Dissertation zur Erlangung des Doktorgrades der Fakultät für
Mathematik und Physik der Albert-Ludwigs-Universität Freiburg
im Breisgau

Search for squarks and gluinos in final state with jets, missing transverse momentum and boosted W bosons with the ATLAS experiment

von Tomáš Javůrek

Betreut von Prof. Dr. Gregor Herten

August 2016

Dekan:	Prof. Dr. Dietmar Kröner
Prodekan:	Prof. Dr. Gregor Herten
Referent:	Prof. Dr. Gregor Herten
Koreferent:	Dr. Stephanie Ulrike Zimmermann
Prüfer (Experiment):	Prof. Dr. Karl Jakobs
Prüfer (Theorie):	Prof. Dr. Stefan Dittmaier
Datum der mündlichen Prüfung:	20.6.2016

Erklärung

Hiermit erkläre ich die vorliegende Arbeit selbständig und nur mit den angegebenen Quellen und Hilfsmitteln verfasst zu haben. Einige Ergebnisse dieser Arbeit wurden in den untenstehenden Publikationen bereits veröffentlicht.

Freiburg, den 17.03.2016
Tomas Javurek

Acknowledgements

Firstly, I would like to thank to Zuzana Rurikova, who has been very patient with me as supervisor as well as colleague and friend. I would like to thank to my wife, colleague in the computing, partner in sport and life Martina Pagacova-Javurkova for cooperation, discussion, patience and great care. Thanks belong also to colleagues at University of Freiburg, primarily to Prof. Gregor Herten and Andrea Di Simone, for discussions, supervision, corrections of the thesis and also for providing me the best environment to study, teach and work as a scientist. The work presented here has been performed in collaboration with the ATLAS SUSY 0-lepton 2-6 jet analysis group. I would like to thank to all colleagues of this group, especially to Renaud Bruneliere, Marija Vranjes Miliosavljevic and Nikola Makovec. Great thanks go to the entire ATLAS Distributed Computing team, but primarily to Jaroslava Schovancova and Alessandro Di Girolamo, for the opportunity to join the team and for supervising me in my first projects in the computing area. Let me thank also to my family. This work would never be finalized without their care and love.

Abstract

At the time of this writing, Supersymmetry is the most discussed extension of the Standard Model. A significant part of the supersymmetric parameter space can be explored with accelerators like the Large Hadron Colider (LHC) at CERN, which provides proton collisions for the ATLAS detector. If Supersymmetry is realized at the weak scale, one of the most promising production channels at the LHC is the direct production of squark and gluino pairs. The search for such events with no lepton, large missing transverse momentum and large transverse momentum of the jets is presented in this thesis. Experimental limits on cross-sections of squark and gluino pair production are interpreted in Simplified models with various assumptions on the masses of these particles.

Reconstruction of specific particles, like W-bosons or τ-leptons, can increase the sensitivity of the analysis in two ways: as a veto in case the desired particle is not in the final state of the signal model, or as a requirement in the opposite case. A study on the W-boson tagging and its usage in both these cases, that lead to published results, is described in this thesis. Alternative approaches to the W-boson reconstruction, such as using the substructure of jets, are discussed as well. Attention is also given to the alternative fit methods of the peak in the invariant mass of the W.

Zusammenfassung

Supersymmetrie ist eine der meist diskutierten Erweiterungen des Standardmodells der Teilchenphysik. Ein großer Bereich des Parameterraums supersymmetrischer Modelle kann an großen Teilchenbeschleunigern, wie dem Large Hadron Collider (LHC) am CERN untersucht werden, der Proton-Proton Kollisionen für den ATLAS Detektor zur Verfügung stellt.

Wenn Supersymmetrie auf schwachen Skalen existiert, ist einer der wichtigsten Produktionskanäle am LHC die direkte Produktion von Squark und Gluino Paaren. In dieser Arbeit wird die Suche nach solchen Ereignissen ohne Leptonen, großem Transversalimpuls der Jets und großer fehlender transversaler Energie präsentiert. Experimentelle Ausschlussgrenzen bezüglich der Wirkungsquerschnitte von Squark und Gluino Produktion werden gesetzt und in vereinfachten Modellen interpretiert, die verschiedene Annahmen hinsichtlich der Massen der supersymmetrischen Teilchen machen.

Die Rekonstruktion spezifischer Teilchenarten, wie W Bosonen oder τ Leptonen kann die Sensitivität einer Analyse in zweierlei Hinsicht erhöhen: Als ein Selektionskriterium, für den Fall dass das relevante Teilchen im Endzustand des Signals zu erwarten ist oder als ein Veto für den umgekehrten Fall. Diese Arbeit präsentiert eine Methode zur Identifikation von W Bosonen, die in beiden beschriebenen Fällen angewandt wurde und in publizierten Resultaten der ATLAS Kollaboration verwendet wird. Alternative Verfahren der W Boson Rekonstruktion, die die Substruktur der Jets ausnutzen, werden ebenfalls vorgestellt. Des Weiteren werden verschiedene Methoden zur Beschreibung der invarianten Masse des W Bosons beschrieben.

Table of contents

1. Introduction

During the past decades, Particle Physics made historical discoveries while at the same time revealing new challenges. The building blocks of Nature can be understood more and more precisely with the ever advancing technology, and vice versa new technologies are developed for the purpose of achieving a deeper understanding of the Nature. This doctoral study is released in such an exciting time presenting the results of the measurement at one of the largest experiment ever.

Although predictions of the Standard Model, introduced in Chapter 5, match with the experimental data with excellent precision, there are still strong motivations for its extensions. The dark matter observed from the mass density distribution in our galaxy remains unexplained, the theoretical issues with the bare mass of the Higgs in higher orders of perturbation theory are difficult to avoid within the Standard Model. The extension of the Standard Model with Supersymmetry has the potential to answer fully or partially these questions. The proton-proton collisions with sufficiently large center of mass energy of the colliding pair could produce the new particles predicted by such an additional symmetry. In the wide range of analyses aiming to discover or disproof some of the broad spectra of Supersymmetric extensions of the Standard Model, the channel with no lepton(s) in the final state and large energy of the final jets is one of the best candidate for an observation of such new particles at the weak scale, as the strong production of sparticles dominates in the widest range of relevant parametric space.

The contribution to the 0-lepton analysis with the ATLAS detector, aiming to discover squarks or gluinos, is reported in this thesis. It consists of background suppression studies and of increasing the discovery potential with the introduction of boosted boson tagging. In addition, a study of shape fit of the W invariant mass peak and a study of the substructure of W's jet or di-jet system, are discussed.

The ATLAS computing grid was used as a source of the computing power for this analysis. The contribution to the field of ATLAS Distributed Computing is also shortly summarized in Chapter 4.

This Ph.D. thesis is organized as follows. The theoretical introduction and motivation is summarized in the Chapter 5. The experimental set-up is described in Chapters 2 and 3. The analysis part of the thesis starts with an overview of the 0-lepton analysis, see Chapter 6. The definition of the objects used in the analysis and the selection of events are presented in Chapter 7. The results of the analysis are presented in Chapters 8, 9 and 10.

2. The Large Hadron Collider

2.1 CERN and the LHC

CERN (Conseil Européen pour la Recherche Nucléaire) and its powerful accelerators such as the LHC (Large Hadron Collider) take upon themselves the greatest scientific challenges of the present time. There are however also great successes in the past scientific program of CERN. To mention few of them: the W/Z-boson discoveries, the World Wide Web, and most recently, observations of the Higgs boson and penta-quark. The European Strategy for Particle Physics is in favour of continuing of all the LHC experiments at least until 2030, see Fig. 2.1 with the planned periods of data taking (Runs) and shutdowns. The strategy has been agreed by the CERN council. Very similar aims are claimed by the Particle Physics Project Prioritization Panel (P5) [1].

The LHC, the largest single machine ever built, accelerates proton bunches in opposite directions in two separated beam pipes. It is situated approximately 100 meters below the ground in the valley of the Geneva lake, together with four large detectors surrounding the four points at LHC where the protons can collide. Apart from these detectors (ATLAS, CMS, ALICE, LHCb, see Fig. 2.2), the LHC hosts also three smaller detectors: TOTEM, MoEDAL and LHCf. ATLAS and CMS are multi-purpose detectors with two complementary general aims: the validation of the Standard Model (SM) and the search for new physics at the accessible energy scale. The objectives of LHCb and ALICE are more specific, namely the search for the origins of the matter-antimatter inequality and the study of the quark-gluon plasma, respectively. The three small experiments are built to measure the total cross section, elastic scattering and diffractive processes, to observe magnetic monopoles and to explore ultra-high-energy cosmic rays.

The proton-proton data produced by the LHC and measured with the ATLAS detector are studied in this thesis. The protons are firstly accelerated from the proton source by a linear accelerator and injected to the proton synchrotron booster, where they gain 1.4 GeV. The next step in acceleration is Proton Synchrotron (PS), where the protons gain an energy of 25 GeV. In the Super Proton Synchrotron (SPS), the energy is increased to 450 GeV, and the protons

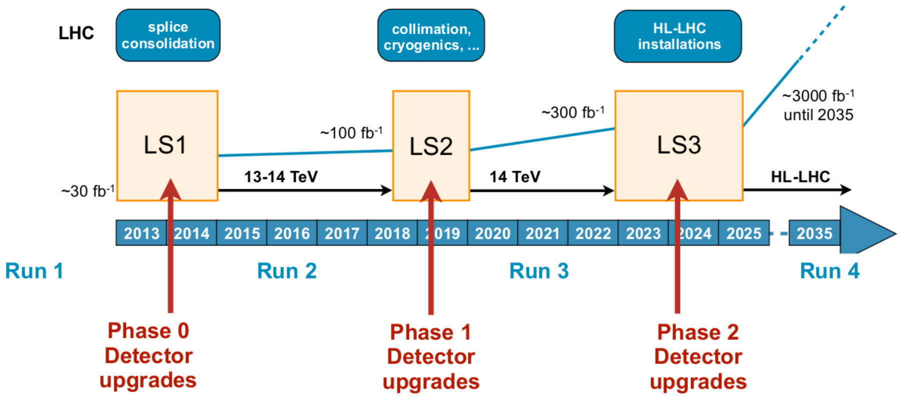


Fig. 2.1 The LHC roadmap, the plan for next decades. [2]

are finally injected to the LHC. The LHC is designed to accelerate protons to 7 TeV, which corresponds to center of mass energy of the collisions of $\sqrt{s}=$14 TeV. The key components of the LHC to achieve such energies are the radiofrequency cavities to accelerate the protons and super-conducting dipole and quadrupole magnets, creating magnetic field with strength up to $\sim$8.3 Tesla, that steer the proton beams. The protons are accelerated in bunches, separated by a time of $\tau=$50 ns (25 ns designed).

However, the designed centre of mass energy was not achieved during Run I. The data analysed in this thesis were produced by the LHC in a short phase of data-taking at $\sqrt{s}=$7 TeV, which was interrupted by an unexpected technical incident, and in the following phase at $\sqrt{s}=$8 TeV.

The rate of a given process with a production cross section σ_{proc} is given by:

$$\frac{dN}{dt}=L_0\sigma_{\text{proc}} \tag{2.1}$$

where L_0 is the instantaneous luminosity, that can be expressed exclusively in terms of beams optics and bunch properties. The total number of events can be derived from integrated luminosity defined as:

$$L_{\text{int}}=\int L_0 dt \tag{2.2}$$

During the first LHC run, and integrated luminosity of 29fb^{-1} were produced, out of which 23fb^{-1} were recorded by the ATLAS experiment and successfully analysed.

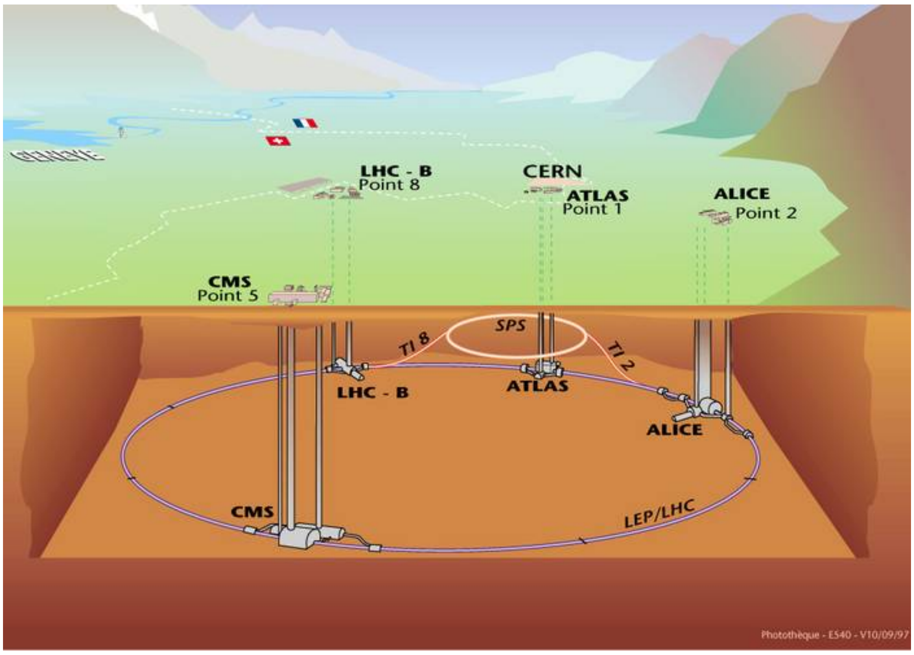


Fig. 2.2 The LHC and its experiments. The tube is placed in a tunnel roughly 100 meters bellow the ground. Source: [3]

3. The ATLAS Detector

The collisions of the protons or ions at the LHC can produce particles reaching energies larger than 1 TeV. In order to detect such energetic particles, a large detector around the collision point has to be built. The ATLAS detector has an architecture (see Fig. 3.1) designed to detect and reconstruct particles in a wide range of energies, from a few GeV to a few TeV. It consists of an Inner Detector, described in Section 3.1, which measures the tracks of charged particles, Electromagnetic and Hadronic Calorimeters, described in Sections 3.2.1 and 3.2.2, designed to measure energy deposits of the showers originating from photons, electrons and hadrons, and the Muon Spectrometer, described in Section 3.3, designed to reconstruct muons. Neutrinos can not be captured by the ATLAS detector, thus leave the detector undetected and contribute to the transverse missing momentum, see Section 7.3. The ATLAS detector, with its roughly 7000 tons of mass, is not the heaviest of the LHC detectors, but with its length of 44 m and 22 m in diameter, it is the largest LHC detector by volume. The Inner Detector and Muon Spectrometer are immersed in a magnetic field generated by super-conducting coils with a strength reaching up to 2 Tesla, 3.5 Tesla for end-cap of Muon Spectrometer.

The coordinate system is by convention set as follows: it is a right-handed coordinate system with the x-axis pointing to the LHC center and the z-axis coinciding with the beam axis. The y-axis is pointing in an upward direction. The polar angle θ is measured with respect to the z-axis. The azimuthal angle ϕ and radial distance R are measured in the xy-plane.

The pseudo-rapidity η is frequently used in place of θ:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{3.1}$$

The separation (angular distance) of two objects is usually expressed by:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \tag{3.2}$$

where $\Delta\phi$ and $\Delta\eta$ are the differences in ϕ and η for these two objects respectively.

The design of all the sub-detectors is briefly described in this chapter. The major upgrades

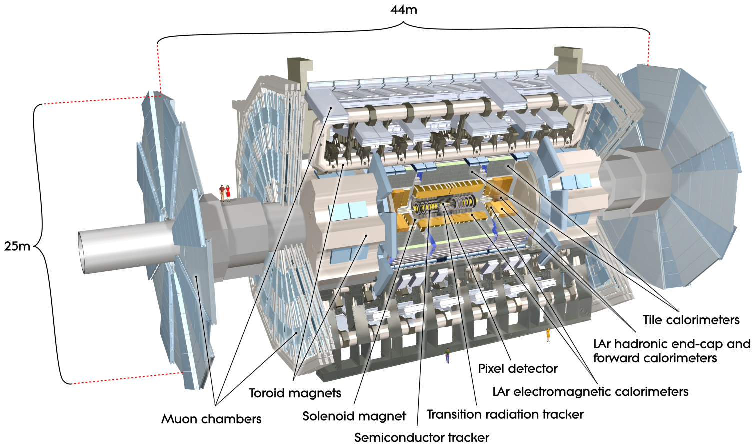


Fig. 3.1 The architecture of the ATLAS detector. Source: [4]

during the LHC shut-down in 2013-2014 are presented at the end of the chapter.

3.1 Inner Detector

The Inner Detector (see [6], page 53), sometimes also called Inner Tracker, is built with the intention of reconstructing the tracks of charged particles using silicon detectors and small gas-filled tubes. This can be achieved only with sufficient granularity of the detector. All of the sub-detectors of the inner detector consist of a barrel surrounding the collision point and two end-caps covering larger η ranges. The detectors in the end-caps are fitted into wheel-shaped disks to ensure that particles with large η hit the surfaces of these detectors almost perpendicularly. The overall geometry is depicted in Fig. 3.2.

1) The innermost part of the Inner Detector is the Pixel Detector. It is arranged into three layers, which are further divided into modules. Each module carries silicon sensors segmented into pixels of size $50{\times}400\ \mu m^2$. The surface of the inner layer is placed at a radial distance of 5.05 cm from the interaction point (see Section 3.4), which provides fine resolution for determination of secondary vertices. About 80 millions of readout channels are involved to read the response of the Pixel Detector, which have to be able to read a signal every 25 ns. The innermost wheel of the end-caps is equipped with a Pixel Detector as well. The spatial resolution of the pixel positioned in the barrel is 100 μm in the z-direction and 12 μm in the xy-plane[1]. The impact parameter resolution for high-p_{T} tracks varies with η, since differently positioned pixels are hit along the track for different η (see [6], page 57).

---

[1] i.e. in the direction of the short pixel pitch

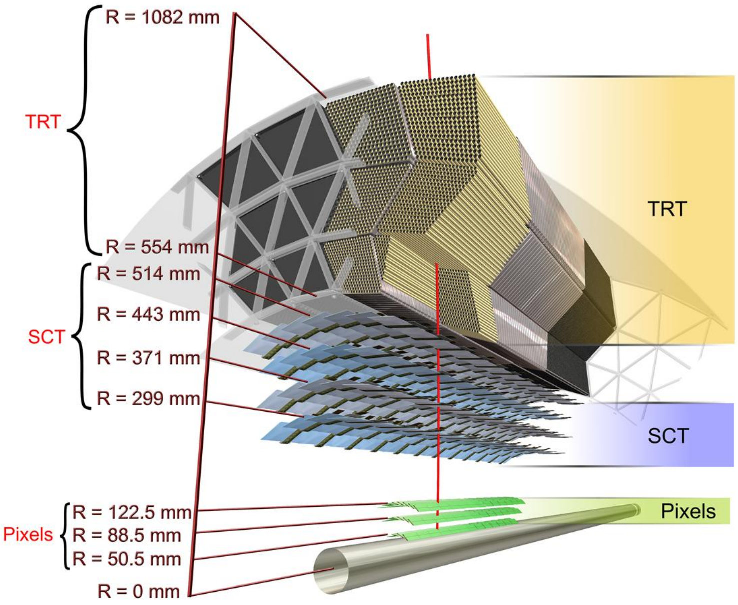


Fig. 3.2 The layers of Inner Detectors. Source: [5]

2) The next-to-innermost part of the Inner Detector is the SCT (Semi-Conductor Tracker). Its building block is a single-sided silicon strip module with an area of 6.36 × 6.40 cm^2, which is joined to a second strip module making a pair with doubled active surface. Such detector pair is glued with another pair under a stereo angle of 40 mrad to make a double sided module. The stereo angle is applied in order to measure both coordinates R and ϕ of the hit when the particle passes through the module. Readout channels are placed on the strip on the top of the module. In total, there are 4088 modules with a coverage of 63 m^2 of active surface. The barrel of the SCT is assembled in four layers of such modules. The end-caps on both sides of the detector consist of 9 wheel disks each. Such a construction of the SCT provides a spatial resolution per module in the barrel of 17 μm in the xy-plane and 580 μm in the z-direction.

3) The outermost part of the Inner Detector is the TRT (Transition Radiation Tracker), constructed from straw detectors placed parallel to the beam in the barrel section, and radially in the end-cap section. The straws have a diameter of 4 mm and a gold-plated W-Re wire is positioned in the center. They are filled with an active medium, the radiator, which is composed at 70% by Xe. The barrel of the TRT carries 52544 of such straws and there is 122880

in each of the end-caps. Such granularity is providing approximately 36 hits for a charged particle passing through the TRT. Each of the straws also performs a drift time measurement providing 130 μm of spatial resolution in the xy-plane. The relatively low spatial resolution of a straw is however compensated by the large number of hits.

3.2 Calorimeter

The Calorimeter of the ATLAS detector, see Fig. 3.3 ([6], page 143), is designed to measure energy deposits form both charged and neutral particles. The technology used and the composition of the Calorimeter allows to distinguish between hadrons, electrons and photons. The Calorimeter can be divided into two parts according to the technology used: LAr (Liquid Argon) named according to its active medium, and TileCal (Tile Calorimeter) built from scintilators (tiles). The Calorimeter can also be divided slightly differently, according to the particles that a given part is designed to measure, namely Electromagnetic Calorimeter (ECAL), Hadronic Calorimeter (HCAL), and Forward Calorimeter (FCAL). The Calorimeter description in this thesis is divided into two sections 3.2.1 and 3.2.2 according to the technology used in a given sub-detector. The energy resolution is however specific for a given type of particle:

$$\frac{\sigma_{\mathrm{E}}}{E} \sim \frac{9\%}{\sqrt{E}} \oplus 0.3\% \quad \text{(ECAL)} \tag{3.3}$$

$$\frac{\sigma_{\mathrm{E}}}{E} \sim \frac{50\%}{\sqrt{E}} \oplus 3\% \quad \text{(HCAL)} \tag{3.4}$$

$$\frac{\sigma_{\mathrm{E}}}{E} \sim \frac{100\%}{\sqrt{E}} \oplus 10\% \quad \text{(FCAL)} \tag{3.5}$$

3.2.1 LAr

The barrels of the LAr consist of copper-tungsten (for the Hadronic Calorimeter in the end-caps) or lead (for EMCAL) absorbers in accordion shape, which are immersed into an active medium (liquid argon). The LAr covers with its two half-barrels and two end-caps the region of η up to 3.2 with small gaps in $|\eta| \sim 0$ and $|\eta| \sim (1.37 - 1.52)$. The LAr barrels are organized in three layers and their granularity decreases with the distance from the beam pipe, since most of the electromagnetic showers is expected to have most of its depositions in the inner parts. Its absorbers are interleaved with layers of electrodes to detect the signal produced in the active medium.

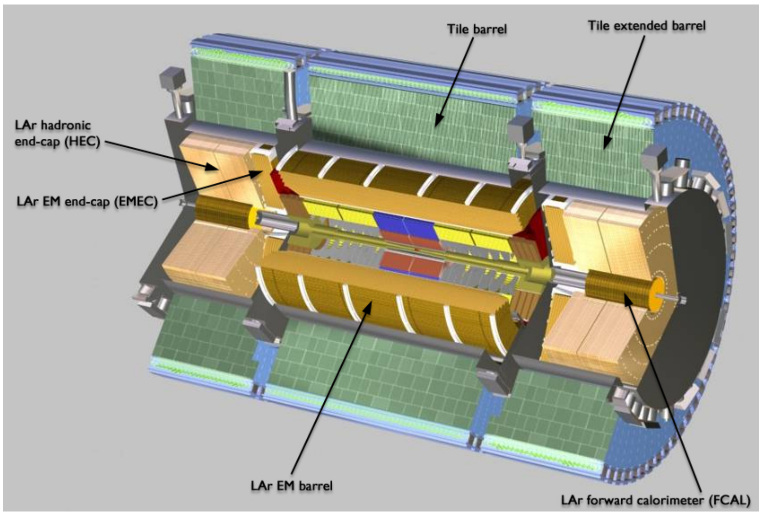


Fig. 3.3 The Inner Detector depicted by layers. Source: [7]

[2]

Cryogenics

The cryostats of the LAr are filled with liquid argon cooled to 87 K. The cooling system is designed so that during operation and cooling down, the lowest possible temperature gradients are created, which prevents material stress damages. Special sensors are used to monitor the temperature fluctuations in the system. A very stable situation has been reached during the past years, with the average fluctuation being less than 70 mK for most of the detector parts.

Electronics

The front-end electronic of the LAr is designed to provide the analogue signal to the back-end electronics at the 75 kHz collection frequency. The amplitude of the signal depends on its strength, i.e. energy deposited in the tower of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, which allows the deposited energy measurement.
The signal is collected by 1524 front-end boards mounted on the cryostat. Each board consists of four sections: a calibration section, which injects a precisely known signal to the detector, a signal collector which collects, amplifies and shapes the signal, a tower builder which sums up the signals from the tower of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and sends the result through a 1.6 Gbits/s link to the L1 trigger (see Section 3.5) and finally a controller board which is responsible for

---

[2] The LAr detector additionally consists of two Hadronic end-caps placed behind Electromagnetic end-caps. Finally, the LAr is hosting also forward calorimeter (FCAL) built to measure the tracks close to beam pipe.

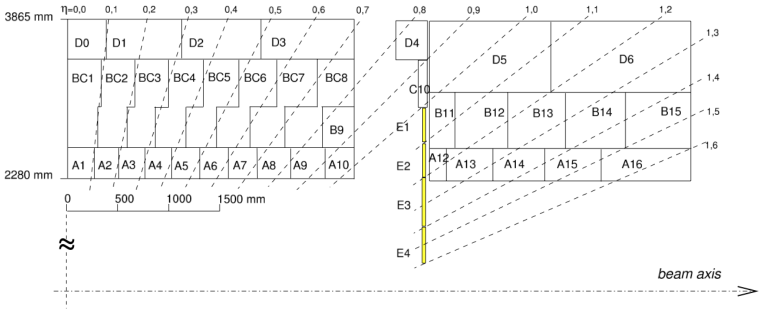


Fig. 3.4 The schematic cross-section of the TileCal. Source: [8]

monitoring and communication with the DAQ, as described in the TDAQ Section 3.5. The signal collector can amplify the signal with three gains, 1/10/100, depending on the initial signal strength.

3.2.2 Tile Calorimeter

The TileCal is the main part of the Hadronic Calorimeter. Although its main purpose is to measure energy deposits of hadrons (jets), muons can also be detected due to a broad spectrum of sensitivity of the TileCal ranging from $\sim$200 MeV to $\sim$1 TeV of energy deposit per cell. A good performance of the Hadronic Calorimeter is crucial for precise determination of the properties of the jets and consequently of the transverse missing momentum.

The TileCal consists of two Long Barrels (LB) with coverage of $|\eta| < 1.0$ and two External Barrels (EB) covering the range of $1.0 < |\eta| < 1.7$. The inner and the outer radius of EB and LB are 228 cm and 423 cm respectively. The barrels are further divided into 64 modules radially in order to cover $\Delta\phi \sim 0.1$ with each module. The modules consist of three layers of read-out units called cells, see Fig. 3.4. The cells differ by their proportions in order to achieve the optimal granularity at a given position in the TileCal. The cells closest to the beam vertex have approximate dimensions 25×30 cm (zy) and the largest (and outermost) cells have roughly 92×38 cm (zy). The cells have a periodical structure where a 3 mm thick scintilator layers, alternates with 14 mm thick steel piece, as a passive medium.

Optics and electronics

The photons emitted in the scintilators are captured by wavelength shifting optical fibres (WLS) attached to the edges of the scintilators and lead further to the photo-multipliers (PMT). The WLS transform the wavelength of the emitted light in order to achieve maximal efficiency of the primary photo-emission on the photo-cathode. The PMTs intensify the light signal and transform it into electric charge and consequently to an analogue pulse. Such a pulse continues to a so called 3-in-1 card, which consists of two parts:

1. The shaper, which amplifies and broadens the analogue signal. The signal is amplified with high gain (HG) or low gain (LG) according to the input signal strength (energy deposit in the cell)
2. The integrator, that integrates the signal used for calibration purposes only (see Section 3.2.2).

There are three output channels of the 3-in-1 card:

- fast signal without amplification directed to LVL1 trigger
- LG and HG signal directed to the digitizer
- output dedicated to calibration and monitoring purposes

The analogue signal is digitized in the Tile Data Management Unit (TileDMU). The second task of the TileDMU is to receive the information from the TTC system (Time, Trigger, Control), to sample the signal with seven intervals of 25ns and synchronize the signal with the bunch crossing. A digitized signal is stored in a pipeline memory waiting for the LVL1 trigger decision, whether it will be sent to further processing or removed.

The whole front-end electronics, as described in the previous paragraph, is placed in the super-drawers mounted on the top of each module. The high voltage needed for the front-end electronics operation is distributed with so-called HV dividers, which are also installed at each super-drawer. The rest of the signal processing is hosted in the back-end electronics placed outside of the detector.

Calibration and Monitoring

The calibration is needed to establish real energy deposits in the cells from the response of the TileCal. The whole system has to be monitored during the data taking to investigate its occasional problematic parts in order to take an appropriate action (e.g. to switch off a cell when it is hot or to exclude it form signal reprocessing). Four calibration methods used by the TileCal are described in following paragraphs.

The **Cesium calibration** is used to control and calibrate the optical system (Tiles, WLS fibres and PMTs). A hydraulic system is used to inject a γ source (137Cs) in tubes parallel to the beam pipe, that lead through small holes made into each tile and each steel plate. The emitted

γs have an energy of 662 keV, that corresponds to an absorption length in steel of $\sim$ 19 mm[3]. The response of each Tile individually can be investigated with this calibration method. In case that some inconsistency in the cell response is observed, it can be adjusted with changing the high voltage on the PMTs.

The PMTs are additionally calibrated with the **Laser calibration** system. A laser pulse with a wavelength of 532 nm and a width of 10 ns is injected into all PMTs. The PMT non-linearities and pulse saturation are monitored and the calibration constants can be derived. Furthermore, the injection of the pulse is synchronized with the TTC system, so that the time instabilities can be uncovered.

While the Cesium and Laser systems are applied deeper in the detector (tiles and PMT's respectively), the third calibration, **Charge Injection** System (CIS), is located at the 3-in-1 card. A well defined charge signal is injected to the convertor which digitizes the analogue signal. After such a pulse is integrated and reconstructed, a calibration factor can be extracted. This procedure is performed for each Tile DMU channel.

Minimum Bias events are the limitation factor of the TileCal's performance. However, their features can be used for calibration purposes. Minimum Bias events are expected to have stable and uniform performance in ϕ and η with small variations given by a different particle flux and amount of detector material in different directions or other unexpected non-uniformities, which needs to be monitored and calibrated.

3.3 Muon Spectrometer

The interactions of high-p_{T} muons in the Inner Detector and Calorimeter lead to relatively weak signals. The energy deposit of a muon passing through these detectors is of the order of few GeV for muons with p_{T} $\sim$20 GeV. Such deposits are not sufficient to capture the muon and therefore a dedicated sub-detector is needed with significant amount of material and strong magnetic field. Note that all the other SM particles (except for neutrinos) are in most of the cases fully absorbed by the Inner Detector or the Calorimeter. The Muon Spectrometer consists of two types of chambers: trigger chambers designed for a fast response and precision chambers with a slower response but a higher precision.

The precision part of the spectrometer consists of Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC). The MDT chambers are organized into three concentric layers in the barrel and small or big wheels in each end-cap of the detector. Each chamber consists of 3-8 layers of drift tubes with a maximum counting rate of up to 500 Hz/cm^2. The MDTs cover the range in η up to 2.7. The range 2.0<$|\eta|$<2.7 is covered by the CSC, multi-wire

---

[3] While absorption length in the scintilator is significantly larger.

proportional chambers with perpendicularly segmented cathode planes in order to provide position measurements. Its designed maximum counting rate is about 1000 Hz/cm^2. The CSC chambers are needed due to the larger particle flux expected in this η range. Both the MDT and CSC use a mixture of Ar and CO$_2$ as an active medium.

The trigger part includes two types of chambers: Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). The RPC cover a range of η up to 1.05 and it uses two parallel resistive plates with a gas mixture between them, that can be ionized by a muon. The TGCs are installed in the end-cap, covering the range $1.05<|\eta|<2.04$. Similarly to the CSC, it is a multi-wire proportional chamber with a smaller gap between the cathode and the wire than the distance between the wires themselves.

Both barrel and end-caps are inserted in a magnetic field induced by superconducting coils. The strength of the magnetic field reaches approximately 0.5 T in the barrel and 1 T in the end-caps.

The momentum resolution of the Muon Spectrometer ranges from 3% at $p_{\mathrm{T}} = 100$ GeV up to roughly 11% at $p_{\mathrm{T}} = 1$ TeV. This is achieved with an excellent precision when determining a hit position: $\sim 80\ \mu m$ for MDT and $\sim 50\ \mu m$ for CSC. The time resolution is the parameter of importance for the RPC and TGC (since they are designed for triggering) being 2 ns and 4 ns respectively.

3.4 Main Upgrades during the Shutdown

In order to cope with the larger impact of the pile-up during RunII, an upgrade of the Inner Tracker is needed to identify the tracks produced in the primary vertex with higher precision. An additional layer of pixel detectors at the inner surface of the barrel of the Inner Detector was installed for these purposes. The Insertable B-Layer (IBL) [9] consist of 280 silicon pixel modules positioned in average of 33 mm from the beam pipe center. The IBL's high granularity is given by the pixel small size of $50\times250\ \mu$m, which makes this layer the component of the ATLAS detector with the greatest intrinsic resolution. Apart from pile-up, this improvement increased the b-tagging efficiency, which has an impact on many analyses.

The beam pipe itself had to be adapted reducing the outer radius to 24.7 mm in order to provide enough space for the IBL. The movements of the ATLAS cavern determine the minimum radius of the beam pipe in the ATLAS. A maximum displacement of about 1 mm had been measured in 2003-2007, which is better than expected. Consequently, the beam pipe radius could be reduced by 4 mm. The position of the beam is controlled by a new Diamond Beam Monitor system.

In order to gain luminosity, the acceptance of the L1 trigger, introduced in next Section 3.5,

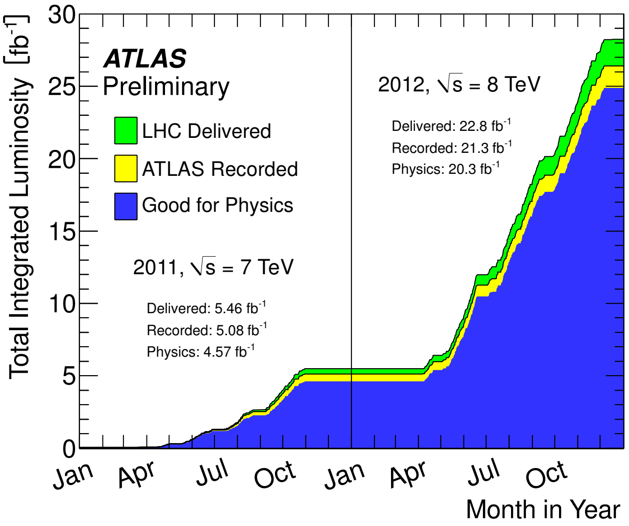


Fig. 3.5 Recorded luminosity of ATLAS during RunI. Source: [11]

has been enhanced from 70 kHz to 100 kHz. Such an improvement has been achieved by adding CPUs to the existing trigger hardware. The trigger menu had to be optimised to the new conditions with increased collision energy [10].

The rest of the upgrades during phase-0 is related to the detector consolidation needed due to the radiation damage during RunI operation. The new power supplies for the calorimeter had to be installed and some components of the power network exchanged. The magnet, cryogenics and cooling of the Inner detector have been upgraded. Wide repairs of the pixel opto-electronics had to be performed.

3.5 Trigger, Data Acquisition and Quality

ATLAS has recorded an integrated luminosity of 21.3 fb^{-1} during the 7/8 TeV runs in 2011 and 2012, see Fig. 3.5. Recording such a large amount of data requires a very efficient system which triggers only the interesting events (from physics point of view) and distributes the data quickly further along the processing chain, namely, to the storage elements in computing centres of the world wide ATLAS grid.

The overall design summary is depicted in Fig. 3.6. The events received by the ATLAS detectors proceed directly to the L1 trigger, which aims to reduce the nominal event rate of 40 MHz (20 MHz during RunI) to less than 100 kHz. The event selection at L1 is based on information from the Muon Spectormeter and the Calorimeters only, since the basic requirement for an event, supposed to be an interesting one, is the presence of large-p_{T} jets or a signal from Muon Spectrometer indicating large-p_{T} muon(s). The information from the Inner Detector

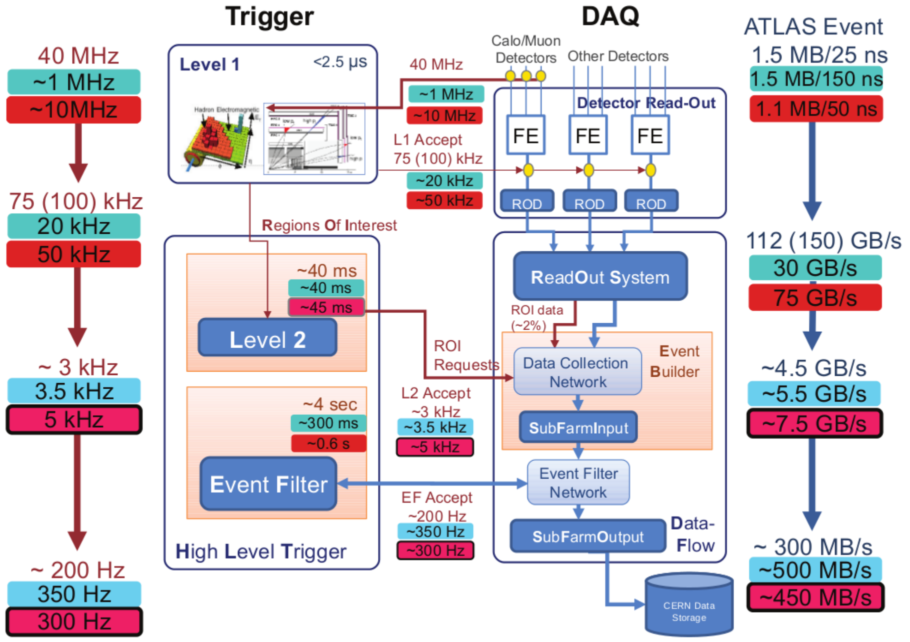


Fig. 3.6 The schematic design of TDAQ system of the ATLAS detector. Soruce: [12]

can not be considered at L1 because of its high granularity, and therefore complicated track reconstruction which requires longer computing time. The events passing through the L1 are stored at Read-Out Systems (ROS) for further processing by the Higher Level Trigger (HLT). The HLT consists of two levels: L2 and Event Filter (EF). In contrast to the with L1 trigger, which uses only electronic units, the HLT uses software running at a computing farm. The Region of Interest (RoI) is an interesting fraction of one event data defined at L1 within specified η/ϕ range. The L2 trigger analyses and judges the event based on detailed information in the RoIs and reduces the data collection rate to 3 kHz appropriately.

