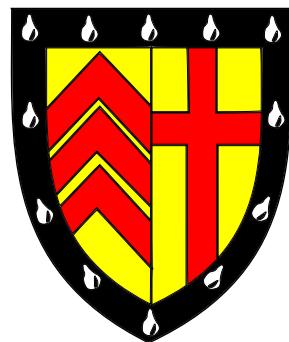


Identifying fake leptons in ATLAS while hunting SUSY in 8 TeV proton-proton collisions

Thomas P. S. Gillam

Clare College



A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy
January 2015

Identifying fake leptons in ATLAS while hunting SUSY in 8 TeV proton-proton collisions

Thomas P. S. Gillam

Abstract

For several theoretically and experimentally motivated reasons, supersymmetry (SUSY) has for some time been identified as an interesting candidate for a theory of fundamental particle physics beyond the Standard Model. The ATLAS collaboration, of which I am a member, possess a detector emplaced in the Large Hadron Collider experiment at CERN. If SUSY does in fact describe our universe, then it is hoped that evidence of it will be visible in data collected in the ATLAS detector.

I present an analysis looking for a particular signature that could indicate the presence of SUSY; events containing two like-charge leptons (e or μ). This signature benefits from having both low Standard Model backgrounds as well as potential to observe several SUSY scenarios, particularly those involving strong production processes. These include pair production of squarks and gluinos. The latter of these are particularly relevant for the analysis presented herein since gluinos are Majorana fermions; hence they can decay to produce like-charge leptons. The analysis considers several SUSY production topologies determined from a variety of simplified and phenomenological models.

One of the core pieces of any ATLAS analysis is estimating the expected backgrounds in the signal regions. These backgrounds arise

both from known Standard Model production processes that can produce the same final state as the SUSY models being targeted, as well as detector mismeasurement effects. One important background in the like-charge analysis is that of “fake” leptons; these are jets that have been misclassified as either electrons or muons by the reconstruction algorithms. A large portion of this thesis is dedicated to introducing novel techniques for robustly estimating these backgrounds, and evaluating their relative performance.

Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This dissertation does not exceed the word limit for the Degree Committee of the Faculty of Physics and Chemistry.

Thomas P. S. Gillam

Acknowledgements

This work would certainly not have been possible, and almost as surely not have been half as much fun, were it not for all those that have supported me and this research over the last few years.

I would especially like to thank my supervisor, Christopher Lester, for providing a steady stream of invaluable ideas, critique, and advice; both directly related to my research, as well as more tangentially. My education in Go, wall-building, and gliding has been as much extended as that in Monte Carlo techniques, probability, and statistics.

The wider support of the Cambridge HEP research group has provided great benefit, and I would like to thank it, as well as our colleagues in DAMTP connected via the SUSY Working Group. It would be impractical to list all those, both in Cambridge and elsewhere, who have been of assistance, however I'd particularly like to thank Teng Jian Khoo, Mireia Crispín Ortuzar, Thibaut Mueller, Dave Sutherland, Sophie Renner, and Damien George for many illuminating discussions, advice, and friendship. I would also like to thank those members of ATLAS with whom I have worked closely – particularly David Côté, Otilia Ducu, Julien Maurer, and Jean-François Arguin.

It also goes without saying that my thanks go to all present and past members of the ATLAS collaboration, without whom the experiment would not exist. Such a large collaborative effort has few parallels in science, and it has been a pleasure to be a part of it.

Outside of my academic work, I have much gratitude for the numerous musical groups in Cambridge that I have been so fortunate as to be able to play with, and also for all those friends who have made my time in Cambridge most enjoyable.

Finally, and by no means least, I would like to thank my parents, without whose support I doubtless would not have had the opportunity to pursue this PhD.

Preface

The problem of estimating the expected number of events which contain fake leptons is not new, and ATLAS analyses have been contending with it throughout the 2012 data-taking period. Throughout this work such a method will be simply referred to as a ‘fake estimation technique’. One of the most widely-used methods of estimating the impact of this background is known as the matrix method, however it is readily admitted that the behaviour of the method is not necessarily always understood by the analyses that use it:

I have seen many analyses that simply used the `FakeLeptBkg` package without even taking a look at the real and fake efficiencies, just assuming the [matrix method] works perfectly.

(Ximo Poveda, ATLAS SUSY Background Forum Convenor, Jan. 2015)

I started working in this area after discovering the problem myself in the context of an analysis searching for evidence of supersymmetry in events with like-charge lepton signatures. Due to the particular selection requirements employed, the existing matrix method could not be directly applied – an extension was therefore necessary. This later led to further investigations into how the statistical robustness of the procedure could be improved through more fundamental changes to the method, experimenting with both maximum likelihood and Bayesian approaches.

The like-charge lepton SUSY analysis, performed using the 8 TeV data collected from ATLAS during 2012, forms a most interesting study and an integral part of this work. The conclusions provide relevant results to help exclude a large variety of supersymmetric signatures, complementary to results from other ATLAS analyses published with the latest dataset. It is a testing time for SUSY, and these results are amongst those that start to put a strain on the desired ‘naturalness’ of the models.

This thesis therefore aims to provide the following:

- A full development of several ‘fake estimation’ techniques, some of them new or extended beyond those used prior to this work. To the author’s knowledge to date, even the existing techniques have not been fully documented publicly.
- A questioning of the assumption that the matrix method will always ‘just work’, and provision of comparisons with alternative methods in controlled scenarios.
- An analysis of the 2012 ATLAS dataset which searches for evidence of supersymmetry in events with like-charge or three leptons, and places strong constraints in many scenarios.

In addition to the phenomenological applications of the physics results presented herein, I hope that future analyses within ATLAS, and in principle CMS, can improve robustness in fake estimation for Run 2 as a result of the studies performed here.

Contents

1	The ATLAS experiment and the LHC	3
1.1	The Large Hadron Collider	3
1.2	The ATLAS detector	5
1.2.1	Co-ordinate system	6
1.2.2	Inner detector	7
1.2.3	Calorimeters	9
1.2.4	Muon system	12
1.2.5	Triggers	16
1.3	Software environment	16
1.4	Defining analysis objects	19
1.4.1	Leptons	19
1.4.2	Jets	21
1.4.3	Overlap removal	23
1.4.4	Missing transverse momentum	24
2	Introduction to Supersymmetry	25
2.1	Limitations of the Standard Model	25
2.1.1	Dark matter	26
2.1.2	Baryon asymmetry	27
2.1.3	Anomalous magnetic moment of the muon	28
2.1.4	Neutrino masses	29
2.1.5	Hierarchy problem	30
2.2	Supersymmetry	31
2.2.1	Theoretical principles and motivation	32
2.2.2	Supersymmetric phenomenology	34
2.2.3	Some specific terminology	39
2.3	Fixing the Standard Model with SUSY	40
2.3.1	Dark matter	40

2.3.2	Anomalous magnetic moment of the muon	40
2.3.3	The hierarchy problem	41
2.3.4	Gauge-coupling unification	42
2.3.5	<i>CP</i> -violation	42
2.4	Constraints on supersymmetry	42
2.4.1	Proton lifetime constraints	42
2.4.2	Flavour-changing neutral currents	43
2.4.3	Naturalness	43
2.4.4	Additional constraints	44
3	The problem of fakes	45
3.1	Motivation and overview	45
3.2	Processes for faking electrons and muons	46
3.2.1	Electrons	46
3.2.2	Muons	47
3.3	Problem statement	47
3.4	Modelling fake events	48
3.4.1	‘Tight and Loose’ model	48
3.4.2	Simple events	48
3.4.3	Complicated events	50
3.5	Potential solutions	52
4	Data-driven methods for estimation of fake backgrounds	55
4.1	Introduction	55
4.2	Common definitions	56
4.3	The vanilla matrix method	57
4.3.1	Single object events	57
4.3.2	Efficiencies varying with kinematic quantities	59
4.3.3	Events with two or more objects	60
4.4	Generalised matrix method	61
4.4.1	Summation formalism	62
4.4.2	Propagation of uncertainties	65
4.4.3	Applications in lepton and jet systems	69
4.5	Likelihood-based fake estimation	71
4.5.1	Frequentist limit setting procedures	71
4.5.2	Limit setting with the matrix method	74
4.5.3	An extended likelihood method	75

4.5.4	Emergence of the matrix method	78
4.6	Bayesian fake rate posterior	79
4.6.1	Sketch of the method	80
4.6.2	Sampling methods	81
5	An analysis of ATLAS data for like-sign lepton events	89
5.1	Motivation	90
5.2	Data and Monte Carlo samples	90
5.2.1	The 8 TeV ATLAS dataset	90
5.2.2	Background samples	91
5.2.3	Signal samples	92
5.2.4	MC sample tuning and corrections	93
5.3	Event selection	94
5.3.1	Trigger selection	95
5.4	Signal regions	96
5.5	Estimating the backgrounds	99
5.5.1	Prompt backgrounds	99
5.5.2	Fake lepton contributions	102
5.5.3	Charge-flipped like-sign lepton events	109
5.5.4	Validating the data-driven background estimates	116
5.5.5	Cross-check of the fake b -jet contribution in SR3b with the matrix method	116
5.6	Systematic uncertainties	129
5.7	Fitting and limit setting procedure	132
5.7.1	Model independent discovery & upper limits	132
5.7.2	Model-dependent exclusions	133
5.8	Results	134
5.8.1	Yields in signal regions	134
5.8.2	Model-independent limits	136
5.9	Interpretation of results	138
5.9.1	Simplified models	139
5.9.2	Phenomenological models	150
6	Comparing fake estimation methods	157
6.1	Toy MC: matrix method vs likelihood	157
6.1.1	Simple example – two leptons, two categories	158
6.1.2	Harder example – two leptons, eight categories	162

6.1.3	Conclusions	165
6.2	ATLAS MC: matrix method vs Bayesian posterior	166
6.2.1	Dealing with many categories	166
6.2.2	Description of tests	167
6.2.3	The status quo: matrix method	169
6.2.4	The Bayesian posterior	172
6.2.5	Comparisons with ‘truth’ efficiencies	176
6.3	Conclusions	178
7	Closing remarks	181
Bibliography		187
Acronyms		203

Chapter 1

The ATLAS experiment and the LHC

1.1 The Large Hadron Collider

As of 2014, the Large Hadron Collider (LHC) [1] is the highest energy proton-proton collider in existence. It is a synchrotron that maintains two counter-rotating beams of proton bunches. Focussing magnets can ‘pinch’ these beams together at any of the four interaction points, where experiments are housed; in these locations interactions will occur between the protons. The design centre-of-mass collision energy of the accelerator is $\sqrt{s} = 14$ TeV, however at the time of writing the majority of the data available for analysis has been taken at $\sqrt{s} = 8$ TeV. Additionally, whilst the LHC is designed to run with a bunch spacing of 25 ns, throughout the 2012 run twice this, i.e. 50 ns was used. This was compensated for by putting more protons in each bunch, however this leads to a larger number of interactions expected per bunch crossing, presenting analysis challenges in the form of ‘pile-up’. With the design parameters, the bunch crossing rate is expected to be 40 MHz, whereas for the 2012 run (and all data used in this thesis), the rate was 20 MHz.

In fact, the LHC is only the final and largest component of a multi-stage process to accelerate the protons from rest – the entire set of accelerators that are used is shown in Figure 1.1. Protons are injected into the LHC ring in bunches at an energy of 450 GeV until the beams are full; that is, enough bunches are created such that the target bunch spacing is reached. At this point both beams are accelerated up to half the target centre of mass energy, which was 4 TeV for the 2012 dataset – this process typically takes

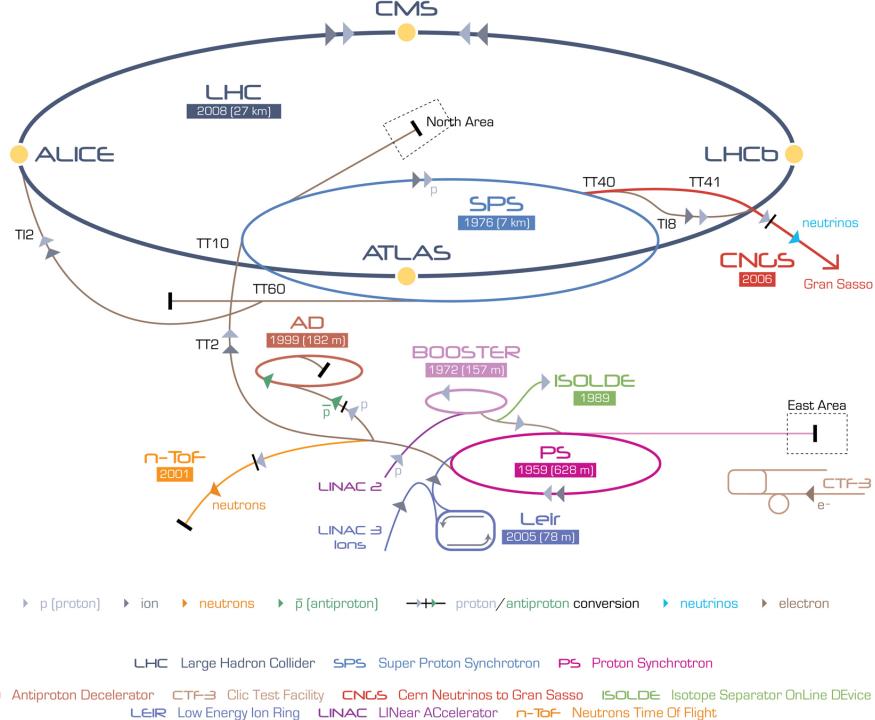


Figure 1.1: Overview of the accelerators supplying the LHC, and the experiments housed at the four interaction points on the main ring [2].

about 20 minutes. The beams are then pinched at each of the interaction points, and interactions subsequently occur for $\mathcal{O}(10)$ hours. At the four interaction points are installed the four main LHC experiments – ATLAS, CMS, ALICE, and LHCb.

The rate of proton interactions occurring is analytically defined in terms of luminosity. Specifically, for a physics process with cross section σ it acts as the constant of proportionality \mathcal{L} to give the event rate,

$$\frac{dN}{dt} = \sigma \mathcal{L}. \quad (1.1)$$

The design luminosity of the LHC is $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Since the precision with which statistical statements can be made typically increases with the amount of data collected, it is useful to know the integral of luminosity over the lifetime of a detector, known simply as ‘integrated luminosity’. The cumulative integrated luminosity delivered to ATLAS is shown in Figure 1.2. The higher delivery rate during the 8 TeV 2012 run compared to that of 2011 is clearly visible.

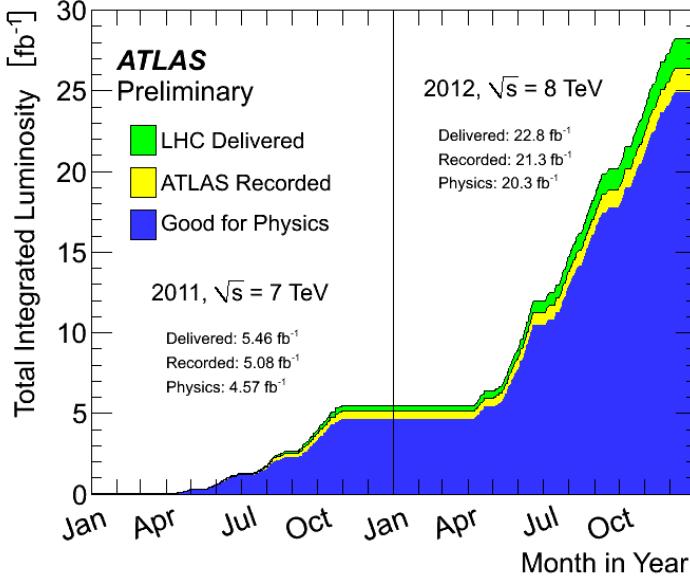


Figure 1.2: Running total of the integrated luminosity delivered to and recorded by ATLAS during the 2011-2012 run of the LHC [3].

1.2 The ATLAS detector

The ATLAS experiment [4] is the largest of those installed at the LHC, weighing in at around 7000 tonnes. Although it is presently, in the public eye at least, very much associated with the 2012 discovery of the Higgs boson, it is designed to be a multi-purpose detector. The other such detector is CMS. Between them they are designed to probe Higgs physics, QCD, flavour physics, as well as a multitude of beyond Standard Model (BSM) physics scenarios including supersymmetry. These capabilities were first described in the second volume of the initial ATLAS design report [5].

ATLAS is designed to capture as much information as possible from any given collision event; it possesses near 4π solid angle coverage, and comprises an array of different detectors to ensure that as many particles as possible are measured with high accuracy. These various subsystems were first described in a technical design report [6], and can be seen in Figure 1.3. In the innermost part of the detector, closest to the interaction point, can be found silicon tracking sensors designed to reconstruct the paths of charged particles. Further out are electromagnetic and hadronic calorimeter cells, to give measurements on the energy of particles, and outermost are the muon chambers, since muons are capable of penetrating the calorimeters. Across the inner detector is a

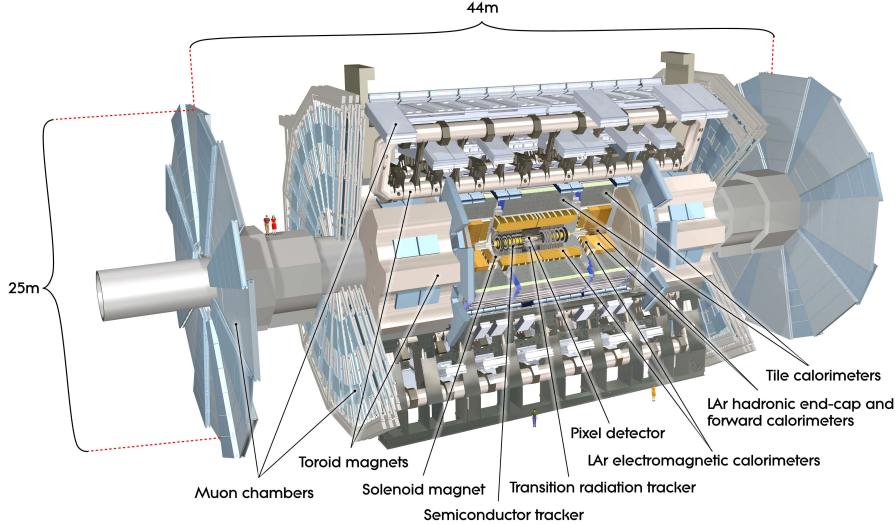


Figure 1.3: Overview of the subsystems that together form the ATLAS detector [4].

applied a magnetic field of 2 T, and a reduced magnetic field of ~ 0.5 T exists in them muon system. This allows for momentum measurements of charged particles through the curvature of their tracks in both the inner detector and muon chambers.

The remainder of this chapter will be dedicated to more detailed descriptions of these components of the ATLAS detector, along with the conventions of the ATLAS supersymmetry (SUSY) group used to define the reconstructed objects with which analyses can be performed.

1.2.1 Co-ordinate system

As is sometimes emphasised [7], the co-ordinate system of ATLAS is right-handed Cartesian, with its origin at the nominal interaction point. The axes are then oriented such that the x -axis is pointing towards the centre of the LHC ring, and y -axis is directed vertically upward. The z -axis thus defines one of the beam directions. The (x, y) plane is referred to as the transverse plane, in which points are frequently given in polar (r, ϕ) co-ordinates, where the azimuthal angle ϕ is, standardly, set to 0 on the x -axis.

Whilst the remaining polar angle $\theta = \arctan(p_z, \sqrt{p_x^2 + p_y^2})$ could be used, it is preferred to use the pseudorapidity

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]. \quad (1.2)$$

It is named thus since in the case of massless particles it is identical to the rapidity

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (1.3)$$

which is invariant under boosts in the z -direction. Since for many of the particles that will be observed in ATLAS it will be the case that $E \approx |\mathbf{p}|$, this property is still desirable even although it does not hold exactly.

Given the definition of pseudorapidity, it is common to describe separations between objects in the detector in terms of

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (1.4)$$

1.2.2 Inner detector

The inner detector, closest to the interaction point, is formed of three subsystems – the pixel detector, SCT (SemiConductor Tracker), and Transition Radiation Tracker (TRT)– as can be seen from Figure 1.4. Overall these give coverage of the solid angle defined by $|\eta| < 2.5$, and occupy the volume with $45.5 < r < 1082$ mm. Using these systems, its purpose is to detect the path taken by charged particles as they bend through the magnetic field, and hence determine their momenta.

Pixel detector

The pixel detector is closest to the beamline, with $45.5 < r < 242$ mm, and as such is the highest resolution detector, containing 140 million semiconductor pixels each of just 50×400 μm . This allows it to achieve measurements of track intersection positions up to a precision of 10×115 μm , which is desirable since it is very close to the interaction point, and as such the area subtended by a given solid angle is at its smallest value for any component in the detector. It is also designed to tolerate the very high radiation doses that must be endured at such proximity to the interaction point. The detector is

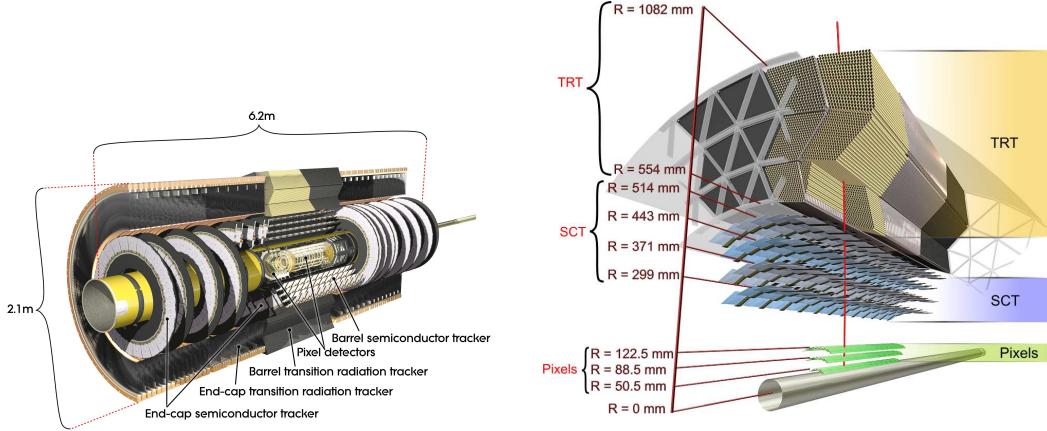


Figure 1.4: Overview of the systems forming the inner detector in ATLAS. The right-hand image shows the radial locations of each component. Both images from [4].

formed of three barrel layers as well as two end-cap structures. Each of the end-caps comprises four discs of sensors, arranged such that most tracks ought to hit pixels in at least three distinct layers. Further details can be found in the corresponding technical design report [8].

SCT

The semiconductor tracker sits outside the pixel layers, and is formed of etched strips in silicon wafers. Whilst not providing the resolution of individual pixels, location in (η, ϕ) is achieved by use of a ‘stereo effect’, whereby wafers are layered with their strips deviated from parallel by 40 mrad. The SCT is formed of four such stereo layers in the barrel, along with nine discs in each end-cap. The pitch of the strips in the barrel permits a resolution of 17 μm in the ϕ -direction, and the stereo effect allows approximately 580 μm in the z -direction. A typical track is expected to cross eight layers of strips, including stereo layers. Further details can be found in the technical design report of the inner detector [9, 10].

TRT

Finally, the transition radiation tracker is a straw chamber, and is the outermost component of the inner detector. Each “straw” is a 4 mm diameter polyimide tube, coated

internally with aluminium to form a cathode, and enclosing a tungsten wire. Each straw is sealed and filled with a gas mixture, formed of 70% Xe, 27% CO₂ and 3% O₂.¹ Charged particles traversing a tube ionise the gas; the ions then drift radially due to the potential difference, and the excess charge is collected and detected. The tubes are arranged parallel to the beam axis in the barrel region, and radially in the end-caps. In total one expects 30 straw tube hits from a typical track. By including measurements of the drift time a resolution of 130 μm is achieved for each straw hit.

1.2.3 Calorimeters

Calorimeters in ATLAS come in two types – liquid argon (LAr) technology in the electromagnetic barrel, end-cap (both electromagnetic and hadronic) and forward calorimeters, and then iron-scintillator ‘tiles’ for the hadronic barrel and extended barrel regions [12]. An overview of the calorimeter system can be seen in Figure 1.5. Overall they cover solid angles up to | η | < 4.9, with the electromagnetic calorimetry providing finer grained measurements to augment the inner detector for electron and photon measurements, whilst the hadronic calorimeter is coarser but sufficient for jet reconstruction and measurements of missing transverse momentum.

All the calorimeters in ATLAS are of the sampling variety, that is they use different materials for the absorber, that triggers a particle shower, to the material that measures the energy of that shower. Whilst this has the advantage that dense materials, e.g. lead, can be used to trigger the shower in a small space, some energy will be lost and unmeasured in the absorber. Thus a calibration must be used to estimate the true energy of any observed shower in the calorimeter. Each calorimeter is also segmented in η and ϕ so as to provide some directional information, although it is coarser than that from the inner detector. Finally, the calorimeter is designed to limit “punch-through” of high energy jets into the muon chambers.

LAr calorimeters

The LAr electromagnetic calorimeters are divided into the barrel section with | η | < 1.475, and two end-cap sections with 1.375 < | η | < 3.2. The barrel section is itself formed of

¹A mixture of 70% Ar and 30% CO₂ has also been tested, and shown to give somewhat inferior electron identification ability. However, due to the high cost of xenon studies checking the impact of using this argon mixture in Run 2 have been made for some analyses. The author was involved in one related to the analysis described later in this work [11].

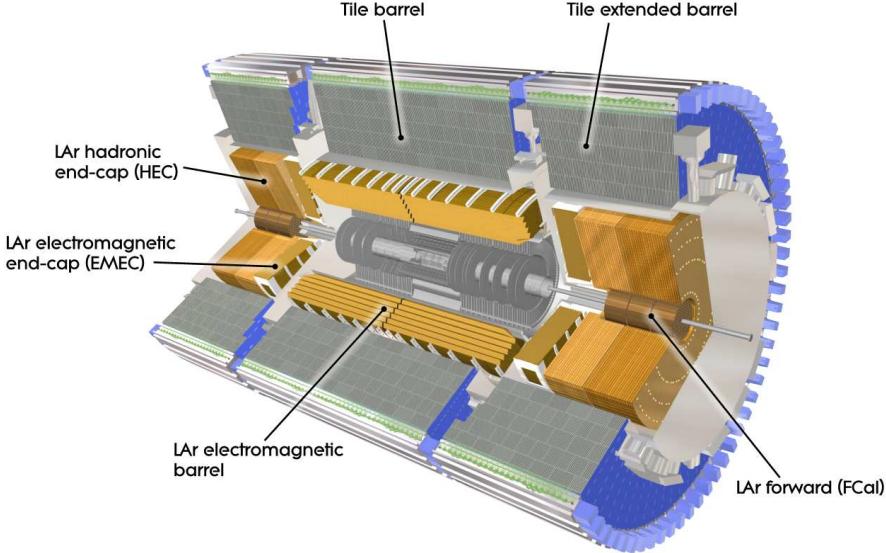


Figure 1.5: Overview of the different calorimeters in ATLAS [4].

two cylindrical halves, joined at $z = 0$, albeit with a gap of 4 mm. Whilst the barrel and end-cap sections do overlap, at the join there exists a region of slightly degraded performance for $1.37 < |\eta| < 1.52$. For the purpose of quality assurance most ATLAS analyses ignore electron and photon candidates falling into this ‘crack’ region.

Barrel Figure 1.6 shows a view of a section of the calorimeter in the middle of the barrel. Notable is that the lead absorbers and electrodes are shaped into an accordion-like pattern, thus providing coverage without cracks in ϕ , as well as meaning that a continuous piece of metal allows for easy signal extraction at either end of the electrode (i.e. at the inside or outside of the calorimeter). The gaps are then filled with liquid argon, and as such the whole system is cooled in a cryostat; separate cryostats are used for the barrel and end-cap sections.

Due to the presence of a significant amount of material in front of the calorimeter, a pre-sampler is placed as the first layer of the calorimeter. This is an instrument LAr cell, however it does not contain a dedicated absorber – rather the material in front of the calorimeter is used as the absorber. It is coarse in ϕ , but has a very fine η granularity, as can be seen in Figure 1.6. The second layer is segmented into square ‘towers’ of $\Delta\eta = \Delta\phi = 0.025$, and the third and final layer is similar, but with a coarser η resolution of $\Delta\eta = 0.05$.

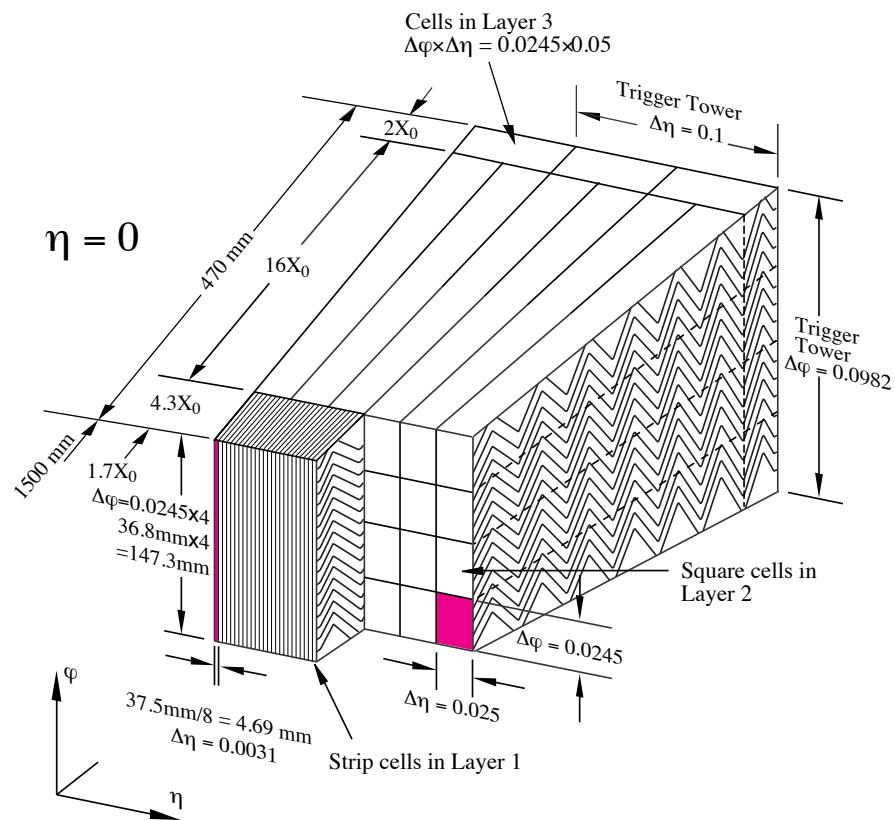


Figure 1.6: Cross-sectional view of a barrel module from the LAr electromagnetic calorimeter [4]. The ‘accordion’ structure of the electrodes can be seen throughout, as can the division of the calorimeter into cells, and clustering of cells into trigger towers.

EMEC The electromagnetic end-cap calorimeter (EMEC) sections have a similar accordion structure to the barrel. In these instances the detector is a disc shaped ‘wheel’, with the accordion folds lying in axial planes. Again, due to material in front of the calorimeter, pre-samplers are used in the same fashion as for the barrel for the region $1.5 < |\eta| < 1.8$; at higher pseudorapidities the combination of higher energy particles together with less dead material means there is not the need for a pre-sampler.

HEC The hadronic end-cap calorimeter (HEC) is placed behind the EMEC and is formed of two wheels, similar in design to the EMEC. The main difference is that copper plates are used instead of lead. These calorimeters contain hadronic showers effectively due to the large amount of material; in total there are about 12 interaction lengths present. The detector is segmented into 32 wedges, as well as two sections in the z -direction.

FCal The forward calorimeter (FCal) is split into three sections, the first of which is intended for electromagnetic measurements, and the latter two for hadronic. These occupy the forward region of $3.1 < |\eta| < 4.83$. The hadronic systems are designed to be very dense so as to minimise the lateral spread of showers, in particular with a view to preventing leakage into the HEC.

Tile calorimeters

Tile calorimeters are placed around the outside of both the barrel and end-cap regions, taking the form of a barrel and extended barrel, as can be seen in Figure 1.5. As with the LAr modules, it is a sampling calorimeter design, however in this case using steel tiles as an absorber and scintillating tiles as the detector. The tiles are connected via wavelength-shifting optical fibres to photomultiplier tubes (PMTs), which are mounted in the support girder at the rear of each tile module. A sketch of the layout can be seen in Figure 1.7. Leakage is minimised in these calorimeters, both to minimise the impact of punch-through on the muon system, as well as to ensure that jet energies are well measured.

1.2.4 Muon system

The muon system forms the outermost part of the ATLAS detector, covering a pseudorapidity range of $|\eta| < 2.7$. The core operating principle of the muon spectrometer [13] is

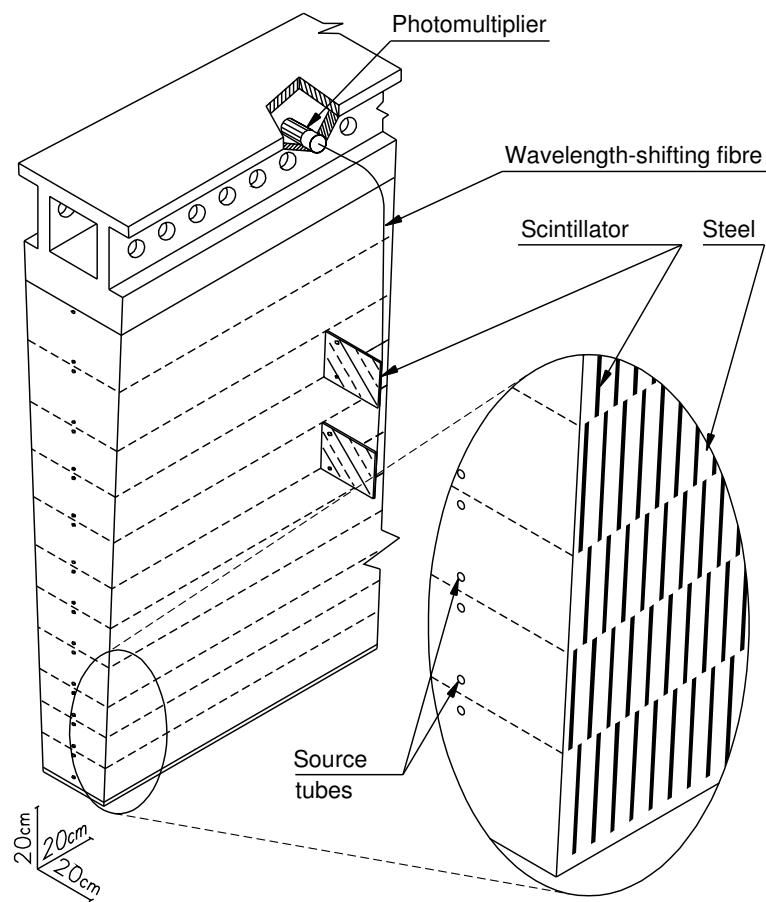


Figure 1.7: Cross-sectional view of a barrel module from the tile hadronic calorimeter [4]. Pictured is the alternating structure of the absorber and scintillator, as well as the optical readout mechanism. Wavelength-shifting fibres are required as the tiles emit ultraviolet photons, which are converted to visible light for detection by the PMTs.

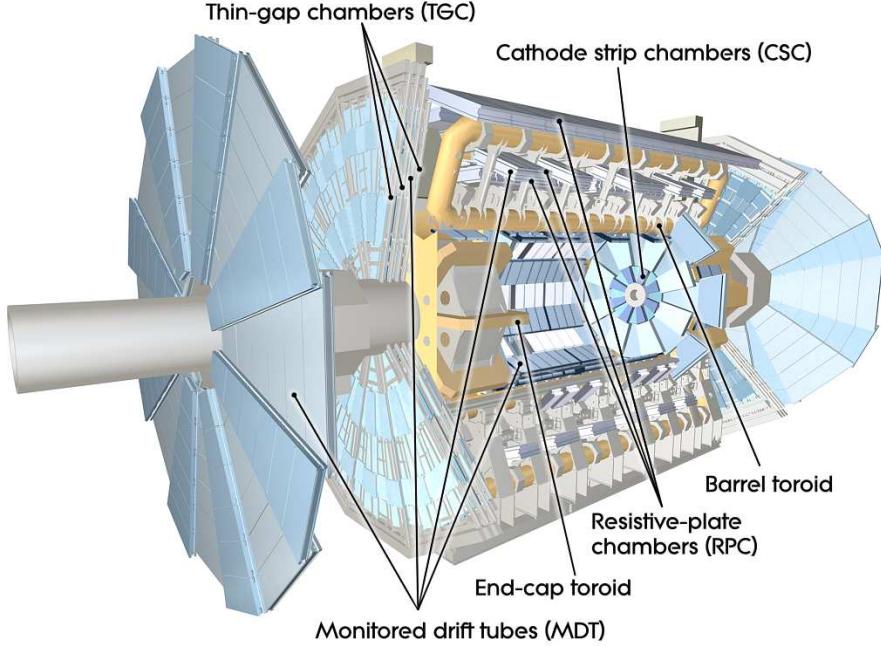


Figure 1.8: Overview of the different components of the muon system in ATLAS [4].

that of measuring the deflection of tracks due to magnetic fields. In the barrel section, $|\eta| < 1.4$, the magnetic field is induced by the main barrel coils, however in the end-cap regions ($1.6 < |\eta| < 2.7$) there are separate coils that induce a toroidal field; these can be seen in Figure 1.8. In the intermediate transition region bending will occur due to the fields from both sources.

As is also shown in Figure 1.8, several different detector technologies are employed. In the barrel region are found resistive-plate chambers (RPCs) for $|\eta| < 1.05$, as well as monitored drift tubes (MDTs) at $|\eta| < 2.0$. The former has the advantage of providing very rapid, ~ 10 ns, information for the purpose of triggering, whilst the MDTs give precise measurements (averaging 35 μm per chamber) in the (η, z) -plane, in which bending occurs. This allows the determining of momenta of muons down to ~ 33 GeV. In the forward section of the detector, cathode strip detectors (CSCs) are placed nearest to the interaction point for $2.0 < |\eta| < 2.7$, followed by thin-gap chambers (TGCs) as well as additional MDTs at $2.0 < |\eta| < 2.7$. The CSCs have a resolution of 40 μm in the (η, z) -plane, and 5 mm in the transverse plane. The detectors are placed in several ‘stations’ throughout the magnetic field, as is most clearly seen in Figure 1.9. In the barrel section this implies three cylindrical layers, whereas in the end-cap region vertical planes (wheels) are used. This sparse design allows sampling of the tracks at several points over

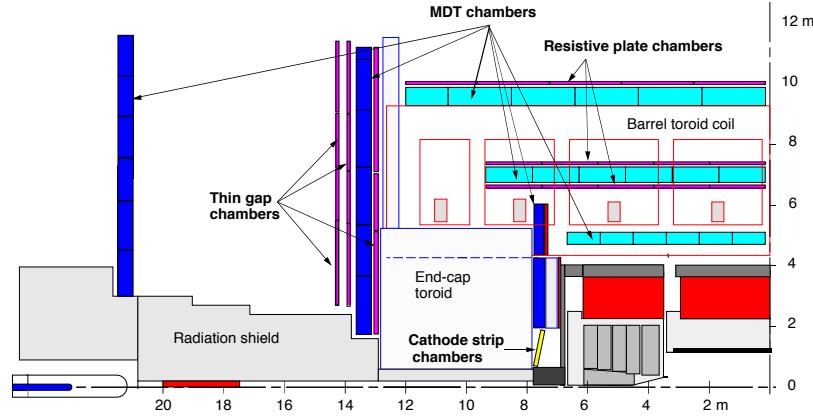


Figure 1.9: Cross-sectional view in the (y, z) -plane of the muon detectors [14].

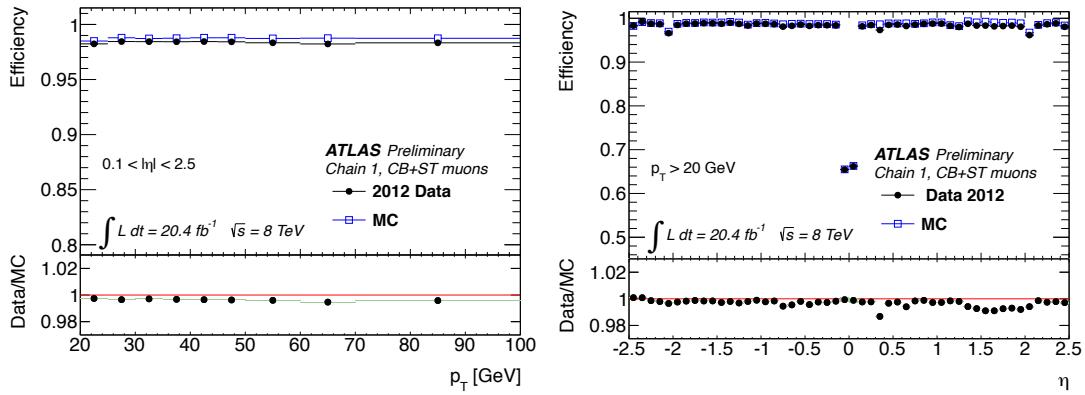


Figure 1.10: Muon reconstruction efficiency measured in 2012 data as functions of p_T and η respectively [15].

a large bending distance without instrumenting the entire region. High-momenta muons, i.e. those with near-straight tracks, are typically expected to be detected by at least three such stations.

Muon reconstruction employs primarily information from the muon system, as well as tracking information from the inner detector. A summary of the reconstruction efficiency for muons is shown in Figure 1.10.

1.2.5 Triggers

Under the design LHC running conditions ATLAS can expect to receive 40 million bunch crossings each second, and for each crossing an expected number of interactions ~ 20 , yielding an overall interaction rate ~ 1 GHz. The trigger is required to reduce this by a factor of 10^7 , since events can only be written to disk at ~ 100 Hz. This is achieved by a multi-stage triggering system; from most fundamental upwards the layers are Level 1 (L1), Level 2 (L2), and the Event Filter (EF). An overview of how they plug together is shown in Figure 1.11.

In essence, the L1 triggers are the fastest, hence simplest, filters, while L2 triggers make use of more information and can afford to be slower, since they are already working on a much reduced data rate. The L1 trigger is implemented entirely in hardware for maximum speed, based purely on coarse muon and calorimeter information, and makes a decision within 2.5 μ s. It is aided in this by the dedicated parts of the muon spectrometer dedicated to triggering, as have been mentioned in section 1.2.4. The L2 trigger is, conversely, software-based which allows for more flexibility in its design, focussing on regions of interest (RoIs) within the event that may contain objects such as leptons or jets. Finally the EF trigger is an additional software layer that is applied to events after they have been fully reconstructed.

The lower-level L1 and L2 triggers are typically combined and used as part of any given EF trigger, which are the objects typically used by physics analyses. The analysis described in the body of this thesis makes use of such triggers, which are specified in more detail in section 5.3.1.

1.3 Software environment

As has already been alluded to in the previous descriptions, large amounts of software are required both to firstly record data, and to then analyse it in a useful sense. A key requirement for the vast majority of analyses is to compare observations to the expected outcomes given a certain physics model (e.g. the standard model, or one of a variety of SUSY scenarios). In order to achieve this, a set of software modules are used:

- **Event generation:** When colliding protons, it is desired to know what new particles will be created, and what their kinematic distributions are. This is typically achieved by means of Monte Carlo samplers, which repeatedly draw samples that represent

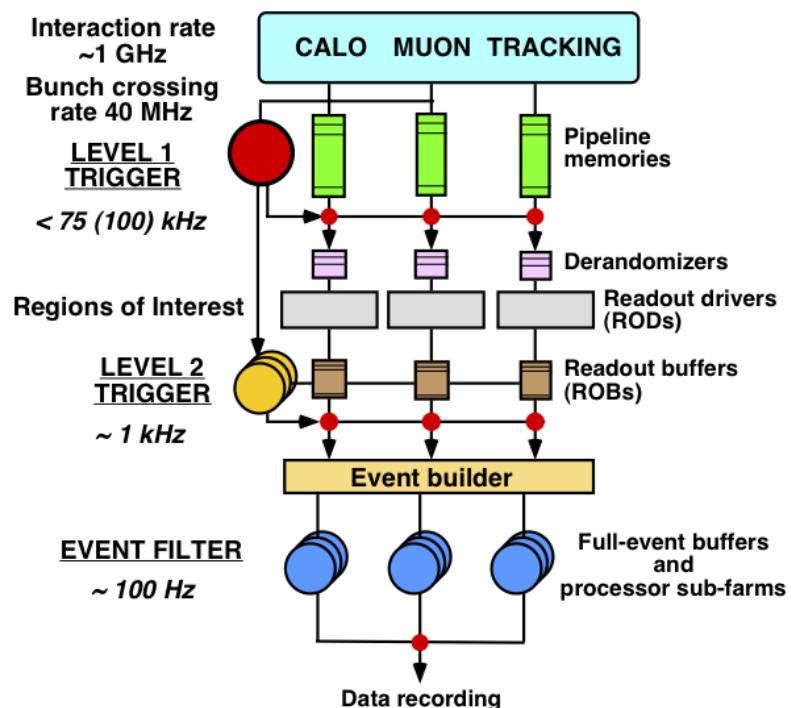


Figure 1.11: Top-level view of the triggering and DAQ system in ATLAS [16].

a possible outcome of a particular physics process. Many such programs exist, designed to simulate different Standard Model, or BSM, scenarios. For example, a new-physics scenario could be encoded in `MadGraph` [17], which gives matrix-element level events to another tool such as `Pythia` [18], which hadronises quarks and gluons into particle jets. The raw output of such generators is forked, with one stream undergoing minimal processing to yield “truth events” which can be used for preliminary analysis.

- **Detector simulation:** Given a record of particle momenta from an event generator, it is desired to know the expected response of various detector components. This is achieved with a detector simulator [19], which simulates the propagation of particles through the different materials comprising the detector, and estimating their energy deposits throughout. The canonical “full” simulation used for the 2012 run was based on `GEANT4` [20], however was typically slow. Hence for many applications faster, but more approximate simulators were used, e.g. for the calorimeter [21]. Significant work is being undertaken to integrate several different simulation methods into a framework to give an optimal trade-off between speed and accuracy [22].
- **Digitisation:** The detector simulation records ‘hits’ and energy deposits in the detector, at which point the digitisation procedure emulates the response of the electrical components given these inputs. The output of this stage is intended to be virtually identical to the data recorded by the real detector in an event.
- **Reconstruction:** This is the entry point for real data recorded in ATLAS. Its purpose is to turn the various activations recorded throughout the detector into objects ideally corresponding to fundamental particles. This includes electrons, photons, muons, and jets, as well as unmatched “soft terms”, which represent energy deposits that don’t easily fit into a hard particle. These are important for calculations of the missing transverse momentum, described later.
- **Analysis:** With reconstructed objects, the data is in a state ready for analysis. The only further processing are sets of quality requirements applied on the reconstructed objects to give so-called ‘analysis objects’. These are described in the next section.

1.4 Defining analysis objects

The objects produced by the reconstruction step in section 1.3 are not suitable for use in analysis. This is because they do not reflect all our knowledge about the calibrations of various sensors, nor the fact that certain parts of the detector may not always work to specification. Furthermore some classes of object are discarded since the detector is known to have limitations, e.g. the ‘crack’ region provides a blind-spot in the electromagnetic calorimeter system. The definitions in this section specifically reflect those used for the analysis in chapter 5, although for the most part they are in agreement with those agreed on by the ATLAS SUSY working group.

1.4.1 Leptons

This section defines the baseline requirements for electrons and muons; taus are not considered since they are not explicitly considered in this thesis. In both cases the baseline requirements are defined globally by the ATLAS SUSY working group. Additionally, a tighter set of requirements defines ‘tight leptons’, or ‘signal leptons’, that are used to define the signal regions in the like-charge lepton analysis considered later. These are not set by the collaboration, but rather were found to be optimal for this particular analysis.

The selection requirements for electrons are summarised in Table 1.1, for both baseline and tight selections. The **Medium++** quality denotes a set of other requirements, namely: it excludes objects falling in the crack region, $1.37 < |\eta| < 1.52$, requires a specific shower shape in the electromagnetic calorimeter, and also places a lower bound on the number of Pixel and SCT hits. **Tight++** has identical requirements to **Medium++**, but additionally requires a smaller $\Delta\phi$ between the inner detector track and the deposits in the EM calorimeter. Moreover the electron must have a hit in the innermost pixel layer (‘b-layer’), and additionally have left hits in the TRT. These quality requirements are specified in more detail in [23].

For both electrons and muons, requirements are placed on the impact parameters of the reconstructed tracks. These are defined to be the minimum (possibly extrapolated) approach to the beam axis, d_0 , as well as the displacement z_0 along the z -axis at the point at which this occurs. This is illustrated by the diagram in Figure 1.12. The selection requirements also use $u(d_0)$, which represents the uncertainty on the d_0 measurement.

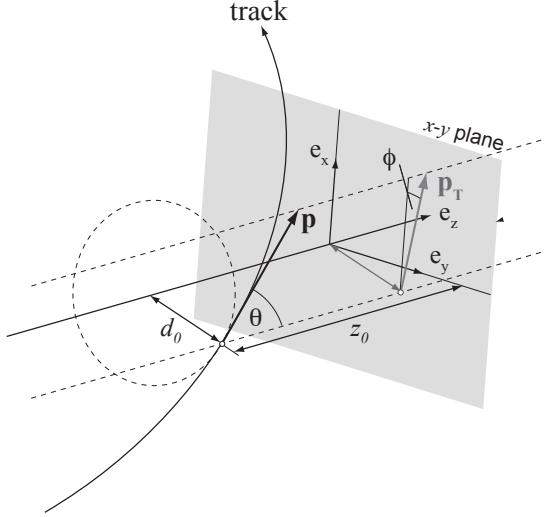


Figure 1.12: Visualisation of the impact parameters used in defining lepton selection requirements, taken from [24].

Their ratio is hence a measure of the significance by which the impact parameter deviates from 0.

Isolation requirements are also used; the “*cone20” and “*cone30” variables represent the summation of all transverse energies/momenta in a region around the nominal track defined by $\Delta R < 0.2$ and $\Delta R < 0.3$ respectively. A prefix of “et” specifies that energies are used, whilst “pt” implies momenta. The variables are separately defined for the inner detector and calorimeters; in the former case the p_T of tracks within the cone are used, whereas in the latter it is the energy deposits in suitably near cells that are summed. In both cases the energy/momenta of the nominal track or energy deposit is not included in the summation. These quantities are most helpful in reducing the rate at which jets can fake lepton objects, since jets tend to be wider than real lepton tracks and hence be less well isolated.

Muons undergo an analogous treatment to electrons, with a summary of the various cuts applied to the objects shown in Table 1.2. Both combined and segment-tagged muons are used, where the former implies that the inner detector and muon spectrometer tracks are consistent and used together in forming the track. By contrast, the segment-tagging algorithm extrapolates the inner detector tracks into the muon system and searches for matching hits in the muon stations. The **Loose** quality places more specific requirements on the recorded hits and energies, as detailed in [25].

Baseline electron	
Acceptance	$p_T > 10 \text{ GeV}, \eta^{\text{clus}} < 2.47$
Quality	Medium++
Impact parameter	$ d_0/u(d_0) < 5$
Tight electron	
Acceptance	$p_T > 15 \text{ GeV}$
Quality	Tight++
Track isolation	$\text{ptcone}20/\min(p_T, 60 \text{ GeV}) < 0.06$
Calorimeter isolation	$\text{topoEtcone}20/\min(p_T, 60 \text{ GeV}) < 0.06$
Impact parameter	$ z_0 \sin(\theta) < 0.4 \text{ mm}$ $ d_0/u(d_0) < 3$

Table 1.1: Object requirements for baseline electrons, as well as the additional requirements for an electron to be labelled ‘tight’. The former are common to the whole SUSY group within ATLAS, whilst the latter are optimised for the analysis presented in chapter 5. Due to the tighter isolation requirements tight electrons are much rarely induced by misclassified jets than the baseline.

1.4.2 Jets

As for leptons, the common definitions used by the ATLAS SUSY working group are shown in the first part of Table 1.3. The jets are reconstructed using the anti- k_T algorithm [26] as implemented in **FastJet** [27]. The distance parameter for this algorithm is set to $\Delta R = 0.4$. It operates on topological clusters from the calorimeter [28], with energies calibrated to take into account the losses inherent in a sampling calorimeter, as well as the presence of other dead material. This is performed using the local cluster weighting (LCW) scheme [29], as well as taking into account jet energy scale (JES) calibrations [30].

A further requirement used in some places by the SUSY working group at the time of the analysis being performed involved placing a cut on the jet vertex fraction (JVF). This is an estimate of the probability that a given jet originated from the primary vertex and aims to reduce the impact of jets from secondary interactions on the analysis. It was not applied as it was found that the 40 GeV requirement sufficed in making the impact from pile-up negligible.

Baseline muon	
Acceptance	$p_T > 10 \text{ GeV}, \eta^{\text{clus}} < 2.5$
Quality	Loose
Pixel hits (*)	# pixel hits ≥ 1
SCT hits (*)	# SCT hits ≥ 5
Pixel, SCT holes	# pixel holes + # SCT holes < 3
B-layer (**)	# b-layer hits ≥ 1
TRT requirements	If $0.1 < \eta < 1.9$: $n_{\text{TRT}} \geq 6$ and $n_{\text{TRT}}^{\text{outliers}} < 0.9n_{\text{TRT}}$ Else if $n_{\text{TRT}} \geq 6$: $n_{\text{TRT}}^{\text{outliers}} < 0.9n_{\text{TRT}}$

Tight muon	
Acceptance	$p_T > 15 \text{ GeV}$
Track isolation	$\text{ptcone30} / \min(p_T, 60 \text{ GeV}) < 0.12$
Calorimeter isolation	$\text{etcone30} / \min(p_T, 60 \text{ GeV}) < 0.12$
Impact parameter	$ z_0 \sin(\theta) < 0.4 \text{ mm}$ $ d_0/u(d_0) < 3$

Table 1.2: Object requirements for baseline muons, as well as the additional requirements for an muons to be labelled ‘tight’. The former are common to the whole SUSY group within ATLAS, whilst the latter are optimised for the analysis presented in chapter 5. n_{TRT} is the total number of hits recorded in the TRT, whereas $n_{\text{TRT}}^{\text{outliers}}$ is the subset of that which are classified as outliers. (*) If a track crosses either a pixel or SCT sensor known to be dead, this counts as a hit. (**) This b-layer requirement is only included if the track is expected to have crossed the b-layer.

Baseline jet	
Algorithm	AntiKt4Topo
Acceptance	$p_T > 20 \text{ GeV}, \eta < 2.8$
Signal jet	
Acceptance $p_T > 40 \text{ GeV}$	
<i>b</i>-jet	
Acceptance	$p_T > 20 \text{ GeV}, \eta < 2.5$
Tagging requirement	MV1 algorithm, 70%

Table 1.3: Object requirements for baseline jets, as well as the additional requirements for the signal jets used in the analysis presented later. The extra requirement for *b*-jets is also included.

1.4.3 Overlap removal

The reconstruction algorithms for electrons, muons, and jets in ATLAS are all independent, and as such it is quite possible for a single set of tracks and energy deposits to be classified as both an electron and a jet, for example. In order to remove this duplication when performing analysis, a procedure denoted ‘overlap removal’ is applied to the objects defined thus far. In the ordered set of cuts below, if e.g. the set of jets is updated at a given step, then the updated collection would be used if referred to in a subsequent step. The recipe is:

1. Remove any jet with $\Delta R < 0.2$ with any electron.
2. Remove any electron with $\Delta R < 0.4$ with any jet.
3. Remove any muon with $\Delta R < 0.4$ with any jet.
4. Remove any electron with $\Delta R < 0.1$ with any muon.

At this stage all of the electrons, muons, and jets are as used in chapter 5.

1.4.4 Missing transverse momentum

One quantity that is possible to measure at near- 4π detectors such as ATLAS is the missing transverse momentum. It is a quantity of particular relevance in analyses searching for new physics signatures expecting a new stable and non-electromagnetically interacting particle; such particles can only leave traces through momentum that *isn't* seen. In principle, one computes the sum of all visible four momenta, and projects it into the transverse plane. In the event that no particles are missed this should be very close to zero, since the beams carry approximately no momentum in this plane. Thus, one defines

$$\mathbf{p}_T^{\text{miss}} \equiv - \sum_{\text{visible}} \mathbf{p}_T \quad (1.5)$$

$$\approx \sum_{\text{invisible}} \mathbf{p}_T, \quad (1.6)$$

where equality holds in the ideal case where all visible particles are measured perfectly. In the subsequent analysis often only the magnitude of this is used, which is written as p_T^{miss} .

For the purposes of this thesis a calculation denoted `MET_Egamma10NoTau_RefFinal` is used, which uses information from all calorimeter cells with $|\eta| < 4.9$ as well as the reconstructed muon objects. The readings from individual cells are calibrated according to which reconstructed object they have been associated. The objects used differ slightly in their acceptances from the analysis objects:

- **Electrons:** All electrons with $p_T > 10$ GeV and satisfying the `Medium++` quality requirement.
- **Muons:** All muons with $p_T > 10$ GeV, and otherwise satisfying the baseline requirements in Table 1.2.
- **Jets:** All jets with $p_T > 10$ GeV, although the JES calibration is not applied for those with $p_T < 20$ GeV, and LCW is used alone.
- **Photons:** This uses reconstructed photons with $p_T > 10$ GeV.
- **CellOut:** This is the term representing all the energy deposits in the calorimeter not associated with any of the previously defined object classes (including objects not meeting their p_T requirements). These are calibrated with the LCW scheme.

Chapter 2

Introduction to Supersymmetry

A significant portion of this thesis is dedicated to the presentation of results from an analysis searching for evidence of SUSY, in chapter 5. In order to motivate this analysis it is first necessary to consider the principles on which SUSY is founded, and then to foray into a discussion of the incongruencies and deficiencies of the Standard Model (SM) which SUSY might be able to address.

