_χ) and the mediator (m_R), and the couplings of the mediator to dark matter (g_χ) and quarks (g_q). The value of g_q is assumed to be flavour independent. A good review of existing constraints can be found in [10]. The existing limits come from the thermal relic dark matter density, direct detection and collider missing E_T + X searches. In order for the simplified model to not overproduce dark matter, the mediator needs to be relatively light. This makes the ME_T+X searches worthwhile, since the collider center of mass energy puts a ceiling on the heaviest mediators that can be produced. However such searches require light dark matter, putting the on-shell decay of the mediator for a sizable cross-section. The direct detection experiments set limits independent of m_R, but require a large local density of dark matter. This also translates to light dark matter. The qualitative shape of the exclusion mainly depends on the masses. Anything left of the line $m_R = 2m_\chi$ is excluded. Also mediator masses above a TeV tend to overproduce dark matter. Changing the couplings

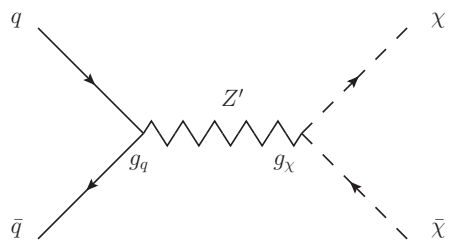


Figure 7.1: The Feynman diagram summarizing the contents of the simplified dark matter model. It extends the Standard Model by adding a single dark matter candidate (χ) of mass m_χ and a mediator with the existing particles (Z') of mass m_R. The mediator has democratic couplings to the quarks of g_q and a coupling to χ of g_χ.

only changes the size of the excluded regions and not their behaviour.

The presence of the mediator in the simplified model allows the ATLAS dark matter program to be expanded with a new class of searches for the mediator itself. This is where the dijet-like searches come in. The mediator can decay to quarks, meaning it should show up as a resonance in the dijet invariant mass spectrum. An advantage of this approach is that the results are independent of the dark matter properties. If no signal is seen, the corresponding limits will eliminate the areas left open by heavy dark matter. But it needs to be noted that if a signal is seen, there is no guarantee that it is the dark matter mediator.

When combined, the TLA and high-mass dijet analyses set limits on mediators with masses of 500 GeV and above. The dijet+ISR search documented here extends the limits down to masses of 200 GeV. It does so by looking for the mediator produced in association with initial state radiation (ISR), as summarized in figure 7.2. The ISR object is then used to pass the trigger, avoiding the limitations of the other dijet searches. Both initial state radiation of a jet (gluon at tree-level) and photon are considered. The downside is that requiring an extra high energy particle reduces the signal acceptance.

The signal samples are produced at tree-level, including the initial state radiation and decay of the Z', using MadGraph5 [3] and showered by Pythia8 [91] using the A14 tune[28].

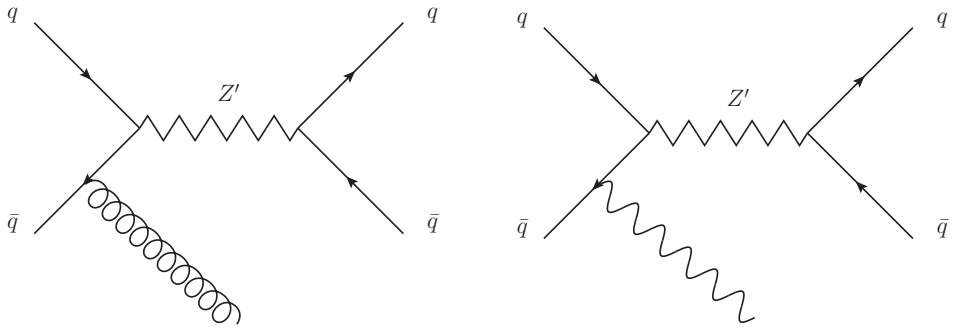


Figure 7.2: The Feynman diagrams of the signal in dijet+ISR searches. Both gluon (left) and photon (jet) are considered for the initial state radiation.

The NNPDF3.0 LO PDF set[4] is applied. Several samples are produced for varying values of m_R and g_q. The mass of dark matter is set high (10 TeV) to not affect the physics of the mediator.

7.2 Background Samples

Simulated background samples are used for selection optimization and validating the analysis code. A multijet Pythia8 [91] sample with the A14[28] tune is used for the dijet+jet background. The decay of charm and bottom quarks is performed using the EvtGen program[77]. Due to the need for large statistics, no detector simulation is performed. Instead the jet finding algorithm is run on stable truth particles not including muons or neutrinos. The parton distribution function is provided by the NNPDF3.0 LO set[4].

The dijet+γ background sample consists of γ+jet events generated using Shepra 2.1.1[70]. Up to 4 additional partons are considered at tree-level and merged by the Shepra parton shower[88]. The CT10 NLO PDF set[76] is used. Full detector simulation is performed.

7.3 Selection

The generic single jet and photon triggers are used to source the events for the two channels. The unprescaled thresholds are 380 GeV for the jet trigger and 140 GeV for photons passing the loose selection criteria. After full calibration, this translates to offline p_T cuts of 430 GeV and 150 GeV for the leading jet and leading photon correspondingly. The photon is also required to be isolated, pass the tight identification criteria and not be located in LAr calorimeter crack (defined using the raw cluster position as $1.37 < |\eta_{\text{cl}}| < 2.37$). All jets considered are required to be within $|\eta| < 2.8$. Due to negligible differences in the operating conditions, the data recorded in 2015 and 2016 are combined to provide 15.5 fb^{-1} of total luminosity.

Further selection requires the definition of the resonance jets used to calculate the invariant mass and the initial state radiation. In the dijet+γ channel, jets overlapping within $\Delta R < 0.2$ of the leading photon are removed and the remaining two leading jets are taken to be produced by the Z'. The situation is more complicated in the dijet+jet channel, where there are three jets to choose from. In the rest frame of the Z', the energy of the resonance jets will be $m_R/2$. If m_R is only a few hundred GeV (the target), the system needs a big boost from the ISR jet to pass the 430 GeV threshold. Thus the simplest heuristic is to assume that the leading jet comes from the initial state radiation and the two subleading jets form the resonance. Figure 7.4 shows how often this assumption is correct. Around $m_R = 600$ GeV, this only works 20% of the time, limiting the upper reach of the dijet+jet channel. For both analyses, the resonance jets are required to be above 25 GeV.