The event which passes the requirements at L1 and L2 proceeds to the EF, which uses dedicated worker nodes (SFI - Sub-Farm Input) to complete the event reconstruction. The final selection reduces the collection rate to 200 Hz and records the interesting event data (O(100 MB)) to the storage center located at CERN.

4. Computing at the ATLAS Experiment

The general concept of the ATLAS Distributed Computing (ADC) [13] is given at the beginning of this chapter. The monitoring of the ATLAS computing centres (sites) and the transfers between them are described in dedicated sections of this chapter. The topic of ADC is included in this thesis because the University of Freiburg (me included) gave significant contributions to the ADC monitoring. The ABCD method, see Section 4.2, was developed by Martina Pagacova and me. The study of the communication protocols, see Section 4.3, has been performed by me with a help and contribution of Alejandro Alvarez. The upgrade of the Sonar functional tests, see Section 4.3, has been done by Vincenzo Lavorini and me. All these topics have been supervised by Alessandro Di Girolamo (ADC co-coordinator for Integration and Commissioning)[1], David Cameron (ADC co-coordinator for Operation)[2] and Jaroslava Schovancova (responsible for CERN Services Operations and Monitoring)[3].

4.1 Scheme of ADC

The EF introduced in Section 3.5 produces order of 100 MB/sec of data which has to be promptly stored for further processing. Furthermore, this data is replicated to ensure data availability and to optimize the analysis throughput. This requires large amount of storage resources placed all over the world, at the universities or in specialised computing centres. Additionally, these centres contain powerful computing elements (CE) to reprocess, analyse or simulate the data. The CE, Storage Elements (SE) and other facilities responsible for processing data define a site. The batch system, such as Load Sharing Facility (LSF), is responsible for the distribution of the tasks between individual worker nodes, managing several CEs to compute a given task. The site contains other essential facilities and software such as: filename space (database), the client for Distributed Data Management (DDM), software for analysis and simulation, communication facilities and monitoring tools. All these components

---

[1] to the date of writing the thesis

[2] to the date of writing the thesis

[3] to the date of development of ABCD

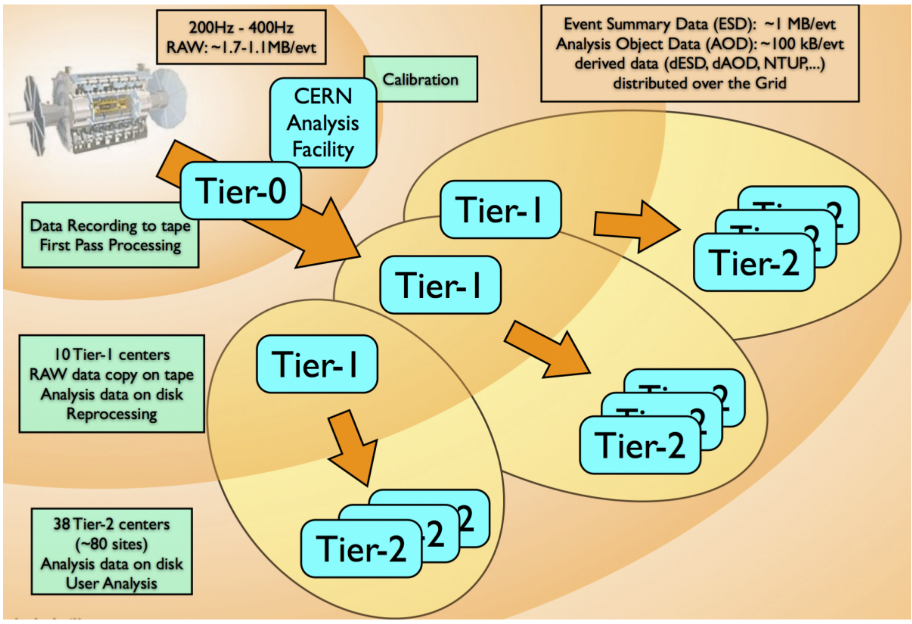


Fig. 4.1 The organization of the sites into the Tier levels [14].

and their structure form an architecture, which is unique for each site.

The entire net of sites dedicated to ATLAS experiment is referred to as ATLAS grid, which is part of the World-wide LHC Computing Grid (WLCG). The sites in the grid are organized into levels according to their role in the data processing, see Fig. 4.1. The Tier-0 centre is only one, located at CERN. Its main task is to make first data processing and to record all raw data on tape and promptly distribute replicas to the Tier-1 centres. The first calibration of the data is performed at the Tier-0 and the events are merged into the outputs: Event Summary Data (ESD) of size $\sim$ 1 MB/event or Analysis Object Data (AOD) of size $\sim$ 100 kB/event [4]. The Tier-1 sites are responsible for storing the raw data on tape and reconstruction outputs on disk for fast access. The Tier-1 centres play a role of head centres for some sets of Tier-2 sites, usually geographically close to a Tier-1 site but more importantly with a good connectivity to the Tier-1. Such a group of sites is referred to as a cloud. The main role of the Tier-2 sites is to provide the CPU and storage for the end-users (physics analysis). The MC simulation is performed at both, Tier-1 and Tier-2 sites. The Central Services nodes, are reserved for operation of the ADC software products responsible for managing the entire grid such as: Rucio - DDM System [15], PanDA - Production and Analysis distribution System [16], [17] or HammerCloud - site operation and stress testing [18].

To complete the general description of the ATLAS computing, a simplified map of the communication between the essential instances is depicted in Fig. 4.2. Each line in the map denotes

---

[4] ROOT accessible format xAOD is used in Run II

a bi-directional communication between the two facilities. The steps for the processing of a computing task are illustrated by the green path, starting at "Jobs" that are defined by the end-user. If the requester is allowed to submit the job, the job is segmented into simpler tasks by the Production System (ProdSys), that is an essential component of PanDA. The tasks are distributed into PanDA queues (PQ) and sent to the most relevant site. The relevance of the site is judged by the PanDA or JEDI brokerage according to the site accessibility to the input data and the site availability. The Wrapper is a mediator between sites and PanDA. A similar path can be followed for the data instead of the jobs, as illustrated by the red path with DDM in place of ProdSys and File Transfer System (FTS) as a mediator between the DDM and the sites. On the background is the ATLAS Grid Information System (AGIS) which populates a database with detailed information about each site. Each of the instances on the figure needs a monitoring of its functionality. Two of the monitoring tools, that have been partially developed and operated by me, are described in the following Sections 4.2 and 4.4.

The monitoring discussed in this thesis uses a platform called Site Status Board (SSB) [19], that provides the data needed by the site performance analysis, and stores its results, allowing easy visualization. The SSB is also used by the sites to monitor themselves and by shifters responsible for monitoring the grid.

4.2 The ATLAS Site Availability Status

The availability of the site is the most important parameter of its performance. If the site is available for user analysis less than 90 % of the time, a user may choose a different site for his analysis. Furthermore, the sites which are not sufficiently available do not offer to the ATLAS experiment the resources that the experiment needs. A fully automated monitoring of site availability status has been developed since the beginning of Run I. The ABCD[5] method described in this section evolved into the ATLAS Site Availability Performance metric (ASAP).

There are two inputs to the ABCD method: the Switcher and the PanDA resources monitoring. The Switcher is an agent that switches the status of a site. The states managed by the Switcher are:

- online (ON) - the site is available for the analysis
- brokeroff (BO) - the jobs are running but no new job is accepted
- offline (OFF) - the downtime of the site

---

[5] the name is described later in the text

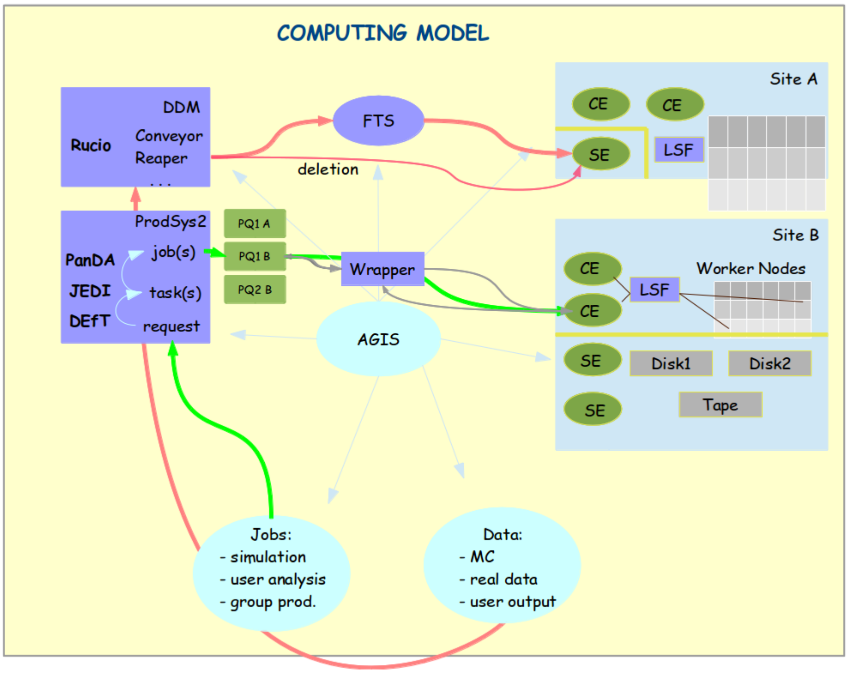


Fig. 4.2 The communication model in the ADC. Each line represents a communication between the two end-points. The arrows show the flow of the job request or data. Communication itself is usually flowing in both directions. The model is schematic and simplified showing the case of a set-up with only two sites.

PR status Switcher2	ON $\vee$ W $\forall$	OFF, T, BO S-$\forall$ $\vee$ W	OFF, T, BO U-$\forall$ $\vee$ no-action	UNK $\forall$	$\forall$ exc ON UNK
Mapping	*good*	*null*	*bad*	*null*	*null*

Table 4.1 The mapping rules between the Switcher and PR.

- the test period (T) - is usually set after the OFF mode by the communication with HammerCloud

All of this states can be either scheduled (S-) by the site in advance or unscheduled (U-). Each site can have more PanDA Resources (PR), computing resources dedicated to the ATLAS experiment. Each PR can employ more PanDA queues where the jobs wait before computation. The PanDA monitoring provides the states of single queues and PRs, which usually correspond to the actions of the Switcher. However, there are cases when the status of a site is changed either intentionally or not (e.g. the administrator set some PR the into test mode). For both, the Switcher and the PR-monitoring, the white (W) or unknown (UNK) period can occur. The W and UNK correspond respectively to the states when the monitoring tools are out of operations and when the monitoring tools are not able to retrieve the status for some reason.

More general states from a user's point of view (*good*, *bad* and *null*) are derived from combinations of the PR states and the Switcher actions. This is summarized in Table 4.1. The W and UNK states are listed individually in this table because they can not be scheduled (S-, U- are useless for them). The status no-action for the Switcher means that no change of the status is reported by the Switcher.

Although T is a bad period from the point of view of a user, the less stringent *null* status is used in this case for the 60 minutes after S-OFF. The site needs some time to be tested before it is switched online again, thus the *bad* label would not be appropriate.

The general states for each PR have to be combined in the general status of the site. If at least one of the PR general states is calculated as *good*, then the site status is marked as *good*. If there is no PR marked *good* but at least one PR is marked *null*, then the site status is marked *null*. Consequently, only if all PRs are marked *bad*, the site is marked *bad*. The *good*, *bad* and *null* states for a given site are determined at a certain point in time. Thus, they do not represent longer-term performance of the site. These states are stored each minute in the SSB.

On the other hand, the A, B, C or D final state is given to each site according to the percentage of the *accumulated good* period in the last 30 days. An accumulated *null* period is subtracted from the calculation. If the availability of a site is more than 90%, the site is labelled A in case of Tier-1 or Tier-2D (defined in the Section 4.4) and B in case of Tier-2. If the site is available for 80-90% of time, it is labelled C. The sites with performance below 80% have status D. This classification was designed to be integrated in PanDA to weight the number of tasks assigned to the site. This has however never been fully automated. Nevertheless, constant (not automated) weights used to be affected by this classification. Furthermore, the state of a site is a good motivation for site improvements. This classification was very quickly replaced by the ASAP metric (not discussed here) based on the methods developed for ABCD.

4.3 Communication Protocols

The communication protocol is an equivalent of a language in the communication between two service endpoints. The information (usually containing a request for certain actions), that is sent between the service endpoints, is expressed in this language and both services have to be able to translate that message. Protocols are used in the ATLAS experiment for communication between sites and also for an internal communication between the components of a site. The protocols supported by most of the ATLAS sites are gridFTP [20] with its subset gsiFTP [21] and HTTP [22] with its extension WebDAV [23].

The utilities installed at the sites managing the communication with the protocols are: Storage Resource Manager (SRM) [24] and Grid File Access Library (GFAL2) [25]. The SRM is a utility officially used by all ATLAS sites (SRM-less are planned to be supported as well). It has its own protocol used only for its internal operation, interfaces are provided to allow interaction with protocols chosen by the user (automatically gridFTP). The SRM is able to manage all of the important communication for all actions with a file, e.g. authentication, space requests, reporting failures, set-up of the transfer, deletion and other actions. All of the communication before the real action on the file is called SRM-overhead, which is monitored and investigated to be as short as possible. The GFAL2, if installed, can replace the SRM almost for all of the actions and thus can serve as an alternative to SRM. Note that GFAL2 usually spends less time in the overhead phase and thus is supposed to be faster. The GFAL2 is however not yet fully supported within the ATLAS Grid but its performance is tested. Both of these utilities support bulk operations (the same action is done on more files with only one communication). The WebDAV and gsiFTP can be operated by both of these managers.

4.4 The ATLAS Site Connectivity

A good connectivity among the ATLAS sites is essential for a good performance of the ATLAS grid in terms of speed. The jobs set by users often need firstly to download the input data from a distant site, the distribution of the raw data has to be very fast, the job options often require partial set-up from the CERN site, the output of the job needs to be replicated to a different site than was used for the computation. All these, and many other operations on the ATLAS grid, require very fast connectivity to the other sites. It is thus natural, that sites are monitored and analysed according to their connectivity and alarmed or penalized in cases when the connectivity is not sufficient. At the time of the writing this thesis, the commercial network monitoring system perfSONAR [26] is still not fully established for all of the ATLAS sites[6]. The DDM Sonar (see [27], slide 11), or shortly Sonar, developed by Simone Campana has been in operation for the purposes of the ATLAS site connectivity monitoring since the beginning of Run I. The next version of this tool designed for Run II is discussed in this section.

Sonar consists of three components: the test files located at the sites and the software to operate them, the Rucio rules (commands) managing the transfers and analysis framework to evaluate the quality of a site connection to the grid. Ten files of 1GB each with a random content are placed at each site. These test files are transferred one by one every week, between each pair of the sites (each channel). The tests are spread over the entire week in order to lower the network load. The transfer rate and the time spent in SRM overhead are provided by the FTS and published in the Dashboard [28]. This data is then fetched and analysed with an analysis framework. The Sonar tests are schematically depicted in Fig. 4.3.

The average connectivities for each channel are stored in the SSB for monitoring and further analysis. Based on these tests, the concept of Tier-2D was established. A Tier-2D site is an enhanced level of Tier-2 site which has connectivity better then 5 MB/sec to at least 8 of the 11 Tier-1 sites. This criterium has to be fulfilled during the last 7 days and over at least three weeks out of the last five weeks. From the two directions in a channel, always the worst is taken into account. These criteria have been studied and optimized for the needs of ADC. This concept is however obsolete, because the connectivities between the sites in the ATLAS grid improved significantly in the past years and almost all of the sites are well within these criteria nowadays. The concept of Tier-2D is simply not needed any more since the tasks that would have priority for Tier-2D sites can be nowadays computed by almost all of the Tier-2 sites in a reasonable time. The Sonar tests are thus used only for monitoring and warning the few sites with the worst connectivity.

---

[6] Only a few DDM endpoints are not yet involved in the perfSONAR tests.

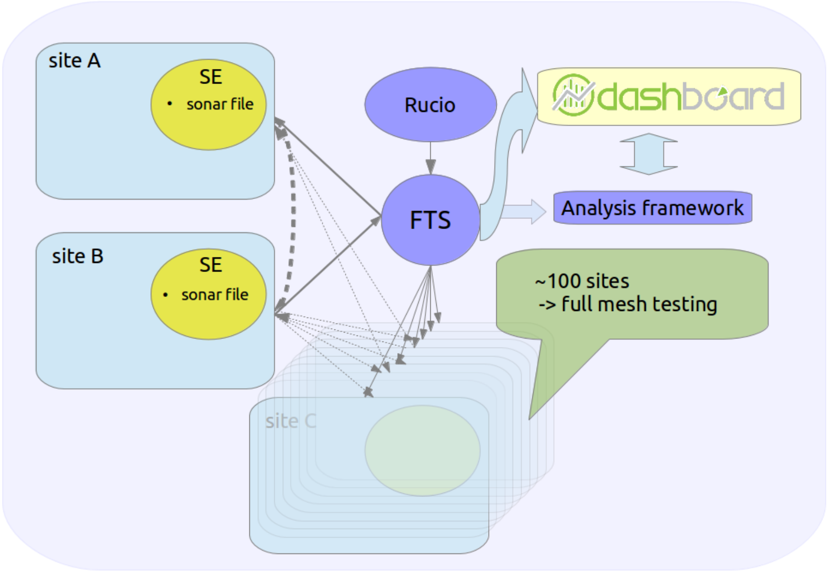


Fig. 4.3 The schema of the Sonar testing.

5. Theoretical Background

5.1 The Standard Model

The set of particles observed in the experiments of the 20th and 21st centuries, involving accelerators and cosmic rays, is very well described by the Standard Model (SM). It is a gauge field theory using a Lagrangian density $\mathscr{L}$ formalism. $\mathscr{L}$ is composed of space-time dependent (x) fields representing the particles, the derivatives of these fields and their complex conjugates. Scalar fields $\Phi(x)$ describe particles without a spin. Column matrix of fields $\psi(x)$, denoted as Dirac spinors, have four components dedicated to two projections of the spin of fermions and theirs anti-fermions. Vector fields $A^{\mu}(x)$ describe the spin 1 particles, the vector bosons.

Each term of $\mathscr{L}$ represents either the particle mass (self-interaction of the particle mass eigenstates) or an interaction. Measurable quantities, such as probability or cross section, can be calculated using an action calculated from $\mathscr{L}$ by inserting it into initial and final state, usually defined in a Fock space[1]. These states act on each term individually and define an amplitude. The amplitudes are added and squared. Since the sum is calculated before the square operation, individual amplitudes are not distinguishable. A Feynman diagram is the visualisation of a single term of $\mathscr{L}$, thus its meaning is purely mathematical.

$\mathscr{L}$ can be built in two ways: up-bottom, when new particles and interactions are needed to suppress some divergence, and bottom-up, when a new symmetry is required. Every physically correct extension of $\mathscr{L}$ can be achieved by both approaches. Thus, a new symmetry is usually investigated in case of need to suppress some divergence. Each of the symmetries is equivalent to the conservation of some quantity. At the beginning, a set of naturally trusted symmetries such as those coming from four-momentum conservation (Lorentz transformation plus translation) or CPT invariance is required. Furthermore, another symmetry can be placed on the basic $\mathscr{L}_0$ involving only scalar and fermion fields and their derivatives:

$$\Phi(x) \to \Phi' = U(x)\Phi(x) \tag{5.1}$$

---

[1] The Hilbert space in the representation of the occupation numbers.

where $U(x)$ is a group member derived from generators of a Lie algebra. It is usually written as a matrix and $\Phi(x)$ in these equations denotes an x-dependent field or a one-column matrix of scalar or fermion fields. However, the invariance of $\mathscr{L}_0$ is not possible to achieve by the transformation from Eq. 5.1. Thus, new terms are introduced to $\mathscr{L}_0$, involving the vector boson fields and their transformation is defined in a way, that such extended $\mathscr{L}$ is now invariant:

$$\mathscr{L}_0 \to \mathscr{L}_0' = \mathscr{L}_0(\Phi'(x), A'^{\mu}(x)) = \mathscr{L}_0(\Phi(x), A^{\mu}(x)) \tag{5.2}$$

Such an invariance is referred to as a gauge symmetry.

Similarly to classical electro-dynamics, the quantum electro-dynamic part of the SM $\mathscr{L}$ is expected to be a gauge $U(1)$[2] invariant. Indeed, the $\mathscr{L}$ of SM obeys a set of gauge symmetries $SU(3)_\mathrm{C} \otimes SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$[3]. The $SU(3)_\mathrm{C}$ is introduced due to color conservation observed in the strong interactions of gluons and quarks. The $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ corresponds to the nature of electro-weak interactions, which conserve the weak isospin I and the hypercharge Y. The subscript L is used, because only left-handed particles have non-zero I. Quarks interact via electro-weak interaction as well and thus carry I and Y, but the eigenstates (the solutions of the electro-weak part of $\mathscr{L}$) are rotated with respect to the mass eigenstates. All of these symmetries can also be motivated with a top-down approach to deal with some high-energy divergence, e.g. introducing four vector boson fields in the electroweak theory.

The complete set of particles (mass eigenstates of the fields) included in the SM $\mathscr{L}$ is shown in Table 5.1 for left-handed fermions and in Table 5.2 for right-handed fermions. The quarks have additionally a color charge, which can be of three values. The gauge bosons, arising with the terms fixing the gauge invariances, are listed in Table 5.3. Each of these particles has its anti-particle, which differs only by the electric charge. Consequently, the particles with zero electric charge are their own anti-particles. The gluon is an intermediate particle of the strong interaction, thus it carries a color and anti-color.

The last missing brick to complete the SM is a Higgs boson. The $\mathscr{L}$ without introducing the Higgs boson can not contain the mass terms of the gauge bosons directly, since they would not be gauge invariant. From the top-down point of view, the need for a neutral scalar field is motivated by the $WW \to WW$ scattering, which suffers from high-energy divergences in the theory of electroweak interactions. A neutral scalar field is introduced via the Goldstone Model [30], by adding a potential with a negative mass term[4]. This leads to a scalar field with non-zero vacuum expectation value (VEV). Figuratively speaking, the other particles

---

[2] Generator is represented by a number

[3] Generators are represented by 3×3 and 2×2 unitary matrices with determinant 1 respectively.

[4] This means that the mass terms can not be interpreted as self-interaction of free particles.

Fermion	Mass	Charge	I_3	Y
e_L	0.51 MeV	$-1e$	-1/2	-1
μ_L	105.65 MeV	$-1e$	-1/2	-1
τ_L	1777.03 MeV	$-1e$	-1/2	-1
ν_{eL}	< 3 eV	0	+1/2	-1
$\nu_{\mu L}$	< 0.19 MeV	0	+1/2	-1
$\nu_{\tau L}$	< 18.2 MeV	0	+1/2	-1
u_L	$1-5$ MeV/c^2	$2/3e$	+1/2	+1/3
d_L	$3-9$ MeV/c^2	$-1/3e$	-1/2	+1/3
c_L	$1.15-1.35$ GeV/c^2	$2/3e$	+1/2	+1/3
s_L	$75-170$ MeV/c^2	$-1/3e$	-1/2	+1/3
t_L	≈ 174 GeV/c^2	$2/3e$	+1/2	+1/3
b_L	$4.0-4.4$ GeV/c^2	$-1/3e$	-1/2	+1/3

Table 5.1 The left-handed fermions and their properties, [29].

Fermion	I_3	Y
e_R, μ_R, τ_R	0	-2
u_R, c_R, t_R	0	+4/3
d_R, s_R, b_R	0	-2/3

Table 5.2 The right-handed fermions and their properties, [29].

Boson	Mass	Charge
W^-	80.4 GeV	-1
Z	91.2 GeV	0
γ	0	0
g	0	0

Table 5.3 The properties of the bosons of the SM, [29].

sailing in the space with the homogeneous VEV acquire mass. Mathematically, quadratic (self-interaction) mass terms of the fields with multiplicative factors containing the VEV occur in the complete SM $\mathscr{L}$. Strictly speaking, it is not an interaction of the particle with the VEV, because the VEV is just a constant and the Higgs field is an excitation of it. Although this construction, referred to as Higgs mechanism, can seem rather difficult, it explains the origin of the masses of the particles in the SM.

The symmetries of $\mathscr{L}$ are often naturally different from the symmetries of its solutions (resulting fields). This is referred to as spontaneous symmetry breaking. The way how the symmetry is broken is defined by the initial and final state, i.e. the interaction that is relevant for a given process. Thus, the fields in one process can be rotated against the fields of the same particles acting in another process. The mixing angle θ_W between the fields of electroweak gauge bosons is determined by the Higgs mechanism. The other mixing angles are not explained by the theory so far and need to be measured experimentally.

5.2 Status of the SM

There are 18 free parameters in the SM. Apart from the mixing angles, there are also several coupling constants, and two free parameters introduced by the Higgs mechanism. These parameters have to be established experimentally. However, there is no other reliable theory that could be able to decrease this number. Although the acceptable number of free parameters is a rather speculative issue, the following problems need to be solved:

1. The neutrinos of the SM are massless, but in a such case, the oscillation between different neutrino flavours is not possible. The oscillations of neutrinos were observed and are under investigation by many experiments. The extension of the SM by terms with massive neutrinos (Majorana particles), without introducing any new field or mechanism, suffers from non-renormalizability [31]. The properties of neutrinos can also be investigated by double-β decay experiments [32] that have already shown promising results [33].

2. The muons show a $\sim 3.6\sigma$ deviation in the anomalous magnetic moment [34]. One of the candidates that could explain this observation is supersymmetry, discussed in next Section 5.3.

3. The gravitational interaction is not included in the SM $\mathscr{L}$ and any extension with gravitons leads to non-renormalizability.

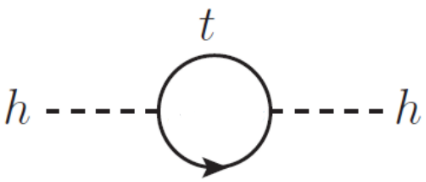


Fig. 5.1 Feynman diagram showing the loop correction to the Higgs mass term.

4. The fermion loop corrections to the Higgs mass, see Fig. 5.2, are quadratically divergent with the energy scale (represented for example by a cutoff Λ). This is not a mathematical issue since it can be cancelled by a large bare mass. However, the cancellation of two large values in order to obtain the observed results is considered by many as unnatural. Additionally, it requires fine-tuning of related couplings.

5. The strong CP problem: In principle, the part of the $\mathscr{L}$ corresponding to the strong interaction allows the violation of a combination of charge conjugation and parity symmetry. However, there is no experimental evidence of such a violation. This is a similar problem as the fine-tuning previously mentioned. The free parameters of the strong $\mathscr{L}$ would have to acquire very precisely chosen values in order to correspond to the experimental observations.

6. The dark matter: the exploration of the dynamics of galaxies indicates a different mass density, than what visually observed or theoretically predicted. The remaining mass could be partially explained by an observation of new particles.

5.3 Supersymmetry, extension of the SM

The fine-tuning issue with the mass correction to the Higgs boson depicted in Fig. 5.2 can be solved by the existence of a scalar particle with the same properties (except from the spin) as the fermion involved in the loop and with a quartic λ and Yukawa coupling g with the Higgs field. After recalculation of the Higgs mass with such new fields, the two leading divergent correction terms vanish. Consequently, the bare mass of the Higgs boson is closer to the observable mass at low energy. In this sense, the hierarchy problem, as well as the fine-tuning problem, are no longer manifest. A similar loop as in Fig. 5.2 can be drawn for each of the fermions and a similar principle of a scalar partner is needed, although the corrections to the Higgs mass are not as substantial as in the case of t-quark due to their lower mass.

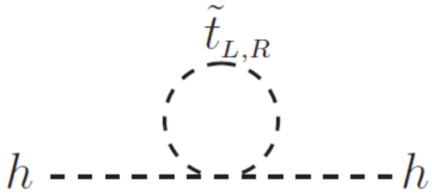


Fig. 5.2 Additional loop correction to the Higgs mass in case of existence of the $\tilde{t}$.

From the bottom-up point of view, the only possible extension of the SM such that the corresponding symmetry is generated by a pseudo Lie algebra[5] is Supersymmetry (SUSY) [35]. Such an extension of the SM was firstly proved to be renormalizable in [36].

The SUSY is defined by its symmetry transformation $Q_\pm$, that acts on the spin component of the field:

$$Q_\pm |j\rangle = |j \pm {}^1/_2\rangle \tag{5.3}$$

since $Q_\pm$ can act on each field, the same number of bosons and fermions in the final model is required. The new field with different spin is referred to as superpartner or more specifically sfermion, slepton, selectron, etc. and is marked by a tilde. Unfortunately, none of the known particles can be a superpartner of the other one, since the properties of the superpartner are "almost" the same as those of the related SM particle. Consequently, a new set of particles is defined, see Table 5.4. Such an extension of the SM by a minimal number of new particles is referred to as Minimal Supersymmetric SM (MSSM), which was firstly discussed in [37].

The particles in Table 5.4 are organized into multiplets according to the chirality in the same manner as their partners in the SM. The Higgs sector has to be enhanced by two doublets of Higgses resulting in two doublets of superpartners. Such enhancement is needed due to gauge anomaly, see [38], page 21. The anomaly occurs when calculating the triangle fermion loop with a $SU(2)$ current situated in the corners of the triangle. The contributions of such a loop of fermions in the SM nicely cancel. Nevertheless, this not the case for SUSY models if only one higgsino is implemented.

The SUSY model in the form introduced above has however an issue. The superpartners can not have the same mass as the SM particles, since they are not observed yet. In order to let such a model survive, some mechanism that increases the masses of the superpartners is needed. This issue is solved by *soft symmetry breaking*, which stands for an addition of terms to $\mathscr{L}$ that break the supersymmetry implicitly, hoping for some mechanism like spontaneous symmetry breaking, which can explain the masses at scale of the Grand Unified Theory[6] (GUT scale).

---

[5] the combination of commutation and anti-commutation relations is required, similarly to SM

[6] The theory where electro-weak and strong interactions are unified. The energy scale is of $\sim 10^{16}$ GeV.

Name	sparticles	spin
squarks	$(\tilde{u},\tilde{d})_{\mathrm{L}}$, $(\tilde{c},\tilde{s})_{\mathrm{L}}$, $(\tilde{t},\tilde{b})_{\mathrm{L}}$, $\tilde{u}_{\mathrm{R}}$, $\tilde{d}_{\mathrm{R}}$, $\tilde{c}_{\mathrm{R}}$, $\tilde{s}_{\mathrm{R}}$, $\tilde{t}_{\mathrm{R}}$, $\tilde{b}_{\mathrm{R}}$	0
sleptons	$(\tilde{e},\tilde{\nu}_e)_{\mathrm{L}}$, $(\tilde{\mu},\tilde{\nu}_\mu)_{\mathrm{L}}$, $(\tilde{\tau},\tilde{\nu}_\tau)_{\mathrm{L}}$, $\tilde{e}_{\mathrm{R}}$, $\tilde{\mu}_{\mathrm{R}}$, $\tilde{\tau}_{\mathrm{R}}$	0
higgsinos	$(\tilde{H}_{\mathrm{u}}^{+}, \tilde{H}_{\mathrm{u}}^{0})$, $(\tilde{H}_{\mathrm{d}}^{0}, \tilde{H}_{\mathrm{d}}^{-})$	$^1/_2$
gluiono	$\tilde{g}$	$^1/_2$
gauginos	$\tilde{W}^{\pm},\tilde{W}^{0},\tilde{B}^{0}$	$^1/_2$

Table 5.4 The particle eigenstates in the gauge invariant $\mathscr{L}$ of the SM extended with SUSY, namely MSSM.

The symmetry breaking has to be by the definition soft, because for the underlying mechanism it could be difficult to return the supersymmetry to the $\mathscr{L}$. At the same time, large differences between the masses of the particles and their superpartners would hardly explain the initial motivation: the hierarchy problem.

The soft symmetry breaking leads to rotated mass eigenstates, that are accessible by the experiments. The charged gauginos and higgsinos: $\tilde{W}^{\pm}$, $\tilde{H}_u^{+}$ and $\tilde{H}_d^{-}$ create the mass eigenstates called charginos: $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^{\pm}$. The neutral ones: $\tilde{H}_u^0$, $\tilde{H}_d^0$, $\tilde{B}_0$ and $\tilde{W}_0$ form neutralinos: $\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$ and $\tilde{\chi}_4^0$. After applying electroweak symmetry breaking, five higgs mass eigenstates occur, additionally to the SM spectrum: h^0, H^0 A^0 and $H^{\pm}$.

Another issue of the new model is that it breaks the conservation of lepton (L) and baryon (B) number. This would potentially lead to a shorter decay-time of the proton, which is not observed. The additional conservation law fixing this issue is R-parity. For each particle, R-parity is defined as:

$$R = -1^{3(B-L)-2s} \tag{5.4}$$

where s is the spin of given particle. In other words, the superpartners carry R-parity equal to -1 and R-parity for the SM particles is 1. The R-parity conservation then stands for the conservation of this number in a vertex. Thus, the SUSY particles are produced in pairs and if a SUSY particle occurs in a decay chain, there has to be some SUSY particle at the end of the decay chain. Such particle is referred to as Lightest Supersymmetric Particle (LSP) and usually corresponds to $\tilde{\chi}_1^0$

5.4 Simplified Models

The MSSM in its full form introduces more than 100 free parameters. This number can be decreased by specific assumptions as for instance in the Constrained MSSM (CMSSM), see

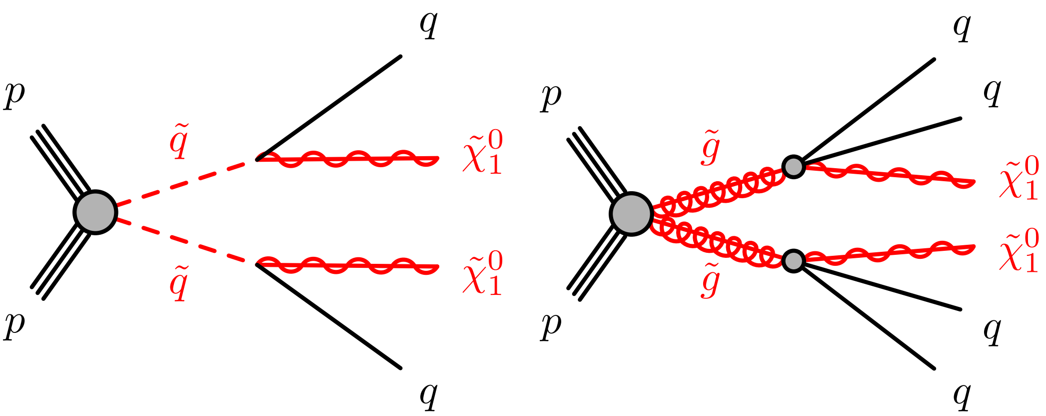


Fig. 5.3 The schematic diagrams for two "direct" SUSY simplified models. Left: $\tilde{q}$-pair production and direct decay, Right: $\tilde{g}$-pair production and direct decay.

Chapter 8 of [39], where five parameters fully determine the mass spectrum of SUSY particles. However, too aggressive constraints placed on the theory could mislead the analysis in such a way that the discoveries are missed although they are at a reachable energy scale. In order to adapt the analysis to the theoretical model without too harsh constraints, simplified models are built. The decay chain with a specific particle content is fixed. Only the masses of the unknown SUSY particles in such decay chain are used as free parameters. The probability of such an event is given by the production cross section of the mother particles and the branching ratios at each node of the decay chain. This is the unknown quantity, on which the analysis can provide limits for each possible set of masses of the SUSY particles appearing in a given decay chain. The limits can be then drawn in a graph with the axe corresponding to the masses of these SUSY particles.

More importantly, the topology of such an event helps us to optimize the analysis in its selection criteria. Finally, even such optimized analysis can still be sensitive to a different theoretical model which contain similar decay chains. The example of four simplified models is shown in Fig. 5.3 and Fig. 5.4. The decay chains in the second figure contain the $W^\pm$ in the final state, which is the object of study of this thesis.