This chapter will not concern itself with the introduction of the SM or quantum field theory – some knowledge of the fundamental principles in these areas is assumed. Many sources covering these concepts already exist, to which the interested reader is encouraged to refer. Rather, the purpose of this chapter is primarily to convince the reader that not only is the SM incomplete, but that SUSY is an interesting extension worthwhile searching for at the LHC.

The theoretical content of this chapter is largely based on the textbooks and extended articles in references [31–36].

2.1 Limitations of the Standard Model

That the Standard Model is incomplete is not in question, since it makes no attempt to include gravity in its modelling of the fundamental forces of nature. The reason for this is that the naïve addition of gravitational terms results in a theory that is not renormalisable. Hence some new physics, such as string theory [37], would need to take over in the regime of the Planck mass where the SM loses predictivity. However, this aside there is still a list of more pressing issues at the energy scales we are currently able

to probe, either with the predictive power of the Standard Model, or its elegance as a theory:

- **Dark matter:** Despite strong evidence suggesting its existence, the SM has no candidate particle for dark matter.
- **Baryon asymmetry:** We observe more matter than antimatter, which suggests a source of CP violation not present in the SM.
- **Anomalous magnetic moment of the muon:** Precision measurements suggest it has a value incompatible with the SM.
- **Neutrino masses:** Neutrino flavour oscillation has been observed, which requires them to have mass; however the SM does not include mass terms for neutrinos.
- **Hierarchy problem:** The apparent ‘miraculous cancellation’ giving rise to the observable Higgs mass at electroweak scales, which suggests fine-tuning in any UV completion at high scales.

The following sections aim to delve into these problems in a bit more detail.

2.1.1 Dark matter

One of the most significant experimental disagreements with the SM is that of the inferred existence of dark matter. Dark matter is defined as matter which is massive, cold, non-relativistic, and has at most very weak couplings to the electromagnetic force. It is generally assumed to be uncharged, however millicharged dark matter has not been ruled out. This is the reason given to the fact that it cannot be directly observed with telescopes. It is theorised that such particles might interact weakly,¹ but this need not be the case.

It is now largely agreed that dark matter is indeed present in our universe, an opinion based primarily upon the functional dependence of the rotational velocities of galaxies with their radius. These studies suggest a lower bound on the average relative density of dark matter, $\Omega_{\text{DM}} > 0.1$, related to the dimensionful density ρ_{DM} by

$$\Omega_{\text{DM}} = \frac{\rho_{\text{DM}}}{\rho_{\text{crit}}}, \quad (2.1)$$

¹These type of dark matter particles are hence called WIMPs (weakly interacting massive particles).

where ρ_{crit} is the density of dark matter that would result in a closed universe [38]. More recent measurements suggest that dark matter comprises $\approx 23\%$ of all energy in the universe [39, 40], with only 4% being formed of the baryonic matter with which we are most familiar.² Measurements of the anisotropy of the cosmic microwave background additionally provide indirect information about the expected dark matter particle mass, and their interaction cross section in the very early stages of the universe's existence [41].

There is currently a large push to run experiments capable of *direct detection* of dark matter. In general these are large quantities of a dense medium with which it is hoped that dark matter particles will interact, typically buried to a great depth to minimise the flux of cosmic rays. These containers would then be surrounded by some form of detector. For example, taking the LUX experiment [42, 43], a container of liquid xenon is surrounded by photomultiplier tubes in order to detect gamma rays emanating from a dark matter interaction.

The majority of such experiments have, as yet, failed to find any direct evidence for dark matter, with the exception of apparent large excesses from the DAMA/LIBRA collaboration, along with other small excesses from CoGeNT, CDMS-Si, and CRESST [44]. These excesses are treated suspiciously due to their apparent inter-incompatibility, in addition to the tension created with results from the XENON-100 experiment.

Despite the current lack of direct evidence for dark matter, the indirect evidence is considered sufficiently strong that its existence is all but confirmed. Unfortunately, the SM does not contain any particle which satisfies the required properties of a dark matter particle, and as such it is strongly suspected that the SM is incomplete.

2.1.2 Baryon asymmetry

It is known that we observe an asymmetry in the ratio of matter to antimatter, with significantly less antimatter than matter [45]. If one assumes that at the moment of the big bang³, matter and antimatter were created in equal quantities, it necessarily follows that the present-day asymmetry must be explained by the presence of CP -violating interactions. The SM does contain such CP -violating terms in the form of the CKM

²The rest of the energy is deemed *dark energy*, and is even less well understood than dark matter. Whilst a very interesting problem, it is not addressed by this work.

³This is not a universally accepted assumption, although is the prevailing opinion. The alternative requires a much larger initial baryon asymmetry due to conversion of baryons to leptons through sphaleron processes [46].

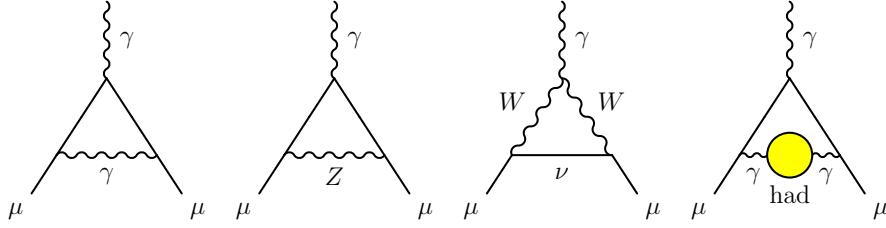


Figure 2.1: Schematic summary of the lowest order loop contributions to the magnetic moment of the muon in the SM. On the far left is the QED contribution, the middle two depict weak contributions, and the right demonstrates the hadronic contribution. Figure taken from [38].

matrix, which describes quark mixing, or the equivalent PMNS matrix for neutrino mixing [47], however the magnitude of the effect is too small to explain the observed asymmetry [45].

2.1.3 Anomalous magnetic moment of the muon

The gyromagnetic ratio of the muon g_μ can be computed at tree level in quantum electrodynamics as being exactly 2. This is then related to the overall magnetic moment \mathbf{M} of the muon by

$$\mathbf{M} = g_\mu \frac{e}{2m_\mu} \mathbf{S}, \quad (2.2)$$

where \mathbf{S} is the spin vector. Loop corrections, for example those shown in Figure 2.1, provide a small deviation from this tree level value. As such it is customary to define the *anomalous magnetic moment*,

$$a_\mu = \frac{g_\mu - 2}{2}. \quad (2.3)$$

Several predictions of the SM value of the anomalous magnetic moment exist, which shall be denoted a_μ^{SM} , which can then be compared to the average observed value a_μ^{exp} . Experiments which can measure this quantity include the E821 experiment at BNL, which studied the precession of muons in a magnetic field whilst contained in a storage ring [48]. The differences between the calculations and prediction is, on average

$$a_\mu^{\text{exp}} - a_\mu^{\text{SM}} = 288(63)(49) \times 10^{-11}, \quad (2.4)$$

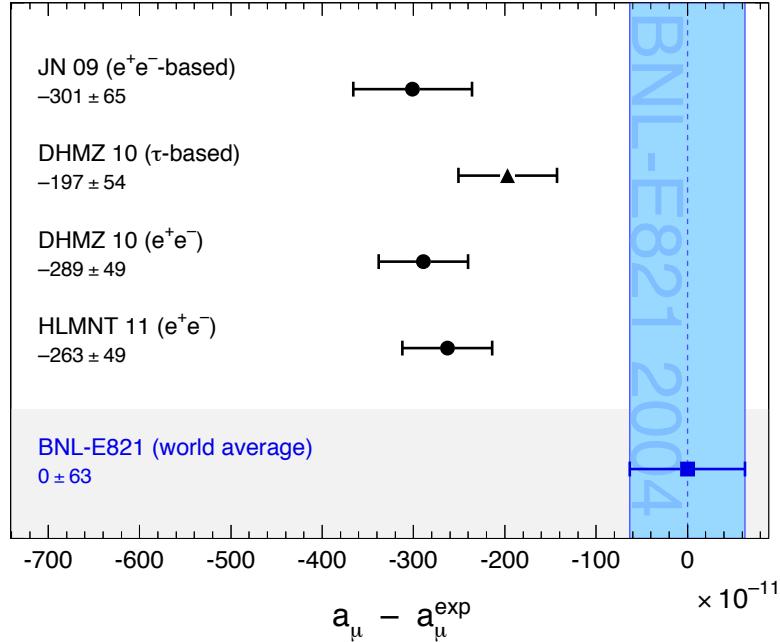


Figure 2.2: Comparison of the experimentally measured anomalous magnetic moment of the muon, a_μ^{exp} , to a selection of the latest SM theory predictions (which vary in the computation of the hadronic component, cf. Figure 2.1). The blue band represents the experimental uncertainty, and the bars on the theory points represent the purely theoretical uncertainty. Figure taken from [38].

where the number in the first bracket is the experimental uncertainty, and in the second the theoretical uncertainty. A graphical comparison of the experimentally observed value to several current theoretical predictions is shown in Figure 2.2. Overall the deviation is 3.6σ , however if τ data is used then the discrepancy is only 2.4σ . Whilst this does not satisfy the 5σ requirement for a discovery, it is an interesting tension, which shall later be shown to be explainable in several supersymmetric scenarios.

This section is largely based on the pertinent article in reference [38].

2.1.4 Neutrino masses

The SM includes only left-handed massless neutrinos, with three flavours. These three neutrinos are, by definition, flavour eigenstates. However, in general it is necessary to compute the *propagation* of any particle in a mass eigenstate, since the propagation speed of waves on a massive field are dependent on this mass. If one postulates for a moment

that neutrinos did indeed have mass, then one can imagine a unitary transformation between the flavour eigenstates $|\nu_\alpha\rangle$ and mass eigenstates $|\nu_i\rangle$,

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i}^* |\nu_i\rangle \quad (2.5)$$

$$|\nu_i\rangle = \sum_\alpha U_{\alpha i} |\nu_\alpha\rangle. \quad (2.6)$$

The transformation U is known as the Pontecorvo-Maki-Nakagawa-Sakata matrix, developed in 1962 [47]. Then, if producing neutrinos of definite flavour, and observing in the flavour basis at a later time some distance from the source, it would be expected that the flavour composition might change if there exist differences between the masses of the $|\nu_i\rangle$.

Observations consistent with this flavour oscillation are well known. For example, in the ‘solar neutrino problem’ it appeared that only about 1/3-1/2 of the expected electron neutrino flux from the Sun was detected. This is now attributed to the Mikheyev-Smirnov-Wolfenstein (MSW) effect, which means that the neutrinos emerging from the Sun are largely in the ν_2 mass eigenstate, which is formed more significantly of ν_μ than ν_e . The ν_1 and ν_2 components propagate through space and arrive, incoherently, at Earth. Subsequent measurements of flavour by experiments will then see the appropriate deficit of ν_e [49, 50]. Neutrino oscillation has also been observed with man-made neutron sources, such as nuclear reactors [51], and long-baseline beam experiments such as OPERA [52].

While not explicable directly by the SM, there are two relatively simple extensions that do allow for it. In one case, one assumes that neutrinos are Majorana particles, and adds a mass term to the Lagrangian [53]. In order to preserve gauge-invariance this is formed by spontaneous symmetry breaking of a dimension-5 Weinberg operator, however it is non-renormalisable. This operator could arise from many different theoretical frameworks, and a review of possible extensions can be found in reference [54]. Alternatively, if neutrinos are Dirac particles it is required to add a right-handed neutrino field to the Lagrangian, and adding a standard Dirac mass term [55].

2.1.5 Hierarchy problem

The ‘hierarchy problem’ pertains to the large difference between the scales of gravity and electroweak physics, namely that the ratio of the Planck mass to the mass of the W boson $M_{\text{Pl}}/M_W \sim 10^{17}$. This fact is equivalently a statement that the gravitational force is much weaker than any of the other known forces. This question can be reduced

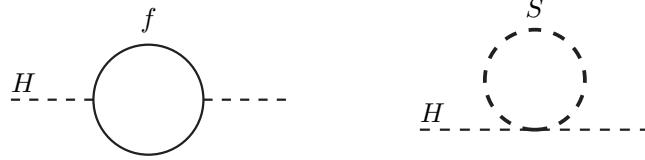


Figure 2.3: Higgs self-coupling terms at one loop contributing to the observed Higgs mass, with (a) terms from fermion loops, and (b) terms from scalar loops. Figure taken from [35].

to a more directly calculable quantity, namely the mass of the Higgs boson. It is found that the physically observed mass, as opposed to the bare mass parameter appearing in the Lagrangian, has significant contributions from loops as shown in Figure 2.3. The correction to the observed squared Higgs mass for a fermion loop of mass m_f is, up to constant factors

$$\Delta m_{H,f}^2 \propto - \int^\Lambda d^4k \left(\frac{1}{k^2 + m_f^2} + \frac{m_f^2}{(k^2 + m_f^2)^2} \right), \quad (2.7)$$

where we choose to apply some cutoff scale, Λ , above which we expect new physics to apply. This integral is hence finite but large, since the first term of the integral will scale like Λ^2 . If the SM were valid up to M_{Pl} , then in order for the Higgs to have a mass ≈ 126 GeV, the *bare* mass in the Lagrangian must be tuned to cancel out this huge correction. This bare mass can in turn be related to the parameters of the higher-energy physics model, and as such this implies a fine-tuning problem in this new theory. Whilst this is possible, it would be more elegant not to require such tuning.

2.2 Supersymmetry

SUSY aims to exploit an additional allowed spacetime symmetry that is not included in the SM. In doing so it is thought to not leave any other possible extensions, and as such is motivated largely by being in some way ‘complete’. The rest of this section will firstly discuss this topic in more detail, and then proceed to demonstrate a selection of supersymmetric theories that have been widely investigated for various reasons. Using these it shall be shown that some of them are capable of fixing, or at least mitigating, several of the known or observed deficiencies in the SM pointed out in section 2.1. Finally

some considerations are made about the current constraints on SUSY, both from collider and other experiments.

2.2.1 Theoretical principles and motivation

The S -matrix

The arguments later in this subsection rely on the meaning of the S -matrix in quantum field theory. The S -matrix is the core part of scattering theory that contains all the *observable* physical properties of a given theory.⁴

In abstract terms, if one has an initial state $|\Phi_i\rangle$ defined at $t = -\infty$, then one can consider the state at time t , $|\Psi(t)\rangle$, which has evolved from this. For a target final state at $t = +\infty$, denoted $\langle\Phi_f|$, the S -matrix represents the transition amplitude to this state from the initial state, namely

$$S_{fi} = \lim_{t \rightarrow +\infty} \langle\Phi_f|\Psi(t)\rangle \quad (2.8)$$

$$\equiv \langle\Phi_f|\mathcal{H}_{\text{int}}|\Phi_i\rangle, \quad (2.9)$$

where S_{fi} is the S -matrix element for the specified initial and final states, and \mathcal{H}_{int} is the interaction Hamiltonian for the system. Quantum field theory allows one to compute the S -matrix, in particular using the perturbative expansion most familiar in its graphical form as Feynman diagrams.

The Coleman-Mandula theorem

In 1967, Coleman and Mandula published a paper [56] containing a ‘no-go’ theorem which claimed to prohibit any additional symmetries of the S -matrix which were not trivial combinations of the known Poincaré space-time symmetry (which includes Lorentz boosts, rotations, and translations), and other internal symmetries. If additional symmetries were enforced, the only allowable S -matrix would be trivially the identity; that is, there would be no interactions. However, the Coleman-Mandula makes a key assumption – that only commuting symmetry generators are considered. Thus the theory only directly

⁴It is possible, and indeed it is the case in the SM, that a theory will contain additional, spontaneously broken, symmetries that are not visible in the S -matrix.

considers bosonic operators, those of integer spin, since those for fermions are defined by *anticommutation* relations.

The Haag-Lopuszanski-Sohnius theorem

Generalising the Coleman-Mandula theorem by additionally allowing these anticommutating symmetry generators, i.e. those operating on fermions, Haag, Lopuszański and Sohnius wrote in 1975 of an additional symmetry that the S -matrix might possess [57]. These are represented by the four supersymmetry generators Q_α and $\bar{Q}_{\dot{\alpha}}$, which belong to the $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ irreducible representations (irreps) of the Lorentz group. α and $\dot{\alpha}$ represent respectively a left-handed and right-handed Weyl spinor index. If acting on a scalar state ϕ , one then finds

$$Q_\alpha |\phi\rangle = |\psi_\alpha\rangle, \quad (2.10)$$

where the resultant object ψ must be a fermion (in this case left-handed) due to the fact that the left-hand side transforms like a left-handed Weyl spinor. Using the anti-commutation identity⁵ $\{Q_\alpha, Q_\beta\} = 0 \forall \alpha, \beta$, it is clear that acting on a scalar state with components of Q_α twice will not leave any free spinor indices. Namely, one finds

$$\varepsilon^{\beta\alpha} Q_\beta |\psi_\alpha\rangle = |\phi\rangle. \quad (2.11)$$

If ϕ is a scalar in a left-handed chiral supermultiplet, then it is defined such that $\bar{Q}_{\dot{\alpha}}\phi = 0$ – as such, since Q s and \bar{Q} s commute, the latter operators cannot be used to add an extra index since the state will be annihilated. However, with an alternative definition one could have a supermultiplet in which

$$\bar{Q}_{\dot{\alpha}} |\psi_\alpha\rangle = |A_{\dot{\alpha}\alpha}\rangle \neq 0, \quad (2.12)$$

where the object $A_{\dot{\alpha}\alpha}$ lives in the $(\frac{1}{2}, \frac{1}{2})$ representation of the Lorentz group, and is hence a vector boson; the pair $(\psi_\alpha, A_{\dot{\alpha}\alpha})$ is then called a vector supermultiplet.

Due to the representations in which the Q operators live, they transform non-trivially under the Lorentz group; as such they do not form an internal symmetry, but in fact supersymmetry is an additional space-time symmetry.

⁵This is true only for $N = 1$ SUSY. For extended SUSY models the result is an antisymmetric matrix Z^{IJ} in the supersymmetry generator indices, known as the ‘central charge’.

It is found that both the momentum operator and any gauge transformation generators corresponding to internal symmetries commute with both Q and \bar{Q} . The net result of this is that supersymmetric states contain both bosonic and fermionic fields due to the presence of Q ; due to this it is commonly referred to as a *supermultiplet*. These supermultiplets are either chiral, containing either a boson and left-handed fermion, or anti-chiral, where the fermion is instead right-handed.

Within a supermultiplet the particles have identical charges under all gauge symmetries, due to the commutation properties. In the context of extending the SM, this means that in a given representation there will be two ‘superpartners’, a boson and a fermion, each of which has the same electric charge, colour charge, and charge under the weak force.⁶ Since P^μ also commutes with Q it can be inferred that each of these superpartners will have an identical mass.

Extended supersymmetry

In the above discussion of the Haag-Lopuszanski-Sohnius theorem, only one supersymmetry operator, Q , was used; however, one is not prevented from inventing N of them, Q_α^I and $\bar{Q}_\dot{\alpha}^J$, with $I, J \in \{1, \dots, N\}$. By increasing N above 1 one places additional constraints on the S -matrix, since there are additional symmetries that must be satisfied by the interaction Hamiltonian. All of the models constrained by the analysis in this thesis are of the $N = 1$ form, however models with $N > 1$ are of interest elsewhere. In general, this latter class of models forms *extended supersymmetry*.

2.2.2 Supersymmetric phenomenology

This section will introduce the common benchmark SUSY model, the Minimal Supersymmetric Standard Model (MSSM), and consider the particle content that one finds after a brief diversion through mechanisms of SUSY breaking. A mention will also be made of R -parity violating extensions to the MSSM, a model based on which is studied later in this work.

⁶In general a supermultiplet need not be restricted to having just one bosonic and one fermionic degree of freedom, although this is the only phenomenological case that will be considered. In the general case it *can* be proved that there must be equal numbers of bosonic and fermionic degrees of freedom.

Minimal Supersymmetric Standard Model

The MSSM is called ‘minimal’ because it introduces the smallest particle content necessary such that the SM particles still exist in their current forms within an $N = 1$ supersymmetric framework. Each SM fermion is placed within a supermultiplet containing an additional spin 0 boson – these new particles have the same name as their fermionic counterpart but with a prepended ‘ s ’ (for ‘scalar’). So an electron (e) is partnered with a selectron (\tilde{e}), a generic quark (q) with a generic squark (\tilde{q}), and so forth. Spin 1 bosons in the SM (before electroweak symmetry breaking (EWSB), so the B^0, W^\pm, W^0 bosons) are placed into gauge supermultiplets with fermionic spin $\frac{1}{2}$ superpartners; these again carry equivalent names but with the ‘-ino’ postfix. One therefore has gluinos (\tilde{g}), winos (\tilde{W}) and binos (\tilde{B}).

The Higgs sector is slightly more complex, with two supermultiplets, in order to ensure anomaly cancellation, which would otherwise make the theory non-renormalisable [58]. This is also required for the reason that each of the Higgs chiral supermultiplets H_u and H_d can give mass to only the up-type and down-type quarks respectively, due to their different charges under $U(1)_Y$.

The overall particle content of the MSSM, before SUSY breaking, is summarised in Table 2.1.

After EWSB, the diminutive *-ino* ending is still used for these gauge fields; in the absence of any supersymmetry breaking terms one would end up with zino (\tilde{Z}^0) and photino ($\tilde{\gamma}$). Since we know that SUSY must be a broken symmetry (as discussed in the next section), these and the Higgs multiplets are mixed into a set of neutralinos, charginos, and Higgs bosons as a result of the SUSY breaking terms that are added. A summary of these mixings is shown in Table 2.2.

The most notable feature here is probably the postulated existence of multiple Higgs bosons – here h^0 is known as the ‘light Higgs’ and H^0 as the ‘heavy Higgs’, since it is specified that $m_{h^0} \leq m_{H^0}$ and these are both CP -even neutral scalars, like the SM Higgs boson. The observation of a boson at 126 GeV [59], consistent with the Higgs, is then one of these two. Moreover there is an additional neutral scalar, A^0 , however this differs in that it is CP -odd. There are then a pair of charged Higgs states, H^\pm , which are the charge conjugates of one another. It is also worth noting that the neutralino

Names	Label	Spin 0	Spin 1/2	Gauge representation
Quarks	Q	$(\tilde{u}_L, \tilde{d}_L)$	(u_L, d_L)	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
	U	\tilde{u}_R^*	u_R^\dagger	$(\bar{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
	D	\tilde{d}_R^*	d_R^\dagger	$(\bar{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
Leptons	L	$(\tilde{\nu}_L, \tilde{e}_L)$	(ν_L, e_L)	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
	E	\tilde{e}_R^*	e_R^\dagger	$(\mathbf{1}, \mathbf{1}, 1)$
Higgs	H_u	(H_u^+, H_u^0)	$(\tilde{H}_u^+, \tilde{H}_u^0)$	$(\mathbf{1}, \mathbf{2}, \frac{1}{2})$
	H_d	(H_d^0, H_d^-)	$(\tilde{H}_d^0, \tilde{H}_d^-)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$

Names	Spin 1/2	Spin 1	Gauge representation
Gluon	\tilde{g}	g	$(\mathbf{8}, \mathbf{1}, 0)$
W bosons	$\tilde{W}^\pm, \tilde{W}^0$	W^\pm, W^0	$(\mathbf{1}, \mathbf{3}, 0)$
B boson	\tilde{B}^0	B^0	$(\mathbf{1}, \mathbf{1}, 0)$

Table 2.1: Supermultiplet particle content in the MSSM. The upper table contains the SM matter content and Higgs, and the lower table the SM force carriers. The matter supermultiplets have conventional labels, as indicated, which shall be used later. In all cases here particles are specified before EWSB; hence the gauge representation is in the form $(SU(3)_C, SU(2)_L, U(1)_Y)$. Note that the $SU(2)_L$ both singlets and doublets are formed of left-handed Weyl spinors, such that the superpotential remains holomorphic; this is why conjugates of the right handed particles are inserted into the supermultiplets. The tables are adapted from [35].

Particle group	Gauge eigenstates	Mass eigenstates
Higgs bosons	$H_u^0, H_d^0, H_u^+, H_d^-$	h^0, H^0, A^0, H^\pm
Neutralinos	$\tilde{B}^0, \tilde{W}^0, \tilde{H}_u^0, \tilde{H}_d^0$	$\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0$
Charginos	$\tilde{W}^\pm, \tilde{H}_u^+, \tilde{H}_d^-$	$\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$
3rd gen. squarks	$\tilde{t}_L, \tilde{t}_R, \tilde{b}_L, \tilde{b}_R$	$\tilde{t}_1, \tilde{t}_2, \tilde{b}_1, \tilde{b}_2$
3rd gen. sleptons	$\tilde{\tau}_L, \tilde{\tau}_R, \tilde{\nu}_\tau$	$\tilde{\tau}_1, \tilde{\tau}_2, \tilde{\nu}_\tau$

Table 2.2: The mass eigenstates observed in the MSSM after SUSY breaking, with an indication of which gauge eigenstates previously described in Table 2.1 are mixed into each group of particles. The precise form of mixing is determined by the SUSY model in question, and its parameters. It is assumed here that negligible mixing occurs between the first and second generation of squarks and sleptons – this is the case for all specific models considered later in this thesis.

and chargino states are ordered in terms of mass, so $m_{\tilde{\chi}_1^0} \leq m_{\tilde{\chi}_2^0} \leq m_{\tilde{\chi}_3^0} \leq m_{\tilde{\chi}_4^0}$, and $m_{\tilde{\chi}_1^\pm} \leq m_{\tilde{\chi}_2^\pm}$.

Whilst the above has detailed the particle content of the MSSM, the precise masses of the particles will be determined by a large number of free parameters. If one considers the SUSY-breaking parameters mentioned in the following discussion, there are a total of 120 new parameters over the SM. Whilst much of this space is already excluded, for the reasons noted in section 2.4, there are many degrees of freedom still remaining.

Much current interest is directed towards a subset of the MSSM known as the phenomenological MSSM (pMSSM), which imposes a number of constraints yielding a model with 19 free parameters [60]. Whilst this forms a large space, it is not infeasible to perform an exhaustive scan with current computer resources, and identify regions where SUSY is still not excluded that might not otherwise have been thought of.

SUSY breaking

Whilst the symmetry structure described above treats supersymmetry on a par with Lorentz symmetry, if SUSY is a valid model then it is necessarily a broken symmetry. This is due to the fact that $[Q_\alpha, P^\mu] = [\bar{Q}_{\dot{\alpha}}, P^\mu] = 0$, which forces particle pairs in a given supermultiplet to have the same mass; however since we have not yet observed any

of the superpartners, there must be a mechanism to break supersymmetry that causes them to be more massive than the corresponding SM particles.

A common scheme is known as *soft supersymmetry breaking*, and is employed by the MSSM. It is particularly tidy since it does not alter the normalisation of the superpotential above the SUSY breaking scale, and additionally allows one to consider separately the issues of what effect actually introduces the breaking of SUSY, and the particle spectra that result from it. This latter feature is possible since soft SUSY breaking simply introduces all possible additional mass terms into the Lagrangian for the gauginos and scalar fields – they break SUSY simply because the resultant addition to the Lagrangian will not be invariant under Q or \bar{Q} .

The causes behind these SUSY-breaking terms will not be discussed here, but two specific examples of gravitationally-mediated symmetry breaking in the mSUGRA (minimal SUper GRAvity) model, and gauge-mediated supersymmetry breaking (GMSB), will be discussed when setting limits on these respective scenarios in section 5.9.2.

R-parity

In general, it is possible to write down supersymmetric interaction terms that violate baryon or lepton number. This is sometimes considered undesirable, since these are perturbatively good symmetries in the standard model, and in all cases $B - L$ is conserved.⁷ The requirement that a supersymmetric theory conserve baryon and lepton number can be recast into the conservation of a quantity called *R-parity*. For baryon number B , lepton number L and particle spin s , we define *R*-parity to be

$$P_R = (-1)^{3(B-L)+2s}. \quad (2.13)$$

As an extension to the MSSM (which is an *R*-parity conserving (RPC) theory), the superpotential for the *R*-parity violating (RPV) interactions is

$$W_{P_R} = \underbrace{\frac{1}{2} \lambda^{ijk} L_i L_j E_k + \lambda'^{ijk} L_i Q_j D_k - \kappa^i L_i H_d}_{\Delta L=1} + \underbrace{\frac{1}{2} \lambda''^{ijk} U_i D_j D_k}_{\Delta B=1}, \quad (2.14)$$

where i , j , and k are flavour indices, and the supermultiplet labels are as denoted in Table 2.1. The formality of the superpotential is not necessary for the discussion at hand;

⁷The one exception to baryon number conservation being non-perturbative sphaleron processes, due to the chiral anomaly, which can change the baryon number by 3.

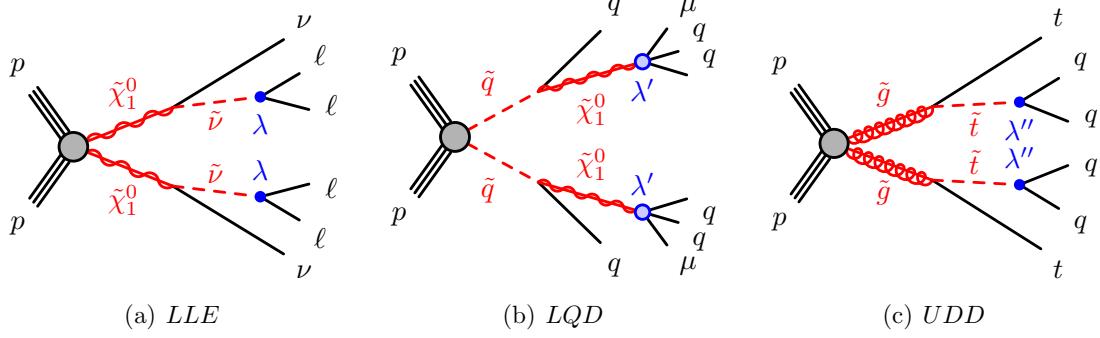


Figure 2.4: Feynman diagrams showing processes with several of the lepton and baryon number RPV couplings. The example for the *LQD* coupling implicitly includes an off-shell squark. Diagrams taken from [61].

merely that it denotes additional allowed couplings between particles such as those shown in Figure 2.4. These are also the only interactions that violate baryon or lepton number conservation, as indicated. If it is required that the product of P_R at any vertex is +1, then it follows that these vertices are disallowed – this would then be an RPC theory. Otherwise, the type of RPV theory would be specified by the values of the coupling constants λ , λ' , λ'' , and κ . The analysis in chapter 5 considers one model with non-zero components of λ'' .

2.2.3 Some specific terminology

This section introduces a few specific terms used later in this work, which do not neatly fit elsewhere.

Simplified models

Simplified models are called such since they typically decouple many particles in the SUSY spectrum – moreover for those that are left typically the branching fraction for one specific decay mode is set to be 100%. This is in the spirit of on-shell effective field theories [62], and is an approach that has been widely used by ATLAS SUSY analyses. Practically, the decoupling is achieved by arbitrarily tweaking the SUSY-broken Lagrangian to include precisely the desired mass terms and couplings; naturally in doing so it is very unlikely one is producing a renormalisable theory, however the results are

useful for a more schematic understanding of what specific types of process are excluded by a given analysis. Work has also been done that shows how exclusions on several different simplified models might be combined to give limits on more complex scenarios [63].

Compressed scenarios

This refers to the situation in RPC SUSY where there is only a small mass gap between two SUSY particles, one of which will decay to the other. Often, particularly in the case of simplified models, these particles will be the next-to-lightest supersymmetric particle (NLSP) and the lightest supersymmetric particle (LSP). This can be a challenging signature to search for at colliders, since if a light SM particle is also produced during this decay (for example a lepton or light quark), then it will have a low energy, and hence also typically have a low p_T .

2.3 Fixing the Standard Model with SUSY

Whilst SUSY is arguably an incredibly elegant theory from a purely mathematical perspective, it is also can serve a pragmatic purpose in fixing some of the limitations in the SM mentioned in section 2.1. Whilst SUSY does not offer solutions to all of these, the particular cases of dark matter, the anomalous magnetic moment of the muon, and the hierarchy problem will be considered.

2.3.1 Dark matter

In an RPC SUSY scenario, in which the $\tilde{\chi}_1^0$ is the LSP, this neutralino is an ideal candidate for dark matter since it is stable, interacts gravitationally, and is electromagnetically neutral. This could provide an explanation for the effects mentioned in section 2.1.

2.3.2 Anomalous magnetic moment of the muon

As previously noted, there is tension between the experimentally observed and predicted values of the anomalous magnetic moment of the muon, a_μ . This tension could be relieved by the introduction of additional diagrams into Figure 2.1. There are several

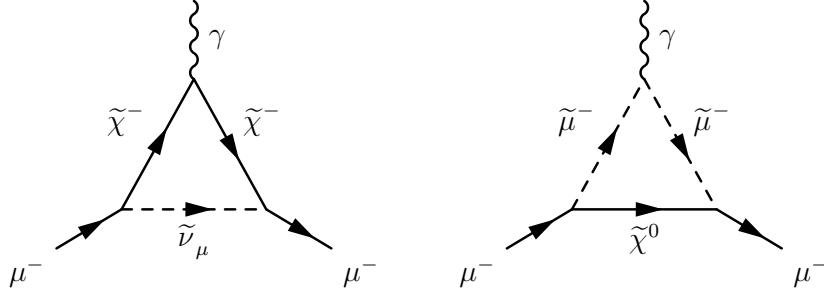


Figure 2.5: Examples of the contributions from particles in the MSSM at one loop to the anomalous magnetic moment of the muon. If the equivalent diagrams for μ^+ are considered, along with a sum over all possible chargino and neutralino states, then these represent all contributions at one loop. The diagrams are inspired by Figure 3 of [64].

SUSY particles which could form these loops, a couple of examples of which are shown in Figure 2.5. It is thought that the presence of SUSY particles in the range $100 - 500$ GeV could add a contribution sufficient to explain the observed deviation of $a_\mu^{\text{exp}} - a_\mu^{\text{SM}}$ from 0 [38].

2.3.3 The hierarchy problem

As noted previously in Figure 2.3, there are diagrams from both fermionic and scalar loops that contribute to corrections to the Higgs mass. It has already been shown that a fermionic loop will have a large negative contribution to m_H^2 , in equation (2.7). However, the contribution from a scalar particle loop of the same mass will, conveniently, be of the same magnitude as this correction but with the opposite sign. Thus it can be seen that unbroken SUSY completely removes the fine-tuning problem, since every loop contributing to the Higgs mass from an SM fermion will be exactly cancelled by the scalar loop from its superpartner.

If it exists, we know that SUSY is a broken symmetry, and thus that this cancellation will not be exact. This leads to the concept of *naturalness* that places approximate limits on the masses of certain SUSY particles, that will be described briefly in the next section. Naturalness simply embodies the desire not to replace the existing fine tuning problem in the SM with another in a supersymmetric model.

2.3.4 Gauge-coupling unification

Whilst not a problem with the SM *per-se*, it is noted that the gauge coupling strengths in the SM do not intersect when extrapolated to increasingly high energies, whereas in the MSSM with superpartners at the TeV scale they do. This is theoretically desirable, since a point of intersection could represent the spontaneous breaking of some larger symmetry group, for instance $SU(5)$ [65] or $SO(10)$ [66]. The couplings we observe at the current energy frontier could then be calculated from a smaller number of parameters of this higher-energy Grand Unified Theory.

2.3.5 CP -violation

Notably, in the MSSM the SUSY-breaking processes are assumed to be diagonal in flavour space; this is done to keep from breaking constraints imposed by measurements of flavour-changing neutral currents (FCNCs) and the electron electric dipole moment (EDM). The result of this is that the MSSM predicts minimal CP -violation, and not enough to explain the observed baryon asymmetry. There are, however, other scenarios which allow for larger CP -violation [67], which could be used to account for this phenomenon.

2.4 Constraints on supersymmetry

Now that the reader is hopefully convinced that SUSY is both a mathematically elegant and phenomenologically useful theory, it is time to temper the enthusiasm. As of the time of writing, no direct evidence has been observed to support the existence of SUSY. This section will give a brief overview of the sources of constraints on supersymmetric models.

2.4.1 Proton lifetime constraints

Some constraints on what RPV couplings are allowed exist due to the observed stability of the proton. In particular, if both $\Delta B = 1$ and $\Delta L = 1$ terms were present, the proton would decay much too quickly, so generally it is assumed that an RPV contribution either violates baryon or lepton number, but not both. Present experimental results

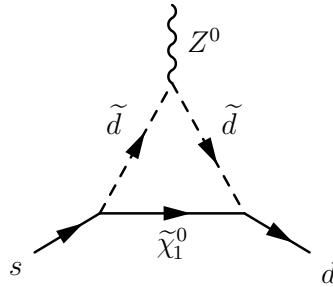


Figure 2.6: Example of a SUSY diagram which shows a flavour-changing neutral current. There are assumed to be non-zero off-diagonal terms in the squark mixing matrix which allows the bottom left vertex.

place constraints on the lepton number violating terms of equation (2.14), however the last, baryon number violating, term is less constrained [68].

2.4.2 Flavour-changing neutral currents

These are, as the name suggests, an interaction which allows a fermion to change to a fermion of a different flavour whilst retaining its electric charge. Whilst observed to occur in the SM, they are strongly suppressed by the GIM mechanism [69], and measurements of the decay of B^0 mesons at BABAR place strong constraints on the magnitude of the effect. SUSY potentially allows for large FCNCs by virtue of couplings like that demonstrated in Figure 2.6. In order to suppress these, models like the MSSM suppress flavour mixing terms, and as such these diagrams are prevented from occurring.

2.4.3 Naturalness

The concept of naturalness is summarised quite concisely by Susskind:

Naturalness ... requires properties of a theory to be stable against minute variations of the fundamental parameters. (Susskind, 1979 [70])

This can be seen simply as a restatement of the hierarchy problem in the SM, and indeed the concepts are related; one wants to avoid the situation where the fundamental parameters of a SUSY model must be fine-tuned in order to reproduce the physics we

observe. Such considerations typically lead one to favour lighter third generation squarks, $\mathcal{O}(1 \text{ TeV})$. Of course, this is not a hard constraint, but rather an aesthetic hope.

2.4.4 Additional constraints

In addition to the items described in slightly more detail above, there are other sources of constraints worth mentioning briefly, namely

- **Dark matter relic density:** Whilst the $\tilde{\chi}_1^0$ is often a good dark matter candidate particle, the parameters of the model should be tuned to ensure a fit to the observed dark matter relic density.
- **EDM measurements:** Measurements of the neutron and electron EDM suggest very strongly that they are near-zero [71]. Without constraining some parameters SUSY could introduce large contributions here.
- **$B_s^0 \rightarrow \mu\mu$ measurements:** Recent research at the LHCb and CMS experiments [72, 73] has observed this decay mode with combined significance greater than 5σ significance. Its branching fraction imposes a strong constraint in some areas of SUSY parameter space [74].

Chapter 3

The problem of fakes

3.1 Motivation and overview

The estimation of ‘fake’ backgrounds to certain types of objects is a common problem through experimental particle physics. A large number of analyses at ATLAS make measurements involving leptons, ranging from those making precision measurements of the Standard Model to those searching for evidence of BSM physics. All of these must hence consider, to some level, the impact of these backgrounds.

Reconstructed objects can be seen as representing some kind of **real** desirable underlying object; for example, an electron. It could also, however, be reconstructed as something it is not – by means of a concrete example, a jet leaving a suitably narrow deposit in the calorimeter could be identified as an electron. This would be denoted a **fake** electron. Any analysis that were to place a requirement on electrons would hence need to model the background where events are accepted due to the presence of one or more fake electrons. A slightly more detailed consideration of the mechanisms by which electrons and muons can be faked is presented in section 3.2.

Several methods have existed for performing such estimates prior to the author’s work in the field. One option is to rely upon the Geant4 detector simulation of ATLAS, and use Monte Carlo (MC) events generated for processes expected to contribute via fake objects. Whilst often a straightforward option for the analyser to adopt, there are two main drawbacks. Firstly, the simulation code might not faithfully reproduce the behaviour of the objects under investigation. Secondly, if looking in a particularly narrow region of phase space, generating a sufficiently large number of events to produce an estimate with a low enough statistical uncertainty might prove problematic. For

these reasons, *data-driven* methods are preferred, which aim to perform complementary measurements in data (sometimes aided by measurements on MC) in order to form a prediction of the fake background.

3.2 Processes for faking electrons and muons

This section summarises some of the more common processes by which electrons and muons can be faked in ATLAS.

3.2.1 Electrons

Electron reconstruction depends both on the observation of a hits in the inner detector, aligning to form a track with appropriate inferred mass, followed by an aligned deposit in the electromagnetic calorimeter. The inner detector can measure momenta of a track by its bending in the magnetic field, and estimate the speed of the particle given the energy loss in the TRT. Photon deposits in the calorimeter are often difficult to distinguish from electrons, and so any process that can give a photon (or more than one if tightly collimated) aligned with a charged track could be mistaken for an electron.

A hadronic jet, primarily formed of charged and neutral pions, can achieve this. The particle multiplicity can fluctuate, so it is possible to find jets with a single charged pion, which will leave an electron-like track in the inner detector – particularly for high-momentum particles the inferred mass measurement is not necessarily precise, allowing for electron and pion tracks to be confused. The decay of a π^0 mesons to photons in the same jet could then leave the required signature in the calorimeter.

Alternatively, electrons can be faked by photons emitted due to Bremsstrahlung from high energy muons. The track left by the muon in the inner detector could be confused with that from an electron, and the photon's subsequent deposit in the calorimeter could result in an electron being reconstructed. In practice this can usually be mitigated by requiring separation between reconstructed muons and electrons at the analysis level, however cases where the muon is not subsequently reconstructed cannot be removed.

3.2.2 Muons

Muons are primarily reconstructed using information from the muon system, although it is also required that the tracks in the muon spectrometer match onto tracks in the inner detector. Faking muons hence requires particles penetrating through the calorimeter layers – this can be achieved by charged hadrons with a lifetime sufficient to traverse the whole detector. Whilst the track of this hadron in the inner detector might not match that expected from a muon, events have enough activity that sometimes a chance match can be found with unrelated hits. In this way a muon can be reconstructed.

3.3 Problem statement

As is so often the case, the devil is in the details; a precise definition of what is required of such a data-driven method is needed.

Imagine there is a signal region, or if one prefers a single bin of a distribution in this region. The analyser suspects that faked leptons¹ are one reason why events are being observed here in the recorded data. At its core, what we would like is *an estimate of how many events we expect to see as a result of one or more leptons being faked*. To be even more explicit, in an ideal world what is desired is the *probability distribution* of this expected number.²

It is perhaps also useful to dispel one particular misconception – for the purposes of the standard ATLAS analysis techniques, and hence this thesis, we are *not* interested in trying to work out how many of the events actually seen in a given experiment are in fact due to faked leptons. Rather, and this is implied by the use of ‘expected’ in the definition, one is interested in the behaviour one would see when running the same experiment many times (if one is a frequentist), or one’s belief on what one would see if one were to perform it (for a Bayesian).

¹Or, in principle, other objects – for the remainder of this chapter ‘lepton’ is used throughout for simplicity.

²Later in chapter 4 it is shown that technically the requirement is slightly different if one aims to integrate the estimation procedure directly with a limit setting procedure. However, from a high-level perspective the difference is rather minor.

3.4 Modelling fake events

Any solution to the problem posed in the previous section will almost certainly contain several features:

1. **Classification:** Whilst we are not ultimately interested in classifying leptons or events as being real or fake, this is a necessary first step. Given quantities observed by the detector, some metric must be defined by which real and fake leptons can be discriminated.
2. **Control regions:** In ATLAS, making measurements commonly boils down to selecting events according to certain criteria, then counting them. In order to make the desired prediction in the signal region, we should therefore expect to be making auxiliary measurements in one or more control regions. In order to be relevant, these control regions must be related to the classification task mentioned previously.
3. **Event rate modelling:** Given a set of measurements in control regions, we desire a prescription to translate this into the fake background estimate.

3.4.1 ‘Tight and Loose’ model

Whilst there are likely many valid approaches, only one model is chosen in this thesis. It is based on a binary classification metric; any observed lepton is deemed to be either ‘tight’ or ‘loose’. This metric could be based on any number of observed quantities, all that is important is that real and fake leptons behave differently – by convention the cut is arranged such that real leptons are more likely to be observed as tight, and vice versa.

A more detailed mathematical formulation of this model is introduced in section 4.2, however for now a more graphical overview is presented.

3.4.2 Simple events

Consider the simplest case, where we are aiming to select events containing exactly one lepton. One can imagine two sets of physics processes contributing in our signal and control regions – one producing real events (those with a real lepton), and the other fake. In a given amount of data, each could be parametrised by the mean of a Poisson distribution, which are denoted ν_R and ν_F respectively. Some number of events will then

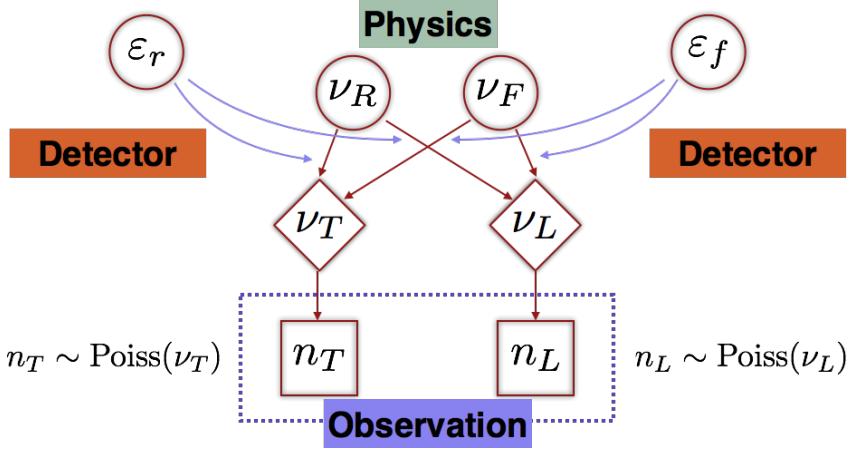


Figure 3.1: Diagram of the relationship between the fundamental parameters ν_R , ν_F , ε_r , ε_f (in circles); the derived parameters ν_T and ν_L (in diamonds); and the observations n_T and n_L (in squares).

be produced, each of which pass through the detector, and are each classified either as tight (T) or loose (L). The probabilities for this happening are, by construction, different for real and fake leptons, and are denoted ε_r and ε_f respectively. The diagram included in Figure 3.1 displays the hierarchy of these parameters graphically.

Commonly the quantities ε_r , ε_f are called ‘efficiencies’, since they measure the fraction of a given population (real or fake) of leptons that satisfies the tight cut. In practice, these are determined experimentally by using control regions enriched, respectively, in real and fake leptons.

The usual choice made at this point, which we shall also make, is to equate n_T with our signal region; that is, signal regions shall be defined exclusively with tight leptons. The other, loose, events will form a control region. In this case, therefore, the problem reduces to measuring n_T and n_L , estimating ε_r and ε_f , and then combining that to yield sufficient information to estimate ν_{FT} , the expected number of events that are both fake and land in the tight, i.e. signal, region.

3.4.3 Complicated events

Multiple categories

Regrettably, things are rarely as simple as in section 3.4.2. The first complication that arises relates to the efficiencies – namely that they are different for different leptons. These can either be fundamentally discrete differences, e.g. electrons vs muons, or continuous, e.g. varying behaviour as a function of p_T .

Continuous variations can be considered as sets of discrete possibilities if we create bins. One can then imagine an extension of the single lepton thought experiment with a set of lepton ‘categories’, divided according to flavour, kinematics, or any other observable quantity. One, quite significant, assumption that can be made is to take each of these categories as being independent. If this is true, then the model now looks more like that in Figure 3.2.

It is perhaps worth giving some more details on what constitutes a lepton category. An easy one is that of flavour – a given lepton can either be an electron or a muon. The mechanism for the formation of fake muons is different to that of fake electrons, and so we should expect the efficiencies to be different. However, we might also observe that the fake efficiency varies as a function of object p_T , and therefore create bins for different p_T ranges, say with index i . Now one can regard all quantities as being functions of this bin, so $\nu_{F,i}$, $\nu_{T,i}$, $\varepsilon_{f,i}$, etc. The assumption of independence discussed in the previous paragraph says that if you know $\nu_{F,i-1}$ and $\nu_{F,i+1}$ then you should have *no idea* what $\nu_{F,i}$ might be (the same goes for the efficiencies or any other quantity). In practice, for a quantity like p_T , this is simply not true: we would expect all quantities to have a reasonably smooth distribution as a function of p_T . Thus the categories are not independent. It is, however, of little importance when one’s knowledge of $\nu_{F,i}$ is primarily driven by information in category i itself rather than by information in ‘adjacent’ categories. This motivates one to choose binning wisely – enough so as to represent the shape of the distribution in each variable, but few enough that inter-category correlations are unimportant.

Essentially, the model from Figure 3.1 has been pasted multiple times, one for each category of event. Each has its own set of efficiencies and rates of production of real and fake events. Correspondingly, rather than just measuring n_T and n_L , we measure both of these quantities for *each* event category.

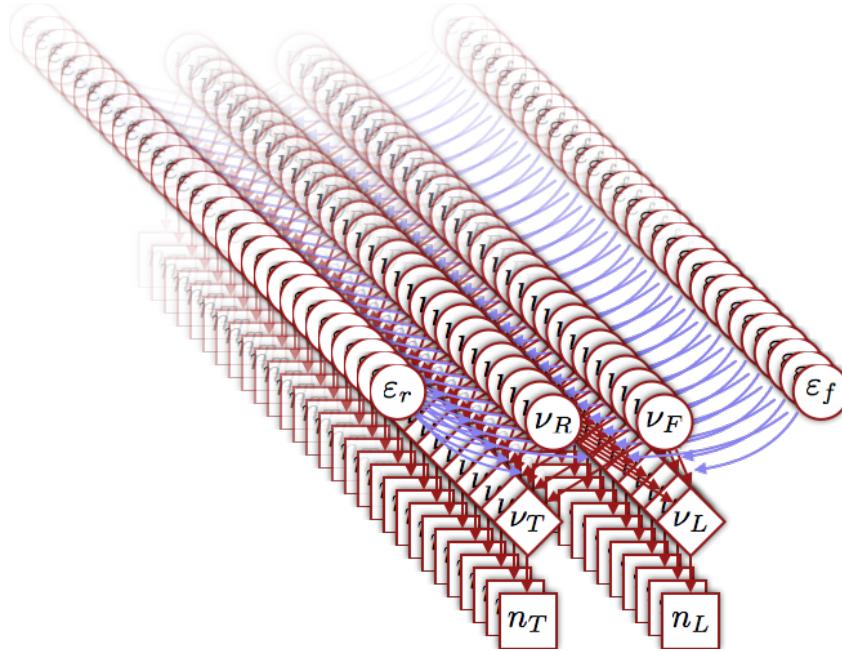


Figure 3.2: Sketch of the event production model in the case of events with a single lepton, where the lepton could fall in one of several independent categories.

Since the signal region is formed by summing over all the tight categories, predictions for each of these categories individually are not *necessarily* of interest. One approach could be to consider each category completely separately, and then sum the predictions at the end, but this could lead to trouble in situations with few events – for a given category one may only observe a single (or zero) events as either tight or loose. Either some form of sensible combination will be necessary, or the previous assumption of independence between categories should be dropped.

Multiple leptons

An alternative way that events can become more complex is with the presence of additional leptons. For example, one might consider the case of events with exactly two leptons; in this case the generalisation is shown in Figure 3.3. The number of parameters again increases, since now each lepton can independently be either real or fake, or tight and loose. Since we wish to estimate expected numbers of *events*, we need to consider all the possible combinations of tight and loose – with m leptons there are 2^m such combinations.

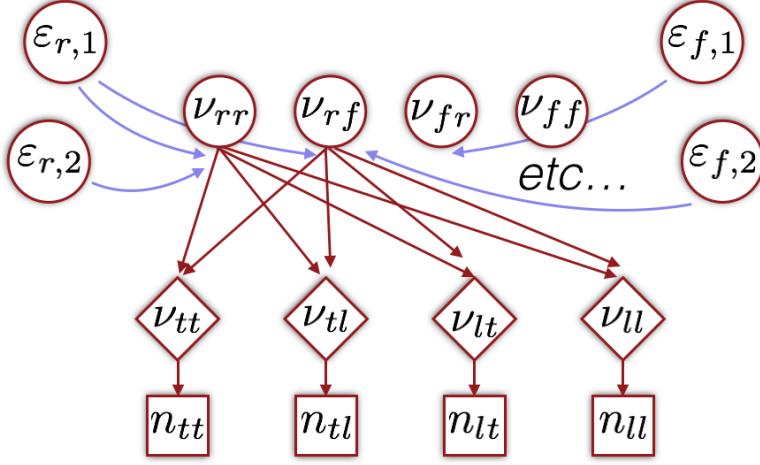


Figure 3.3: Generalising to an event with two leptons (right). It is necessary to consider the two leptons ordered in some fashion, even if it is randomly. Some arrows on the right hand diagram are omitted for clarity.

Finally, one can imagine what happens when additionally adding multiple lepton categories, as in Figure 3.4. To consider this, it helps to additionally consider the concept of an *event category*. This concept assigns a unique identifier to an event based upon the lepton category of each of its constituent leptons, with some definite ordering (such as by p_T). Thus if each lepton can be in one of N_ω categories, and an event has m leptons, there are N_ω^m possible event categories.

This can clearly be a very large number. In the analysis presented in chapter 5, for events with two leptons there are more than 30k such event categories – given that the analysis additionally considers three lepton events (with over 6 million event categories), it is clear that a truly independent treatment of each event category is simply impossible, for the reasons stated earlier.

This potentially enormous category multiplicity is the core problem that fake estimation methods must address.

3.5 Potential solutions

Whilst the detailed descriptions of the methods will be presented in chapter 4, these all must have an answer for the “hard problem” that has been outlined above. Possible approaches are:

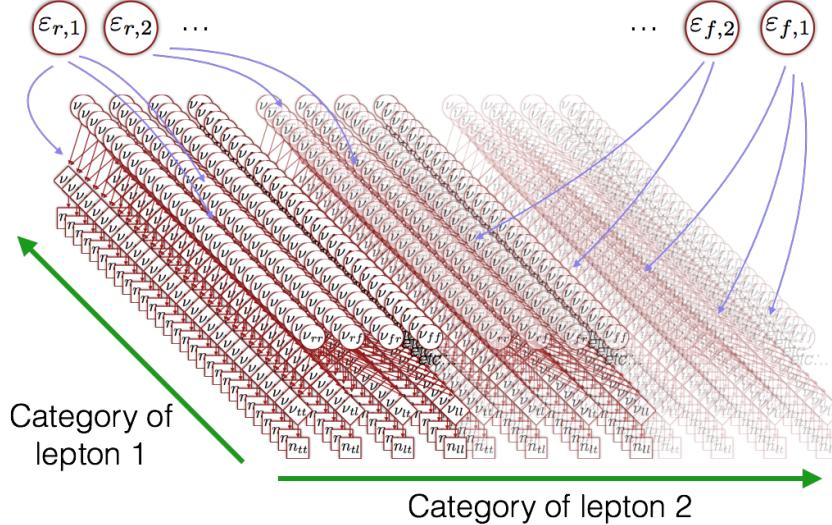


Figure 3.4: Visualisation of how multiple lepton categories are handled in events with two leptons. If there are n lepton categories then n^2 event categories emerge, each with their own rate parameters and observed numbers of events. Efficiencies are shared between the different event categories, as is indicated.

- **Combining all event categories:** This is the approach aimed for by the matrix method, albeit in an approximate fashion. This reflects quite closely what is desired, since the signal region itself is just a sum over all contributing event categories.
- **Enforce smoothness between event categories:** Often, the fundamental problem will be the implicit assumption that each lepton efficiency category is independent. When the categories are subsets of e.g. (p_T, η) , as is commonly the case, the efficiencies will likely be rather smooth functions, $\varepsilon(p_T, \eta)$, and similarly for the underlying rates e.g. $\nu_R(p_T, \eta)$. The “right” way to approach the problem then might be to use techniques such as kernel density estimation [75] to find plausible *functions* $\nu_T(p_T, \eta)$. Whilst elegant, unfortunately these techniques are computationally expensive, and as such are not pursued in this thesis.
- **Combining lepton categories:** Rather than combining event categories, by reducing the number of lepton categories sufficiently the problem could be made tractable again. This approach is needed to make the Bayesian sampling method feasible, as demonstrated in chapter 6.
- **Form a parametric model with a reduced parameter set:** This is used in the tests of the likelihood method, as described in section 4.5.3. The potential

drawbacks are that the parameter set must be sufficient to describe the physics processes producing the event, and so performing a sufficient reduction may still be difficult. It also creates a more complex network that might be tricky to work with computationally.

The methods mentioned naturally differ in other ways, but this discussion will be left for the next chapter.

Chapter 4

Data-driven methods for estimation of fake backgrounds

4.1 Introduction

The problem of fake estimation, and the reasons that a solution is desired, have already been detailed in chapter 3. A widely used data-driven method used to estimate fake-backgrounds is the so-called *matrix method*, first employed in ATLAS in [76]. It has since been employed by many ATLAS analyses, including [77–105]; CMS have used the method less prolifically [106–108], and claim to originally describe it in [109]. This chapter endeavours to distil the aims of the method, present a novel syntax that clarifies the intent, and then demonstrates how the method may be generalised to cater for a wider class of events simultaneously. Further to this, a larger methodological improvement is presented which uses a likelihood-based method to provide more accurate, and statistically meaningful, estimates of the fake background, and ultimately a more reliable limit on a postulated signal process.¹ A final method is presented which is related to the likelihood method, but aims instead to produce the Bayesian posterior estimate of the expected fake contamination in a signal region.

This chapter makes use of both frequentist and Bayesian statistical methods. For a useful introduction to statistics and inference, particularly from a Bayesian viewpoint, see MacKay’s excellent book [110]. It also contains a brief discussion on confidence intervals, which shall be used in this section.