The remaining selection is used to optimize the signal sensitivity. The effect of several variables on $s/\sqrt{b}$ are studied, with the most discriminating ones shown in figure 7.5. By far the largest increase in sensitivity comes from cutting on y^* (< 0.6 for dijet+jet and < 0.8 for dijet+γ) between the two resonance jets. The optimal value is fairly independent of the

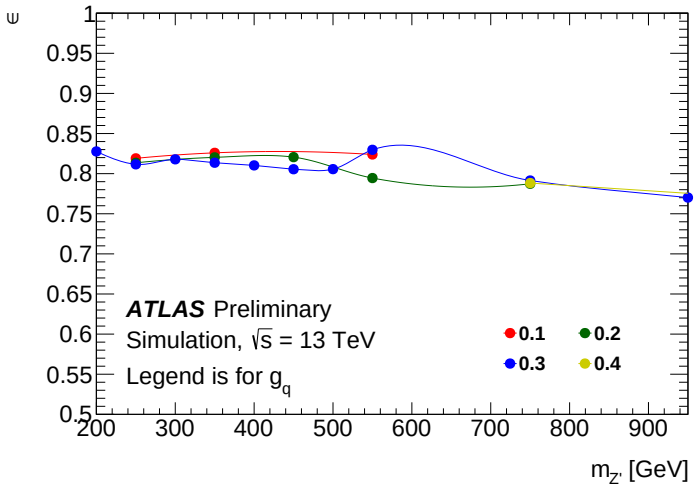


Figure 7.3: The efficiency of photon identification (tight ID and isolation) for different signal masses.

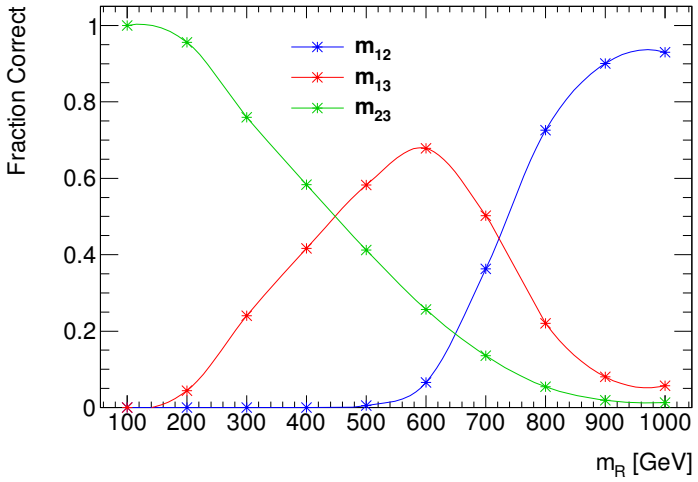


Figure 7.4: The fraction of time different jet combinatorics correctly form the Z' in the dijet+jet selection.

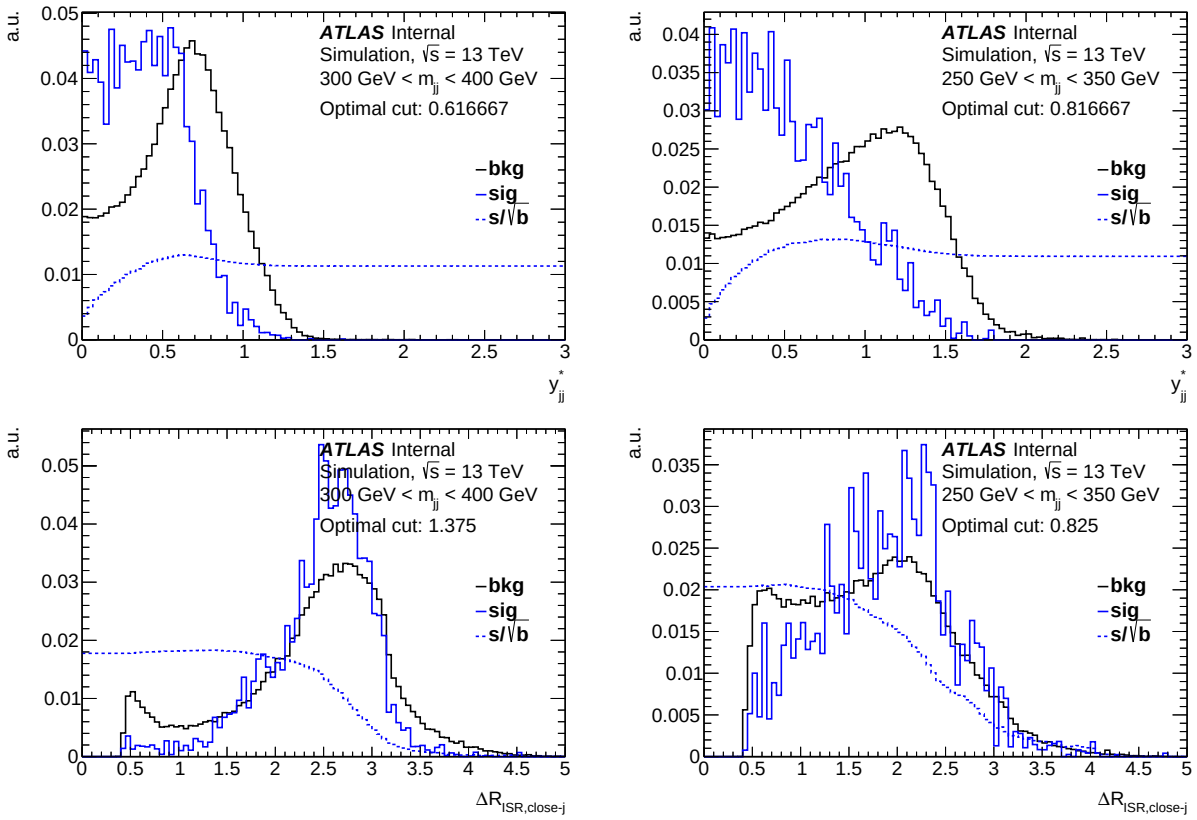


Figure 7.5: The optimization of the dijet+jet with $m_R = 300$ GeV (left) and dijet+γ with $m_R = 300$ GeV (right). The top figure shows y^* of the resonance jets with respect to the kinematic selection. The bottom figure shows $\Delta R_{\text{ISR,close jet}}$ after the application of the y^* cut.

resonance mass. The dijet+γ selection also removes events where the photon is close to a resonance jet by requiring $\Delta R_{\gamma,\text{close jet}} > 0.85$. The complete list of cuts is summarized in table 7.1.