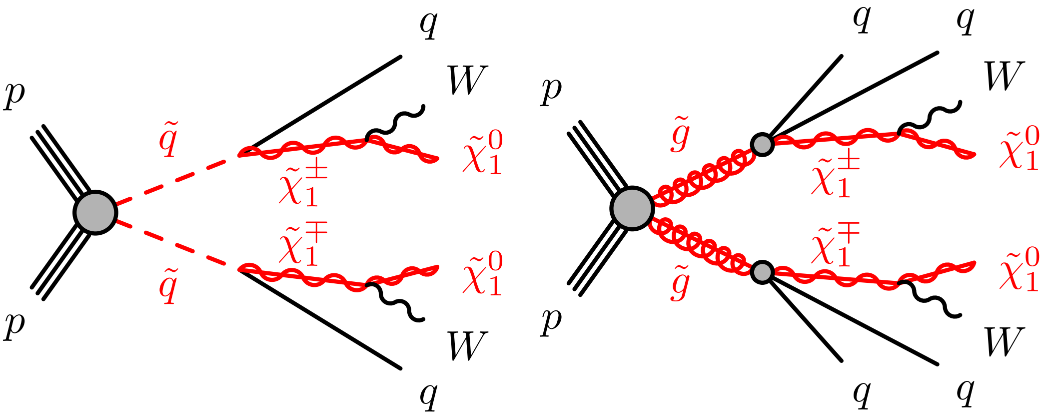


Fig. 5.4 The schematic diagrams for two "1step" SUSY simplified models. Left: $\tilde{q}$-pair production followed by one-step decay over $\tilde{\chi}_1^\pm$ to $\tilde{\chi}_1^0$. Right: $\tilde{g}$-pair production followed by one-step decay over $\tilde{\chi}_1^\pm$ to $\tilde{\chi}_1^0$. Two W's are produced in the final state of both of these models.

6. Overview of the 0-lepton Analysis

The convention followed in this thesis is that the vector variables are written bold ($\boldsymbol{A}$) and their amplitudes are written with normal font (A). The signature in the ATLAS detector associated with a particle is indicated by the term "object". The objects measured by the ATLAS detector are listed and described in chapter 7.

The most important definitions of the variables used in the analysis are given in this chapter. When the essential variables are defined, the techniques to operate with them are explained as well.

6.1 Dedicated Variables

The variables used in this analysis are selected based on their separation power between the signal and the background according to long-standing experience gained in the previous versions of this analysis. The transverse momentum $\boldsymbol{p}_\mathrm{T}$ of an object can be used for these purposes. It denotes the momentum projection into the xy-plane (the plane perpendicular to the beam).

The exact definition of the missing transverse momentum $\boldsymbol{E}_\mathrm{T}^\mathrm{miss}$ is given in a dedicated section in Chapter 7. In order to introduce other variables, it is however worth to provide a general definition of $\boldsymbol{E}_\mathrm{T}^\mathrm{miss}$:

$$\boldsymbol{E}_\mathrm{T}^\mathrm{miss} = -\sum_\mathrm{obj} \boldsymbol{p}_\mathrm{T}(\mathrm{obj}) \tag{6.1}$$

where obj stands for the objects. The missing transverse momentum itself is not included among the objects in this definition but is treated as an object later on, as it can be associated with invisible particles as ν or non-SM $\tilde{\chi}_1^0$.

The other important variable is effective mass m_eff defined as:

$$m_\mathrm{eff}(N_\mathrm{obj}) = \sum_\mathrm{obj} p_\mathrm{T}(\mathrm{obj}) + E_\mathrm{T}^\mathrm{miss} \tag{6.2}$$

where obj runs over a well defined set of the objects N_{obj}, which in this analysis corresponds to some specific sub-set of the jets with p_{T} >40 GeV. If N_{obj} is not specified, the entire set of the jets with $p_{\mathrm{T}} > 40\mathrm{GeV}$ is used. Such a variable is often referred as $m_{\mathrm{eff}}(\mathrm{incl.})$.
The previous two quantities can be combined into the ratio $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(N_{\mathrm{j}})$, where N_{j} specifies the number of jets involved. This ratio has usually lower values for the SM events and larger values for most of the SUSY signal models. In particular the multi-jet background is largely suppressed by applying even a looser cut on this variable.
Although the mentioned variables have already excellent discrimination power against all types of SM background, a third variable $\Delta\phi$ plays an important role for the suppression of the multi-jet background. It is defined as:

$$\Delta\phi(jet_i) = |\phi(\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}) - \phi(\mathrm{jet}_i)| \tag{6.3}$$

where jet_i stands for the jet (defined in Chapter 7) with the i-th largest p_{T}. In this sense, $\Delta\phi$ gives more freedom when optimizing the selection rules by varying cuts on previously defined variables since it is not correlated to them. Note that the multi-jet background produces many jets and large $E_{\mathrm{T}}^{\mathrm{miss}}$ occurs when these jets are not precisely measured or calibrated. In this case, the $E_{\mathrm{T}}^{\mathrm{miss}}$ can be close in ϕ to the leading jet. Consequently, the $\Delta\phi_{1,2,3}$ defined as $\min(\Delta\phi(\mathrm{jet}_1), \Delta\phi(\mathrm{jet}_2), \Delta\phi(\mathrm{jet}_3))$ can be used as a selection criteria.
The last variable defined here is the transverse mass between $E_{\mathrm{T}}^{\mathrm{miss}}$ and some object with $p_{\mathrm{T}}(\mathrm{obj})$:

$$m_{\mathrm{T}}\left(E_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{obj}\right) = \sqrt{2E_{\mathrm{T}}^{\mathrm{miss}} p_{\mathrm{T}}(\mathrm{obj})\left(1 - \cos\Delta\phi\left(E_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{obj}\right)\right)} \tag{6.4}$$

6.2 Signal Selection and Background Suppression

The Signal Region (SR) is the key ingredient of each "cut-and-count" analysis aiming to search for an excess above well known background or to disprove some specific model. It is defined as a set of selection rules optimized for the maximal sensitivity to the studied signal model with maximal possible reduction of the background at the same time. The expected number of events can be calculated from the MC simulation either directly, or by an indirect semi-data-driven method based on Control Regions (CRs) which is generally considered to be more reliable. The CRs are defined by another set of criteria with the following properties:

- each CR is kinematically close to some SR
- each CR is dedicated to one of the major background process (e.g. W+jets, Z+jets, $t\bar{t}$, QCD), i.e. a given background process has significant event count in a given CR

- the expected event count for signal is negligible in every CR

Since this analysis applies requirements in its SRs on large $E_{\mathrm{T}}^{\mathrm{miss}}$, large p_{T} of the jets and lepton veto, the dominant background composition is:

- W+jets: whichever process when a W boson accompanied with jets is in the final state
- Z+jets: Z boson and jets in the final state
- $t\bar{t}$: production of top quark pairs
- QCD jets: the events with jets originating from the strong interaction only

Consequently, each SR used in this analysis employs at least four CRs. The final event yield in the SR is then estimated with the formula:

$$N(\text{SR,proc,est}) = N(\text{CR,proc,obs}) \left[\frac{N(\text{SR,proc,raw})}{N(\text{CR,proc,raw})} \right] \tag{6.5}$$

where proc, est, obs, stand for process, estimated, observed respectively. The last parameter defines the type of data. The "raw" by this convention means the simulation data without any sampling. The term in square brackets is often referred to as Transfer Factor (TF). The TFs are established from the MC simulation only, as shown in formula 6.5, the only exception being the TF for the multi-jet background where a data-driven method is used. The term $N(\text{CR,proc,obs})$ is estimated by a simultaneous fit of the equations $N(\text{CR,proc,obs}) = N(\text{CR,all proc,obs}) - N(\text{CR,rest proc,obs})$ where the yields in all CRs (not only those for main backgrounds) are employed. This fit procedure[1] runs additionally to the background, exclusion and observation fits mentioned in Section 6.3. The result of the eq. 6.5 serves as an input to the likelihood, see eq.6.6.

The main CRs are listed in Table 6.1.

Contrary to direct MC simulation yields in the SR, this method takes advantage of the ratio in TF where the effect of systematic uncertainties is largely cancelled. Since there are four main backgrounds, additional CRs are needed in order to have less parameters in the fit than the number of fitted bins (event counts in CRs). This is achieved with the development of CRZ and CRτ aiming to be dominated by the events containing Z and τ respectively.

Another set of selection criteria is defined as Validation Region (VR). Such a set is supposed to be free of signal. The same fit procedure as in the case of SRs is used to estimate event counts in the VRs, which are compared to the observed data. In case of sufficient agreement, the whole procedure is validated. The fit methods and statistical tools used for hypothesis (SUSY model) testing are described in Section 6.3.

---

[1] In other words, it is a normalization of the estimated events in the CRs to the observed data.

CR	SR background	CR process	CR isolation
CRγ	$Z \to \nu\nu$+jets	γ+jets	Isolated lepton
CRQ	Multi-jets	Multi-jets	SR with reversed cut on $\Delta\phi$ and $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$
CRW	$W \to l\nu$+jets	$W \to l\nu$+jets	30 GeV$< m_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell) <$100 GeV, b-veto
CRT	$t\bar{t}$ and single-t	$t\bar{t} \to b\bar{b}qq'\ell\nu$	30 GeV$< m_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell) <$100 GeV, b-tag

Table 6.1 The main Control Regions used in this analysis.

6.3 Statistical Apparatus

Sophisticated statistical treatment is provided by an a tool called *HistFitter* [70], which is a user-friendly framework based on the official packages *RooStats* [71] and *HistFactory* [72]. The mathematical background is briefly described in this chapter, see the doctoral work of Michael Rammensee [73] for the precise equations and detailed description of the full procedure adapted to the 0-lepton analysis.

A likelihood fit is performed in this analysis. The likelihood function has the general form:

$$L(n|\mu, s, b, \theta) \quad = \quad P_{\mathrm{SR}} \times P_{\mathrm{CRW}} \times P_{\mathrm{CRT}} \times P_{\mathrm{CRY}} \times P_{\mathrm{CRQ}} \times C_{\mathrm{Syst}}\,. \tag{6.6}$$

each Poisson function P_i on the right is dependent on the number of observed events n_i and expected events λ_i in a given SR or CR labeled by index i. The parameters of the likelihood function (n, μ, s, b, θ) stand for the sets of observed number of events in the SR, signal/background strength estimated by the fit, signal yield in SR, background yield in the SR and "hidden" nuisance parameters, respectively.

The event yield λ_i is calculated from formula:

$$\lambda_i(\mu, s_i, \mathbf{b}, \theta) = s_i(\theta) \cdot \mu_s + \sum_{j}^{W,Z,t\bar{t},QCD} b_{i,j}(\theta) \cdot \mu_j + b_i^{VV}(\theta) \tag{6.7}$$

where the additional index j denotes the background process. The terms $b_{i,j}$ and $s_{i,j}$ correspond to the estimates of the event yields in the region (SR, CR or VR) for the background and signal respectively. In the likelihood fit, these parameters are multiplied by a factor of the form $1 + \Delta\theta$ where Δ is the initial shift of the statistics in a given region caused by some systematic uncertainty and θ is a dedicated nuisance parameter. The parameters μ_s, μ_j are the signal

strengths of the signal and the background, that are the fitted parameters of interest.
Three different modes of the fit are performed in this analysis:

- **background fit**: The parameter μ_s is set to 0 and the fit runs only over CRs. The event counts estimated from the fit can be transferred to the SR with 6.5. The one-sided p-value is used to verify that the estimated background is in good agreement with the observed data in a given SR.
- **observation fit**: Sometimes called model-independent fit. The parameter μ_s is assumed to be non-negligible and is a free parameter of the fit. The fit runs additionally also over the SR.
- **exclusion fit**: The parameter μ_s is set to 1 for a given signal. This hypothesis is compared to the observation fit and the exclusion limit on the cross-section of a given signal model is established.

Statistical properties such as significance or Confidence Level (CL) are calculated with the asymptotic formula [74]. Note that in some cases when rough quick estimate of the significance is needed, the simple formula is used:

$$Z = \frac{s}{\sqrt{s+b}} \tag{6.8}$$

The hypothesis with μ_s set to 0 is compared with the hyphothesis with μ_s as free parameter of the fit to prove that there is no significant excess in SR and to set an upper limit on signal event count in SR. The exclusion of the specific model proceeds with comparison of the hypothesis with μ_s set to 1 against the hypothesis with free μ_s with default value starting at 1.

6.4 Simulation Samples

The simulation of the proton collisions followed by multiple decays or interactions is essential to compare the theory with the measured data. The different generators differ mostly by the approximations employed. The simulation usually proceeds in many phases depending on the scale of the simulated effect.
Cross-section of a given process, when two protons collide and primary particles are produced, is approximated with Factorisation Theorem:

$$\sigma(s,\mu_R,\mu_f) = \sum_{i,j} \int dx_i dx_j \hat{\sigma}(\hat{s},\mu_R) f_i(x_i,\mu_f) f_j(x_j,\mu_f) \tag{6.9}$$

where s is center-of-mass energy squared, $\hat{\sigma}$ is cross-section for the two partons from the sum, $\hat{s}$ is a fraction of s carried by these two partons, and f_i is parton distribution function (PDF), which is a probability density function that a given parton i carries momentum fraction x_i of the colliding proton. The μ_R is the renormalization scale. The μ_f is an energy scale factor, which defines the energy range at which non-perturbative QCD can be used for calculation of PDFs. The early phase of an event when this calculation is applied is called factorization phase.

The factorization phase typically produces primary particles which further evolve via QCD shower. This evolution connects the previous hard scale with the hadronization scale, where the hadrons are produced (in simulation) using phenomenological fragmentation models with parameters fitted to the data. Further decays of the hadrons are simulated with effective theories which are able to operate at the first order of perturbation theory at the time of the writing of this thesis. The issue of such a model is the lack of the possibility of a second parallel (underlying) hard or semi-hard event, which can produce partons missing in the simulation.

The modern MC generators can be split in two kinds: Matrix-Element (ME) generators and general purpose generators. The ME generators are designed to generate one specific hard process with the differential cross section calculated at fixed order from the Feynman rules. The outcome of a ME generator is usually combined with a general purpose generator to reconstruct the parton shower. Such a combination has to be corrected for double counting, i.e. the hardest leg of the parton shower can overlap with the legs from a higher order calculation. The general purpose generators are designed to generate a broad spectra of events and evolve them through the factorization and hadronization phase. In the most recent versions of such generators, underlying events and their overlap with the leading event are implemented as well. The cross sections of the processes are hard-coded in the general-purpose generators or calculated in the cases of two-to-two processes.

The MC generators used here are listed in the table 6.2 and briefly described in this paragraph. PYTHIA [40] is the most established general-purpose generator, optimized over the past decades. The evolution variables in the parton showers can be either the p_{T} of the parton (the shower is p_{T}-ordered) or the virtuality (the on-shell vertexes are preferred). The shower evolves to the a certain cutoff of the p_{T} of its constituents which determines the beginning of the hadronization phase.

HERWIG++ [41] is a general-purpose generator with angular-ordered parton showers. The hadronization is performed via a cluster model [42]. The underlying event is not simulated, thus the outcome has to be interfaced with some other generator.

SHERPA [43] is another general-purpose generator with a dipole formulation used for parton showering, and a cluster model for hadronization. The underlying event is simulated similarly to PYTHIA.

POWHEG-Box [44], [45] is a ME generator operating at NLO precision. It provides a method to merge the outcome with p_{T}-ordered parton showers, i.e. PYTHIA, by simple replacement of the hardest component of the shower by NLO calculation.

As already mentioned in Chapter 6, this analysis uses CRs for background estimation. The inputs to the fits are the MC expectations provided by different generators. Table 6.2 summarizes the different generators used for this analysis, with details on their implementation.

sample	Generator	fragmentation/hadronization	odrer in α_s	PDF
W+jets	SHEPRA-1.4.0	HERWIG-6.520	NNLO [46]	CT10 [47]
Z/γ+jets	SHEPRA-1.4.0	HERWIG-6.520	NNLO	CT10
γ+jets	SHEPRA-1.4.0	HERWIG-6.520	LO	CT10
$t\bar{t}$	POWHEG-BOX-1.0	PYTHIA-6.426	NNLO+NNLL	CT10
Diboson	SHERPA-1.4.0	-	NLO	CT10

Table 6.2 The most important MC generators used for this analysis with the technical information.

The SUSY models are simulated with HERWIG++-2.5.2 or MADGRAPH-5.0 [48] matched to PYTHIA-6.426, using PDF set CTEQ6L1. The schematic decay diagrams of the most relevant SUSY models for this analysis are depicted in figure 5.4, for which MADGRAPH-5.0 has been deployed. These diagrams correspond to $\tilde{g}$-pair or $\tilde{q}$-pair production followed by the 1-step decay over the $\tilde{\chi}_1^{\pm}$ to $\tilde{\chi}_1^0$.

7. Object Definition and Event Cleaning

An "object" in the ATLAS nomenclature is the response of the detector to one or more particles. This chapter lists and describes all the objects relevant for this analysis.

7.1 Jets

The jet is an object originating from particles carrying colour charge and from their hadronization. The key terms are introduced at the beginning of this section, followed by a brief description of the reconstruction and calibration of the jets as well as the discussion of the experimental uncertainties assigned to the jets.

Topological clusters

A topological cluster (Topocluster) [49] is a set of neighbouring calorimetric cells with larger energy deposit with respect to their average noise. The average electronic noise is monitored for each TileCal cell individually. Topoclusters are iteratively built starting from the seed cell, which is defined as a cell with the Signal/Noise ratio (S/N) larger than 4. The neighbouring cells of the seed cell are added to the cluster in the case of $S/N > 2$. The last step is repeated until there is no neighbouring cell to the cluster with $S/N > 2$. Finally, all neighbouring cells to those ones which are already in the cluster are included as well, without any threshold requirement.

The anti-$k_\perp$ algorithm

Many jet algorithms have been developed in the past years in order to deal with issues related to hadronization and underlying event contamination, impacting mostly the final jet boundaries. The anti-$k_\perp$ [50], $k_\perp$ [51] and the Cambridge/Aachen (CA) [52] algorithms are examples of jet algorithms based on sequential recombination. The infra-red and collinear soft radiations cause theoretical inaccuracies in the definition of the jets, which is however not the case for these three algorithms.

The anti-$k_\perp$ algorithm [50] is briefly described in the following paragraph. As an input, a group of objects[1] is used, labelled by indexes i or j. Two distance parameters are defined: $d_{i,j} = \min(p_{\mathrm{T}}^{-2}(i), p_{\mathrm{T}}^{-2}(j))\Delta R_{ij}^2/R^2$ with a free parameter R, referred to as cone size (see Eq. 3.2) and $d_{iB} = p_{\mathrm{T}}(i)$ where B stands for the beam. In this sense, the d_{iB} represents the distance of the object i to the beam. Equipped with these parameters, one can loop over all the objects and for each object i calculate the $d_{\mathrm{min},i} = \min d_{ik}$ where $k \in (j,B)$. The minimum value d_{min} of the $d_{\mathrm{min},i}$ among the objects i is found. If $d_{\mathrm{min}} = d_{ij}$ for some j, the objects i and j are grouped into new object. If $d_{\mathrm{min}} = d_{iB}$ the object i is declared as a final jet and is removed from the list of the objects for further iterations. This loop is run until none of the objects remains.

The Topoclusters play the role of the objects in our case. The only difference between CA, $k_\perp$ and anti-$k_\perp$ is the way how the distance parameters are defined. The anti-$k_\perp$ algorithm has the feature that the final jet areas are rather regular. Thus, it is technically easier to deal with the jet energy calibration and subtraction of the pile-up and the underlying event. This is the main reason that the anti-$k_\perp$ algorithm is the algorithm of choice for most of the ATLAS analyses.

Calibration

The calibration of the jets is needed to fit their response in calorimeters to the actual energy they carry. Several techniques have been developed by ATLAS to deal with this issue with the highest possible accuracy [53], [54].

A simple method called Jet Energy Scale (JES) is used to correct the jets for a difference between real and true simulated energy deposits of the jets in the Calorimeters[2]. This method applies an easy correction factor to the jet energy derived from either comparison of simulated data and test-beam data acquired in 2010/2011 or from comparison of MC simulation and a well calibrated object.

The Local Cluster Weighting (LCW) method is used in this analysis. This method takes advantage of the identification of hadronic and electromagnetic topoclusters, for which different calibration corrections are applied. The final correction is affected by energy deposits in the dead material, invisible energy deposits (neutrinos + nuclear reactions) and finally also by out-of-cluster depositions (OOC) which are initially not included in the Topocluster. The LCW is fully derived from MC simulation. Compared to other calibration schemes, it has better

---

[1] The input can be also clusters of objects or jets.

[2] The "true information" stands for a simulation at generator level

performance in terms of jet resolution. The final calibrated jets are often called LCW+JES jets.

In order to ensure the correctness of the calibration, the events with the jets propagating in the very forward direction $|\eta| > 2.8$ are rejected. The jets with $p_{\mathrm{T}} < 20$ GeV are not involved in the analysis for the same reason.

Systematic uncertainties on JES calibration

The following in-situ methods have been used to assess and validate the systematic uncertainties coming from JES calibration:

- comparison to the momentum carried by tracks associated to a jet
- direct p_{T} balance between a photon and a jet
- photon p_{T} balance to hadronic recoil
- balance between low-p_{T} and high-p_{T} jet system

The uncertainties resulting from these methods are combined using HVPTools [55] based on a Toy MC method. The relative JES uncertainty is roughly of 1-3 % [56] and JER ranges between 10 and 20 % [57].

To illustrate the evaluation of systematic uncertainties, the first method is briefly described. The charged-to-total momentum ratio is defined as $\mathrm{r}_{trk} = \frac{\sum p_T^{trck}}{p_T^{jet}}$. The double ratio is defined to compare the data and MC simulation as: $R_{\mathrm{r}_{trk}} = \frac{[<\mathrm{r}_{trk}^{data}>]}{[<r_{trk}^{MC}>]}$, and it is evaluated for several $|\eta|$ regions and p_{T} bins. The systematic sources for this method are:

- the physics behind the MC generator affects the fraction of coloured particles and colour distribution, and therefore r_{trk}
- the interactions of charged particles passing through a given material are not described with absolute accuracy by the simulation
- the track reconstruction efficiency
- high-p_{T} jets suffer from lower track p_{T} resolution in the Inner Detector
- low-p_{T} jets suffer from lower track p_{T} resolution in the Calorimeter

b-tagging

The results of the analysis can be improved by tagging specific particle in many ways. For instance, as a veto when such particle is not expected in the final state of the signal model under consideration or, exactly in the opposite way, as a selection rule in a signal region in the case where the particle is expected in the final state. Note that there are both such cases presented in this analysis, for example b-quarks appear as a decay products of t or $\tilde{t}$ and b-veto has been tested as well within ATLAS SUSY 0-lepton analysis in the past. The b-tagging plays also an important role for building our Control and Validation regions, see Chapter 8.

The key of the b-tagging of the jet lies in the precise determination of the displacement between the secondary and the primary vertex. The typical displacement of a b-jet is expected to be of the order of mm (due to the longer lifetime of B-hadrons) which can be established with sufficient precision with the ATLAS detector as long as there are at least 5 tracks associated with the secondary vertex [58]. The jets associated to other flavours are distinguished from b-jets by training a neural network. More accurately, the MV1 algorithm [58] is used in this analysis, which provides also the uncertainties assigned to the b-tagging.

The efficiency of the b-tagging is defined as the the number of correctly reconstructed b-jets divided by a total number of the true b-jets [59]. This ratio can have different values for a different MC sample and MC generator. The operation point of the b-tagging is defined as the efficiency of the b-tagging for a $t\bar{t}$ sample. This analysis uses the b-tagging at operation point of 70%.

7.2 Leptons

Since a lepton veto is applied in this analysis and Control Regions with leptons are used, proper lepton identification is needed to reject the events carrying leptons originating from the primary vertex, and therefore, with large transverse momentum. A description of the reconstruction of all three types of charged leptons (e, μ, τ) is given in this chapter.

7.2.1 Electrons

Similar methods are used for reconstruction of both electrons and photons. The E/gamma performance group [60] of the ATLAS experiment is responsible for the maintenance the reconstruction code for both these particles and releases the official recommendations and tools that are followed in this analysis.

Reconstruction

A particle entering the LAr causes many interactions in its absorbers, that lead to a shower of secondary particles. The ionization of the material in the gap between the absorbers produces electrons, which drift to the electrodes and cause an electric signal brought by cables to the front-end electronics, where the signal is amplified and sampled. Further reprocessing and merging of the signal from all affected LAr cells provides a map of towers, i.e groups of neighbouring cells with significant response. If the cluster of close-by towers is matched well with a track originating in the primary vertex, this cluster and track are assumed to be an electron response.

The Electron identification yields three qualities: *loose*, *medium* and *tight*. The *loose* selection applies criteria to quantities such as the shower shape and the fraction of energy deposited in the Hadronic Calorimeter with respect to the EM Calorimeter. These criteria are enhanced by requirements on the associated track quality and track-cluster matching. The *medium* selection adds selections on the transverse impact parameter and on the number of hits in the TRT associated with the track. Furthermore, at least one measured hit in the innermost layer of the Pixel Detector is required, to exclude unconverted photons. The *tight* selection applies criteria on the ratio of the energy and track momentum and a veto on photon conversions.

These regimes differ by the number of the variables used in a multivariate method to gain the most efficient electron selection. The final *tight* electrons are a subset of the *medium* electrons which are a subset of the *loose* electrons. In this sense, the tighter selection has high background rejection but the efficiency of identification is lower than for the looser selection. The detailed description of this procedure is given in [61].

Calibration

The energy of the electron or photon is calculated from clusters of cells of in the EM Calorimeter. Such a measurement needs to be calibrated to fit correctly to the real energy of the particle. The calibration proceeds with the following steps:

1. Calibration constants are derived from a MC simulation as an output of a multivariate optimization comparing the particle's original energy with the energy of EM cluster.

2. The differences between longitudinal layers of the EM calorimeter have to be equalised in data with respect to simulation.

3. The calibration constants derived above are applied to the reconstructed cluster energies for both MC and data.
4. A set of corrections dedicated to non-uniformity and non-stability of the EM Calorimeter is applied.
5. The response in data is calibrated to the response in simulation using a $Z \to ee$ sample. The electron energy in simulation is smeared to fit with the resolution derived from the data.
6. The final energy scale is validated with MC simulation for $J/\psi \to ee$ for electrons and Z$\to$ $ll\gamma$ for photons.

The Z resonance is used to set the absolute energy scale of the electrons as the electrons from the Z decays achieve a calibration inaccuracy of less than 0.05%.

Uncertainties

The uncertainties affecting the electron energy measurement are classified as Electron Energy Resoulution (EER) and Electron Energy Scale (EES). The sources of these uncertainties within the range $|\eta| < 2.47$ are:

1. the (non)uniformity of the LAr
2. inter-calibration of LAr layers
3. simulation inaccuracy
4. setting of the absolute energy scale from the Z resonance
5. pile-up
6. tracking inefficiency

The non-uniformity due to mechanical and high-voltage defects has been established to be less than 0.5% in barrel and less than 0.75% in endcap. The amount of material upstream is established from the shower depth with typical accuracy of 3-10% of radiation length.

Table 7.1 shows the final uncertainty on calibration resulting from the procedure described above. The numbers are derived from [61].

A multivariate method optimized on MC simulation has been used to improve the resolution of the electron/photon energy measurement in the Calorimeter. The improvement is of the order

	$\lvert\eta\rvert < 1.37$	$1.37 < \lvert\eta\rvert < 1.52$	$1.52 < \lvert\eta\rvert < 1.82$	$1.82 < \lvert\eta\rvert$
e[10GeV]	$0.4-1.0\%$	1.1%	1.1%	0.4%
e[40GeV]	0.04%	0.2%	0.2%	0.05%
converted-γ	$0.2-0.3\%$	-	0.4%	0.2%
unconverted-γ	$0.2-0.3\%$	-	0.9%	0.2%

Table 7.1 The table of inaccuracies caused by calibration for the electron and the photon.

of 10-20% for photons with respect the to previous calibration approaches and 5-30% for electrons depending on the amount of material upstream of the Calorimeter's active medium. The pile-up has very small effect on the calorimeter energy response to electrons and photons and great stability is also achieved in performance as a function of time, both of about 0.05%. In the case of tracking, the higher the energy of electron, the larger the inaccuracy, due to the smaller curvature of its track in the Inner Tracker. This effect is compensated by the calorimetric measurement of energy, which becomes more precise as energy increases.
The relative energy resolution can be parametrized using a form:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{7.1}$$

where all the energies are in GeV. The a in this equation is called sampling term, the b is a noise term and c is a constant term. All of these terms are dependent from η. Their rough values are: $a \sim 10\%$, $b \sim 0.35\cosh(\eta)$, $c \sim 0.7\%$. The EER uncertainty is roughly 10% for the electrons with E_{T}<50GeV and increasing asymptoticaly with E_{T} to 40%.

7.2.2 Muons

Two types of muon reconstruction have been tested or used in this analysis: "combined" and "segment-tagged" [62]. A "combined" muon requires identification of the muon in the Inner Detector and in the Muon Spectrometer independently and the two tracks to be matched together. A "segment-tagged" muon requires hits from the muon spectrometer matched to a track from the Inner Detector, but a full reconstruction of the track from the Muon Spectrometer is not required.
Additional requirements are applied to both types of muons. The acceptance selection criteria p_{T} > 10GeV and $|\eta| < 2.4$ are applied. At least one pixel hit (or the track passing through a dead pixel sensor) in the Pixel Detector is required. At least 5 hits in the SCT and less than 3 pixel or SCT holes (expected but not detected hit) are required. At least one hit has to be recorded in the b-layer, if it is expected from the track. If the track passes the acceptance

$0.1 < |\eta| < 1.9$, a special condition for track quality has to be fulfilled: $n_{\mathrm{TRT}} > 6$ and $n_{\mathrm{TRT}}^{\mathrm{outliers}} < 0.9 n_{\mathrm{TRT}}$ where n_{TRT} is the number of hits in the TRT sub-detector and $n_{TRT}^{outliers}$ is the number of hits neighbouring with the track. The condition $\Delta R(\mathrm{jet}, \mathrm{muon}) > 0.4$ is used to reduce the overlap between jets and muons. The signal muons used in the CRs with leptons (see Chapter 8) have three more requirements: $|z_{\mu} - z_{\mathrm{PV}}| < 1$ mm, $d_0 < 0.2$ mm and the sum of p_{T} of the tracks (excluding the muon track itself) in a cone of $\Delta R < 0.2$ has to be less than 1.8 GeV. Here PV denotes the primary vertex and d_0 and transverse impact parameter to the PV.

An event which contains such a muon is vetoed in this analysis (0-lepton). The additional criteria required for the signal muon in the CRs are not applied for vetoing in SRs.

Uncertainties such as Muon energy scale (MES), the uncertainty on the energy resolution (MER), and the uncertainty on the muon identification efficiency (MEFF) are considered in this analysis . Detailed information on the uncertainties can be found in [62].

Tau Leptons

The τ-leptons decaying leptonically are suppressed by the lepton veto. The τ-leptons decaying into hadrons are reconstructed for two purposes:

- two VRs with τ-leptons in the final state have been developed to validate the CRW and CRT
- potential usage of τ-veto has been studied

The instructions for τ reconstruction using the multivariate methods described in [63] have been followed in the case of VRWτ and VRTτ with few improvements optimized for this analysis, that are summarized in [64]. Alternative methods for the tau reconstruction designed to veto the events with τ-lepton in the final state are described in Appendix B.

7.2.3 Photons

The photons are investigated by the same ATLAS performance group as electrons and the methods used for their identification, calibration and derivation of the uncertainties were already partially discussed in Section 7.2.1.

The photon can convert into an electron pair (converted photon). In such a case we identify an electron and positron with the same momentum originating from a secondary vertex. The opposite case, when the photon does not decay into electron-positron pair (unconverted photon), is identified from its shower shape in the EM Calorimeter.

The granularity of the first layer of the LAr is sufficient to distinguish between single photon shower and two overlapping showers produced by neutral hadrons. If the LAr towers match with a single/double track originating from a well reconstructed vertex, such response is assumed to be a converted photon. If the cluster of towers does not match to any track, it is assumed to be an unconverted photon candidate.

The photon identification is based on shower shape and on the calorimeter isolation transverse energy, which is defined as the sum of energy deposited in the calorimeter cells in a given cone around the track but excluding the cells hit by track itself.

Photons are not used in the SR selection, but are used to constrain the Z+jets background using CRγ. Furthermore, the events where jets and high-p_{T} γ candidates lie within $\Delta R < 0.2$ from each other are rejected (see [64], page 5).

7.3 Missing Transverse Momentum

The missing transverse momentum/energy is defined as an imbalance in the vector sum of the transverse momenta of all the reconstructed objects in the event. Such an observable has origin in the SM from neutrinos, cosmic rays, beam-halo and, as a fake $E_{\mathrm{T}}^{\mathrm{miss}}$, from inactive or bad-performing parts of the detector. Its large values can indicate the presence of particles beyond SM, such as neutralinos in SUSY models. Therefore, rather high cuts are placed on $E_{\mathrm{T}}^{\mathrm{miss}}$ in many analysis attempting to search for a new particles. A precise estimate of $E_{\mathrm{T}}^{\mathrm{miss}}$ is needed to suppress the impact of the uncertainties related to this observable.

The prescription from [65] is followed in the analysis in this thesis except for a few differences described below. The clusters in the event are associated with a particle in the following order: electrons, photons, jets and muons. The electrons and the photons are required to have a $p_{\mathrm{T}} > 10$ GeV and a requirement $p_{\mathrm{T}} > 20$ GeV is applied to the jets. The rest of the clusters are summed to the soft term. The negatives of the sum of transverse momenta of off all objects are added, resulting in the final $E_{\mathrm{T}}^{\mathrm{miss}}$, symbolically written as:

$$E_{\mathrm{T}}^{\mathrm{miss}} = E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{electrons}) + E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{photons}) + E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{jets}) + E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{muons}) + E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{soft}) \quad (7.2)$$

Each of the terms alone has no physical meaning, as there is no reason why individual sets of objects should be balanced in the momentum. The reason to treat them separately is that for each of the objects, a different calibration has to be applied. The local topoclusters calibration mentioned in Section 7.1 is used for the topoclusters of the soft term. The main difference with this method and the method defined in [65] is that the term for hadronically decaying tau leptons is not specified here. The jets corresponding to hadronic taus are included either in $E_{\mathrm{T}}^{\mathrm{miss}}$(jets) or $E_{\mathrm{T}}^{\mathrm{miss}}$(soft) depending of their p_{T}. The second difference is that the term $E_{\mathrm{T}}^{\mathrm{miss}}$(muons) is based on the reconstruction and calibration used in this analysis as described in Section 7.2.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ from Eq. 7.2 can be calculated for both data and MC samples and this method can be validated by comparing the results. Furthermore, MC simulation provides us the true-$E_{\mathrm{T}}^{\mathrm{miss}}$ information defined as the sum of the transverse momenta of the neutrinos. This feature can be used for the validation as well.

The uncertainties on $E_{\mathrm{T}}^{\mathrm{miss}}$ are calculated term by term from the propagation of the uncertainties placed on each individual object. Two in-situ methods using $Z \rightarrow \mu\mu$ have been used in case of the soft term [66].

7.4 Removing Overlapping Objects

Some of the real physical objects can be misleadingly identified multiple times as different objects. This can be however recognised by the similar directions of such objects. The ΔR is used for this purpose. All but one possibilities have to be removed from the set of overlapping objects. The removal of the objects obeys the following conditions:

1. If ΔR(electron,jet)$<$0.2, the electron is preferred and the overlapping jet is removed.
2. If ΔR(muon,jet)$<$0.4, the object is identified as a jet and the muon is removed.
3. If 0.2$<$ ΔR(electron,jet)$<$0.4, the object is treated as a jet and the nearby electron is removed.

The overlap between jets and photons is treated in CRY such that the photon is preferred if the jet is within $\Delta R < 0.2$.

7.5 Event Cleaning

One of the challenges of the ATLAS experiment, when working with such an enormous amount of collisions, is to select the interesting events. Most of the reduction proceeds already at the trigger level (see Section 3.5) which has to be very fast. Additional selection criteria based on event quality can arise from the full reconstruction of the event at the EF level. At this stage, the events with an uncorrected $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV, with at least one uncorrected jet with $p_{\mathrm{T}} > 80$ GeV are selected.

The event can be affected by background effects not originating from the bunch crossing of interest. The dominant non-collision backgrounds are:

- coherent noise in the electromagnetic calorimeters
- spike noise of the hadronic LAr calorimeter end-cap
- cosmic rays and beam halo
- the events originating from proton collision with gas molecules in the not-ideal vacuum of the pipe

The event cleaning from these types of background is discussed in this chapter.

An event is discarded, when none of the jets in the event satisfies quality requirements on low detector noise and non-collision background, and on well-reconstructed primary vertex [67]. The official recommendations [68] for Tile and LAr cleaning are followed in this analysis. The LAr DAQ group releases a flag to every event whether its status is OK, WARNING or ERROR. The ERROR status means serious data integrity issue like hardware breakdown or noise bursts occurring in the same time interval when the event happened. The events with the ERROR flag are subtracted from the integrated luminosity recorded by ATLAS. The WARNING status is assigned to the event when it suffers from significant noise in EM calorimeters. The Tile DAQ group releases a quality flag for each event in a similar way as the LAr DAQ which is considered in this analysis as well. A TTC (see Section 3.5) restart could cause an incomplete event right after the restart. Such events are rejected from the analysis.

The event is selected or rejected according to the origin of the reconstructed jets. The jets originating from non-collision background can be of two types: "real" jets produced by some physics process from a different bunch crossing and "fake" jets produced by mis-measurement of the energy deposits in the Calorimeter. An event containing such jets is rejected applying the jet 'Looser' cleaning criteria recommended by the ATLAS JetEtmiss performance group [69].