¹It would also have been possible to frame the discussion in terms of discovery significance – this choice reflects the author’s guess as to which procedure would prove more relevant in Run 2.

4.2 Common definitions

The terminology described here is equally applicable to several types of objects in certain scenarios (for example electrons, muons or b -jets). The core concepts to all variants of matrix methods are:

- **Real** (r): An object that represents the intended entity.
- **Fake** (f): An object that does not represent the intended entity.
- **Tight** (t): An object that passes tighter quality requirements.
- **Loose** (l): An object that passes baseline and not tighter quality requirements.
- $\tilde{l} = l \cup t$: An object that passes baseline quality requirements.
- $t \cap f$: An object that is both tight and fake – this is the type of object whose presence we wish to predict.

With these definitions, one of the most important quantities in methods such as the matrix method are the *efficiencies*, defined thus:

- $\varepsilon_r = P(t|r\tilde{l})$: The ‘real efficiency’ – the probability that an object is tight given that it is real and passes baseline quality.
- $\varepsilon_f = P(t|f\tilde{l})$: The ‘fake efficiency’ – the probability that an object is tight given that it is fake and passes baseline quality.
- $\bar{\varepsilon}_r = P(l|r\tilde{l}) = 1 - \varepsilon_r$: Symbol used for convenience.
- $\bar{\varepsilon}_f = P(l|f\tilde{l}) = 1 - \varepsilon_f$: Symbol used for convenience.

Since these efficiencies are often expected to vary as a function of kinematic quantities such as object p_T , they are typically subdivided into several bins, or ‘categories’. This shall be explored in more depth later. Later in this chapter such categories will be labelled $\omega_1, \omega_2, \dots$, with the efficiencies gaining an additional subscript e.g. $\varepsilon_{\omega_1 r}$.

Given such object-level quantities, it is necessary to classify events containing these objects. For a given event containing m leptons, each lepton is observed to be either l or t , and will have some category ω_i . If there are N_ω possible categories for each lepton, then the number of measurable *event categories* will be $N_\Omega = 2^m \times N_\omega^m$; note that this includes tightness and looseness in addition to the object category information. If the number of leptons can differ between events, then one introduces a sum with appropriate

limits over m . Each of these event categories will correspond to an event that is either tight (T) or loose (L), and real (R) or fake (F). That is, events are classified by the same symbols as for the object level, but in capitals, hence the set of all baseline events is \tilde{L} .

Experimentally, one counts how many events fall into each of the N_Ω sub-regions of a given signal region, yielding the set of integers $\{n_{\Omega_i}\}$. For the purpose of the physics analysis being performed, one is most often interested in the total number of tight events, $n_T = \sum_{\Omega_i \subset T} n_{\Omega_i}$. Usually this is the quantity with which a limit on the cross section of a new physics model is placed.

Further to this, the observed numbers of events are often assumed to be the particular values of a Poisson distributed random variable, with means specified by ν . For example, one can have $n_T \sim \text{Poiss}(\nu_T)$; in general the indices on the rate ν correspond to those on the observation n .

Explicitly, we define

- $n_{\mathcal{X}}$: The number of events in set \mathcal{X} .
- $\langle n_{\mathcal{X}} \rangle$: The expected value $\mathbb{E}[n_{\mathcal{X}}|C]$, for some possible condition C that will depend on the context, and will be specified when necessary.
- $\nu_{\mathcal{X}}$: The rate of events being produced in set \mathcal{X} , assuming that $n_{\mathcal{X}} \sim \text{Poiss}(\nu_{\mathcal{X}})$.

4.3 The vanilla matrix method

4.3.1 Single object events

Following the notation introduced in section 4.2, in the case where the efficiencies and n_R , n_F are known, it is claimed that the following statement holds true:

$$\begin{pmatrix} \langle n_T \rangle \\ \langle n_L \rangle \end{pmatrix} = \begin{pmatrix} \mathbb{E}[n_T | n_R, n_F] \\ \mathbb{E}[n_L | n_R, n_F] \end{pmatrix} = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ 1 - \varepsilon_r & 1 - \varepsilon_f \end{pmatrix} \begin{pmatrix} n_R \\ n_F \end{pmatrix}. \quad (4.1)$$

The result follows by considering the real/fake event counts to be known integers, which are then divided into tight and loose components according to a binomial distribution using the probabilities contained in the efficiencies. In fact, it can be noted that equation (4.1)

is a special case of a relation between the means of Poisson distributions

$$\begin{pmatrix} \nu_T \\ \nu_L \end{pmatrix} = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ \bar{\varepsilon}_r & \bar{\varepsilon}_f \end{pmatrix} \begin{pmatrix} \nu_R \\ \nu_F \end{pmatrix}. \quad (4.2)$$

This shall be made use of later when discussing the likelihood formulation of the method.

At this point it is worth remarking that the predictive power of this method stems from real and fake objects have different composition in the tight and loose sets; that is, the ability to distinguish fake from real comes from having $\varepsilon_f \neq \varepsilon_r$. Moreover, in all scenarios considered in this paper, the tight region will be considered to favourably contain real objects, and the loose region to favourably contain fake objects. It can be seen that this corresponds to having $\varepsilon_f \ll \varepsilon_r$.

In this method, and also in the other variants that follow, a significant task is that of estimating the efficiencies ε_r and ε_f . In this chapter we will not concern ourselves with this issue; rather it shall be addressed in the more specific contexts within which this method is applied.

It is now claimed that, given measurements of the efficiencies and of n_T and n_L , an estimator $\hat{n}_{T \cap F}$ of $\langle n_{T \cap F} \rangle = \mathbb{E}[n_{TF}|n_T, n_L]$ can be formed. The first step is to invert the relation found in equation (4.1) to obtain

$$\begin{pmatrix} n_R \\ n_F \end{pmatrix} = \frac{1}{\varepsilon_r - \varepsilon_f} \begin{pmatrix} \bar{\varepsilon}_f & -\varepsilon_f \\ -\bar{\varepsilon}_r & \varepsilon_r \end{pmatrix} \begin{pmatrix} \langle n_T \rangle \\ \langle n_L \rangle \end{pmatrix}. \quad (4.3)$$

The matrix is invertible so long as $\varepsilon_f \neq \varepsilon_r$, however as previously explained we typically expect $\varepsilon_r \gg \varepsilon_f$ so this condition will be assumed to hold true. Following this, estimates of $\langle n_R \rangle$ and $\langle n_F \rangle$ are obtained by a heuristic swapping of which values are observed and expected, and upon which variables the expectation is conditioned. Note also that the expectation $\langle n_R \rangle$ in this context is conditioned on different quantities to $\langle n_T \rangle$, as shown explicitly below.

$$\begin{pmatrix} \mathbb{E}[n_R|n_T, n_L] \\ \mathbb{E}[n_F|n_T, n_L] \end{pmatrix} \approx \begin{pmatrix} \hat{n}_R \\ \hat{n}_F \end{pmatrix} = \frac{1}{\varepsilon_r - \varepsilon_f} \begin{pmatrix} \bar{\varepsilon}_f & -\varepsilon_f \\ -\bar{\varepsilon}_r & \varepsilon_r \end{pmatrix} \begin{pmatrix} n_T \\ n_L \end{pmatrix}. \quad (4.4)$$

The hatted quantities, \hat{n}_R and \hat{n}_F represent the matrix method estimators for the expectation values on the far left-hand side. Formally, the expectation values should be

given by an expression akin to equation (4.1), but with the efficiency terms replaced; e.g. $\varepsilon_r = P(l|r\tilde{l})$ should be replaced by $P(r|l\tilde{l})$ using Bayes' theorem. It can also be seen that the estimators do not always reflect the expectation value because it permits negative values for \hat{n}_R or \hat{n}_F . For example, consider the case where $n_T = 7, n_L = 1, \varepsilon_r = \frac{4}{5}$, and $\varepsilon_f = \frac{1}{5}$. One can then show that $\hat{n}_R = 9$, and $\hat{n}_F = -1$; this negative estimator for the expectation of the number of fake events is clearly unphysical. Despite this shortcoming being possible, as observed in [79], in other analyses it has been found to be less of a problem when all regions in question have moderately high statistics.

By further selecting the expected fake component from here, and applying equation (4.1) again (specifically the identity that $\hat{n}_{T\cap F} = \varepsilon_f \hat{n}_F$), one obtains an estimator for the *expected* number of tight and fake events

$$\hat{n}_{T\cap F} = \frac{\varepsilon_f}{\varepsilon_r - \varepsilon_f} (\varepsilon_r(n_T + n_L) - n_T). \quad (4.5)$$

This method has been used in recent publications, including [101], to form a fake estimate in signal regions with one lepton. The method will, in section 4.5.4, be seen to emerge in many cases as a maximum likelihood estimator of the background contribution. As has already been noted, it is possible for the matrix method *estimator* \hat{n}_F , and hence the estimator $\hat{n}_{T\cap F}$ above to be negative; in fact it is also possible to have $\hat{n}_{T\cap F} > n_T + n_L$. In these cases it is later shown that this maximum likelihood property no longer holds, and as such the method ought not be trusted if results are yielded in these regimes.

4.3.2 Efficiencies varying with kinematic quantities

It is commonly the case that the efficiencies will vary in terms of kinematic properties of the objects being studied; for example the values of p_T and $|\eta|$. It is assumed that the efficiencies will vary slowly as a function of such parameters, and as such can be measured in a fixed number of categories, N_ω . As mentioned in section 4.2, *events* with m leptons will have N_Ω categories. For the single object case described here, there is a one-to-one mapping between object and event categories.

More generally however, one would define quantities such as n_{tl} , the number of events with the first² lepton tight and the second loose – others are defined similarly.

²The definition of “first” can depend on the analysis. Often it is chosen to be the hardest according to p_T .

The number of tight *events* would now be denoted $n_T \equiv n_{tt}$. In order to include the possible categories for each lepton, n_{tt} must be further subdivided to take into account all combinations, that is

$$\begin{aligned} n_{tt} &= n_{tt}^{\omega_1\omega_1} + n_{tt}^{\omega_1\omega_2} + \dots \\ &= \sum_{i,j} n_{tt}^{\omega_i\omega_j}. \end{aligned} \tag{4.6}$$

In this notation, $n_{tt}^{\omega_1\omega_2}$ indicates the number of events with two tight leptons, where the first is in category ω_1 , and the second in ω_2 . The result in equation (4.5) can then be applied to each orthogonal subset of events, and the overall prediction will be the sum of all contributions. The efficiencies themselves would then be written with an extra index representing the category which they represent; the real efficiency for category ω_1 would be $\varepsilon_{\omega_1 r}$.

The same end result can be achieved by computing the estimate on an event-by-event basis, yielding a weight for each event with an object in \tilde{L} . These weights would then be summed together in a final step. The weight is given by equation (4.5) in the special case that one or the other of n_T and n_L is 1, and the other is 0. This has the advantage of automatically making predictions about distributions in variables that have not explicitly been considered in producing the fake estimate, since each weight is associated to an event with specific properties, e.g. m_{eff} or $p_{\text{T}}^{\text{miss}}$. That is, histograms can be produced in any desired variable that can be determined for each event in data and filled with the computed weight of each event in \tilde{L} to give the background estimate.

A further advantage to the weight-based workflow will be seen in subsequent sections, since the additional subdivisions for events with larger numbers of objects might otherwise become tricky to process.

4.3.3 Events with two or more objects

In section 4.3.1 a presentation was made of the commonly-used single-object matrix method. The method is readily extended to events with exactly two objects, and was done in [111–113].

Care must be taken, however, when defining what precisely is meant by a ‘two object event’. In particular, it is difficult to have a consistent procedure unless one requires that only events with *exactly* two objects in \tilde{L} are considered by the analysis. Otherwise

there exists an ambiguity whereby an event with two objects in t might have its highest p_T (and arguably most significant) object in l – this is problematic since it is not clear whether it should be counted as passing the nominal cuts (tt), or whether it would just be used as an lt event in forming the background estimate³.

As mentioned in the previous section, efficiencies are assumed implicitly to be functions of properties of the objects such as their p_T . Therefore, the efficiencies of the two objects in the event will, in general, be different, and as such are denoted (for example, for the real efficiencies) as ε_{r1} and ε_{r2} . Given this, a similar identity to equation (4.1) can be formed:

$$\begin{pmatrix} \langle n_{tt} \rangle \\ \langle n_{tl} \rangle \\ \langle n_{lt} \rangle \\ \langle n_{ll} \rangle \end{pmatrix} = \begin{pmatrix} \varepsilon_{r1}\varepsilon_{r2} & \varepsilon_{r1}\varepsilon_{f2} & \varepsilon_{f1}\varepsilon_{r2} & \varepsilon_{f1}\varepsilon_{f2} \\ \varepsilon_{r1}\bar{\varepsilon}_{r2} & \varepsilon_{r1}\bar{\varepsilon}_{f2} & \varepsilon_{f1}\bar{\varepsilon}_{r2} & \varepsilon_{f1}\bar{\varepsilon}_{f2} \\ \bar{\varepsilon}_{r1}\varepsilon_{r2} & \bar{\varepsilon}_{r1}\varepsilon_{f2} & \bar{\varepsilon}_{f1}\varepsilon_{r2} & \bar{\varepsilon}_{f1}\varepsilon_{f2} \\ \bar{\varepsilon}_{r1}\bar{\varepsilon}_{r2} & \bar{\varepsilon}_{r1}\bar{\varepsilon}_{f2} & \bar{\varepsilon}_{f1}\bar{\varepsilon}_{r2} & \bar{\varepsilon}_{f1}\bar{\varepsilon}_{f2} \end{pmatrix} \begin{pmatrix} n_{rr} \\ n_{rf} \\ n_{fr} \\ n_{ff} \end{pmatrix}. \quad (4.7)$$

This factorisation makes the assumption that the processes of each object in the event becoming tight or loose are independent. Whilst not necessarily true, in practice it is found to work, primarily because the objects are separated in the detector by an overlap removal procedure. The result of passing tight cuts is only dependent on spatially local properties, hence it is logical that this assumption should be applicable. An inversion procedure can again be applied, and it can be shown that the condition for the inverse of the matrix in equation (4.7) existing is that $\varepsilon_{r1} \neq \varepsilon_{f1}$ and $\varepsilon_{r2} \neq \varepsilon_{f2}$. Whilst quantities such as n_{rr} are used in the above discussion, in practice the event-weight workflow will be used, as described in section 4.3.2.

At this point one can see that the method can be extended for m objects by considering an analogous identity to that in equation (4.7) but with a $2^m \times 2^m$ efficiency matrix. Performing explicit inversion of such matrices either symbolically or numerically is possible, however a much more compact approach is considered in section 4.4.

4.4 Generalised matrix method

Firstly some new notation will be used to describe the existing methods, and having introduced this it will become clear how to extend to multi-object systems in a general

³Such ambiguities are addressed more generally by the method in section 4.4

way. Finally specific workflows will be considered for two concrete examples; one using leptons and the other using b -jets.

4.4.1 Summation formalism

The first insight is to note that equation (4.7) can be re-written in terms of a Kronecker product [114]:

$$\begin{aligned} \begin{pmatrix} \langle n_{tt} \rangle \\ \langle n_{tl} \rangle \\ \langle n_{lt} \rangle \\ \langle n_{ll} \rangle \end{pmatrix} &= \begin{pmatrix} \varepsilon_{r1}\varepsilon_{r2} & \varepsilon_{r1}\varepsilon_{f2} & \varepsilon_{f1}\varepsilon_{r2} & \varepsilon_{f1}\varepsilon_{f2} \\ \varepsilon_{r1}\bar{\varepsilon}_{r2} & \varepsilon_{r1}\bar{\varepsilon}_{f2} & \varepsilon_{f1}\bar{\varepsilon}_{r2} & \varepsilon_{f1}\bar{\varepsilon}_{f2} \\ \bar{\varepsilon}_{r1}\varepsilon_{r2} & \bar{\varepsilon}_{r1}\varepsilon_{f2} & \bar{\varepsilon}_{f1}\varepsilon_{r2} & \bar{\varepsilon}_{f1}\varepsilon_{f2} \\ \bar{\varepsilon}_{r1}\bar{\varepsilon}_{r2} & \bar{\varepsilon}_{r1}\bar{\varepsilon}_{f2} & \bar{\varepsilon}_{f1}\bar{\varepsilon}_{r2} & \bar{\varepsilon}_{f1}\bar{\varepsilon}_{f2} \end{pmatrix} \begin{pmatrix} n_{rr} \\ n_{rf} \\ n_{fr} \\ n_{ff} \end{pmatrix} \\ &= \begin{pmatrix} \varepsilon_{r1} & \varepsilon_{f1} \\ \bar{\varepsilon}_{r1} & \bar{\varepsilon}_{f1} \end{pmatrix} \otimes \begin{pmatrix} \varepsilon_{r2} & \varepsilon_{f2} \\ \bar{\varepsilon}_{r2} & \bar{\varepsilon}_{f2} \end{pmatrix} \begin{pmatrix} n_{rr} \\ n_{rf} \\ n_{fr} \\ n_{ff} \end{pmatrix}. \end{aligned} \quad (4.8)$$

This is possible since the matrix method treats each object in the event independently. Whilst the set of events in which there is one high and one low p_T object are treated separately to those where both objects are high p_T , fundamentally the matrix method is formulated in terms of probabilities of a given *object* to pass certain selections, from which a statement about the event can be inferred.

Whilst the Kronecker product neatens some of the notation, in order to achieve greater freedom the set of 4 variables n_{tt} , n_{tl} , n_{lt} , and n_{ll} should be rewritten as a rank 2 ‘tensor’⁴, $\mathcal{T}_{\alpha_1\alpha_2}$. Similarly the tensor corresponding to the ‘real or fake’ combinations shall be labelled $\mathcal{R}_{\alpha_1\alpha_2}$. Each index α_i corresponds to one object, and it can take values $\alpha_i \in \{t, l\}$ for \mathcal{T} , or $\alpha_i \in \{r, f\}$ for \mathcal{R} . Now the Kronecker product can be expressed in terms of contracting a 2×2 matrix with each index of the tensor. This matrix is that formed by the efficiencies shown above, denoted $\phi_{1\beta_1}^{\alpha_1}$, where ϕ s with different indices are implicitly assumed to take appropriate values given the kinematics of the object which they describe (the first subscript will subsequently be dropped when the intention is

⁴Strictly speaking the objects in this section are not tensors, rather just mathematical objects with several indices.

clear). In this way, the result in equation (4.8) can be written more compactly as

$$\langle \mathcal{T}_{\beta_1 \beta_2} \rangle = \phi_{1 \beta_1}^{\alpha_1} \phi_{2 \beta_2}^{\alpha_2} \mathcal{R}_{\alpha_1 \alpha_2}, \quad \phi_1 = \begin{pmatrix} \varepsilon_{r1} & \varepsilon_{f1} \\ \bar{\varepsilon}_{r1} & \bar{\varepsilon}_{f1} \end{pmatrix}. \quad (4.9)$$

Finding the inverse relation is now much more straightforward, since instead of having to invert a 4×4 matrix one merely uses the inverse of each 2×2 matrix. Given a measurement $\mathcal{T}_{\alpha_1 \alpha_2}$, one can hence write the full background estimate as $\langle \mathcal{T}_{\alpha_1 \alpha_2}^F \rangle$, where

$$\langle \mathcal{T}_{\nu_1 \nu_2}^F \rangle = \phi_{\nu_1}^{\mu_1} \phi_{\nu_2}^{\mu_2} \zeta_{\mu_1 \mu_2}^{\beta_1 \beta_2} \phi^{-1}{}_{\beta_1}^{\alpha_1} \phi^{-1}{}_{\beta_2}^{\alpha_2} \mathcal{T}_{\alpha_1 \alpha_2}.$$

Of this tensor, one is typically most interested in the component $\langle \mathcal{T}_{tt}^F \rangle$, corresponding to the prediction in the tt region. The logic that ensures that this is a prediction of the *fake* and tight component is encoded in the ζ tensor, which selects out the expected fake component. For example, if $rr \equiv R$ and $\{rf, fr, ff\} \equiv F$ then one would choose $\zeta_{12}^{12} = \zeta_{21}^{21} = \zeta_{22}^{22} = 1$, and all other components 0. It is most easily visualised in the outer product basis, as below, and is seen to be equivalent to an identity matrix with the upper left hand 1 changed to a 0. This prevents the rr component from contributing, whilst retaining all others that have at least one object being fake:

$$\zeta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

With the expression of the matrix method in this form, it is evident how the method can be generalised to events with any number of objects. In particular, one finds the expression for those events containing exactly m objects to be

$$\langle \mathcal{T}_{\nu_1 \dots \nu_m}^F \rangle = \phi_{\nu_1}^{\mu_1} \dots \phi_{\nu_m}^{\mu_m} \zeta_{\mu_1 \dots \mu_m}^{\beta_1 \dots \beta_m} \phi^{-1}{}_{\beta_1}^{\alpha_1} \dots \phi^{-1}{}_{\beta_m}^{\alpha_m} \mathcal{T}_{\alpha_1 \dots \alpha_m}. \quad (4.10)$$

In the preceding equation the tensor ζ is of the general form

$$\zeta_{\mu_1 \dots \mu_m}^{\beta_1 \dots \beta_m} = \delta_{\mu_1}^{\beta_1} \dots \delta_{\mu_m}^{\beta_m} h(\beta_1, \dots, \beta_m, \nu_1, \dots, \nu_m),$$

where the function h takes values 0 or 1 to pick out the sets of indices β_i that, when considered together, make for an event that should be classified as ‘fake’ (e.g. in the exactly two object case this might be all those events with fewer than 2 real objects). In general ζ can also depend on the output tight/loose configuration being computed, as shown in the functional form of h through its dependence on the ν_i indices.

The result in equation (4.10) is most readily applied on an event-by-event basis, where the input tensor $\mathcal{T}_{\alpha_1 \dots \alpha_m}$ has exactly one component = 1, and the rest 0. Depending on the precise selection, it will typically be necessary to consider each event as having a different value of m , and then assign a weight for each element of $\langle \mathcal{T}^F \rangle$ which represents a configuration that would be considered a signal event. Further details on how this is done are provided in the case studies in section 4.4.3.

Results with explicit categories

In section 4.5, the more explicit notation for object categories shall be used. As such it shall be useful to note the results for the generalised matrix method in this form also. Specifically, if one has exactly two leptons, then the analogous relation to equation (4.9) is

$$\begin{aligned} \langle \mathcal{T}_{\beta_1 \beta_2}^{\omega_i \omega_j} \rangle &= \phi_{\omega_i \beta_1}^{\alpha_1} \phi_{\omega_j \beta_2}^{\alpha_2} \mathcal{R}_{\alpha_1 \alpha_2}^{\omega_i \omega_j}, \\ \phi_{\omega_i \{t,l\}}^{\{r,f\}} &= P(\{t,l\} | \{r,f\} \omega_i \tilde{l}), \quad \phi_{\omega_i} = \begin{pmatrix} \varepsilon_{\omega_i r} & \varepsilon_{\omega_i f} \\ \bar{\varepsilon}_{\omega_i r} & \bar{\varepsilon}_{\omega_i f} \end{pmatrix}, \end{aligned} \quad (4.11)$$

where the same notation as equation (4.6) is adopted to label object categories on the \mathcal{T} object. Subsequently the background estimate for events that are fake is denoted as before, but this time will contain sums over all possible categories of the leptons

$$\langle \mathcal{T}_{\nu_1 \nu_2}^F \rangle = \sum_{i,j} \left(\phi_{\omega_i \nu_1}^{\mu_1} \phi_{\omega_j \nu_2}^{\mu_2} \zeta_{\mu_1 \mu_2}^{\beta_1 \beta_2} \phi_{\omega_i \beta_1}^{-1} \phi_{\omega_j \beta_2}^{-1} \mathcal{R}_{\alpha_1 \alpha_2}^{\omega_i \omega_j} \right). \quad (4.12)$$

In the case with more leptons, corresponding summations over categories will hence be required.

4.4.2 Propagation of uncertainties

Another important consideration is the propagation of uncertainties to the final estimate of the fake rate. It will be assumed that there are one or more distinct sources of uncertainty on each measured efficiency, in addition to the statistical uncertainty in the values of the components of \mathcal{T} . In this section it shall be assumed that the functions transforming the underlying random variables can be taken to be approximately linear in the vicinity (about $\pm 1\sigma$) of the measured point – in fact in many cases the relationship is exactly linear, and so no approximation is being made.

Derivatives of an event weight

In this section the methods of propagation of standard uncertainty shall be used [115]. The main requirement to do this is to be able to compute the first derivative of the result with respect to each variable. For compactness, let the tensors Φ and Φ^{-1} be defined to be the products of all ϕ and ϕ^{-1} terms respectively for a given event, such that

$$\langle \mathcal{T}_{\nu_1 \dots \nu_m}^F \rangle = \Phi_{\nu_1 \dots \nu_m}^{\mu_1 \dots \mu_m} \zeta_{\mu_1 \dots \mu_m}^{\beta_1 \dots \beta_m} \Phi_{\beta_1 \dots \beta_m}^{-1} \mathcal{T}_{\alpha_1 \dots \alpha_m}.$$

The derivatives with respect to ε_r for each lepton can then be shown to be

$$\begin{aligned} \frac{\partial}{\partial \epsilon_{r_i}} \langle \mathcal{T}_{\nu_1 \dots \nu_m}^F \rangle &= \Phi_{\nu_1 \dots \nu_{i-1} \nu_{i+1} \dots \nu_m}^{\mu_1 \dots \mu_{i-1} \mu_{i+1} \dots \mu_m} \zeta_{\mu_1 \dots \mu_m}^{\beta_1 \dots \beta_m} \Phi_{\beta_1 \dots \beta_{i-1} \beta_{i+1} \dots \beta_m}^{-1} \\ &\quad \left(\frac{\partial \phi_{\nu_i}^{\mu_i}}{\partial \epsilon_{r_i}} \phi_{\beta_i}^{-1} + \phi_{\nu_i}^{\mu_i} \frac{\partial \phi_{\beta_i}^{-1}}{\partial \epsilon_{r_i}} \right) \mathcal{T}_{\alpha_1 \dots \alpha_m}, \end{aligned}$$

since ζ is independent of the efficiencies. An analogous expression can be derived for ε_f . At this point it is useful to note that the derivatives $\frac{\partial \phi}{\partial \epsilon_r}$, $\frac{\partial \phi^{-1}}{\partial \epsilon_r}$ etc. are 2×2 matrices that are easy to compute, for example in this case we have

$$\phi = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ \bar{\varepsilon}_r & \bar{\varepsilon}_f \end{pmatrix} \Rightarrow \frac{\partial \phi}{\partial \epsilon_r} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, \quad \frac{\partial \phi^{-1}}{\partial \epsilon_r} = \frac{1}{(\varepsilon_r - \varepsilon_f)^2} \begin{pmatrix} -\bar{\varepsilon}_f & \varepsilon_f \\ \bar{\varepsilon}_f & -\varepsilon_f \end{pmatrix}.$$

Uncertainty scheme on ε

In general the same efficiency will be used for more than one object in an event, and as such this introduces a correlation that needs to be taken into account. This section

considers the mechanism by which the efficiencies are measured to introduce correlations in a way that allows uncertainties to be propagated efficiently even when using the event-weight technique, where it is desirable to have the sum over events to happen last. The most general final result can be found in equation (4.15) with common approximations in limiting cases following. Whilst the techniques and results herein are not in any way revolutionary, the author is unaware of any previous document describing these prescriptions.

Consider a set of efficiencies $\{\varepsilon_i\}$ ⁵ which have been computed as a function of some data and MC simulations, with i denoting the category of the efficiency. In general they can be parametrised by

$$\varepsilon_i = f(\mathbf{a}_i, \mathbf{b}_i, \mathbf{x})$$

where the variables \mathbf{a}_i , \mathbf{b}_i , and \mathbf{x} differ in terms of the correlation structures that exist internally between the values of the vectors for different i . Each element of a given vector is assumed to be a random variable independent of all other elements, that is if $\mathbf{a}_i = a_{i\mu}$ then $u(a_{i\mu}, a_{j\nu}) = \delta_{\mu\nu} u(a_{i\mu}, a_{j\mu})$, and similarly for \mathbf{b} and \mathbf{x} . Furthermore, there is no correlation between the cross terms, that is

$$u(\mathbf{a}_i, \mathbf{b}_j) = u(\mathbf{a}_i, \mathbf{x}) = u(\mathbf{b}_i, \mathbf{x}) = 0 \quad \forall i, j,$$

where the notation $u(x, y)$ indicates the covariance between two random variables x and y , as used in [115]. The \mathbf{a} term is intended to correspond to a ‘statistical’ component of the uncertainty, where each efficiency ε_i is using a different, uncorrelated, set of values. As such, representing the vector as $\mathbf{a}_i = \{a_{i\mu}\}$, we find that only the diagonal (variance) term is non-zero. The \mathbf{b} term represents some variables that vary systematically with unknown correlation between different ε_i – at the end of the computation one can either apply a conservative estimate assuming maximal correlation, or alternatively one might be able to justify that the correlation is negligible on physical grounds. Finally, the \mathbf{x} variable represents parameters that are common to the computation of all efficiencies. In this case the same variables are used explicitly in the functional form of ε_i , as above.

⁵The ‘real’ and ‘fake’ subscripts are dropped in this section, since the same argument applies to both cases, and it should be clear how to generalise the result.

The covariance scheme can hence be summarised as follows:

$$\begin{aligned} u(a_{i\mu}, a_{j\nu}) &= \delta_{\mu\nu}\delta_{ij}u^2(a_{i\mu}) \\ u(b_{i\mu}, b_{j\nu}) &= \delta_{\mu\nu}u(b_{i\mu}, b_{j\mu}) \\ u(x_\mu, x_\nu) &= \delta_{\mu\nu}u^2(x_\mu). \end{aligned}$$

Given these, one can now write down the full covariance matrix for the efficiencies. Through use of the chain rule, it can be shown that for some set of functions $f_i(a_\alpha)$, with some known $u(a_\alpha, a_\beta)$, the covariance on the f s is given by

$$u(f_i, f_j) = \sum_\alpha \sum_\beta \frac{\partial f_i}{\partial a_\alpha} \frac{\partial f_j}{\partial a_\beta} u(a_\alpha, a_\beta).$$

Using this, after some computation one finds that

$$u(\varepsilon_i, \varepsilon_j) = \delta_{ij} \underbrace{\sum_\mu \left(\frac{\partial \varepsilon_i}{\partial a_{i\mu}} \right)^2 u^2(a_{i\mu})}_{u_{\text{stat}}^2(\varepsilon_i)} + \sum_\mu \underbrace{\frac{\partial \varepsilon_i}{\partial x_\mu} \frac{\partial \varepsilon_j}{\partial x_\mu} u^2(x_\mu)}_{\pm u_{\text{corr},\mu}(\varepsilon_i) u_{\text{corr},\mu}(\varepsilon_j)} + \sum_\mu \underbrace{\frac{\partial \varepsilon_i}{\partial b_{i\mu}} \frac{\partial \varepsilon_j}{\partial b_{j\mu}} u(b_{i\mu}, b_{j\mu})}_{\leq u(b_{i\mu}) u(b_{j\mu})}, \quad (4.13)$$

where $u_{\text{stat}}^2(\varepsilon_i)$ has been packaged, since it is clear that this quantity is sufficient for subsequent propagation. If, in general, we expected that $\frac{\partial \varepsilon_i}{\partial x_\mu} > 0$ we could also safely combine the fully correlated part in terms of $u_{\text{corr},\mu} = \left| \frac{\partial \varepsilon_i}{\partial x_\mu} \right| u(x_\mu)$, although there is no reason *a priori* that this will be the case.

The final weight W , indicating the overall estimator for the expected number of fake and tight and events, is given by the sum of weights on a set of events, that is $W = \sum_\alpha w_\alpha$. For each event α , w_α is computed using an expression like that in equation (4.12), and then as previously noted the final result is obtained by summing the estimators from all events.

Given that each w_α is a function of the efficiencies whose derivatives can be computed, as has been shown previously⁶, the squared uncertainty on W due to the propagated

⁶The event weight w corresponds to the sum of the relevant components of $\langle \mathcal{T}^F \rangle$, the derivatives of which were calculated in the previous section.

uncertainties on the efficiencies is given by

$$u^2(W) = \sum_i \sum_j \sum_\alpha \sum_\beta \frac{\partial w_\alpha}{\partial \varepsilon_i} \frac{\partial w_\beta}{\partial \varepsilon_j} u(\varepsilon_i, \varepsilon_j).$$

This can then be combined with equation (4.13) to yield the final propagated uncertainty, which is found to be

$$u^2(W) = \sum_i \left(\sum_\alpha \frac{\partial w_\alpha}{\partial \varepsilon_i} u_{\text{stat}}(\varepsilon_i) \right)^2 + \sum_\mu^{n_x} \left(\sum_\alpha \sum_i \frac{\partial w_\alpha}{\partial \varepsilon_i} \frac{\partial \varepsilon_i}{\partial x_\mu} u(x_\mu) \right)^2 \quad (4.14)$$

$$+ \sum_\alpha \sum_\beta \sum_i \sum_j \sum_\mu^{n_b} \frac{\partial w_\alpha}{\partial \varepsilon_i} \frac{\partial w_\beta}{\partial \varepsilon_j} \frac{\partial \varepsilon_i}{\partial b_{i\mu}} \frac{\partial \varepsilon_j}{\partial b_{j\mu}} u(b_{i\mu}, b_{j\mu}). \quad (4.15)$$

The last term of this expression can be rewritten in the two limits that would be used in a computation where the full covariance matrix is not known; that is assuming no correlation

$$\sum_i \left(\sum_\alpha \frac{\partial w_\alpha}{\partial \varepsilon_i} u_{\text{uncorr}}(\varepsilon_i) \right)^2, \quad u_{\text{uncorr}}^2(\varepsilon_i) = \sum_\mu^{n_b} \left(\frac{\partial \varepsilon_i}{\partial b_{i\mu}} \right)^2 u^2(b_{i\mu}), \quad (4.16)$$

and assuming full correlation

$$\sum_\mu^{n_b} \left(\sum_\alpha \sum_i \frac{\partial w_\alpha}{\partial \varepsilon_i} \frac{\partial \varepsilon_i}{\partial b_{i\mu}} u(b_{i\mu}) \right)^2. \quad (4.17)$$

The final result in equation (4.15), and the two limits for the third term in equations (4.16) and (4.17) are written in a fashion that is readily computable in a weight-based workflow. In general, there are some components that can be summed up for each individual event, for example the contributions sum over efficiencies i in the second term of equation (4.15), which is then followed by a sum over events α . There are several such independent terms that need to be summed separately, and these are then combined in quadrature.

Given that the efficiency computation is a separate procedure from the application process that is focused upon in this section, it is useful to identify which uncertainties or related quantities need to be provided alongside the efficiencies. For the uncorrelated uncertainty components, $u_{\text{stat}}(\varepsilon_i)$ and $u_{\text{uncorr}}(\varepsilon_i)$ (in the limit of equation (4.16)) are clearly necessary and sufficient, however in the case of the correlated component a signed

quantity, $\frac{\partial \varepsilon_i}{\partial x_\mu} u(x_\mu)$, is ideally necessary. However, if it is the case that the only quantity that can feasibly be computed is

$$u_{\text{corr},\mu}(\varepsilon_i) = \left| \frac{\partial \varepsilon_i}{\partial x_\mu} \right| u(x_\mu),$$

then this is insufficient to properly perform the propagation in the second term of equation (4.15). In this case, the term must be replaced by the conservative expression

$$\sum_{\mu}^{n_x} \left(\sum_{\alpha} \sum_i \left| \frac{\partial w_{\alpha}}{\partial \varepsilon_i} \right| u_{\text{corr},\mu}(\varepsilon_i) \right)^2.$$

4.4.3 Applications in lepton and jet systems

Example: fakes in m lepton region

In a case where we are interested in identifying the contribution from fake leptons, we are often interested in whether a *particular* lepton is real or fake; for example, the analysis described in the following chapter needs to be aware of whether the two leptons it contains are of the same charge. As such, after having computed \mathcal{T}^F using equation (4.10) each component of it that *might* correspond to a tight event is treated separately. Additionally, variables such as the dilepton invariant mass and m_{eff} will be re-computed for each of these weights.

For example, if one were to measure an event with three leptons in \tilde{l} , $e^+e^-\mu^+$, with configuration tll , then the matrix method will produce the following

Input	Output			
	lll	w_{lll}	$e_l^+ e_l^- \mu_l^+$	Fails cuts
		
$e^+e^-\mu^+, tll \longrightarrow$	ttl	w_{ttl}	$e_t^+ e_t^- \mu_l^+$	Fails cuts
	tlt	w_{tlt}	$e_t^+ e_l^- \mu_t^+$	2 lepton SS
	ltt	w_{ltt}	$e_l^+ e_t^- \mu_t^+$	Fails cuts
	ttt	w_{ttt}	$e_t^+ e_t^- \mu_t^+$	> 2 lepton

Of the possible combinations, only two pass the channel selection cuts, presuming additional requirements such as trigger matching are also satisfied.

At the LHC, it is very rare to have more than 4 leptons in an event. As such, it is feasible to use the above method in an event-by-event workflow by simply storing each event multiple times, according to how many tight/loose permutations pass the cuts of the signal region. However, considering events multiple times in this fashion necessitates a correction to the usual handling of the *statistical* component of the uncertainty, due to the splitting of each input event into multiple streams that are each processed separately with unique weights. If two such streams from one event, with weights w_{ttl} and w_{tlt} , were to fall into one signal region, then the total statistical uncertainty should be

$$(w_{ttl} + w_{tlt})^2,$$

whereas if they were treated as completely separate events (as software utilities such as HistFitter by default will do), they would be combined as $w_{ttl}^2 + w_{tlt}^2$.

Example: fakes in 3 b -jet region

The analysis in the next chapter also uses a matrix method to evaluate the contribution of fake b -jets in one signal region, with results shown in section 5.5.5. Whilst in principle the same workflow could be used here as for the leptonic case, storing the multiple t/l permutations for each event, it is neither necessary nor computationally feasible to do so.

In this particular example, one is attempting to identify whether the b -tagged jets in an event do in fact correspond to b -jets, or if one or more of them are in fact mis-tagged jets from other sources. Since, unlike in the leptonic case, there are no particular event selections on *which* of the jets in an event are b -jets, it is safe to simply form one weight per event. This is also the only feasible approach; it is not uncommon to have ≥ 10 jets in an event, at which point the number of ways of selecting three or more of the jets as tight is ~ 1000 . As such, after having computed \mathcal{T}^F , one simply sums those components that correspond to tight events (three or more b -jets), and takes the result as the weight for the given event.

4.5 Likelihood-based fake estimation

As an alternative to the matrix method discussed previously, a likelihood-based method is now presented. Since this concept ties in closely with that of the frequentist limit-setting procedures used in ATLAS, a digression into the CL_s method is made before returning to the topic of fake background estimation.

4.5.1 Frequentist limit setting procedures

ATLAS analyses use the CL_s method [116, 117] to place an upper limit on the parameter that sets the mean of the Poisson distribution controlling the appearance of events in one or more signal regions. A short overview of the method is presented here.

One starts by defining a likelihood for a set of parameters (which shall be defined subsequently), given some observed data \mathbf{X} ; that is, the probability of observing the data given the parameters

$$\mathcal{L}(\mu, \boldsymbol{\theta} | \mathbf{X}) = P(\mathbf{X} | \mu, \boldsymbol{\theta}). \quad (4.18)$$

The data term \mathbf{X} should contain all data that might have been observed, including that from auxiliary experiments (such as control regions), which might constrain the model parameters. In order to set a limit, there must be a single parameter $\mu \in \mathbb{R}$, $\mu > 0$, representing the ‘strength’ of the signal process. In this context ‘signal’ refers to the physics model which is being tested, which is of a form that is expected to increase the expected number of events in the signal region if present. The μ strength parameter is typically used to scale the nominally expected cross section for the signal process, however there is no reason that it could not represent the overall rate of appearance of signal events.⁷ The other parameter, $\boldsymbol{\theta}$ represents a set of *nuisance* parameters – so called simply because these are not the parameters upon which we are trying to set a limit. Finally, \mathbf{X} is a placeholder for one or more real (in our case, typically integer) values corresponding to experimental measurements. At this point it is not necessary to consider the functional form of \mathcal{L} .

⁷The likelihood will of course behave differently with respect to other parameters, for example those controlling luminosity uncertainty, but it remains a valid expression.

Next one defines a set of test statistics representing the ‘profiled likelihood ratio’, q_{μ_t} , defined to be

$$q_{\mu_t}(\mathbf{X}) = \begin{cases} -2 \ln \left(\frac{\mathcal{L}(\mu_t, \hat{\boldsymbol{\theta}}_{\mu_t} | \mathbf{X})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}} | \mathbf{X})} \right) & \mu_t > \hat{\mu} \\ 0 & \text{otherwise} \end{cases}, \quad (4.19)$$

where here it is written explicitly as a function of the observed data \mathbf{X} . For a given \mathbf{X} , the likelihood is maximised when one chooses the parameter values $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$. If one also fixes the value of μ to some proposed value μ_t , then the set of values $\hat{\boldsymbol{\theta}}_{\mu_t}$ is defined to be that which maximises the likelihood for the particular value of μ_t .⁸ It is always the case that $q_{\mu_t} \geq 0$, and qualitatively it can be seen that small values correspond to cases where μ_t is ‘about right’, and large values to the case where μ_t is too large to explain the data (that is, the numerator is decreasing relative to the denominator). The precise form of q_{μ_t} is motivated by the desire to be able to apply approximations derived by Wald and Wilks [118, 119] in high-statistics scenarios.

It is then necessary to construct a p -value for a particular model; in this case a particular assumed *true* value for the parameter of interest, μ^* . This p -value is defined to be the probability that the data could have been more discrepant than that observed, given the hypothesis defined by μ^* . By definition, this general case is represented by

$$\begin{aligned} p(\mu_t, \mu^*, \mathbf{X}) &= P(q_{\mu_t}(\mathbf{X}') \geq q_{\mu_t}(\mathbf{X}) | \mathbf{X}' \sim \mu^*) \\ &= \sum_{\mathbf{X}'} \theta(q_{\mu_t}(\mathbf{X}') - q_{\mu_t}(\mathbf{X})) P(\mathbf{X}' | \mu^*), \end{aligned} \quad (4.20)$$

where the unemboldened θ is the Heaviside step function,

$$\theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4.21)$$

Conceptually this involves a sum over the set of all possible datasets we could have recorded, weighted according to the probability of getting such a dataset were the signal strength μ^* . The result necessarily involves the slightly problematic term, $P(\mathbf{X}' | \mu^*)$; this can be expressed in terms of the likelihood by introducing an integral over the nuisance

⁸This use of maximal values of the likelihood is sometimes referred to as *profiling* over the nuisance parameters. Common terminology is to say the the nuisance parameters are ‘profiled away’, since the value of the maximum is no longer a function of these variables.

parameters:

$$\begin{aligned} P(\mathbf{X}'|\mu^*) &= \int P(\mathbf{X}'|\mu^*, \boldsymbol{\theta}') P(\boldsymbol{\theta}'|\mu^*) d\boldsymbol{\theta}' \\ &= \int \mathcal{L}(\mu^*, \boldsymbol{\theta}'|\mathbf{X}') P(\boldsymbol{\theta}'|\mu^*) d\boldsymbol{\theta}' \\ &\approx \mathcal{L}\left(\mu^*, \hat{\boldsymbol{\theta}}_{\mu^*}|\mathbf{X}'\right). \end{aligned} \quad (4.22)$$

In this last approximate step, $\hat{\boldsymbol{\theta}}_{\mu^*}$ is the value that maximises $\mathcal{L}\left(\mu^*, \hat{\boldsymbol{\theta}}_{\mu^*}|\mathbf{X}\right)$, the likelihood conditioned on the *observed* data. The rationale behind this is that the prior term, $P(\boldsymbol{\theta}'|\mu^*)$, ought to be peaked around ‘reasonable’ values of $\boldsymbol{\theta}'$, and that hence the most significant contribution to the integral will come from this point. Whilst in principle this point should be computed without knowledge of \mathbf{X} , for example with MC simulations, in practice the correlations introduced by doing so are not significant; the important thing is to have a reasonable model for the dependence of $\boldsymbol{\theta}'$ on μ^* . We can then write an approximate form for the p -value which is easily computed by performing pseudoexperiments

$$\begin{aligned} p(\mu_t, \mu^*, \mathbf{X}) &\approx \frac{\sum_{\mathbf{X}'} \theta(q_{\mu_t}(\mathbf{X}') - q_{\mu_t}(\mathbf{X})) \mathcal{L}(\mu^*, \hat{\boldsymbol{\theta}}_{\mu^*}|\mathbf{X}')}{\sum_{\mathbf{X}'} \mathcal{L}\left(\mu^*, \hat{\boldsymbol{\theta}}_{\mu^*}|\mathbf{X}'\right)} \\ &= \mathbb{E}[\theta(q_{\mu_t}(\mathbf{X}') - q_{\mu_t}(\mathbf{X}))], \quad \mathbf{X}' \sim \left(\mu^*, \hat{\boldsymbol{\theta}}_{\mu^*}\right), \end{aligned} \quad (4.23)$$

where this expectation is easily computed by throwing ‘toy’ datasets \mathbf{X}' , distributed according to the likelihood with μ^* as the signal parameter, and the corresponding maximum likelihood estimate of the nuisance parameters. In the literature, equation (4.20) is often written as an integral over the test statistic – in practice this is useful when considering the asymptotic approximations described in [120], however this chapter will not aim to develop these.

The two terms defined in the limit setting framework used by ATLAS are the ‘signal plus background’ hypothesis, or CL_{s+b} , and the ‘background only’ hypothesis, CL_b . In the generally accepted notation, these are then defined

$$CL_{s+b}(\mu_t, \mathbf{X}) = p(\mu_t, \mu_t, \mathbf{X}) \quad (4.24)$$

$$1 - CL_b(\mu_t, \mathbf{X}) = p(\mu_t, 0, \mathbf{X}), \quad (4.25)$$

where the use of “ $1 - CL_b$ ” is such that CL_b itself refers to the probability that the data would *not* have been more discrepant than that observed if there were no signal present. From these, a third quantity is defined,

$$CL_s(\mu_t, \mathbf{X}) = \frac{CL_{s+b}(\mu_t, \mathbf{X})}{1 - CL_b(\mu_t, \mathbf{X})}. \quad (4.26)$$

An upper bound with a confidence level of α can then be set on the strength parameter, defined by

$$\max_{\mu_{\max,\alpha}} \left[CL_s(\mu_{\max,\alpha}, \mathbf{X}) \leq 1 - \alpha \right]. \quad (4.27)$$

The motivation for this variable stems from a calculation by Zech [121]. The usual frequentist procedure would be to place a limit using the CL_{s+b} *p*-value, however when applied to a system consisting of a signal and background component it can give results that were deemed difficult to interpret; specifically that the upper limit on the signal strength parameter would be 0 if the estimated background contribution were equal to or greater than the number of observed events. To avoid this, Zech proposed to alter the definition of the *p*-value to include a condition that the (unmeasurable) number of observed background events be \leq the total number of observed events; applying basic probability theory then yields the result above.

4.5.2 Limit setting with the matrix method

In the context of limit setting, the output from the matrix method is treated on a par with those irreducible background components estimated from MC samples. Once the central value and corresponding uncertainty is estimated as described in section 4.4, the background mean \bar{b} and uncertainty σ_b are fed into a joint likelihood for the signal and background rates, μ and b , given the number of events observed in the signal region n_T . In this simple case it would take the form

$$\mathcal{L}(\mu, b | n_T) = \text{Poiss}(n_T; \mu + b) \text{Gauss}(\bar{b}; b, \sigma_b). \quad (4.28)$$

When setting the limit, the nuisance parameter b is profiled away to form the test statistic q_{μ_t} , as defined in equation (4.19), and then upper limits (CL_s or CL_{s+b}) at some confidence level are then formed by following the recipe outlined in section 4.5.1.

Method	Cost of calculation	Limit quality	Other names
A	low	poor, frequently undercovers	matrix method
B	very high	very good	likelihood matrix method
C	quite low	good	—

Table 4.1: An overview of the three methods discussed in this chapter, and their relative strengths and weaknesses. ‘Limit quality’ refers to whether CL_{s+b} limits tend to have the correct frequentist coverage properties, and also avoid unnecessary overcoverage.

4.5.3 An extended likelihood method

In this section, it will be convenient to introduce abbreviated names for three variations on the combined background estimation and limit setting procedure, as also used by the author in [122]. Method A denotes the matrix method which has previously been described. Method B labels the concept of setting a limit with a likelihood that includes all information from the background estimate itself; this shall be the focus of this section.

Finally, Method C is used to denote a half-way house where the full likelihood of Method B is used to produce a maximum likelihood estimate (MLE) for the fake rate and corresponding uncertainty. This uncertainty represents both an uncertainty with which the efficiencies are known, as well as statistical limitations of the observed data. It is estimated with the MINOS method [123], by taking the values of the fake rate where the minimum of the negative log likelihood with respect to the *remaining* parameters increases by 0.5 from its minimum value. A limit is then placed using an expression identical to that in equation (4.28), where \bar{b} and σ_b take the aforementioned MLE fake rate and uncertainty.

All of these methods have relative advantages and disadvantages, which have been described in the text but are also summarised in Table 4.1.

Whilst the matrix method can yield limits that suffer from undercoverage (a confidence interval whose statistical coverage is below that intended), as noted further in the subsection below, this can be avoided for a purely data-driven background if the *full* likelihood, including all datasets used to make the measurement, is used in the limit

setting procedure. That is, one should use

$$\mathcal{L}(\mu, \boldsymbol{\theta}|n_t, n_l, n_{tt}, \dots), \quad (4.29)$$

where $\boldsymbol{\theta}$ represents the set of nuisance parameters. If the leptons can fall into more than one category, quantities should be replaced with the separate terms from equation (4.6). Each of these quantities can be considered as independent random variables with a Poisson distribution. The means of these Poisson distributions will be denoted as functions of the parameters e.g. $\nu_{tt}^{\omega_1\omega_1}(\mu, \boldsymbol{\theta})$; the likelihood then factorises and takes a form similar to equation (4.28)

$$\mathcal{L}(\mu, \boldsymbol{\theta}| \dots, n_{tt}^{\omega_1\omega_1}, \dots) = \dots \text{Poiss}(n_{tt}^{\omega_1\omega_1}; \nu_{tt}^{\omega_1\omega_1}(\mu, \boldsymbol{\theta})) \dots P(\bar{\boldsymbol{\theta}}|\boldsymbol{\theta}). \quad (4.30)$$

The final term represents constraints placed on the nuisance parameters by external measurements.

Coverage of frequentist limits

There is some discussion in the literature regarding how the incorporation of background components with some uncertainty affects the frequentist coverage properties of p -value limits [124]. In particular, when one is considering a background that is constrained e.g. from an MC sample, the acceptance region for the hypothesis test in the full Neyman construction will vary according to the value assumed by the nuisance parameter(s) controlling the strength of the background. In an approximated scheme, such as the profile likelihood method used in the computation of CL_{s+b} and CL_s , the coverage can hence deviate from that nominally expected; potentially significantly if the background overestimates the data. Since both Methods A and C feed information into the likelihood in the same way (and hence have the shortcoming that the likelihood used in the limit-setting procedure is not the likelihood for *all* the data), one should not be surprised if one or both methods under-cover. It is later shown, however, that by virtue of the MLE fake rate being more ‘sensible’ than that from the matrix method, any deviations in coverage from that nominally expected would be less extreme in Method C than with Method A.

Choice of parametrisation

The efficacy of any likelihood method depends on a sensible choice of θ . In this case, they must completely describe how events from both signal and background are expected to be divided between the different categories. For example, one *could* directly use $\nu_{tt}^{\omega_1\omega_1}$ etc. as the free parameters, but this removes all predictive power!

A method has been trialled within the ATLAS collaboration [125, 126] that uses a similar parametrisation to the matrix method – the efficiencies described before, in addition to the rates separated both by object category and real/fake-ness. Whilst this has an advantage of making minimal assumptions about how a given background process distributes itself between these categories, it does lead to a very large parameter space. For example, even with three objects coming from only three possible categories, there are already 80 such parameters (before considering efficiencies). Since any form of prediction will require a maximisation of the likelihood over this input parameter space, and since such global maximisations become computationally more expensive as dimensionality increases, the author has chosen to use an alternative parametrisation.

Decision tree parametrisation

Diagrammatically, the parametrisation used in this work is displayed in Figure 4.1. For every event that is generated, it is first decided how many leptons that event ought to contain. This is controlled by a set of parameters $\{\alpha_m\}$, each of which corresponds to the probability of forming an event with m leptons. As noted in the caption these must sum to 1 in an appropriate fashion. For each lepton, a category ω_i is assigned to it with probability β_i , and it is then further assigned to be either f with probability π_i or r with probability $1 - \pi_i$. Formally, $\beta_i \equiv P(\omega_i|\tilde{l})$ and $\pi_i \equiv P(f|\omega_i\tilde{l})$. Efficiencies are then used in the usual way to select objects as being t or l .

Using these terms, together with one extra parameter denoting the mean of the Poisson distribution controlling the total production of tight *events*⁹, one can compute the terms such as $\nu_{tt}^{\omega_1\omega_1}$ in equation (4.30). It should be noted that one of these trees must exist for every separate ‘component’ that is being fitted – that is, at least one for the hypothesised signal process and one for the fake component of the background, and

⁹One could alternatively use the overall production of \tilde{L} events, however it is essential to have the rate of T events as a parameter for any signal component, since this is the quantity upon which one wishes to place a limit.

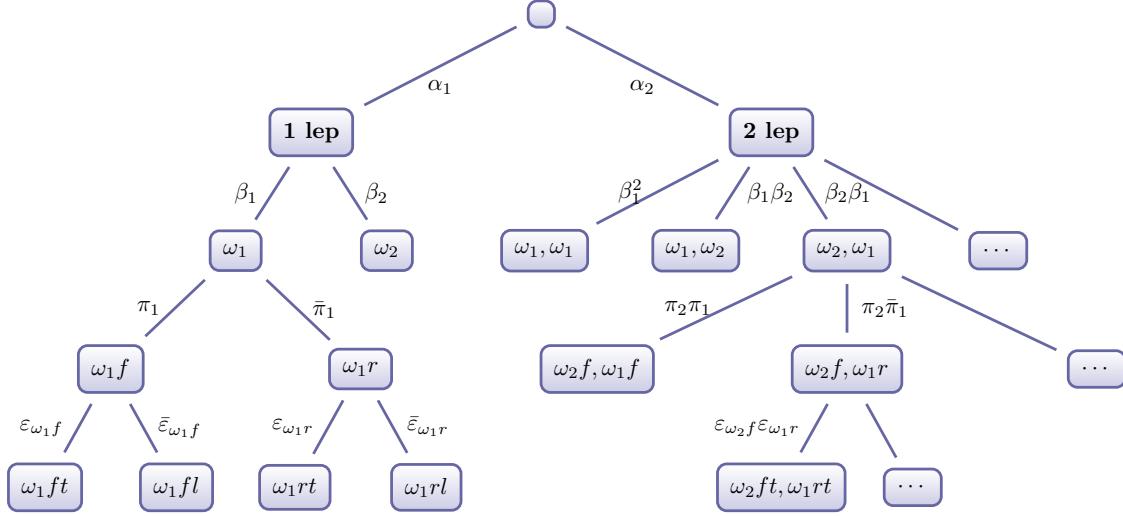


Figure 4.1: Parametrisation proposed to be used with a likelihood matrix method, to determine how events should be split between different categories, due to its constituent objects themselves falling into different categories. The leftmost branch is complete, the others are not. In general one could allow both more lepton categories, as well as more leptons in the event. Note that $\sum_{m=0}^{m_{\max}} \alpha_m = \sum_{i=1}^{N_\omega} \beta_i = 1$, where m_{\max} is the largest number of leptons that can be produced in a given event. Additionally, the abbreviation $\bar{\pi}_i = 1 - \pi_i$ is used.

optionally others for additional background components that have been estimated using MC samples. This is because for *every* component it is necessary to say something about how its events are divided up between the different categories.

4.5.4 Emergence of the matrix method

One interesting side effect of the parametrisation trialled by ATLAS is that it can be used to justify the approximate result in equation (4.4) in certain circumstances. The following discussion presents this justification, along with its limitations.

When considering a likelihood as a product of Poisson terms as in equation (4.28), and neglecting the Gaussian terms involving the efficiencies, the negative log likelihood will be

$$-\ln \mathcal{L} = \sum_{\omega, \beta} (\langle T_\beta^\omega \rangle - T_\beta^\omega \ln \langle T_\beta^\omega \rangle), \quad (4.31)$$

where for a set of m leptons the categories and tight/looseness information are compacted into vectors ω and β of length m respectively. From equation (4.2), one finds

$$\langle \mathcal{T}_\beta^\omega \rangle = \sum_{\alpha'} \Phi_{\omega\beta}^{\alpha'} \langle \mathcal{R}_{\alpha'}^\omega \rangle, \quad \Phi_{\omega\beta}^{\alpha'} = \phi_{\omega_{i_1}\beta_1}^{\alpha'_1} \phi_{\omega_{i_2}\beta_2}^{\alpha'_2} \dots, \quad (4.32)$$

where α' is a vector representing whether each lepton is real or fake.

One can now differentiate equation (4.31) with respect to $\langle \mathcal{R}_\alpha^\omega \rangle$, $\forall \omega, \alpha$, using the identity in equation (4.32). In order to locate the minimum of the negative log likelihood, one sets all these derivatives to 0, yielding

$$\sum_{\beta'} \left(1 - \frac{\mathcal{T}_{\beta'}^\omega}{\langle \mathcal{T}_{\beta'}^\omega \rangle} \right) \Phi_{\omega\alpha}^{\beta'} = 0, \quad \forall \omega, \alpha. \quad (4.33)$$

These are satisfied if $\langle \mathcal{T}_{\beta'}^\omega \rangle = \mathcal{T}_{\beta'}^\omega \forall \beta'$, the result of which being that upon inversion equation (4.32) will look like

$$\langle \mathcal{R}_\alpha^\omega \rangle = \sum_{\beta'} \Phi_{\omega\beta'}^{-1} \mathcal{T}_{\beta'}^\omega, \quad (4.34)$$

analogously to equation (4.4).