Comparison between the data and simulated events is done for key distributions. A scale factor is applied to the simulated events. It is derived by matching the normalization in the the leading jet ϕ distribution. The dijet+jet distributions, with an extra scale factor of 0.82 for the background, are shown in figure 7.6. The dijet+γ distributions, with an extra scale factor of 1.5 for the background, are shown in figure 7.7. There are no major deviations that could indicate problems with the detector.

Selection	dijet+jet	dijet+γ
Standard	GRL, event cleaning	
ISR object	Leading jet $p_T > 430$ GeV	Leading photon $p_T > 150$ GeV
Resonance jets	Subleading and third jet $p_T > 25$ GeV	Leading and subleading jet $p_T > 25$ GeV
y^*	< 0.6	< 0.8
$\Delta R_{\gamma,\text{close jet}}$	–	> 0.85
Fit Range	(303 GeV, 611 GeV)	(169 GeV,1493 GeV)

Table 7.1: The dijet+jet and dijet+γ analysis selection.

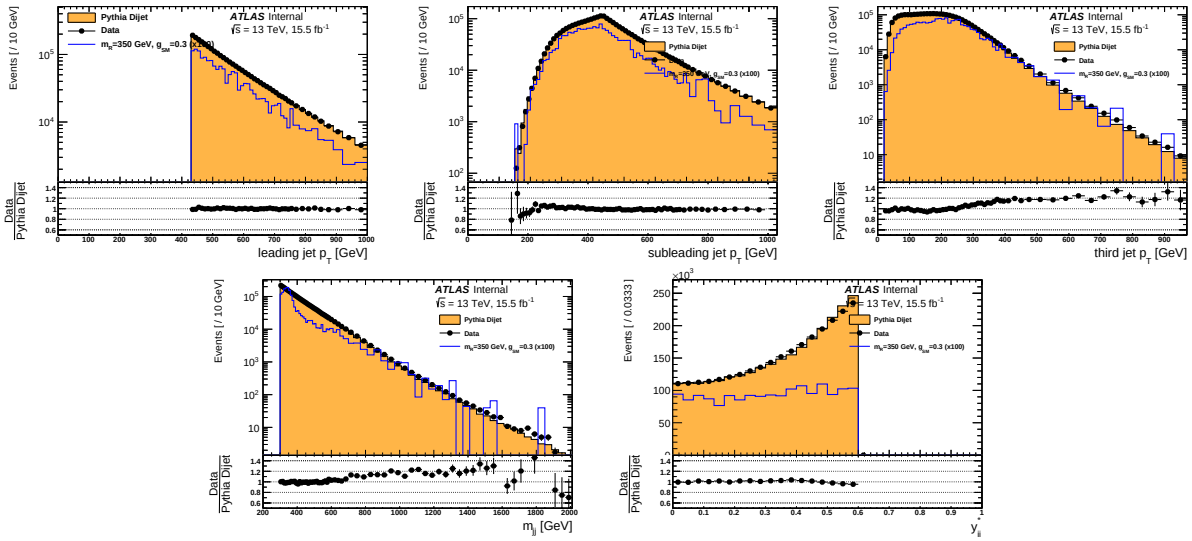


Figure 7.6: Comparison of key distributions between the data (black), simulated PYTHIA8 8 multijet (yellow) and Z' events (blue) in the dijet+jet analysis.

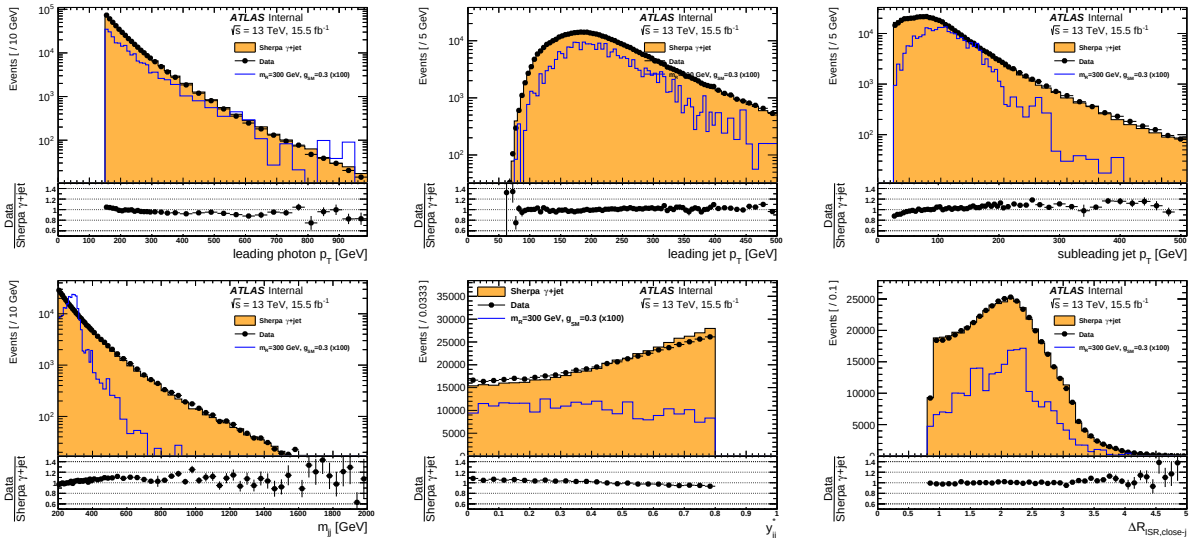


Figure 7.7: Comparison of key distributions between the data (black), simulated PYTHIA8 8 multijet (yellow) and Z' events (blue) in the dijet+γ analysis.