The jet cleaning criteria contain cuts on the jet charge fraction f_{ch} and jet electromagnetic fraction f_{em} defined as ratio of the energy of the jet measured in the inner tracker or EM Calorimeter respectively and the total energy of the jet. In addition to this selection, the event is rejected if the leading jet within $|\eta| < 2.0$ satisfies: 1) $f_{\text{ch}} < 0.02$ or 2) $f_{\text{ch}} < 0.05$ and $f_{\text{em}} > 0.9$. Such cuts are effective in rejecting events affected by tracks from minimum bias event, cosmic background and events with the leading jet not produced by hard scattering. In addition, the event is rejected when any of the leading jets satisfies $f_{\text{ch}} < 0.3$ and $f_{em} < 0.25$. This conditions were designed to select the events with jets not affected by any hot cells of the

Calorimeter. Besides this rule, the recommendation of the ATLAS DAQ performance groups are followed precisely and all of the events identified as significantly affected by hot cells are excluded from the analysis.

Data corruption in the readout of a TileCal cell can cause a fake measurement of negative energy. The topoclusters containing such cells can not be clustered in any object and they contribute to the last term $E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{soft})$ of the missing transverse energy, see Eq. 7.2. Denoting such contribution as $E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{Cellout})$, one can build a condition to reject an event affected by such mis-measurement:

$$\frac{E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{Cellout})}{E_{\mathrm{T}}^{\mathrm{miss}}}\cos\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{Cellout}),E_{\mathrm{T}}^{\mathrm{miss}}) > 0.5 \tag{7.3}$$

Furthermore, a cut on the energy-weighted mean of the propagation time of the two leading jets after overlap removal is applied: $| < t > | > 4\mathrm{ns}$. Such requirement has no effect on the event counts in the SRs and is applied to reject the residual non-collision background.

A fake muon is a muon reconstructed as a combined muon, see Section 7.2.2, but not belonging to any real muon produced by the pp-collision. Such a fake particle can be caused either by a wrong matching of the hits in the Muon Spectrometer with the track in the Inner Detector or by a residual particle of a very energetic jet escaping the Hadronic Calorimeter. Bad reconstruction of muons affects the analysis, since the lepton veto is applied in SRs, and due to potential appearance of muon in the CRT and CRW. A good quality of muon reconstruction is provided by vetoing an event with:

$$\frac{\sigma(q/p)}{|q/p|} > 0.2 \tag{7.4}$$

$$\frac{E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{muon})}{E_{\mathrm{T}}^{\mathrm{miss}}}\cos\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{muon}),E_{\mathrm{T}}^{\mathrm{miss}}) > 0.5 \tag{7.5}$$

where $E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{muon})$ denotes the transverse missing energy calculated only from reconstructed muons and q, p are charge and momentum of a muon respectively. These cuts are very efficient in rejecting events, where a large fraction of $E_{\mathrm{T}}^{\mathrm{miss}}$ originates from muons. In case of the CRs containing muons, the cosmic background is suppressed by requirements on impact parameters, as discussed in Section 7.2.2.

8. Results of the 0-lepton Analysis

In 2013/2014, the 0-lepton analysis was very well established with several publications released [75], [76]. The work that lead to the publication [64] starting from the previously released conference note [77] is described in this and the following chapters.

The input to this analysis is provided by the datasets produced by the EF (see Chapter 3.5), where the events are already pre-selected according to their quality and relevance to this analysis as discussed in Section 7.5. The further selection criteria applied additionally to the pre-selection criteria define the SRs.

The final set of SRs is listed in Table 8.1. The selections employed here have been optimized and the variables involved have been chosen according to their efficiency in selecting the SUSY signal. The lepton veto, listed in the first row, is introduced due to the constraint of this analysis to stay orthogonal to the other ATLAS analyses with lepton(s) produced in the final state. Furthermore, most of the dominant decays of the SUSY particles produced in the strong interactions are non-leptonic. Note that due to the lepton veto, the electro-weak background is largely suppressed and thus the 0-lepton SUSY analysis is generally considered to be one of high sensitivity. This feature is valid only when analysing signal models which are not themselves suppressed by the lepton veto. After the row with the cut on the $E_{\mathrm{T}}^{\mathrm{miss}}$, the requirement on the p_{T}s of the jets follows in the next rows of this Table. These cuts are motivated by the production of the $\tilde{g}$-pair or $\tilde{q}$-pair with rather large mass, thus large-p_{T} jets are expected in the final state. Due to R-parity conservation introduced in Chapter 5, the LSP in the final state does not decay further [1]. The $\tilde{\chi}_1^0$ is assumed to be the LSP, which is the most likely situation in most SUSY models. The rest of the selection criteria are motivated by their separation power between the signal and the background.

A widely used discriminating variable is $m_{\mathrm{eff}}(\mathrm{incl.})$ (see Eq. 6.2). In the way it is defined, it serves as a variable sensitive to the center-of-mass energy of colliding partons, and consequently, to the masses of the produced particles. The performance of $m_{\mathrm{eff}}(\mathrm{incl.})$ for 2j SR is shown in Fig. 8.1. Two benchmark $\tilde{q}\tilde{q}$-direct models (i.e. $\tilde{q} \to q\tilde{\chi}_1^0$) show different behaviour with respect to the background. The typical shape of this distribution for other simplified

---

[1] R-parity Violating (RPV) theories are not discussed here.

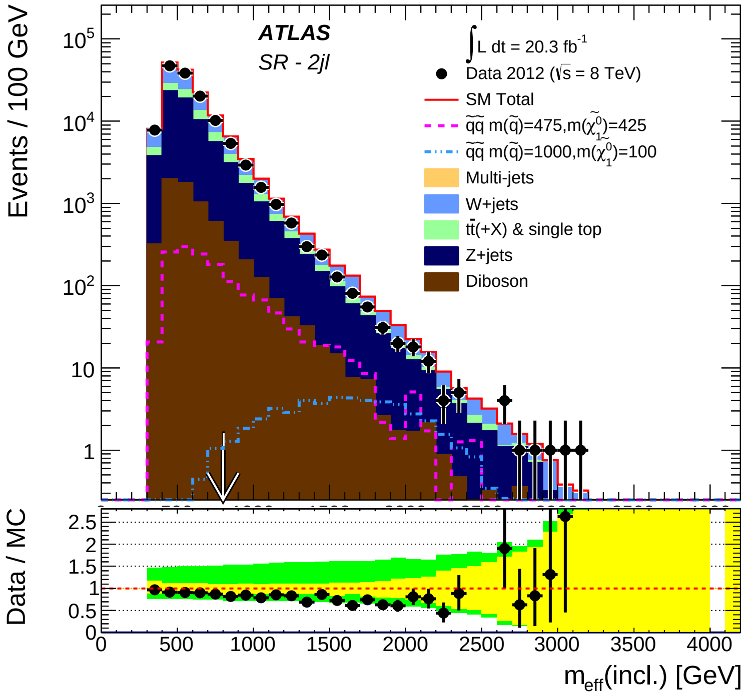


Fig. 8.1 The $m_{\rm eff}$(incl.) distribution for the SR 2j with no $m_{\rm eff}$(incl.)-cut applied. The arrow shows the optimal value of the $m_{\rm eff}$(incl.) cut. The histograms denote the MC background expectations, normalised to the integrated luminosity of the data. In the lower panels the light (yellow) error bands denote the experimental systematic and MC statistical uncertainties, while the medium dark (green) bands also include the theoretical modelling uncertainty. This pattern of depicting systematic uncertainties is kept in the whole thesis.

models is somewhere in between the two benchmark models that are shown. The same distribution for the W SRs is depicted in Fig. 8.2. The requirement of two Ws in the final state is responsible for the lack of statistic, but an evident discriminating power of the $m_{\rm eff}$-cut is present also for these two SRs. Different benchmark models are used in this plot, since these SRs have been designed for 1-step models containing W boson in the final state.

The event yields in all SRs, which are considered as the main result of this analysis, are shown in Fig. 8.3. The event yields of the expected SM background in the SR are calculated by the fit with the method discussed in Chapter 6.

The systematic uncertainties placed on the objects discussed in Chapter 7 affect the event yields in the CRs and VRs which further impact the TFs (see Eq. 6.5). To quantify that, the original object is transformed within the given systematic uncertainty and the effect on CRs, VRs and TFs is evaluated together with the correlation matrix, monitoring the relevance of the systematic uncertainty with respect to the others. The resulting shift serves as the Δ, the input to the fit described in Section 6.3. The usage of Eq. 6.5 and the fit procedure

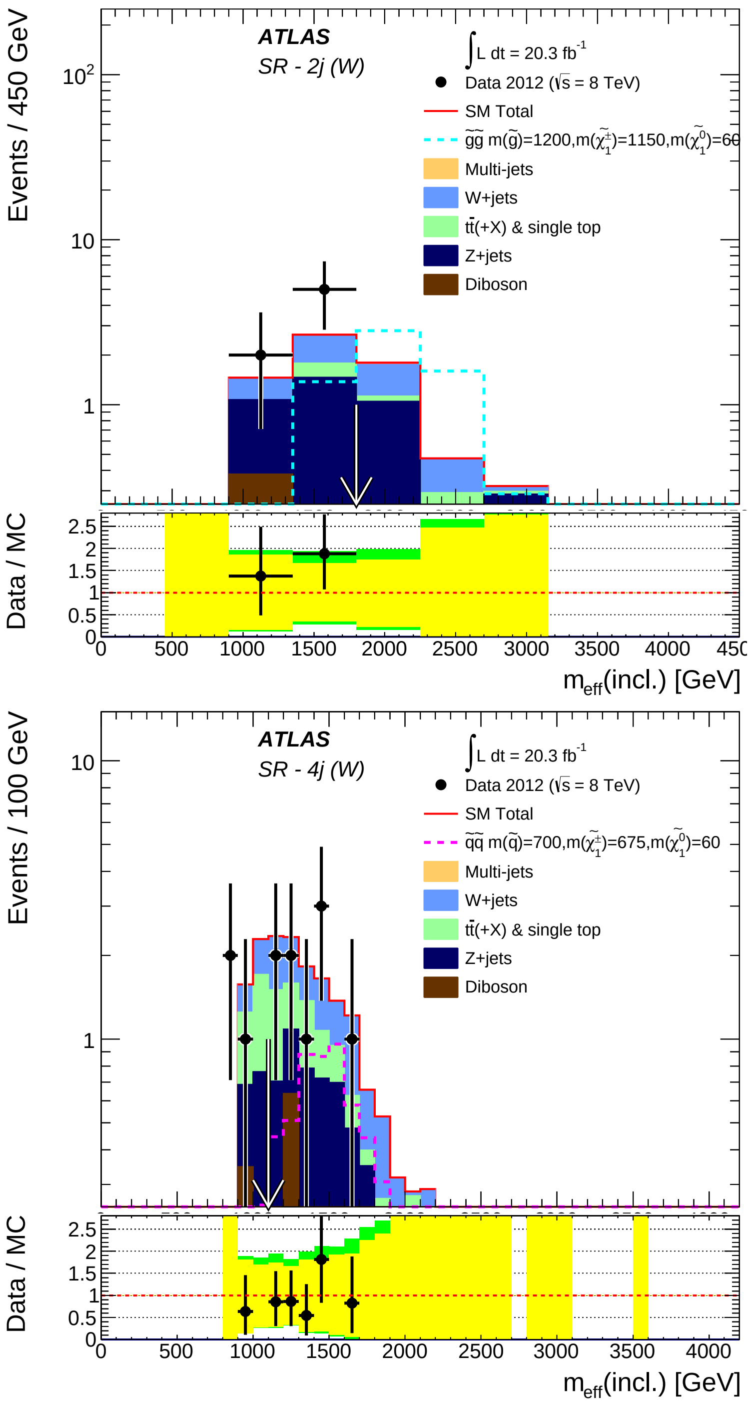


Fig. 8.2 The m_{eff}(incl.) distribution for the SR 2j(W) and SR 4j(W). The arrows indicate the optimal values of the m_{eff}(incl.) cut.

Cut	Channel														
	Description	2j			3j	4j					5j	6j			
	2jl	2jm	2jt	2j(W)		4j(W)	4jl-	4jl	4jm	4jt		6jl	6jm	6jt	6jt+
Lepton veto	No selected e/μ after overlap removal with $p_T > 10$ GeV.														
E_T^{miss} [GeV] $>$	160														
$p_T(j_1)$ [GeV] $>$	130														
$p_T(j_2)$ [GeV] $>$	60														
$p_T(j_3)$ [GeV] $>$	–				60	40	60				60	60			
$p_T(j_4)$ [GeV] $>$	–					40	60				60	60			
$p_T(j_5)$ [GeV] $>$	–						–				60	60			
$p_T(j_6)$ [GeV] $>$	–						–				–	60			
W candidates	–			$2\ W \rightarrow j$ 60<m(W)<100 GeV	–	$W \rightarrow j +$ $W \rightarrow jj$ 60<m(W)<100 GeV	–								
$\Delta\phi(j_{1,2,(3)}, E_T^{miss}) >$	0.4														
$\Delta\phi(j_{i>3}, E_T^{miss}) >$	–					0.2									
$E_T^{miss}/\sqrt{H_T} >$	8	15	15				10	10							
$E_T^{miss}/m_{eff}(Nj) >$				0.25	0.3	0.35			0.4	0.25	0.2	0.2	0.2	0.25	0.15
m_{eff}(incl.) [GeV] $>$	800	1200	1600	1800	2200	1100	700	1000	1300	2200	1200	900	1200	1500	1700

Table 8.1 The SRs used in the analysis. See text for the definitions of the quantities used in the cuts and their motivations. The channel labels 'l-', 'l', 'm', 't', 't+' correspond to 'very loose', 'loose', 'medium', 'tight' and 'very tight' m_{eff} selections respectively. Special regions 2j(W) and 4j(W) are dedicated to the $\tilde{g}\tilde{g}$-1step and $\tilde{q}\tilde{q}$-1step models respectively.

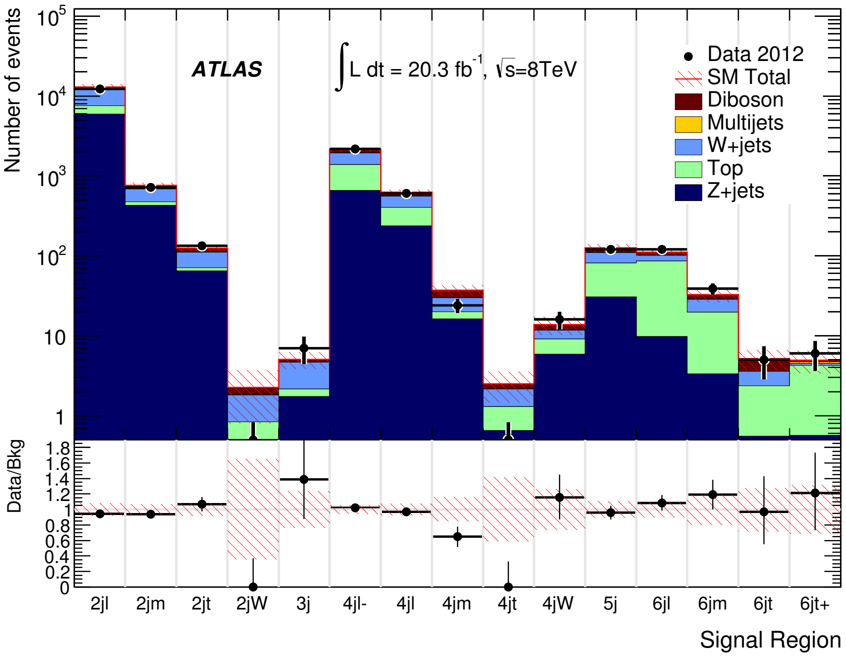


Fig. 8.3 The event counts in all SRs. The background is determined by the fit of the simulated MC events to the measured data.

affect the systematic uncertainties such that they can differ from the initial values Δ. The final systematic uncertainties are listed in Table 8.2. Note that the systematic uncertainties are not always treated as correlated and thus their quadratic sum does not necessarily correspond to the total error reported in the third row. The breakdown of the uncertainties for the other SRs of the 0-lepton analysis can be found in [64].

Channel	SR 2j(W)	4j(W)
Total bkg.	2.3	14
Total bkg. unc.	±1.4[61%]	±4 [29%]
CR stats Z/γ^*+jets	±0.4 [17.4%]	1.3 [9.3%]
CR stats W+jets	±0.7 [30.4%]	±1.0 [7.1%]
CR stats top quark	±0.35 [15.2%]	±0.5 [3.6%]
MC statistics	±0.34 [14.8%]	±0.7 [5.0%]
Jet/MET	±0.27 [11.7%]	±0.6 [4.3%]
Leptons	±0.04 [1.7%]	±0.06 [0.4%]
Z/γ-TFs	±0.028 [1.2%]	±0.5 [3.6%]
Theory Z/γ^*+jets	±0.03 [1.3%]	±1.3 [9.3%]
Theory W+jets	±0.1 [4.3%]	±0.9 [6.4%]
Theory top quark	±0.9 [%39.1]	±2.8 [20.0%]
Theory diboson	±0.2 [8.7%]	±1.0 [7.1%]
Theory scale unc.	±0.13 [5.7%]	±0.12 [0.9%]
Multi-jets method	±0.04 [1.7%]	–
Others	±0.24 [10.4%]	±0.12 [0.9%]

Table 8.2 The breakdown of the systematic uncertainties for the background estimates in the two W SRs.

In comparison to other SRs, the W-SRs (especially 2j(W)) suffer from low statistics and thus the statistical errors in the CRs dominate. The looser the selection, the lower the effect of the statistical uncertainty and thus the lower the total uncertainty is. Apart from the statistical uncertainty (referred to in Table 8.2 as CR stats), the theoretical uncertainty placed on the $t\bar{t}$/single-t background is one of the dominating ones. This is due to the similar topology of the $t\bar{t}$ events to our signal models, which makes $t\bar{t}$/single-t one of the major backgrounds in the W-SRs. The initial JMS and JMR uncertainties have been set to 10% and 20% respectively, see Chapter 9. The JMS, JMR, JES and JER uncertainties are all included in the term Jet/MET and their impact, decreased by the usage of TFs, is not so large with respect to the other uncertainties. The uncertainties quoted as "Others" in Table 8.2 contain minor terms as the pile-up effect on MC simulation or b-tagging.

Fig. 8.3 shows that there is no unexpected deviation between the measured data and the background estimation from the simulation. Thus, the exclusion fit can be reliably used for the signal models which this analysis was optimized for. The statistical apparatus discussed in Section 6.3 has been used to calculate the CLs for each signal model. For the $\tilde{q}\tilde{q}/\tilde{g}\tilde{g}$-direct simplified models, only two free parameters appear, namely $m_{\tilde{g}} \vee m_{\tilde{q}}$ and $m_{\tilde{\chi}_1^0}$. In such a case, all of the CLs can be placed in one graph with the two parameters on the axes and a contour line can be drawn through the points corresponding to 95% CL. The models represented by points inside such a curve are excluded by this analysis. The exclusion contours and the strongest SR at each point for the direct simplified models are shown in Fig. 8.4. The exclusion contours of these models have not been derived by me but they play an essential role for the whole analysis. However, supplementary studies on background reduction have been done by myself for these models, see Appendix B. The contours in Fig. 8.4 show that the excluded mass of the $\tilde{q}$ or $\tilde{g}$ depends on the mass of the $\tilde{\chi}_1^0$. Furthermore, the compressed scenarios (i.e. the models where $m_{\tilde{q}} \vee m_{\tilde{g}} \sim m_{\tilde{\chi}_1^0}$) are difficult to exclude at all. To determine the final limit by the ATLAS experiment on the $m_{\tilde{q}}$ or $m_{\tilde{g}}$, the result provided by this analysis needs to be combined with other ATLAS analyses.

The case of the 1step models (e.g. $\tilde{g} \to \tilde{\chi}_1^{\pm} q\bar{q} \to \tilde{\chi}_1^0 W q\bar{q}$) is complicated by three free parameters, namely $m_{\tilde{g}} \vee m_{\tilde{q}}$, $m(\tilde{\chi}_1^{\pm})$ and $m(\tilde{\chi}_1^0)$. In Fig. 8.5, the mass of the $\tilde{\chi}_1^0$ is fixed at 60 GeV. Although the analysis is performed with this constraint, the exclusion limits are assumed to be valid even for slightly different values of $m_{\tilde{\chi}_1^0}$. The two remaining mass parameters are used in the exclusion plots on the axes. The x parameter of the form $x = \Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0)/\Delta m(\tilde{g}/\tilde{q}, \tilde{\chi}_1^0)$ is used for the y-axis. A value of $x \sim 1$ corresponds to the compressed scenario for the 1step model, i.e. cases where $m_{\tilde{\chi}_1^{\pm}} \sim m_{\tilde{g}} \vee m_{\tilde{q}}$. The W SRs evidently reach the best performance for the compressed scenarios that is a great contribution of this thesis to the 0-lepton analysis. In order to investigate the compressed scenarios, extra points had to be produced to gain sufficient information in the relevant region of the parametric space (for technical details see Appendix E).

A different approach is to set the ratio $x = \Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0)/\Delta m(\tilde{g}/\tilde{q}, \tilde{\chi}_1^0)$ to a specific value, 0.5 in this case. As can be seen in Fig. 8.6, the W SRs do not perform well for these models, since there is no sufficient boost provided to the Ws in the $x = 0.5$ scenario.

Although the 0-lepton SUSY analysis is concentrated on SUSY extensions of the SM by the optimization of the SRs, this analysis can be, besides SUSY, sensitive also to other physics models beyond SM. The minimal Universal Extra Dimensions (mUED) [78] has been investigated by our analysis team at the University of Freiburg. The exclusion limit contour depicted in the plane of its free parameters $R \cdot \Lambda$ and $1/R$ is given in Fig. 8.7. Although the mUED is dominated by multi-leptonic final states, the 0-lepton analysis provides competitive results to the di-lepton, three-lepton and two same sign lepton ATLAS analyses, see [79]. The combined exclusion limit of the three lepton analyses mentioned above reaches approximately 900 GeV

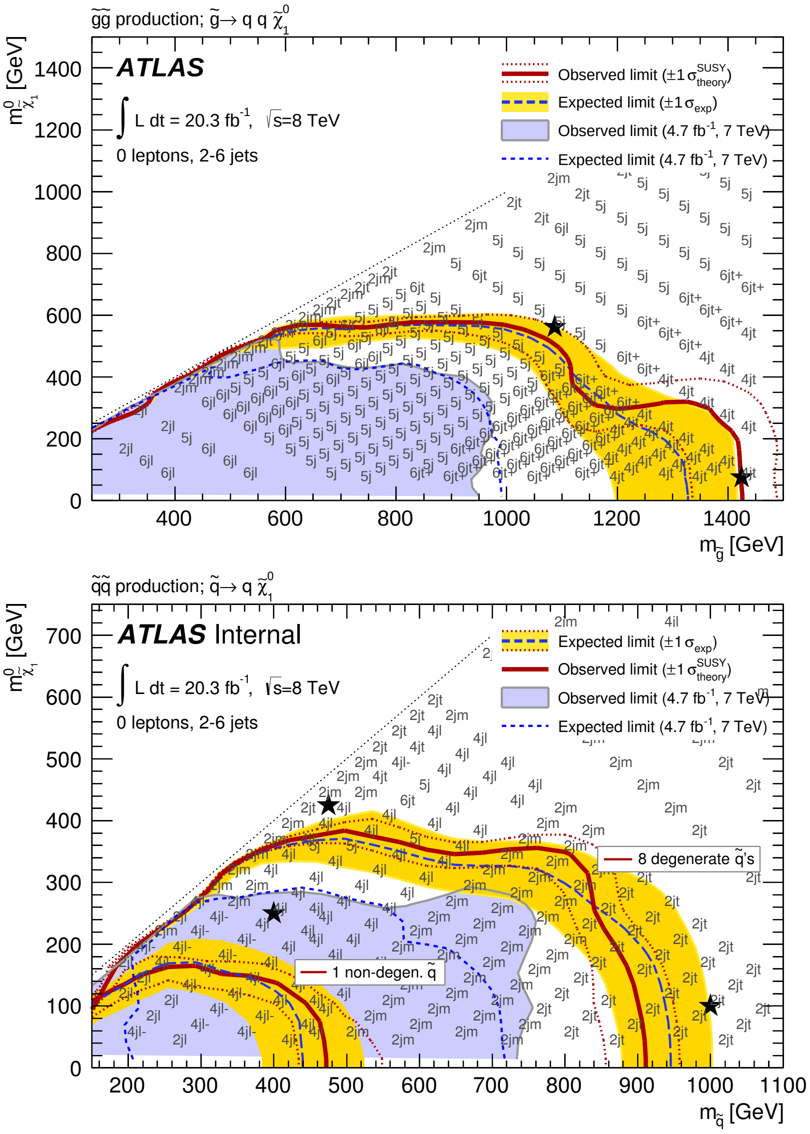


Fig. 8.4 The exclusion limits for the $\tilde{g}\tilde{g}$-direct signal model (upper plot) and $\tilde{q}\tilde{q}$-direct signal model (bottom plot). The red line denotes the observed limit with doted lines as 1σ band. The blue line with the yellow 1σ band corresponds to expected exclusion limit.

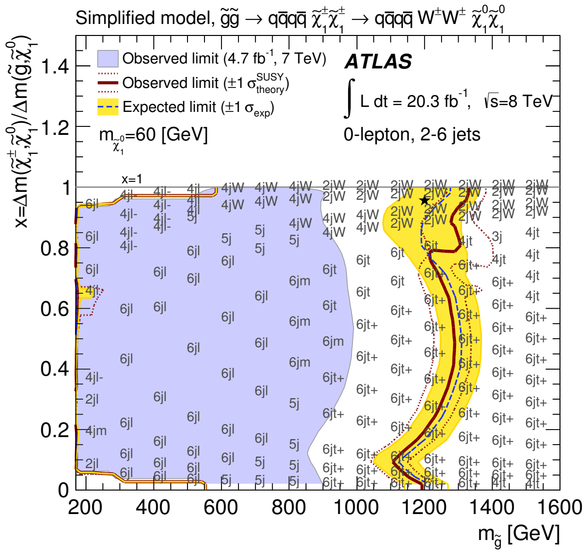


Fig. 8.5 The exclusion limits for the $\tilde{g}\tilde{g}$-1step signal model on the left and the $\tilde{q}\tilde{q}$-1step signal model on the right. The mass of $\tilde{\chi}_1^0$) is fixed to 60 GeV.

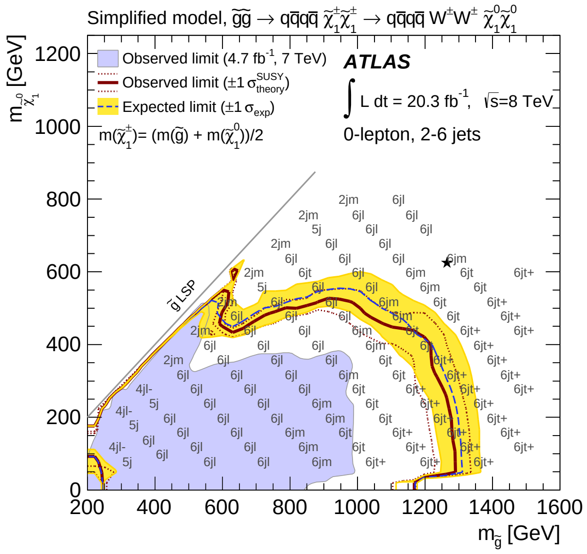


Fig. 8.6 The exclusion limits for the $\tilde{g}\tilde{g}$-1step signal model on the left and the $\tilde{q}\tilde{q}$-1step signal model on the right. The parameter $x = \Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0)/\Delta m(\tilde{g}/\tilde{q}, \tilde{\chi}_1^0)$ is fixed to 0.5.

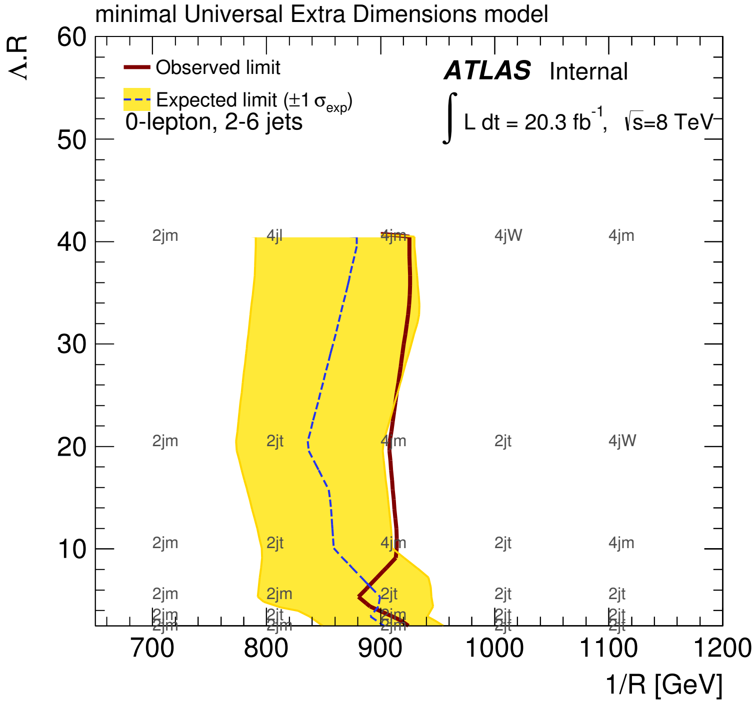


Fig. 8.7 The exclusion limits for the MUED model.

in the $1/R$ parameter for all values of $R\cdot\Lambda$. A more detailed explanation can be found in thesis [80].

9. Tagging of Boosted W Bosons

The schematic diagrams of the $\tilde{g}\tilde{g}$-1step and $\tilde{q}\tilde{q}$-1step simplified decay chains, see Fig. 5.4, contain two W bosons in the final state accompanied by other jets originating from the produced quarks. The number of accompanying jets produced in the $\tilde{q}\tilde{q}$-1step is expected to be lower due to only one daughter quark being produced in the decay of $\tilde{\chi}_1^{\pm}$. The Ws can decay into leptons, but such decays are suppressed by the lepton veto, applied as a baseline cut of this analysis. The W bosons decaying into hadrons (68 %, [29]) can be identified by the invariant mass of the jets produced by these decays, which must have a value close to the mass of the W. Furthermore, the only background with two real W bosons are $t\bar{t}$ and Diboson processes. The rest of the background components are expected to be largely suppressed, when requiring two Ws in the final state. The boost of the W can also provide additional separation power between the signal and the background, as the Ws are expected to be more boosted for some points in the signal parametric space. The aim of the study introduced in this chapter is to increase the sensitivity of the 0-lepton analysis to the 1step simplified models. The final results, i.e the exclusion limits depicted in the parametric space for both simplified models, were already shown in Fig. 8.5.

9.1 Boost of the W

The quarks originating from the decay of the Ws can produce either di-jet or single-jet systems. The number of jets depends on the boost of the W and on the jet cone size in the anti-$k_{\perp}$ jet algorithm used. When $p_{\mathrm{T}}(W)$ is smaller or comparable to the mass of the W, the opening angle between the daughter quarks is rather large. Consequently, the two jets are visible separately in the detector. This di-jet system is referred to as a **resolved** W. In the opposite case, when the momentum of the W is large and two quarks fall into the same jet cone, the W is labelled as **unresolved**. Both cases are schematically depicted in Fig. 9.1.

The balance between unresolved and resolved cases is strongly dependent on the boost of the W, i.e. on the point in the signal parametric space. The two $\tilde{g}\tilde{g}$-1step models with the parameters: $m_{\tilde{g}} = 1200$ GeV, $m_{\tilde{\chi}_1^{\pm}} = 360/1060$ GeV and $m_{\tilde{\chi}_1^{0}} = 60$ GeV have been used to

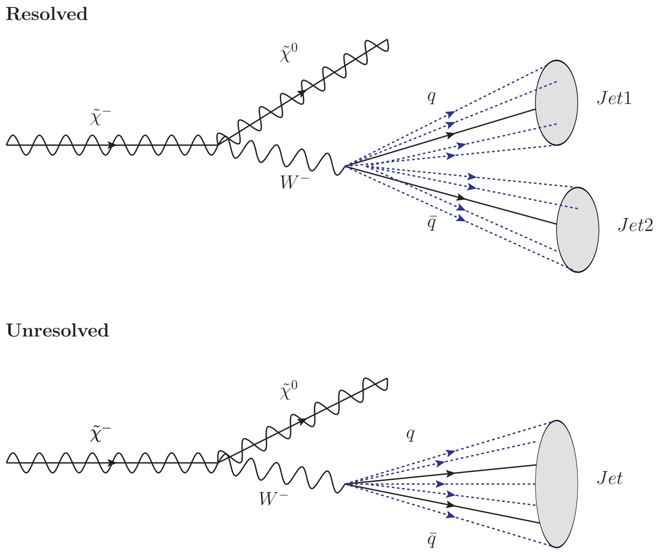


Fig. 9.1 The schematic diagrams of the decay of $\tilde{\chi}_1^\pm$ into W bosons. The resolved case in the upper scheme produces two distinguishable jets in the detector. The unresolved case in the lower scheme has a highly boosted W, and its products can not be resolved.

draw the distribution of the opening angle between two quarks coming from the same W, see Fig. 9.2. The angle is represented by the ΔR on the y axis and its dependence on the $p_\mathrm{T}(W)$ on the x axis is shown. The truth information at generator level (no detector simulation) about the properties of individual particles is used for these purposes. The following conclusions can be derived from these figures:

- The ΔR is less than 0.4 for Ws with $p_\mathrm{T} > 400$ GeV.
- $m_{\tilde{\chi}_1^\pm} = 1060$ GeV: the p_T of the W is peaking at $\sim$500 GeV.
- $m_{\tilde{\chi}_1^\pm} = 360$ GeV: the p_T of the W is peaking below 300 GeV.

The first conclusion is especially relevant for the anti-$k_\perp$ with $\Delta R = 0.4$, which is used in the published results. The signal models will be dominated by the unresolved W when they contain most of the Ws with $p_\mathrm{T} > 400$ GeV. The other two conclusions refer to the two discussed $\tilde{g}\tilde{g}$-1step models. The models with small difference between $m_{\tilde{\chi}_1^\pm}$ and $m_{\tilde{g}}$ (but large $m_{\tilde{g}}$ at the same time) will be most likely dominated by the unresolved cases. One could ask, why the boost of the W is larger when $m_{\tilde{\chi}_1^\pm} \sim m_{\tilde{g}}$: the reason is that in this case, most of the energy from the $\tilde{g}$ is transferred to the $\tilde{\chi}_1^\pm$ and then almost equally split between W and $\tilde{\chi}_1^0$. In the case of small $m_{\tilde{\chi}_1^\pm}$: the $\tilde{\chi}_1^\pm$ shares the energy from the $\tilde{g}$ with the quarks and thus only a smaller amount of energy remains for the $\tilde{\chi}_1^0$ and the W.

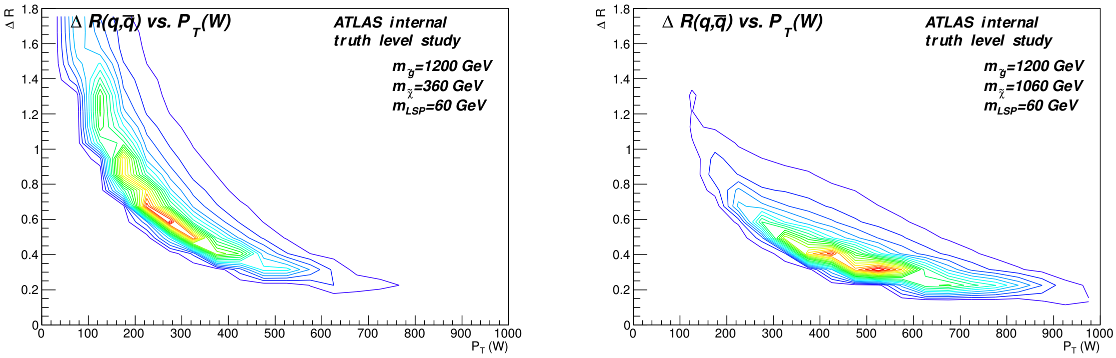


Fig. 9.2 The $\Delta R(q,\bar{q}')$ dependence on the W's p_{T} is depicted for two points of the $\tilde{g}\tilde{g}$-1step grid.

9.2 Reconstruction of the Ws

Two specific signal models, often referred to as signal points, are investigated in this chapter. The signal point with $m_{\tilde{g}} = 1200$ GeV, $m_{\tilde{\chi}_1^{\pm}} = 1150$ GeV and $m_{\tilde{\chi}_1^0} = 60$ GeV in the case of the $\tilde{g}\tilde{g}$-1step model, and the point with $m_{\tilde{q}} = 700$ GeV, $m_{\tilde{\chi}_1^{\pm}} = 675$ GeV and $m_{\tilde{\chi}_1^0} = 60$ GeV in case of $\tilde{q}\tilde{q}$-1step model. These benchmark points are picked in order to achieve a large boost of the W in the final state. Note that these points are chosen so that they lay on the border between "excluded" and "not-excluded" on the exclusion limits plots.

The motivation to investigate the boson-tagging in the framework of the 0-lepton analysis originally arose with the distribution of the invariant mass of the leading jet. A very clear peak is visible around m_W for some of the signal samples of the 1-step simplified models, namely for those where large energy is available for the W. The mass distribution published in [64] is shown in Fig. 9.3 on the left. The selection of the events populating this histogram is the same as in the column labelled 2j(W) in Table 8.1, except for the cuts on the jet invariant mass introduced to reconstruct the W (the row labelled W candidates). This mass distribution contains all jets with $p_{\mathrm{T}} > 40$ GeV. The signal samples contain boosted Ws, which are represented by a peak in the range 60-100 GeV (white arrows in Fig. 9.3). The rest of the jets, corresponding to the other quarks, peak at lower values of the invariant mass. A similar distribution corresponding to the selection in the column labelled 4j(W) in Table 8.1 is shown in Fig. 9.3 on the right.