Whilst this is a valid operation for the problem as stated above, it should be noted that the minimum of $-\ln \mathcal{L}$ is represented by equation (4.33) *only* when the components of $\langle \mathcal{R}_\alpha^\omega \rangle$ are > 0 . It is also only useful in the case where the components of $\langle \mathcal{R}_\alpha^\omega \rangle$ are readily assigned to either signal plus other ‘real’ backgrounds (those typically estimated from MC samples) and the fake background.

4.6 Bayesian fake rate posterior

The new likelihood methods discussed in section 4.5 revolved around considering the full likelihood with which one wished to set limits on a new physics scenario, and performing the fake estimation as part of that limit setting process. The more approximate approach of using the MLE fake rate and estimated uncertainty as the input for a simpler likelihood for limit setting was also presented; it is this latter concept that shall now be improved upon.

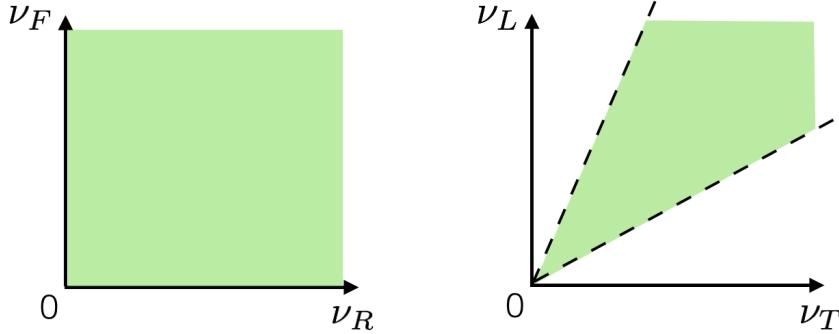


Figure 4.2: Diagram representing the transformation from the valid quarter plane in real/fake space to the valid region in tight/loose space. The angles of the dotted lines on the right plot with respect to the axes are determined by the efficiencies.

An alternative approach to using an MLE and uncertainty is to sample from the Bayesian posterior distribution on the fake rate. With the use of appropriately uninformative priors, this distribution will have a peak corresponding to the desired MLE estimate. By doing this one gains not only the ability to more reliably identify the location of this peak (since one can never be sure if only a local minimum has been located), but also visualise the entire posterior probability density function (pdf). This extra information could, in principle, be included in the limit setting stage. Moreover, rather than use an MLE of the fake rate it might be more desirable to estimate the median value of the posterior, and use a credible interval as an estimate of the asymmetric uncertainty band.

4.6.1 Sketch of the method

By means of an introduction to the sampling method presented in this section, it is instructive to first consider the simple case of events containing a single lepton with only one possible category previously presented in Figure 3.1.

Recalling the discussion from section 4.3.1, the matrix method is capable of returning negative fake rate estimates. This can be understood in a more graphical way: in the (ν_R, ν_F) plane only the positive quadrant contains valid values of the parameters. For a given pair of efficiencies $\varepsilon_r, \varepsilon_f$, this quarter plane valid region can be drawn as a triangular shape in the (ν_T, ν_L) plane, as shown in Figure 4.2. This follows directly from the relation in equation (4.2). In the case of events containing m leptons, the space in which this wedge shape appears has 2^m dimensions.

Whilst our observed values of n_T and n_L place constraints on our belief of the values of ν_T and ν_L , it is also necessary to respect the boundaries exemplified in Figure 4.2. Coupled with some belief on the values of the efficiencies, if one can draw points in the *valid region* of the tight and loose plane, this can be directly turned into a value of the fake and tight rate.

4.6.2 Sampling methods

In this section we develop the numerical methods necessary to sample from the posterior. The derivations shall be shown for the single lepton case, however the generalisation to arbitrary lepton number should hopefully be clear.

The quantity we are aiming to obtain is

$$P(\nu_{TF}|n_T, n_L, \tilde{\boldsymbol{\varepsilon}}), \quad (4.35)$$

where $\tilde{\boldsymbol{\varepsilon}}$ represents measurements made in some control regions that constrain the efficiencies, and other quantities defined as in chapter 4. This can be defined as marginalisation of a joint probability of a sufficient set of parameters to uniquely define the system. That is, we can write

$$\begin{aligned} P(\nu_{TF}|n_T, n_L, \tilde{\boldsymbol{\varepsilon}}) &= \int d\boldsymbol{\varepsilon} \int_{\Psi_{RF}} d\nu_R d\nu_F \delta(\nu_{TF} - f_{RF}(\nu_R, \nu_F, \boldsymbol{\varepsilon})) \cdots \\ &\quad \cdots P(\nu_R, \nu_F, \boldsymbol{\varepsilon}|n_T, n_L, \tilde{\boldsymbol{\varepsilon}}) \end{aligned} \quad (4.36)$$

$$\begin{aligned} &= \int d\boldsymbol{\varepsilon} \int_{\Psi_{TL}(\boldsymbol{\varepsilon})} d\nu_T d\nu_L \delta(\nu_{TF} - f_{RF}(\nu_T, \nu_L, \boldsymbol{\varepsilon})) \cdots \\ &\quad \cdots P(\nu_T, \nu_L, \boldsymbol{\varepsilon}|n_T, n_L, \tilde{\boldsymbol{\varepsilon}}), \end{aligned} \quad (4.37)$$

where Ψ represents the allowed region of integration in the rate space *RF* for real/fake space, and *TL* for tight/loose space; f represents the transformation from the collection of rates and efficiencies to the tight and fake rate. Each of these are defined in the real/fake space (equation (4.36)) and the tight/loose space (equation (4.37)). One observation worth noting is that Ψ_{TL} is a function of $\boldsymbol{\varepsilon}$; this represents the fact that the allowed region in tight/loose space is defined by the efficiencies, as was shown in Figure 4.2.

Neither of these integrals can be evaluated analytically, however various numerical techniques exist that could be applied to perform the integration. Monte Carlo methods

seem appropriate, given the potentially high dimensional nature of the integral, and the most efficient methods would aim to draw samples from the probability distributions under the integral, thus not spending time in “unimportant” places. It shall be shown how Gibbs sampling may be used to compute the integral in equation (4.37), and how a more general Metropolis-Hastings algorithm could be used in equation (4.36)

Evaluating integral in tight/loose space with Gibbs sampling

One can now consider evaluating the probability contained in equation (4.37),

$$\begin{aligned} P(\nu_T, \nu_L, \boldsymbol{\varepsilon} | n_T, n_L, \tilde{\boldsymbol{\varepsilon}}) &= P(\boldsymbol{\varepsilon} | \tilde{\boldsymbol{\varepsilon}}) P(\nu_T, \nu_L | n_T, n_L, \boldsymbol{\varepsilon}) \\ &\propto P(\boldsymbol{\varepsilon} | \tilde{\boldsymbol{\varepsilon}}) P(n_T, n_L | \nu_T, \nu_L) P(\nu_T, \nu_L | \boldsymbol{\varepsilon}) \\ &\propto P(\boldsymbol{\varepsilon} | \tilde{\boldsymbol{\varepsilon}}) P(n_T | \nu_T) P(n_L | \nu_L) P(\nu_T, \nu_L | \boldsymbol{\varepsilon}), \end{aligned} \quad (4.38)$$

where the prior on the tight and loose rates is necessarily dependent on $\boldsymbol{\varepsilon}$, due to the restrictions in Figure 4.2. Note that the last of these three steps contains an assumption – namely that the efficiencies are much more tightly constrained by $\tilde{\boldsymbol{\varepsilon}}$ than by any information in (n_T, n_L) . Thus, the integral to evaluate is

$$\int d\boldsymbol{\varepsilon} P(\boldsymbol{\varepsilon} | \tilde{\boldsymbol{\varepsilon}}) \int_{\Psi_{TL}(\boldsymbol{\varepsilon})} d\nu_T d\nu_L \delta(\nu_{TF} - \dots) P(n_T | \nu_T) P(n_L | \nu_L) P(\nu_T, \nu_L | \boldsymbol{\varepsilon}). \quad (4.39)$$

There are two evidently straightforward Poisson terms, and then a prior term that requires closer inspection. In choosing $P(\nu_T, \nu_L | \boldsymbol{\varepsilon})$, it is desired to be largely uninformative, except that it must only have support in $\Psi_{TL}(\boldsymbol{\varepsilon})$. As such, Gamma distributions are used for each of ν_T and ν_L , truncated so as to remain within $\Psi_{TL}(\boldsymbol{\varepsilon})$. This is not exactly the required probability distribution – an alternative might be to enforce ν_R and ν_F to have Gamma priors, which then avoids the need for truncation, however this could not be used with a Gibbs sampler. Figure 4.3 presents an illustrative example that suggests that the difference between these scenarios is not too great. As specified in more detail in the caption, one imagines placing Gamma distribution priors in either the tight/loose or real/fake planes, and making equivalent observations of event counts in each of these planes respectively. The resulting distributions are not identical, however it is argued that the differences are of minor importance. It seems very likely that it will only influence the tails of the final fake rate posterior, given that the differences are largely in very low-probability density areas, and as such is of limited importance for the intended application.

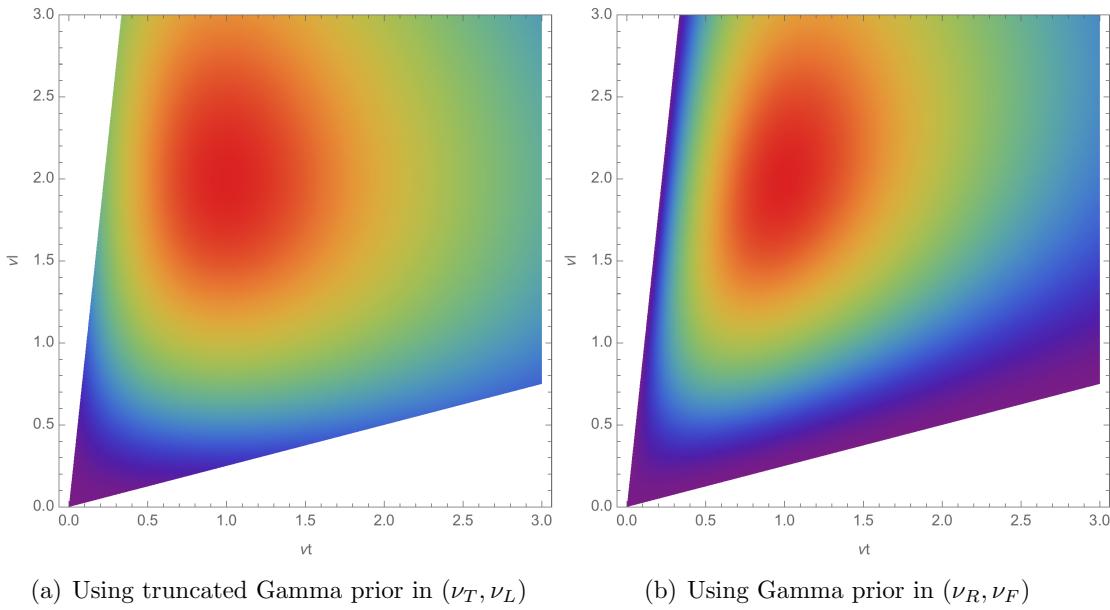


Figure 4.3: Illustration of the truncation procedure used in the model, compared to an alternative model with a Gamma distribution used in (ν_R, ν_F) -space. An example is shown here with a single lepton, single category model, and known efficiencies of $\varepsilon_r = 0.8$, $\varepsilon_f = 0.2$. For a) it is assumed that one observation of $n_T = 1$, $n_L = 2$ has been made, and for b) that an equivalent observation of $n_R = 1$, $n_F = 2$ – the two observations are exactly related by the matrix method approximation, equation (4.1). The former uses the observation to determine a Gamma posterior on (ν_T, ν_L) , truncating to remain in the allowed region. The latter uses the equivalent observation in the (ν_R, ν_F) , with the distribution subsequently plotted in the (ν_T, ν_L) -plane. Red indicates high probability mass, down to dark purple at 0. The disallowed region is shown in white.

At this point it is difficult to express the prior directly, due to the truncation procedure, so instead we define it through the complete set of conditional probabilities. These are, up to constant factors:¹⁰

$$P(\nu_T|\nu_L, \boldsymbol{\varepsilon}) \propto \theta(\nu_T|\nu_L, \boldsymbol{\varepsilon}) \text{ Gamma}(\nu_T|\alpha_0, 0) \quad (4.40)$$

$$P(\nu_L|\nu_T, \boldsymbol{\varepsilon}) \propto \theta(\nu_L|\nu_T, \boldsymbol{\varepsilon}) \text{ Gamma}(\nu_L|\alpha_0, 0), \quad (4.41)$$

where the notationally abused Heaviside step functions θ are such as to perform the necessary truncation in each case, and α_0 is a prior constant to be discussed later. Explicitly, $\theta(\nu_T|\nu_L, \boldsymbol{\varepsilon})$ is 1 inside the green region of Figure 4.2, and 0 elsewhere; the same logic applies for $\theta(\nu_L|\nu_T, \boldsymbol{\varepsilon})$. Note that the ‘shape and rate’ parametrisation of the Gamma distribution is used throughout, that is

$$\text{Gamma}(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}. \quad (4.42)$$

Through application of Bayes’ theorem one can show that if a random variable λ has prior $\text{Gamma}(\lambda|\alpha, \beta)$, and another random variable $M \sim \text{Poiss}(\lambda)$, then an observation m of M yields a posterior on λ of $\text{Gamma}(\lambda|\alpha + m, \beta + 1)$.

The conditional probabilities above are sufficient for the Gibbs sampling algorithm [110] to be applied. In the simple example above, it can be written out as follows:

1. Choose a random point, (ν_{T0}, ν_{L0}) , in $\Psi_{TL}(\boldsymbol{\varepsilon})$
2. Draw $\nu_{Ti} \sim \theta(\nu_{Ti}|\nu_{Li-1}, \boldsymbol{\varepsilon}) \text{ Gamma}(\nu_{Ti}|\alpha_0 + n_T, 1)$
3. Draw $\nu_{Li} \sim \theta(\nu_{Li}|\nu_{Ti}, \boldsymbol{\varepsilon}) \text{ Gamma}(\nu_{Li}|\alpha_0 + n_L, 1)$
4. Present pair (ν_{Ti}, ν_{Li})
5. Repeat steps 2-4 until burn in completes, and then samples are approximately independent draws from the joint distribution of the tight and loose rates.

Drawing from truncated distributions is straightforward using inverse transform sampling, so long as the inverse of the cumulative density function (CDF) is computable [127]. An implementation of the inverse of the Gamma distribution CDF can be found in the Boost C++ libraries [128].

¹⁰Here, ‘constants’ are also taken to include conditioning variables for a given expression.

Computing the truncation limits For any given draw from one of the conditional distributions above, it is necessary to compute the upper and lower allowed bounds for the rate in question. Whilst this can easily be derived analytically in the single lepton case, it becomes less trivial with more leptons. However, since it is a standard linear programming problem – maximise and minimise one rate subject to linear constraints imposed by the other rates and efficiencies – the Simplex algorithm can be used to find the bounds [129]. For the results shown later in this chapter the implementation in the GNU Linear Programming Kit [130] is used.

Choosing α_0 It is desired to choose a value of α_0 such that the priors are uninformative. In a model with a single random variable λ , which is the mean of another random variable, $M \sim \text{Poiss}(\lambda)$, then the usual uninformative prior is $\text{Gamma}(\lambda|1, 0)$, i.e. $\alpha_0 = 1$.

However, now consider $\Lambda = \sum_{i=1}^N \lambda_i$; if one again chooses priors on each λ_i of $\text{Gamma}(\lambda_i|1, 0)$, then this effectively places a prior on Λ of $\text{Gamma}(\Lambda|N, 0)$. This is appropriate if each of the λ_i are completely independent, however if in fact the quantity of interest is Λ , and if one is unsure which of the λ_i might be constrained by data, and N is big, then the prior of $\text{Gamma}(\Lambda|N, 0)$ on Λ is potentially problematic. A workaround is to choose the prior on each λ_i to be $\text{Gamma}(\lambda_i|1/N, 0)$, which then leaves an effective prior on Λ of $\text{Gamma}(\Lambda|1, 0)$, which would be more desirable due to its resembling the uninformative choice mentioned previously.

An alternative in this scenario could be to use a Jeffreys prior, which is proportional to the square root of the Fisher information. In the case of a Poisson distribution with scale parameter λ this would be $\propto \frac{1}{\sqrt{\lambda}}$. Considering the model above one could then use a joint prior

$$p(\lambda_1, \dots, \lambda_N) \propto \frac{1}{\sqrt{\Lambda}} = \frac{1}{\sqrt{\sum_{i=1}^N \lambda_i}}. \quad (4.43)$$

In practice this form could be more difficult to work with than the approximation described in the previous paragraph, since it does not easily factorise into separate terms for each λ_i .

This toy example is similar to the case with multiple categories of events. Since the quantities we are interested in are fundamentally sums over all event categories, we choose $\alpha_0 = 1/N_\Omega$, where N_Ω represents the number of event categories. Ideally it would be necessary to ensure the priors are overwhelmed, that is that the posterior distribution

is not strongly affected by choice of prior, such that by definition this choice should not be too important. In practice this isn't always achieved, and as such the approximate form above avoids biasing the estimated fake and tight rate higher than one might expect due to the presence of many categories.

Varying efficiencies The discussion above demonstrates an algorithm to sample from the joint posterior on tight and loose rates *given* that $\boldsymbol{\varepsilon}$ is known. However, the actual situation is that we must also integrate over $P(\boldsymbol{\varepsilon}|\tilde{\boldsymbol{\varepsilon}})$. If this distribution can be sampled from,¹¹ then the Gibbs sampling algorithm can be updated as follows:

1. Draw efficiencies $\boldsymbol{\varepsilon}$ from $P(\boldsymbol{\varepsilon}|\tilde{\boldsymbol{\varepsilon}})$
2. Choose a random point, (ν_{T0}, ν_{L0}) , in $\Psi_{TL}(\boldsymbol{\varepsilon})$. Set $i = 0$
3. Increment i
4. Draw $\nu_{Ti} \sim \theta(\nu_{Ti}|\nu_{Li-1}, \boldsymbol{\varepsilon})$ Gamma $(\nu_{Ti}|\alpha_0 + n_T, 1)$
5. Draw $\nu_{Li} \sim \theta(\nu_{Li}|\nu_{Ti}, \boldsymbol{\varepsilon})$ Gamma $(\nu_{Li}|\alpha_0 + n_L, 1)$
6. Repeat steps 3-5 until burn in completes
7. Present pair (ν_{Ti}, ν_{Li})
8. Repeat steps 1-7 until the desired number of samples have been drawn

Multiple event categories Given the prescription above it is straightforward to include multiple event categories in the Gibbs sampling process. After one has selected $\boldsymbol{\varepsilon}$, the burn-in loop is repeated for each event category, yielding a sample of (ν_T, ν_L) for each.

Burning in Gibbs sampling is a particular type of Markov Chain Monte Carlo (MCMC) algorithm, and like other MCMC methods it can take a while for the Markov chain to “burn in”; that is, for the samples being drawn to be truly representative of the target distribution. The number of steps required for burn in has to be determined empirically. It is clear that the number of burn in steps should only depend strongly on

¹¹In practice this is a very reasonable assumption, since one typically finds mean values through control region measurements, and assigns statistical and systematic uncertainties, with potential correlations between efficiencies. These correlation structures can then be modelled by drawing a suitable number of (potentially truncated) normally distributed random variables and determining the efficiencies.

the number of leptons, since this determines the dimensionality of the allowed region, and the approximate “shape” of this region is always the same.

Evaluating integral in real/fake space with Metropolis-Hastings

An alternative approach is to start from equation (4.36), which involves drawing samples in the joint space of real and fake rates. Since this method is not used in the remainder of the chapter a mathematical development is not included, however a sketch of important points follows.

In this case there is no need to compute limits of truncated distributions, since the allowed region is always the positive quadrant (e.g. as on the left of Figure 4.2). However, the ‘likelihood’ terms are necessarily Poisson distributions in tight/loose space, i.e. terms like $P(n_T|\nu_T(\nu_R, \nu_F, \epsilon))$. Thus it is not possible to write down the necessary conditional probabilities to use Gibbs sampling. As such a more general Metropolis-Hastings algorithm has to be used, which has the downside of necessitating some effort to go into the choice of an effective proposal distribution.

In the case under investigation here, where each category is independent one efficiencies have been drawn, the use of Gibbs sampling seems preferable since it allows avoidance of tuning a proposal distribution. However, if tighter links were introduced between the categories then this would render the Gibbs mechanism impossible. As such Metropolis-Hastings would be the natural fallback, and then operating in real/fake space would be logical so as to avoid the necessity of computing the appropriate boundaries of the allowed region at each step.

Chapter 5

An analysis of ATLAS data for like-sign lepton events

The chapters thus far have set the scene for the search described herein. A motivation for the signature is presented, and the details of the event selection procedures are described. The mechanisms used for estimating the various backgrounds are discussed in detail, and the results of these are presented. Finally the analysis is put into the context of several simplified and popular phenomenological supersymmetric models, with limits being set in their respective parameter spaces.

The author was involved in all aspects of the analysis and performed much of the critical lower-level development and maintenance required for the implementation and running of the analysis. The most significant physics contributions were centred around the background estimation procedure used for the fake leptons, the theory of which has already been discussed. This method was introduced for the latest publication of this analysis [100], which included the three-lepton signatures that necessitated the use of a generalised matrix method. The author also performed a cross-check background estimation procedure for a signal region with three b -jets, as well as various other specific cross checks to ensure our results were correct and consistent – these are not described here. The analysis has had four publications [100, 111–113], and the author has been a member on the analysis team for the last two of these. This chapter focusses on the latest publication, [100], to which the author made the largest contributions.

5.1 Motivation

When designing a ‘general’ analysis for new physics signatures, the primary considerations are to ensure that the signal regions have acceptance for a range of well-motivated models, whilst simultaneously reducing the contribution from SM processes as much as possible. The use of like-charge leptons achieves both of these goals.

Firstly, the fact that gluinos are Majorana fermions and are pair produced, means that each branch of the decay can produce a lepton of either charge. Thus the possibility of observing like-charge leptons is not significantly lower than that of seeing those of opposite-charge.

Secondly, since the SM does not predict production of particles with these properties, like-charge signatures are highly suppressed. Contributions occur from processes such as $t\bar{t} + V$, that is $t\bar{t}$ in association with a vector boson, and di- and tri-boson production. However, since $t\bar{t}$ production is kinematically similar to, and has a significantly higher cross-section than the typical signatures for which we search, two significant backgrounds come from a mis-reconstruction of this process. A semi-leptonic decay of $t\bar{t}$, where one top quark decays leptonically, and the other hadronically, can be misidentified as a leptonic decay mode if the hadronic decay products ‘fake’ a lepton in the detector. Likewise, the fully leptonic decay mode can contribute to our background if one of the leptons has its charge mis-reconstructed. As such a mixture of MC samples and data-driven methods have been used to model both the reducible and irreducible backgrounds.

This analysis additionally makes use of events containing three leptons. Whilst producing little additional contribution from the SM the extra leptons allow the analysis to target supersymmetric models with longer decay chains. By allowing them to be soft (a low p_T requirement, $p_T > 15$ GeV) this also increases sensitivity to compressed scenarios (see section 2.2.3).

5.2 Data and Monte Carlo samples

5.2.1 The 8 TeV ATLAS dataset

The data used in this analysis corresponds to the 2012 run of the LHC, with a centre of mass energy of 8 TeV. The total integrated luminosity collected by the ATLAS detector

was 20.3 fb^{-1} , with an uncertainty of 2.8% [131, 132]. This uses the centrally-provided offline calculation normalised with Van der Meer scan data, using techniques similar to those described in [133]. The average pileup during this run, i.e. the number of pp interactions within each bunch crossing, lies in the approximately range 10–20, with a mean of 20.7. Further quality selections were applied to this dataset to ensure that any period where a necessary component of the detector was considered faulty or unstable was removed. The analysis uses data from the `Egamma`, `Muons` and `JetTauEtmiss` streams.

5.2.2 Background samples

Standard model processes that contribute to the background of this analysis that cannot be accounted for using the data driven methods described in section 5.5 are estimated using centrally produced MC samples, corresponding to the MC12 production run. Samples corresponding to processes that *are* normally estimated by the data driven methods were additionally used for optimisation studies, as well as the testing of the data-driven background estimation methods themselves. After matrix element generation and parton showering using the generators described below, the ATLAS detector is simulated using either ‘full’ simulation (based on GEANT4) or Atlfast-II simulation, as described in section 1.3. Full simulation is assumed for these samples, unless otherwise stated.

Backgrounds with prompt leptons: The backgrounds with two prompt like-charge leptons, or three leptons, comprise $t\bar{t}$ in association with bosons (W , Z , H), single top in association with a Z boson (tZ) and diboson production plus jets (including $W\gamma$, WH and ZH). These are collectively referred to as the ‘irreducible’ backgrounds, since these processes are capable of producing the targeted signature directly.

The matrix elements for $t\bar{t} + V + n$ partons, where $n \in \{0, 1, \geq 2\}$, in addition to tZ , $t\bar{t} + WW$ and $VVV + \text{jets}$ were generated using `MadGraph-5.1.4.8` [17], followed by parton showering using `Pythia-6.426` [18]. Here and henceforth the abbreviation for a vector boson $V \in \{W, Z\}$ is used. The $t\bar{t} + H$, WH and VH processes are modelled in their entirety by `Pythia-8.165` [134], with the mass of the Higgs boson set to 125 GeV. Diboson samples for WW and WZ are generated using `Sherpa-1.4.1` [135], which produces matrix elements including up to three final state partons. The default parametrisations for renormalisation and factorisation scales are used, and a dilepton invariant mass cutoff of $m_{ll} > 0.1 \text{ GeV}$ is applied. The ZZ diboson sample is produced by the `Powheg-1.0` [136] interface to `Pythia8`. Finally, the $W\gamma$ sample is produced using `Alpgen-2.13` [137],

interfaced to **Herwig**-6.510 [138] for parton showering and fragmentation, and to **JIMMY** [139] for underlying event simulation.

Further to these samples used in the core measurement, additional samples representing diboson and $t\bar{t} + V$ processes were used to estimate systematic uncertainties caused by the choice of generator. These used **Sherpa** for the ZZ sample and **Powheg** for the WZ and WW processes. Furthermore, **Alpgen** was used to generate an alternative $t\bar{t} + V$ sample.

Samples for reducible backgrounds: In the process of developing the background estimates, and for the purpose of cross-checking, MC samples corresponding to the ‘reducible’ backgrounds are also used. In this context, a reducible background is one that can only produce the experimental signature through a reconstruction error, and as such includes $t\bar{t} + \text{jets}$, $W/Z + \text{jets}$ and single-top production. The $W/Z + \text{jets}$ sample is modelled using **Sherpa**, with the same configuration as for the diboson samples previously described, however with matrix elements containing up to five final state partons rather than three. In order to increase the accuracy of the prediction, the **MENLOPS** [140] treatment is applied. This applies a weight to every event that not only causes the overall normalisation to correspond to a next-to-leading order (NLO) treatment, but also introduces corrections to the shapes of distributions of kinematic variables. $t\bar{t} + \text{jets}$ samples are produced using **Powheg** with **Pythia**, and finally the single-top process is created using the **MC@NLO** [141] generator for the Wt process and s channel, and **AcerMC** [142] for the t channel.

Parton distribution functions: Two different sets of parton distribution function (PDF) sets were used with the generators described above to describe the quark and gluon content of the proton. With the **Sherpa**, **Powheg** and **MC@NLO** generators the NLO [143] PDF set is used, whilst with **MadGraph**, **Pythia** and **Alpgen** the PDFs were taken from the CTEQ6L1 [144] prediction.

Cross sections: For all background processes, the most precise theoretical available cross sections [145–147] were used to normalise the background distributions found from the above samples.

5.2.3 Signal samples

The samples corresponding to the SUSY models considered were either simulated with **Herwig++**-2.5.2 [148] or **MadGraph** interfaced with **Pythia**, using the CTEQ6L1 PDF set

in both cases. Cross sections were calculated at NLO furthermore adding the resummation of soft gluon emission to next-to-leading-logarithmic (NLL) order [149–153].

Uncertainties on these cross sections were obtained by propagating the uncertainties from the PDF sets, in addition to varying the generator factorisation and renormalisation scales, as described in [154]. Additionally, a set of samples corresponding to an mUED model were used; these were also generated with `Herwig++` with the CTEQ6L1 PDF set, and in this case the leading-order cross section computed by `Herwig++` was used. The majority of these samples were simulated using `Atlfast-II`, with the exception of the gluino-stop models in the off-shell $t\tilde{\chi}_1^0$ and on-shell $b\tilde{\chi}_1^\pm$ channels, for which the full detector simulation was applied.

A thorough account of the different SUSY models considered, along with the resulting exclusions from this analysis, can be found in section 5.7.2.

5.2.4 MC sample tuning and corrections

ATLAS data that was known to be sensitive to initial- and final-state QCD radiation, colour reconnection, hadronisation and multiple parton interactions were used to determine the parton shower parameters of the simulated samples. Specifically, the set of tuned parameters AUET2 [155] is used with `Pythia`, `Pythia8` and `Herwig`, with the exception of the `Powheg +Pythiat $t\bar{t}$` sample for which the P2011C [156] tune is applied. For samples generated with `Herwig++`, the UEEE3 [157] tune is used.

Since multiple protons can interact on each bunch crossing, an effect known as “pile-up”, corrections are applied as follows to account for the extra energy deposits that are likely to appear alongside the desired primary interaction. This effect is modelled by overlaying minimum-bias events, simulated with `Pythia8` using the AUET2 tune, onto the hard scatter event produced by the generator for the sample in question. Subsequently, simulated events are weighted such that the distribution for the average number of collisions per bunch crossing matches that observed in the recorded dataset using the `ExtendedPileupReweighting` [158] tool. The optimal agreement in the distribution of the number of primary vertices between MC and data is found when MC samples are reweighted to $(1/1.09) \times \langle \mu \rangle$, where $\langle \mu \rangle$ is the average number of interactions per bunch crossing.

Whilst the simulated samples are reconstructed using the same algorithms as used to reconstruct data, corrections are applied for known deficiencies in the detector simulation

procedure, largely derived from data control samples. The properties corrected for include the leptonic trigger efficiencies, lepton reconstruction efficiencies, object momentum scale and resolution, and finally the real and fake efficiency for the tagging of jets originating from b -quarks. Some of these are discussed in more detail in section 5.6.

5.3 Event selection

Using the basic objects reconstructed and selected as described in section 1.4, one can define the procedure by which events are either accepted or rejected for use in the analysis. This is represented by a series of cuts, every one of which must be passed. The same procedure is applied to both data and MC samples, unless stated otherwise.

The cuts can be separated into those largely associated with ‘data quality’, which tend to be common between many ATLAS analyses, and those looking to isolate those events of interest to this particular analysis. The quality cuts, with the exception of the trigger, are as follows:

1. **Good Runs List:** Isolate data-taking runs known to be good – corresponds to a total working luminosity of 20.3 fb^{-1} .
2. **Trigger skim:** Require any of the triggers mentioned in section 5.3.1 to have passed; this is a fast requirement to remove many events, although the necessary overlap removal is performed later.
3. **LAr and Tile Error:** Inhibit unwanted effects from noise bursts and data corruption in the LAr and Tile calorimeters.
4. **Incomplete events:** Reject incomplete events due to the timing, trigger and control (TTC) [159] restart procedure.
5. **Fake p_T^{miss} veto:** Reject events where p_T^{miss} is induced by jets pointing towards dead calorimeter cells.
6. **Jet cleaning:** Events required to pass the “looser” selection, limiting contribution from jets caused by detector noise, beam-induced particles, or cosmic rays.
7. **Primary vertex:** The primary vertex must have at least five tracks with $p_T > 0.4 \text{ GeV}$ associated with it.

8. **Bad & cosmic muon veto:** Remove events with poorly reconstructed muons, or muons with impact parameters implying they were likely to have been created by cosmic rays.

The event selection procedure continues, but now with those cuts which are more specific to this analysis:

9. **Two leptons:** Events are required to contain at least two signal leptons (as defined in section 1.4.1) with $p_T > 15$ GeV, the leading of which must additionally satisfy $p_T > 20$ GeV.
10. **Trigger processing:** Described further in section 5.3.1.
11. **Trigger matching:** The trigger objects must match to within $\Delta R < 0.15$ of a corresponding reconstructed lepton – if not the event is vetoed, as it suggests the trigger was fired mistakenly.
12. **Same sign:** If there are *exactly* two leptons, it is required that they have the same charge. Otherwise no requirement is imposed.
13. **Z-veto:** To reduce the contribution from Z production, veto events where the invariant mass m_{ll} of same-flavour and opposite-charge leptons are in the window $84 < m_{ll} < 98$ GeV. By definition this will only affect events with more than two signal leptons.
14. **Invariant mass:** Require leading lepton invariant mass $m_{ll} > 12$ GeV to avoid heavy flavour meson resonances.

5.3.1 Trigger selection

Due to the requirement of our signal regions on both lepton multiplicity as well as moderate p_T^{miss} , the analysis is designed to use a combination of pertinent triggers to maximise the number of potentially interesting events we accept. This step is also used to classify events as being in one of five non-intersecting categories; high p_T^{miss} , dielectron, electron-muon, muon-electron, or dimuon; each of which comes from exactly one of the overlapping data streams. Hence this ensures that no data event that appears in more than one of the `Egamma`, `Muons` or `JetTauEtmiss` is used more than once. Events falling into each category have additional requirements applied for the purpose of ensuring that trigger efficiencies are near the plateau (i.e., are approximately constant regardless of

event kinematics). A summary of all the categories and requirements can be found in Table 5.1.

By operating in the kinematic regions where the efficiency has plateaued, it allows for trigger-dependent weight corrections to be applied to MC datasets correcting for deficiencies in the simulated trigger efficiency. These are found to be approximately $\pm 2 - 4\%$ for the p_T^{miss} trigger when lepton $p_T > 20 \text{ GeV}$. For dielectron events in the ‘soft’ case when the subleading electron has $15 < p_T < 20 \text{ GeV}$, the MC events are weighted by +9%.

5.4 Signal regions

Signal regions were chosen through optimisation, based on a combination of those from previous versions of this analysis and also an older dedicated three lepton search [160, 161]. The aim is to achieve broad sensitivity for a variety of supersymmetric scenarios, broadly separated into three signal regions separated by their b -jet multiplicity, and a further two focusing on the three lepton signatures.

Event accepted into any signal region are required to have passed the event selection procedure as described in section 5.3. For each event a number of quantities are computed:

- N_{leps} – the number of leptons passing the signal requirements.
- N_{jets} – the number of jets with $p_T > 40 \text{ GeV}$.
- $N_{b\text{-jets}}$ – the number of b -jets with $p_T > 20 \text{ GeV}$.
- p_T^{miss} – the magnitude of the missing transverse momentum, as defined in section 1.4.4.
- m_T – the transverse mass, defined to be $\sqrt{2 \cdot p_T \cdot p_T^{\text{miss}} \cdot (1 - \cos \Delta\phi)}$, where p_T is the transverse momentum of the leading lepton and $\Delta\phi$ is the difference in ϕ between the lepton and the missing transverse momentum vector.
- m_{eff} – the ‘effective mass’, defined to be the scalar sum of missing transverse momentum and all signal lepton and jet p_T ’s. $m_{\text{eff}} = p_T^{\text{miss}} + \sum_l p_T^l + \sum_{\text{jet}, p_T > 40 \text{ GeV}} p_T^{\text{jet}}$.
- **Z veto** – remove events whose same-flavour opposite-charge lepton invariant mass, m_{ll} , satisfies $84 < m_{ll} < 98 \text{ GeV}$.

Category	Data stream	Trigger requirements
High p_T^{miss}	JetTauEtmiss	<code>EF_xe80T_tclcw_loose</code> for data-taking periods A-B5, or <code>EF_xe80_tclcw_loose</code> otherwise. Require <code>LoCHadTopo</code> (LHT) $p_T^{\text{miss}} > 125$ GeV.
Dielectron	Egamma	If leading electron $p_T > 65$ GeV, require single electron trigger <code>EF_e60_medium1</code> (plateau efficiency $> 95\%$). Else check the dielectron trigger <code>EF_2e12_Tvh_loose1</code> with baseline lepton p_T cuts (plateau efficiency 97%).
Electron-muon	Egamma	If leading lepton is an electron with $p_T > 65$ GeV, require single electron trigger <code>EF_e60_medium1</code> (efficiency $> 95\%$). Else check the dileptonic trigger <code>EF_e12Tvh_medium1_mu8</code> with baseline lepton p_T cuts (plateau efficiency 95%).
Muon-electron	Muons	If leading lepton is a muon with $p_T > 36$ GeV, require single muon trigger <code>EF_mu36_tight</code> (plateau efficiency for $p_T > 40$ GeV is 75% [90%] in the barrel [end-cap]).
Dimuon	Muons	If leading muon has $p_T > 36$ GeV, require single muon trigger <code>EF_mu36_tight</code> . Else check the dimuon trigger <code>EF_mu18_tight_mu8_EFFS</code> . The <code>EF_mu18_tight</code> component plateaus for $p_T > 20$ GeV at 75% [90%] in the barrel [end-cap], and the <code>EF_mu8_EFFS</code> is fully efficient for $p_T > 10$ GeV.

Table 5.1: Summary of the trigger requirements placed on events from the different data streams. Each requirement is tried in the order presented, and if failed the next is attempted. If all fail then the event is rejected. The data stream indicates the requirement placed on only events from data; all the possible categories are tried for MC events.

SR	N_{leps}	$N_{b\text{-jets}}$	N_{jets}	$p_{\text{T}}^{\text{miss}}/\text{GeV}$	Other	$m_{\text{eff}}/\text{GeV}$
SR3b	≥ 2	≥ 3	≥ 5	–	–	> 350
SR0b	$= 2$	$= 0$	≥ 3	> 150	$m_{\text{T}} > 100 \text{ GeV}$	> 400
SR1b	$= 2$	≥ 1	≥ 3	> 150	$m_{\text{T}} > 100 \text{ GeV}$	> 700
SR3Llow	≥ 3	–	≥ 4	$\in [50, 150]$	Z veto	> 400
SR3Lhigh	≥ 3	–	≥ 4	> 150	–	> 400

Table 5.2: Definition of the signal regions in terms of the quantities defined in the text. In order to remove overlap between SR3b and other regions, events are checked against each set of requirements in turn, and accepted into whichever region whose requirements they meet first. The m_{eff} requirement is *only* applied when running the fit in ‘discovery’ mode. A ‘–’ indicates no requirement.

SR	m_{eff} bin limits/GeV			
	Bin 1	Bin 2	Bin 4	Bin 4
SR3b	[190, 845]	[845, ∞]		
SR0b	[300, 600]	[600, 900]	[900, 1200]	[1200, ∞]
SR1b	[300, 700]	[700, 1100]	[1100, ∞]	
SR3Llow	[255, 727.5]	[727.5, ∞]		
SR3Lhigh	[355, 1077.5]	[1077.5, ∞]		

Table 5.3: When running in ‘exclusion’ mode, signal regions are divided into two or more bins in m_{eff} , in place of the basic cut on m_{eff} described in Table 5.2. These binning are shown here

Five signal regions are created using these quantities, as shown in Table 5.2. As will be further discussed in context of the fitting and interpretation procedure in section 5.7, the analysis has two modes; an m_{eff} shape-fit ‘exclusion’ mode for placing limits on specifically simulated models, and a ‘discovery’ mode used for quantifying any generic excess in one or more of the signal regions. For the shape fit the signal regions are divided into bins of m_{eff} , these are shown in Table 5.3. The latter discovery mode modifies the signal regions with additional constraints on m_{eff} , as denoted in the table. These regions with additional cuts are sometimes referred to as *e.g.* SR3b_{disc}.

The SR0b region targets those SUSY models that do not have a tendency to produce b -quarks, which are largely those either directly producing first or second generation squarks, or those mediated by gluino production in the case where stops are sufficiently massive to not be a favoured decay mode of the gluino. Naturalness arguments tend to favour a lighter stop; In this case direct or gluino mediated stop production will lead to events containing at least two b -quarks. Since the tagging process is not fully efficient, SR1b is designed to pick up these events. Finally, a handful of the models discussed in section 5.9 are expected to produce at least four b -quarks, and these are targeted by SR3b.

In cases where there is a tendency for multi-step cascades to occur, such as in the gluino-mediated $\tilde{q} \rightarrow qWZ\tilde{\chi}_1^0$ scenario, either of both of leptons and jets could occur with large multiplicity. The latter of these motivates the moderately high jet multiplicity cuts present in many of the signal regions, whilst the former is the prime motivation behind the two three-lepton signal regions. These are split into both low and high p_T^{miss} variants, to target models producing, respectively, off- and on-shell W and Z bosons in cascades respectively. The background from Z boson production is suppressed by a Z veto in SR3Llow, and in SR3Lhigh is already reduced by the higher p_T^{miss} requirement.

5.5 Estimating the backgrounds

5.5.1 Prompt backgrounds

There are a number of processes contributing to the appearance of events in our signal regions. Those with real, or ‘prompt’ leptons (not occurring as the result of e.g. a mis-identified jet) turn out to not be readily susceptible to the use of data-driven methods; these are irreducible backgrounds. Instead, the expected contribution is estimated using those MC samples described in section 5.2.2. To re-iterate, the most significant processes contributing are $t\bar{t} + V$, diboson and triboson production processes, and single top with a Z boson.

In order to ensure that the modelling of the effective mass distribution is good here, three control regions are defined which aim to demonstrate good modelling for each of the major irreducible backgrounds; these are VRttW (for $t\bar{t} + W$), VRttZ (for $t\bar{t} + Z$) and VRVV (primarily for $WZ + \text{jets}$). These regions are defined with the same quantities introduced in section 5.4, with the addition of:

Region	N_{leps}	$N_{b\text{-jets}}$	N_{jets}	$p_T^{\text{miss}}/\text{GeV}$	Other
VRttW	$\mu^\pm\mu^\pm$	= 2	≥ 1 (30 GeV)	$\in [20, 120]$	$m_T > 80 \text{ GeV}$
VRttZ	≥ 3	$\in \{1, 2\}$	≥ 2 (40 GeV)	$\in [20, 120]$	$m_{\text{eff}} > 300 \text{ GeV}, Z \text{ accept}$
VRVV	$\mu^\pm\mu^\pm$	= 0	≥ 2 (20 GeV)	$\in [20, 120]$	$m_T > 100 \text{ GeV}$

Table 5.4: Definition of the validation regions in terms of the quantities defined in the text. To avoid ambiguity, note that $\mu^\pm\mu^\pm$ indicates a requirement for like-charge muon pairs.

- **Z accept** – keep *only* events whose same-flavour opposite-charge lepton invariant mass, m_{ll} , satisfies $83 < m_{ll} < 96 \text{ GeV}$.

The cuts placed for each of these are recorded in Table 5.4, and were optimised to ensure that each region was enriched in events from the targeted background process whilst remaining kinematically similar to the signal regions¹. It should also be noted that the p_T requirement on jets is varied between the validation regions, unlike in the signal regions where the lower limit was fixed at 40 GeV.

The distributions in effective mass for each of the validation regions are shown in Figure 5.1. Whilst each region is rich in the targeted process, the full background estimate is applied (including the data-driven methods discussed in the subsequent sections) so as to account from contamination from the other sources of background. In all cases good agreement is observed between data and MC, giving us good confidence in the validity of our estimates.

It has also been verified that the contamination in the validation from signal models near exclusion, but not excluded is small. Had this not been the case, it could have lead to an overestimate of the backgrounds, perhaps nullifying our ability to see the same new physics in the signal regions. VRttZ and VRVV both have negligible contamination, and VRttW has contamination of at most 25%.

Since the modelling of b -jets is suspected to be one potential weakness of both the MC generators and detector simulation procedure, an additional check is carried out to verify our confidence in the modelling of the background in SR3b. Since there are few events with three b -jets, a validation region is designed requiring events with opposite-sign

¹Checks were performed to ensure minimal contamination from SUSY signal models near the exclusion threshold.

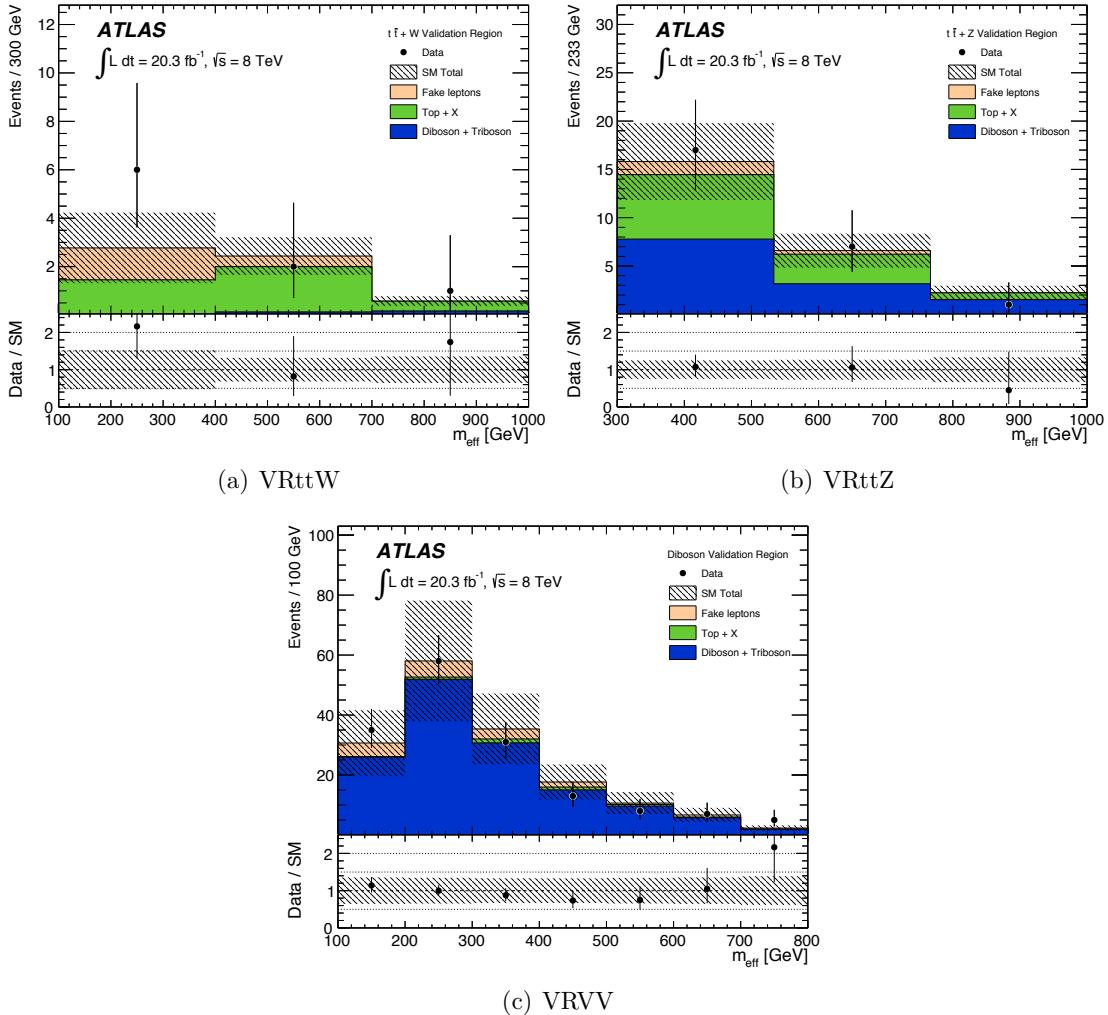


Figure 5.1: Comparison of data with the full background estimate for the effective mass (m_{eff}) distribution in each of the three validation regions, as labelled. The statistical and systematic uncertainties on the background prediction are included in the uncertainty band. The last bin of each histogram includes the overflow, and the lower region shows the ratio of data to the total standard model background estimate.

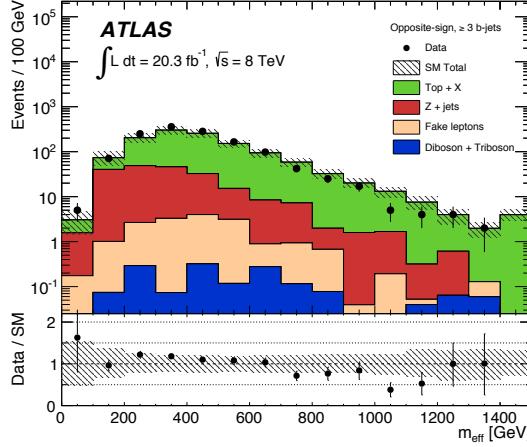


Figure 5.2: Effective mass (m_{eff}) distribution for a control region requiring opposite sign leptons in addition to three b -jets. The statistical and systematic uncertainties on the background prediction are included in the uncertainty band. The last bin includes overflows. The lower part of the figure shows the ratio of data to the total standard model background prediction.

lepton pairs, but otherwise the same selection described in section 5.3. Due to requiring opposite-sign lepton pairs, it is also necessary to include a $Z + \text{jets}$ sample, modelled with `Sherpa`. The comparison can be seen in Figure 5.2, and demonstrates that at least in this region agreement is mostly good, and within uncertainties. From the ratio plot a small bias can be seen to overestimate at high m_{eff} and underestimate at low m_{eff} , however this was not deemed significant enough to be a cause for concern.

5.5.2 Fake lepton contributions

As has been discussed to some length in chapter 4, events requiring one or more leptons are prone to being faked by events containing non-prompt leptons or mis-reconstructed jets. In the older versions of the analysis [112] it was found that the primary sources of these fakes were non-isolated leptons produced in the decays of B hadrons, as well as conversions in the inner detector.

The methods therein, specifically those developed by the author in section 4.4, were applied to this analysis. A crucial part of forming the estimate is the measurement of the real and fake efficiencies; this was performed by other members of the analysis team, but is documented herein.

Isolation choice

In order to use the matrix method, it is required to discriminate between tight and loose leptons; such a distinction has already been made between baseline and signal leptons, i.e. loose-or-tight and tight. The difference between the two is largely the choice of isolation variables, as has already been discussed in section 1.4.1.

Fake efficiency computation

As defined in section 4.2, the fake efficiency is $\varepsilon_f = P(t|f\tilde{l})$. In order to estimate this quantity, one firstly wants to use control regions as kinematically close to the signal regions as possible, to ensure that the efficiencies derived are applicable there. The rest of the procedure follows from the definition; one wants to pick a control region with a set of leptons that are almost surely fake, and then approximate the *probability* of a given lepton passing the tight requirements as the *fraction* of those leptons that do. In practice there will be a contamination from real leptons in this sample, and as such MC is used to estimate this component and subtract it before taking the ratio. That is:

$$\varepsilon_f \approx \frac{n_T^{\text{Data}} - n_{TR}^{\text{MC}}}{n_T^{\text{Data}} - n_R^{\text{MC}}}, \quad (5.1)$$

where $n = n_T + n_L$, n_T^{Data} represents the number of events measured in data in the control region that are tight, and n_{TR}^{MC} are those events from MC in the control region that are both tight and known to be real. Fake efficiencies are measured separately for electrons and muons, using different control regions for each. For each of these they are then subdivided into categories distinguished by p_T and $|\eta|$ ranges.

Electron fake efficiency: The control region used searches for like-sign μe pairs, using the muon as a real ‘tag’ object. To make it likely that the muon is real, it is required to pass the tight selection and have $p_T > 40$ GeV.² This is, of course, an assumption, and as such an uncertainty is associated with it, to be described forthwith. The electron is then used as a ‘probe’ with which to measure the fake efficiency, since it is more likely than not to be a fake; this is because processes that will produce like-sign $e\mu$ events have very low cross section by comparison.

²The selection procedure from section 5.3 is followed up to the point of the analysis-specific “two leptons” cut - from this point forward the selection for computing both the fake and real efficiencies diverges from that described in that section.

Region	ε_f	Statistical uncertainty			
		Total	Data	MC subt.	Syst.
$p_T \in [15,20) \text{ GeV}, \eta \leq 1.5$	0.055	0.015	0.015	0.002	0.022
$p_T \in [15,20) \text{ GeV}, 1.5 < \eta \leq 2.5$	0.059	0.024	0.024	0.002	0.032
$p_T \in [20,25) \text{ GeV}, \eta \leq 1.5$	0.101	0.028	0.028	0.003	0.049
$p_T \in [20,25) \text{ GeV}, 1.5 < \eta \leq 2.5$	0.083	0.045	0.045	0.004	0.036
$p_T \in [25,35) \text{ GeV}, \eta \leq 1.5$	0.035	0.018	0.018	0.001	0.019
$p_T \in [25,35) \text{ GeV}, 1.5 < \eta \leq 2.5$	0.045	0.040	0.040	0.005	0.039
$p_T \in [35,45) \text{ GeV}$	0.080	0.037	0.033	0.004	0.035
$p_T \in [45,65) \text{ GeV}$	0.043	0.042	0.042	0.004	0.042
$p_T > 65 \text{ GeV}$	0.055	0.072	0.072	0.009	0.092

Table 5.5: Measured electron fake efficiencies (ε_f) including statistical and systematic uncertainties, in the presence of at least one b -jet. ‘MC subt.’ refers to the systematic associated with the MC subtraction procedure.

Since it was noted that a particularly prevalent source of fakes are from B hadron decays, the most likely *true* fake efficiency for a given electron will likely vary as a function of heavy flavour activity in the event. Due to the fact that signal regions place varying requirements on b -jet multiplicity it is therefore prudent to extract the efficiencies in two scenarios – with a b -jet veto, and requiring at least one b -jet.³ In each of these scenarios the efficiencies are binned in p_T , and for $p_T < 35 \text{ GeV}$ also in η ; for higher p_T there are insufficient statistics to perform this split, causing the uncertainties to become too large. The efficiencies as well as a breakdown of the uncertainties are shown in the b -jet scenario (Table 5.5) and the scenario vetoing b -jets (Table 5.6).

Muon fake efficiency: A like-sign muon control region is used, additionally requiring two jets of $p_T > 25 \text{ GeV}$. It has been verified that the muon fake efficiencies are not strongly dependent on whether or not these jets are required to also be b -jets. As for the electron case, the hardest muon is used as a tag object and is required to have $p_T > 40 \text{ GeV}$; the precise method employed then depends on the p_T of the softer probe muon. For $p_T < 40 \text{ GeV}$ the same method is employed as for electrons, except that statistics permit only binning in p_T and not η . The results in this case are

³There were found to be insufficient statistics to further separate this latter region into bins of b -jet multiplicity. A correction factor is noted in due course.

Region	ε_f	Statistical uncertainty			
		Total	Data	MC subt.	Syst.
$p_T \in [15,20) \text{ GeV}, \eta \leq 1.5$	0.035	0.006	0.006	0.001	0.006
$p_T \in [15,20) \text{ GeV}, 1.5 < \eta \leq 2.5$	0.055	0.008	0.008	0.001	0.017
$p_T \in [20,25) \text{ GeV}, \eta \leq 1.5$	0.052	0.010	0.010	0.002	0.022
$p_T \in [20,25) \text{ GeV}, 1.5 < \eta \leq 2.5$	0.075	0.013	0.013	0.002	0.059
$p_T \in [25,35) \text{ GeV}, \eta \leq 1.5$	0.032	0.009	0.009	0.002	0.020
$p_T \in [25,35) \text{ GeV}, 1.5 < \eta \leq 2.5$	0.070	0.013	0.013	0.002	0.039
$p_T \in [35,45) \text{ GeV}$	0.100	0.014	0.014	0.002	0.061
$p_T \in [45,65) \text{ GeV}$	0.107	0.019	0.019	0.004	0.070
$p_T > 65 \text{ GeV}$	0.131	0.028	0.028	0.006	0.085

Table 5.6: Measured electron fake efficiencies (ε_f) including statistical and systematic uncertainties, requiring there be no b -jets in the events. ‘MC subt.’ refers to the systematic associated with the MC subtraction procedure.

Region	ε_f	Statistical uncertainty			
		Total	Data	MC subt.	Syst.
$p_T \in [15,20) \text{ GeV}$	0.107	0.019	0.019	0.003	0.042
$p_T \in [20,25) \text{ GeV}$	0.087	0.032	0.031	0.006	0.064
$p_T \in [25,40) \text{ GeV}$	0.128	0.051	0.050	0.011	0.148

Table 5.7: Measured muon fake efficiencies (ε_f) including statistical and systematic uncertainties. ‘MC subt.’ refers to the systematic associated with the MC subtraction procedure.

shown in Table 5.7. For the probe having $p_T > 40 \text{ GeV}$, the fake efficiency from the $p_T \in [25,40] \text{ GeV}$ bin is used, but scaled according to a factor derived from a $t\bar{t}$ MC sample.

Correction for SR3b: As noted previously, for neither the electron nor muon efficiency calculations were there sufficient statistics to directly extract a fake efficiency directly applicable to a signal region containing three b -jets. In order to circumvent this limitation, a correction factor for the final fake *rate* is derived following measuring the number of real and fake events in several $t\bar{t}$ MC samples. In order to increase the statistical power,

Generator	Electrons, f_e	Muons, f_μ
Powheg +Pythia	1.24 ± 0.40	1.07 ± 0.68
Powheg +JIMMY	1.31 ± 0.47	1.65 ± 1.17
MC@NLO +JIMMY	1.13 ± 0.66	0.62 ± 1.08
Alpgen +JIMMY	1.39 ± 0.63	0.68 ± 1.20

Table 5.8: Correction factors for events with at least one b -jet to those with at least three b -jets for various $t\bar{t}$ MC samples for both electrons and muons. Due to limited statistics any former divisions on p_T and $|\eta|$ have been removed, and as such these would be applied to the combined efficiency in those regions. Only the statistical uncertainties are included.

samples produced with MC@NLO, Alpgen and Powheg were combined after checking that they had consistent behaviour; these factors are shown in Table 5.8.