7.4 Results

The search in dijet+jet channel is performed for m_{jj} ranging from 303 GeV to 611 GeV. The lower range is limited by requiring a smooth distribution. The upper range is limited by the choice of the second and the third jet to form the invariant mass no longer being correct. The likelihood ratio test, shown in figure 7.8, indicates that enabling parameters past three in the background function makes no significant difference. Figure 7.9 shows the result of using BUMPHUNTER. It indicates that the largest deviation from the background model is around 500 GeV, but with a large p-value of 0.6.

The dijet+γ channel has no restriction on the highest resonance mass it can probe. However it is more difficult to fit a larger range with a generic smooth function. Thus the m_{jj} range is limited from 169 GeV to 1493 GeV. This gives enough overlap with the TLA analysis to confirm any possible signal. Figure 7.8 shows that the likelihood ratio test prefers the four parameter function to model the background. The result of BUMPHUNTER is shown in figure 7.9. The most significant deviation is at $m_{jj} = 900$ GeV and also has a large p-value

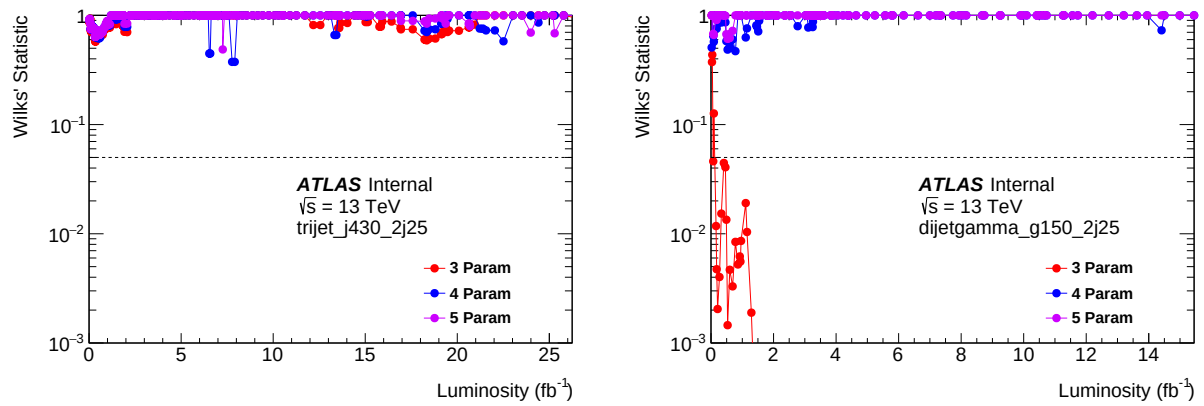


Figure 7.8: The result of the likelihood ratio test on the dijet+jet (left) and dijet+γ (right) data. The likelihood ratio for the function with n parameters (legend) is calculated with respect to the function $n+1$ parameters.

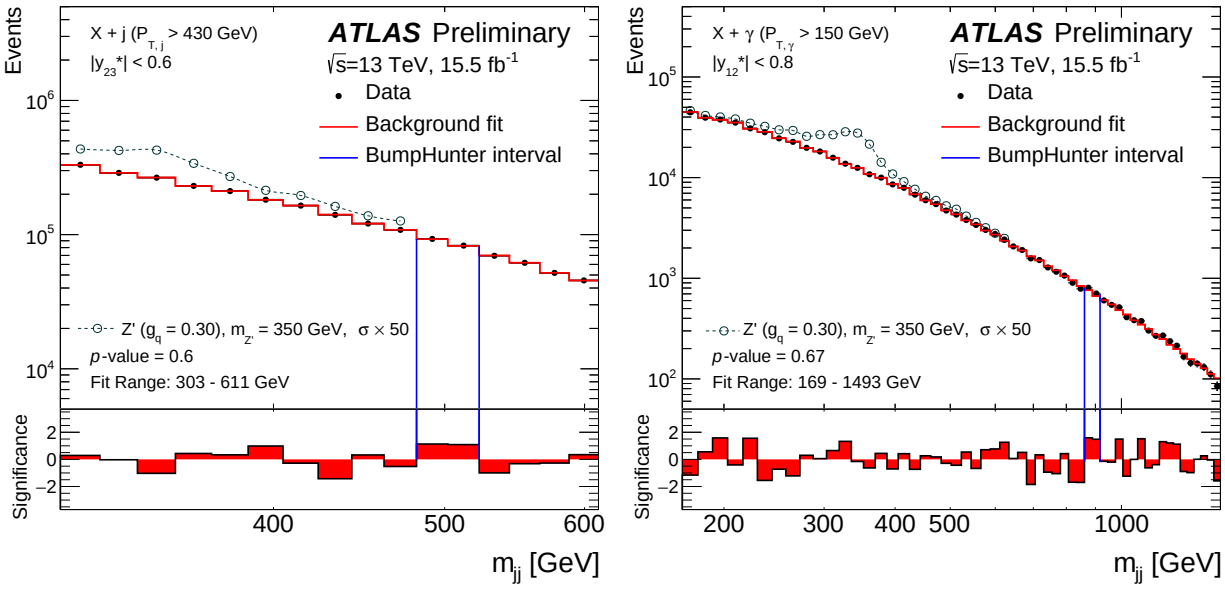


Figure 7.9: The result of the BumpHunter algorithm applied to the dijet+jet (left) and dijet+γ (right) data.