Considering the information gained from the mass distribution for the 2j(W) SR, the first unresolved W can be found by looping over the set of the jets searching for the jet within the mass range 100-60 GeV. Both signal samples, $\tilde{g}\tilde{g}$-1step and $\tilde{q}\tilde{q}$-1step, contain exactly two Ws in the final state. Thus, the invariant mass peak is still expected when the jet corresponding to the W already tagged is removed from the set of the jets we are looping over. The mass distribution

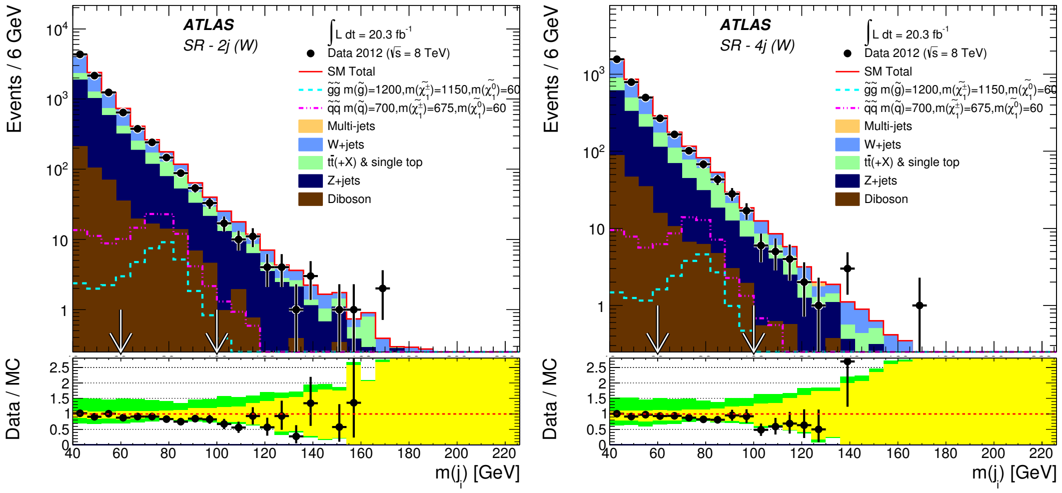


Fig. 9.3 The mass distributions of the jets after the selections corresponding to the SRs used in the publication: 2j(W) on the left, 4j(W) on the right. A suggested mass window cut is indicated by the white arrows.

of all the remaining jets, with the condition that one unresolved W has been already found, is shown in Fig. 9.4. This distribution shows a peak for the signal around the W mass as well, thus options of SRs with the two unresolved Ws were tested in the case of 2j SR.

In the case of 4j(W) SR, the resolved type of W was investigated when reconstructing the second W candidate. After removal of the jet corresponding to the first W (same as for 2j(W)), the set of remaining jets (still ordered in p_{T}) is examined, looking for pairs of close-by jets. The invariant mass of all such combinations is shown in Fig. 9.5. The selection was done according to the 4j(W) SR except for the mass window applied to the second W, see Table 8.1. The peak in this distribution occurs for the signal samples at m_W as expected.

Given the results discussed in the previous paragraphs, the following procedure for W tagging is established:

- The procedure loops over the set of jets with $p_{\mathrm{T}} > 40$ GeV ordered in p_{T} to search for jets within the mass window of 60-100 GeV.
- Always, when such a jet is found, it is labelled as an unresolved W and it is removed from the set of jets for the further steps.
- The set of the remaining jets is investigated in order to search for resolved W candidates. The closest jet is found to each jet and their invariant mass is calculated. If the mass of such a pair is within 60-100 GeV, the pair of jets is labelled as resolved W.
- The procedure terminates when all the jets have been investigated. The final cuts applied in the SR are placed on the $N_{\mathrm{res}-W}$ or $N_{\mathrm{unres}-W}$ returned by this procedure.

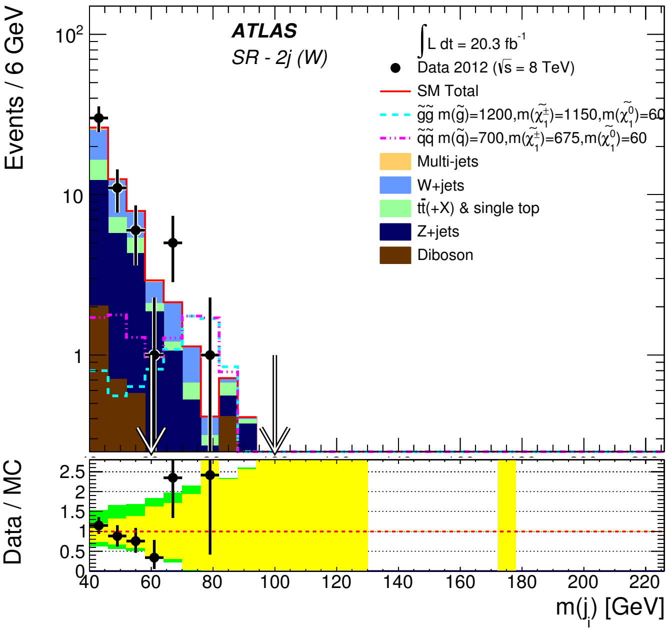


Fig. 9.4 The mass distribution of the jets after one unresolved W candidate is already found. The selection corresponding to 2j(W) is applied.

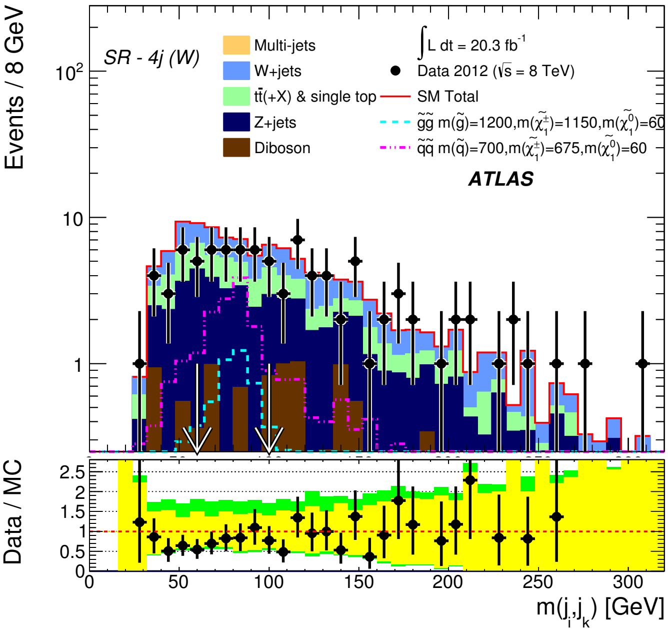


Fig. 9.5 The invariant mass distribution of the jet pairs after one unresolved W candidate is already found. The pair is made out of a jet and its closest jet. The selection corresponding to 4j(W) is applied.

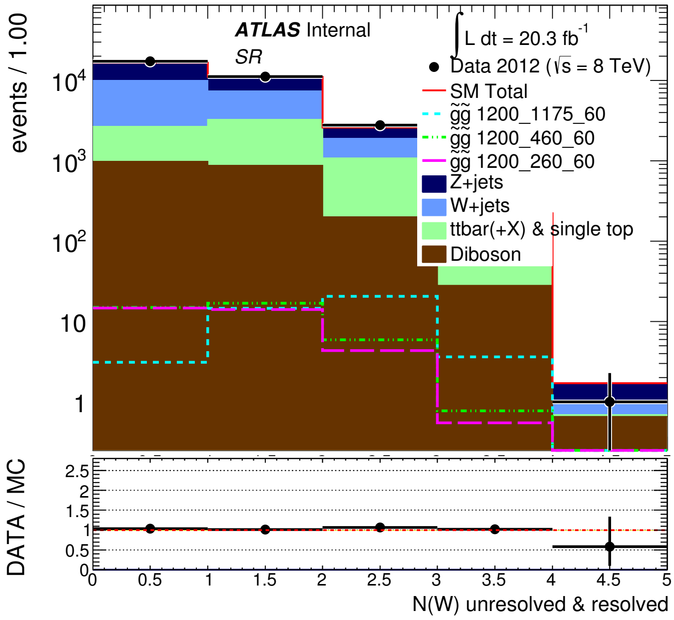


Fig. 9.6 The final count of Ws found with the W-finder.

This procedure is referred to as "W-finder" in the following. The unresolved Ws are preferred in this procedure because: 1) A jet with such a large invariant mass is most likely a real W while the resolved case suffers from large combinatorial background; 2) The unresolved W is more rare than the resolved W. The building of pairs of jets in this procedure does not suffer from double-counting, as the jets recognized as belonging to a W are always removed from the internal list of the jets. Other methods of pair building (e.g. different ordering of the jets or different approach to the second jet in the pair) have been tested with comparable or worse results. An optional SR built from two resolved Ws is discussed in Section 9.3.

The number of either resolved or unresolved Ws found by W-finder is depicted in Fig. 9.6. This figure was done in an earlier phase of the analysis with the loosest selection of the 2jl SR, see Table 8.1. There are three signal points of the $\tilde{g}\tilde{g}$-1step grid. The distribution for the signal point with the masses: $m_{\tilde{g}} = 1200$ GeV, $m_{\tilde{\chi}_1^\pm} = 1175$ GeV and $m_{\tilde{\chi}_1^0} = 60$ GeV has a maximum at 2 reconstructed Ws, which is in agreement with expectations.

The efficiency of the W-tagging has been studied for two points in the $\tilde{g}\tilde{g}$-1step parametric space, see Fig. 9.7. However, the p_{T} of the reconstructed W has a large impact on the efficiency, more than the specific decay chains of the tested sample. The efficiency is plotted with respect to the p_{T} of the true W. The W is assumed to be reconstructed correctly if it lays within $\Delta R < 0.2$ from the truth W. The efficiency is calculated as the number of correctly reconstructed Ws of either resolved or unresolved type divided by total number of Ws in the samples. The efficiency reaches almost 80% for anti-$k_{\perp}$ with $R = 0.4$ for large p_{T} of the Ws.

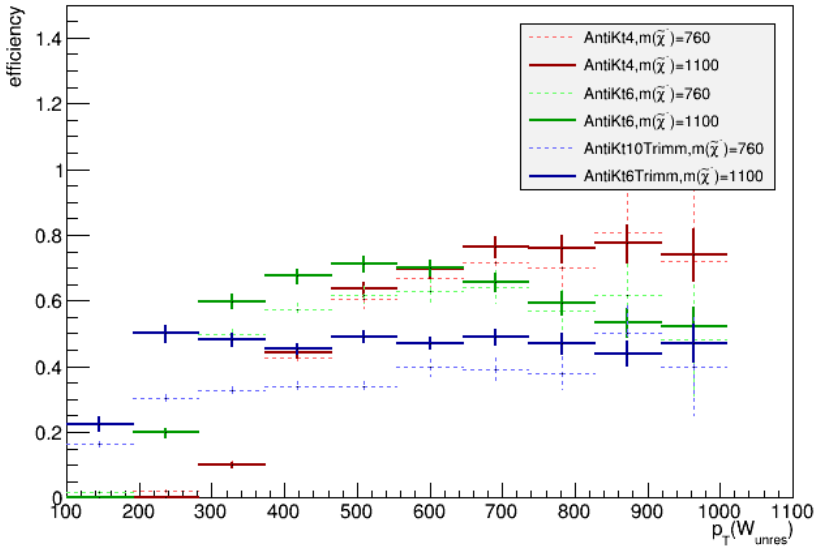


Fig. 9.7 The W reconstruction efficiency for several anti-$k_\perp$ algorithms. The label Trimm refers to trimmed jets where the trimming conditions have been applied, see chapter 10. The signal models are specified by the $m_{\tilde{\chi}_1^\pm}$ in top right legend, the other two mass parameters are fixed: $m_{\tilde{g}} = 1200$ GeV and the $m_{\tilde{\chi}_1^0} = 60$ GeV. The error bars are calculated only from the statistical uncertainty.

However, in the lower p_{T} range the larger cone size anti-$k_\perp$ jets performs better. This can be explained by dependence of the balance between the resolved and unresolved cases on the cone size of the anti-$k_\perp$ algorithm. The jets with larger cone size are expected to form an unresolved W more frequently, and thus the efficiency increases as it does not suffer from the combinatorial background. This observation can also be considered as a motivation for the reclustering studies described in Chapter 10.

9.3 Optimization of the W Signal Regions

The optimization of the SR proceeds in a simple way:

- A set of cuts is chosen, defining a candidate SR.
- The expected exclusion line is plotted for all SR candidates.
- Only the SRs with the best performance are chosen.

The variables tested during the optimization are $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ with cut values 0.15, 0.2, 0.25, 0.3, 0.35 and 0.4 [$\sqrt{\mathrm{GeV}}$] and the $m_{\mathrm{eff}}(\mathrm{incl.})$ with cut values 700, 800, 1000, 1200, 1400,

1600, 1800 and 2000 [GeV]. The SRs candidates with two or more Ws of the resolved type, restricting the presence of the unresolved W, were checked without any beneficial result. The selection of the SRs from the last bullet proceeds in such a way, that they do not overlap between themselves, i.e. if there are two SRs with roughly the same exclusion power over all of the analysed SUSY models, only one of them is used. Fig. 9.8 shows an example of this procedure for $\tilde{g}\tilde{g}$-1step and $\tilde{q}\tilde{q}$-1step simplified models and some of the most promising SRs containing W(s). These figures were prepared at the stage of the analysis where not all of the systematic uncertainties were considered as in the publication [64]. The two W SRs listed in Table 8.1 resulted from the optimization described above. A significant improvement is achieved in the compressed regions: $x \in (0.8, 1.0)$. Hypothetical SRs built from two resolved W would not improve our exclusion limits and thus they are not involved in the publication.
Note that this method of optimization is very CPU-intensive, thus not many of the SR options can be tested by running the entire statistical procedure. One can however roughly estimate the optimal cuts by checking the distribution of the quantity on which the cut is placed.

9.4 Systematic Uncertainties in the W SRs

In addition to the usual set of systematics uncertainties used in each SR, the jet mass scale (JMS) and the jet mass resolution (JMR) uncertainties affecting the invariant mass of the W (jet or di-jet system) were used in case of 2jW and 4jW SRs. These uncertainties have to be considered in these SRs due to the invariant mass cut placed on the jets. Note that there were no officially released uncertainties or prescriptions on how to deal with them to the date of the writing this thesis. So this work can be considered as pioneering in this field within ATLAS. In the published analysis, the JMS was fixed to 10% and JMR to 20%. Technically, this means that the masses of the jets were coherently shifted up and down by 10% with respect to the nominal value in the case of JMS and randomly smeared with a Gaussian distribution by 20% of the nominal to simulate the JMR uncertainty. All the rest of the kinematic properties of the jets remained unchanged. The effect of such a transformation of the jet properties on the event yield across the entire statistics defines the resulting uncertainty, propagated to the likelihood in the usual manner as a nuisance parameter.
The comparison between the shape of the W invariant mass peak for the data and simulation has been studied to support the conservativeness of the fixed values of the JMS and JMR, see Fig. 9.9. The events from $t\bar{t}$ production have been selected for these purposes since the W appears as product of top decay together with a b quark and b-tagging is highly efficient in suppressing the background.

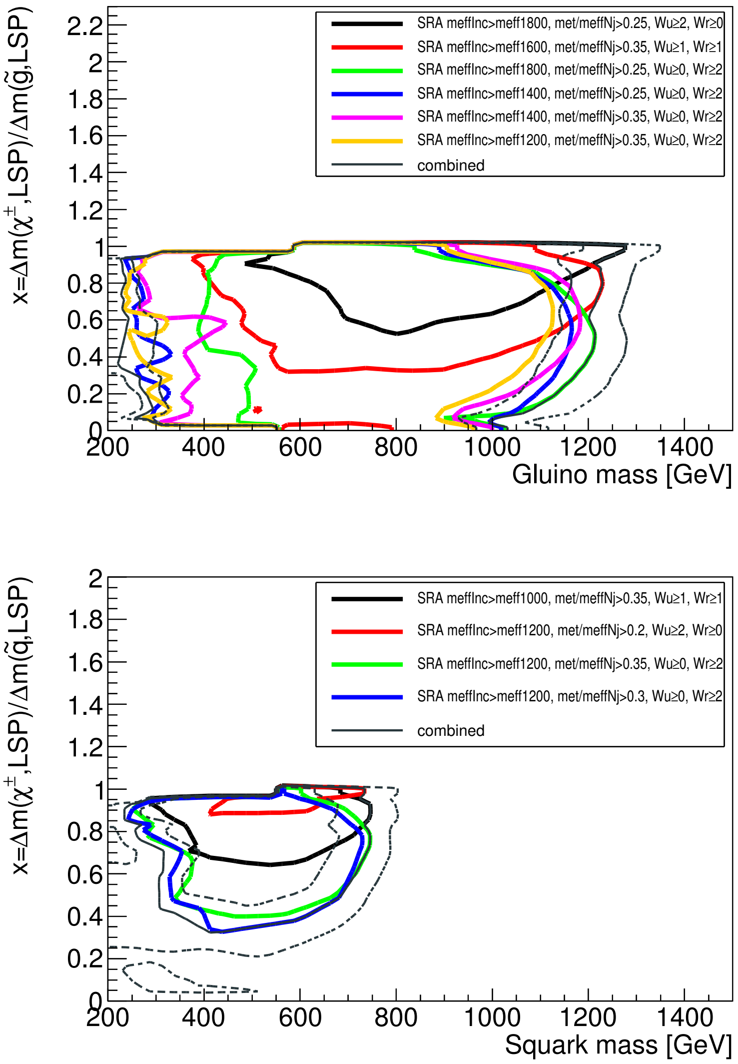


Fig. 9.8 The optimization of the boosted *W* SRs for the two simplified models: $\tilde{g}\tilde{g}$-1step on the upper plot, $\tilde{q}\tilde{q}$-1step on the bottom plot.

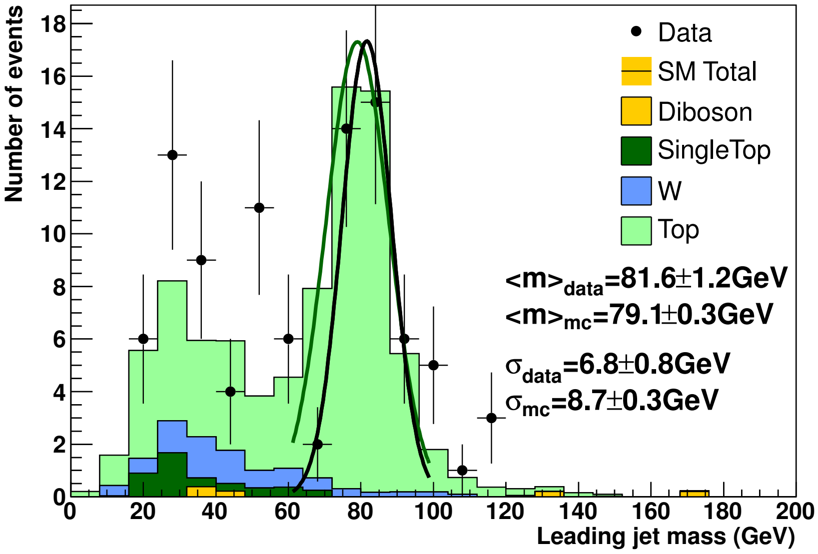


Fig. 9.9 The Gaussian fit of the W mass peak for the simulation and data.

The requirements used to select $t\bar{t}$ are:

- exactly one lepton (electron or muon)
- exactly three jets with $p_{\mathrm{T}} > 40$ GeV
- $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV
- a leading jet with $p_{\mathrm{T}} > 400$ GeV
- the leading jet is not a b-tagged jet
- $\Delta\phi$ (leading jet, b-jet)>1.0

Despite the fact that the *W* mass peak position and width show good agreement between data and simulation, the JMS and JMR were fixed to very conservative values. Their impact on the final contour line of the exclusion limit is illustrated in Fig. 9.10, where the contour lines for several values of the JMS uncertainty are depicted in the parametric space of the $\tilde{g}\tilde{g}$-1step simplified model. The statistical uncertainty in the SR 2j(W) dominates with respect to the systematic ones, thus even very conservative values of the JMS and JMR do not change the position of the contour line.

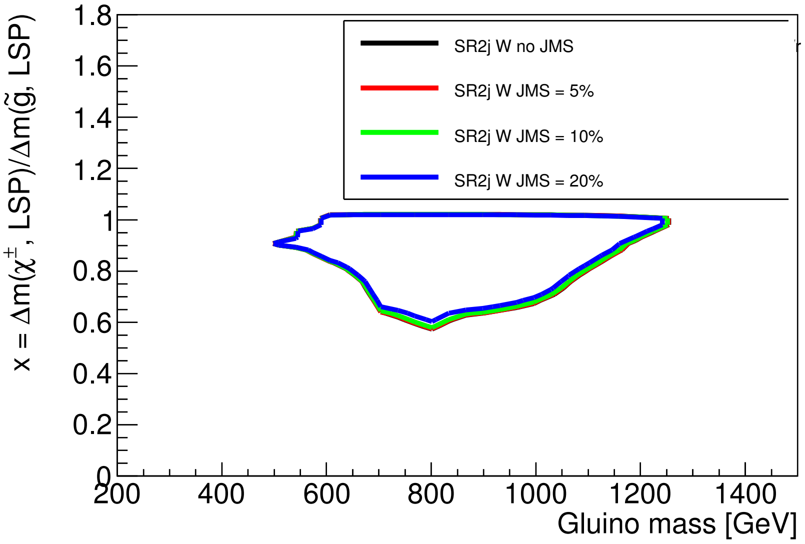


Fig. 9.10 The contour lines for several fixed values of the JMS uncertainty, which has clearly only small effect on the results.

9.5 Other Applications of W-tagging

9.5.1 M_{T2}, Enhanced Transverse Mass

The 1step simplified models studied here contain a decay $\tilde{\chi}_1^{\pm} \rightarrow W^{\pm}\tilde{\chi}_1^0$ in each of the $\tilde{g}$ branches. The aim of this chapter is to approximate the $\tilde{\chi}_1^{\pm}$ mass from this decay with the M_{T2} variable and investigate the potential of this variable to discriminate the signal from the background.

After the successful W reconstruction, the four-vector of the W is known. Additionally, the transverse part of the four-vector of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is measured as well. However, only the transverse information is not enough to reconstruct the mass of the particle decaying in the particles contributing to $E_{\mathrm{T}}^{\mathrm{miss}}$ (i.e. whenever a $\tilde{\chi}_1^0$ or a ν occurs, the masses of the particles deeper in the decay chain can not be reconstructed). This is the case of the $\tilde{\chi}_1^{\pm}$. Although it is not feasible to reconstruct the invariant mass of the $\tilde{\chi}_1^{\pm}$, the transverse part of it can be approximated with M_{T2} defined as:

$$M_{\mathrm{T2}} = \min_{p_1+p_2=\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}} [\max[m_{\mathrm{T}}^2(W_1,p_1), m_{\mathrm{T}}^2(W_2,p_2)]] \tag{9.1}$$

where p_1 and p_2 are free four-vectors in the transversal plane with a condition on their four-vector sum: $p_{1\mathrm{T}} + p_{2\mathrm{T}} = \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ (only the transverse part). W_1 and W_2 correspond to the four-

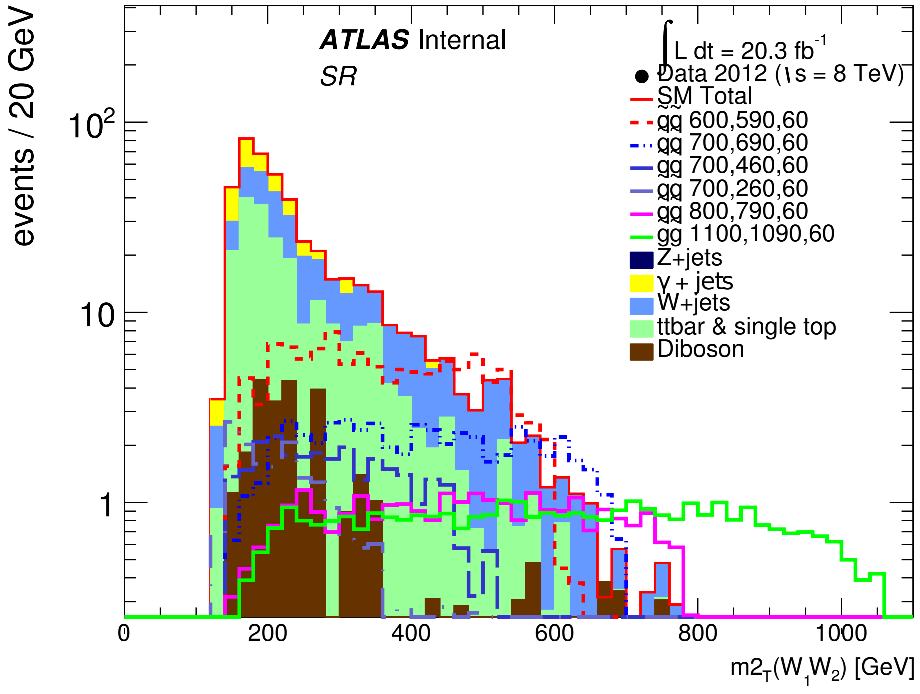


Fig. 9.11 The M_{T2} distribution calculated from $E_{\mathrm{T}}^{\mathrm{miss}}$ and the pair of reconstructed Ws.

vectors of the reconstructed Ws. The mass components of the p_1 and p_2 are free parameters, set to 60 GeV for this study. The distribution of this variable is shown in Fig. 9.11. In addition to all pre-selection criteria, the following selection has been used: $p_T(j_1) > 130$ GeV, $p_T(j_2) > 60$ GeV, $E_T^{miss} > 160$ GeV. Several $\tilde{g}\tilde{g}$-1step models have been tested varying the $\tilde{g}$ mass, to achieve a larger boost of the Ws. The distribution is rather flat for the signal, with an abrupt cut-off edge at the $\tilde{\chi}_1^{\pm}$ mass. The background instead is falling steeply, reaching only up to 800 GeV. Thus, a large suppression of the background of all kinds could be achieved with M_{T2}, which would be especially worthwhile for the signal models with $\tilde{g}$ (or $\tilde{q}$) mass above 800 GeV. Furthermore, even a loose cut on M_{T2} could be helpful for rejecting the Diboson and γ+jets background.

The impact of selection rules containing several combination of $m_{\mathrm{eff}} > 1500$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}} > 0.3$ and $M_{\mathrm{T2}} >500$ GeV cuts are presented in Table 9.1. The values of the significance, calculated from the simplified formula (see Eq. 6.8), show that the usage of M_{T2} as discriminating variable could increase our exclusion power in the W SRs. A similar technique could be used for other SRs dedicated to some specific model (e.g. $\tilde{t}\bar{\tilde{t}}$ production followed by decay $\tilde{t} \to t\tilde{\chi}_1^0 \to W^+ b\tilde{\chi}_1^0$). This study was introduced after the publication [64] was released and was strongly recommended for further investigations during Run II analysis.

	m_{eff}	m_{eff}, M_{T2}	E_T^{miss}/m_{eff}	$E_T^{miss}/m_{eff}, M_{T2}$	$E_T^{miss}/m_{eff}, m_{eff}$	$E_T^{miss}/m_{eff}, m_{eff}, M_{T2}$
N(SM bg.)	21770	149.2	546.5	70.3	36.1	12.4
$N(\tilde{q}\tilde{q})$	14.3	8.2	16	11.3	8.1	7.0
$N(\tilde{g}\tilde{g})$	14	10.5	10.5	10.1	9.9	9.3
$s/\sqrt{b+s}(\tilde{q}\tilde{q})$	0	0.7	0.7	1.3	1.2	1.6
$s/\sqrt{b+s}(\tilde{g}\tilde{g})$	0	0.8	0.5	1.1	1.5	2.0

Table 9.1 Impact of different combinations of cuts on the final significance. The signal points chosen are:
$\tilde{g}\tilde{g}$-1step: $m_{\tilde{g}} = 1200$ GeV, $m_{\tilde{\chi}_1^\pm} = 1190$ GeV, $m_{\tilde{\chi}_1^0} = 60$ GeV.
$\tilde{q}\tilde{q}$-1step: $m_{\tilde{q}} = 800$ GeV, $m_{\tilde{\chi}_1^\pm} = 790$ GeV, $m_{\tilde{\chi}_1^0} = 60$ GeV.
At least 1 unresolved and 1 resolved Ws or 2 unresolved Ws are required.

9.5.2 W-veto

One can take advantage of the W reconstruction also for signal models where the W is not produced in a final state. All the events from $t\bar{t}$, Diboson and W-jets have one or more Ws in the final state and therefore could be suppressed by W-veto. This is however complicated by the insufficient boost of the Ws produced in the SM processes. The reconstruction efficiency with our approach is not high enough in the case of SM events (our exclusion limits are extended only in regions of 1-step models with a large boost of W). Fig. 9.12 shows the number of reconstructed Ws in the final state for the background and two $\tilde{q}\tilde{q}$-direct signal models: $m(\tilde{q}) = 850$ GeV, $m(\tilde{\chi}_1^0) = 100$ GeV and $m(\tilde{q}) = 450$ GeV, $m(\tilde{\chi}_1^0) = 400$ GeV. Both of the models are located very close to the exclusion limit curve of the SUSY 0-lepton analysis. There is not any obvious value of the number of Ws that could be used for vetoing the events exceeding this number. Note that once such a cut would be used in a SR, the effect of the JMS and JMR does not have to be necessarily negligible, which would worsen the results even more.

The issues arising from a small boost are relevant for hadronically decaying Ws. The leptonically decaying Ws are suppressed by the lepton veto applied in the SUSY 0-lepton analysis except of the Ws decaying into τ-leptons. The tagging of τ-lepton and its potential usage as τ-veto is discussed in the Appendix B.

9.6 Calorimetric versus Truth W Invariant Mass

The truth jet is the jet constructed from nearby particles produced during the simulation at the generator level. Its construction includes the hadronization and factorization phases followed by the anti-$k_\perp$ algorithm when all particle vectors are clustered into the truth jet (i.e. without detector simulation). The invariant mass of the jet (real, truth or simulated) or di-jet system

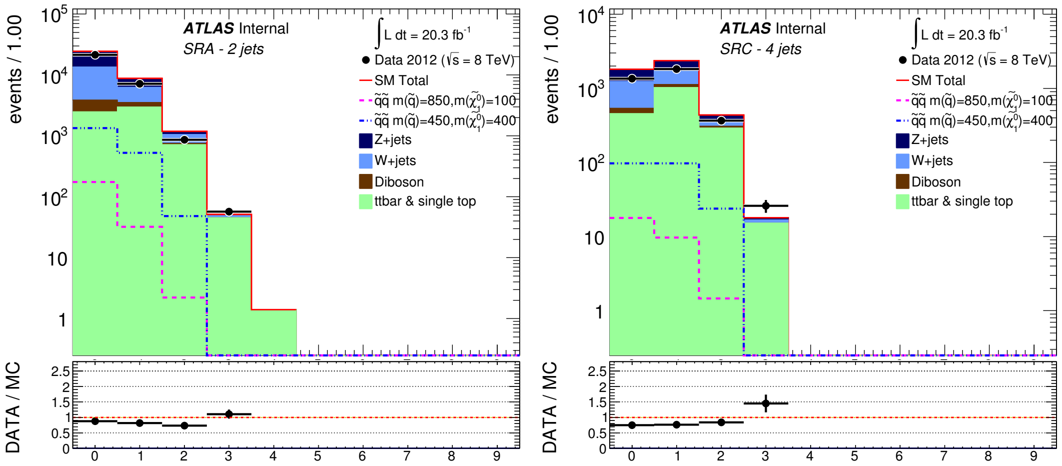


Fig. 9.12 The number of reconstructed Ws of both types: resolved and unresolved. The signal points are two direct decay chains, i.e. without Ws in the final state. There is no clear indication on how to use this variable to suppress the SM background.

originating from the W does not correspond to the W mass but is smeared around it. The reconstructed jet undergoes the same smearing but additionally, the shower and the detector-related effects determine the final shape of jet invariant mass distribution. Building the W mass ratio: $m_{\mathrm{reco}}(W\mathrm{jet(s)})/m_{\mathrm{truth}}(W\mathrm{jet(s)})$ one can study of the effects that smear the W reconstructed mass, namely: impact of the reconstruction and calibration of the jets, mis-reconstruction of the W and so called "close-by" effects. The last effect corresponds to the overlaps and interactions of the jets which are propagating in very close directions. Any large discrepancy of this ratio from 1 could eventually mean difficulties in terms of the effects mentioned above.

The mass ratio has been studied with W' simulated samples where the complete truth information about particles as well as truth jets is available. Two-dimensional histograms are populated with the mass ratio on the y-axis in Fig. 9.13 and Fig. 9.14. Only the jets with the $p_{\mathrm{T}} > 20$ GeV are involved in this study and the reconstructed W (either resolved or unresolved) is required to be within $\Delta R < 0.2$ from the truth W in order to study primarily the correctly reconstructed Ws. The x-axis of the histograms represents the ΔR between the jets in the W's di-jet system and the p_{T} of the truth W respectively. The ΔR calculation is not feasible in the unresolved case, thus this case is counted into a single bin, i.e into the first bin of the the first histogram. The conclusion of this study can be summarized as follows:

- Both the histograms do not show any unexpected dependence of this ratio with the p_{T}(truth W) or ΔR.
- The mean value of this ratio seems to be slightly shifted upwards from one. However, it is compatible with one within 5%.

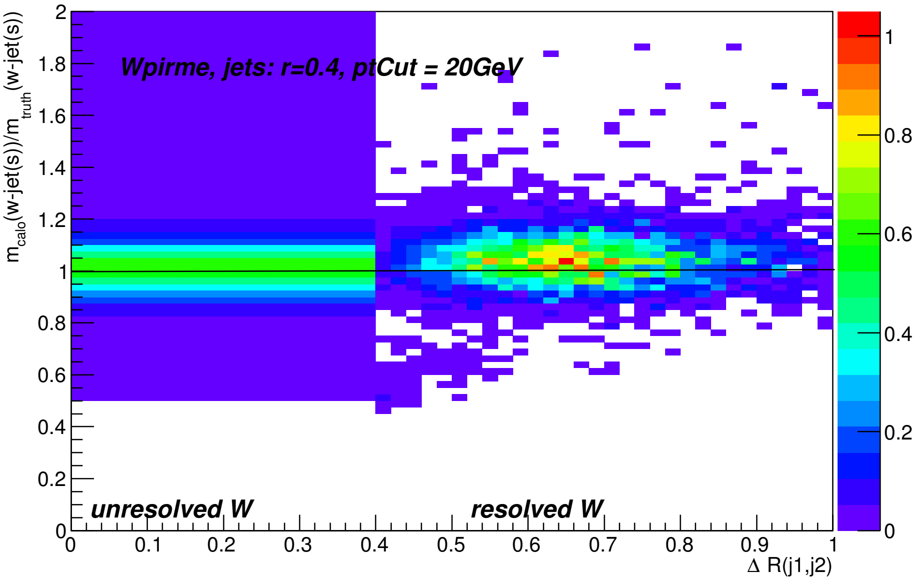


Fig. 9.13 The mass ratio of the calorimetric response and the truth jets with respect to the ΔR between the jets in the W di-jet system. The first bin with $\Delta R < 0.4$ is populated by the unresolved W events.

The second conclusion is a small effect not considered by the ATLAS SUSY group so far. However, this effect should be investigated for each analysis individually in the Run II.

9.7 Shape Fit of the *W* Mass Distribution

To date, the 0-lepton analysis does not profit from the shape of the distribution of any variable. The W peak in the invariant mass distribution is a convenient candidate to perform shape fits, that could increase the sensitivity of the analysis.

In Appendix A, the fit of the background with a quadratic function is performed, for which an explanation is missing from the physics point of view. Furthermore, there is not any large improvement in terms of sensitivity of the 0-lepton analysis by the usage of this simple fit. A somewhat different approach is tested in this section: the current SR 4jW is segmented into bins in invariant mass and such bins are treated as an individual SRs in the likelihood fit expressed by the form 6.6. All of the bins are involved in the hypothesis testing procedure to set the exclusion limits instead of one bin only as it is introduced in the 0-lepton analysis. The fit is performed in the distribution of the invariant mass of the di-jet system, the same distribution used to define the resolved W in SR 4jW, see Fig. 9.3.

The performance of this method is shown in Fig. 9.15, where the bins filled by red color

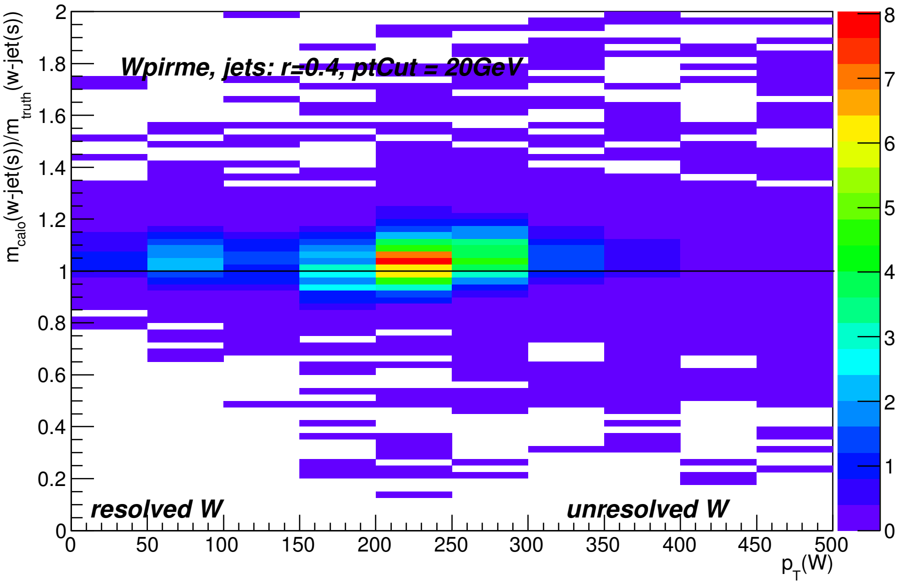


Fig. 9.14 The mass ratio of the calorimetric response and the truth jets with respect to the p_{T} of the associated truth W.

correspond to the $\tilde{q}\tilde{q}$-1step model with mass parameters $m_{\tilde{q}} = 700$ GeV, $m_{\tilde{\chi}_1^{\pm}} = 690$ and $m_{\tilde{\chi}_1^0} =$ 60 GeV. The signal is strongly suppressed after the fit, which can be interpreted in the sense that this SUSY model is more strongly excluded.