The scaling factors, f_e and f_μ , are applied directly to the expected fake component derived for events in the dielectron and dimuon channel. When the leptons are of different flavour, a combined weight $f_{e\mu}$ is used, formed by taking a weighted average of f_e and f_μ according to the number of events with either fake electrons or fake muons events expected in the MC sample, denoted n_e and n_μ respectively:

$$f_{e\mu} = \frac{f_e n_e + f_\mu n_\mu}{n_e + n_\mu}. \quad (5.2)$$

After the samples were combined the overall correction factors were $f_e = 1.27 \pm 0.25$, $f_{e\mu} = 1.24 \pm 0.31$ and $f_\mu = 1.16 \pm 0.51$. These were applied during the fitting procedure described in section 5.7.

Uncertainties: In addition to the natural uncertainties arising above due to statistical limitations in both data and MC in the control regions, three systematic effects are considered which introduce additional contributions to the overall uncertainty on the fake efficiencies:

- **Kinematic dependence of ϵ_f :** Since the control regions have different kinematics to the signal regions, there is no guarantee that the true fake efficiencies remain the same when extrapolated. A study is made varying the kinematic requirements in the control region, and the maximum difference observed is symmetrised and taken as the uncertainty; it is found to be 33.7%.

- **MC subtraction:** The cross sections of the diboson and $t\bar{t} + V$ MC samples used for subtracting unwanted real components are not known exactly; conservative uncertainties of $\pm 30\%$ and $\pm 50\%$ are used for each of these samples respectively.
- **Methodological assumptions:** The methods assume that the tag object is real, which is not always the case. By looking at the number of dilepton events with the higher p_T lepton loose and the other tight, compared to the converse isolation, this assumption is estimated to be wrong only 2% (3%) of the time for electrons (muons), and is hence negligible compared to the other uncertainties.

Real efficiency computation

The real efficiency is defined to be $\varepsilon_r = P(t|r\tilde{l})$, therefore contrary to the fake efficiency computation one wishes to use a sample enriched in real leptons. This is achieved by imposing cuts to select $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events. As before for the fake efficiencies, the basic event selection procedure is followed for the cleaning cuts, at which point two opposite-sign same-flavour leptons are required passing the loose isolation cuts. The dilepton invariant mass m_{ll} is then required to satisfy $80 < m_{ll} < 100$ GeV. Since by and large both leptons are expected to be real, each lepton is considered as a possible tag, and the other as a probe, lowering the probability of an unintended systematic bias arising.

Figure 5.3 shows the invariant mass distribution in these control regions, demonstrating that most of the baseline leptons are also real. As noted with regards to the FSR photons in the caption, this is not always the case; it is, however, a fairly small effect, and no specific correction is applied. Instead, the real efficiency is computed in both data and MC and the difference between the two is taken as a systematic uncertainty. Whilst conservative, in practice the uncertainty on the final fake rate prediction is dominated by the uncertainty on the fake efficiencies rather than the real efficiencies.

The real efficiencies are then extracted as a function of both p_T and η , displayed in Figure 5.4. Since these are extracted in Z control region with limited hadronic activity, there is potentially an issue in applying them to the signal regions which all require several jets. A systematic uncertainty was introduced to take this into account after studies in the variation of real efficiency with jet multiplicity and m_{eff} in $t\bar{t}$ MC. The results of this study were that a uniform uncertainty of 3% was applied across all measurements.

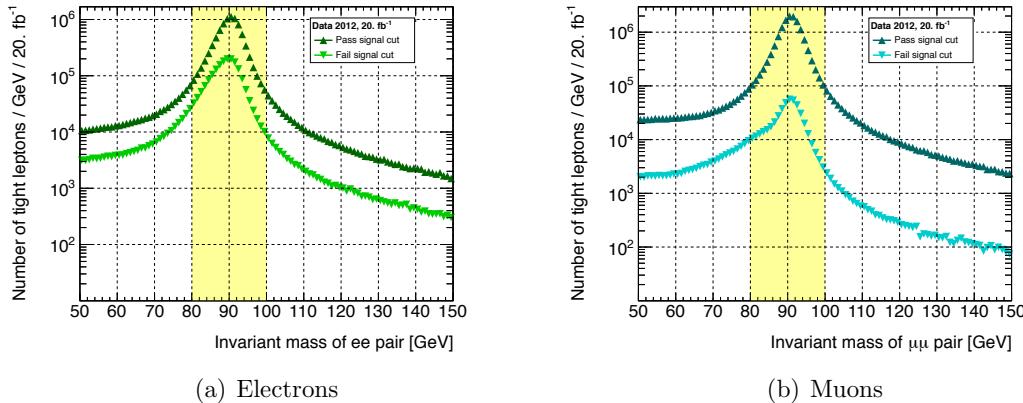


Figure 5.3: Invariant mass distribution of two opposite-sign muons and electrons where at least one *tag* lepton passes the signal requirements. The data are then split between the scenarios where the other *probe* lepton passes or fails the signal requirements. The shoulder in the looser selection for muons was identified as being caused by fake muons originating from soft FSR photons – these were then removed by the calorimeter isolation cut in the tight requirement.

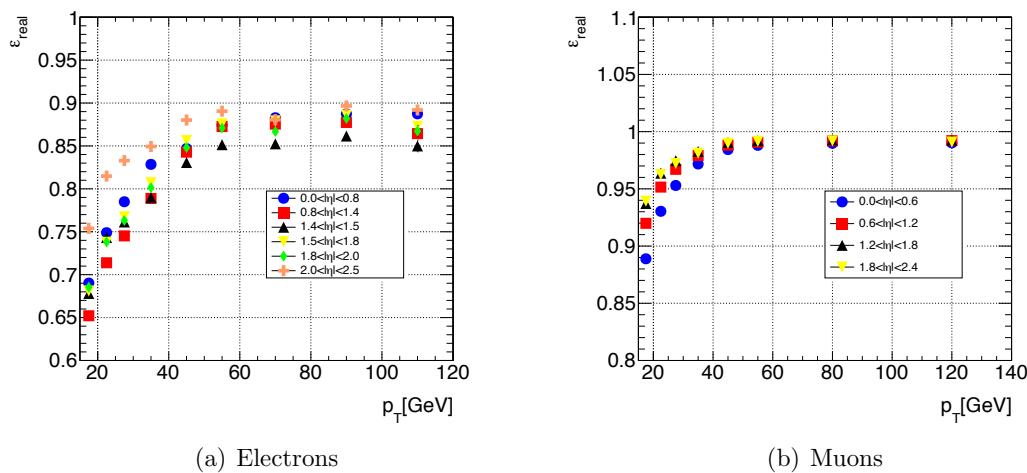


Figure 5.4: Real efficiencies measured in data for electrons and muons in bins of p_T and η .

Forming the background estimate

Having measured the efficiencies and estimated the uncertainties on them, it only remains to define the procedure by which an estimate of the expected number of events due to fakes in a given signal region is formed from the observed data.

The first example of section 4.4.3 demonstrates use of the generalised matrix method in exactly the case required here; the selection can require either exactly two like sign leptons, or three leptons with no requirement on charge. Efficiencies have been measured across a large number of categories with uncertainties split into systematic components that are correlated between the different categories (for example the uncertainties in the MC sample cross sections), as well as uncorrelated statistical components; hence the uncertainty in the final estimate is found by following the steps in section 4.4.2.

5.5.3 Charge-flipped like-sign lepton events

Another background that also arises due to detector effects is that of charge-flip; that is, an event with like-charge leptons is observed when in fact an opposite-charge event was produced in the hard scatter.⁴ This effect is significant for electrons, but negligible for muons. Most commonly this is caused by so-called “trident electrons”, where one electron in a dielectron event undergoes Bremsstrahlung in the inner detector, radiating an off-shell photon. This photon subsequently decays into an electron and positron: depending on the relative p_T of each object, it is possible that only the positron track will be reconstructed, which will of course be determined to have the opposite charge of the initial electron. The impact of this process is significantly reduced by the track isolation requirements imposed on electrons, however it is not removed entirely. Charge mis-identification, where a track is merely identified as having the opposite charge than its true value, is also possible though a smaller effect. Figure 5.5 demonstrates the presence of charge-flip in dielectron events by examining the Z peak – notably the structure seen in the same-sign dielectron invariant mass is not present in the dimuon invariant mass.

In order to estimate the expected contribution of this effect to the events in the signal region, a fully data-driven procedure is applied to reweight opposite-sign dielectron events otherwise kinematically identical to those that would have been recorded in the signal region. Each of these events is given a weight according to the flip probabilities for each

⁴Or vice versa, but in practice this effect is quite negligible due to the low cross section of like-charge events.

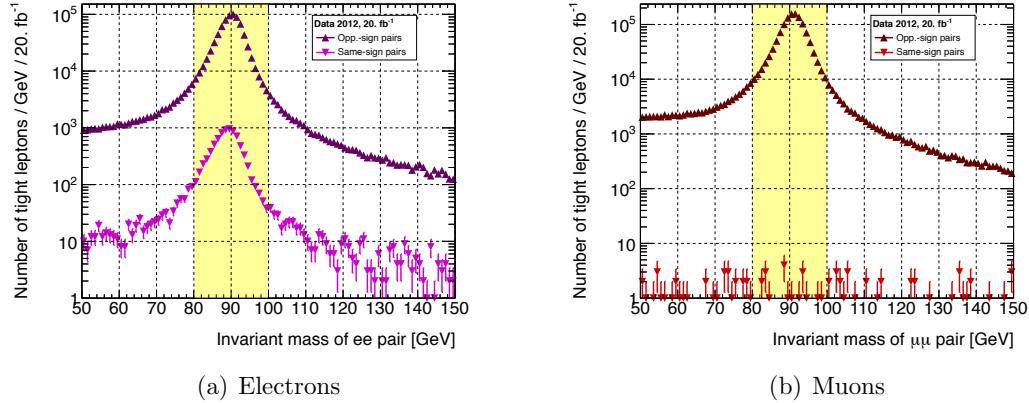


Figure 5.5: Inclusive opposite- and same-charge dilepton invariant mass spectrum for electron and muon pairs. In both cases all events passing the baseline selection are used where the ‘probe’ lepton has $|\eta| > 2.1$, since this selects events for which charge-flip is more prevalent. Note the presence of charge-flip in the electron sample, but not the muon sample.

of its constituent electrons. These probabilities are estimated using a likelihood-based method as described in the next section.

Estimation of flip probability

Let us define the probability of an electron of category i flipping to be ζ_i . In this case categories will comprise bins in p_T and η . Pseudorapidity is particularly important since trident electron formation depends strongly on the material present, which in turn varies significantly with η . Figure 5.6 demonstrates how the electrons in opposite- and same-sign electron pair events are distributed in p_T and η . This demonstrates that higher p_T causes a slight increase in flip probability, and more significantly that there is a large increase towards large $|\eta|$ (as one might expect since the electrons are forced to traverse more material).

If one assumes that each of these categories is independent, and also that all same-sign events arise from charge-flip in the given control region, then one can write an expression for $\langle n_{ij}^{SS} \rangle = \mathbb{E} [n_{ij}^{SS} | n_{ij}, \zeta_i, \zeta_j]$, the expected number of same-sign events with leptons in categories i and j respectively, in terms of n_{ij} , the total number of ‘actual’ opposite-sign

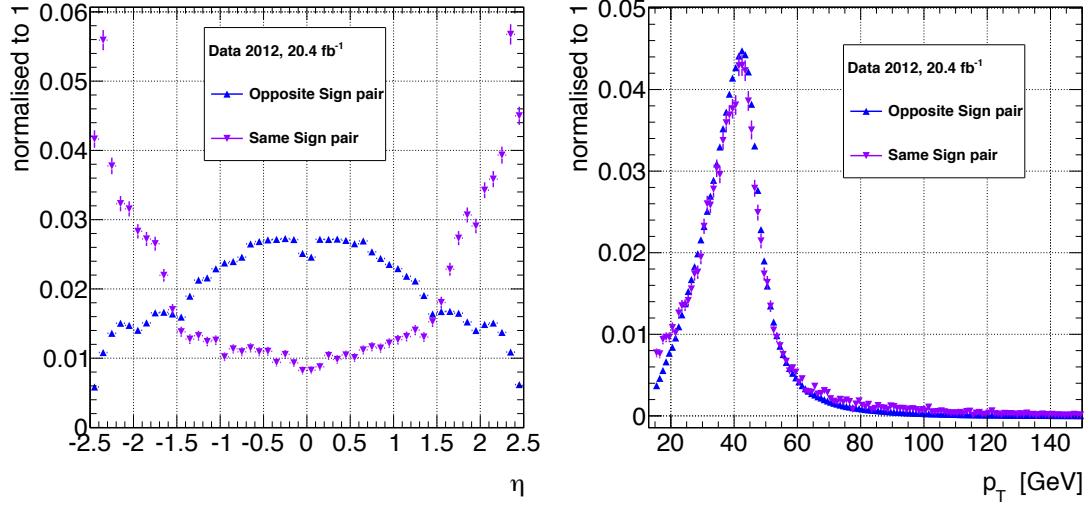


Figure 5.6: Inclusive lepton p_T and η spectra found for opposite- and same-charge dielectron events.

events, is

$$\langle n_{ij}^{SS} \rangle = n_{ij} (\zeta_i (1 - \zeta_j) + (1 - \zeta_i) \zeta_j) \quad (5.3)$$

$$\approx n_{ij} (\zeta_i + \zeta_j). \quad (5.4)$$

The approximation follows in the case that all ζ are small; this is shown to be the case in the scenarios of interest to this analysis.

One further assumes that the act of an electron flipping charge is a Bernoulli trial, such that for a given number of events in an dielectron control region, the number of events in its same-sign subset will be governed by a Bernoulli process; i.e. it will follow a binomial distribution. Since both the flip probabilities are small, and the number of events in each bin large, the Poisson limit theorem can be applied [162]. As such the full probability distribution is

$$P(n_{ij}^{SS} | \lambda_{ij}) \approx \frac{\lambda_{ij}^{n_{ij}^{SS}} e^{-\lambda_{ij}}}{n_{ij}^{SS}!}, \quad (5.5)$$

where $\lambda_{ij} = \langle n_{ij}^{SS} \rangle = n_{ij} (\zeta_i + \zeta_j)$. Therefore the joint probability across all event permutations from m lepton categories is

$$\mathcal{L}(\zeta_1, \dots, \zeta_m | \dots) = P(n_{11}^{SS}, n_{12}^{SS}, \dots, n_{mm}^{SS} | \zeta_1, \dots, \zeta_m) = \prod_{i=1}^m \prod_{j=1}^m P(n_{ij}^{SS} | \lambda_{ij}) \quad (5.6)$$

$$\Rightarrow -\ln \mathcal{L} = -\sum_i \sum_j \ln P(n_{ij}^{SS} | \lambda_{ij}) \quad (5.7)$$

$$= \sum_i \sum_j \lambda_{ij} - n_{ij}^{SS} \lambda_{ij} + \text{constants} \quad (5.8)$$

$$= \sum_i \sum_j n_{ij} (\zeta_i + \zeta_j) (1 - n_{ij}^{SS}) + \text{constants}. \quad (5.9)$$

For a given set of observations of $\{n^{SS}\}$ and n , this negative log likelihood function can be minimised (neglecting the constant terms) to give the MLE values of ζ_1, \dots, ζ_m .

A Z boson control region is chosen as a region dominated by opposite-sign dilepton events, selected by requiring the invariant mass of the two leptons to be in the range $75 < m_{ll} < 100$ GeV. The asymmetric window is chosen because, due to the trident process described above, charge-flip leptons have a smeared p_T with respect to the electrons that induced them. This is found to result in the Z peak in the same-sign channel being shifted to lower invariant mass.

The measurement is performed for both tight and loose leptons separately⁵; in the latter case the contamination of fake leptons in the same-sign Z peak is particularly significant. In both cases the fake component is estimated in 25 GeV sidebands above and below the Z peak; an MC estimate of the Z production background is subtracted from the observed data, yielding an estimate of the fake component. The average of the fake component in the upper and lower bands is taken to be the contribution within the Z peak itself, since it is not expected to have significant m_{ll} dependence.

Uncertainties: In the procedures described above several assumptions and approximations are made; as usual these are accounted for in several uncertainties in the estimated values of the flip probabilities:

- **Statistical uncertainty:** the data sample in the Z peak has limited statistics; inherent uncertainty then follows in any quantity derived from event counts therein, which are propagated in the usual way.

⁵It shall later be seen to be necessary to know the flip probabilities for loose electrons for the purposes of removing overlap with the fake lepton estimate.

Electron type	Weighted OS	SS	Difference / %
tight	11315.33 ± 45.7	11473	1.4 ± 0.4
loose	64265.76 ± 181.2	67703	5.1 ± 0.3

Table 5.9: Results of the charge-flip probability closure test. The weighted opposite-sign event multiplicity in the Z peak, weighted using the flip probabilities as described in the text, is compared to the observed same-sign multiplicity. The percentage difference between these is shown in the final column; this is taken as the systematic uncertainty.

- **Background subtraction:** a very conservative estimate is given to the subtraction of the estimated fake component, taken to be the symmetrised absolute difference in flip probability obtained with and without the subtraction procedure. It is computed independently in each of the p_T and η bins, and found to be $\approx 90\%$ for low p_T and η , and only a few % at higher values.
- **Closure:** A closure test was performed in order to demonstrate how well the measured flip probabilities reproduced the same-sign Z peak – the percentage normalisation difference between opposite-sign events weighted with a factor ω_{ij} and the observed same-sign events was then used as a systematic uncertainty. The weight ω_{ij} is defined to be

$$\omega_{ij} = \frac{\zeta_i + \zeta_j}{(1 - \zeta_i)(1 - \zeta_j)}, \quad (5.10)$$

and hence varies depending on the categories i and j of each lepton in the event (determined by their p_T and η). This expression is derived later and shown in equation (5.13). The procedure was repeated for both tight and loose electrons, as summarised in Table 5.9.

Results: The results of the flip probability measurements are shown for tight and loose leptons in Figure 5.7, along with the corresponding systematic uncertainties in Figure 5.8; all flip probabilities are in the range of 10^{-5} to 0.05 (0.07) for tight (loose) electrons.

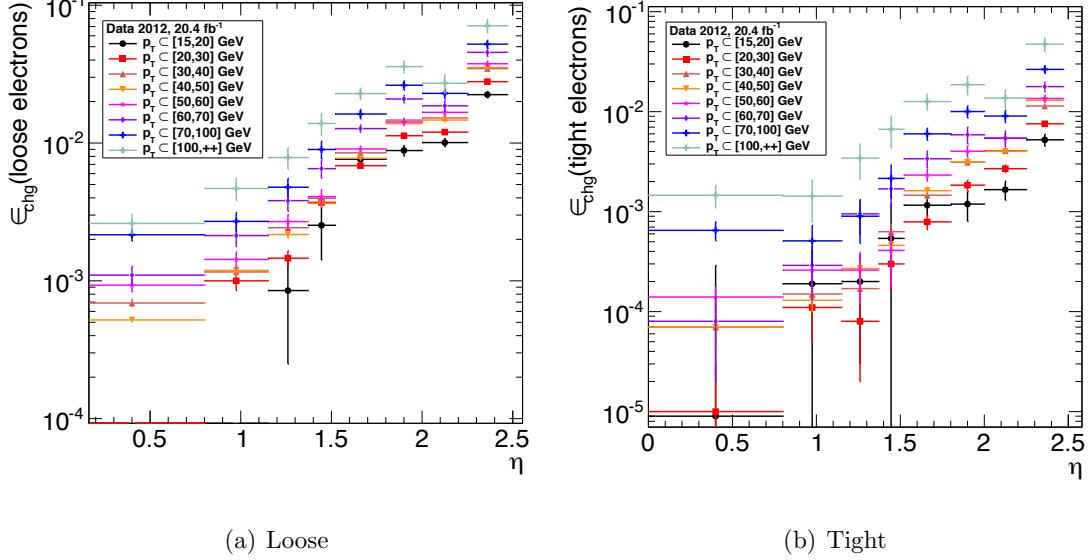


Figure 5.7: Measured charge flip probabilities for loose and tight electrons in bins of p_T and η . Uncertainties shown are statistical only; systematic uncertainties are shown in Figure 5.8.

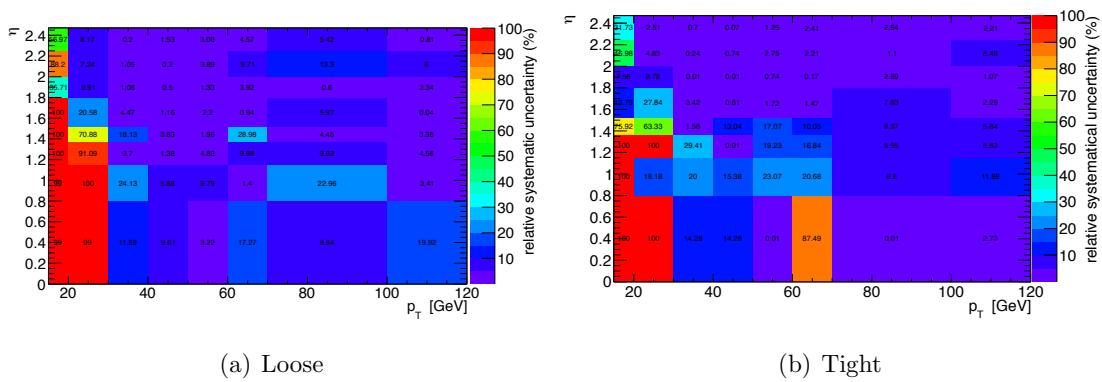


Figure 5.8: Combined relative systematic uncertainty on the measured charge flip probabilities for loose and tight electrons in bins of p_T and η .

Forming the background estimate

Earlier in equation (5.12) it was shown how the expected number of observed same-sign events would be related to the total number of opposite-sign events produced, assuming that charge-flip was the only mechanism by which same-sign events are produced. An analogous relation can be written down for the number of observed opposite-sign events (considering that either neither lepton must flip charge, or both must flip; this latter scenario can be neglected on grounds that all ζ are small, however).

$$\langle n_{ij}^{OS} \rangle = n_{ij} ((1 - \zeta_i)(1 - \zeta_j) + \zeta_i \zeta_j) \quad (5.11)$$

$$\approx n_{ij}(1 - \zeta_i)(1 - \zeta_j). \quad (5.12)$$

By taking a ratio of equations (5.4) and (5.12) the dependence on n_{ij} is eliminated, leaving the approximate dependence of $\langle n_{ij}^{SS} \rangle$ on $\langle n_{ij}^{OS} \rangle$ to be

$$\langle n_{ij}^{SS} \rangle = \omega_{ij} \langle n_{ij}^{OS} \rangle, \quad \omega_{ij} = \frac{\zeta_i + \zeta_j}{(1 - \zeta_i)(1 - \zeta_j)}. \quad (5.13)$$

Since we never have access to n_{ij} , equation (5.13) is ideal since we take $\langle n_{ij}^{OS} \rangle$ to be the number of opposite-sign events we *actually* saw. Therefore, in order to model the charge-flip component of an arbitrary distribution composed of same-sign events, the same distribution is formed from opposite-sign events, but with every event weighted by the factor ω_{ij} defined here.

Removing overlap with the fake lepton prediction

The fake estimate described in section 5.5.2 will have an overlap with the charge-flip background described in this section. This is because the set of observed opposite-sign events used in the derivation, $\langle n_{ij}^{OS} \rangle$, is partially formed of events with at least one fake lepton. Therefore, when reweighted to form $\langle n_{ij}^{SS} \rangle$, this prediction also includes the effect of this component.

In order to remove it, the matrix method procedure described in the previous section is used. Since the basic charge flip estimation in equation (5.13) can be written as an explicit sum of weights event by event,

$$\langle n_{ij}^{SS} \rangle = \sum_{\alpha \in \text{OS tight}} \omega_{ij}, \quad (5.14)$$

then a correction can be applied to remove the fake component. Here, α denotes any information included within each event – thus i and j in the equation above can be thought of as functions of α . Similarly, every event α (both loose and tight) have some weight under the matrix method, which we shall denote \mathcal{T}_α^F . Since a prediction for the total fake and tight component in some region is formed by summing the weights from corresponding tight and loose events, the *corrected* expression for the charge flip estimate, $\langle n_{ij,\text{corr}}^{SS} \rangle$ is

$$\langle n_{ij,\text{corr}}^{SS} \rangle = \sum_{\alpha \in \text{OS tight}} \omega_{ij}(1 - \mathcal{T}_\alpha^F) - \sum_{\alpha \in \text{OS loose}} \omega_{ij}\mathcal{T}_\alpha^F \quad (5.15)$$

$$= \sum_{\alpha \in \text{OS tight}} \omega_{ij} - \sum_{\alpha \in \text{OS}} \omega_{ij}\mathcal{T}_\alpha^F. \quad (5.16)$$

The last term of the second line of this expression represents the correction term to the original estimate.

5.5.4 Validating the data-driven background estimates

It is necessary to validate that both of the data-driven estimates, together with the MC generators validated previously, describe the backgrounds well. This is particularly relevant since both of the data-driven methods involve the use of control regions kinematically different (by necessity) to the signal regions; as such it should be verified that similar extrapolations to validation regions (not expected to contain signal) yield background estimates consistent with the observed data.

Inclusive validation regions demonstrating the proficiency of the fake lepton and charge-flip background estimates are shown in figures 5.9, 5.10, and 5.11, which include regions comprising subsets of those events selected according to the procedure described in section 5.3. These are selected to probe a variety of distributions, and in every case good agreement between the measured data and the prediction is observed.

5.5.5 Cross-check of the fake b -jet contribution in SR3b with the matrix method

Our signal region SR3b, due to requiring at least three b -tagged jets, has a significant background from events with one mistagged b -jet (or a ‘fake’ b -jet). Typically these form

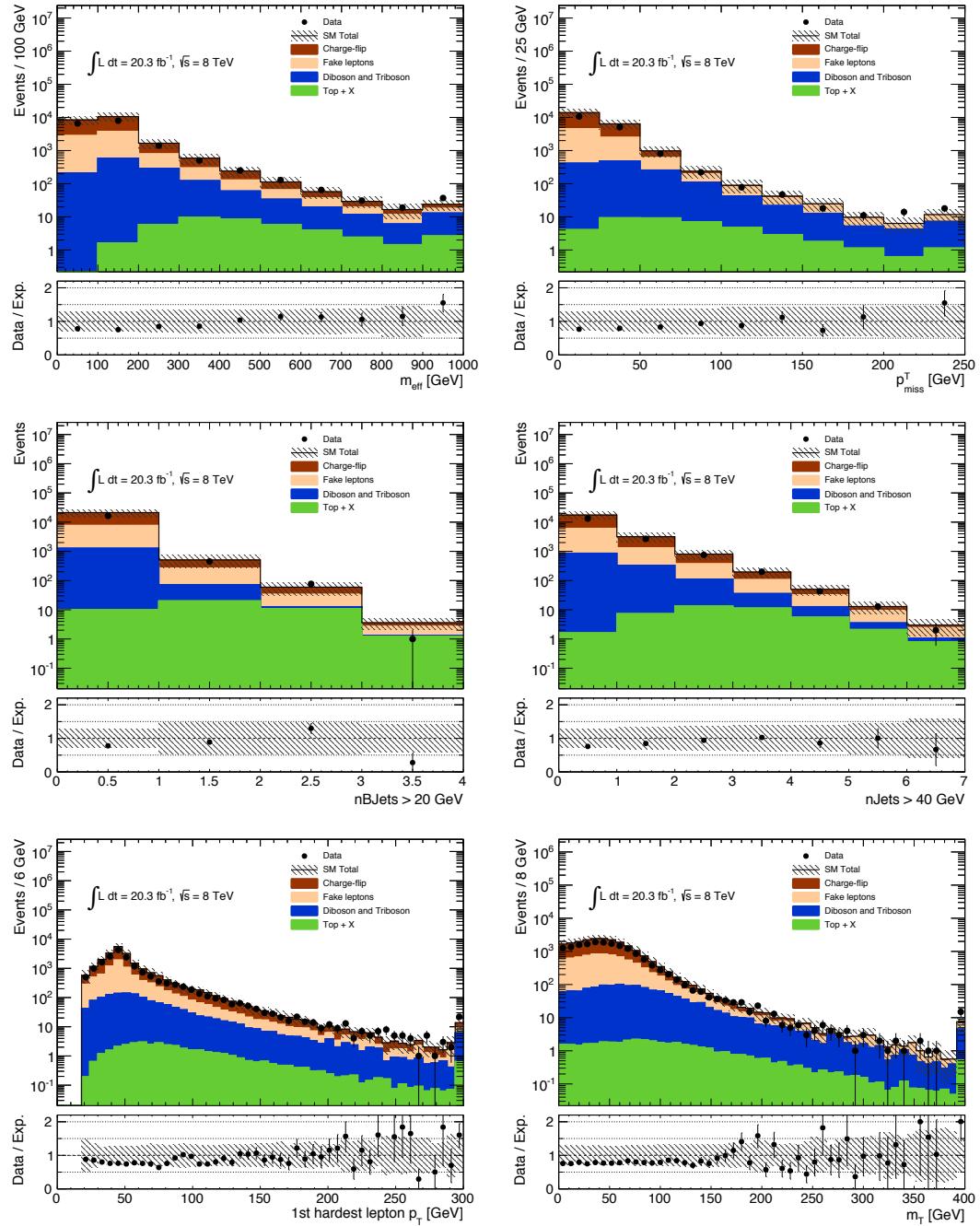


Figure 5.9: Distributions of kinematic variables in ee same-sign inclusive validation regions. The statistical and systematic uncertainties on the background prediction are included in the uncertainty band. The last bin includes overflows. The lower part of the figure shows the ratio of data to the total standard model background prediction.

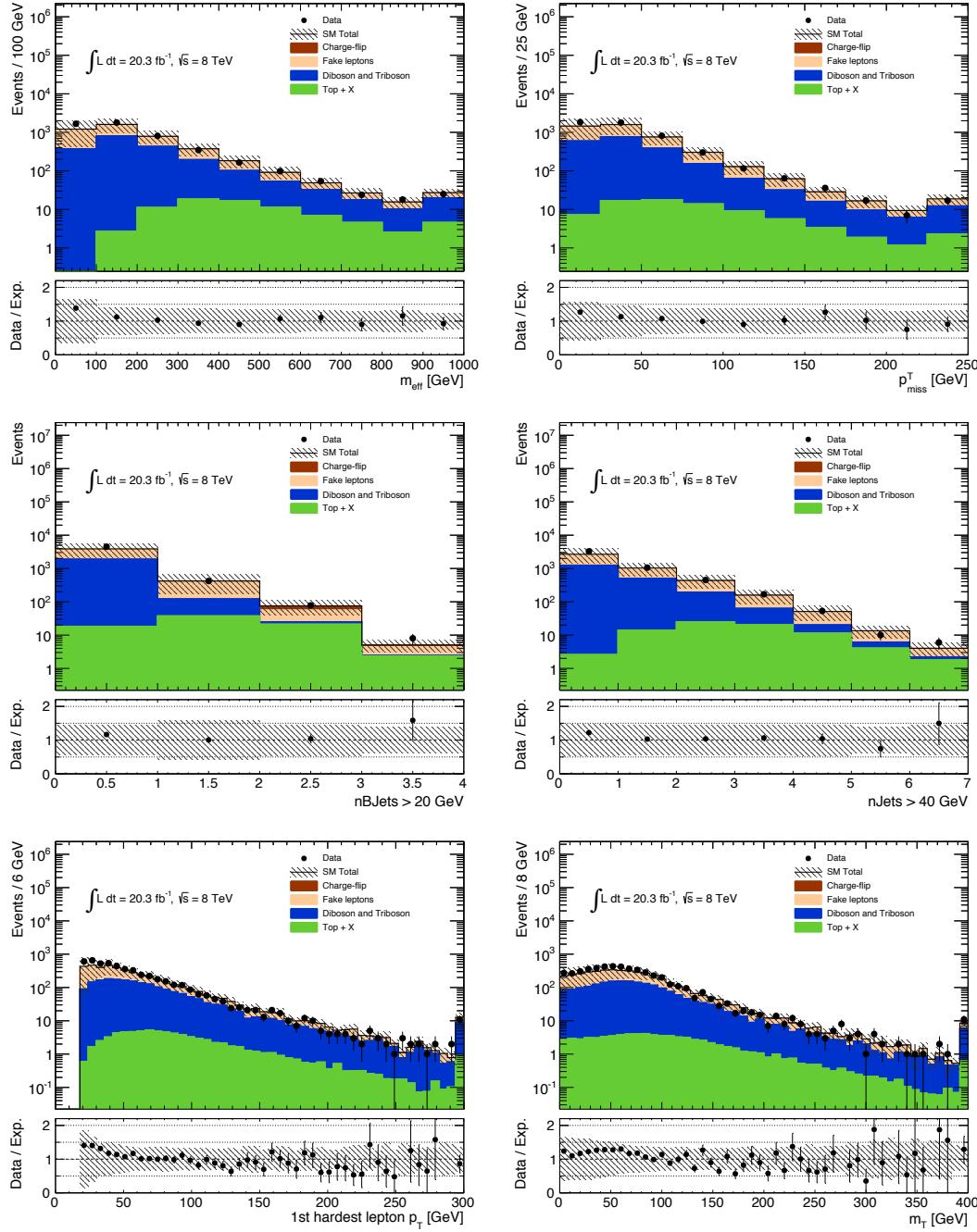


Figure 5.10: Distributions of kinematic variables in $e\mu$ same-sign inclusive validation regions. The statistical and systematic uncertainties on the background prediction are included in the uncertainty band. The last bin includes overflows. The lower part of the figure shows the ratio of data to the total standard model background prediction.

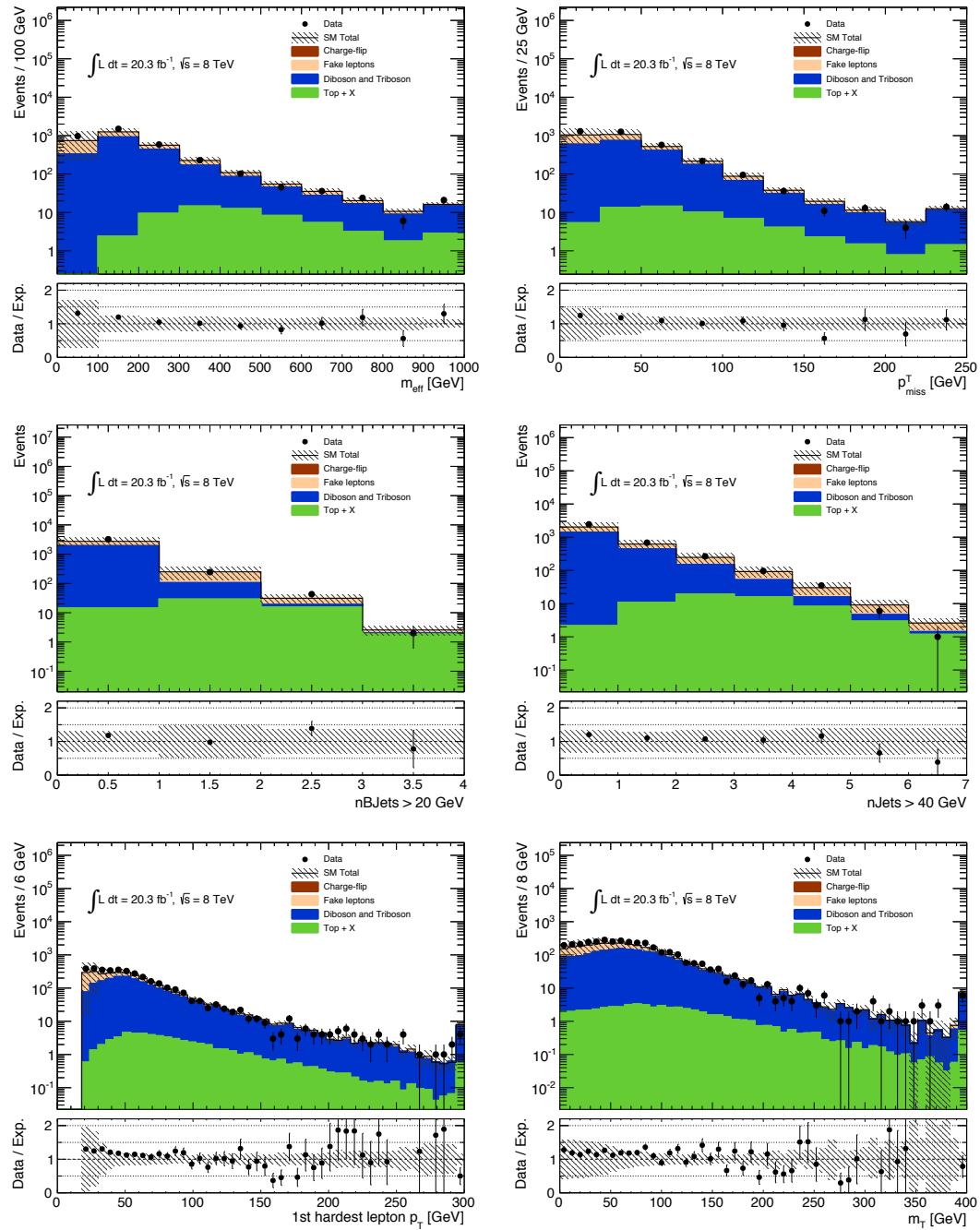


Figure 5.11: Distributions of kinematic variables in $\mu\mu$ same-sign inclusive validation regions. The statistical and systematic uncertainties on the background prediction are included in the uncertainty band. The last bin includes overflows. The lower part of the figure shows the ratio of data to the total standard model background prediction.

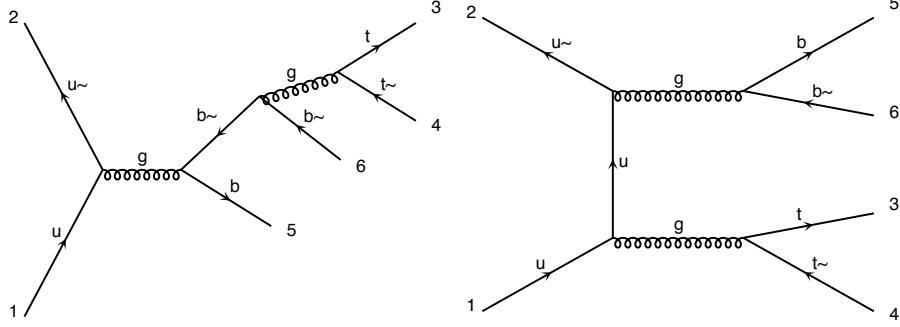


Figure 5.12: Two of the seven diagrams contributing to $t\bar{t}$ in association with $b\bar{b}$ production ($t\bar{t}b\bar{b}$). This process is responsible for producing events with three real b -jets. Figures produced by MadGraph [17].

from semileptonic $t\bar{t}$ events, where the decay products of the hadronically decaying top fake both a lepton and a b -jet. Alternatively processes such as $t\bar{t} + V$ can yield two real leptons in addition to one or more fake b -jets. The nominal background estimate for this analysis focusses on estimating the fake leptonic contribution using the data-driven matrix method, and trusting the detector simulation to model the fake b -jet component. In order to cross-check this method, a data-driven method has been applied to estimate the fake b -jet contribution directly.

The method described in this section will yield the background components which include only fake b -jets, as well as fake b -jets in addition to fake leptons. However, it will itself not predict the contribution from events with three real b -jets but one or more fake or charge-flipped leptons. Processes that can contribute to this include $t\bar{t}b\bar{b}$, as shown in Figure 5.12. Nominally it will produce three real b -jets and two opposite sign leptons, so either one top will decay hadronically and create a fake lepton, or both tops will decay leptonically with one of the resulting leptons being charge-flipped. Since neither the data-driven fake lepton or charge-flip estimation have been designed to be used *together* with the fake b -jet estimate, there would be double counting associate with using both together. As such, and since the background from $t\bar{t}b\bar{b}$ and others is expected to be very small, this component of fake & charge flipped leptons will be taken from MC for the purposes of this cross-check.

The b -jet matrix method

A previous ATLAS search for supersymmetry using three b -jets documented use of a matrix method for computing the contribution of fake b -jets in their signal regions [163]. Their method shares some similarities with the generalised matrix method as has been described in section 4.4. When compared to the procedure used to estimate the fake lepton contribution in this analysis, the main difference is the multiplicity of objects; whilst most events typically only had four or fewer leptons passing the looser set of isolation cuts, it is not unusual to find events with ten or more loose jets. This significantly increases the computational complexity of the problem, and means that the most thorough method of considering all of possible the tight/loose combinations for every object that could make for a tight event is infeasible. Instead, one simply computes the total weight for every event – this is sufficient in this case since none of the computed variables used in defining our signal regions are sensitive to the b -jet multiplicity (given of course that it is ≥ 3 , which is the case that the estimate is catering for), nor indeed upon *which* of the jets in an event are tagged as b -jets.

Computing efficiencies

As in the fake leptonic matrix method, the key measurement that needs to be performed is that of the efficiency for real and fake b -jets to pass a tighter set of cuts. In this case, ‘tight’ and ‘loose’ objects correspond to those whose values of the MV1 tagger output correspond to a b -tagged jet or not, respectively. These quantities will depend on the origin of the fake b -jet – that is, c -jets, τ -jets, light jets have different probabilities for passing a given MV1 operating point. Therefore, in order to compute the fake efficiency control regions are used that aim to emulate the relative contribution of these processes in the signal region as best as possible. The real efficiencies are taken from a central calibration effort, described in further detail below.

Moreover, it has already been described how both $t\bar{t}$ and $t\bar{t} + V$ processes can be responsible for contributions with one or more fake b -jets, possibly in association with a fake or charge-flipped lepton. Depending on the origin of the leptons, the origin of the fake b -jets will also be altered; for example a semi-leptonic decay mode of $t\bar{t}$ is most likely when considering events with a fake lepton, or indeed with two real leptons in $t\bar{t} + V$. Conversely, the fully leptonic $t\bar{t}$ decay mode would be responsible for events passing the selection with one of the leptons having been charge-flipped. Therefore it is prudent to

Index of b -jet	Opposite-sign	Same-sign	1 lepton
0	22%	23.3%	29%
1	28.6%	39.1%	31.8%
2	48.6%	37.4%	38.7%

Table 5.10: The probability that a b -tagged jet of given index, sorted in order of decreasing object p_T , is a fake. Studies are performed in Powheg $t\bar{t}$ and MadGraph $t\bar{t} + V$ MC for opposite-sign and same-sign dilepton, as well as single lepton control regions.

compute the fake b -jet efficiencies in both single and di-leptonic signal regions, although the single lepton control region is thought to be most appropriate for the same-sign signal region.

Methodology for fake efficiencies: Given a $t\bar{t}$ control region that requires *at least* two b -tagged jets, it is most likely that two of the total number of tags observed are real, and that the presence of additional b -jets indicates that one or more of them are mistags. A ‘tag and probe’ method is then used to compute the desired efficiencies as functions of p_T and η as described below.

Firstly consider the case where there are more than two b -jets in an event. If one is performing this procedure on an MC sample, then one has the luxury of perusing the truth record to determine *which* of the b -jets that have been observed are in fact fake. However, we wish to perform this procedure on data where this is not an option, forcing one to pick one of the b -jets as a probe. Although it is impossible to pick correctly with certainty, a study using the MC samples showed that softer b -jets were more likely to be fake. The results are summarised in Table 5.10.

When computing efficiencies in data the probe b -jet is selected at random with probability as shown in this table. For each kinematic bin in which efficiencies are measured, the number of these probe objects are counted and denoted $n_{t,\text{raw}}$. This number is then corrected for events with three real b -jets; a Powheg inclusive $t\bar{t}$ MC sample is processed, truth matching to find those events with ≥ 3 real b -jets in order to find an estimate of $n_{t,\text{real}}$, and is subtracted from the raw estimate. The number used in the final calculation is $n_t = n_{t,\text{raw}} - n_{t,\text{real}}$.

For any event, it is assumed that all real b -jets are tagged properly – that is, any jet that is not tagged is assumed to be fake and loose, in matrix method terminology. These

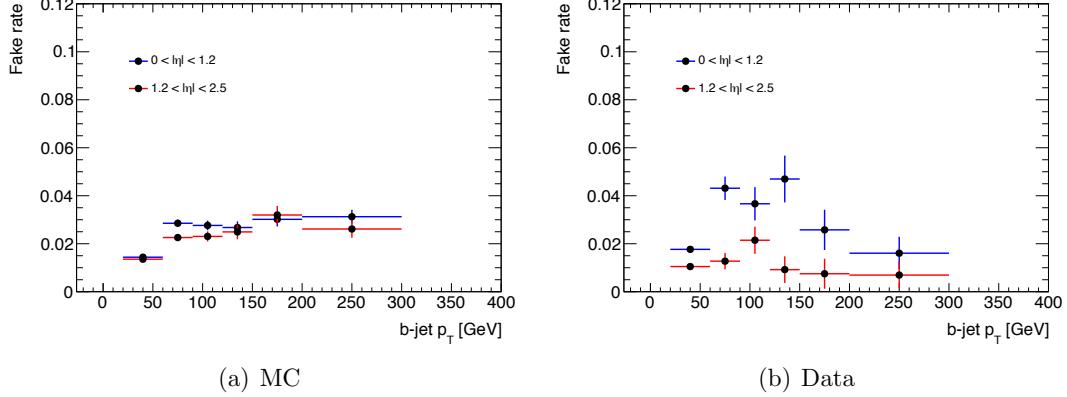


Figure 5.13: Fake efficiencies computed using both MC and data for an opposite-sign dilepton control region. The last p_T bin extends to infinity.

jets are counted, and for a given p_T and η bin there are n_l of them. The estimate of the efficiency for this bin is then $\varepsilon_f = n_t/n_l$. In total there are two bins in pseudorapidity, $0 \leq |\eta| < 1.2$ and $1.2 \leq |\eta| < 2.5$, each of which is divided into six p_T regions.

Since the fake efficiency is a primarily data driven source, with the real three b -jet subtraction constituting only a minor correction, the only relevant source of uncertainty is of a statistical nature. These uncertainties are computed, and propagated through the matrix method procedure.

Fake efficiency measurements: Two control regions are used to estimate the fake efficiencies, building upon the basic event selection described in section 5.3 with a requirement for $N_{\text{jets}} \geq 3$ and $N_{b\text{-jets}} \geq 2$. The first additionally requires an opposite-sign lepton pair (with leading lepton having $p_T > 20$ GeV, and the sub-leading with $p_T > 15$ GeV), the results for which are shown in Figure 5.13. The latter requires only a single lepton with $p_T > 20$ GeV, and additionally that there be at least four jets, and that the missing transverse energy satisfies $100 < p_T^{\text{miss}} < 200$ GeV. The efficiencies computed from this control regions are shown in Figure 5.14.

Each of these figures demonstrates efficiencies computed both with the algorithm described above, using data, and also an alternative computation directly with MC samples. Whilst agreement is reasonable, it is not necessarily expected to be perfect – if it were the background estimation process would not be needed! Uncertainties are quite large in the dilepton control region due to low statistics, however fortunately the single

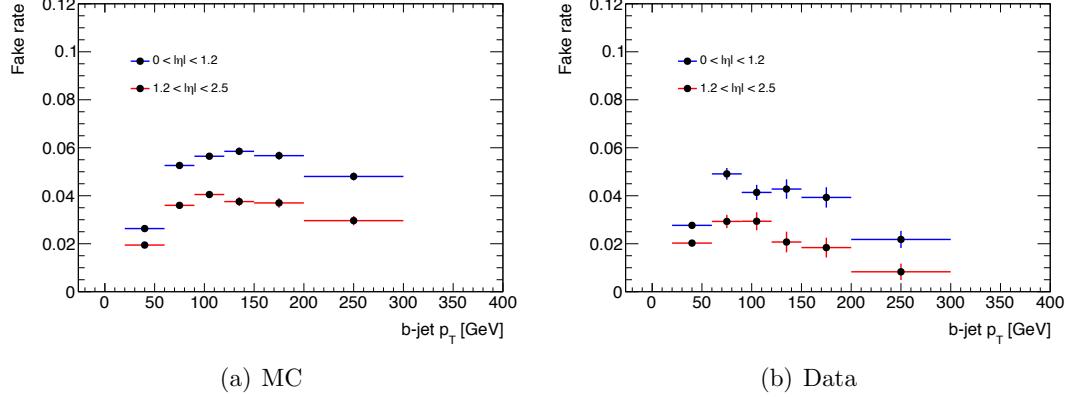


Figure 5.14: Fake efficiencies computed using both MC and data for a single lepton control region. The last p_T bin extends to infinity.

lepton region was already deemed most appropriate for a same-sign signal region, as here the statistics are significantly higher.

Real efficiencies: The probabilities for real b -jets to pass the MV1 operating point, as functions of p_T and η , can be taken from a centrally performed measurement, since there is not the same influence of the c -jet, τ -jet and light jet composition that effects an accurate computation of the fake efficiency. The ATLAS internal documentation [164] contains these efficiencies as measured in a $t\bar{t}$ control region; further documentation on the method can be found in [165].

Validation of the method

Closure tests: In order to verify that the method is unbiased, closure tests are performed in both the opposite-sign dilepton and single lepton scenarios. The basic procedure of this test is to measure efficiencies in a particular MC sample, and then apply the method to the whole sample blindly, pretending it is data. This ought to fairly represent the component of the ‘pretend data’ that has at least one fake b -jet. Truth matching is then used to extract the component with three real b -jets. When these backgrounds are summed they ought to match very well with the pretend data – any significant deviations could indicate a bias in the fake estimation procedure.

The opposite-sign dilepton closure tests uses the MC results from Figure 5.13, and the comparison in several distributions after following the procedure described above is

shown in Figure 5.15. Similarly for the single lepton case, efficiencies from Figure 5.14 are propagated through to the comparisons included in Figure 5.16. Overall the agreement is excellent, save for an apparent bias in the high jet multiplicity tail in the single lepton scenario. This is a moderately small effect, and since it acts in the direction of giving a background estimate that is slightly conservative, it is deemed that the method is sufficiently unbiased for the intended purpose.

Validation in data: Constructing a sensible validation region in data is difficult due to the already very limited statistics in the volume of phase space which requires three b -jets. The only change that is plausible is to invert the jet multiplicity requirement with respect to SR3b itself – thus we define

$$\begin{aligned} \text{SR3b : } & N_{\text{leps}} \geq 2, \quad N_{\text{jets}} \geq 5, \quad N_{b\text{-jets}} \geq 3 \\ \text{VR3b : } & N_{\text{leps}} \geq 2, \quad N_{\text{jets}} < 5, \quad N_{b\text{-jets}} \geq 3. \end{aligned}$$

The b -jet matrix method is then used to make a prediction in VR3b using the efficiencies computed from the single lepton region in both MC and data. A comparison, including this computation as well as the real three b -jet component from truth-matched MC is shown in Figure 5.17. It is very encouraging that either set of efficiencies produces an estimate that agrees with data within the uncertainties.

Results

The matrix method has then been applied to the SR3b signal region, using the single lepton efficiencies measured both in MC and in data, with the results presented in Table 5.11. The three rows correspond to the two bins of the signal region in the exclusion mode fit, and the ‘discovery’ mode signal region. The first two data columns exemplify the different results one obtains when using the MC vs the data derived efficiencies. Whilst there is a noticeable difference it is within the estimated uncertainties. Each of these is then combined with the real contribution in the third column, to give a number which can be compared to the nominal background estimate in the final column. These can finally be compared with the observed results of 0 events in the low m_{eff} bin of SR3b, one event in the high m_{eff} bin, and one event in SR3b_{disc}.

Comments on results: Firstly, it is interesting to note that the uncertainties on the b -jet matrix method background shown in Table 5.11 are clearly dominated by the

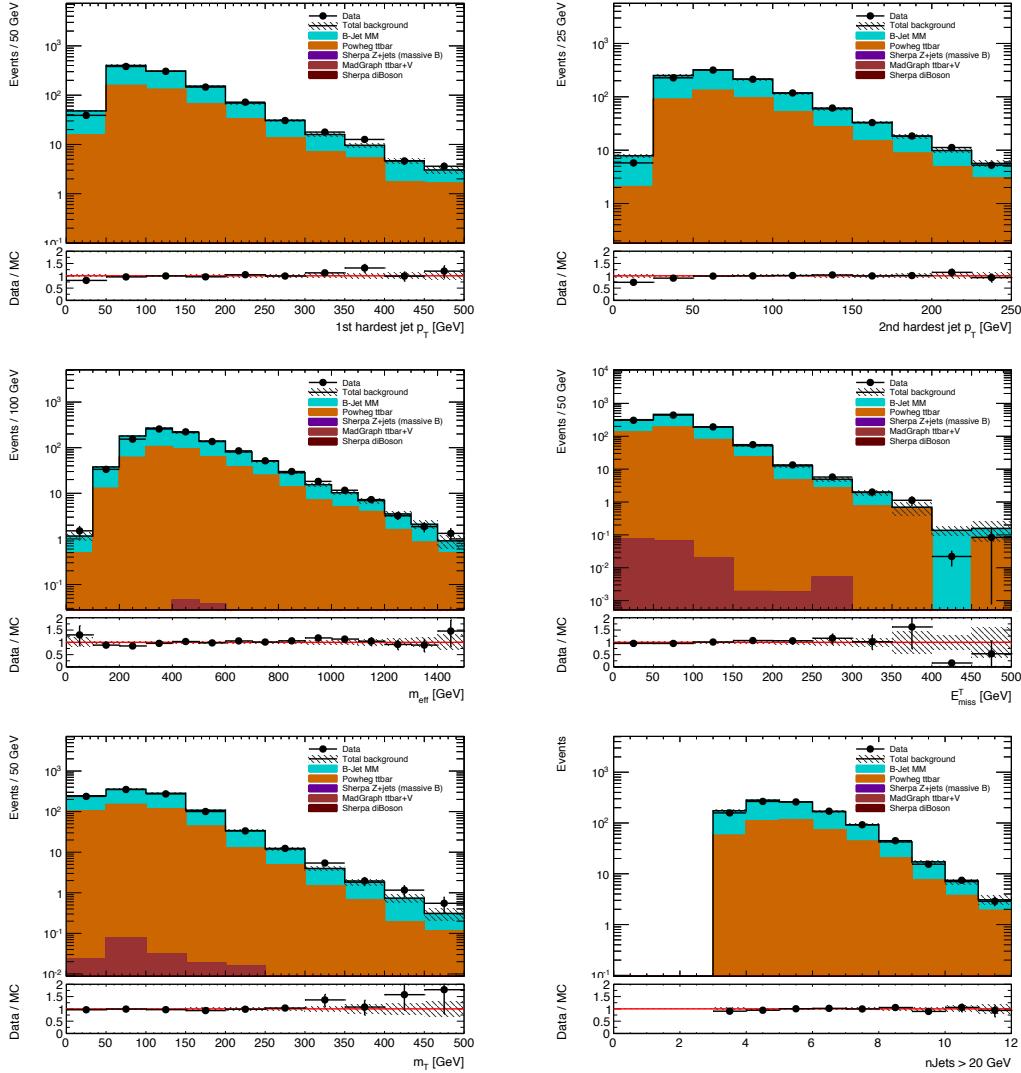


Figure 5.15: Closure test performed in opposite-sign control region using Powheg $t\bar{t}$ and MadGraph $t\bar{t} + V$ MC samples. The points labelled ‘Data’ are all events in the MC and the other MC components are truth matched to pertain to the ‘real 3b’ component. The matrix method component is computed using the efficiencies computed from MC in this region.

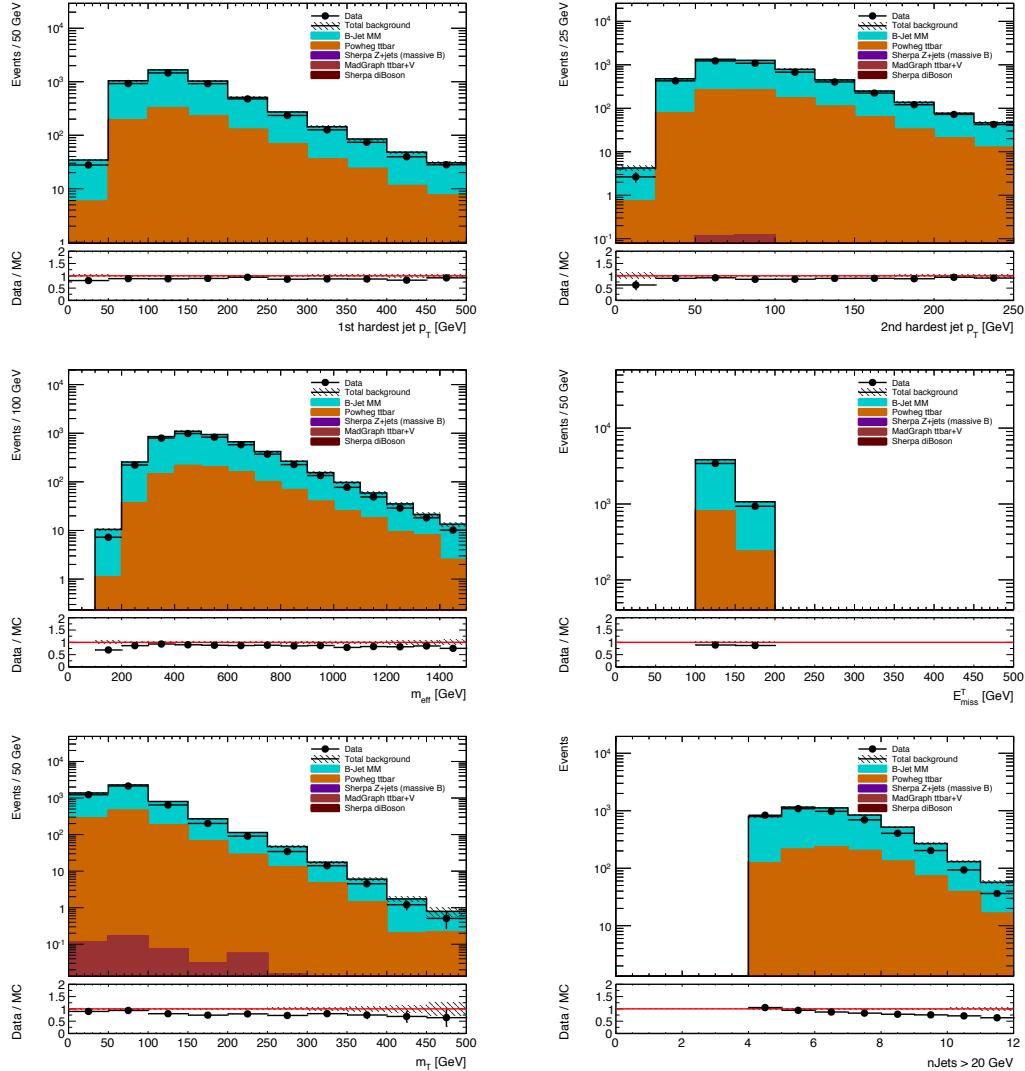


Figure 5.16: Closure test performed in single lepton control region requiring $100 < p_T^{\text{miss}} < 200$ GeV and at least 4 jets $p_T > 20$ GeV, using Powheg $t\bar{t}$ and MadGraph $t\bar{t} + V$ MC samples. The points labelled ‘Data’ are all events in the MC and the other labelled MC samples are truth matched to pertain to the ‘real 3b’ component. The matrix method component is computed using the efficiencies computed from MC in this region.