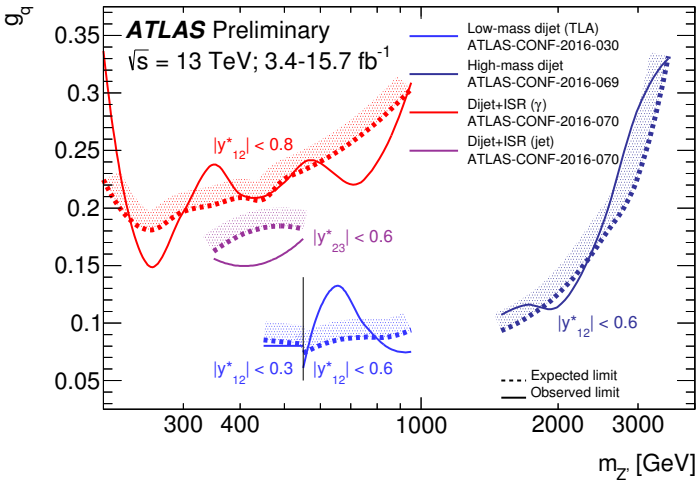


Figure 7.10: The limits on the g_q coupling due to the dijet+γ (red), dijet+jet (purple), TLA dijet (light blue) and high-mass dijet (dark blue) analyses.

of 0.67.

Due to the lack of a signal, the limit setting procedure is performed on both datasets. To convert the limits on $\sigma \times \mathcal{A}$ to limits on the coupling g_q, the event yield predicted by the simulation is scaled by g_q^2. The results are compared with the other dijet-like searches in figure 7.10. The limits from the dijet+ISR searches are weaker than from the dijet searches, however they are only analyses able to access mediator masses below 400 GeV. Both eliminate coupling values around 0.2, with the dijet+jet analysis setting slightly stronger constraints.

Figure 7.11 shows the same limits recast on the $m_\chi - m_R$ plane. Dedicate samples with varying m_χ values are generated at truth-level and reinterpreted using the Gaussian limits.

The limits on $\sigma \times \mathcal{A}$ for a generic Gaussian signal are shown in figure 7.12. The efficiency due to the photon identification, shown in figure 7.3, is corrected for by comparing the event yield between the full selection with detector simulation and kinematic selection at truth-level. The dijet+jet and dijet+γ analysis eliminate any new resonances with $\sigma \times \mathcal{A}$ around

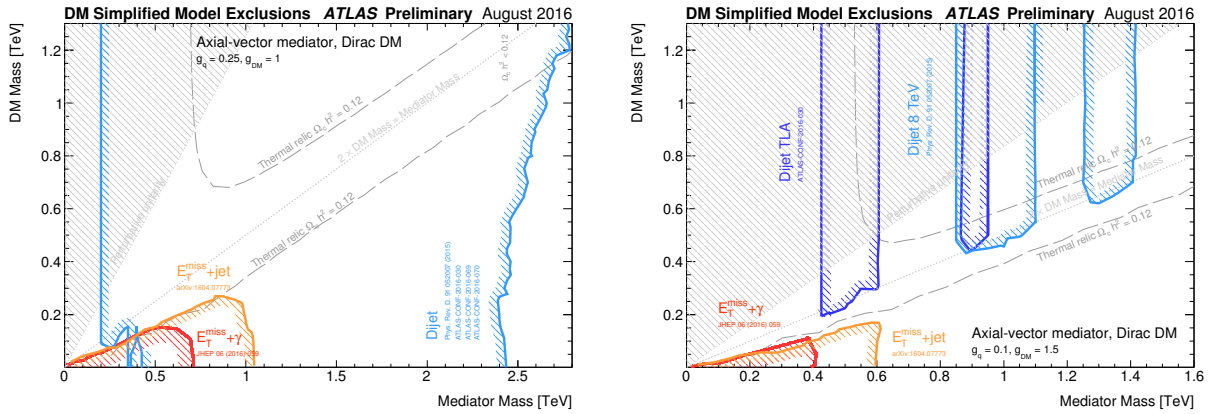


Figure 7.11: The limits from the MET+jet (yellow), MET+γ (red) and dijet-like (blue) searches for the simplified model as function of m_χ and m_R for different couplings values of g_q and g_χ. The contour where the dark mater relic density is reproduced by the simplified model is shown as the dashed line.

a pb and 0.1 pb correspondingly.

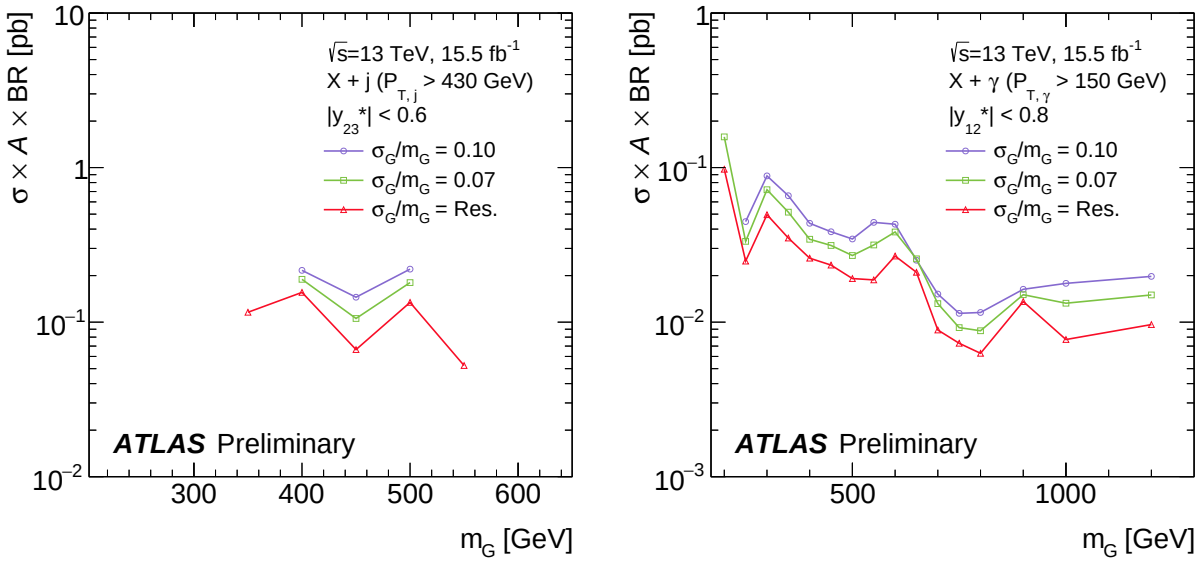


Figure 7.12: The limits on $\sigma \times \mathcal{A}$ for generic Gaussian signals using the dijet+jet (left) and dijet+γ (right) data.