The fit has been processed over the entire set of available $\tilde{q}\tilde{q}$-1step simplified models and the expected exclusion limit has been plotted in Fig. 9.16. There is a clear extension of the expected exclusion limit line in the region rich on boosted Ws. Furthermore, the method is fully transparent as it is the same method officially used in the 0-lepton analysis with the only addition of the SR binning.

The only issue arises with the monitoring of the nuisance parameters, see Fig. 9.17. The nuisance parameters representing the uncertainties of the shape fit are largely shifted from their nominal value, compared to the fit of the cut-and-count method, which is often called *over-profiling* of the systematic uncertainties. If the nuisance parameters remain at their nominal value as it happens for the cut-and-count method, no more investigations are needed. When the opposite happens, it can point to non-complete information included for a given uncertainty (e.g. correlations are not correctly described). In such case, the recommended procedure is to divide the systematic uncertainty in more pieces (e.g. the JES uncertainty into all of its components). In the case of the shape fit tested here, almost all of the uncertainties undergo truly large over-profiling. Note, that dividing of the uncertainties into their components would require a large amount of computing time.

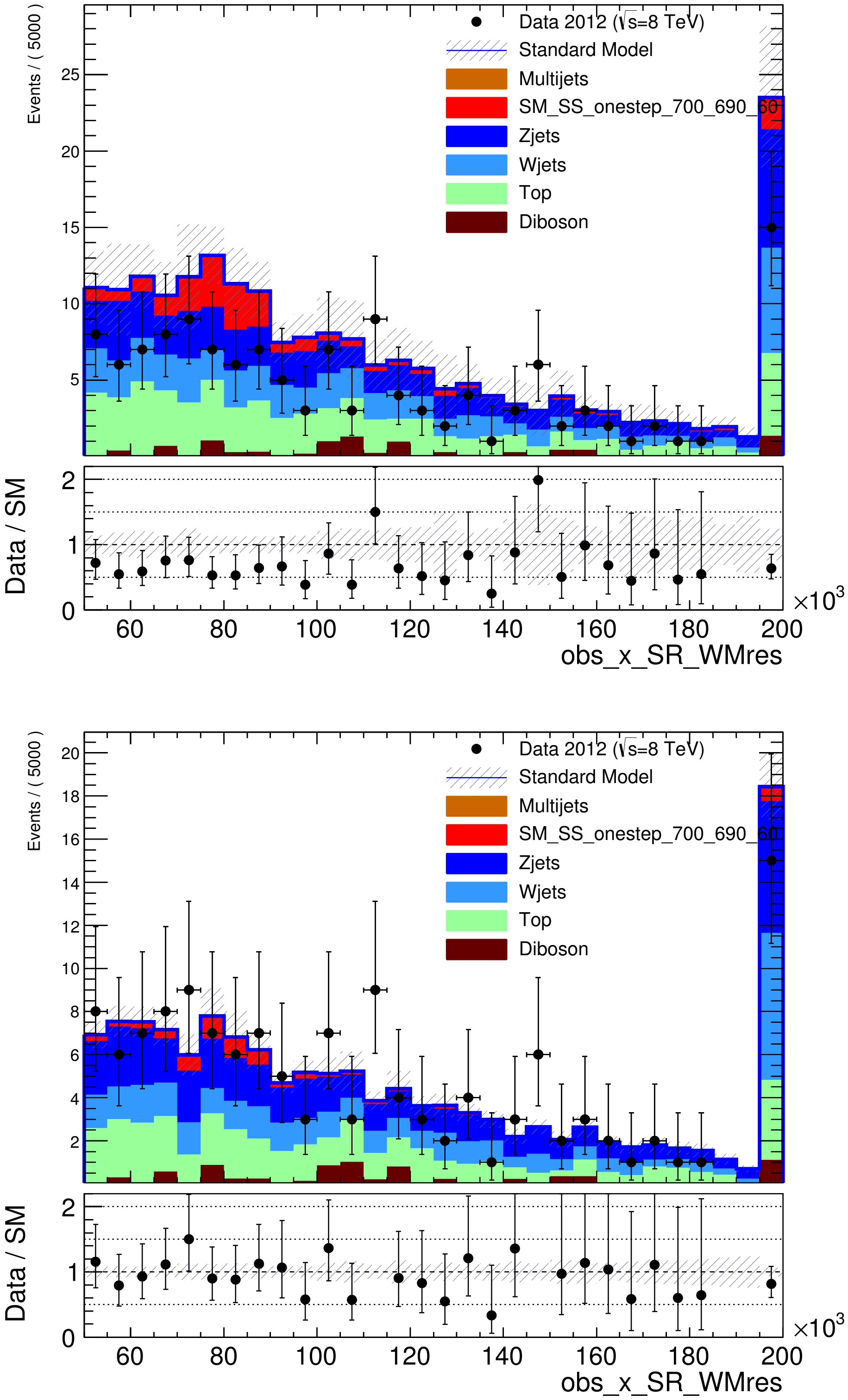


Fig. 9.15 The di-jet invariant mass. The upper plot shows the distribution before the fit and the lower plot after the fit.

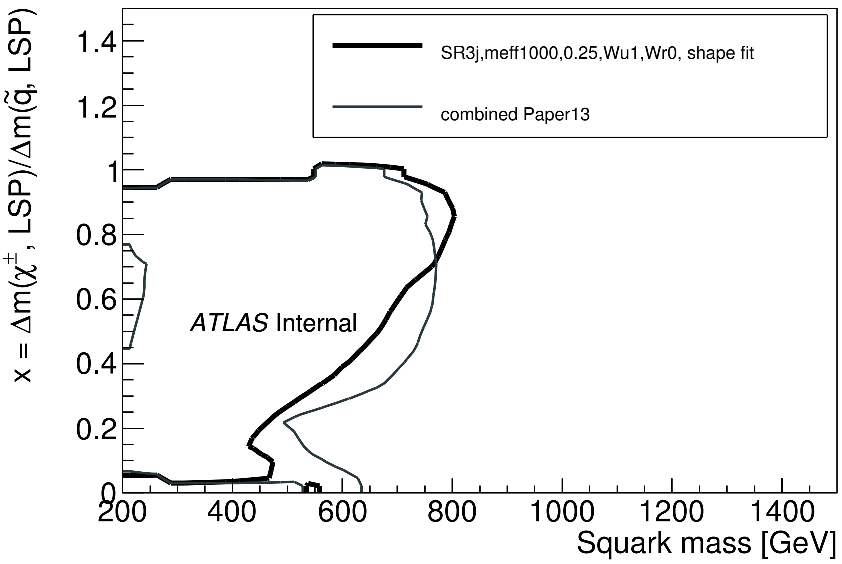


Fig. 9.16 The expected exclusion limit for the $\tilde{q}\tilde{q}$-1step simplified model. The thin black line denotes the best combined performance of the optimized 0-lepton SRs. The thick black line is obtained using the shape fit of the di-jet invariant mass distribution in the SR with the selection: $E_{\mathrm{T}}^{\mathrm{miss}} > 1000$ GeV,$E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}} > 0.25$ and at least 1 unresolved W.

.

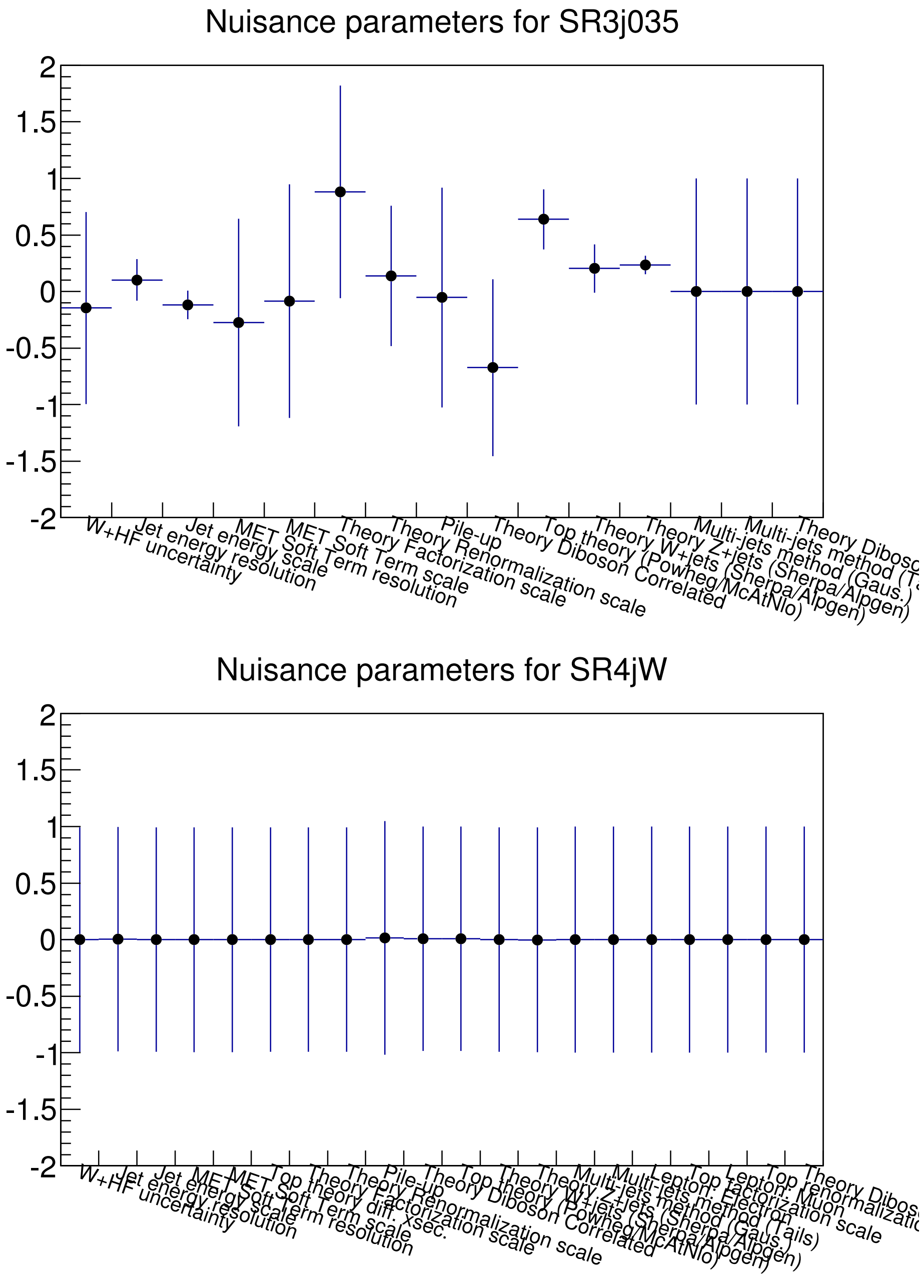


Fig. 9.17 The monitoring of the nuisance parameters representing the systematic uncertainties in two cases: the shape fit in the upper figure, the cut-and-count method in the bottom figure. Each bin is dedicated to one systematic uncertainty. The expected value should stay close to 0 and the error bar length close to 1, as in the bottom figure.

10. Usage of Reclustering Techniques for W boson Tagging

The boosted W tagging described in Chapter 9 suffered from the JMS and JMR uncertainties, for which a clear prescription has not been studied yet within the ATLAS experiment. Hence, the natural development is to investigate these uncertainties and develop some strategy to handle them that could be also used in other analyses. The reclustering described in this chapter is one of the possible approaches to solve this issue.

The reclustering is a method of building jets with larger jet cones starting from the calibrated jets of smaller cones using the anti-$k_\perp$ algorithm[1] described in Chapter 7. Additionally, a trimming condition (specific selection) can be applied to the small jets as well as a requirement on their p_T. Freedom in these two parameters provides a space for the optimization of this method by each analysis independently. The trimming condition is applied via a parameter f defined as the ratio p_T(subjet)/p_T(jet), where jet is the final large jet. The trimming condition and p_T-cut are applied in order to suppress the pile-up or underlying event contamination of the final large jet. The reclustering is schematically depicted in Fig. 10.1.

The same W' simulated sample as already mentioned in Chapter 9 has been used for this study and the multi-jet (QCD) simulated sample has been used for the background. Both the samples contain the information about anti-$k_\perp$ jets with smaller jet cones. The cone radius of the small jets shown in the following figures was set to $R = 0.2$ and the small jets are reclustered into the large jets of cone radius $R = 1.0$. At the time of this study, the official calibration of anti-$k_\perp$ jets with $R = 0.2$ was expected to be released at some point during Run II. A preliminary released calibration is applied to $R = 0.2$ jets, where pile-up suppression is implemented. The JES and JER uncertainties officially released for the $R = 0.4$ jets are applied to the tested $R = 0.2$ jets as they were not yet available for smaller cone sizes. The trimming conditions are set to the reasonable (non-extreme) values $f > 0.05$ and p_T(subjet)$>$ 15 GeV.

After the reclustering procedure is applied to each event, a new set of the fat jets, referred here as RT jets (reclustered and trimmed), is available, together with information about their

---

[1] The small jets are used instead of topo-clusters as an input to the reclustering.

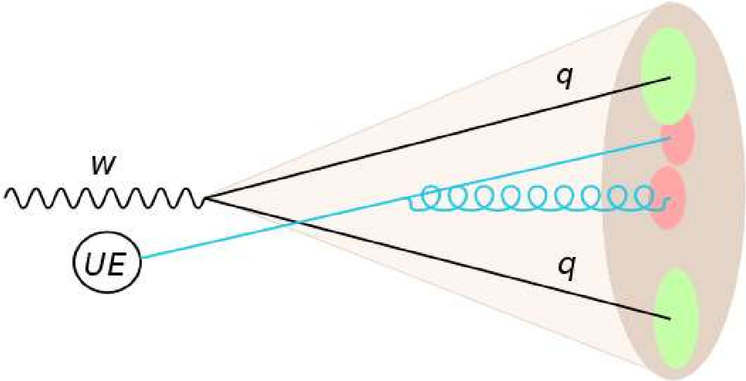


Fig. 10.1 The reclustering procedure. The small jets (red) coming from the underlying event or pile-up are excluded by the application of the trimming conditions. The two "interesting" sub-jets (green) are building the final fat jet.

structure, i.e. the constituting jets (sub-jets). All of the histograms published in this chapter are populated with the RT jet with largest p_{T} in the event. Thus, each event corresponds to one (or none) entry in the histogram, weighted for the cross-section of the given sample.

Firstly, the number of sub-jets in the RT jets is investigated, see Fig. 10.2. The histogram of the number of sub-jets is shown in three p_{T} regimes of the truth jet associated to the RT jets. From the point of view of the 0-lepton SUSY analysis, the low p_{T} regime is suffering from the low boost of the W[2]. In the high p_{T} regime on the other hand, the fraction of unresolved Ws, which contains only one sub-jet, increases. In such a case, the JMS and JMR can not be derived from the JES and JER uncertainties of the sub-jets as it is discussed further in this chapter. This effect is clearly visible in the last p_{T} regime as a shift of the mean value of the number of sub-jets towards lower values. However, the high p_{T} regime is still dominated by the cases with two sub-jets in the jet, keeping in mind that some of them can indeed originate from the underlying event or pile-up and not from the decay of the W.

The JMS and JMR uncertainties could be treated in the set of RT jets in the following way: **If there are two or more sub-jets in the RT jet, the JMS and JMR uncertainties should be dominated by the uncertainties on the p_{T} measurements of the sub-jets, i.e. their JES and JER uncertainties.** This hypothesis is supported by the fact that the invariant mass of the RT jet is mostly given by the p_{T} of the sub-jets rather than their invariant masses. Furthermore, two prongs (two sub-jets) are expected from the W decaying into $q\bar{q}'$ pair. This statement is quantified in the following way: fixed values of the JMS and JMR uncertainties (5% and 10% respectively) are placed on the small jets and their impact on the RT jets is evaluated with respect to the effect of sub-jet's JES and JER uncertainties. The result of this study is

---

[2] Sensitivity of the SUSY searches is not significantly increased by tagging low-pt Ws.

discussed further in this chapter.

For those RT jets with two or more sub-jets, the impact of the JER and JES uncertainties with respect to fixed JMS and JMR uncertainties can be studied using the invariant mass distributions of the RT jets, see Fig. 10.3. Three p_{T} regimes are employed as in the previous case, since the impact is expected to be different for different boosts of the sub-jet system. Indeed, the yellow bands corresponding to JES are significantly larger than the orange error bands corresponding to JMS for the two lower p_{T} regimes. The resolution uncertainties are behaving differently. There is only a small impact of both the JMR and JER in the medium p_{T} regime. The impact of the JER is noticeable in the low p_{T} regime and a small, but non-negligible, impact of both resolution uncertainties is visible for the high p_{T} regime.

To evaluate this effect quantitatively, several statistical properties of the invariant mass peak are summarized in Fig. 10.4, Fig. 10.5 and in Fig. 10.6. The MEAN and RMS are defined in an usual manner. The rest of the quantities refer to WINDOW, which is defined as the narrowest interval containing 68% of the signal statistics. One can calculate then WINDOW MEAN, WINDOW RMS and WINDOW WIDTH as the length of the interval. The WIN./ALL refers to the ratio of the integral of the signal events in the WINDOW and entire range of the invariant mass distribution. This is expected to be at 68% but it can slightly differ due to the binning (this variable was included as a check). The WINDOW QCD ACCEPTANCE refers to the percentage of the QCD events entering the WINDOW. The WINDOW SIGNIFICANCE is calculated as $s/\sqrt{s+b}$ in the WINDOW. Note that this significance is derived using only a partial description of the background (only QCD) and is therefore used here only as an indicator of the performance of the selection in terms of the uncertainties. The WINDOW 11 BIN MAX. POSITION is the position of the maxima of the peak calculated as a weighted average of the 11 neighbouring bins with the largest statistics within the WINDOW.

The conclusion from Figs. 10.4, 10.5 and 10.6 can be summarized as follows:

- The statement that the JMS and JMR uncertainty can be derived from the JES and JER of the sub-jets for the reclustered $R = 1.0$ jets from the small jets with $R = 0.2$ is valid up to $p_{\mathrm{T}} = 500$ GeV of the RT jet. This is supported by Figs. 10.4, 10.5 for low and medium p_{T} regimes, where the yellow bands representing JES and the green lines representing JER significantly exceed the orange bands representing JMS and the blue lines for JMR with fixed conservative values of 5% and 10% respectively.

- It is mostly the properties related to the peak width that are suffering from JMS and JMR for the largest p_{T} regime. Note that none of the uncertainties determines the peak properties (i.e. the final jet mass uncertainties) directly, as the condition $N_{\text{sub-jets}} \geqslant 2$ is applied. The same figures were plotted when the tighter condition $N_{\text{sub-jets}} = 2$ is applied

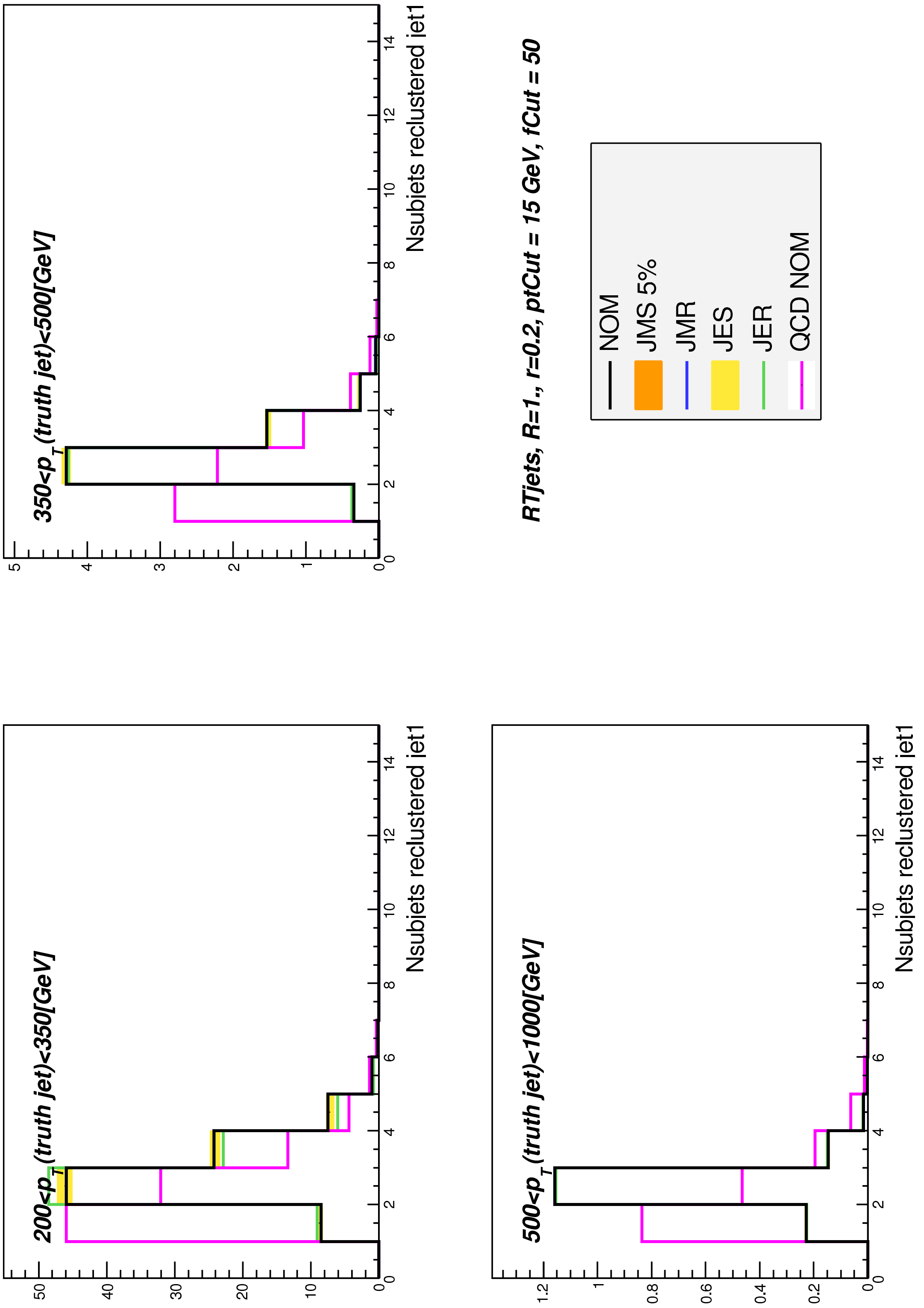


Fig. 10.2 The number of the sub-jets inside the RT jets divided in three regions of p_T of the truth jet associated to the RT jet.

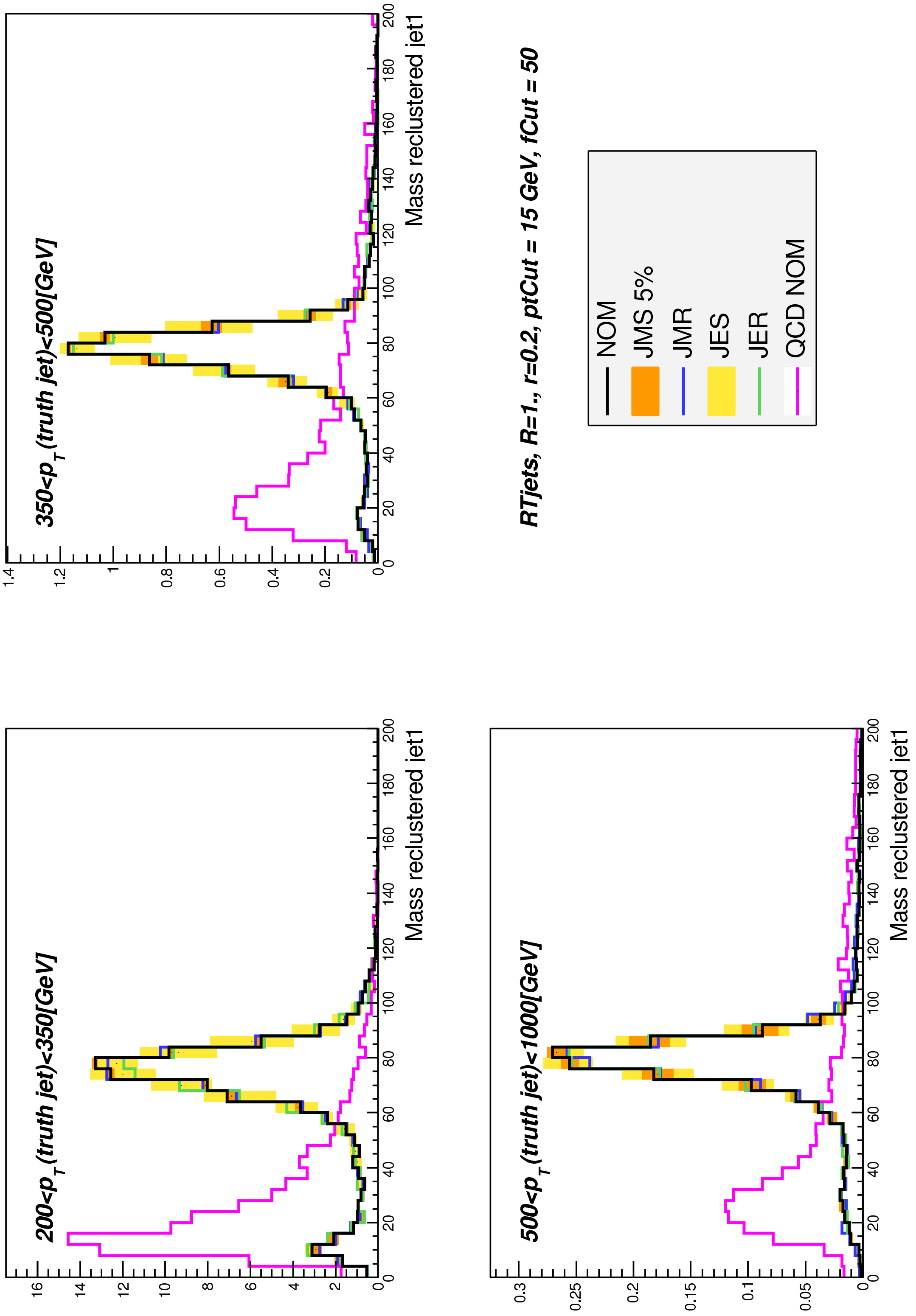


Fig. 10.3 The invariant mass distributions of the RT jets for three regions of p_{T} of the RT jet.

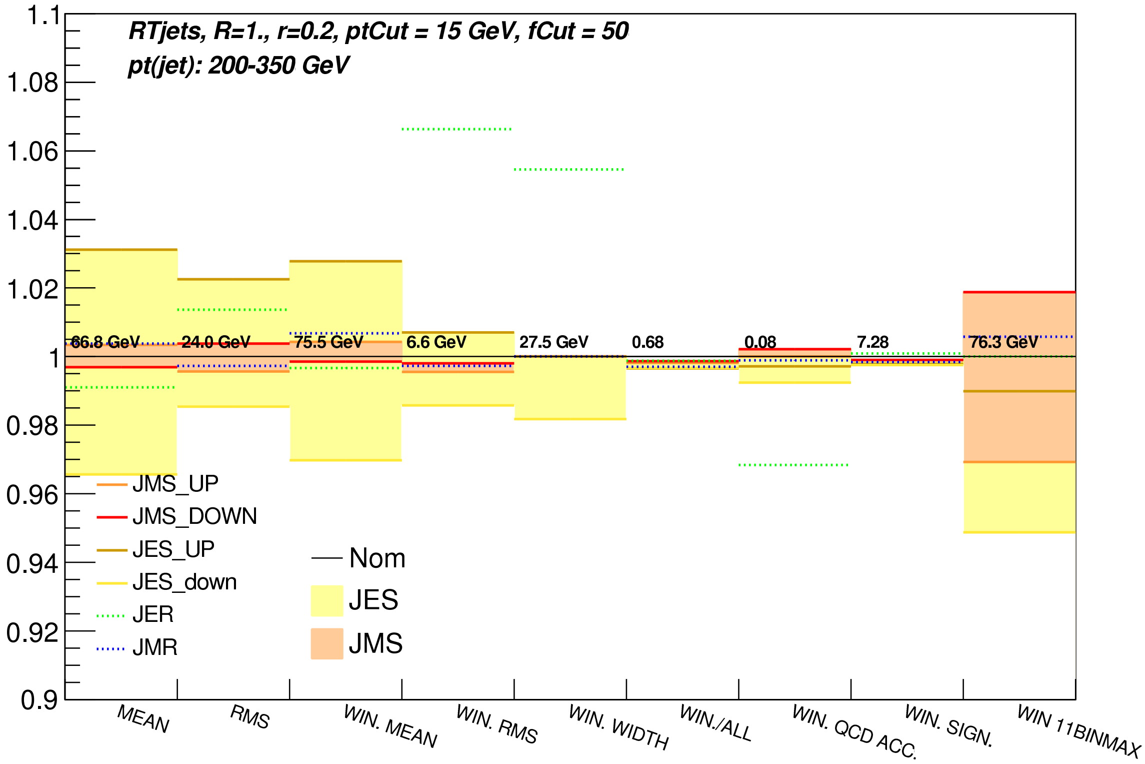


Fig. 10.4 The statistical properties of the invariant mass peak for the low p_T regime. The relative change of the given statistical property is shown on the y-axis.

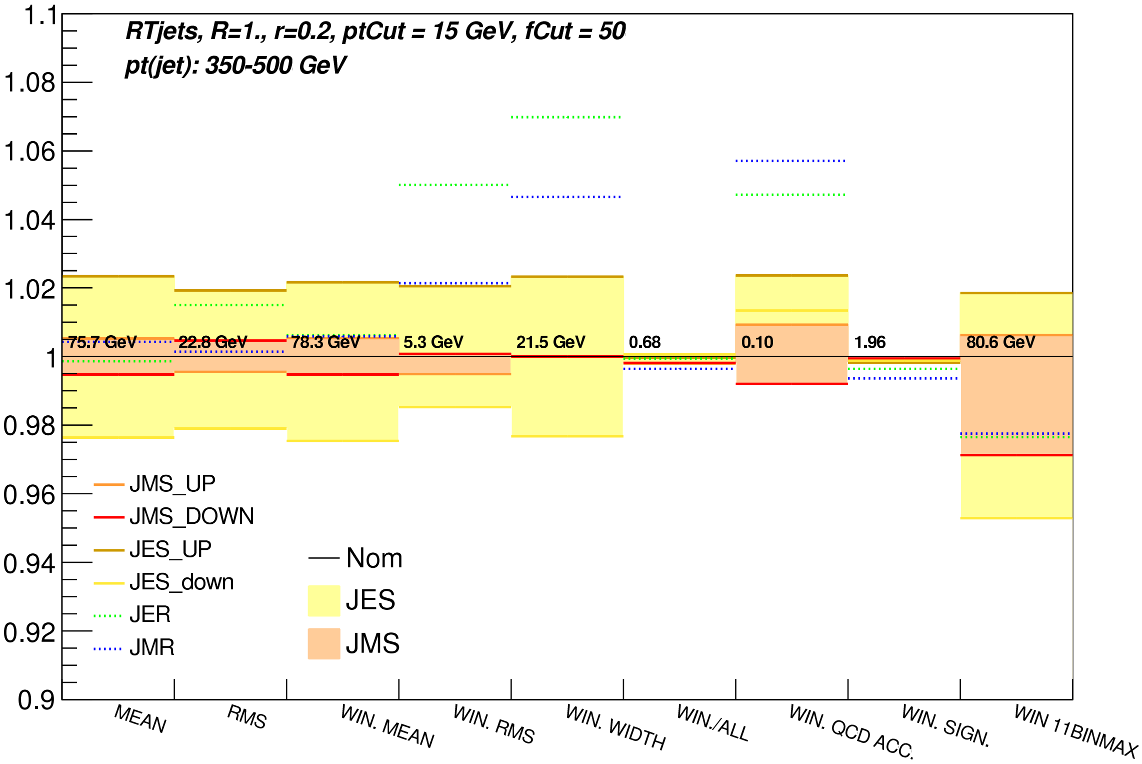


Fig. 10.5 The statistical properties of the invariant mass peak for the medium p_T regime. The relative change of the given statistical property is shown on the y-axis.

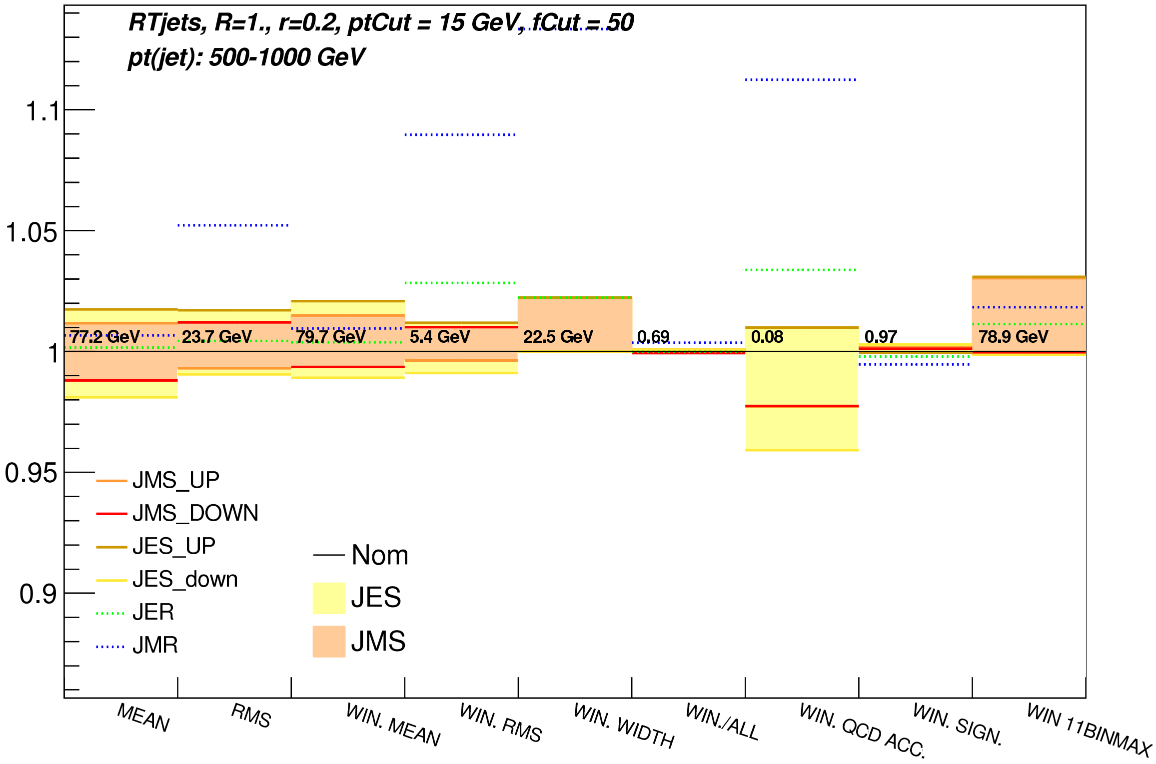


Fig. 10.6 The statistical properties of the invariant mass peak for the large p_{T} regime. The relative change of the given statistical property is shown on the y-axis.

with very comparable results.

- All the uncertainties are at relatively low values, less than 10% for the low and medium p_{T} regimes. Only two properties exceed slightly 10 % in the large p_{T} bin. However, the WINDOW SIGNIFICANCE stays well within 2%, which supports the conservativeness of the JMS and JMR used in the 0-lepton analysis.

Furthermore, the mass uncertainties are not negligible for larger cone sizes of the the anti-$k_{\perp}$ small jets. Details and explanations are given in Appendix D.

Although the results look promising for the $R = 0.2$ cone size of the small jets, these jets could not be officially used at the time of the writing of this thesis, because of missing official calibration of jets of small cone sizes. This method of treating JMS and JMR would probably require some upper limit on the p_{T} of the RT jet, because the JMS and JMR have larger impact in larger p_{T} ranges. Thus, other techniques have been investigated. **One of the options to exclude the impact of the JMS and JMR completely would be to reconstruct the W without using any quantity dependent on the mass.** The reclustering tool provides jet-substructure variables that could be potentially used for this purpose. Namely the p_{T}-asymmetry defined as:

$$p_{\mathrm{T}}\text{-asymmetry} = \frac{p_{\mathrm{T}}(j_1) - p_{\mathrm{T}}(j_2)}{p_{\mathrm{T}} j_1 + p_{\mathrm{T}} j_2} \tag{10.1}$$

j_i denotes the sub-jet with index i, where sub-jets are sorted descending in p_{T}. The second sub-structure variable is the jet-width defined as:

$$\text{jet-width} = \frac{\sum_i \Delta R(RT\,jet, j_i) p_{\mathrm{T}}(j_i)}{\sum_i p_{\mathrm{T}}(j_i)} \tag{10.2}$$

and the third variable studied here is the mass splitting scale $\sqrt{d_{1,2}}$:

$$\sqrt{d_{1,2}} = p_{\mathrm{T}}(j_2) \Delta R(j_1, j_2) \tag{10.3}$$

The j_1 and j_2 in this definition are defined as follows: all small jets falling into the fat jet are reclustered until exactly two sub-jets remains (the two sub-jets before the last recombination step of anti-$k_{\perp}$ algorithm), which create j_1 and j_2.