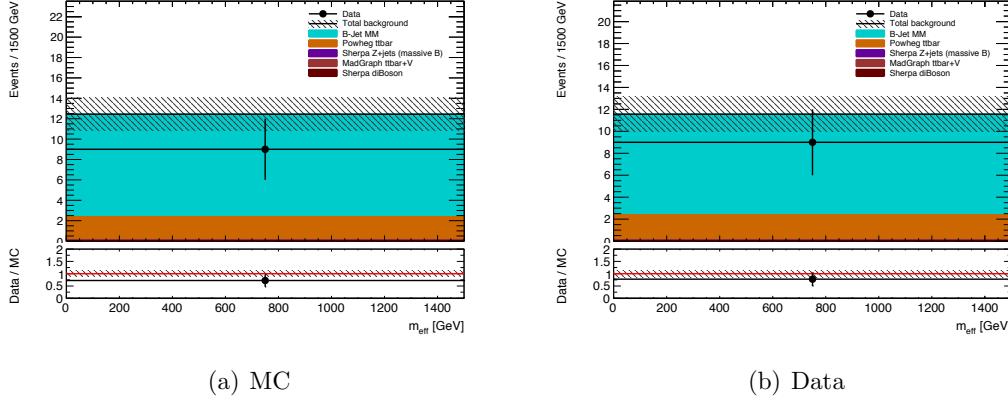


Figure 5.17: Agreement found in VR3b when using efficiencies computed in a single lepton region from both MC (left) and data (right). The MC samples included are truth matched so as to pertain to the ‘real 3b’ component.

Bin/region	MM (data eff.)	MM (MC eff.)	Real 3b (MC)	Nominal
Bin 1	1.51 ± 0.72	1.90 ± 0.89	0.29 ± 0.19	1.55 ± 0.63
Bin 2	1.37 ± 0.60	1.90 ± 0.83	< 0.01	0.45 ± 0.12
SR3b _{disc}	2.88 ± 0.94	3.80 ± 1.21	0.29 ± 0.19	1.99 ± 0.69

Table 5.11: Predictions in the bins of SR3b and SR3b_{disc} (as defined in Table 5.3), with fake efficiencies computed from data and MC in the single lepton control region. The first two columns are the data-driven matrix method described above, and the third is the component with three real b -jets estimated from truth-matched MC ($t\bar{t}$ and $t\bar{t} + V$; the $t\bar{t} + V$ component is negligible, however.). The ‘nominal’ values listed refer to estimates made using the normal fake lepton based matrix method, and MC samples, as presented in the results section.

statistical component in the application procedure, rather than by those in the efficiencies. This can be seen since the MC-based efficiencies have a significantly lower uncertainty than their data-driven counterparts, yet in the final estimate the overall uncertainties are of similar magnitude. As such, the method cannot be made more predictive only through an improved efficiency measurement; without fundamentally altering the method the uncertainty cannot be reduced. The combined uncertainty is larger than that for the nominal background estimate (from the fake lepton matrix method), driving the decision that this method be used as a cross-check for this signal region.

Whilst agreement between the purely data-driven fake b -jet background and the fake lepton estimate is good in the low m_{eff} region, in the other regions it is significantly larger. It is therefore also systematically overestimating the data in all regions. Whilst it is possible that this stems from the same issue observed in the single lepton closure test, it is hard to form any conclusion given the limited statistics. One therefore tentatively concludes that both fake lepton and fake b -jet matrix methods support the hypothesis that the observations in SR3b are consistent with background.

5.6 Systematic uncertainties

The measurement and reconstruction procedures are all prone to systematic error, and as such these sources of systematic uncertainty must be included when forming our result. One benefit of relying on data-driven backgrounds as much as possible is that these estimates are already subject to the same systematic *error* as the data to which we are comparing, and as such the effect of the *uncertainty* need only be considered for those backgrounds estimated with MC samples and passed through the detector simulation. These uncertainties are broken down as recommended by the ATLAS SUSY working group as follows.

- **Jet Energy Scale:** The energies of jets used in the analysis have been corrected to take into account the inefficiencies of the calorimeter cells, and their differing response to charged and neutral particles passing through them. A recommended calibration was derived in ATLAS using a combination of simulation, as well as test beam and *in situ* data [30, 166]. This procedure carries an inherent uncertainty correlated between all events – as such all distributions used in the final result are produced not only with the nominal calibration but with ‘Up’ and ‘Down’ variations, where the energy scale factors are varied accordingly within its uncertainty band.

Whilst the full calibration is available broken down into several independent sources that can be varied separately, our analysis uses the simplified (combined) version to reduce the number of nuisance variables in the fitting procedure.

- **Jet Energy Resolution:** In addition to the systematic calibration that must be applied to jet energies, the measured energy of a given jet will be smeared stochastically – the jet energy resolution (JER). The width of this distribution is not always correctly modelled by the detector simulation; as such a systematic uncertainty is created to estimate the effect on the final result of this defect. The JER in data has previously been estimated by ATLAS in dijet events [167].
- **B-tagging:** The simulation does not quite correctly model the b -tagging efficiency of the detector. Correction factors derived from $t\bar{t}$ and dijet measurements were applied to MC samples [168–170]. These correction factors are then varied within their uncertainties to produce ‘Up’ and ‘Down’ variations as for JES.
- **Lepton energy scale, resolution and ID efficiencies:** Similarly to the three sources of systematic uncertainty discussed regarding jets, electrons and muons have corresponding energy scale and resolution systematic uncertainties, as well as corrections to account for differing identification efficiency in the detector and simulation [15, 23].
- **Missing transverse momentum:** The primary source of uncertainty in the p_T^{miss} calculation is due to the calibration of ‘soft terms’, *i.e.* those calorimeter deposits that are not associated to a reconstructed object [171]. Both scale and resolution effects are considered, as for JES and JER however they are treated as uncorrelated to the jet uncertainties.
- **Simulation of pile-up:** The pile-up reweighting procedure described in section 5.2.4 depends on the value of $\langle \mu \rangle$. In order to reflect the uncertainty in this procedure this value is varied up and down by 10%, and treated as a systematic.
- **Luminosity:** As previously mentioned, the measured luminosity has an uncertainty of 2.8% [132]. The overall normalisation of MC backgrounds are scaled up and down by this amount to yield the systematic variations in the final distributions.
- **Trigger efficiency:** The trigger efficiency in MC samples was observed to be different to that in data, and as such factors were derived to correct for this. The uncertainty on these was propagated as a systematic uncertainty, as well as an additional 2% to account for inefficiencies observed in the plateau of the p_T^{miss} trigger.

Background	Systematic	Signal regions		
		SR0b,SR1b	SR3b	SR3L low,high
$t\bar{t} + V$	Cross section	22%	22%	22%
	Shape uncertainty	12%	12%	12%
$WZ + \text{jets}$	Cross section	7%	7%	7%
	Shape uncertainty	17%	29%	23%
	Parton multiplicity	30%	163%	56%
$ZZ + \text{jets}$	Cross section	5%	5%	5%
	Shape uncertainty	47%	23%	7%
	MC generator uncertainty	37%	78%	82%

Table 5.12: Summary of theoretical systematic uncertainties on $t\bar{t} + V$ and diboson samples arising from cross section, shape uncertainty from factorisation and normalisation scales, and the effects of ISR and FSR.

- **Theoretical:** The $t\bar{t} + V$ and diboson MC samples have associated normalisation and shape uncertainties, derived by considering the effect of varying the factorisation and renormalisation scales in the generators, as well as the overall normalisation variation introduced by the uncertainty on the theoretical cross sections for the processes. The cross sections have uncertainties of 22% for $t\bar{t} + W$ [145] and $t\bar{t} + Z$ [146], and 7% for diboson production [147]. Additionally the impact of initial state radiation (ISR) and FSR jets is considered for the $WZ + \text{jets}$ process by producing additional samples with `MadGraph` that include extra jets in the matrix element, and symmetrising differences that arise. These systematics were estimated at truth level to save the computational overhead of a full detector simulation. Finally, a generator uncertainty is assigned to the $ZZ + \text{jets}$ sample by symmetrising the largest difference in distributions produced by the nominal `Powheg` sample to those same distributions generated with `MC@NLO +Pythia6`, `MC@NLO +Pythia8`, and `Powheg +Pythia8`.

A summary of these theoretical uncertainties in each signal region can be found in Table 5.12. Additionally, processes with small contributions, $t\bar{t} + H$ and $t\bar{t}t\bar{t}$, are given a conservative 100% uncertainty, justified since this has no measurable impact on the results.

The lepton, p_T^{miss} , and pile-up performance related systematic uncertainties, as well as that on the luminosity, are found to be small compared to the others, and have negligible impact on the final results. All of the above are fed into the fitting and limit setting procedure, which is described in section 5.7. In doing so it is assumed that each of these sources of uncertainty are uncorrelated, and can hence be treated independently.

5.7 Fitting and limit setting procedure

Recall that limit setting using the CL_s method was previously discussed in a fairly abstract context in section 4.5.1; now this shall be applied to the problem in hand. In equation (4.18) (reproduced below) the observed data, \mathbf{X} , represents a vector of measured values dependent on a set of parameters $\boldsymbol{\theta}$, as well as the signal strength μ .

$$\mathcal{L}(\mu, \boldsymbol{\theta} | \mathbf{X}) = P(\mathbf{X} | \mu, \boldsymbol{\theta}). \quad (5.17)$$

In this analysis we aim to achieve two distinct goals:

- **Model independent discovery/upper limits:** Inspect each signal region individually for generic excesses, or if there are no excesses provide a 95% upper limit on the cross section \times efficiency of any new physics process.
- **Model dependent exclusions:** With a specific model in mind, try to form the most constraining CL_s value possible through not only using all signal regions simultaneously, but dividing each signal region into two or more bins in m_{eff} .

In each case we shall consider the form the likelihood function takes, at which point the limit setting procedure previously described can be applied. The analysis uses the HISTFITTER [172] package, which is based on ROOSTATS [173] and HISTFACTORY [174]. The HISTFITTER reference [172] provides further details on how the systematic uncertainties are incorporated into the fit.

5.7.1 Model independent discovery & upper limits

This is the simplest case, since here \mathbf{X} has a one-dimensional component with value X , representing the count of events in the signal region. In addition to X , one should also include auxiliary measurements that constrain the nuisance parameters – these

observations are packaged into a variable denoted $\tilde{\boldsymbol{\theta}}$. Given statistical independence of each background estimate i , one writes

$$P(\mathbf{X}, \tilde{\boldsymbol{\theta}} | \mu, \boldsymbol{\theta}) = g(X | \nu_s(\mu, \boldsymbol{\theta})) \times \prod_i g(X | \nu_i(\boldsymbol{\theta})) \times \prod_{\alpha} \pi_{\alpha}(\tilde{\theta}_{\alpha} | \theta_{\alpha}), \quad (5.18)$$

where ν_i represents the mean of a Poisson distribution representing the contribution of background process i in the signal region, and ν_s is the same thing for signal. Both depend on $\boldsymbol{\theta}$ which will contain parameters controlling all independent sources of uncertainty; these underlying parameters are what will ultimately be profiled over in the CL_s procedure. Each term $g(X | \nu) = e^{-\nu} \nu^X / X!$ is the Poisson probability density function. The subsequent terms $\pi(\tilde{\theta} | \theta)$ represent the likelihood of θ given the auxiliary measurement $\tilde{\theta}$ that has been made. The precise form of this likelihood will depend on the nuisance parameter in question. For example, the luminosity uncertainty would be expressed by a Gaussian constraint, centred on the known value and with the experimentally determined uncertainty as its width. Its effect would be to scale *all* of the ν_i rates in the same way, since it is correlated across all samples. Alternatively, the theoretical uncertainty in a given cross section calculation would only result in scaling the *one* sample to which it corresponds.

Since the upper-limit mode has a comparatively simple likelihood function, it is feasible to do the integral in equation (4.23) by performing pseudo-experiments – that is generating datasets drawn from a distribution that is consistent with those observed. In this process 5000 such pseudo datasets are used.

5.7.2 Model-dependent exclusions

As defined in Table 5.3, signal regions are divided into multiple regions in m_{eff} for the purpose of making a stronger exclusion of simulated signal models. Intuitively this can be done because the *shape* of the distributions of signal and background sources can be different within a given signal region. Each of these bins from every signal region is considered at once in the likelihood, so equation (5.18) needs to be modified to include a product over these regions. The observed data in each region are now, again, packaged into a vector \mathbf{X} . The overall expression for the likelihood is then

$$P(\mathbf{X}, \tilde{\boldsymbol{\theta}} | \mu, \boldsymbol{\theta}) = \prod_q g(X_q | \nu_{s,q}(\mu, \boldsymbol{\theta})) \times \prod_i g(X_q | \nu_{i,q}(\boldsymbol{\theta})) \times \prod_{\alpha} \pi_{\alpha}(\tilde{\theta}_{\alpha} | \theta_{\alpha}). \quad (5.19)$$

Here the vector of nuisance parameters $\boldsymbol{\theta}$ has now been expanded in equation (5.19) with respect to equation (5.18), since it must not only encode the information about all background processes in all regions, but also the predicted signal process contributions in each region. Note in fact that, by convention, one leaves μ as a free ‘signal strength’ parameter, without additional constraints, with respect to which the limit setting is performed. It is exactly equivalent to instead use this to directly represent the cross section of the signal process, if one includes the corresponding constraint term from theory on μ .

For the model dependent case, more regions are being fit simultaneously resulting in a more complex likelihood function than before. The procedure also needs to be repeated many times for each parameter point of each model. Thus, when evaluating equation (4.23), the asymptotic approximations are used rather than the more accurate, but slower, pseudo-dataset method.

5.8 Results

With the analysis performed as described in the preceding sections, the results can now be presented. Most fundamental are the combined background predictions in the signal regions, compared with the observed data. As shall be seen, no excesses are observed with respect to the SM background prediction. In light of this, upper limits on generic processes are computed and presented for each discovery mode signal region.

5.8.1 Yields in signal regions

Running in ‘discovery mode’, the predicted background contributions are compared to the observed yields in Table 5.13. This shows that whilst most of the signal region results are consistent with the SM prediction, there are some small tensions in SR0b and SR1b. This has been quantified in the table through computation of the p value for the excess. One can transform these p values to find that the significances of these excesses are 1.8 and 1.5 standard deviations respectively, and the combined region SR0b+SR1b has an excess with a significance of 2.1 standard deviations.

The uncertainties in Table 5.13 also demonstrate that the statistical uncertainties on the MC driven background estimates are a major factor, along with the uncertainties on

	SR3b	SR0b	SR1b	SR3Llow	SR3Lhigh
Observed events	1	14	10	6	2
Expected background events	2.2 ± 0.8	6.5 ± 2.3	4.7 ± 2.1	4.3 ± 2.1	2.5 ± 0.9
$p(s = 0)$	0.50	0.03	0.07	0.29	0.50
Background components					
$t\bar{t}V, t\bar{t}H, tZ$ and $t\bar{t}t\bar{t}$	1.3 ± 0.5	0.9 ± 0.4	2.5 ± 1.7	1.6 ± 1.0	1.3 ± 0.7
Dibosons and tribosons	< 0.1	4.2 ± 1.7	0.9 ± 0.4	1.2 ± 0.6	1.2 ± 0.6
Fake leptons	0.7 ± 0.6	$1.2^{+1.5}_{-1.2}$	$0.8^{+1.2}_{-0.8}$	1.6 ± 1.6	< 0.1
Charge-flip electrons	0.2 ± 0.1	0.2 ± 0.1	0.5 ± 0.1	—	—
Systematic uncertainties					
Fake-lepton background	± 0.6	$^{+1.5}_{-1.2}$	$^{+1.2}_{-0.8}$	± 1.6	< 0.1
Theory: dibosons	< 0.1	± 1.5	± 0.3	± 0.4	± 0.4
Jet and p_T^{miss} scale, resolution	± 0.1	± 0.7	± 0.4	± 0.4	± 0.3
Monte Carlo statistics	± 0.1	± 0.5	± 0.2	± 0.4	± 0.4
b -jet tagging	± 0.2	± 0.5	± 0.1	< 0.1	± 0.1
Theory: $t\bar{t}V, t\bar{t}H, tZ$ and $t\bar{t}t\bar{t}$	± 0.4	± 0.3	± 1.7	± 1.0	± 0.6
Trigger, luminosity, pile-up	< 0.1	± 0.1	± 0.1	± 0.1	± 0.1
Charge-flip background	± 0.1	± 0.1	± 0.1	—	—
Lepton identification	< 0.1	± 0.1	< 0.1	± 0.1	± 0.1

Table 5.13: Tabulation of the number of observed data events together with the expected backgrounds predictions and a summary of the corresponding systematic uncertainties for the discovery signal regions SR3b, SR0b, SR1b, SR3Llow and SR3Lhigh (with the additional m_{eff} cut of Table 5.2). The p -value of the observed dataset for the background-only hypothesis is denoted by $p(s = 0)$. By convention, the $p(s = 0)$ value is truncated at 0.50 when the number of observed data events is smaller than the expected background prediction. The breakdown of the systematic uncertainties on the expected backgrounds is also shown; however it should be noted that correlations exist between them that will cause them not to add in quadrature to the total quoted uncertainty.

the fake background estimate and the theoretical uncertainty on the cross section of the diboson processes. This suggests that these are the two areas in which improvements would most aid the constraining power of the analysis.

Additionally the distributions of observed events and predictions in m_{eff} for the exclusion mode signal regions are shown in Figure 5.18. Although impossible to draw any firm conclusions given the limited available data, there is a suggestion that the excess in SR0b mentioned previously is consistent with an overall scaling error in the background prediction. More suggestively, a benchmark simplified model is chosen for each signal region and its m_{eff} distribution is overlaid. The benchmark points are chosen to be demonstrative of the signature for which the given signal region is most sensitive, and to be at the point of being not-quite excluded.

5.8.2 Model-independent limits

Given the raw results presented in the previous section, the fitting procedure described in section 5.7.1 can be used to set model-independent upper limits on the visible cross section in each signal region, or equivalently a limit on the number of events that could have been produced with 20.3 fb^{-1} at the LHC. The visible cross section is defined to be the product of acceptance, reconstruction efficiency, and production cross section. The results can be seen in Table 5.14.

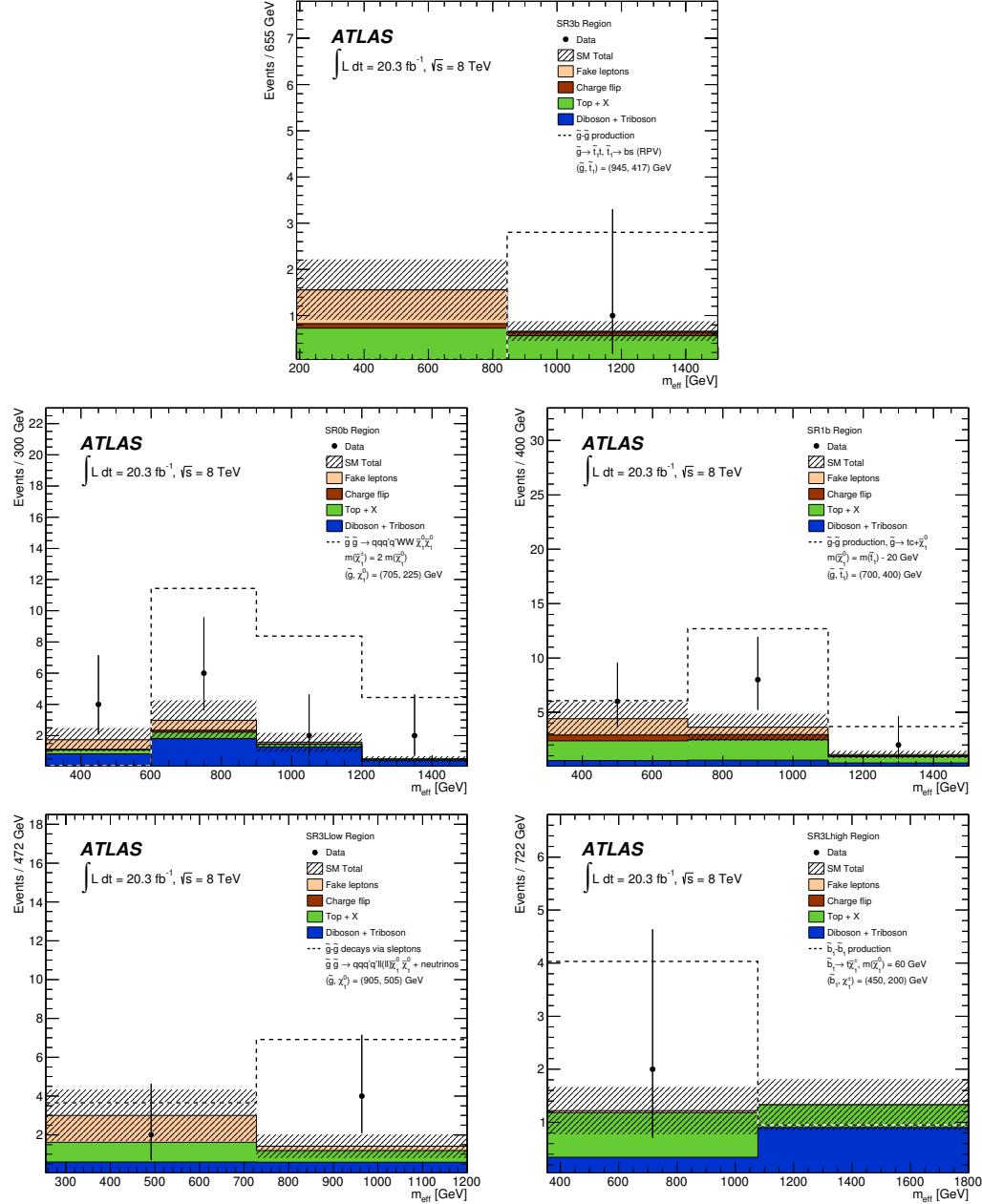


Figure 5.18: Yields in the five signal regions defined in the analysis, shown after the fitting procedure has been performed. The hatched regions denote the statistical and systematic uncertainties on the overall background prediction. The dashed lines are the overlaid distribution in m_{eff} of selected simplified model points – the descriptions refer to the final states and parameter values of models described in section 5.9.1.

Signal region	$\langle \sigma_{\text{vis}} \rangle_{\text{obs}}^{95} / \text{fb}$	S_{obs}^{95}	S_{exp}^{95}
SR3b	0.19	3.9	$4.4^{+1.7}_{-0.6}$
SR0b	0.80	16.3	$8.9^{+3.6}_{-2.0}$
SR1b	0.65	13.3	$8.0^{+3.3}_{-2.0}$
SR3Llow	0.42	8.6	$7.2^{+2.9}_{-1.3}$
SR3Lhigh	0.23	4.6	$5.0^{+1.6}_{-1.1}$

Table 5.14: The 95% CL_s upper limits on the visible cross section ($\langle \sigma_{\text{vis}} \rangle_{\text{obs}}^{95}$), and the observed and expected 95% CL_s upper limits on the number of BSM events (S_{obs}^{95} and S_{exp}^{95}).

5.9 Interpretation of results

In section 5.7.2, the procedure for setting CL_s limits on given models was described. In this section several simplified models, as well as more phenomenologically complete scenarios, are tested with the results from this analysis. In each case the model typically has a set of free parameters, allowing a two dimensional exclusion to be drawn. These plots typically all have the same form:

- A dashed line representing the expected exclusion under the ‘no signal hypothesis’. That is, it uses the nominal value of the background estimate in place of the observed data. The limit curve is interpolated from a grid of points, for each of which a CL_s value will have been calculated.
- A yellow band around the expected exclusion. This represents the estimate of the $\pm 1\sigma$ uncertainty in the expected exclusion due to the underlying systematic and statistical uncertainties in the background predictions; this is computed by the HISTFITTER package. Note that while this does include systematic uncertainties on the signal samples, for example the effect of JES and JER, it does not include the theoretical uncertainty on model cross sections.
- A solid red line representing the observed exclusion. It is interpolated as before, but this time uses the observed data.
- Dotted red lines either side of the solid line represent the $\pm 1\sigma$ effect of the uncertainty in the signal cross section.

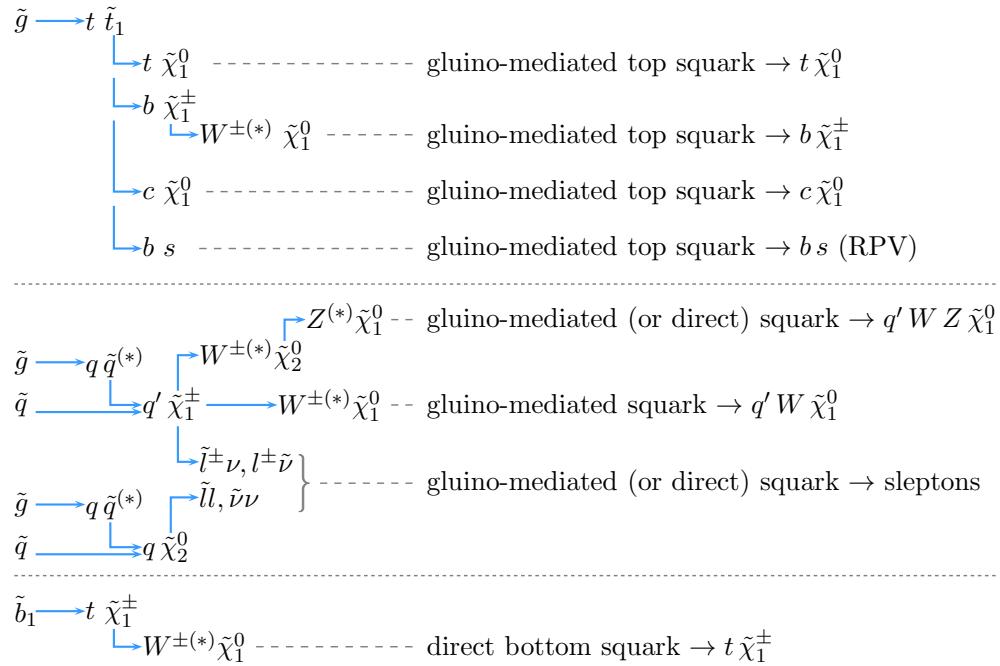


Figure 5.19: Overview of the decay modes present in the simplified SUSY models that are considered by this analysis.

- Coloured dotted lines in some plots represent the expected limit curve from individual signal regions, rather than their combination.
- Grey numbers in some of the simplified model plots represent the excluded cross section for the process.
- Coloured solid lines in a few plots are used to make a comparison with the observed exclusions of other published analyses.

5.9.1 Simplified models

A summary of the simplified models studied in this analysis are shown in Figure 5.19. These naturally fall into three categories which are described below, together with their results. A description of the basic concepts behind simplified models can be found in section 2.2.3.

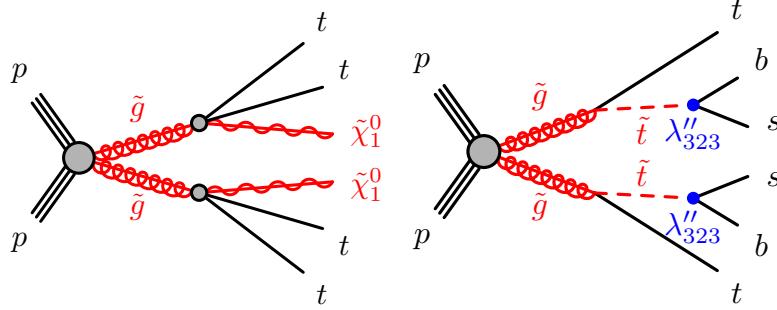


Figure 5.20: Examples of simplified models featuring gluino-mediated stop production. On the left the case of an off-shell stop decaying to $t\tilde{\chi}_1^0$ is shown, and on the right the RPV model with a non-zero UDD coupling, producing tbs in the final state from each gluino decay. Diagrams taken from [61].

Gluino-mediated top quarks

The first tranche of models in Figure 5.19 represent gluino pair production followed by a decay $\tilde{g} \rightarrow t\tilde{t}_1^{(*)}$ with a branching fraction of 100%. The \tilde{t}_1 is taken to be the lightest squark, as favoured by naturalness arguments; all other squarks are decoupled in these scenarios. The four models detailed below differ in the decay mode of the stop, two examples of which are shown in Figure 5.20. In the subsequent discussion, we shall refer to \tilde{t}_1 simply as \tilde{t} .

Gluino-stop ($t\tilde{\chi}_1^0$) off-shell. In this model the top squark is produced off-shell and decays to $t\tilde{\chi}_1^0$ with branching fraction 100%, as featured in the left-hand side of Figure 5.20. That is, the decay chain for each gluino is $\tilde{g} \rightarrow t\tilde{t}^* \rightarrow t\bar{t}\tilde{\chi}_1^0$, with tops decaying either leptonically or hadronically. The neutralino is the LSP in this scenario. Since the gluinos are required to be on-shell in this model, there exists an additional constraint that $m_{\tilde{g}} > 2m_t + m_{\tilde{\chi}_1^0}$. The stop mass is set to $m_{\tilde{t}} = 2.5$ TeV, and all other squarks are decoupled with much higher masses. Limits are then set in the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$ -plane, as shown in Figure 5.21. As might be expected due to the high multiplicity of b quarks in the final state, the sensitivity is dominated by SR3b.

Gluino-stop ($b\tilde{\chi}_1^\pm$) on-shell. In this model the top squark is produced on-shell and subsequently decays to $b\tilde{\chi}_1^\pm$, where it is specified that $m_{\tilde{\chi}_1^\pm} = 118$ GeV. Since we additionally constrain $m_{\tilde{\chi}_1^0} = 60$ GeV, making it the LSP, the chargino will decay via an off-shell W , $\tilde{\chi}_1^\pm \rightarrow W^*\tilde{\chi}_1^0$. Again, the gluino is required to be on shell, resulting in

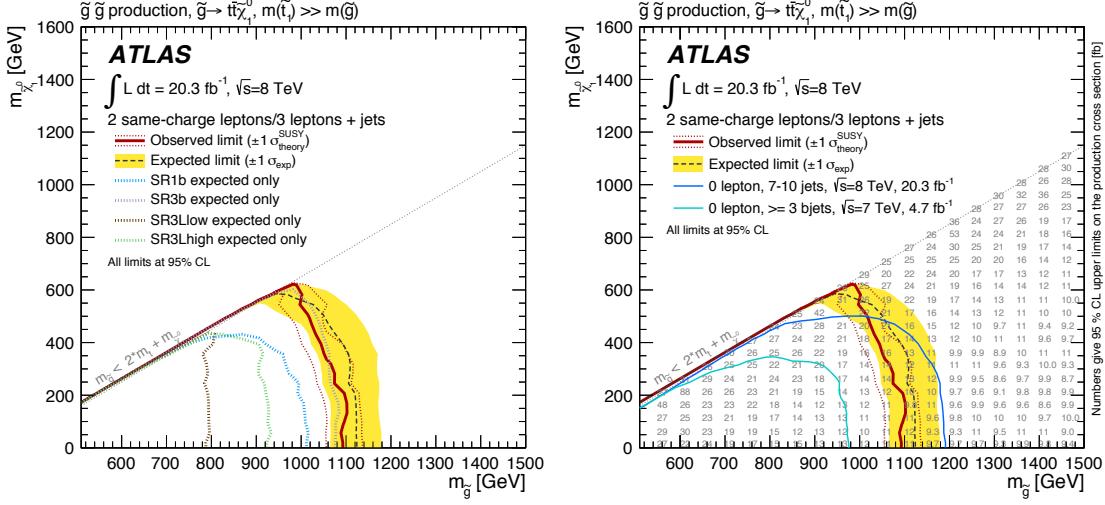


Figure 5.21: Exclusions for a gluino-mediated stop production model, with the off-shell stop decaying to $t\tilde{\chi}_1^0$. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point. The right hand plot also shows a comparison with limits from multijet [175] and three b -jet [176] analyses.

the mass requirement $m_{\tilde{g}} > m_t + m_{\tilde{t}}$. Limits are set in the $(m_{\tilde{g}}, m_{\tilde{t}})$ -plane, as shown in Figure 5.22. As in the $\tilde{t} \rightarrow t\tilde{\chi}_1^0$ case, sensitivity is dominated by SR3b.

Gluino-stop ($c\tilde{\chi}_1^0$) on-shell. In this model there is a small mass splitting between the top squark and neutralino (which is the LSP again), $\Delta(m_{\tilde{t}}, m_{\tilde{\chi}_1^0}) = 20 \text{ GeV}$. This prevents the stop decaying to a top quark, however the channel to $c\tilde{\chi}_1^0$ is still allowed. This is the sole decay mode considered in this scenario. In order to keep the gluino on shell, it is required that $m_{\tilde{g}} > m_t + m_c + m_{\tilde{\chi}_1^0}$. Limits are set in the $(m_{\tilde{g}}, m_{\tilde{t}})$ -plane, as shown in Figure 5.23. Due to fewer top quarks in the final state compared to the other models in this section, SR1b has the dominant sensitivity for this model.

Gluino-stop (bs) RPV. This model is also shown in Figure 5.20, and features gluino-mediated production of on-shell stops, which are then decayed by an RPV UDD coupling (see section 2.2.2), whose coupling strengths are denoted by the matrix λ'' . This case considers only $\lambda''_{323} = \lambda''_{332} = 1$ (equality follows from symmetries of λ'') to be non-zero, as proposed in [177]. This value is large enough, by several orders of magnitude, to make the stop lifetime small enough to prevent the appearance of displaced vertices in

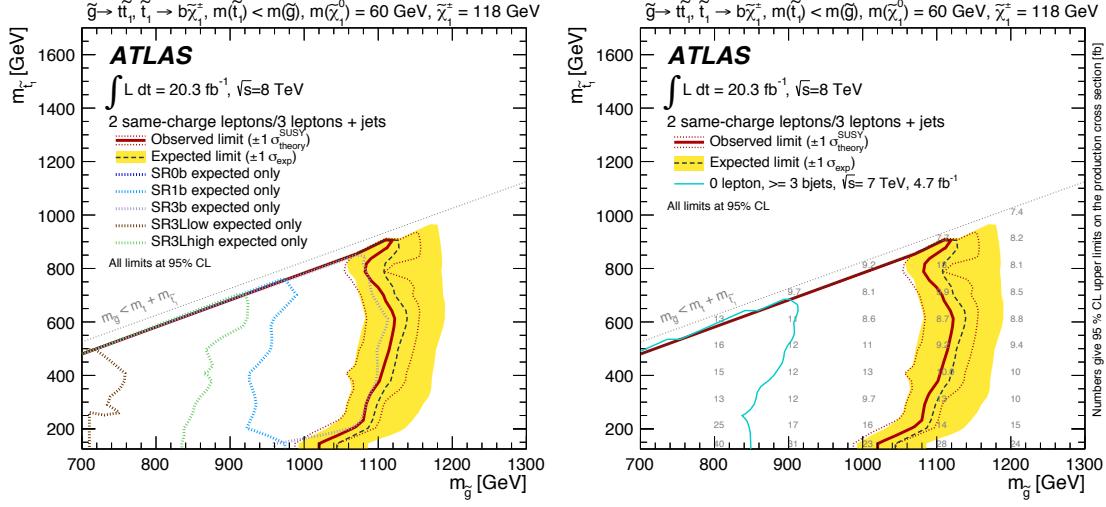


Figure 5.22: Exclusions for a gluino-mediated stop production model, with the stop decaying to $b\tilde{\chi}_1^{\pm}$. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point. The right hand plot also shows a comparison with limits from a three b -jet [176] analysis.

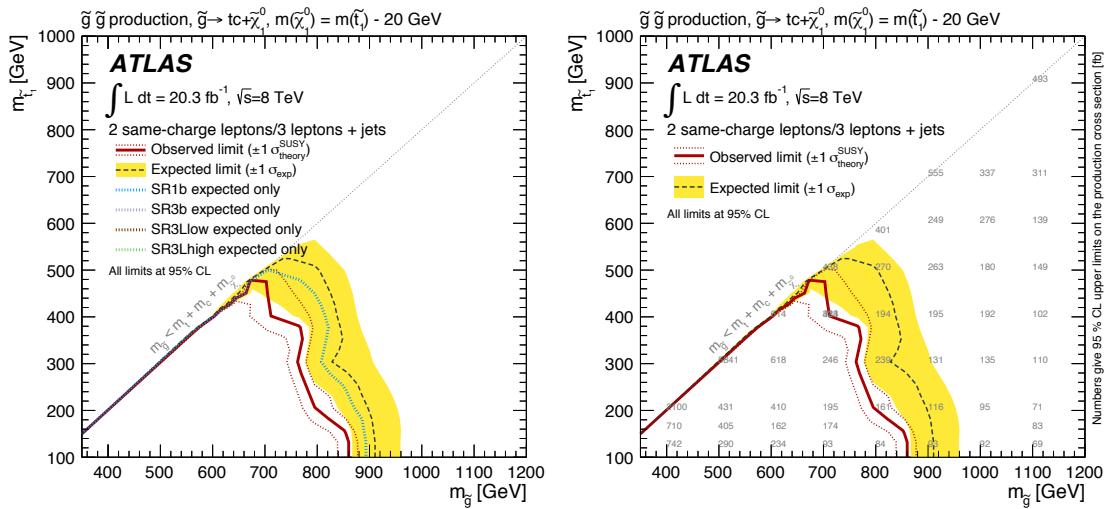


Figure 5.23: Exclusions for a gluino-mediated stop production model, with the stop decaying to $c\tilde{\chi}_1^0$. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point.

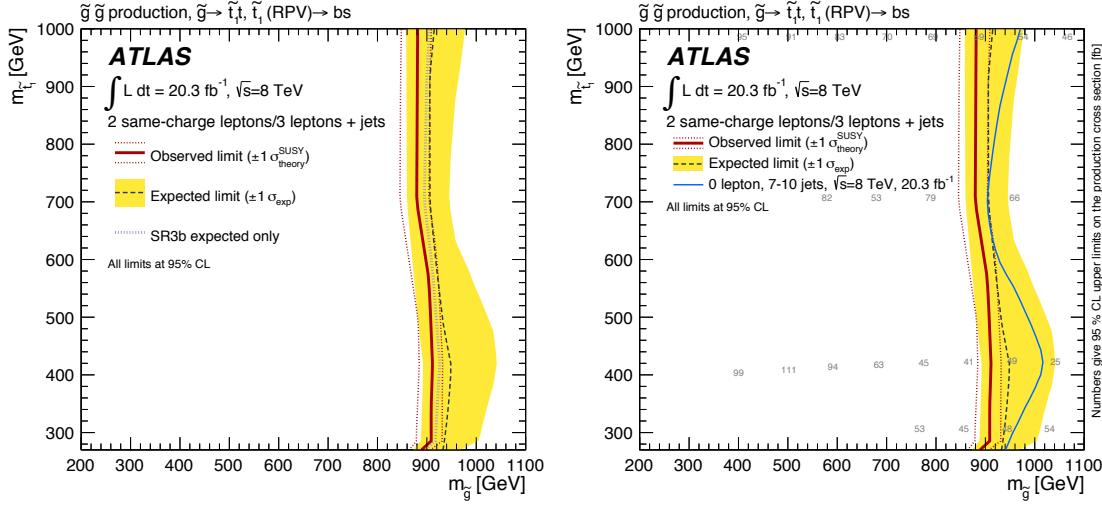


Figure 5.24: Exclusions for an RPV model with non-zero λ''_{323} . The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point. The right hand plot also shows a comparison with limits from a multijet [175] analysis.

the detector. Results are interpreted in the $(m_{\tilde{g}}, m_{\tilde{t}})$ -plane, as shown in Figure 5.24. The final state hence has four b -quarks, but limited p_T^{miss} , and hence sensitivity is again dominated by SR3b.

Summary. SUSY scenarios involving the top squark are favoured by naturalness, and with this analysis significant bounds have been placed across several key signatures. Due to the large multiplicity of b -quarks that appear in the final states, SR3b is particularly effective, however sensitivity is also gained in RPC scenarios with SR1b, and the three lepton region SR3Lhigh. Gluino masses below 850 GeV are excluded independently of top squark mass in the RPV case, and up to 950 GeV is excluded in the RPC scenarios considered where b -jets are produced; in the trickier $\tilde{t} \rightarrow c\tilde{\chi}_1^0$ scenario it is still possible to rule out gluinos below 840 GeV.

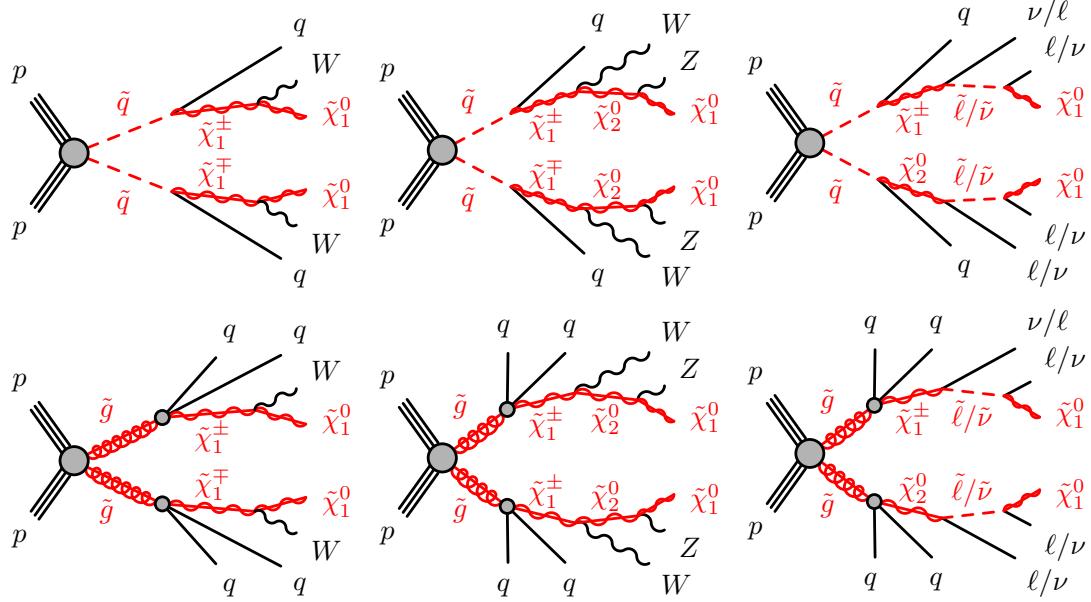


Figure 5.25: Examples of simplified models featuring 1- and two-step decays of light (first and second generation) squarks. The squarks are produced either directly, as shown in the first row, or via gluinos, as shown in the second row. Gluino-mediated production yields extra jets in the final state. From left to right is shown the one-step decay, two-step decay via charginos, and two-step decay via sleptons. Diagrams are taken from [61].

First and second generation squark production

Six simplified models, shown in Figure 5.25, are considered here, each featuring either direct or gluino-mediated production of a pair of first or second generation squarks.⁶ Exclusions are then presented for five of these. The models otherwise differ in the decay mode allowed for the squarks, and are discussed in further detail in the ensuing paragraphs.

Strong production one-step decay These models are shown on the far left-hand side of Figure 5.25, that is we consider the decay $\tilde{q} \rightarrow qW\tilde{\chi}_1^0$, via a chargino. We choose the neutralino and chargino masses to be related by $m_{\tilde{\chi}_1^\pm} = 2m_{\tilde{\chi}_1^0}$, and specify the neutralino to be the LSP. Other particles are decoupled. However, one can note that the

⁶in this section \tilde{q} is used to exclusively denote these lighter squarks, and not those from the third generation

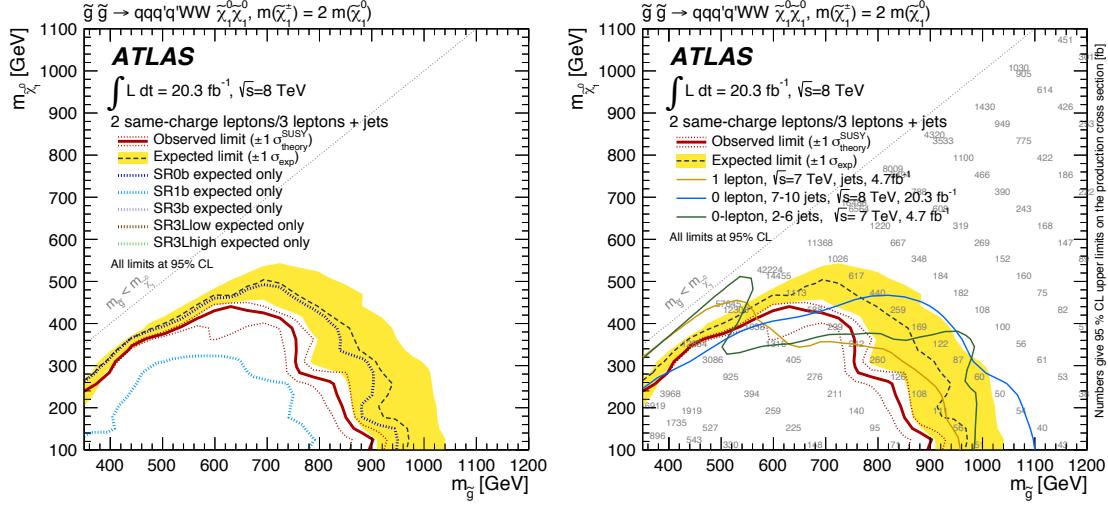


Figure 5.26: Exclusions for the light squark production (from gluino pair production) model with one intermediary SUSY particle. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point. The right hand plot also shows a comparison with limits from zero lepton [178], multijet [175] and single lepton [179] analyses.

final state in the case of direct squark production is

$$\tilde{q}\tilde{q} \rightarrow qq W^{(*)} W^{(*)} \tilde{\chi}_1^0 \tilde{\chi}_1^0, \quad (5.20)$$

and hence can only contain two opposite-sign leptons, rather than the same-sign pair required for this analysis to be sensitive. Hence this variant is discounted for this scenario, and limits are only set on the gluino-mediated squark production, where the final state is

$$\tilde{g}\tilde{g} \rightarrow qqqq W^{(*)} W^{(*)} \tilde{\chi}_1^0 \tilde{\chi}_1^0. \quad (5.21)$$

The limits are shown in the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$ -plane in Figure 5.26, with the largest sensitivity from SR0b due to the lack of b -quarks in the final state.

Strong production two-step decay via gauginos These are the central two models in Figure 5.25, and consider the decay $\tilde{q} \rightarrow qWZ\tilde{\chi}^0$, which is mediated by first a chargino and then a neutralino ($\tilde{\chi}_2^0$). In order to preserve the correct hierarchy in the spectrum, the masses of these two particles are set in between the gluino and neutralino masses

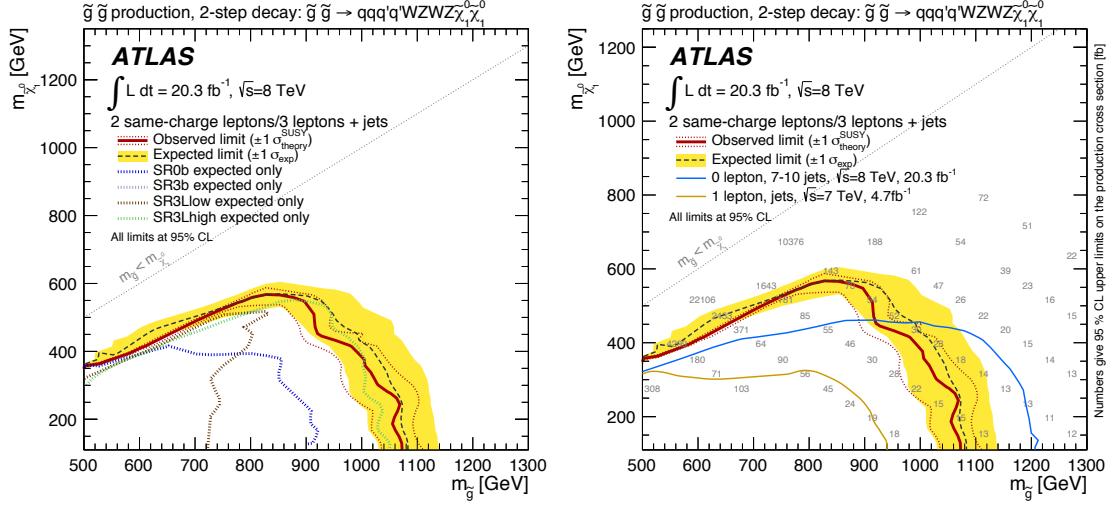


Figure 5.27: Exclusions for the light squark production (from gluino pair production) model with a two-step decay involving gauginos. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point. The right hand plot also shows a comparison with limits from multijet [175] and single lepton [179] analyses.

according to

$$m_{\tilde{\chi}_1^\pm} = \left(m_{\tilde{g}/\tilde{q}} + m_{\tilde{\chi}_1^0} \right) / 2 \quad (5.22)$$

$$m_{\tilde{\chi}_2^0} = \left(m_{\tilde{\chi}_1^\pm} + m_{\tilde{\chi}_1^0} \right) / 2, \quad (5.23)$$

where the squark mass is used in the direct production scenario, and the gluino mass otherwise. The final states are most easily read from the Feynman diagrams in Figure 5.25, although as in the one-step case it should be noted that the direct squark production variant will always produce oppositely charged W bosons. Unlike the one-step case, a same-sign pair can always be formed from the Z bosons. The W and Z bosons may be off-shell, depending on the mass splitting $\Delta m(\tilde{g}, \tilde{\chi}_1^0)$ (or $\Delta m(\tilde{q}, \tilde{\chi}_1^0)$ in the direct production case). Results are plotted in the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$ -plane in Figure 5.27 for the gluino-mediated case, and in the $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0})$ -plane in Figure 5.28 for the direct squark case. The three lepton signal regions are most sensitive in this scenario, with SR3Lhigh doing best at large mass splitting, and SR3Llow doing perhaps marginally better in the more compressed scenario.

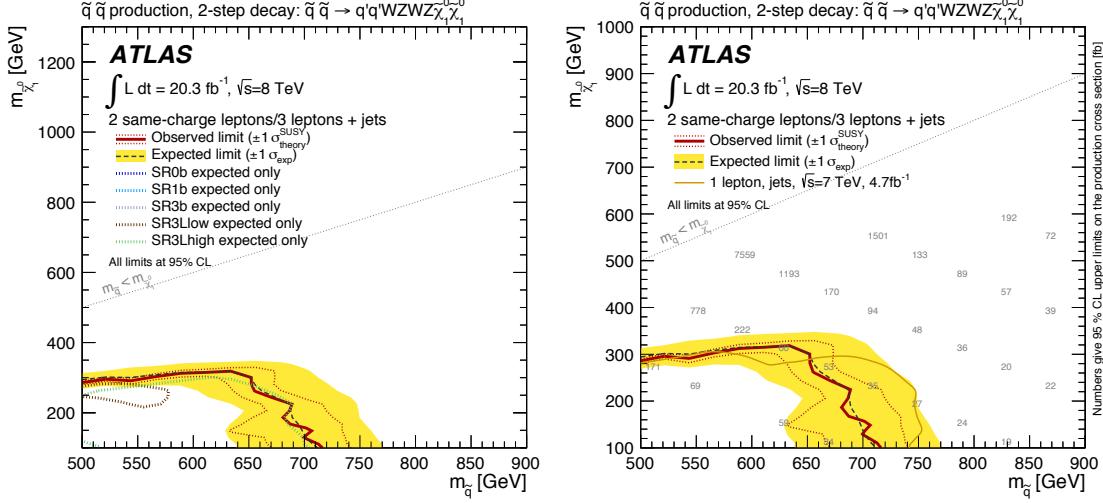


Figure 5.28: Exclusions for the light squark production (from squark pair production) model with a two-step decay involving gauginos. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point.

Strong production two-step decay via sleptons In this model the squarks have two modes of decay available, $\tilde{q} \rightarrow q\tilde{\chi}_1^\pm$ and $\tilde{q} \rightarrow q\tilde{\chi}_2^0$, which are defined to have equal branching fractions. The diagrams on the far right of Figure 5.25 attempt to encode this for both direct and gluino-mediated squark production. Subsequently both the second neutralino and chargino have two decay modes open to them, which again occur with equal probability:

$$\tilde{\chi}_1^\pm \rightarrow \tilde{\ell}^\pm \nu \quad \tilde{\chi}_2^0 \rightarrow \ell^\pm \tilde{\ell}^\mp \quad (5.24)$$

$$\tilde{\chi}_1^\pm \rightarrow \ell^\pm \tilde{\nu} \quad \tilde{\chi}_2^0 \rightarrow \nu \tilde{\nu}. \quad (5.25)$$

The sleptons always decay via $\tilde{\ell} \rightarrow \ell \tilde{\chi}_1^0$, and the sneutrinos as $\tilde{\nu} \rightarrow \nu \tilde{\chi}_1^0$. It is defined that all three flavours of slepton are degenerate in mass, and that

$$m_{\tilde{\chi}_1^\pm} = m_{\tilde{\chi}_2^0} = \left(m_{\tilde{g}/\tilde{q}} + m_{\tilde{\chi}_1^0} \right) / 2 \quad (5.26)$$

$$m_{\tilde{\ell}} = m_{\tilde{\nu}} = \left(m_{\tilde{\chi}_2^0} + m_{\tilde{\chi}_1^0} \right) / 2. \quad (5.27)$$

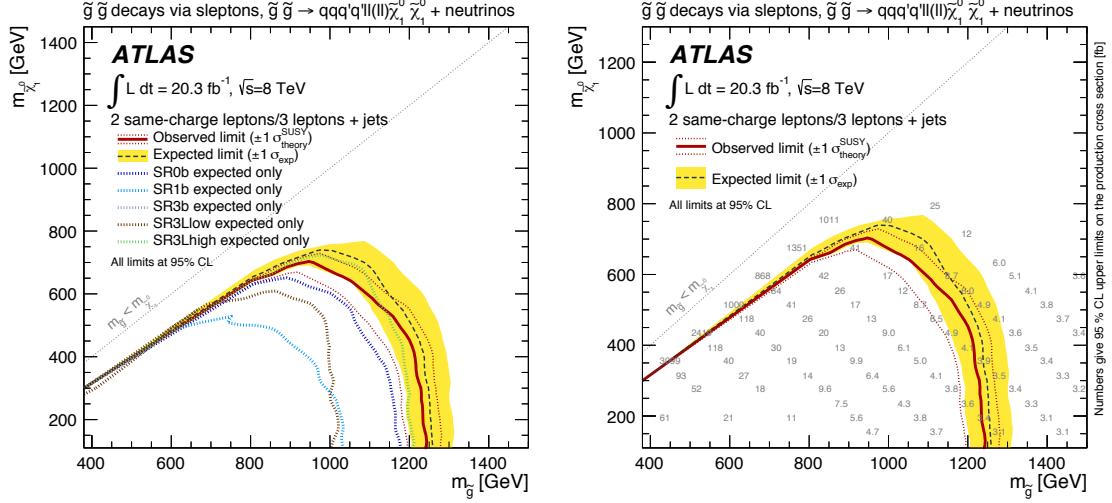


Figure 5.29: Exclusions for the light squark production (from gluino pair production) model with a two-step decay involving sleptons. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point.

Thus there is some variety in the final states that are allowed, enumerating the possible final states for a squark decay yields:

$$\tilde{q} \rightarrow q \ell \nu \tilde{\chi}_1^0 \quad (5.28)$$

$$\tilde{q} \rightarrow q \ell \ell \tilde{\chi}_1^0 \quad (5.29)$$

$$\tilde{q} \rightarrow q \nu \nu \tilde{\chi}_1^0, \quad (5.30)$$

the first of which is twice as likely as the other two.

The entire event can thus comprise missing transverse momentum, two or four light jets, and up to four leptons. Results are plotted in the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$ -plane in Figure 5.29 for the gluino-mediated case, and in the $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0})$ -plane in Figure 5.30 for the direct squark case. The best sensitivity is gained by SR3Lhigh in all parts of parameter space shown, most likely due to the possibility of both significant p_T^{miss} and high lepton multiplicity.