CHAPTER 8

CONCLUSION

Hadron collider experiments are a great way to search for new physics at the subatomic scale. They provide a controlled environment where any physics that can happen will happen and can be detected. The only limitation is the available energy, which puts an upper bound on the heaviest particles that can directly be studied. This results in a drive to build colliders with more energetic beams. The latest machine that realizes this is the Large Hadron Collider. In 2011 it started operating at 7 TeV center-of-mass energy, beating the TeVatron energy by a factor of 3.5. In 2015, another factor of two was gained with record collision energies of 13 TeV. This allowed limits to be set at a whole new energy scale. But this came at a cost. The higher energies result in a busier environment, implying that the treshholds triggering the recording of collision events have to be raised. This leaves a large and unexplored space at lower energies.

For example, the a new Higgs-like particle has been discovered. Yet there is no measurement of its most dominant decay mode into two bottom quarks. The Higgs boson is relatively light ($m_H = 125$ GeV), but the Large Hadron Collider operates at $\mathcal{O}\,(\text{TeV})$ energies. This makes it nearly impossible to even record the events containing the boson. Another example is the search for dark matter, where viable models contain a light mediating particle.

This thesis explored different techniques that can be used to bypass the issues related to new sub-TeV physics. The solution in most cases is to produce new particles in association with others. This makes the final state richer, allowing to define interesting collisions happening a recordable rate.

The first analysis considered is the search for the $H \to b\bar{b}$ decay, where the Higgs boson is produced using vector boson fusion. The VBF process results in the presence of two additional jets with a certain topology. This is enough auxiliary information to reduce the

background and make the analysis viable. Using collisions recorded at 8 TeV, the ATLAS experiment is able to constrain the VBF $H \to b\bar{b}$ signal strength to

$$\mu_{H\to b\bar{b}} < 4.7. \tag{8.1}$$

Approximately 300 fb^{-1} of data, available by the end of Run 2, will be required to make an observation in this channel.

The next analysis considered is a sub-TeV search for diBjet resonances. If a heavy Higgs exists, it should have a preferential decay to heavy quarks. By using b-tagging information in the trigger, the ATLAS experiment is able to set limits of

$$\sigma \times \mathrm{BR}(X \to b\bar{b}) <\approx 2 \text{ pb} \quad \text{with} \quad 600 \text{ GeV} < m_X < 1.2 \text{ TeV}. \tag{8.2}$$

The final analysis considered is the search for a generic dijet resonance. A simple motivation for such a final state is that if a new particle is produced from parton collisions, then it should also decay to them. A more physics-based motivation comes from the simplified dark matter models, where a new particle mediates between the Standard Model and the dark sector. The measurements of the cosmological constant indicate that if such a new particle exists, then it should be between 100 GeV and 1 TeV. However the standard dijet search is limited to $m_R > 1$ TeV by the trigger. This can be bypassed by producing the mediator in association with initial state radiation and triggering on that. Using a photon or a jet as the ISR object, the ATLAS experiment is able to set new limits on mediators with masses below 500 GeV. This is an important achievement, as the only collaborations that studied this region previously are the long-defunct SPS and TeVatron experiments.

REFERENCES

[1] Jalal Abdallah et al. Simplified Models for Dark Matter Searches at the LHC. pages 9–10, March 2015.

[2] D S Akerib, S Alsum, H M Araújo, X Bai, A J Bailey, J Balajthy, P Beltrame, E P Bernard, A Bernstein, T P Biesiadzinski, E M Boulton, R Bramante, P Brás, D Byram, S B Cahn, M C Carmona-Benitez, C Chan, A A Chiller, C Chiller, A Currie, J E Cutter, T J R Davison, A Dobi, J E Y Dobson, E Druszkiewicz, B N Edwards, C H Faham, S Fiorucci, R J Gaitskell, V M Gehman, C Ghag, K R Gibson, M G D Gilchriese, C R Hall, M Hanhardt, S J Haselschwardt, S A Hertel, D P Hogan, M Horn, D Q Huang, C M Ignarra, M Ihm, R G Jacobsen, W Ji, K Kamdin, K Kazkaz, D Khaitan, R Knoche, N A Larsen, C Lee, B G Lenardo, K T Lesko, A Lindote, M I Lopes, A Manalaysay, R L Mannino, M F Marzioni, D N McKinsey, D M Mei, J Mock, M Moongweluwan, J A Morad, A St J Murphy, C Nehrkorn, H N Nelson, F Neves, K O'Sullivan, K C Oliver-Mallory, K J Palladino, E K Pease, P Phelps, L Reichhart, C Rhyne, S Shaw, T A Shutt, C Silva, M Solmaz, V N Solovov, P Sorensen, S Stephenson, T J Sumner, M Szydagis, D J Taylor, W C Taylor, B P Tennyson, P A Terman, D R Tiedt, W H To, M Tripathi, L Tvrznikova, S Uvarov, J R Verbus, R C Webb, J T White, T J Whitis, M S Witherell, F L H Wolfs, J Xu, K Yazdani, S K Young, C Zhang, et al. Results from a search for dark matter in LUX with 332 live days of exposure. Technical Report arXiv:1608.07648, Aug 2016. Comments: A combined analysis of these results with previous LUX results is forthcoming, and will be submitted to PRL.

[3] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. page JHEP, July 2014.

[4] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[5] Andy Buckley, James Ferrando, Stephen Lloyd, Karl Nordstrom, Ben Page, Martin Ruefenacht, Marek Schoenherr, and Graeme Watt. LHAPDF6: parton density access in the LHC precision era, March 2015.

[6] A. D. Bukin. Fitting function for asymmetric peaks, December 2007.

[7] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.