The potential of these variables has been firstly tested by looking at the correlation between each of them and the invariant mass, to check that they carry the same information. The correlation between the three variables themselves has also been tested, to evaluate if the set is actually redundant. An example of correlation study is shown in Fig. 10.7, illustrating the position of the peak of the 2-dimensional distribution. Significant separation power in the lowest-p_{T} regime is visible in this example for both variables, invariant mass and $\sqrt{d_{1,2}}$. The peaks for the multi-jet background (green) and W' (blue) are overlapping in the high-p_{T} regime. Nevertheless, the $\sqrt{d_{1,2}}$ seems to perform comparably with the invariant mass.

To quantify the statement that these three variables could potentially perform as good as the invariant mass during the reconstruction, the following significance test is employed. The 2-dimensional interval giving the best significance (see Eq. 6.8) in the correlation histogram is found for each pair of the jet-substructure variables plus invariant mass. The significance of such an interval is then quantifying the separation power between W jets (each event corresponds to one largest p_{T} RT jet) and jets originating from multi-jet background, since we know that the jets in W' simulation sample originate from Ws. Unfortunately, other background simulation samples could not be used when this test was performed, since the samples did not contain all the necessary information. Thus, not the absolute value of the significance but the comparison of the value between the pairs of variables is the object of the study. The result is shown in Fig. 10.8. Multi-jet acceptance and W' acceptance are written in the two bottom lines of the figure. The aim is to keep QCD acceptance low and W' acceptance as large as possible at the same time. The conclusion from these numbers is that the pairs which contains the invariant mass still perform slightly better. On the other hand, there is not any impact of the JMS and JMR in terms of significance in the case where the combinations are built purely from the jet-substructure variables.

The usage of jet-substructure to identify the W has been already investigated in the publication [81]. All of the studies support the idea of reclustering and provide the ATLAS JetEtmiss working group a good motivation to calibrate anti-$k_{\perp}$ jets with cone size 0.2, which is the best candidate to be used as an input to the reclustering procedure.

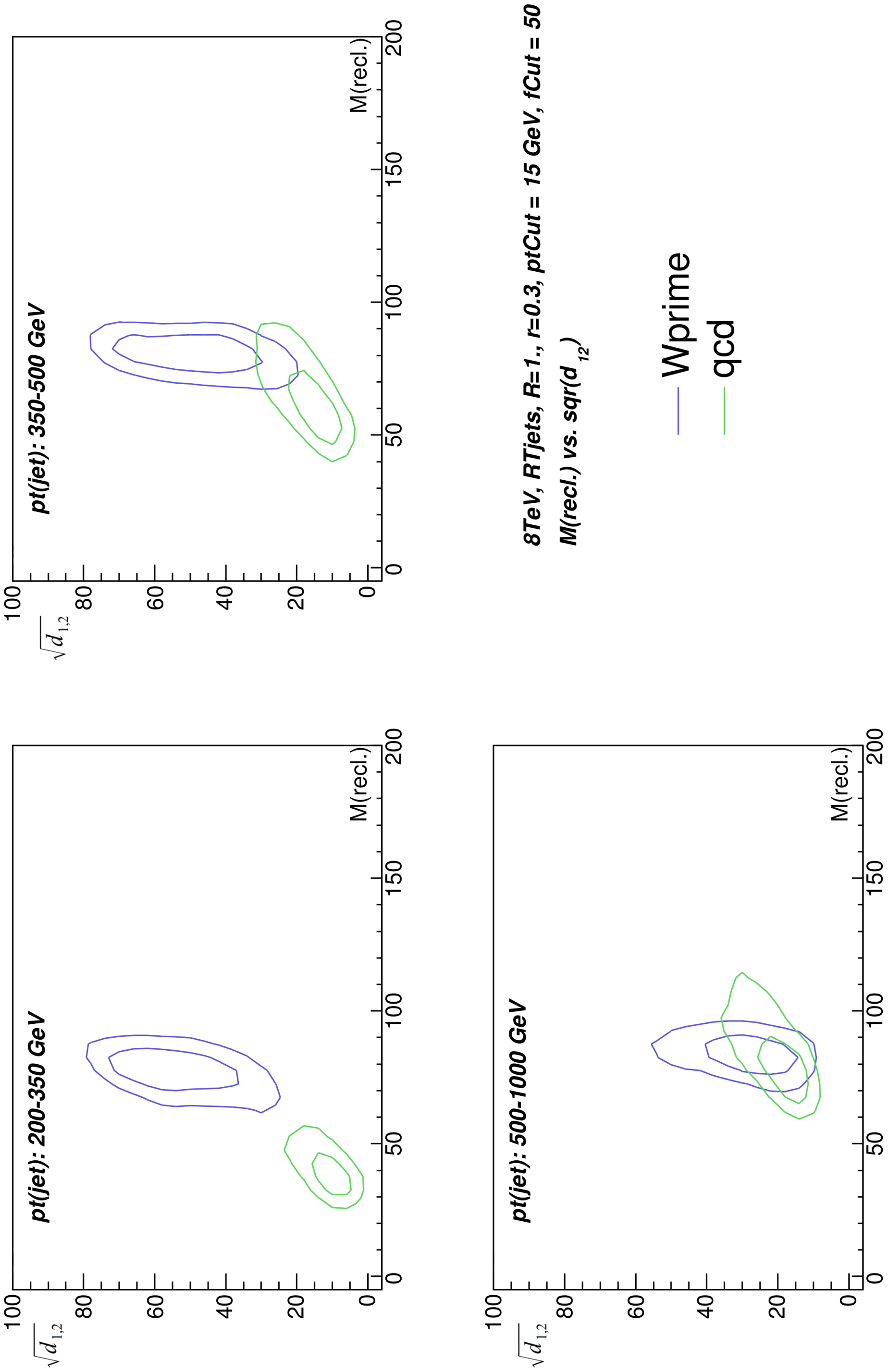


Fig. 10.7 The correlation of the invariant mass of the RT jet and mass splitting scale $\sqrt{d_{12}}$, also suitable for the W reconstruction.

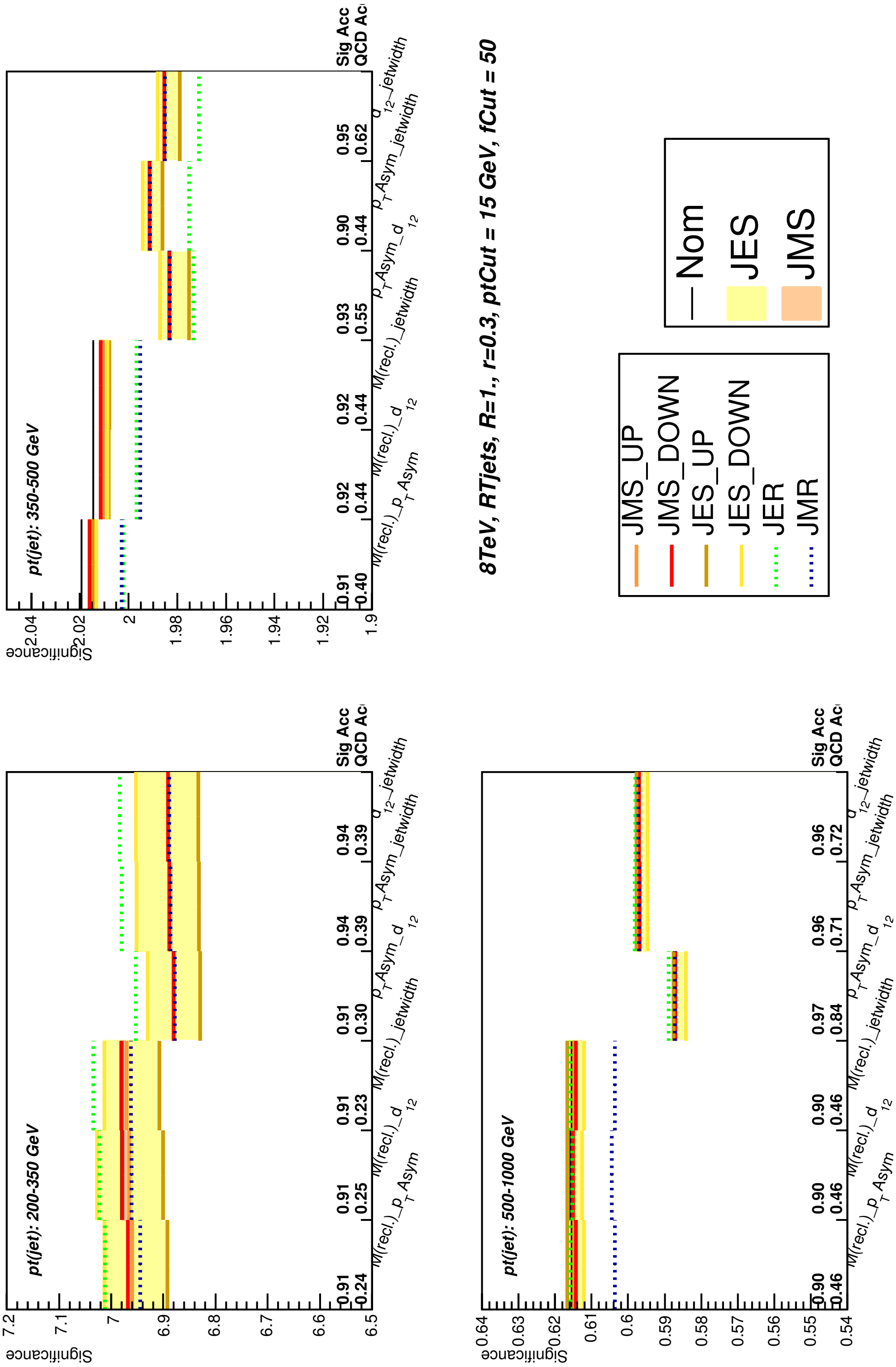


Fig. 10.8 Significance for different pairs of variables, calculated using an optimal two-dimensional cut, as described in the text.

References

[1] http://www.usparticlephysics.org/p5/.

[2] I. M. Gregor. ATLAS Upgrades for the next Decades. Aug 2014.

[3] http://www.weltderphysik.de/gebiet/teilchen/experimente/teilchenbeschleuniger/lhc/.

[4] http://cds.cern.ch/record/1095924.

[5] http://inspirehep.net/record/940611/files/ATLAS_ID_Barrel.png.

[6] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999. Electronic version not available.

[7] http://www.atlas.ch/photos/calorimeters-tile-barrel.html.

[8] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTile.

[9] Collaboration ATLAS. ATLAS Insertable B-Layer Technical Design Report Addendum. Technical Report CERN-LHCC-2012-009.ATLAS-TDR-19-ADD-1, CERN, Geneva, May 2012. Addendum to CERN-LHCC-2010-013, ATLAS-TDR-019.

[10] Collaboration ATLAS. The ATLAS Level-1 Trigger System. (J. Phys.: Conf. Ser. 396 012010), 2012.

[11] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[12] S Ballestrero, W Vandelli, and G Avolio. ATLAS TDAQ system: current status and performance. Technical Report ATL-DAQ-PROC-2011-020, CERN, Geneva, Sep 2011.

[13] G. Duckeck, D. Barberis, R. Hawkings, R. Jones, N. McCubbin, G. Poulard, D. Quarrie, T. Wenaus, and E. Obreshkov. ATLAS computing: Technical design report. 2005.

[14] I Ueda, T Kubes, L Goossens, G Stewart, S Jezequel, A Nairz, G Negri, S Campana, A Klimentov, and A Di Girolamo. ATLAS Operations: Experience and Evolution in the Data Taking Era. Technical Report ATL-COM-SOFT-2010-049, CERN, Geneva, Oct 2010. CHEP2010, 18 Oct 2010.

[15] T Beermann, V Garonne, M Lassnig, M Barisits, R Vigne, C Serfon, G A Stewart, L Goossens, A Nairz, and A Molfetas. Rucio - The next generation large scale distributed system for ATLAS Data Management. Mar 2014.

[16] K De, A Klimentov, T Maeno, P Nilsson, D Oleynik, S Panitkin, A Petrosyan, J Schovancová, A Vaniachine, and T Wenaus. The Future of PanDA in ATLAS Distributed Computing. Technical Report ATL-SOFT-PROC-2015-047, CERN, Geneva, May 2015. Proceedings paper of the talk presented at CHEP2015 - International Conference on Computing in High Energy and Nuclear Physics (CHEP), Okinawa, Japan, 13-17 Apr 2015.

[17] T Maeno. Panda: distributed production and distributed analysis system for atlas. *Journal of Physics: Conference Series*, 119(6):062036, 2008.

[18] D C van der Ster, J Elmsheuser, M Úbeda García, and M Paladin. Hammercloud: A stress testing system for distributed analysis. *Journal of Physics: Conference Series*, 331(7):072036, 2011.

[19] http://dashb-atlas-ssb.cern.ch/.

[20] http://toolkit.globus.org/toolkit/docs/latest-stable/gridftp/.

[21] http://toolkit.globus.org/toolkit/docs/2.4/datagrid/deliverables/gsiftp-tools.html.

[22] https://tools.ietf.org/html/rfc1945.

[23] www.webdav.org/.

[24] https://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.html.

[25] https://dmc.web.cern.ch/projects/gfal2-utils.

[26] http://www.perfsonar.net/.

[27] J Schovancova, F H Barreiro Megino, C Borrego, S Campana, A Di Girolamo, J Elmsheuser, J Hejbal, T Kouba, F Legger, E Magradze, R Medrano Llamas, G Negri, L Rinaldi, G Sciacca, C Serfon, and D C Van Der Ster. ATLAS Distributed Computing Automation. Jul 2012.

[28] http://dashb-wlcg-transfers.cern.ch/ui/.

[29] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

[30] J. Goldstone. Field theories with « superconductor » solutions. *Il Nuovo Cimento (1955-1965)*, 19(1):154–164, 2008.

[31] F R Klinkhamer. Neutrino mass and the Standard Model. Technical Report arXiv:1112.2669. KA-TP-41-2011, Dec 2011. Comments: 4 pages.

[32] J. Schechter and J. W. F. Valle. Neutrinoless double-β decay in su(2)×u(1) theories. *Phys. Rev. D*, 25:2951–2954, Jun 1982.

[33] H. V. Klapdor-Kleingrothaus, A. Dietz, H. L. Harney, and I. V. Krivosheina. Evidence for neutrinoless double beta decay. *Mod. Phys. Lett.*, A16:2409–2420, 2001.

[34] Michel Davier, Andreas Hoecker, Bogdan Malaescu, and Zhiqing Zhang. Reevaluation of the Hadronic Contributions to the Muon g-2 and to alpha(MZ). *Eur. Phys. J.*, C71:1515, 2011. [Erratum: Eur. Phys. J.C72,1874(2012)].

[35] Rudolf Haag, Jan T. Lopuszanski, and Martin Sohnius. All Possible Generators of Supersymmetries of the s Matrix. *Nucl. Phys.*, B88:257, 1975.

[36] J. Wess and B. Zumino. A Lagrangian Model Invariant Under Supergauge Transformations. *Phys. Lett.*, B49:52, 1974.

[37] Pierre Fayet. Supersymmetry and Weak, Electromagnetic and Strong Interactions. *Phys. Lett.*, B64:159, 1976.

[38] Michael E. Peskin. Supersymmetry in Elementary Particle Physics. pages 609–704, 2008.

[39] H.P. Nilles. Supersymmetry, supergravity and particle physics. *Physics Reports*, 110(1):1 – 162, 1984.

[40] S. Mrenna T. Sjostrand and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP 05 (2006) 026*, (arXiv:hep-ph/0603175), 2006.

[41] G. Corcella et al. HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes). *JHEP 01 (2001) 010*, (arXiv:hep-ph/0011363), 2001.

[42] Daniele Amati and Gabriele Veneziano. Preconfinement as a property of perturbative QCD. *Phys. Lett. B*, 83(CERN-TH-2620):87–92. 13 p, Feb 1979.

[43] T. Gleisberg et al. Event generation with SHERPA 1.1 . *JHEP 02 (2009) 007*, (arXiv:0811.4622 [hep-ph]), 2009.

[44] P. Nason S. Frixione and C. Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method . *JHEP 11 (2007) 070*, (arXiv:0709.2092 [hep-ph]), 2007.

[45] C. Oleari S. Alioli, P. Nason and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP 06 (2010) 043*, (arXiv:1002.2581 [hep-ph]), 2010.

[46] S. Catani et al. Vector boson production at hadron colliders: A Fully exclusive QCD calculation at NNLO. *Phys. Rev. Lett. 103 (2009) 082001*, (arXiv:0903.2120 [hep-ph]), 2009.

[47] H.-L. Lai et al. New parton distributions for collider physics. *Phys. Rev. D 82 (2010) 074024*, (arXiv:1007.2241 [hep-ph]).

[48] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. MadGraph 5: Going Beyond. *JHEP 06 (2011) 128*, (arXiv:1106.0522 [hep-ph]), 2011.

[49] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008.

[50] G. P. Salam M. Cacciari and G. Soyez. The anti-k t jet clustering algorithm. JHEP(arXiv:0802.1189 [hep-ph]), 2008.

[51] M. H. Seymour S. Catani, Y. L. Dokshitzer and B. R. Webber. Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions. (Nucl. Phys. B 406 (1993) 187), 1993.

[52] S. Moretti Y. L. Dokshitzer, G. D. Leder and B. R. Webber. Better Jet Clustering Algorithms. (arXiv:hep-ph/9707323), 1997.

[53] ATLAS Collaboration. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. (arXiv:1112.6426 [hep-ex]), 2008.

[54] ATLAS Collaboration. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV. *Eur. Phys. J. C*, 73(arXiv:1112.6426. CERN-PH-EP-2011-191):2304. 111 p, Dec 2011. Comments: 100 pages plus author list (111 pages total), 93 figures, 17 tables, submitted to European Physical Journal C.

[55] M. Davier, A. Hoecker, B. Malaescu, C.Z. Yuan, and Z. Zhang. Reevaluation of the hadronic contribution to the muon magnetic anomaly using new $e^+e^- \rightarrow \pi^+\pi^-$ cross section data from BABAR. (arXiv:0908.4300 [hep-ex]), 2010.

[56] ATLAS Collaboration. Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC. *Eur. Phys. J. C*, 73(arXiv:1203.1302. CERN-PH-EP-2012-005):2305. 36 p, Mar 2012. Comments: 24 pages plus author list (36 pages total), 23 figures, 1 table, submitted to European Physical Journal C.

[57] ATLAS Collaboration. Jet energy resolution in proton-proton collisions at $\sqrt{s}$ = 7 TeV recorded in 2010 with the ATLAS detector. *Eur. Phys. J. C*, 73(arXiv:1210.6210. CERN-PH-EP-2012-191):2306. 26 p, Oct 2012. Comments: 13 pages plus author list (26 pages total), 15 figures, 1 table, submitted to European Physical Journal C.

[58] Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011.

[59] ATLAS Collaboration. Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector. Technical Report ATLAS-CONF-2012-043, CERN, Geneva, Mar 2012.

[60] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J. C*, (arXiv:1407.5063. CERN-PH-EP-2014-153):74. 51 p, Jul 2014. Comments: 39 pages plus author list + cover pages (51 pages total), 42 figures, 8 tables, published in EPJC, All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2013-05/.

[61] Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data. Technical Report ATLAS-CONF-2014-032, CERN, Geneva, Jun 2014.

[62] ATLAS Collaboration. Muon reconstruction efficiency and momentum resolution of the ATLAS experiment in proton–proton collisions at $\sqrt{s} = 7$ TeV in 2010. Technical Report arXiv:1404.4562. CERN-PH-EP-2013-154, Apr 2014. Comments: 22 pages plus cover

page plus author list (37 pages total), 18 figures, 0 tables,submitted to EPJC, All figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2011-01/.

[63] Performance of the Reconstruction and Identification of Hadronic Tau Decays in ATLAS with 2011 Data. Technical Report ATLAS-CONF-2012-142, CERN, Geneva, Oct 2012.

[64] ATLAS Clooaboration. Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using $\sqrt{s} = 8$ TeV proton-proton collision data. *J. High Energy Phys.*, 09(arXiv:1405.7875. CERN-PH-EP-2014-093):176. 52 p, May 2014. Comments: 36 pages plus author list + cover pages (54 pages total), 13 figures, 5 tables, submitted to JHEP, All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/SUSY-2013-02/.

[65] Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV with ATLAS. *Eur. Phys. J. C*, 72(arXiv:1108.5602. CERN-PH-EP-2011-114):1844. 33 p, Sep 2011. Comments: 22 pages plus author list (33 pages total), 18 figures, 5 tables, submitted to European Physical Journal C.

[66] Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at $sqrts = 7$ TeV. Technical Report ATLAS-CONF-2012-101, CERN, Geneva, Jul 2012.

[67] Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data. Technical Report ATLAS-CONF-2012-020, CERN, Geneva, Mar 2012.

[68] https://twiki.cern.ch/twiki/bin/view/Atlas/DataPreparation.

[69] Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data. Technical Report ATLAS-CONF-2012-020, CERN, Geneva, Mar 2012.

[70] M. Baak, G. J. Besjes, D. Côte, A. Koutsman, J. Lorenz, and D. Short. HistFitter software framework for statistical data analysis. *Eur. Phys. J.*, C75:153, 2015.

[71] Lorenzo Moneta, Kevin Belasco, Kyle Cranmer, Alfio Lazzaro, Danilo Piparo, Gregory Schott, Wouter Verkerke, Matthias Wolf, Kevin Belasco, Kyle Cranmer, Alfio Lazzaro, Danilo Piparo, Gregory Schott, Wouter Verkerke, and Matthias Wolf. The RooStats Project. *PoS*, ACAT2010(arXiv:1009.1003):057, Sep 2010. Comments: 11 pages, 3 figures, ACAT2010 Conference Proceedings.

[72] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical Report CERN-OPEN-2012-016, New York U., New York, Jan 2012.

[73] Michael Rammensee. *Search for supersymmetric particles in final states with jets and missing transverse momentum with the ATLAS detector.* PhD thesis, Freiburg U., 2013.

[74] Cowan, Glen and Cranmer, Kyle and Gross, Eilam and Vitells, Ofer. Using the Profile Likelihood in Searches for New Physics. *Eur. Phys. J. C*, 71(arXiv:1007.1727):1554, Jul 2010. Comments: fixed typo in equations 75 and 76.

[75] ATLAS Clooaboration. Search for squarks and gluinos using final states with jets and missing transverse momentum with the ATLAS detector in $\sqrt{s} = 7$ TeV proton-proton collisions. *Phys. Lett. B*, 701(arXiv:1102.5290. CERN-PH-EP-2011-022):186–203. 17 p, Feb 2011. Comments: 5 pages plus author list (17 pages total), 3 figures, 2 tables, submitted to Physics Letters B.

[76] ATLAS Clooaboration. Search for squarks and gluinos using final states with jets and missing transverse momentum with the ATLAS detector in $\sqrt{s} = 7$ TeV proton-proton collisions. *Phys. Lett. B*, 710(arXiv:1109.6572. CERN-PH-EP-2011-145):67–85. 21 p, Sep 2011. Comments: 9 pages plus author list (21 pages total), 6 figures, 3 tables, submitted to Physics Letters B.

[77] Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum and 20.3 fb^{-1} of $\sqrt{s} = 8$ TeV proton-proton collision data. Technical Report ATLAS-CONF-2013-047, CERN, Geneva, May 2013.

[78] Alexander Belyaev, Matthew Brown, Jesús Moreno, and Chloé Papineau. Discovering Minimal Universal Extra Dimensions (MUED) at the LHC. *J. High Energy Phys.*, 06(arXiv:1212.4858):080, Dec 2012. Comments: 46 pages, 19 figures.

[79] ATLAS Collaboration. Summary of the searches for squarks and gluinos using $\sqrt{s} = 8$ TeV pp collisions with the ATLAS experiment at the LHC. *J. High Energy Phys.*, 10(arXiv:1507.05525. CERN-PH-EP-2015-162):054. 91 p, Jul 2015. Comments: 91 pages plus author list (107 pages total), 43 figures, 23 tables, submitted to JHEP, All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/SUSY-2014-06/.

[80] Ljiljana Morvaj and Makoto Tomoto. *Search for Minimal Universal Extra Dimensions in the final state involving muons, jets and missing transverse energy in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*. PhD thesis, Nagoya U., Nov 2014. Presented 04 Feb 2015.

[81] Chunhui Chen. A new approach to identify boosted hadronically-decaying particle using jet substructure in its center-of-mass frame. Technical Report arXiv:1112.2567, Dec 2011. Comments: 6 pages, 7 figures.

[82] Rene Brun and Fons Rademakers. Root - an object oriented data analysis framework. See also https://root.cern.ch/.

[83] http://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=2&confId=126652.

[84] S.S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9, 1938. 60-2.

A. Shape Fits of the W Invariant Mass

A.1 Quadratic Fit of the Background

The invariant mass distributions of a jet/di-jet system show significant peaks corresponding to the presence of two Ws in the final state of the $\tilde{g}\tilde{g}/\tilde{q}\tilde{q}$-1step models. The second W in the final state is very rare for the background except for the $t\bar{t}$ and Diboson events. However, the W is not boosted enough to be reconstructed with large efficiency for such events (that is a benefit in this case). The following appendix investigate the possibility of fitting the shape of the mass distribution of the second (resolved) W.

The fit of the shape of MC samples to data is discussed in Section 9.7. A simpler (but faster) approach is to fit the background with a well-defined function (of a simple explicit form) and use the simulated samples only for the signal. The set of selection criteria have to be chosen carefully for this purpose in order to have the background shape correctly described by the function. Although this approach could not be officially used, until the behaviour of the background following the function is explained, it can at least serve as a quick test on whether some sensitivity can be gained by the usage of the shape fits.

Building Transverse Mass Distribution For Fit

The pre-selection criteria of the 0-lepton analysis have been applied as discussed in Section 7.5. In addition to them, there are requirements on:

- the transverse momentum of the first jet $p_{\mathrm{T}} > 100$ GeV
- the transverse momentum of the second jet $p_{\mathrm{T}} > 60$ GeV
- the presence of exactly one unresolved W in the event
- the difference between transversal angles of Missing Energy and nearby jet: $\min_i \Delta\phi(\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}, j_i) > 0.2$

The fit is then performed on the transverse mass, see Eq. 6.4, of the second W candidate. The invariant/transverse mass distribution for the first W candidate is suffering excessively from background, i.e. there are usually significantly less signal events with respect to the background and one can not profit from the shape. However, no large background suppression is needed, since the value of $\Delta\chi^2$ is not dependent on the amplitudes in the fitted bins but only on their deviations.

If the requirement on the presence of an unresolved-W is fulfilled, the set of the remaining jets (without the jet identified as the unresolved W) is investigated. In this case, a slightly different approach to the di-jet mass of the second (resolved) W is used than the one described in Chapter 9: two jets with the highest p_{T} in the event are found and the transverse mass is constructed from them. The transverse mass is chosen due to its better shape that can be easily parametrized using an analytical function. This distribution makes a clear peak for the signal points with large $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0)$. The peak can be interpreted as a consequence of the presence of the second (resolved) W in the decay chain.

Extracting Expected Exclusion Limits

The aim of the following procedure is to find the expected exclusion limits for the $\tilde{q}\tilde{q}$-1step model for which the SR 4jW has the largest impact. The total SM background is composed of W+jets, Z+jets, Diboson and $t\bar{t}$. They are used to represent the data, marked by blue points with error bars in the following figures. The signal benchmark point is chosen such that it contains boosted Ws in the final state. This is achieved with the choice of its mass parameters: $m_{\tilde{q}} = 600$ GeV, $m_{\tilde{\chi}_1^{\pm}} = 590$ GeV and $m_{\tilde{\chi}_1^0} = 60$ GeV.

The m_{T} distribution computed as described in the previous paragraphs is used for a fit and a χ^2 is derived. The formula used for the fit consists of the quadratic function of data and a signal term: $ax^2 + bx + c$ + $N\times$signal, where N plays the role of the strength parameter μ_s and (a, b, c) are the parameters of the fit. Firstly, the distribution is fitted with the strength parameter N set to 1, which corresponds to the case when the given SUSY model is realized (*null hypothesis*)[1]. Secondly, N is left free as a parameter of the fit (*alternative hypothesis*). ROOT [82] is the analysis framework used to perform the fit. Both hypothesis have been fitted with ROOT::TGraph::Fit method and the difference of χ^2 between these hypothesis $\Delta\chi^2$ has been calculated using ROOT::TF1::GetChisquare. The fit for the given benchmark point is shown in Fig. A.1 on the right. To ensure that the quadratic part describes well the total background, we provide also the fit with N fixed to 0, see Fig.A.1 on the left.

---

[1] In case that we would like to disprove MC background only hypothesis than *null-hypohtesis* would be set with $N = 0$. The procedure performed here corresponds to the exclusion fit.

The same procedure can be performed for all available points of the $\tilde{q}\tilde{q}$-1step signal grid. The p-value is provided by ROOT::TMath::Prob method based on the known χ^2 distribution followed by ROOT::TMath::NormQuantile method in order to compute significance and set the expected exclusion limits, see Fig. A.2. One can see that the expected limits of the cut-and-count method of the 0-lepton analysis are not reached. The reasons for that can be: 1) The SRs from the previous publication [76] used for this plot performed very efficiently even without any re-optimization. Our selection instead, is not optimized at all and one can not be entirely sure about its efficiency. 2) The fit details (binning, intervals) would have to be investigated to get the best result.

Note to the statistical apparatus standing behind the fit

The χ^2 test was used in the simplified shape fit method instead of the recommended likelihood ratio testing (see Section 6.3 described in [83]). The main difference between the p-value derived here (ROOT::TMath::Prob) and the p-value calculated from the CL used officially, is a proper treatment of the systematic uncertainties in the official likelihood method. Thus, the p-value calculated here is slightly overestimated. On the other hand, there is a larger effect of the systematics uncertainties entering the official method based on MC simulations for the background whereas for this simplified method, the background is determined directly from data. Note that the likelihood method without implementing systematic uncertainties converges to the χ^2 method with increasing statistics according to Wilk's theorem [84].

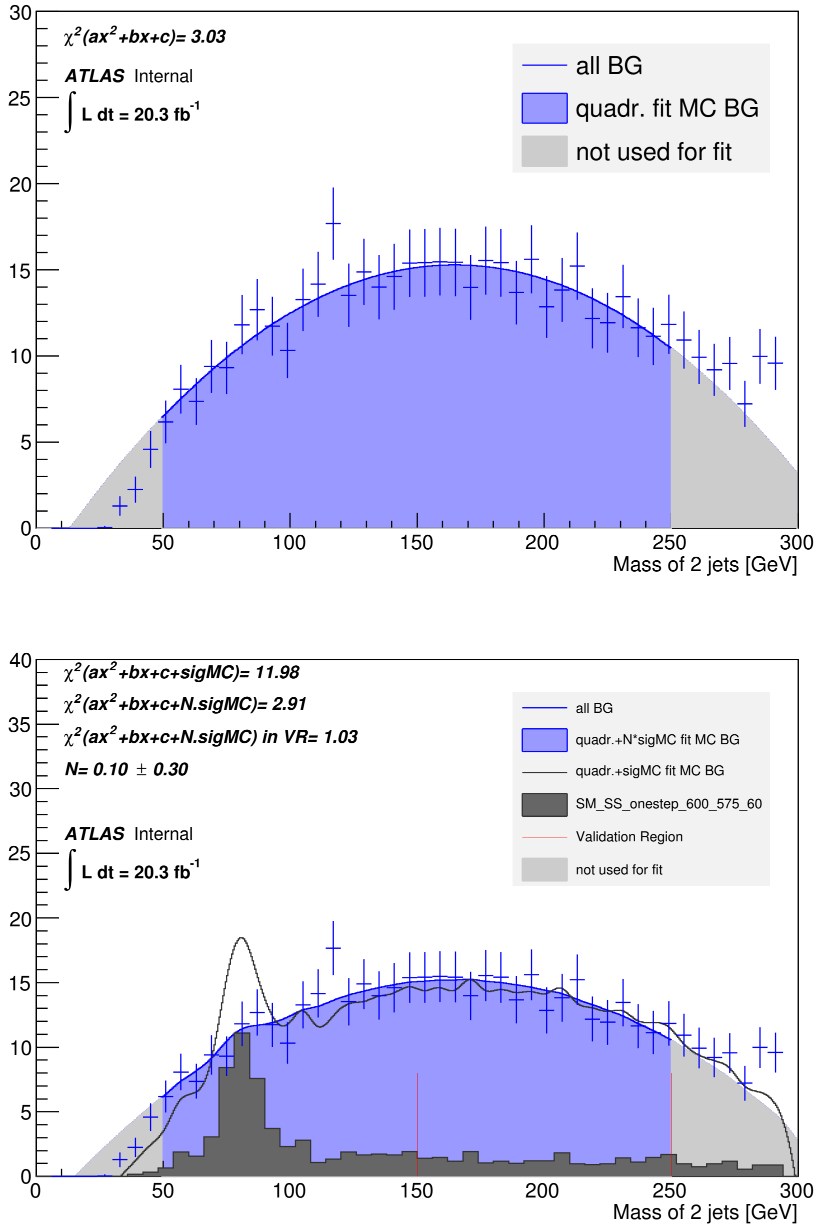


Fig. A.1 At the top: the fit of the total MC m_{T} distribution with a quadratic function ($N = 0$). At the bottom: example of the fit for one special set of signal parameters ($m_{\tilde{q}} = 600$ GeV, $m_{\tilde{\chi}^{\pm}} = 575$ GeV, $m_{LSP} = 60$ GeV) for both *null/alternative* hypothesis. Values of χ^2 are shown at the top left corner of the figure.

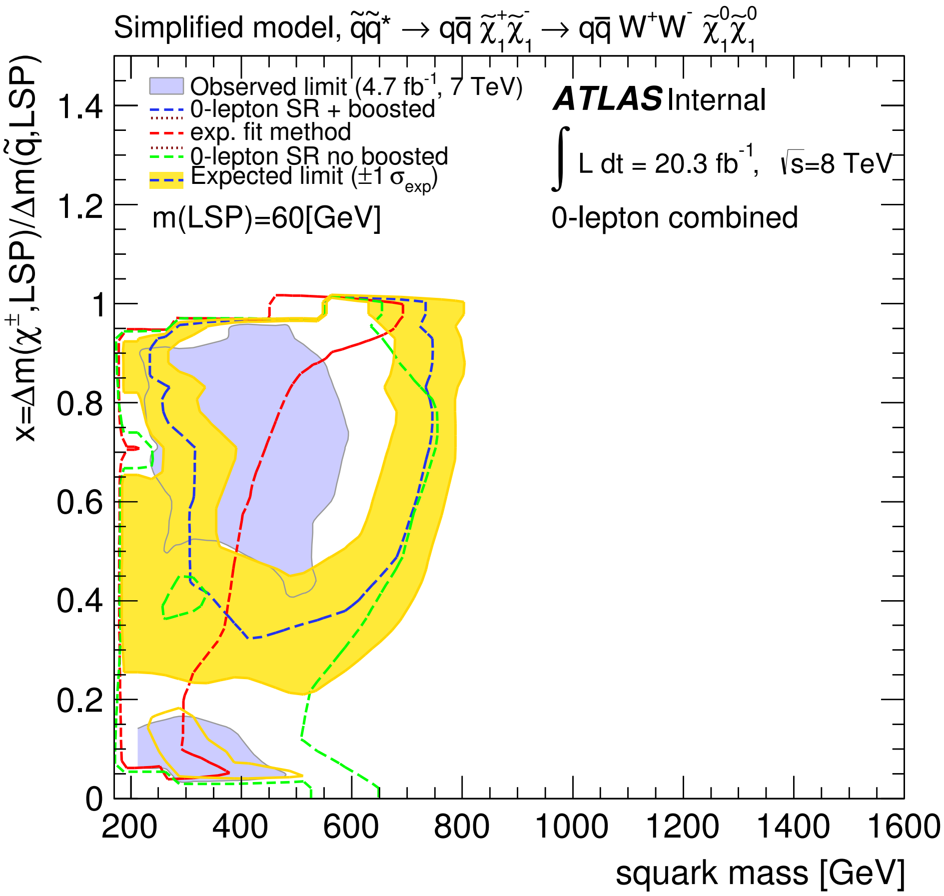


Fig. A.2 Expected exclusion limits for a simplified model $\tilde{q}\tilde{q}$-1step: the red line represents the shape fit method discussed in this appendix, the green line represents 0-lepton analysis without W-SRs, the blue line represents the current 0-lepton analysis with W-SRs. The results are not in full agreement with the final result from the publication [64], since this test was performed during an earlier phase of the 0-lepton analysis.

B. Study of τ-veto

The suppression of the SM background events is always very welcome from two reasons: firstly, that the significance of the potential signal rises with a lower background; secondly, the effect of the systematic uncertainties on the background estimation can be minimized. The background can be suppressed simply by requiring some tighter selection, while the event yield of the signal in SRs should remain unchanged. This can be achieved by vetoing the events for which there are clear indications that they belong to SM. Such an indication can be the presence of a specific particle, the τ lepton in our case. The τ can be present also in the final states of several signal models, but let us assume that we aim to improve the sensitivity for $\tilde{q}\tilde{q}/\tilde{g}\tilde{g}$-direct models only with this veto. The following study was performed during the early phase of my Ph.D., thus the framework of the 0-lepton to that date was used [77] (see also Table of SRs C.1). The integrated luminosity available to the date of this study was $L = 14$ fb^{-1}.