Summary These models, with longer decay chains, are those that especially benefit from the existence of the three lepton signal regions SR3Lhigh and SR3Llow. Gains made in the compressed regions of parameter space (small $\Delta m(\tilde{g}, \tilde{\chi}_1^0)$), for example in Figure 5.27 compared to the multijet analysis, can be attributed to the low threshold of

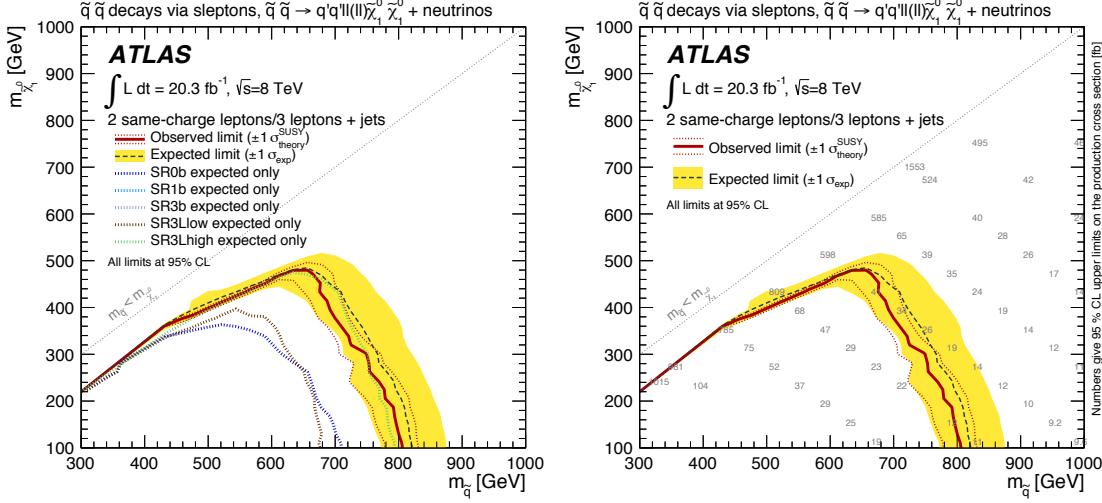


Figure 5.30: Exclusions for the light squark production (from squark pair production) model with a two-step decay involving sleptons. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point.

15 GeV on the transverse momenta of subleading leptons. Overall the reach in gluino mass varies considerably, by more than 300 GeV, between the ‘easiest’ and ‘hardest’ scenarios described above.

Direct sbottom production

In this scenario only the sbottom, neutralino and chargino SUSY particles are considered (others are decoupled). As shown in Figure 5.31 the model involves direct production of sbottom squarks, followed by a decay $\tilde{b} \rightarrow t\tilde{\chi}_1^\pm$, the chargino subsequently decaying $\tilde{\chi}_1^\pm \rightarrow W^{(*)}\tilde{\chi}_1^0$.

Exclusion limits are displayed for the two following mass scenarios:

1. Chargino mass always twice the neutralino mass, $m_{\tilde{\chi}_1^\pm} = 2m_{\tilde{\chi}_1^0}$. Limits set in the $(m_{\tilde{b}}, m_{\tilde{\chi}_1^0})$ -plane.
2. Neutralino mass is fixed to 60 GeV, and limits are set in the $(m_{\tilde{b}}, m_{\tilde{\chi}_1^\pm})$ -plane.

These cases are shown in Figure 5.32 and Figure 5.33 respectively. It can be seen that in both scenarios we exclude up to about $m_{\tilde{b}} = 460$ GeV, and that in each case SR1b is

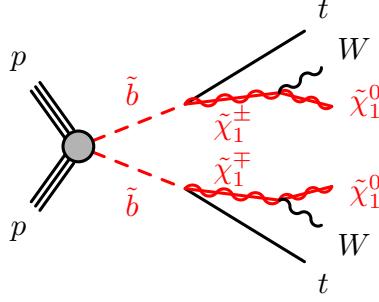


Figure 5.31: Feynman diagram demonstrating the direct sbottom simplified model. The diagram is taken from [61].

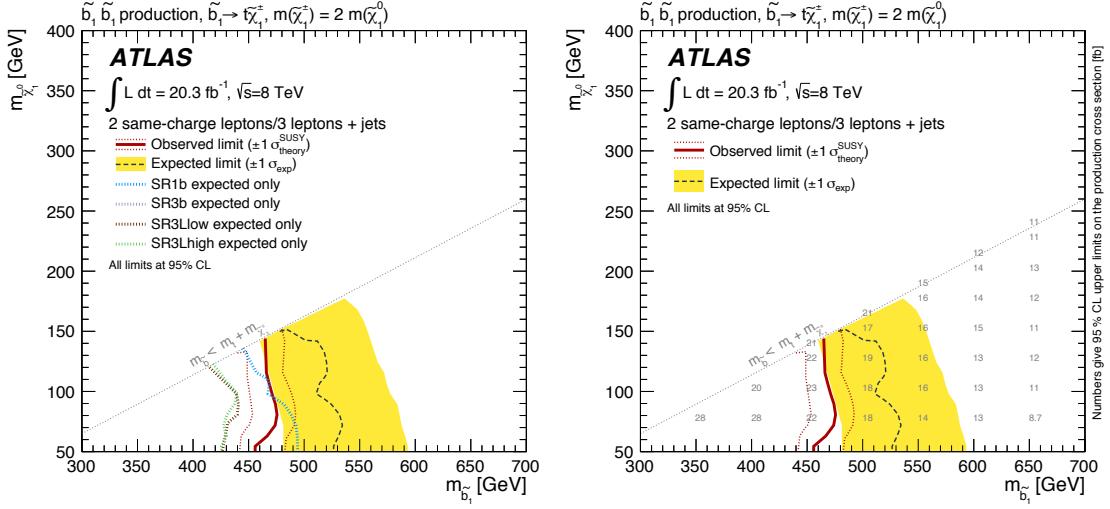


Figure 5.32: Exclusions for the direct sbottom simplified model in the case where $m_{\tilde{\chi}_1^\pm} = 2m_{\tilde{\chi}_1^0}$ always. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point.

the most constraining signal region. This is to be expected given that one expects two b -quarks in the final state.

5.9.2 Phenomenological models

In addition to the simplified scenarios considered above, the sensitivities to a small selection of models of a more ‘realistic’, phenomenologically viable, nature are investigated. There are three models that fit into the SUSY framework: mSUGRA/ constrained

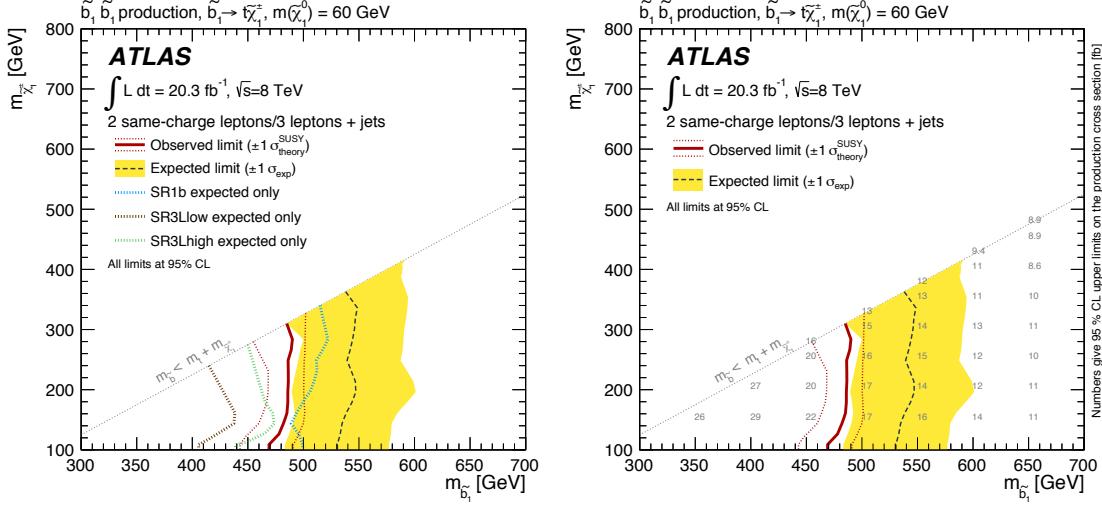


Figure 5.33: Exclusions for the direct sbottom simplified model in the case where $m_{\tilde{\chi}_1^0} = 60$ GeV always. The left hand plot includes a breakdown of sensitivity by individual signal region, whereas on the right is displayed the upper limit on the production cross section for each simulated point.

Minimal Supersymmetric Standard Model (cMSSM), GMSB, and bilinear RPV (bRPV). Additionally an exotic non-SUSY model is also presented, minimal Universal Extra Dimensions (mUED). These are described in the following sections.

mSUGRA/cMSSM The mSUGRA model, also known as the cMSSM[180–185], uses a hidden sector via which gravity can mediate supersymmetry breaking. It is defined as a subset of the MSSM, which has over 100 free parameters, and instead only allows five parameters to be varied:

- $\mathbf{m_0}$: the mass of all scalar particles at the grand unified theory (GUT) scale.
- $\mathbf{m_{1/2}}$: the mass of all gauginos at the GUT scale.
- $\mathbf{A_0}$: the trilinear coupling strengths.
- $\mathbf{\tan \beta}$: the ratio of the vacuum expectation values of the Higgs doublets.
- $\mathbf{\text{sgn } \mu}$: the sign of the SUSY Higgs mass parameter.

We set limits on a subset of this space, where $A_0 = -2m_0$, $\mu > 0$, and $\tan \beta = 30$. This combination of parameters allows for a Higgs mass in the range $122 < m_H < 128$ GeV, as is shown in the overlay in the right hand plot of Figure 5.34. Exclusions are set in

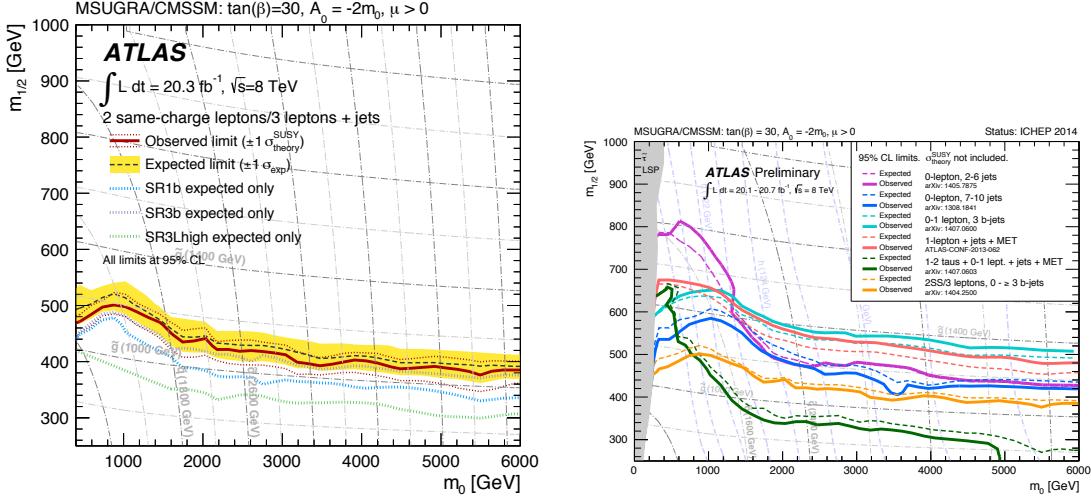


Figure 5.34: Exclusions in the mSUGRA/cMSSM scenario described in the text. The left hand plot includes a breakdown of sensitivity by individual signal region. On the right shows the current best public ATLAS limits in the same scenario [186]; this analysis is shown in orange.

the plane of the remaining two free parameters, m_0 and $m_{\frac{1}{2}}$, as can be seen in the same figure. Overall the most sensitive signal region is SR3b, although unlike in the simplified scenarios it is less obvious (and perhaps less useful) to consider which sub-processes are responsible for this.

bRPV The bRPV model [187] is based on mSUGRA as described previously, but additionally allows non-zero couplings for the bilinear terms in the superpotential in such a way that the LSP is unstable, and can decay to mixtures of leptons, quarks, and neutrinos [188]. Two example decay modes are shown in Figure 5.35.

The choice of the original mSUGRA parameters is the same as previously ($A_0 = -2m_0$, $\mu > 0$, and $\tan \beta = 30$, scan the $(m_0, m_{\frac{1}{2}})$ -plane), and the bilinear couplings are determined as a function of these under the tree-level-dominance scenario [189, 190]. The LSP decays within the detector, however for $m_{\frac{1}{2}} < 200$ GeV it has a sufficiently long lifetime such that the lepton acceptance by this analysis' criteria is significantly reduced – as such, these situations have not been considered. The exclusion is shown in the right hand plot of Figure 5.35.

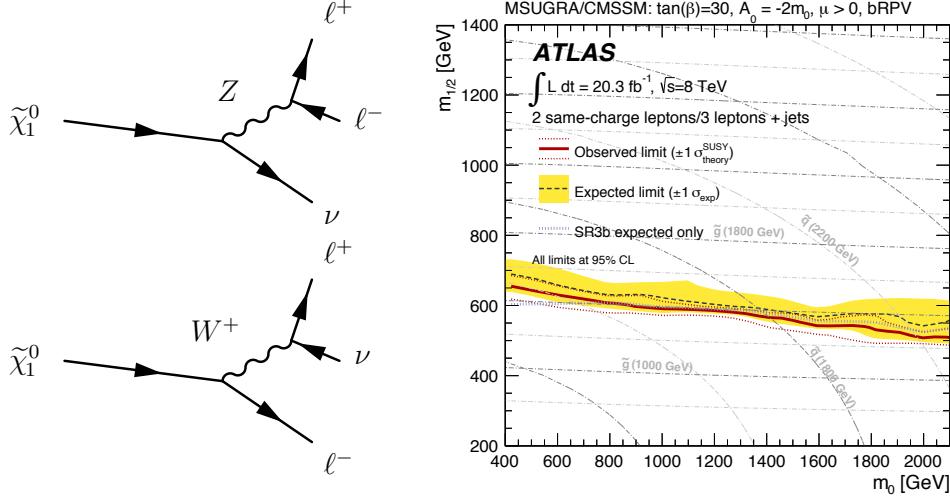


Figure 5.35: Example decays of the neutralino via RPV bilinear couplings (left), and exclusion (right) in the bRPV scenario. The exclusion plot includes a breakdown of sensitivity by individual signal region.

GMSB As mSUGRA performs the breaking of SUSY via the gravitational sector, GMSB achieves the same goal but via the SM’s gauge interactions [191–196]. It is a subset of the MSSM where new chiral supermultiplets are introduced, so-called ‘messenger fields’, that couple the MSSM to the source of SUSY breaking. Whilst still present, gravitational effects leading to SUSY breaking are overwhelmed by the gauge coupling. The model has six free parameters, which are:

- Λ : the SUSY-breaking mass scale.
- M_{mess} : the mass of the messenger field(s).
- N_5 : the number of $SU(5)$ messenger fields.
- $\tan \beta$: the ratio of vacuum expectation value (GUT)s of the Higgs doublets.
- $\text{sgn } \mu$: the sign of the SUSY Higgs mass parameter.
- C_{grav} : the scale factor for the gravitino mass.

Following [179, 197, 198], we fix $M_{\text{mess}} = 250$ TeV, $N_5 = 3$, $\mu > 0$ and $C_{\text{grav}} = 1$, with the exclusion then being presented in the $(\Lambda, \tan \beta)$ -plane, as displayed in Figure 5.36.

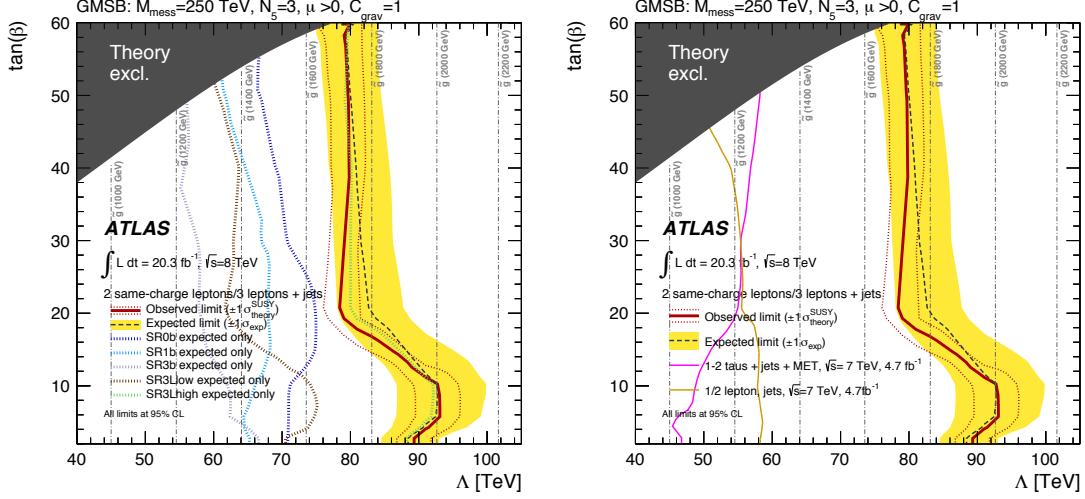


Figure 5.36: Exclusions in the GMSB scenario described in the text. The left hand plot includes a breakdown of sensitivity by individual signal region. On the right the same exclusion is shown but with comparison to ATLAS analyses with one or two taus [89], and one or two leptons [179]. The greyed out area is excluded theoretically, since it leads to the existence of tachyonic states.

mUED This is the only non-supersymmetric model considered here; mUED postulates the existence of an extra spatial dimension is postulated with compactification radius R , and cut-off scale Λ [199, 200]. The decays of the Kaluza-Klein quarks produce similar decay chains to squarks decaying down to the LSP, and as such this analysis has sensitivity to it. Exclusions are plotted in the plane of $(1/R, \Lambda R)$, as shown in Figure 5.37.

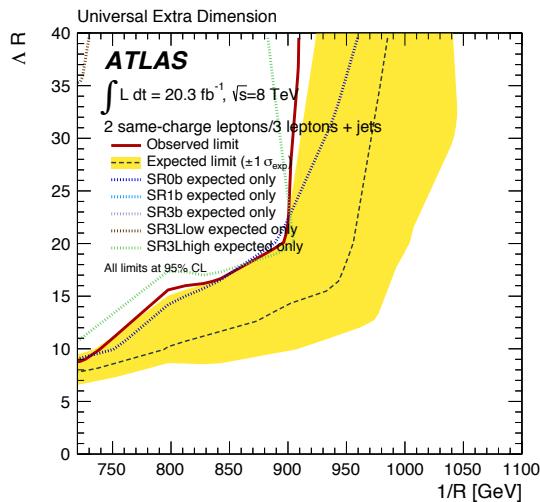


Figure 5.37: Exclusions in the mUED scenario. The plot includes a breakdown of sensitivity by individual signal region. No upper and lower bounds are displayed on the observed limit, since an uncertainty on the cross section calculation was unavailable.

Chapter 6

Comparing fake estimation methods

This chapter presents two sets of tests that have been performed on the fake estimation methods detailed in chapter 4. The first compares the matrix method to the two likelihood methods described in section 4.5, considering the effect on a limit setting procedure performed with a toy MC generator. It concludes that the likelihood methods are indeed more robust than the matrix method, but that more specifically the “approximate” form of the method using just the MLE fake rate and an estimated uncertainty is nearly as good as the “full” method, when performing the whole limit-setting procedure.

In light of this conclusion, since it follows *a priori* that the Bayesian sampling method should be no worse than the MLE method for a given model, a second comparison is then performed in a more realistic setting which pits the matrix method against the sampling method. From studies in ATLAS MC in two like-charge lepton regions similar to those studied in chapter 5, it is found that the Bayesian method outperforms the matrix method, except in very low statistics cases.

6.1 Toy MC: matrix method vs likelihood

This section will study the two likelihood methods, and perform a comparison with the matrix method. The study aims to test the full statistical procedure used in a search analysis, including setting a limit on a hypothetical signal sample (see section 4.5.1), and evaluate what differences, if any, exist in these final results when using the different fake estimation methods. The parameter model for the likelihood method used here is that depicted in Figure 4.1; whilst this particular choice is not necessarily expected to

necessarily generalise to a more realistic scenario, it is sufficient to draw some conclusions, particularly on the relative performance of the two variations of the likelihood method.

This section will use the ‘Method A, B, and C’ which was terminology introduced in section 4.5.3.

Using a toy event generator, datasets are produced using the same model shown in Figure 4.1. The parameters are set such that datasets are produced containing a mixture of fake and real ‘signal’ events. The values of these parameters are specified in the subsequent sections. For each of several configurations, 19000 independent datasets were formed using the generator. Each of these was subsequently processed using Methods A, B and C where possible – for the harder example it is found that Method B was computationally infeasible, and so results for Method B in this case are not shown. In all cases the necessary minimisation of a negative log likelihood is performed using the Minuit2 library [123]. The result are 95% CL_{s+b} and CL_s upper limits on the signal strength parameter.¹

6.1.1 Simple example – two leptons, two categories

Firstly, a configuration is used that produces events always with exactly two leptons, each of which can be in one of two categories. There are separate configurations for a signal process, which produces only real leptons ($\pi_1 = \pi_2 = 0$), and a fake process which produces only fake leptons ($\pi_1 = \pi_2 = 1$). The full set of parameters used in this example can be found in Table 6.1. In each dataset, 100 events are produced using the tree in Figure 4.1. As such the number of T events is approximately the sum of two Poisson random variables; one representing the signal component with mean 0.706, and another representing the fake background with a mean of 1.94.

The CL_{s+b} and CL_s limits from each of the 19000 generated datasets is shown in Figure 6.1. There is significant overcoverage in the CL_s limit, however this is expected due to the definition $CL_s = \frac{CL_{s+b}}{1-CL_b}$. In low statistics regimes, often $(1 - CL_b) < 1$, meaning that $CL_s > CL_{s+b}$ by a potentially significant margin. The CL_{s+b} limit is also seen to over-cover, particularly with Method C, and to a lesser extent with Method A, and least of all with Method B. It is indeed expected that Method B should have the

¹The p -values used to compute CL_s and CL_{s+b} are computed by performing pseudoexperiments, rather than using asymptotic methods [120], since it is known that the latter are only a good approximation for scenarios with a large number of events.

Object category	Signal			Background			Effs.	
	$\nu_{\bar{L}}$	β	π	$\nu_{\bar{L}}$	β	π	ε_r	ε_f
ω_1	0.01	0.6	0	0.99	0.6	1	0.8	0.1
ω_2	–	0.4	0	–	0.4	1	0.9	0.2

Table 6.1: Parameters controlling the simple example with exactly two leptons, and two categories for each lepton. The parameters are as described in Figure 4.1, however $\alpha_2 = 1$ and $\alpha_i = 0 \forall i \neq 2$. The overall production rate of events is $\nu_{\bar{L}}$, each one of which is filtered through the decision tree. Components marked with a ‘–’ are not applicable in the context.

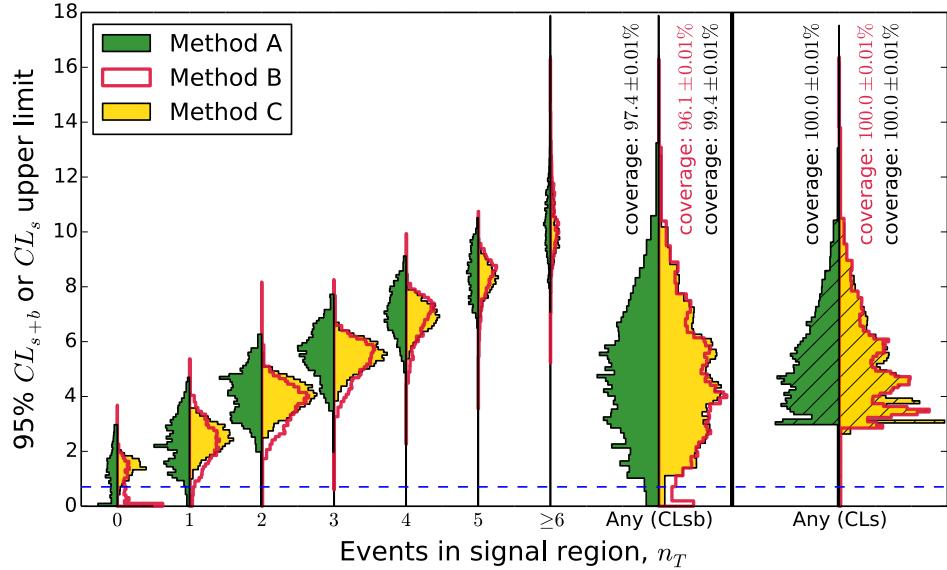


Figure 6.1: Using the two lepton, two category configuration, the 95% CL_{s+b} and CL_s upper limits on the rate of T signal events for each of 19000 independent toy datasets are histogrammed. For each ‘column’, histograms are made for each of Methods A, B and C, and plotted back-to-back. Method B and Method C are plotted overlapping on the right hand side of each column. The CL_{s+b} results are further divided into bins of observed n_T ; in all cases the area of each histogram is proportional to the number of toy datasets used to create it. The dashed blue line indicates the true signal production rate, $\nu_{TR} = 0.706$. The coverage of the observed limits of this truth rate are noted for the overall CL_{s+b} and CL_s results.

best behaviour, and any deviation from accurate coverage must be attributable to the approximation introduced in equation (4.22). Moreover, this figure also contains a division of the CL_{s+b} limit according to the number of events observed in the signal region, n_T . This demonstrates a clear, approximately linear, dependence of the limit on the number of events in the signal region for all methods; again this behaviour should be expected since the background estimate will be more tightly constrained by the (much larger) number of *loose* events, meaning that the remaining tight events are attributed to the signal. This is corroborated by Figure 6.3, discussed in more detail later, which shows that the measured fake rate is approximately independent of n_T .

Finally, Figure 6.2 demonstrates that, for $n_T \gtrsim 2$, limits for CL_s and CL_{s+b} are not biased towards being more constraining in any one of the methods. For lower n_T , Method B is biased towards placing the most aggressive limit, followed by Method A and then Method C is the least aggressive. CL_s exhibits much less of a spread at lower n_T , and in fact shows a tendency to have a *reduction* in the spread of the ratios for the lowest values of n_T .

Another interesting observation is the significantly larger spread, and the ‘fatter tails’ signified by the larger number of outliers, in the Method A vs Method B limit ratio plot than that for Method A and Method C. This can be explained by recognising that Method B makes the fullest use of the available information in the whole limit setting procedure, which evidently results in a greater tendency to have larger differences between its limit and that from Method A for any given event; Method C, being a half-way house has some discrepancy but less than that found in Method B.

A further comparison that can be made is of the fake rate that is the output of the matrix method in Method A, against the MLE of the fake rate obtained in Method B (by maximising the likelihood function) and Method C; such a comparison is made in Figure 6.3. This demonstrates the property that Method A can predict a negative fake rate, as seen in a handful of the generated datasets. It also shows that Method B and Method C produce fake rates that cluster more closely around the true value, even at low n_T . The larger fake rates that Method A predicts are responsible for the dip in ratio towards low n_T in Figure 6.2.

From this information, it can foremost be seen that, overall, rather similar limits are being placed by all methods. Method B and Method C have a tendency to produce more plausible MLEs of the fake rate, corresponding to slightly more tightly clustered limits in Figure 6.1. On the whole, there is a tendency to be very slightly more constraining

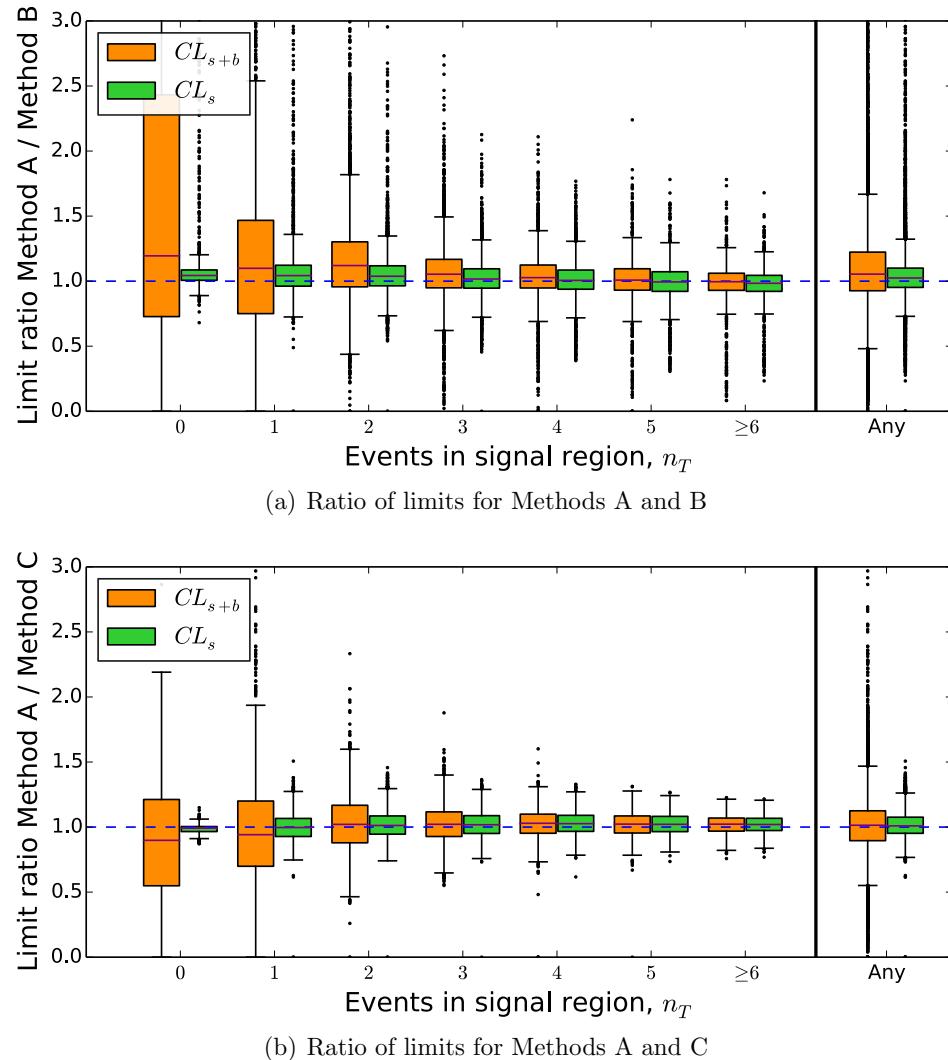


Figure 6.2: Using the two lepton, two category configuration, sub-figure (a) shows the ratio between the CL_{s+b} and CL_s limits obtained with Methods A and B respectively; the box plots are described further in Figure 6.3. Finally sub-figure (b) shows the same limit ratio for Methods A and C.

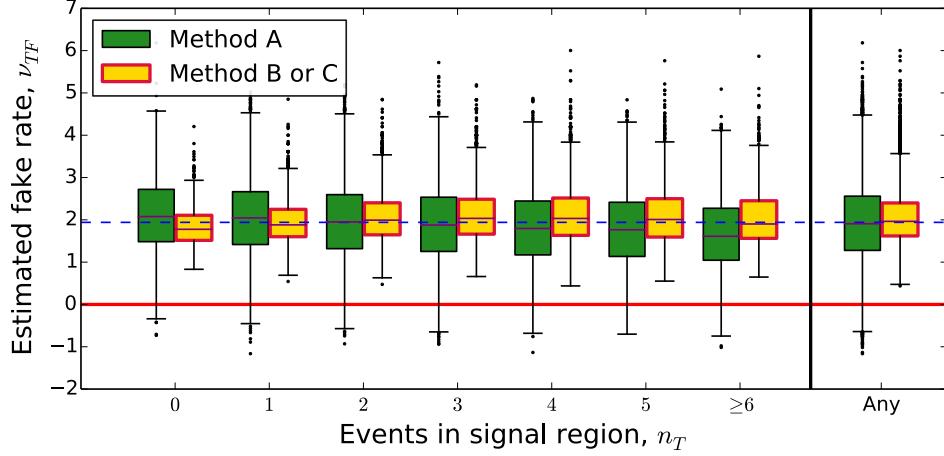


Figure 6.3: Using the two lepton, two category configuration, the estimated fake rate for each of 19000 independent toy datasets are shown as a function of n_T , comparing Methods A and B with box plots. These indicate the median and lower & upper quartiles with the box, while the whiskers extend to most extreme datum within $1.5 \times$ inter-quartile range of the nearest quartile; this corresponds to the $k = 1.5$ case as detailed in [201]. Black dots are used to mark data points outside the range of the whiskers. The dashed blue line marks the true value of $\nu_{TF} = 1.94$, and the red line delimits the unphysical $\nu_{TF} < 0$ region.

than one would expect to find from Method A. One therefore ought to slightly favour Method B in this scenario, all else being equal, to be confident that the limit one obtains is more likely to be representative of the limit one would *expect* to obtain from performing the experiment.

6.1.2 Harder example – two leptons, eight categories

The simple example of section 6.1.1 has been extended to use eight categories instead of two. As per the parametrisation being used, this involves the addition of 24 extra parameters – twelve each for the signal and fake background from the addition of six β and six π terms. These parameters can be referred to in Table 6.2. As before, 100 events were generated in each dataset, corresponding to a signal rate of 0.748 and a fake background rate of 2.77.

It was found that the increase in parameter space dimensionality was sufficient to increase the computation time for the minimisation to such an extent that producing

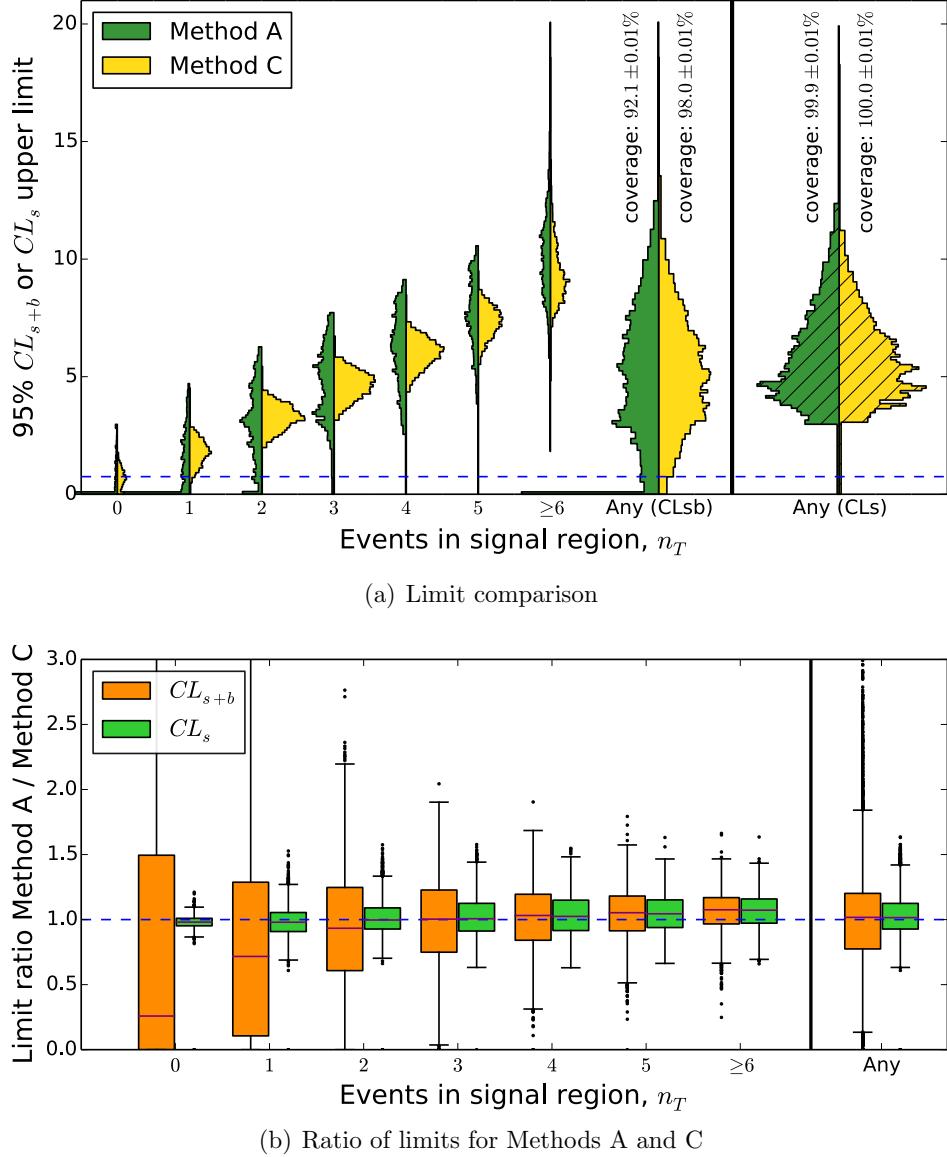


Figure 6.4: Using the two lepton, eight category configuration, the 95% CL_{s+b} and CL_s upper limits on the rate of T signal events for each of 19000 independent toy datasets are histogrammed, and plotted in (a). For each ‘column’, histograms are made for each of Methods A and C and plotted back-to-back. The CL_{s+b} results are further divided into bins of observed n_T ; in all cases the area of each histogram is proportional to the number of toy datasets used to create it. The dashed blue line indicates the true signal production rate, $\nu_{TR} = 0.748$. The coverage of the observed limits of this truth rate are noted for the overall CL_{s+b} and CL_s results. Sub-figure (b) shows the ratio between the CL_{s+b} and CL_s limits obtained with Methods A and C respectively; the box plots are described further in Figure 6.3.

Object category	Signal			Background			Effs.	
	$\nu_{\tilde{L}}$	β	π	$\nu_{\tilde{L}}$	β	π	ε_r	ε_f
ω_1	0.01	0.086	0	0.99	0.184	1	0.8	0.1
ω_2	–	0.143	0	–	0.008	1	0.8	0.2
ω_3	–	0.110	0	–	0.182	1	0.8	0.1
ω_4	–	0.010	0	–	0.123	1	0.8	0.3
ω_5	–	0.092	0	–	0.102	1	0.9	0.2
ω_6	–	0.284	0	–	0.081	1	0.9	0.1
ω_7	–	0.245	0	–	0.106	1	0.9	0.4
ω_8	–	0.030	0	–	0.214	1	0.9	0.1

Table 6.2: Parameters controlling the simple example with exactly two leptons, and eight categories for each lepton. The parameters are as described in Figure 4.1, however $\alpha_2 = 1$ and $\alpha_i = 0 \forall i \neq 2$. The overall production rate of events is $\nu_{\tilde{L}}$, each one of which is filtered through the decision tree. Components marked with a ‘–’ are not applicable in the context.

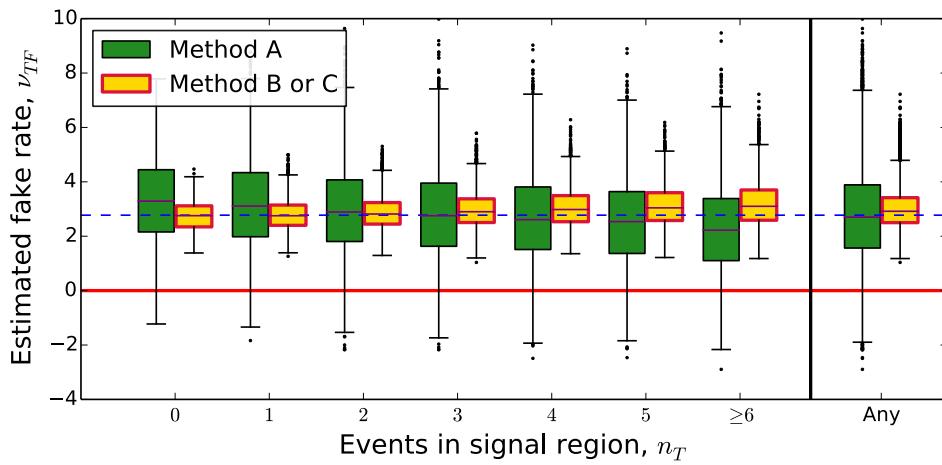


Figure 6.5: Using the two lepton, eight category configuration, the estimated fake rate for each of 19000 independent toy datasets are shown as a function of n_T , comparing Methods A and C with box plots similarly to Figure 6.3. The dashed blue line marks the true value of $\nu_{TF} = 2.77$, and the red line delimits the unphysical $\nu_{TF} < 0$ region.

limits with Method B became infeasible using the resources at the authors' disposal. As such only Method C is compared to Method A.

Figure 6.5 shows that the MLE fake rate for Method C is much more tightly constrained around the true value than the Method A estimate; moreover Method A gives even more significant deviations into negative values than with the simple example. Moreover, as n_T increases, the median fake rate from Method A decreases slightly, whereas that from Method C is stable for low event counts, only increasing slightly for larger n_T ; the Method C behaviour seems more desirable here. Secondly, Figure 6.4(a) shows that the CL_{s+b} limits derived in Method A suffer from undercoverage; the upper limit only bounds the true rate 92% of the time rather than the expected 95%. Finally the ‘upper tails’ of the CL_{s+b} limit are significantly more pronounced in Method A than in Method C, as can be seen when the limits are separated by n_T , as also included in Figure 6.4(a).

When additionally looking at Figure 6.4(b), one can see that, as in section 6.1.1, the CL_{s+b} ratio dips for low n_T and for the same reason. The new feature is the increase in ratio above 1 for both CL_{s+b} and CL_s limits for larger n_T – this corresponds to Method C systematically placing a more constraining limit than Method A. Therefore, both for this reason, and the greater consistency shown in Figure 6.4(a), one would favour Method C over Method A here.

6.1.3 Conclusions

Whilst the tests run in this section have been of a ‘toy’ nature, and might not accurately represent the behaviour in actual data, it does serve to demonstrate the drawbacks of Method A, and how both Method B and Method C could be expected to deliver a more accurate CL_{s+b} or CL_s upper limit. In particular, Method B can in some senses be considered to be the “best” limit setting procedure one can do, making full use of the available data. Method C has been shown by the author to be a much more readily used replacement for analyses perhaps already using Method A, due to significantly reduced computational complexity, whilst still providing rather similar behaviour to Method B in the tests conducted here.

6.2 ATLAS MC: matrix method vs Bayesian posterior

Having just demonstrated the theoretical performance of a likelihood-based fake estimation technique, we now turn to the Bayesian method described in section 4.6. Given that the previous section concluded that Method C – using just the MLE – gave comparable performance to limit setting with the full likelihood, we might hope that the Bayesian method will give another small improvement over this, in robustness if nothing else. Additionally, rather than using a toy MC generator, this section will use the ‘real world’ MC from the 8 TeV ATLAS like-charge lepton analysis, as described in section 5.2.2. As such the tests within this section should represent more accurately how one might expect the new method to perform in a real analysis.

6.2.1 Dealing with many categories

In the likelihood method tested previously, the issue of many categories was mitigated somewhat through use of a reduced parameter space in a tree hierarchy. This comes with two main problems – firstly, reducing the parameter space in any way implies that one believes some information is redundant; the assumptions made in Figure 4.1 are not necessarily generally applicable, which could cause the method to provide a poor fit in some situations. Secondly, the nature of this parametrisation makes the likelihood landscape rather “lumpy”, increasing the risk of a minimisation algorithm becoming stuck in local minima. Indeed, during preliminary tests it was found that the tree parametrisation was unusable given the number of categories in the like-charge analysis (in chapter 5) for this second reason.

The methodology described for the Bayesian sampling process in section 4.6 did not include any explicit method to solve the many category problem. One of the methods alluded to in section 3.5 shall hence be used – namely the combining of lepton categories.

Since ultimately we are interested in the integral of the fake and tight rate over all p_T and η in a given bin of a signal region, an attractive alternative proposition is simply to merge categories until we are left with a set of categories where each hopefully contains sufficient events to overwhelm the Gamma distribution priors. To a good approximation, the correct efficiencies for a merged category are an average of the efficiencies from the original categories, weighted by the number of real/fake leptons respectively for

the real/fake efficiencies. Of course, we do not have access to these numbers, and so approximations to them must be used.

In the remainder of this chapter it is noted that since, for at least the like-charge lepton regions considered, the measured real efficiencies are all large and the fake efficiencies all small, with relatively small variations between most categories, the weighting procedure can be performed to a reasonable approximation using the numbers of *tight* leptons to compute the average ε_r , and the numbers of *loose* leptons for ε_f . This should also be the case more generally, and as such this approximation should be fairly widely applicable.

6.2.2 Description of tests

The set of MC samples described in section 5.2.2 were used to model the SM contributions in a region requiring exactly two leptons of the same charge. More precisely, every event is required to have exactly two leptons with $p_T > 15$ GeV that are either tight or loose, and then a tight event will have both of its constituent leptons tight. This is done so as to simplify the problem – considering the more general problem with larger numbers of loose leptons is possible but introduces additional complexity. A truth cut is also performed to remove events accepted due to charge-flip (see section 5.5.3), thus allowing a cleaner test of the fake estimation procedures.

A region is then defined with $N_{b\text{-jets}} \geq 1$ and $N_{\text{jets}} \geq 2$, resembling the kinematics of the signal regions described in section 5.4, and then bins are formed in m_{eff} . Due to these similarities it is considered appropriate to use the real and fake efficiencies measured and shown in section 5.5.2. This assumption is later tested by also calculating the efficiencies for this region directly from the MC samples, in a form of closure test.

Having done this, the rates from various processes can be computed and histogrammed in this region, as has been done in Figure 6.6. This shows that $t\bar{t}+V$, diboson, and triboson processes are responsible for the real events, while the fake events come predominantly from $t\bar{t}$, as well as a small contribution from $Z+\text{jets}$. It is also clear that fake events occur much more frequently in the loose regions, whereas tight events are slightly favoured in the tight region.

In order to best simulate the application of the fake estimation methods to a real dataset, these MC samples are used to generate a set of pseudo-data. Precisely, a given MC sample contains a set of events each with a weight w_i . The sum $\sum_i w_i$ in a given region gives the expected rates normalised to luminosity, as shown in Figure 6.6.

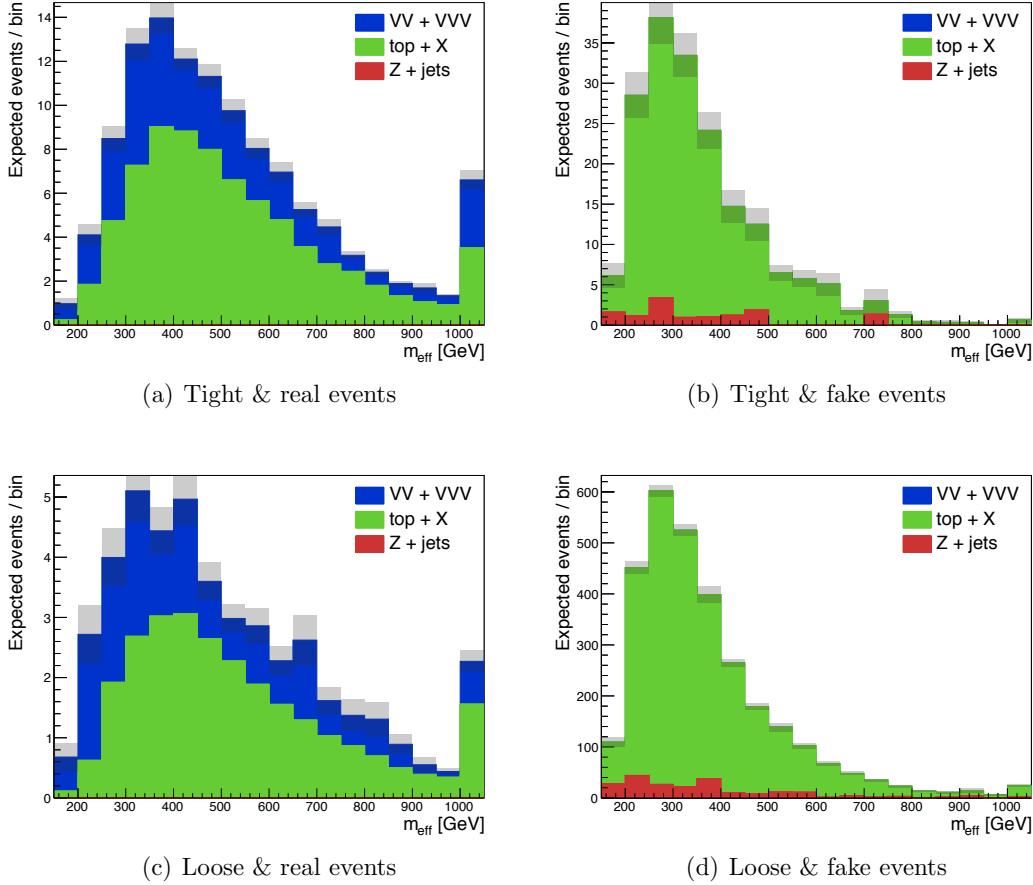


Figure 6.6: Plots of the expected rates of production of tight and loose events in each m_{eff} bin split by background process, normalised to correspond to a luminosity of 20.3 fb^{-1} . The left hand plots only include the production of ‘real’ events is considered (where both leptons are real), and conversely in the right hand plot only ‘fake’ events are considered (if either lepton is or both leptons are fake). The shaded band denotes the overall statistical uncertainty – systematic effects are not considered for the purposes of this study. The rightmost bin in each plot includes overflows.

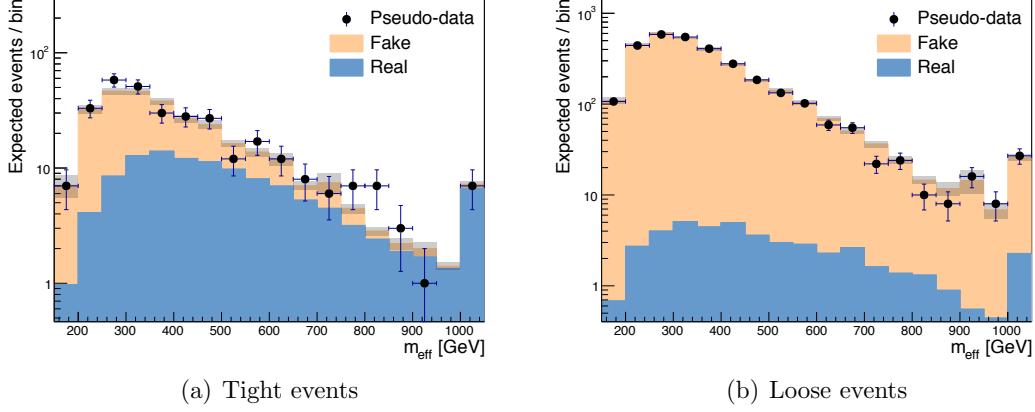


Figure 6.7: Plots of the expected rates of production of tight (left) and loose (right) events in each m_{eff} bin split by whether the events are real or fake, normalised to correspond to a luminosity of 20.3 fb^{-1} . The shaded band denotes the overall statistical uncertainty – systematic effects are not considered for the purposes of this study. The rightmost bin in each plot includes overflows. Additionally a pseudo-dataset is shown, computed as described in the text.

A pseudo-dataset is then generated using the properties of event i , with each event i included $n_i \sim \text{Poiss}(w_i)$ times. Since in most cases $w_i \ll 1$, it is rare for any event to be included more than once. Hence, to a good approximation, all events included in this pseudo-dataset are independent. The benefit of producing the test dataset in this fashion is that it is easy to retain the properties of individual events, *i.e.* information like the individual lepton p_T and η , which is required for fake estimation techniques.

An example of such a pseudo-dataset is shown in the region of interest in Figure 6.7. This plot additionally shows, as one expects, the tight region to have a much higher proportion of real events than the loose region. The goal of any fake estimation method will be to estimate the expected fake contribution in a given bin; this corresponds exactly to the cream-coloured area of the left-hand plot.

6.2.3 The status quo: matrix method

Firstly the generalised matrix method was applied to the pseudo-dataset; the results can be seen in Figure 6.8. From referring to the ratio plot on the right-hand side of the figure, there appears to be a small but systematic overestimate at low m_{eff} , and then a more significant overestimate in the highest few m_{eff} bins.

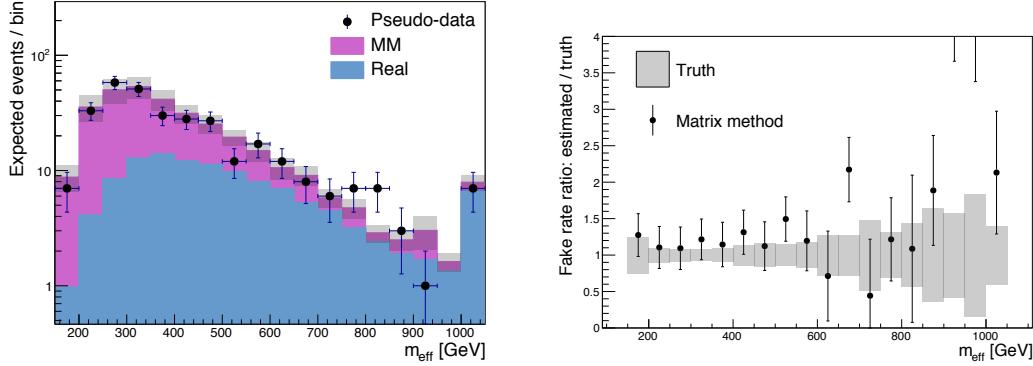


Figure 6.8: On the left-hand side, a plot of the pseudo-dataset of tight events overlaid on the nominal matrix method (MM) estimate of the fake and tight rate, using the efficiencies from section 5.5.2. The real rate plotted is that from MC, as would have been done if this were a real dataset. The right-hand plot shows the ratio of the matrix method prediction to the truth fake and tight rate. Uncertainties on both the truth and estimated fake rate are also scaled by a factor of $1/(\text{truth fake rate})$.

If there are two categories which can each be considered to have diagrams as in the Figure 3.1, the ‘correct’ way to combine the efficiencies for events with a single lepton is to take

$$\varepsilon_r = \frac{\nu_{R1}\varepsilon_{r1} + \nu_{R2}\varepsilon_{r2}}{\nu_{R1} + \nu_{R2}} \quad (6.1)$$

$$\varepsilon_f = \frac{\nu_{F1}\varepsilon_{f1} + \nu_{F2}\varepsilon_{f2}}{\nu_{F1} + \nu_{F2}}. \quad (6.2)$$

This is correct in the sense that one can then compute e.g. $\nu_T = \varepsilon_r\nu_R + \varepsilon_f\nu_F$ and find that it is the same as $\nu_{T1} + \nu_{T2}$ computed analogously. In practice we of course do not have access to these underlying rates, however it is noted that efficiencies do not vary too significantly in adjacent bins, so even an approximate weighting should be sufficient for practical purposes. In order to avoid any chicken-and-egg issues that might be created by trying to estimate the real and fake rates for the purpose of weighting efficiencies (which are required for said estimation), a very simple scheme is used here; both ε_r and ε_f are weighted using the total number of *leptons* in a given category. Naturally this procedure is also applicable to events with multiple leptons.

Using this heuristic method of category combination, the matrix method is applied three more times with different merging strategies. These are:

- All categories are merged into one category.

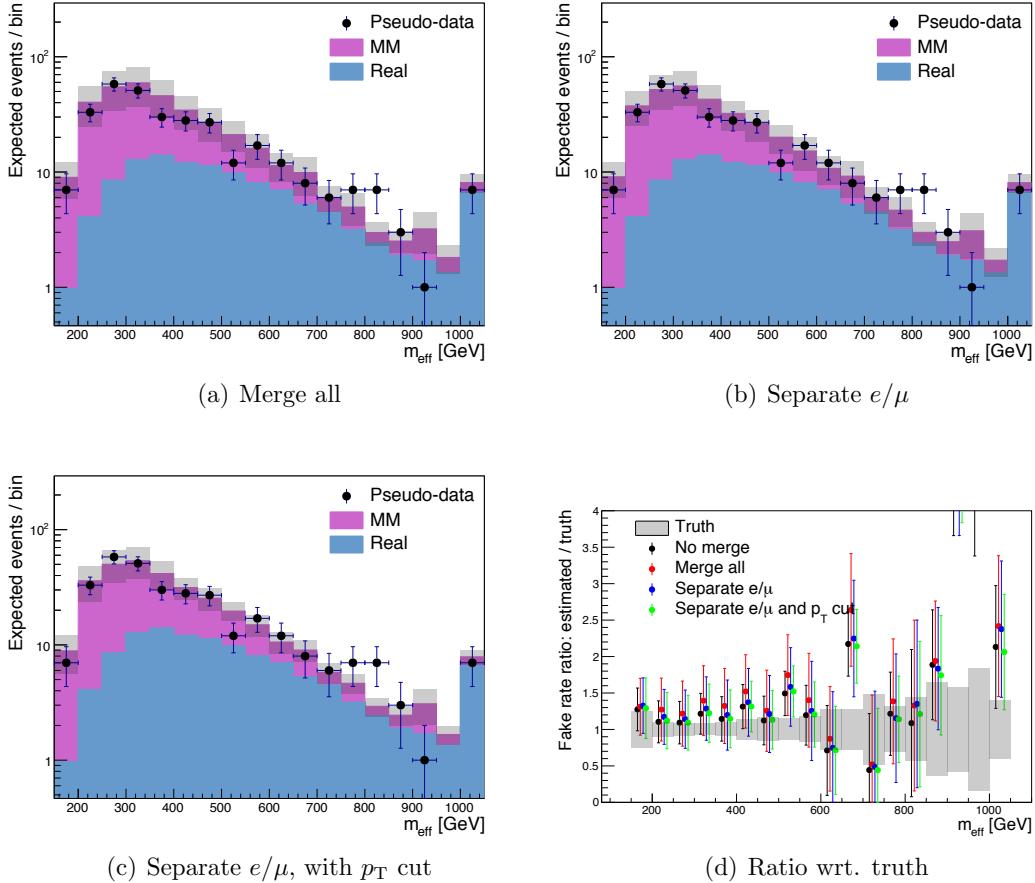


Figure 6.9: Plot of the pseudo-dataset of tight events overlaid on the matrix method (MM) estimate of the fake rate, merging efficiencies in the manners stated in the text. The real rate plotted is that from MC, as would be done if this were a real dataset. The last panel shows the ratio of each of the matrix method predictions to the truth fake and tight rate. Displacements of each point horizontally within each bin is for legibility reasons only. All uncertainties are also scaled by a factor of $1/(truth\ fake\ rate)$.