[8] John M. Campbell, R. Keith Ellis, and Giulia Zanderighi. Next-to-Leading order Higgs + 2 jet production via gluon fusion. *JHEP*, 10:028, 2006.

[9] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and Antimatter in the Universe. *New J. Phys.*, 14:095012, 2012.

[10] Mikael Chala, Felix Kahlhoefer, Matthew McCullough, Germano Nardini, and Kai Schmidt-Hoberg. Constraining Dark Sectors with Monojets and Dijets. *JHEP*, 1507:089, June 2015.

[11] Georgios Choudalakis. On hypothesis testing, trials factor, hypertests and the BumpHunter, January 2011.

[12] Mariano Ciccolini, Ansgar Denner, and Stefan Dittmaier. Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC. *Phys. Rev.*, D77:013002, 2008.

[13] Mariano Ciccolini, Ansgar Denner, and Stefan Dittmaier. Strong and electroweak {NLO} corrections to Higgs-boson production in vector-boson fusion at the {LHC} . *Nuclear Physics B - Proceedings Supplements*, 183:103–108, 2008. Proceedings of the 9th {DESY} Workshop on Elementary Particle Theory.

[14] F. E. Close. *Antimatter.* OUP Oxford, 2009.

[15] ATLAS Collaboration. *ATLAS liquid-argon calorimeter: Technical Design Report.* Technical Design Report ATLAS. CERN, Geneva, 1996.

[16] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, 2008.

[17] ATLAS Collaboration. ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.

[18] ATLAS Collaboration. Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev. D*, 83:052005, Mar 2011.

[19] ATLAS Collaboration. New ATLAS event generator tunes to 2010 data. Technical Report ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr 2011.

[20] ATLAS Collaboration. Search for New Physics in Dijet Mass and Angular Distributions in pp Collisions at $\sqrt{s} = 7$ TeV Measured with the ATLAS Detector. *New J. Phys.*, 13:053044, March 2011.

[21] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012.

[22] ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett. B*, 726:120, July 2013.

[23] ATLAS Collaboration. Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C*, 73:2518, July 2013.

[24] ATLAS Collaboration. Jet energy scale and its systematic uncertainty in proton-proton collisions at sqrt(s)=7 TeV with ATLAS 2011 data. Technical Report ATLAS-CONF-2013-004, CERN, Geneva, Jan 2013.

[25] ATLAS Collaboration. Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Phys. Lett. B*, 726:88, October 2013.

[26] ATLAS Collaboration. Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System. Technical Report CERN-LHCC-2013-018. ATLAS-TDR-023, Sep 2013. Final version presented to December 2013 LHCC.

[27] ATLAS Collaboration. A neural network clustering algorithm for the ATLAS silicon pixel detector. *Journal of Instrumentation*, 9(09):P09009, 2014.

[28] ATLAS Collaboration. ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014.

[29] ATLAS Collaboration. Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data. Technical Report ATLAS-CONF-2014-046, CERN, Geneva, Jul 2014.

[30] ATLAS Collaboration. Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J.*, C74(8):3023, 2014.

[31] ATLAS Collaboration. Search for new particles in events with one lepton and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2014(9):37, 2014.

[32] ATLAS Collaboration. Commissioning of the ATLAS b-tagging algorithms using $t\bar{t}$ events in early Run-2 data. Technical Report ATL-PHYS-PUB-2015-039, CERN, Geneva, Aug 2015.

[33] ATLAS Collaboration. Expected performance of the ATLAS b-tagging algorithms in Run-2. Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015.

[34] ATLAS Collaboration. Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva, Jul 2015.

[35] ATLAS Collaboration. Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV. Technical Report ATLAS-CONF-2015-002, CERN, Geneva, Mar 2015.

[36] ATLAS Collaboration. Search for dark matter in events with heavy quarks and missing transverse momentum in $$pp$$ p p collisions with the ATLAS detector. *The European Physical Journal C*, 75(2):92, 2015.

[37] ATLAS Collaboration. Search for new phenomena in events with a photon and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. D*, 91:012008, Jan 2015.

[38] ATLAS Collaboration. Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. page Eur. Phys. J. C, September 2015.

[39] ATLAS Collaboration. Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector. *JHEP*, 01:069, 2015.

[40] ATLAS Collaboration. Discrimination of Light Quark and Gluon Jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector. Technical Report ATLAS-CONF-2016-034, CERN, Geneva, Jul 2016.

[41] ATLAS Collaboration. Measurement of jet charge in dijet events from sqrt(s)=8 TeV pp collisions with the ATLAS detector. *Phys. Rev. D*, 93:052003, June 2016.

[42] ATLAS Collaboration. Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data. *Eur. Phys. J.*, C76(12):666, 2016.

[43] ATLAS Collaboration. Performance of the ATLAS Trigger System in 2015. 2016.

[44] ATLAS Collaboration. Photon identification in 2015 ATLAS data. Technical Report ATL-PHYS-PUB-2016-014, CERN, Geneva, Aug 2016.

[45] ATLAS Collaboration. Search for light dijet resonances with the ATLAS detector using a Trigger-Level Analysis in LHC pp collisions at $\sqrt{s} = 13$ TeV. Technical Report ATLAS-CONF-2016-030, CERN, Geneva, Jun 2016.

[46] 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. Technical Report ATLAS-CONF-2016-070, CERN, Geneva, Aug 2016.

[47] ATLAS Collaboration. Search for New Phenomena in Dijet Events with the ATLAS Detector at $\sqrt{s}$=13 TeV with 2015 and 2016 data. Technical Report ATLAS-CONF-2016-069, CERN, Geneva, Aug 2016.

[48] ATLAS Collaboration. Search for new phenomena in dijet mass and angular distributions from pp collisions at with the {ATLAS} detector . *Physics Letters B*, 754:302–322, 2016.

[49] ATLAS Collaboration. Search for resonances below 1.2 TeV from the mass distribution of b-jet pairs in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector. Technical Report ATLAS-CONF-2016-031, CERN, Geneva, Jun 2016.

[50] ATLAS Collaboration. Search for resonances in diphoton events at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2016(9):1, 2016.