Although the τ is a lepton, it is often reconstructed as a jet in a detector due to its dominant decays into one (45.65 %, [29]) or three (12.74 %, [29]) charged π mesons. Thus, one can expect to see one or three tracks in the Inner Detector (or in any case a low number of tracks from the other hadronic decays of the τ). The τ is mostly produced in the decay of the W boson: $W^{\pm} \rightarrow \tau^{\pm}\nu_{\tau}$ which can be produced directly in the W+jets events, in the Diboson events or in the decays of the top quark from the $t\bar{t}$ or single-t production processes. In all of these cases, the W is expected to carry significant (more than 100 GeV) amount of energy and thus the direction of the τ-jet is close to the direction of the missing energy (created by the ν_{τ} only, if it is the only ν in the event). Furthermore, the invariant mass of the system of the missing energy and the τ, and consequently the transverse mass, rarely exceeds 100 GeV. These assumptions lead to the possibility of reconstructing the τ-jet using the number of tracks N_{trk}, the $\Delta\phi_{min}$ corresponding to the difference of the transversal angles of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and its nearby jet (supposed to be the τ) and the transverse mass m_{T} of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and its closest jet. The m_{T} (see Eq. 6.4) is defined in a way that the jet not related to the $E_{\mathrm{T}}^{\mathrm{miss}}$ (not originating in the same mother particle) usually acquires a larger value of m_{T}. If the requirements on such variables are fulfilled, the closest jet to the $E_{\mathrm{T}}^{\mathrm{miss}}$ is supposed to be the τ. After the

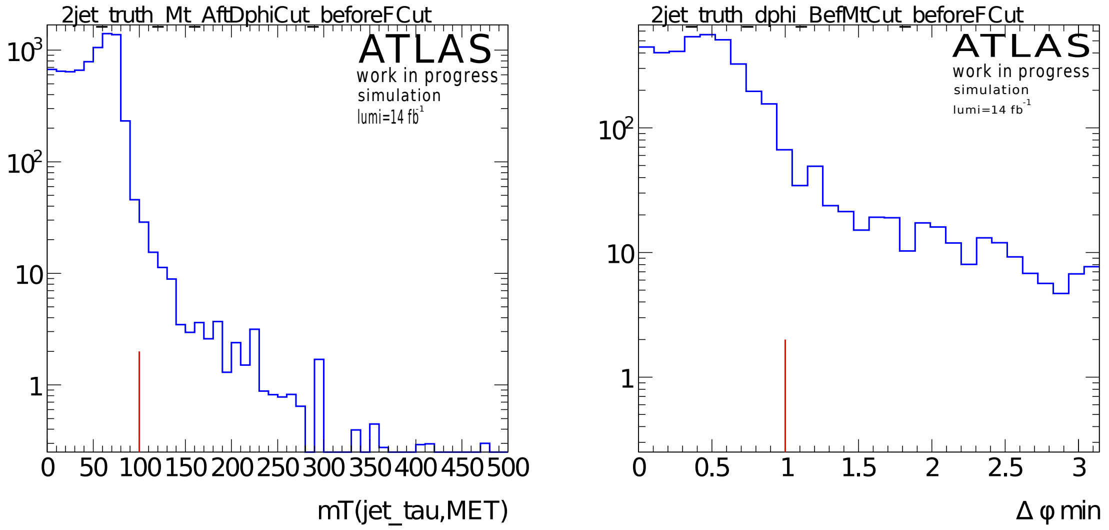


Fig. B.1 The m_{T} and the $\Delta\phi_{\min}$ distributions of the jets with $N_{\mathrm{trk}} = 1 \vee 3$. The selection corresponds to 0-lepton analysis in former SRA, see Table C.1.

identification of the events with the τ candidate, the veto can be applied.
The distributions of the $\Delta\phi_{\min}$ and m_{T} are studied for the $t\bar{t}$ and W+jets samples in the histograms in Fig. B.1 to take a decision about the proper values of the cuts on these variables. These samples have been used in order to have rich statistic of events with the τ in the final state. The histograms are populated only by the jets satisfying the condition $N_{\mathrm{trk}} = 1 \wedge 3$. The choice of the reconstruction criteria is depicted by the red vertical line.
Consequently, well-motivated τ identification criteria can be defined as follows:

- it is a closest jet to the $E_{\mathrm{T}}^{\mathrm{miss}}$
- $N_{\mathrm{trk}} = 1 \vee 3$
- $m_{\mathrm{T}} < 100\mathrm{GeV}$
- $\Delta\phi_{\min} < 1.0$

The jet that passes this criteria is referred to as a τ candidate in the following. In case the veto on the events with the reconstructed τ is applied, the reconstruction itself needs to be validated in terms of agreement of data and MC simulation for all of the kinematic distributions. The η, ϕ and p_{T} distributions of the τ candidates are monitored in Fig. B.2 as well as the distribution of m_{eff}(incl.) in the events where a τ candidate is present. All the pre-selection criteria of the 0-lepton analysis [77] are applied, but none of the SR specific selection criteria are used. Thus the multi-jet background is not reduced as none of the $\Delta\phi$ requirement from Table C.1

are involved. The monitoring of the kinematic distribution did not uncover any unexpected discrepancy between data and MC simulation.

The simulation samples of $t\bar{t}$ and W+jets contain generator level (truth) information about the produced τ leptons. The ΔR between the τ candidate and the truth τ should show a significant peak around $\Delta R = 0$ when the reconstruction runs correctly. This expectation is validated in Fig. B.3.

Once the τ-reconstruction is developed, the τ-veto can be applied additionally to the SRs criteria in Table C.1. The impact of the τ-veto on the upper limits on the excluded cross section for the $\tilde{q}\tilde{q}$-direct ($\tilde{q} \rightarrow \tilde{\chi}_1^0 q$) and $\tilde{g}\tilde{g}$-direct ($\tilde{g} \rightarrow \tilde{\chi}_1^0 q\bar{q}$) simplified models is shown in Fig. B.4 and Fig. B.5. The lower the number in the plot at the given point, the better the exclusion power. The τ-veto lowers the factors at almost each point of both signal models. However, the exclusion limit line, the line passing the points with value 1.0 in the plots, remains unchanged and no extension of the excluded area is observed. Additionally, the systematic uncertainties potentially affecting τ reconstruction, such as tracking efficiency, were not applied, which would worsen the results even more. Thus, the τ-veto investigated here has never been officially used in the publications, in order to keep the analysis as simple as possible.

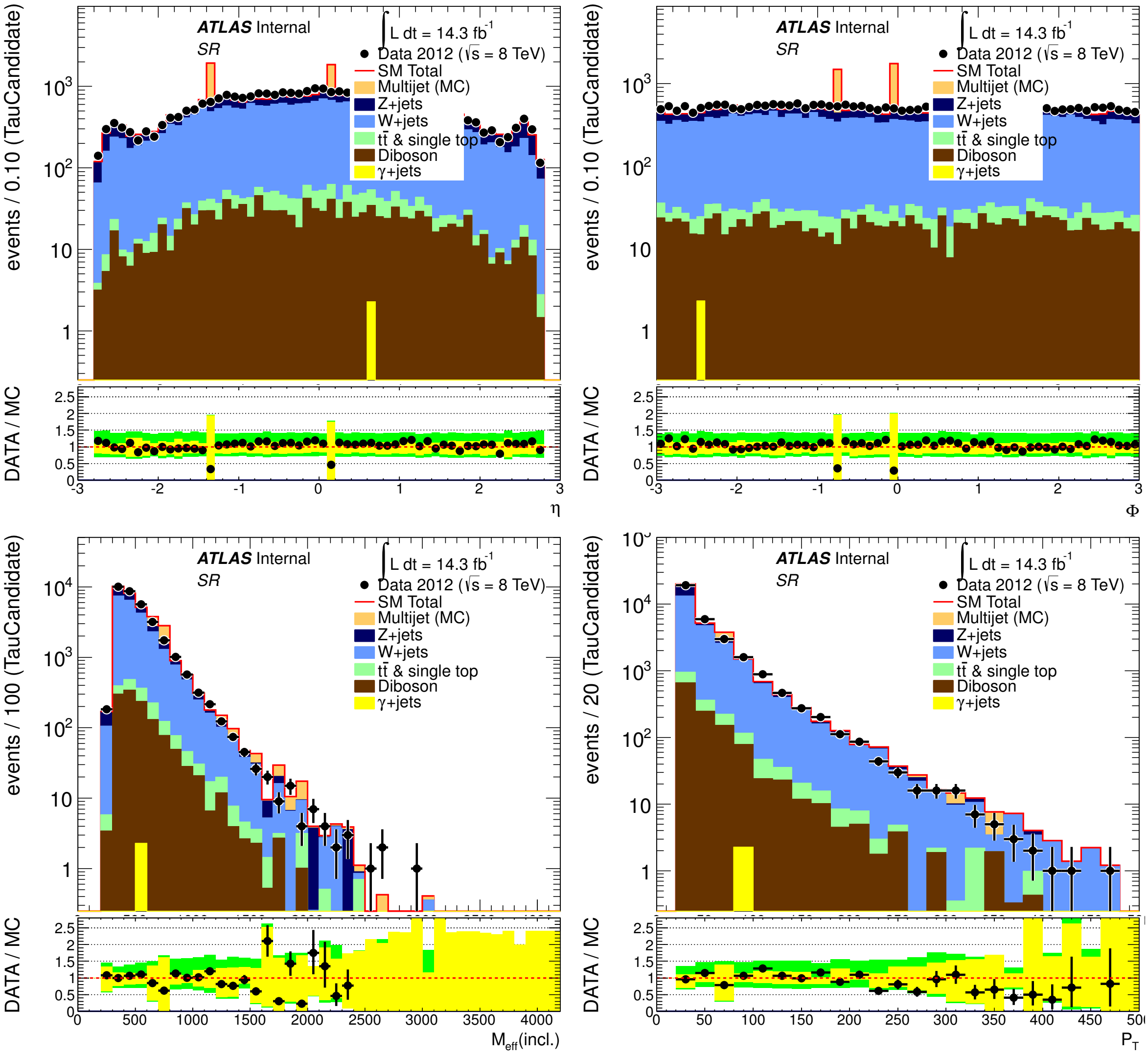


Fig. B.2 The kinematic distributions of the τ candidate: η upper left, ϕ upper right, m_{eff}(incl.) bottom left and p_{T} bottom right. All the pre-selection criteria are applied, without any SR specific selection criteria.

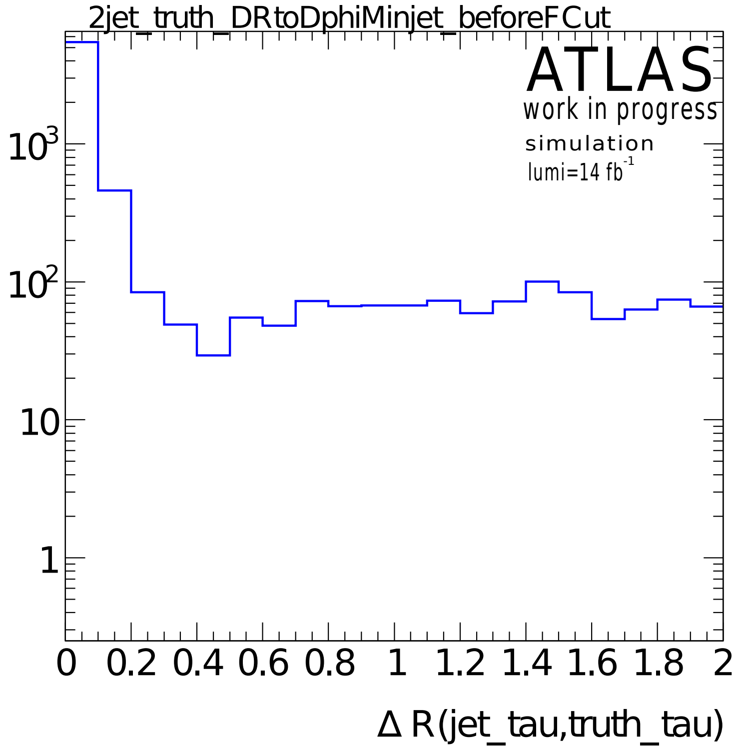


Fig. B.3 The ΔR distribution between truth τ and reconstructed τ for $t\bar{t}$ and W+jets simulation samples. Most of the statistics (92 %) populates the first two bins, which indicates the correctness of the τ reconstruction.

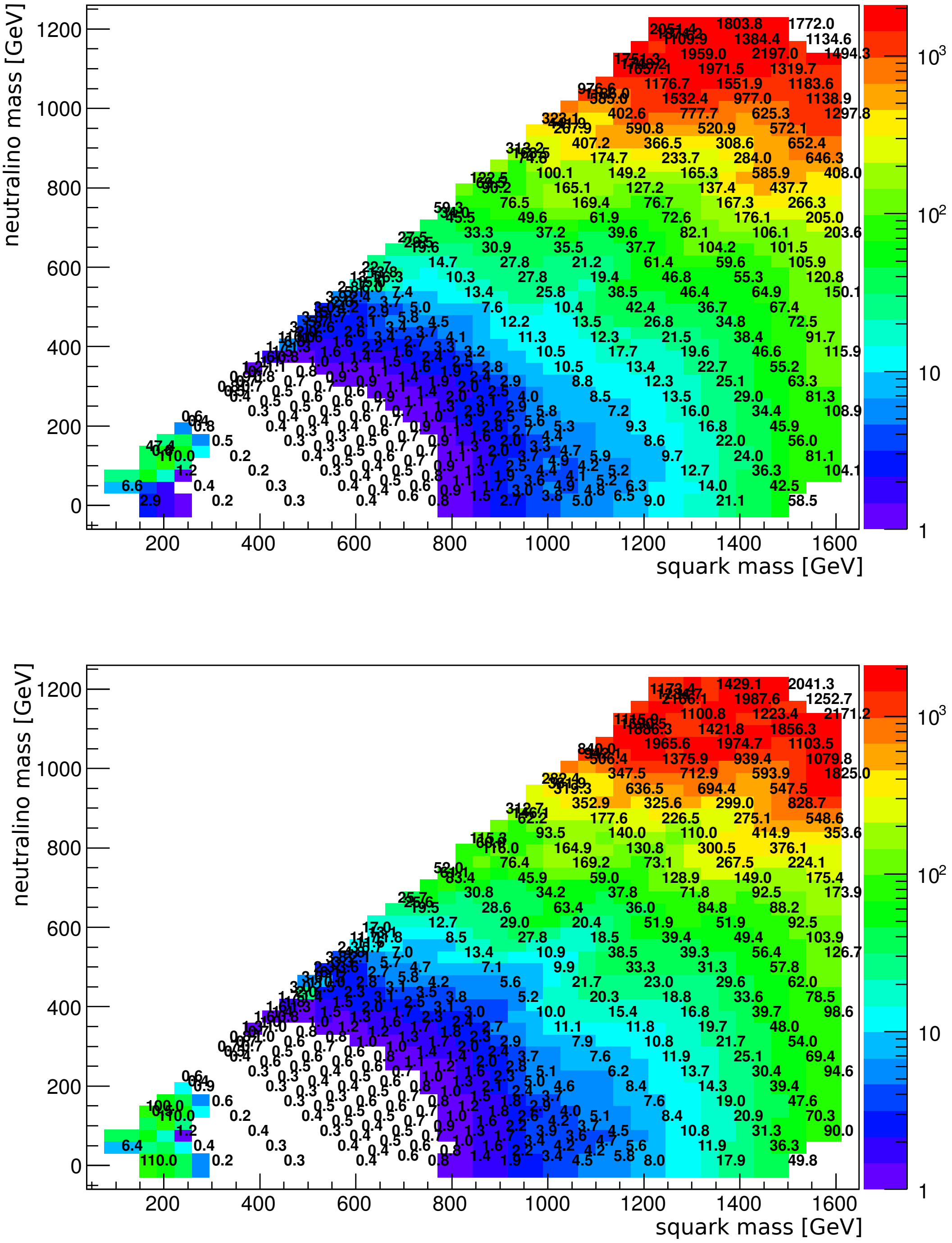
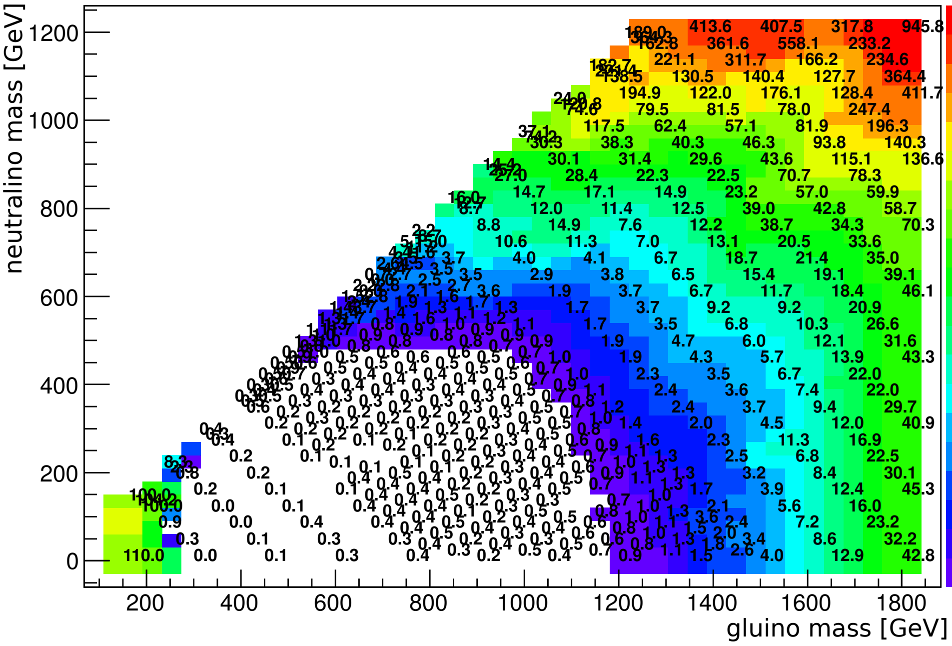


Fig. B.4 The comparison of the upper limits on the excluded cross section for the *ss*-direct model: without the τ-veto in the upper plot, with the τ-veto in the bottom plot.



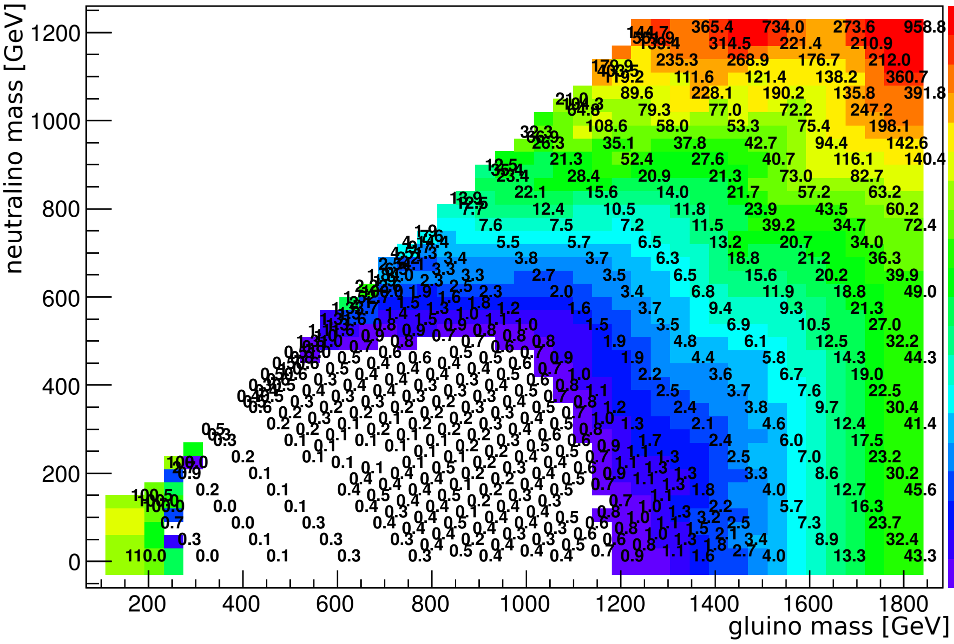


Fig. B.5 The comparison of the upper limits on the excluded cross section for the *gg*-direct model: without the τ-veto in the upper plot, with the τ-veto in the bottom plot.

C. Earlier Selection Criteria in Detail

While the pre-selection criteria discussed in Section 7.5 remained almost the same, the selection criteria evolved due to the development of new SRs and due to the optimisation procedure. The set of the selection criteria used in the earlier phase of this thesis is listed in Table C.1. One can compare it with the newer set in Table 8.1.

Requirement	Channel				
	SRA	SRB	SRC	SRD	SRE
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV] $>$	160				
$p_{\mathrm{T}}(j_1)$ [GeV] $>$	130				
$p_{\mathrm{T}}(j_2)$ [GeV] $>$	60				
$p_{\mathrm{T}}(j_3)$ [GeV] $>$	-	60	60	60	60
$p_{\mathrm{T}}(j_4)$ [GeV] $>$	-	-	60	60	60
$p_{\mathrm{T}}(j_5)$ [GeV] $>$	-	-	-	60	60
$p_{\mathrm{T}}(j_6)$ [GeV] $>$	-	-	-	-	60
$\Delta\phi(j_{1,2,(3)}, E_{\mathrm{T}}^{\mathrm{miss}}) >$	0.4				
$\Delta\phi(j_{i>3}, E_{\mathrm{T}}^{\mathrm{miss}}) >$	–		0.2 (if $p_{\mathrm{T}} > 40$[GeV])		
$E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(Nj) >$	0.3/0.4/0.4	0.25/0.3/–	0.25/0.3/0.3	0.15	0.15/0.250.3
$m_{\mathrm{eff}}(\mathrm{incl.})$ [GeV] $>$	1900/1300/1000	1900/1300/–	1900/1300/1000	1700/–/–	1400/1300/1000

Table C.1 The definitions of SRs used in the conference note [77]. The five basic SRs are subdivided into tight/medium/loose selection according to their m_{eff} and $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ requirements specified in the last two rows, separated by "/" symbol, respectively.

D. Additional Information on Reclustering

This appendix contains additional (explanatory) information to Chapter 10.

D.1 JMS and JMR

One of the outcomes of the study of the reclustering procedure is that the JMS and JMR of the sub-jets can not be neglected with respect to the JES and JER for a large-p_T RT jet. Furthermore, this statement is even stronger when we use a larger size of the small jet cone, see Fig. D.1 and D.2. There are three major effects causing these two phenomena, listed in this appendix.

Not Proper Cone Size

Let us imagine that the jet was created by two quarks with $\Delta R < 0.2$. Then most of the physics of interest happens inside the cone $R = 0.2$ and the use of larger cones can only lead to contamination, such as from pile-up and UE. Such additional clusters have a large impact on the mass of the jet, since they are distributed equally over the entire distance from the cone center. This effect is suppressed by proper cleaning and calibration of the jets but one can never expect 100% efficiency. Therefore it is worth to use a suitable cone size, otherwise the mass of the jet is increased.

Number of Sub-jets

As we could learn from Fig. 10.2, the number of sub-jets is getting smaller with larger p_T and related boost of the RT jet. Evidently, the larger the number of sub-jets in the RT jet, the lower the impact of the mass of the sub-jets on the final mass of the RT jet. The same logic applies to mass and energy uncertainties.

Kinematic Reasons

The mass M of the final jet, composed of two sub-jets with properties m_1, E_1, m_2, E_2 and closing angle θ, can be calculated as:

$$M^2 = 2E_1E_2 + m_1^2 + m_2^2 - 2\sqrt{E_1^2 - m_1^2}\sqrt{E_2^2 - m_2^2}\cos\theta \tag{D.1}$$

let us assume that there is an uncertainty σ_E on E_1 and an uncertainty σ_m on m_1 (e.g. JES and JMS). The uncertainties of E_2 and m_2 behave symmetrically and therefore can be neglected for a while. The total σ_{M^2} can be derived from Eq. D.1:

$$\sigma_{M^2}^2 = \left(2E_2 - 2E_1\frac{\sqrt{E_2^2 - m_2^2}}{\sqrt{E_1^2 - m_1^2}}\cos\theta\right)^2 \sigma_E^2 + \left(2m_2 + 2m_1\frac{\sqrt{E_2^2 - m_2^2}}{\sqrt{E_1^2 - m_1^2}}\cos\theta\right)^2 \sigma_m^2 \tag{D.2}$$

Since the θ angle expresses the boost of the RT jet, the impact of the second term gets larger when the boost is larger, while the impact of the first term gets smaller with the larger boost (due to the signs). Thus, the impact of mass uncertainties is larger when the boost is larger. Note that this argument is valid only in the case that m_1 or m_2 are not negligible.

D.2 The W-tagging Efficiency

The W-finder described in Chapter 9 works efficiently in terms of comparison between reconstructed and truth Ws. Such a study has been performed in the later phase of the doctoral studies with W' simulated samples, merged together and weighted properly in order to have rich statistic over broad spectra of the $p_{\mathrm{T}}(W)$. The W' stands for a hypothetical heavy boson decaying into pair of W and Z bosons. Note that the model itself is not the object of the study. The complete truth information from the generator about individual particles is available in this sample. Thus, the W reconstructed with the calorimetric jets can be matched (if correctly reconstructed) to the truth W. Furthermore, this sample has exactly one large p_{T} W in the final state decaying hadronically. The Z-bosons are assumed to decay leptonically. Finally, the efficiency of the W reconstruction is assumed to depend on the p_{T} of the W and not on the physics simulated in the sample.

The result of such a study is depicted in Fig. D.3. Each W in the final state means one entry into the histogram weighted properly for the cross-section of the sample. The left two bins correspond to the cases when W was correctly reconstructed either as resolved or unresolved, which means that the jet four-vector of the W, or the sum of the four-vectors of the W's jets in the resolved case, matches to the truth W within $\Delta R < 0.2$. The two bins on the right correspond to the case when the reconstructed and the truth Ws do not match within $\Delta R <$

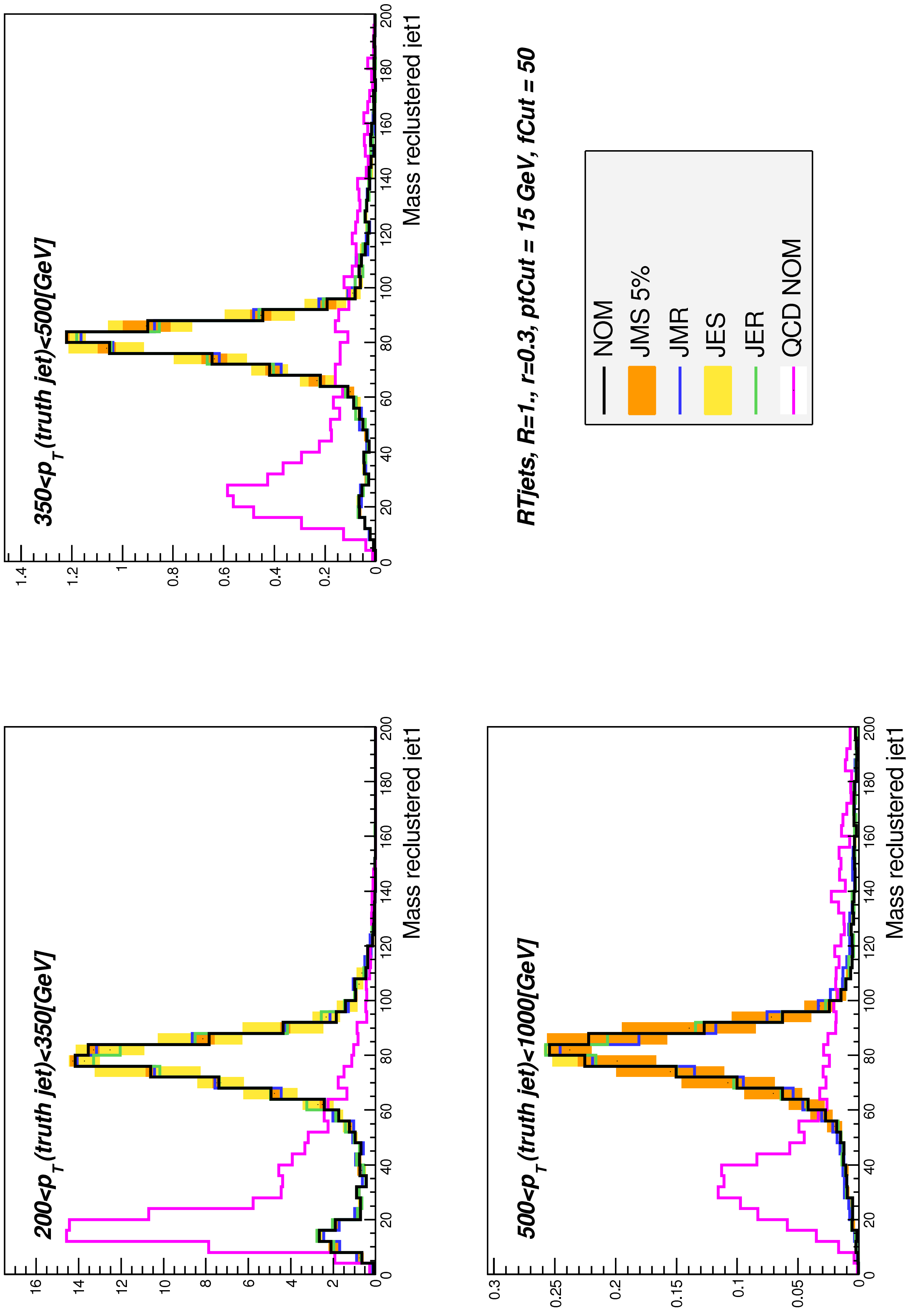


Fig. D.1 The invariant mass distributions of the RT jets for three regions of p_{T} of the RT jet for small jet cone size 0.3.

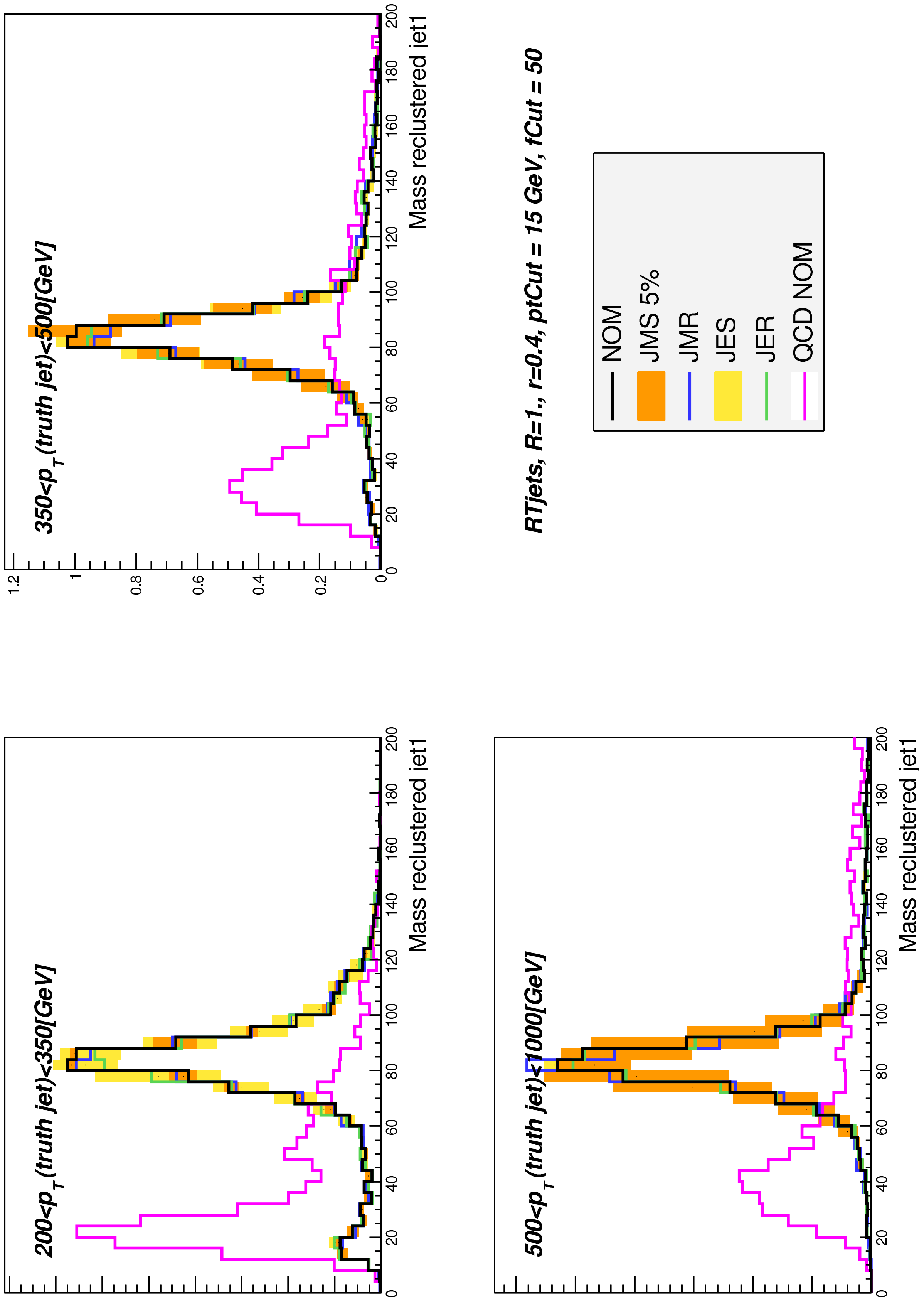


Fig. D.2 The invariant mass distributions of the RT jets for three regions of p_T of the RT jet for small jet cone size 0.4.

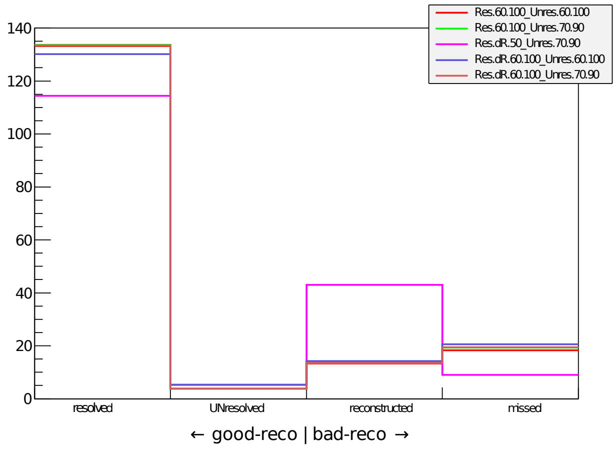


Fig. D.3 Correctly and not correctly reconstructed W bosons for different options of the W mass window cut and specifics of the resolved W reconstruction.

0.2. The published version of the W-finder corresponds to the red color. The other colors represent different options of the mass cuts implemented in W-finder. One can see, that there are approximately 130 correctly reconstructed Ws, while roughly 15 are reconstructed by W-finder but not matched to any truth W. Considering the last bin, out of all $\sim$160 (integral of the histogram) truth Ws only 20 have not been correctly reconstructed. Note that these numbers are very rough. In the analysis, the selection criteria for a given SR can impact these numbers as well as the method of the transfer factors used to calculate the event yield in the SRs.

E. p_{T}(jet1) Distribution for Specific Signal Models

In order to have more dense coverage of the signal 1step models containing the boosted Ws, a few new points in the signal parametric space were generated with Pythia [40] + MadGraph [48]. The complication arose when providing the mandatory plots of the p_{T} distribution of the hardest jet before the detector simulation. Several signal points, also the one studied here: $m(\tilde{g}) = 1200$ GeV, $m(\tilde{\chi}_1^{\pm})$ =972 GeV and $m(\tilde{\chi}_1^0) = 60$ GeV, show a double peak, see the pink distribution in Fig. E.1. This effect could be explained by the origin of the first jet. If the jet originates from the W (green color), it clearly corresponds to the second peak at larger values of p_{T}. If the first jet does not originate from the W (blue color) or from one of its decay products (red color), the p_{T} is peaking at lower values. Note that this study is dependent on the cone used in the anti-$k_{\perp}$ algorithm. The cone size of 0.2 was used for the Fig. E.1.

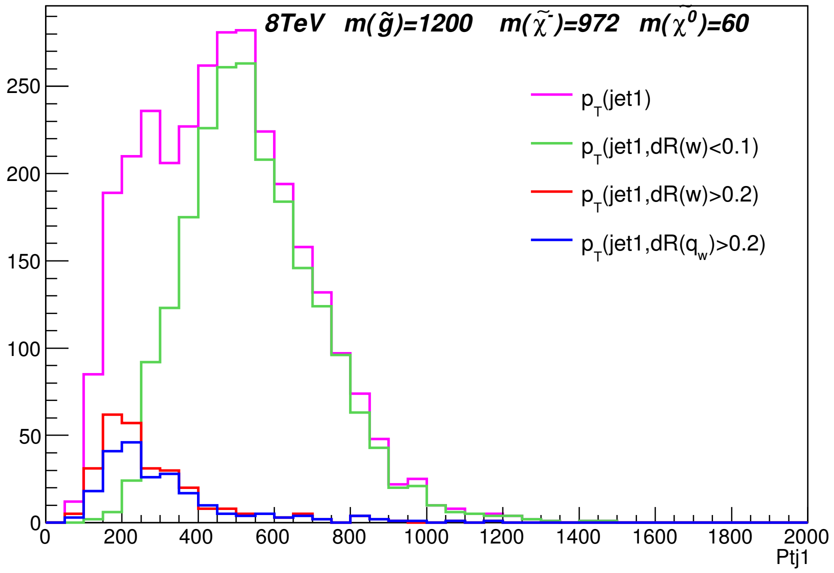


Fig. E.1 The p_{T} of the first jet for the requested signal $\tilde{g}\tilde{g}$-1step model in terms of its origin distinguished by the color. The truth jets (after Pythia+MadGraph simulation) are clustered with anti-$k_{\perp}$ algorithm with the cone radius 0.2.

.

List of figures

List of tables