- All bins in p_T and η are averaged over, retaining only independent categories for electrons and muons.
- Independent categories are retained for electrons and muons, and within each there are two categories differing by requiring $p_T < 40$ GeV and $p_T \geq 40$ GeV respectively.

The findings can be seen in Figure 6.9, and are most readily apparent in the last plot, which shows the ratio of each of the four (nominal, and three merging schemes)

matrix method predictions to the truth fake and tight rate. Looking firstly at low m_{eff} , there is a slight systematic overestimate of the fake rate, which is worsened in the more aggressively merged efficiency scenarios. The least merged case has results which are very similar to the case without any merging, apart from the fact that they have slightly larger uncertainties. In general it appears that the additional merging only worsens the matrix method; uncertainties are larger, and agreement becomes worse. This suggests that the nature of the matrix method already produces an approximate, but appropriate, efficiency category merging procedure.

At large m_{eff} there is clearly a much larger discrepancy. However, this seems more likely attributable to a physics reason – namely that the efficiencies are no longer accurate in this region. The overall values of ε_r and ε_f are, respectively, estimated in the $950 < m_{\text{eff}} < 1000$ GeV bin to be 0.879 and 0.0898 respectively. This arises from merging the efficiencies using the approximate weighting scheme described above. These values can be compared to the truth efficiencies, computed using all the MC events in the same m_{eff} bin. This yields real and fake efficiencies of 0.851 and 0.0103 respectively. In conclusion, the fake efficiency is in fact much smaller in this region than was estimated – as such it is no surprise that the matrix method produces a significant overestimate. Further demonstrations that all issues in this region are rectified by using the correct efficiencies are included in section 6.2.5.

6.2.4 The Bayesian posterior

Using the sampling method described in section 4.6.2, the pseudo-data in each bin can be used to form the Bayesian posterior on the fake and tight rate. Examples of the distribution obtained in two m_{eff} bins can be seen in Figure 6.10. It is desired to extract a small but sufficient amount of information from these plots, both for visualisation in histograms as well as to pass on to any subsequent limit setting procedure.

The first point to note is that the distributions are asymmetric, particularly in the region with the lower predicted rate – as such simply computing the mean and standard deviation is inappropriate. Instead, I choose to use the median of the sample as a measure of centrality, along with the bounds of the middle 68% probability mass. These are visualised in Figure 6.10 as the dotted cyan line and shaded region, respectively.

One could, of course, choose to use an alternative measure, such as the maximum of the posterior distribution and some form of peak width from a fit. Whilst this *maximum*

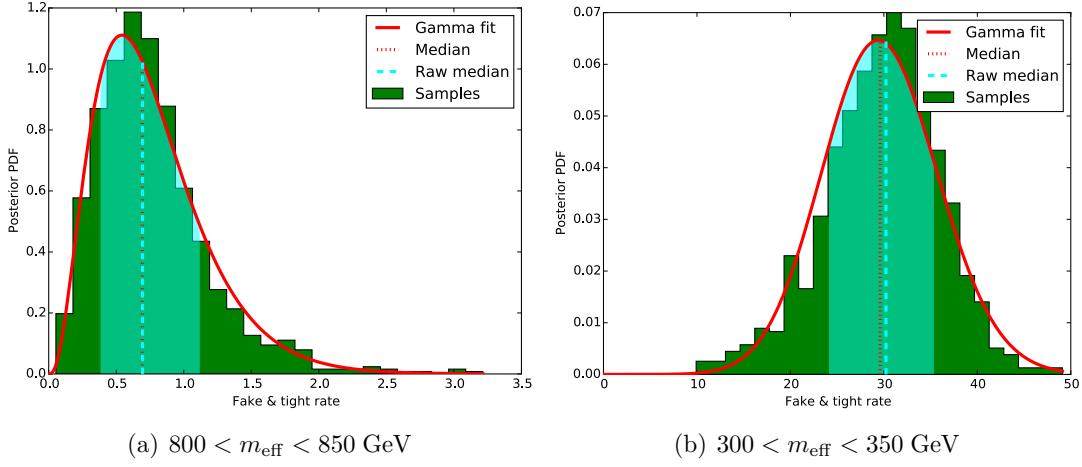


Figure 6.10: The Bayesian posterior distribution of the fake and tight rate, in the two-category merging scheme for two different m_{eff} bins. The green histogram represents a sample of 1000 draws using the Gibbs sampling mechanism with 25 steps of burn-in. Also shown is a fit of a Gamma distribution, with its median, as well as the median of the sampled data and shaded 68% probability mass region centred on it.

a posteriori technique is widely used in statistics, and can be viewed as a regularisation of a maximum likelihood method, it is avoided by the author. This is because it tends to throw away the additional information one has already procured in the form of the full posterior distribution; in practice it is used when computing the posterior distribution in full is computationally unviable. The results shown in the later plots would change if an alternative centrality and width measure were used, however given that the distribution in Figure 6.10 is representative of the posteriors found in these tests, the mean, median, and mode are all within the 68% credible interval.

The plots also include a fit of a Gamma distribution to the drawn samples. Whilst it is a good fit for the $800 < m_{\text{eff}} < 850 \text{ GeV}$ bin, where there is a longer positive tail, the $300 < m_{\text{eff}} < 350 \text{ GeV}$ bin has a longer *negative* tail, for which the Gamma distribution can be seen to not fit so well. As such this fit is only provided for illustrative purposes, and the median and $\pm 1\sigma$ (68%) credible interval of the raw sample is propagated forwards.²

²Frameworks such as HistFitter deal with asymmetric uncertainties by means of two-sided Gaussian distributions. To a good approximation the 68% credible interval could be used to define such ‘up’ and ‘down’ uncertainties.

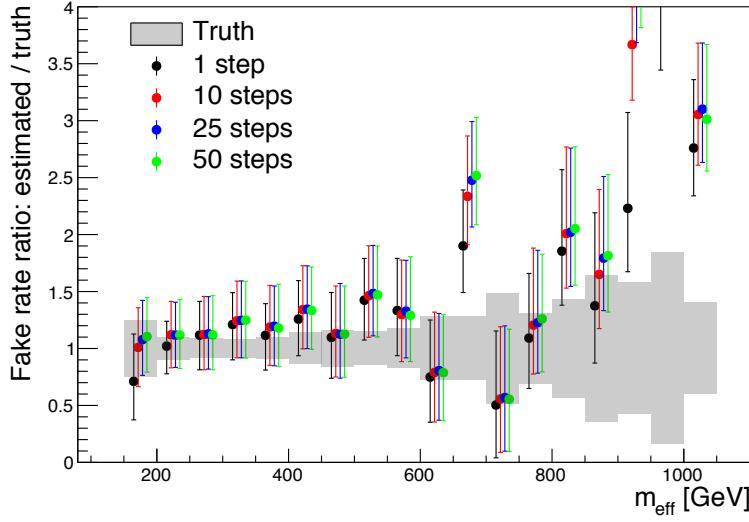


Figure 6.11: The ratios of estimates of the fake and tight rate using the sampling method to the truth value. Efficiency categories are merged into one region each for electrons and muons. Each data series uses a different number of burn in steps, as noted in the legend.

Determining the optimal amount of burn in In performing these fits 25 Gibbs steps were used for burn in, before drawing a sample. This number was determined experimentally as follows. It is known that the more steps that are taken, the closer the method is to sampling from the true distribution – as such the burn-in period should be chosen such that few significant changes occur if the period is increased further. One can hence compute the limits across all m_{eff} bins for several choices of burn in length, and pick the value that seems to have the best trade-off of accuracy against performance. The results of doing this are presented in Figure 6.11, which shows the ratios of estimates to truth for several choices of burn in. It can be seen that changes tend to become small after having taken 25 steps, but are more sizeable before that.

Results As has already been discussed, the Bayesian sampling method suffers from the presence of many more event categories than events to support them, since unlike the matrix method it doesn't come with an in-built merging scheme. As such only the performance of the three merged schemes described previously is considered. For each bin in m_{eff} the median and 68% credible interval are used to plot the central value and asymmetric uncertainty bands, as shown in Figure 6.12.

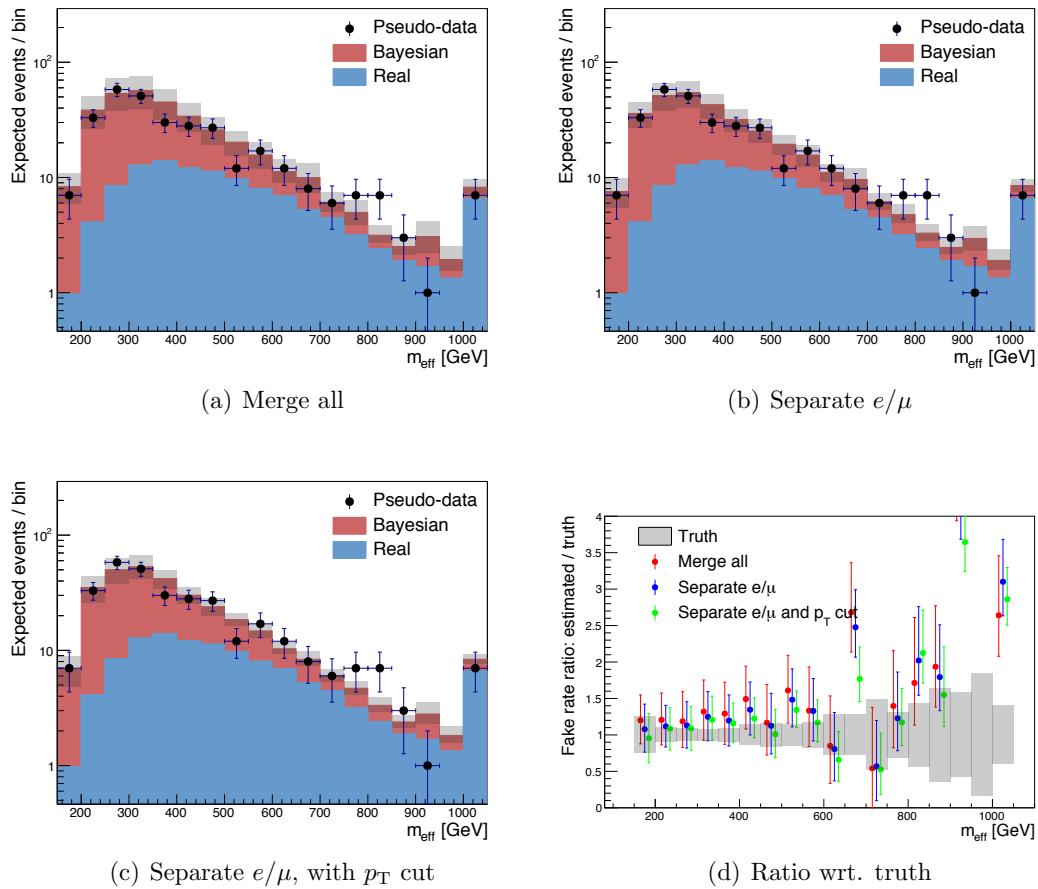


Figure 6.12: Plot of the pseudo-dataset of tight events overlaid on the Bayesian posterior estimate of the fake rate, merging efficiencies in the manners stated in the text. The real rate plotted is that from MC, as would be done if this were a real dataset. The last panel shows the ratio of each of the matrix method predictions to the truth fake and tight rate. Displacements of each point horizontally within each bin is for legibility reasons only. All uncertainties are also scaled by a factor of $1/(truth\ fake\ rate)$.

Similarly to the matrix method, there is relative stability in the fake rate estimates between all three efficiency merging schemes tested here. As before with the matrix method, at low m_{eff} there is a slight bias towards overestimating which improves as more finely separated efficiency categories are used. There is also still a large discrepancy at high m_{eff} , which seems likely due to the measured efficiencies being inappropriate in this region. Comparing back to Figure 6.9 it can be seen that the uncertainty bands for the Bayesian posterior are considerably narrower than those for the matrix method across the board, whilst still seeming to give appropriate coverage of the truth value (with the exception of the aforementioned significantly discrepant regions). This seems to be the main improvement gained by using the Bayesian method over the matrix method.

6.2.5 Comparisons with ‘truth’ efficiencies

Another test of interest is to look again at the predictions in the test region, but this time to use real and fake efficiencies computed directly from the MC samples, which hence correspond to the true underlying values. The three merging schemes defined previously are again used, where for any given scheme the truth efficiencies are determined by counting all pertinent events from the whole set of MC samples (rather than just the subset that forms the pseudo-dataset). Strictly speaking the efficiencies are not merged for these tests – rather they are separately calculated in each of the bins defined by each merging scenario. Both the matrix and sampling methods are then used on the pseudo-dataset in each case, with results shown in Figure 6.13.

From these plots it can be seen that both methods have similar behaviour. However, there are a few differences worth highlighting:

- **Uncertainties:** As before, the Bayesian estimate has smaller uncertainties than that from the matrix method, which for the most part also seem appropriate for the observed agreement with the truth. The effect is also more pronounced when using a finer binning of categories.
- **Coping with uncertain efficiencies:** In the m_{eff} bins above about 800 GeV, the efficiencies are being estimated with increasingly low statistics samples, and as such have larger uncertainties. Increasing the number of categories for a given m_{eff} bin will only increase these uncertainties – as such looking at the relative behaviour of the matrix method and Bayesian posterior method in any given bin as the number of categories increases can be instructive. In the $750 < m_{\text{eff}} < 800$ GeV bin, for

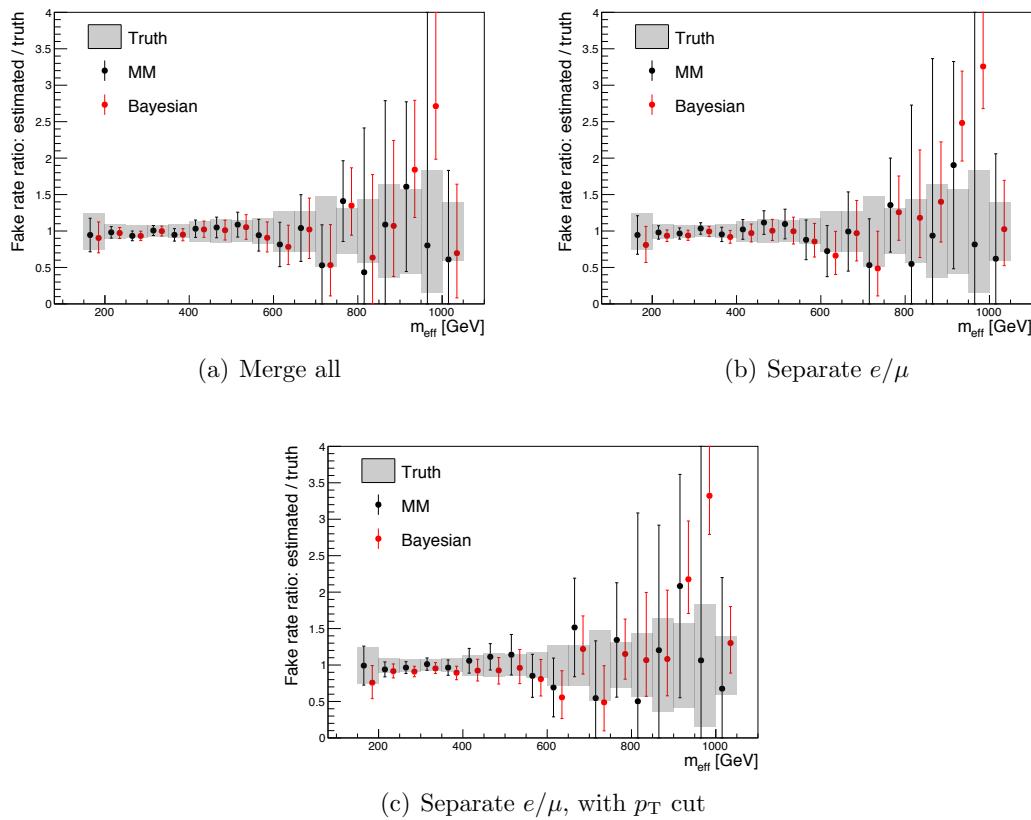


Figure 6.13: Ratios of both sampling and matrix method fake and tight rate predictions to the truth value, when using efficiencies computed from the overall MC sample from which the pseudo-dataset has been drawn. These efficiencies are computed separately in several broad bins, as have been discussed previously.

example, the uncertainty on the matrix method estimate increases, whereas that on the Bayesian estimate is relatively stable.

- **Low-stats behaviour:** In the 3rd and 2nd last bins, the Bayesian method starts to overestimate the fake rate. The overestimation becomes more prevalent with more efficiency categories. These bins have the lowest statistics, with only a handful of events per bin, and as such there will be several event categories with just one event. In these categories the estimated fake rate will be very prior dependent. The matrix method produces an estimate which agrees well with truth, although with a very large estimated uncertainty (particularly in the $950 < m_{\text{eff}} < 1000$ GeV bin).

Given these features, it seems that the Bayesian method is clearly superior when there are sufficient statistics to overwhelm the priors of each category independently, in that it produces more precise estimates, and can cope with uncertain efficiencies somewhat better than the matrix method.

However, in the case of extremely low statistics regions, which would include the signal regions of the like-charge ATLAS analysis, it is less clear that the Bayesian method can be recommended. The treating of each category independently is its undoing, since with few events the answer is almost entirely dependent on the choice of prior. On the other hand the matrix method has an implicit combination of the different regions which appears to allow it to produce more accurate answers here.

6.3 Conclusions

In this chapter it has been set out to verify the claims made in chapter 4 that a likelihood-based or Bayesian approach can be more robust than the matrix method, by using tests in both toy and ATLAS MC.³

The essential points to take away are:

- With sufficient statistics, say $\mathcal{O}(10)$ events in a two lepton signal region, a Bayesian or likelihood method will outperform the matrix method.
- With very low statistics, treating each event category separately is not possible – this is why the Bayesian method fails. The matrix method is less biased, but the

³ In this context a ‘likelihood method’ refers to Method C as defined in Table 4.1, and the ‘matrix method’ is Method A. Due to the speed issues with Method B it is not mentioned in these conclusions. The Bayesian technique does not map onto any of the lines in Table 4.1.

best results are presented in the first half of this chapter when using an appropriate reduced parametrisation.

- Merging efficiency categories, in addition to improving predictive power, also significantly reduces computation time for likelihood and Bayesian methods. It however does not provide any benefit for the matrix method.

Whilst the new methods present do somewhat mitigate the many-category problem described in section 3.4.3, it has clearly not been an unmitigated success. Further improvements would either need to tackle the concept of adding smoothness constraints between event categories, or the problem needs to be approached from a new, as yet unconsidered, angle.

Chapter 7

Closing remarks

This thesis set out on a two-pronged mission: both to improve the methods of estimating the contributions of fake events in ATLAS signal regions, as well as to use these methods in a search for SUSY.

With respect to the former of these, three distinct methods were presented – an extended version of the matrix method, a maximum likelihood method, and a Bayesian method.

The extension to the matrix method was the start of my interest in this particular background. It was designed to be of particular use to the SUSY search that was presented; namely to be able to deal seamlessly with events containing varying numbers of leptons, even when the selection could depend on which leptons in the event were in fact real. However, both the likelihood and Bayesian methods are certainly more statistically well-defined than the matrix method, as was demonstrated from a theoretical perspective. Both of these were also shown to offer modest improvements in robustness and accuracy over the matrix method in simulations. Apart from suffering a lack of predictivity in regions with very low event counts, these new methods seem suitable as replacements for the matrix method.

The SUSY search for like-charge leptons was shown to place competitive, sometimes the best, limits in several SUSY scenarios, as well as provide model-independent exclusions for the kinematic regions in which it searched.

The simplified signatures investigated in section 5.9.1 are used across ATLAS, and where appropriate in the given sections comparisons were made to the exclusions from other analyses. Whilst not always setting the strongest limit, a particularly beneficial feature is the complementarity of the results presented here to other ATLAS SUSY

analyses. For example, in the gluino-mediated stop scenario (Figure 5.21) the like-charge analysis extends the limit closer to the diagonal, despite the fact that the multijet analysis has a greater reach in the regions with less compressed spectra. A similar complementarity can be seen in Figure 5.27. The analysis also proved its wider applicability through its strong exclusion of the GMSB model, for which it was not optimised – at the time of writing it is the strongest such public exclusion on this scenario. This application to hitherto untested or unforeseen models is enabled through the publication of the model-independent limits; these can be used by the phenomenological community to test the viability of new models which are expected to produce events with like-charge lepton signatures.

Overall, in many of the simplified SUSY scenarios considered the existence of gluinos of masses up to $\mathcal{O}(1 \text{ TeV})$ was excluded. These limits are starting to push the bounds of the masses we would expect to observe given naturalness constraints [202]. Whilst theorists are now identifying unexplored areas of parameter space where SUSY could still be hiding, one must surely start to be concerned at the fine-tuning such theories often require.

Colophon

This thesis was made in L^AT_EX 2 _{ε} using the “hepthesis” class [203].

Bibliography

- [1] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** (2008), no. 08 S08001.
- [2] “CERN accelerator complex.”
<http://te-dep-epc.web.cern.ch/te-dep-epc/machines/general.stm>.
- [3] “ATLAS Luminosity Public Results.”
<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- [4] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [5] ATLAS Collaboration, *ATLAS: Detector and physics performance technical design report. Volume 2*, .
- [6] ATLAS Collaboration, *ATLAS: Detector and physics performance technical design report. Volume 1*, .
- [7] A. J. Barr and C. G. Lester, *A search for direct heffalon production using the ATLAS and CMS experiments at the Large Hadron Collider*, [arXiv:1303.7367](https://arxiv.org/abs/1303.7367).
- [8] N. Wermes and G. Hallewel, *ATLAS pixel detector: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1998.
- [9] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 1*, .
- [10] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 2*, .
- [11] D. Cote, O. A. Ducu, T. P. S. Gillam, J. Maurer, T. J. Mueller, and J. Poveda, *Impact of TRT transition radiation detection in the search for SUSY with two same-sign leptons or three leptons at $\sqrt{s} = 8$ TeV*, Tech. Rep. ATL-COM-PHYS-2013-1550, CERN, Geneva, Nov, 2013.
- [12] ATLAS Collaboration, *ATLAS calorimeter performance: Technical Design Report*. CERN, Geneva, 1996.

- [13] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. CERN, Geneva, 1997.
- [14] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512.
- [15] ATLAS Collaboration, *Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data*, ATLAS-CONF-2013-088, <https://cds.cern.ch/record/1580207> (2013).
- [16] W. Buttinger, *The ATLAS Level-1 Trigger System*, *Journal of Physics: Conference Series* **396** (2012), no. 1 012010.
- [17] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **1106** (2011) 128, [arXiv:1106.0522].
- [18] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175].
- [19] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, [arXiv:1005.4568].
- [20] GEANT4 Collaboration, *GEANT4: A simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250–303.
- [21] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, <https://cds.cern.ch/record/1300517> (2010).
- [22] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, Tech. Rep. ATL-SOFT-PROC-2012-065, CERN, Geneva, Jun, 2012.
- [23] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton–proton collision data*, *Eur. Phys. J. C* **72** (2012) 1909, [arXiv:1110.3174].
- [24] T. G. Cornelissen, N. Van Eldik, M. Elsing, W. Liebig, E. Moyse, N. Piacquadio, K. Prokofiev, A. Salzburger, and A. Wildauer, *Updates of the ATLAS Tracking Event Data Model (Release 13)*, Tech. Rep. ATL-SOFT-PUB-2007-003. ATL-COM-SOFT-2007-008, CERN, Geneva, Jun, 2007.
- [25] “ATLAS Muon Combined Performance Guidelines for Analyses of 2012 Data.” https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCPAnalysisGuidelinesData2012#Muon_Types.

- [26] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- $k(t)$ jet clustering algorithm*, *JHEP* **0804** (2008) 063, [[arXiv:0802.1189](https://arxiv.org/abs/0802.1189)].
- [27] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur.Phys.J.* **C72** (2012) 1896, [[arXiv:1111.6097](https://arxiv.org/abs/1111.6097)].
- [28] 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*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008.
- [29] T. Barillari, E. Bergeaas Kuutmann, T. Carli, J. Erdmann, P. Giovannini, K. J. Grahn, C. Issever, A. Jantsch, A. Kiryunin, K. Lohwasser, A. Maslennikov, S. Menke, H. Oberlack, G. Pospelov, E. Rauter, P. Schacht, F. Span, P. Speckmayer, P. Stavina, and P. Strzenec, *Local Hadronic Calibration*, Tech. Rep. ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001, CERN, Geneva, Jun, 2008.
Due to a report-number conflict with another document, the report-number ATL-LARG-PUB-2009-001-2 has been assigned.
- [30] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *Eur.Phys.J.* **C73** (2013) 2304, [[arXiv:1112.6426](https://arxiv.org/abs/1112.6426)].
- [31] B. R. Martin and G. Shaw, *Particle Physics*. Wiley, 2nd ed., 1997.
- [32] C. Burgess and G. Moore, *The Standard Model, A Primer*. Cambridge University Press, 2007.
- [33] J. Binney and D. Skinner, *The Physics of Quantum Mechanics*. Cappella Archive, 1st ed., 2008.
- [34] A. Zee, *Quantum Field Theory in a Nutshell*. Princeton University Press, 2nd ed., 2010.
- [35] S. P. Martin, *A Supersymmetry primer*, *Adv.Ser.Direct.High Energy Phys.* **21** (2010) 1–153, [[hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356)].
- [36] M. Bertolini, “Lectures on Supersymmetry.”
<http://people.sissa.it/~bertmat/teaching.htm>.
- [37] K. Stelle, *String Theory, Unification and Quantum Gravity*, *Lect.Notes Phys.* **863** (2013) 3–30, [[arXiv:1203.4689](https://arxiv.org/abs/1203.4689)].
- [38] Particle Data Group, *Review of particle physics*, *Phys. Rev. D* **86** (Jul, 2012) 010001.
- [39] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik,

- D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, *Astrophysical Journal Supplement* **192** (Feb., 2011) 18, [[arXiv:1001.4538](https://arxiv.org/abs/1001.4538)].
- [40] B. A. Reid, W. J. Percival, D. J. Eisenstein, L. Verde, D. N. Spergel, R. A. Skibba, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. R. Gott, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi, *Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies*, *Monthly Notices of the Royal Astronomical Society* **404** (May, 2010) 60–85, [[arXiv:0907.1659](https://arxiv.org/abs/0907.1659)].
- [41] M. S. Madhavacheril, N. Sehgal, and T. R. Slatyer, *Current Dark Matter Annihilation Constraints from CMB and Low-Redshift Data*, *Phys.Rev.* **D89** (2014) 103508, [[arXiv:1310.3815](https://arxiv.org/abs/1310.3815)].
- [42] LUX Collaboration, *The Large Underground Xenon (LUX) experiment, Nuclear Instruments and Methods in Physics Research A* **704** (Mar., 2013) 111–126, [[arXiv:1211.3788](https://arxiv.org/abs/1211.3788)].
- [43] LUX Collaboration, *First results from the LUX dark matter experiment at the Sanford Underground Research Facility*, *Phys.Rev.Lett.* **112** (2014), no. 9 091303, [[arXiv:1310.8214](https://arxiv.org/abs/1310.8214)].
- [44] P. Cushman, C. Galbiati, D. McKinsey, H. Robertson, T. Tait, et al., *Working Group Report: WIMP Dark Matter Direct Detection*, [arXiv:1310.8327](https://arxiv.org/abs/1310.8327).
- [45] L. Canetti, M. Drewes, and M. Shaposhnikov, *Matter and Antimatter in the Universe*, *New J.Phys.* **14** (2012) 095012, [[arXiv:1204.4186](https://arxiv.org/abs/1204.4186)].
- [46] F. R. Klinkhamer and N. Manton, *A Saddle Point Solution in the Weinberg-Salam Theory*, *Phys.Rev.* **D30** (1984) 2212.
- [47] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, *Progress of Theoretical Physics* **28** (1962), no. 5 870–880, [<http://ptp.oxfordjournals.org/content/28/5/870.full.pdf+html>].
- [48] Muon G-2 Collaboration, *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, *Phys.Rev.* **D73** (2006) 072003, [[hep-ex/0602035](https://arxiv.org/abs/hep-ex/0602035)].
- [49] A. Bellerive, *Review of solar neutrino experiments*, *Int.J.Mod.Phys.* **A19** (2004)

- 1167–1179, [[hep-ex/0312045](#)].
- [50] A. Y. Smirnov, *The MSW effect and solar neutrinos*, [hep-ph/0305106](#).
- [51] C. Mariani, *Review of Reactor Neutrino Oscillation Experiments*, *Mod.Phys.Lett.* **A27** (2012) 1230010, [[arXiv:1201.6665](#)].
- [52] OPERA Collaboration, *Search for $\nu_\mu \rightarrow \nu_e$ oscillations with the OPERA experiment in the CNGS beam*, *JHEP* **1307** (2013) 004, [[arXiv:1303.3953](#)].
- [53] F. Klinkhamer, *Neutrino mass and the Standard Model*, *Mod.Phys.Lett.* **A28** (2013) 1350010, [[arXiv:1112.2669](#)].
- [54] S. M. Bilenky and S. T. Petcov, *Massive neutrinos and neutrino oscillations*, *Rev. Mod. Phys.* **59** (Jul, 1987) 671–754.
- [55] M. Drewes, *The Phenomenology of Right Handed Neutrinos*, *Int.J.Mod.Phys.* **E22** (2013) 1330019, [[arXiv:1303.6912](#)].
- [56] S. Coleman and J. Mandula, *All possible symmetries of the S matrix*, *Phys. Rev.* **159** (Jul, 1967) 1251–1256.
- [57] R. Haag, J. T. Lopuszański, and M. Sohnius, *All possible generators of supersymmetries of the S-matrix*, *Nuclear Physics B* **88** (1975), no. 2 257 – 274.
- [58] “Higgs and supersymmetric higgs phenomenology.”
http://www.physics.gla.ac.uk/~dmiller/doc/YETI08_DJM_short.pdf, 2008.
- [59] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys.Lett.* **B716** (2012) 1–29, [[arXiv:1207.7214](#)].
- [60] MSSM Working Group, *The Minimal supersymmetric standard model: Group summary report*, [hep-ph/9901246](#).
- [61] “SUSY Feynman Diagrams.” <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/SUSYFeynmanDiagrams>.
- [62] N. Arkani-Hamed, P. Schuster, N. Toro, J. Thaler, L.-T. Wang, et al., *MARMOSET: The Path from LHC Data to the New Standard Model via On-Shell Effective Theories*, [hep-ph/0703088](#).
- [63] J. Barnard and B. Farmer, *A simple technique for combining simplified models and its application to direct stop production*, *JHEP* **1406** (2014) 132, [[arXiv:1402.3298](#)].
- [64] D. Stockinger, *The Muon Magnetic Moment and Supersymmetry*, *J.Phys.* **G34** (2007)

- R45–R92, [[hep-ph/0609168](#)].
- [65] J. Ellis, A. Mustafayev, and K. A. Olive, *Constrained Supersymmetric Flipped SU(5) GUT Phenomenology*, *Eur.Phys.J.* **C71** (2011) 1689, [[arXiv:1103.5140](#)].
- [66] D. Miller and A. Morais, *Supersymmetric SO(10) Grand Unification at the LHC and Beyond*, *JHEP* **1412** (2014) 132, [[arXiv:1408.3013](#)].
- [67] M. Dugan, B. Grinstein, and L. J. Hall, *CP Violation in the Minimal N=1 Supergravity Theory*, *Nucl.Phys.* **B255** (1985) 413.
- [68] C. Brust, A. Katz, and R. Sundrum, *SUSY Stops at a Bump*, *JHEP* **1208** (2012) 059, [[arXiv:1206.2353](#)].
- [69] L. Maiani, *The GIM Mechanism: origin, predictions and recent uses*, [arXiv:1303.6154](#).
- [70] L. Susskind, *Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory*, *Phys. Rev. D* **20** (Nov, 1979) 2619–2625.
- [71] C. Baker, D. Doyle, P. Geltenbort, K. Green, M. van der Grinten, et al., *An Improved experimental limit on the electric dipole moment of the neutron*, *Phys.Rev.Lett.* **97** (2006) 131801, [[hep-ex/0602020](#)].
- [72] LHCb Collaboration, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment*, *Phys.Rev.Lett.* **111** (2013) 101805, [[arXiv:1307.5024](#)].
- [73] CMS Collaboration, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ with the CMS Experiment*, *Phys.Rev.Lett.* **111** (2013) 101804, [[arXiv:1307.5025](#)].
- [74] A. Arbey, M. Battaglia, F. Mahmoudi, and D. Martinez Santos, *Supersymmetry confronts $B_s^0 \rightarrow \mu^+ \mu^-$: Present and future status*, *Phys.Rev.* **D87** (2013) 035026, [[arXiv:1212.4887](#)].
- [75] K. S. Cranmer, *Kernel estimation in high-energy physics*, *Comput.Phys.Commun.* **136** (2001) 198–207, [[hep-ex/0011057](#)].
- [76] ATLAS Collaboration, *Background studies for top-pair production in lepton plus jets final states in $\sqrt{s} = 7$ TeV ATLAS data*, Tech. Rep. ATLAS-CONF-2010-087, CERN, Geneva, Oct, 2010.
- [77] ATLAS Collaboration, *Search for supersymmetry using final states with one lepton, jets, and missing transverse momentum with the ATLAS detector in $\sqrt{s} = 7$ TeV pp*, *Phys.Rev.Lett.* **106** (2011) 131802, [[arXiv:1102.2357](#)].

- [78] ATLAS Collaboration, *Search for an excess of events with an identical flavour lepton pair and significant missing transverse momentum in $\sqrt{s} = 7$ TeV proton-proton collisions with the ATLAS detector*, *Eur.Phys.J.* **C71** (2011) 1647, [[arXiv:1103.6208](#)].
- [79] ATLAS Collaboration, *Search for supersymmetric particles in events with lepton pairs and large missing transverse momentum in $\sqrt{s} = 7$ TeV proton-proton collisions with the ATLAS experiment*, *Eur.Phys.J.* **C71** (2011) 1682, [[arXiv:1103.6214](#)].
- [80] ATLAS Collaboration, *Search for pair production of first or second generation leptoquarks in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, *Phys.Rev.* **D83** (2011) 112006, [[arXiv:1104.4481](#)].
- [81] ATLAS Collaboration, *Measurement of the cross section for the production of a W boson in association with b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys.Lett.* **B707** (2012) 418–437, [[arXiv:1109.1470](#)].
- [82] ATLAS Collaboration, *Search for anomalous production of prompt like-sign muon pairs and constraints on physics beyond the Standard Model with the ATLAS detector*, *Phys.Rev.* **D85** (2012) 032004, [[arXiv:1201.1091](#)].
- [83] ATLAS Collaboration, *Search for pair-produced heavy quarks decaying to Wq in the two-lepton channel at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys.Rev.* **D86** (2012) 012007, [[arXiv:1202.3389](#)].
- [84] ATLAS Collaboration, *Observation of spin correlation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *Phys.Rev.Lett.* **108** (2012) 212001, [[arXiv:1203.4081](#)].
- [85] ATLAS Collaboration, *A search for $t\bar{t}$ resonances with the ATLAS detector in 2.05 fb^{-1} of proton-proton collisions at $\sqrt{s} = 7$ TeV*, *Eur.Phys.J.* **C72** (2012) 2083, [[arXiv:1205.5371](#)].
- [86] ATLAS Collaboration, *A search for $t\bar{t}$ resonances in lepton+jets events with highly boosted top quarks collected in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **1209** (2012) 041, [[arXiv:1207.2409](#)].
- [87] ATLAS Collaboration, *ATLAS search for a heavy gauge boson decaying to a charged lepton and a neutrino in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur.Phys.J.* **C72** (2012) 2241, [[arXiv:1209.4446](#)].
- [88] ATLAS Collaboration, *Search for resonant top plus jet production in $t\bar{t} + \text{jets}$ events with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys.Rev.* **D86** (2012) 091103, [[arXiv:1209.6593](#)].

- [89] ATLAS Collaboration, *Search for Supersymmetry in Events with Large Missing Transverse Momentum, Jets, and at Least One Tau Lepton in 7 TeV Proton-Proton Collision Data with the ATLAS Detector*, *Eur. Phys. J.* **C 72** (2012) 2215, [[arXiv:1210.1314](#)].
- [90] ATLAS Collaboration, *Measurement of ZZ production in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous ZZZ and ZZ γ couplings with the ATLAS detector*, *JHEP* **1303** (2013) 128, [[arXiv:1211.6096](#)].
- [91] ATLAS Collaboration, *Multi-channel search for squarks and gluinos in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS detector*, *Eur.Phys.J.* **C73** (2013) 2362, [[arXiv:1212.6149](#)].
- [92] ATLAS Collaboration, *Search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV*, *Phys.Rev.* **D88** (2013), no. 1 012004, [[arXiv:1305.2756](#)].
- [93] ATLAS Collaboration, *Measurement of Top Quark Polarization in Top-Antitop Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector*, *Phys.Rev.Lett.* **111** (2013), no. 23 232002, [[arXiv:1307.6511](#)].
- [94] ATLAS Collaboration, *Measurement of the top quark pair production charge asymmetry in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **1402** (2014) 107, [[arXiv:1311.6724](#)].
- [95] ATLAS Collaboration, *Search for a Multi-Higgs Boson Cascade in $W^+Wb\bar{b}$ events with the ATLAS detector in pp collisions at $s = 8$ TeV*, *Phys.Rev.* **D89** (2014) 032002, [[arXiv:1312.1956](#)].
- [96] ATLAS Collaboration, *Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **1405** (2014) 068, [[arXiv:1402.6263](#)].
- [97] ATLAS Collaboration, *Search for direct top-squark pair production in final states with two leptons in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **1406** (2014) 124, [[arXiv:1403.4853](#)].
- [98] ATLAS Collaboration, *Search for direct top squark pair production in events with a Z boson, b-jets and missing transverse momentum in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, *Eur.Phys.J.* **C74** (2014) 2883, [[arXiv:1403.5222](#)].
- [99] ATLAS Collaboration, *Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **1405** (2014) 071, [[arXiv:1403.5294](#)].

- [100] ATLAS Collaboration, *Search for supersymmetry at $\sqrt{s} = 8$ TeV in final states with jets and two same-sign leptons or three leptons with the ATLAS detector*, [arXiv:1404.2500](#).
- [101] ATLAS Collaboration, *Search for microscopic black holes and string balls in final states with leptons and jets with the ATLAS detector at $\sqrt{s} = 8$ TeV*, *JHEP* **1408** (2014) 103, [[arXiv:1405.4254](#)].
- [102] ATLAS Collaboration, *Search for top squark pair production in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, [arXiv:1407.0583](#).
- [103] ATLAS Collaboration, *Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three b-jets at $\sqrt{s} = 8$ TeV proton-proton collisions with the ATLAS detector*, [arXiv:1407.0600](#).
- [104] ATLAS Collaboration, *Search for supersymmetry in events with large missing transverse momentum, jets, and at least one tau lepton in 20 fb^{-1} of $\sqrt{s} = 8$ TeV proton-proton collision data with the ATLAS detector*, [arXiv:1407.0603](#).
- [105] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section as a function of jet multiplicity and jet transverse momentum in 7 TeV proton-proton collisions with the ATLAS detector*, [arXiv:1407.0891](#).
- [106] CMS Collaboration, *Measurement of the top-quark mass in $t\bar{t}$ events with dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur.Phys.J.* **C72** (2012) 2202, [[arXiv:1209.2393](#)].
- [107] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, *JINST* **8** (2013) P04013, [[arXiv:1211.4462](#)].
- [108] CMS Collaboration, *Search for a Higgs boson decaying into a b-quark pair and produced in association with b quarks in proton-proton collisions at 7 TeV*, *Phys.Lett.* **B722** (2013) 207–232, [[arXiv:1302.2892](#)].
- [109] CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **1211** (2012) 067, [[arXiv:1208.2671](#)].
- [110] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002.
- [111] ATLAS Collaboration, *Search for gluinos in events with two same-sign leptons, jets and missing transverse momentum with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys.Rev.Lett.* **108** (2012) 241802, [[arXiv:1203.5763](#)].
- [112] ATLAS Collaboration, “Search for Supersymmetry in final states with two same-sign

- leptons, jets and missing transverse momentum with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV.” ATLAS-CONF-2012-105, 2012.
- [113] ATLAS Collaboration, “Search for strongly produced superpartners in final states with two same sign leptons with the ATLAS detector using 21 fb^{-1} of proton-proton collisions at $\sqrt{s} = 8$ TeV.” ATLAS-CONF-2013-007, 2013.
- [114] A. J. Laub, *Matrix Analysis For Scientists And Engineers*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.
- [115] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, *Guide to the Expression of Uncertainty in Measurement*. International Organization for Standardization, 1995.
- [116] A. L. Read, *Presentation of search results: The $CL(s)$ technique*, *J.Phys.* **G28** (2002) 2693–2704.
- [117] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl.Instrum.Meth.* **A434** (1999) 435–443, [[hep-ex/9902006](#)].
- [118] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, *Transactions of the American Mathematical Society* **54** (1943), no. 3 pp. 426–482.
- [119] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *The Annals of Mathematical Statistics* **9** (03, 1938) 60–62.
- [120] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *European Physical Journal C* **71** (Feb., 2011) 1554, [[arXiv:1007.1727](#)].
- [121] G. Zech, *Upper limits in experiments with background or measurement errors*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **277** (1989), no. 23 608 – 610.
- [122] T. P. S. Gillam and C. G. Lester, *Improving estimates of the number of ‘fake’ leptons and other mis-reconstructed objects in hadron collider events: BoB’s your UNCLE*, *JHEP* **1411** (2014) 031, [[arXiv:1407.5624](#)].
- [123] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput.Phys.Commun.* **10** (1975) 343–367.
- [124] K. Cranmer, *Frequentist Hypothesis Testing with Background Uncertainty*, in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology* (L. Lyons, R. Mount, and R. Reitmeyer, eds.), p. 261, 2003. [physics/0310108](#).

- [125] Kyle Cranmer, ATLAS Collaboration. Private communication, 2013.
- [126] James Beacham, ATLAS Collaboration. Private communication, 2013.
- [127] N. Hyvönen, J. Heino, and V. Kaarnioja, “Computational methods in inverse problems.” https://noppa.aalto.fi/noppa/kurssi/mat-1.3626/luennot/Mat-1_3626_lecture12.pdf.
- [128] “Boost C++ libraries.” <http://www.boost.org>.
- [129] Wolfram MathWorld, “Simplex Method.” <http://mathworld.wolfram.com/SimplexMethod.html>.
- [130] “GNU Linear Programming Kit.” <https://www.gnu.org/software/glpk/>.
- [131] “Luminosity Uncertainty.” <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/DataMCForAnalysis#Luminosity>.
- [132] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ using the ATLAS detector at the LHC*, *Eur. Phys. J. C* **73** (2013) 2518, [[arXiv:1302.4393](https://arxiv.org/abs/1302.4393)].
- [133] ATLAS Collaboration, *Updated Luminosity Determination in pp Collisions at $\sqrt{s} = 7 \text{ TeV}$ using the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2011-011, CERN, Geneva, Mar, 2011.
- [134] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [[arXiv:0710.3820](https://arxiv.org/abs/0710.3820)].
- [135] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [[arXiv:0811.4622](https://arxiv.org/abs/0811.4622)].
- [136] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [[arXiv:0709.2092](https://arxiv.org/abs/0709.2092)].
- [137] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001, [[hep-ph/0206293](https://arxiv.org/abs/hep-ph/0206293)].
- [138] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010, [[hep-ph/0011363](https://arxiv.org/abs/hep-ph/0011363)].
- [139] J. Butterworth, J. Forshaw, and M. Seymour, *Multiparton interactions in photoproduction at hera*, *Z. Phys. C* **72** (1996) 637–646, [[hep-ph/9601371](https://arxiv.org/abs/hep-ph/9601371)].
- [140] S. Hoeche and *et al.*, *NLO matrix elements and truncated showers*, *JHEP* **08** (2011) 123,

- [hep-ph/1009.1127].
- [141] S. Frixione, B. R. Webber, and P. Nason, *MC@NLO generator version 3.4*, hep-ph/0204244 and hep-ph/0305252.
- [142] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, *Comput.Phys.Commun.* **184** (2013) 919–985, [hep-ph/0405247].
- [143] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024, [arXiv:1007.2241].
- [144] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012, [hep-ph/0201195].
- [145] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^\pm$ production and decay at NLO*, *JHEP* **07** (2012) 052, [arXiv:1204.5678].
- [146] M. Garzelli, A. Kardos, C. Papadopoulos, and Z. Trocsanyi, *$t\bar{t}W^\pm$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, *JHEP* **11** (2012) 056, [arXiv:1208.2665].
- [147] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **07** (2011) 018, [arXiv:1105.0020].
- [148] M. Bähr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639, [arXiv:0803.0883].
- [149] W. Beenakker, R. Hopker, M. Spira, and P. Zerwas, *Squark and gluino production at hadron colliders*, *Nucl. Phys. B* **492** (1997) 51, [hep-ph/9610490].
- [150] A. Kulesza and L. Motyka, *Threshold resummation for squark-antisquark and gluino-pair production at the LHC*, *Phys. Rev. Lett.* **102** (2009) 111802, [arXiv:0807.2405].
- [151] A. Kulesza and L. Motyka, *Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC*, *Phys. Rev. D* **80** (2009) 095004, [arXiv:0905.4749].
- [152] W. Beenakker et al., *Soft-gluon resummation for squark and gluino hadroproduction*, *JHEP* **12** (2009) 041, [arXiv:0909.4418].
- [153] W. Beenakker et al., *Squark and gluino hadroproduction*, *Int. J. Mod. Phys. A* **26** (2011) 2637, [arXiv:1105.1110].
- [154] M. Krämer et al., *Supersymmetry production cross sections in pp collisions at*

- $\sqrt{s} = 7 \text{ TeV}$, arXiv:1206.2892.
- [155] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and PYTHIA 8 for MC11*, ATL-PHYS-PUB-2011-009, <http://cds.cern.ch/record/1363300> (2011).
- [156] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D 82** (2010) 074018, [arXiv:1005.3457].
- [157] S. Gieseke, C. Rohr, and A. Siodmok, *Colour reconnections in Herwig++*, Eur. Phys. J. **C 72** (2012) 2225, [arXiv:1206.0041].
- [158] “Physics Analysis Tools, ExtendedPileUpReweighting.” <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ExtendedPileupReweighting>.
- [159] S. Ask, D. Berge, P. Borrego-Amaral, D. Caracinha, N. Ellis, P. Farthouat, P. Glbn, S. Haas, J. Haller, P. Klofver, A. Krasznahorkay, A. Messina, C. Ohm, T. Pauly, M. Perantoni, H. P. L. Junior, G. Schuler, D. Sherman, R. Spiwoks, T. Wengler, J. M. de Seixas, and R. T. Teixeira, *The ATLAS central level-1 trigger logic and TTC system*, Journal of Instrumentation **3** (2008), no. 08 P08002.
- [160] ATLAS Collaboration, *Search for SUSY using events with three leptons, multiple jets, and missing transverse momentum at $\sqrt{s} = 8 \text{ TeV}$* , tech. rep., 2012.
- [161] ATLAS Collaboration, *Search for supersymmetry using events with three leptons, multiple jets, and missing transverse momentum in 13.0 fb^{-1} of pp collisions with the ATLAS detector at $\sqrt{s} = 8 \text{ TeV}$* , tech. rep., 2012.
- [162] A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill series in electrical and computer engineering. McGraw-Hill Education, 2002.
- [163] ATLAS Collaboration, *Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three b-jets using 20.1 fb^{-1} of pp collisions at $\sqrt{s} = 8 \text{ TeV}$ with the ATLAS Detector.*, Tech. Rep. ATLAS-CONF-2013-061, CERN, Geneva, Jun, 2013.
- [164] “Flavour Tagging Performance Group, Calibration Data Interface.” <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BTaggingCalibrationDataInterface>.
- [165] ATLAS Collaboration, *Measuring the b-tag efficiency in a top-pair sample with 4.7 fb^{-1} of data from the ATLAS detector*, Tech. Rep. ATLAS-CONF-2012-097, CERN, Geneva, Jul, 2012.
- [166] ATLAS Collaboration, *Single hadron response measurement and calorimeter jet energy*

- scale uncertainty with the ATLAS detector at the LHC,* *Eur. Phys. J. C* **73** (2013) 2305, [[arXiv:1203.1302](https://arxiv.org/abs/1203.1302)].
- [167] 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** (2013) 2306, [[arXiv:1210.6210](https://arxiv.org/abs/1210.6210)].
- [168] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment,* ATLAS-CONF-2014-004, <https://cdsweb.cern.ch/record/1664335> (2014).
- [169] 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,* ATLAS-CONF-2012-043, <https://cds.cern.ch/record/1435197> (2012).
- [170] ATLAS Collaboration, *Measurement of the Mistag Rate of b-tagging algorithms with 5 fb^{-1} of Data Collected by the ATLAS Detector,* ATLAS-CONF-2012-040, <https://cds.cern.ch/record/1435194> (2012).
- [171] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS,* *Eur. Phys. J. C* **72** (2012) 1844, [[arXiv:1108.5602](https://arxiv.org/abs/1108.5602)].
- [172] M. Baak, G. Besjes, D. Cote, A. Koutsman, J. Lorenz, et al., *HistFitter software framework for statistical data analysis*, [arXiv:1410.1280](https://arxiv.org/abs/1410.1280).
- [173] L. Moneta et al., *The RooStats Project*, [arXiv:1009.1003](https://arxiv.org/abs/1009.1003).
- [174] “HistFactory package.” <https://twiki.cern.ch/twiki/pub/RooStats/WebHome/HisFactoryLikelihood.pdf>.
- [175] ATLAS Collaboration, *Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS experiment,* *JHEP* **10** (2013) 130, [[arXiv:1308.1841](https://arxiv.org/abs/1308.1841)].
- [176] ATLAS Collaboration, *Search for top and bottom squarks from gluino pair production in final states with missing transverse energy and at least three b-jets with the ATLAS detector,* *Eur. Phys. J. C* **72** (2012) 2174, [[arXiv:1207.4686](https://arxiv.org/abs/1207.4686)].
- [177] B. Allanach and B. Gripaios, *Hide and Seek With Natural Supersymmetry at the LHC,* *JHEP* **1205** (2012) 062, [[arXiv:1202.6616](https://arxiv.org/abs/1202.6616)].
- [178] ATLAS Collaboration, *Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using 4.7 fb^{-1} of $\sqrt{s} = 7$ TeV*

- proton-proton collision data, Phys. Rev. **D** **87** (2013) 012008, [[arXiv:1208.0949](#)].*
- [179] ATLAS Collaboration, *Further search for supersymmetry at $\sqrt{s} = 7$ TeV in final states with jets, missing transverse momentum and isolated leptons with the ATLAS detector, Phys. Rev. **D** **86** (2012) 092002, [[arXiv:1208.4688](#)].*
- [180] A. H. Chamseddine, R. L. Arnowitt, and P. Nath, *Locally Supersymmetric Grand Unification, Phys. Rev. Lett. **49** (1982) 970.*
- [181] R. Barbieri, S. Ferrara, and C. A. Savoy, *Gauge Models with Spontaneously Broken Local Supersymmetry, Phys. Lett. **B** **119** (1982) 343.*
- [182] L. E. Ibanez, *Locally Supersymmetric SU(5) Grand Unification, Phys. Lett. **B** **118** (1982) 73.*
- [183] L. J. Hall, J. D. Lykken, and S. Weinberg, *Supergravity as the Messenger of Supersymmetry Breaking, Phys. Rev. **D** **27** (1983) 2359.*
- [184] N. Ohta, *Grand Unified Theories Based on Local Supersymmetry, Prog. Theor. Phys. **70** (1983) 542.*
- [185] G. L. Kane, C. F. Kolda, L. Roszkowski, and J. D. Wells, *Study of constrained minimal supersymmetry, Phys. Rev. **D** **49** (1994) 6173–6210, [[hep-ph/9312272](#)].*
- [186] “ATLAS Supersymmetry Public Results.” <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults>.
- [187] S. Roy and B. Mukhopadhyaya, *Some implications of a supersymmetric model with R-parity breaking bilinear interactions, Phys. Rev. **D** **55** (1997) 7020, [[hep-ph/9612447](#)].*
- [188] W. Porod, M. Hirsch, J. Romao, and J. Valle, *Testing neutrino mixing at future collider experiments, Phys. Rev. **D** **63** (2001) 115004, [[hep-ph/0011248](#)].*
- [189] Y. Grossman and S. Rakshit, *Neutrino masses in R-parity violating supersymmetric models, Phys. Rev. **D** **69** (2004) 093002, [[hep-ph/0311310](#)].*
- [190] D. Carvalho, M. Gomez, and J. Romao, *Charged lepton flavor violation in supersymmetry with bilinear R-parity violation, Phys. Rev. **D** **65** (2002) 093013, [[hep-ph/0202054](#)].*
- [191] M. Dine and W. Fischler, *A Phenomenological Model of Particle Physics Based on Supersymmetry, Phys. Lett. **B** **110** (1982) 227.*
- [192] L. Alvarez-Gaume, M. Claudson, and M. Wise, *Low-energy supersymmetry, Nucl. Phys. **B** **207** (1982) 96.*
- [193] C. R. Nappi and B. A. Ovrut, *Supersymmetric Extension of the $SU(3) \times SU(2) \times U(1)$*

- Model*, *Phys. Lett.* **B 113** (1982) 175.
- [194] M. Dine and A. E. Nelson, *Dynamical supersymmetry breaking at low-energies*, *Phys. Rev.* **D 48** (1993) 1277.
- [195] M. Dine, A. E. Nelson, and Y. Shirman, *Low-energy dynamical supersymmetry breaking simplified*, *Phys. Rev.* **D 51** (1995) 1362.
- [196] M. Dine, A. E. Nelson, Y. Nir, and Y. Shirman, *New tools for low-energy dynamical supersymmetry breaking*, *Phys. Rev.* **D 53** (1996) 2658.
- [197] ATLAS Collaboration, *Search for events with large missing transverse momentum, jets, and at least two tau leptons in 7 TeV proton-proton collision data with the ATLAS detector*, *Phys. Lett.* **B 714** (2012) 180, [[arXiv:1203.6580](https://arxiv.org/abs/1203.6580)].
- [198] ATLAS Collaboration, *Search for supersymmetry with jets, missing transverse momentum and at least one hadronically decaying τ lepton in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Lett.* **B 714** (2012) 197, [[arXiv:1204.3852](https://arxiv.org/abs/1204.3852)].
- [199] J. A. Cembranos, J. L. Feng, and L. E. Strigari, *Minimal Universal Extra Dimensions*, [arXiv:0708.0239](https://arxiv.org/abs/0708.0239).
- [200] A. Belyaev, M. Brown, J. Moreno, and C. Papineau, *Discovering Minimal Universal Extra Dimensions (MUED) at the LHC*, *JHEP* **1306** (2013) 080, [[arXiv:1212.4858](https://arxiv.org/abs/1212.4858)].
- [201] M. Frigge, D. C. Hoaglin, and B. Iglewicz, *Some implementations of the boxplot*, *The American Statistician* **43** (1989), no. 1 pp. 50–54.
- [202] N. Craig, *The State of Supersymmetry after Run I of the LHC*, [arXiv:1309.0528](https://arxiv.org/abs/1309.0528).
- [203] A. Buckley, “hepthesis - a LaTeX document class for academic theses.”
<http://www.ctan.org/tex-archive/help/Catalogue/entries/hepthesis.html>.

Acronyms

ATLAS A Toroidal Lhc ApparatuS. 5–10, 12, 14, 16–19, 21–24, 39, 45, 46, 48, 90, 91, 93, 94, 121, 124, 129, 130, 152, 154, 157, 166, 178, 181

bRPV bilinear RPV. 151–153

BSM beyond Standard Model. 5, 18, 45

CDF cumulative density function. 84

cMSSM constrained Minimal Supersymmetric Standard Model. 151, 152

EDM electric dipole moment. 42, 44

EWSB electroweak symmetry breaking. 35, 36

FCNCs flavour-changing neutral currents. 42, 43

FSR final state radiation. 107, 108, 131

GMSB gauge-mediated supersymmetry breaking. 38, 151, 153, 154, 182

GUT grand unified theory. 151, 153

ISR initial state radiation. 131

JER jet energy resolution. 130, 138

JES jet energy scale. 21, 24, 130, 138

LAr liquid argon. 9–12, 94

LHC Large Hadron Collider. 3–6, 16, 25, 90, 136

LHT LocHadTopo. 97

LSP lightest supersymmetric particle. 40, 140, 141, 144, 152, 154

MC Monte Carlo. 45, 46, 66, 73, 74, 76, 78, 79, 90–94, 96, 97, 99, 100, 103–107, 109, 112, 116, 120, 122–131, 134, 157, 166, 167, 170–172, 175–178

MCMC Markov Chain Monte Carlo. 86

MLE maximum likelihood estimate. 75, 76, 79, 80, 112, 157, 160, 165, 166

MSSM Minimal Supersymmetric Standard Model. 34–38, 41–43, 151, 153

mSUGRA minimal SUper GRAvity. 38, 150–153

MSW Mikheyev-Smirnov-Wolfenstein. 30

mUED minimal Universal Extra Dimensions. 151, 154, 155

NLL next-to-leading-logarithmic. 93

NLO next-to-leading order. 92, 93

NLSP next-to-lightest supersymmetric particle. 40

PDF parton distribution function. 92, 93

pdf probability density function. 80

pMSSM phenomenological MSSM. 37

RPC R -parity conserving. 38–40, 143

RPV R -parity violating. 38, 39, 42, 140, 141, 143, 153

SCT SemiConductor Tracker. 7, 8

SM Standard Model. 25–32, 34–38, 40–43, 90, 134, 153, 167

SUSY supersymmetry. 6, 19, 21, 22, 25, 31–35, 37–44, 92, 93, 99, 100, 129, 139, 143, 145, 149–151, 153, 181, 182

TRT Transition Radiation Tracker. 7

TTC timing, trigger and control. 94