[51] ATLAS Collaboration. Search for resonances in the mass distribution of jet pairs with one or two jets identified as b-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Letters B*, 759:229–246, June 2016.

[52] ATLAS Collaboration. Search for scalar diphoton resonances with 15.4 fb^{-1} of data collected at $\sqrt{s}$=13 TeV in 2015 and 2016 with the ATLAS detector. Technical Report ATLAS-CONF-2016-059, CERN, Geneva, Aug 2016.

[53] ATLAS Collaboration. Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 05:160, 2016.

[54] ATLAS Collaboration. Search for the Standard Model Higgs boson produced by vector-boson fusion in 8 TeV pp collisions and decaying to bottom quarks with the ATLAS detector. 2016.

[55] CDF Collaboration. Search for New Particles Decaying to $b\overline{b}$ in $p\overline{p}$ Collisions at $\sqrt{s} =$ 1.8TeV. *Phys. Rev. Lett.*, 82:2038–2043, Mar 1999.

[56] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, 2008.

[57] CMS Collaboration. Performance of quark/gluon discrimination in 8 TeV pp data. 2013.

[58] CMS Collaboration. Search for Heavy Resonances Decaying into bb and bg Final States in pp Collisions at sqrt(s) = 8 TeV. Technical Report CMS-PAS-EXO-12-023, CERN, Geneva, 2013.

[59] CMS Collaboration. Observation of the diphoton decay of the Higgs boson and measurement of its properties. *Eur. Phys. J. C*, 74:3076, October 2014.

[60] CMS Collaboration. Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks. *Phys. Rev. D*, 89:012003, February 2014.

[61] CMS Collaboration. Search for the standard model Higgs boson produced through vector boson fusion and decaying to b bbar. *Phys. Rev. D*, 92:032008, September 2015.

[62] Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. *Astronomy & Astrophysics*, 594:A13, 2015. cite arxiv:1502.01589Comment: Abstract severely abridged.

[63] SNO Collaboration. Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by 8B solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87(nucl-ex/0106015. UPR-0240-E):071301. 6 p, Jun 2001.

[64] Daniel de Florian, Giancarlo Ferrera, Massimiliano Grazzini, and Damiano Tommasini. Higgs boson production at the LHC: transverse momentum resummation effects in the H$\to \gamma\gamma$, H $\to$ WW $\to$ lνlν and H $\to$ ZZ $\to$ 4l decay modes. *Journal of High Energy Physics*, 2012(6):132, 2012.

[65] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, Oct 1993.

[66] Roberto Franceschini, Gian F. Giudice, Jernej F. Kamenik, Matthew McCullough, Alex Pomarol, Riccardo Rattazzi, Michele Redi, Francesco Riva, Alessandro Strumia, and Riccardo Torre. What is the gamma gamma resonance at 750 GeV?, February 2016.

[67] J. H. Friedman. Data Analysis Techniques for High Energy Particle Physics. Technical report, 1974.

[68] Jason Gallicchio and Matthew D. Schwartz. Quark and Gluon Tagging at the LHC. *Phys.Rev.Lett.*, 107:172001, October 2011.

[69] Gian Francesco Giudice. Naturally Speaking: The Naturalness Criterion and Physics at the LHC. 2008.

[70] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 0902:007, November 2009.

[71] Jessica Goodman, Masahiro Ibe, Arvind Rajaraman, William Shepherd, Tim M. P. Tait, and Hai-Bo Yu. Constraints on Dark Matter from Colliders. *Phys.Rev.D*, 82:116010, April 2010.

[72] Massimiliano Grazzini and Hayk Sargsyan. Heavy-quark mass effects in Higgs boson production at the LHC. *Journal of High Energy Physics*, 2013(9):129, 2013.

[73] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[74] Vardan Khachatryan et al. Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method. *Eur. Phys. J.*, C75(6):251, 2015.

[75] David Krohn, Tongyan Lin, Matthew D. Schwartz, and Wouter J. Waalewijn. Jet Charge at the LHC. *Phys. Rev. Lett.*, 110:212001, June 2013.

[76] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys.Rev.D*, 82:074024, August 2010.

[77] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[78] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy correlation functions for jet substructure. *Journal of High Energy Physics*, 2013(6):1, 2013.

[79] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN-2013-004*, CERN, Geneva, 2013.

[80] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56. *Astrophys.J.*, 606:819–824, January 2004.

[81] STEPHEN MYERS. THE LARGE HADRON COLLIDER 2008–2013. *International Journal of Modern Physics A*, 28(25):1330035, 2013.

[82] S. H. Neddermeyer and C. D. Anderson. Note on the Nature of Cosmic Ray Particles. *Phys. Rev.*, 51:884–886, 1937.

[83] G. P. S. Occhialini and C. F. Powell. Nuclear disintegration produced by slow charged particles of small mass. 1987. [Nature159,186(1947)].

[84] G. Paál, I. Horváth, and B. Lukács. Inflation and compactification from Galaxy redshifts? *Astrophysics and Space Science*, 191(1):107–124, 1992.

[85] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.

[86] Jon Pumplin. How to tell quark jets from gluon jets. *Phys. Rev.*, D44:2025–2032, 1991.

[87] V. C. Rubin, N. Thonnard, and W. K. Ford, Jr. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *Astrophys. J.*, 238:471, 1980.

[88] Steffen Schumann and Frank Krauss. A Parton shower algorithm based on Catani-Seymour dipole factorisation. *JHEP*, 03:038, 2008.

[89] M Shochet, L Tompkins, V Cavaliere, P Giannetti, A Annovi, G Volpi, et al. Fast TracKer (FTK) Technical Design Report. Technical Report CERN-LHCC-2013-007. ATLAS-TDR-021, Jun 2013. ATLAS Fast Tracker Technical Design Report.

[90] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[91] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A Brief Introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178:852–867, October 2008.

[92] Marek Tasevsky. Differences between Quark and Gluon jets as seen at LEP, October 2001.

[93] Wouter Verkerke and David P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003. [,186(2003)].

[94] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pages 270–279, 1998.