Searches for Supersymmetric Signatures in all Hadronic Final States with the α_T Variable.

Darren Lee Burton

High Energy Physics
Blackett Laboratory
Imperial College London

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy and the Diploma of Imperial College

Abstract

A search for supersymmetric particles in events with high transverse momentum jets and a large missing transverse energy signature, is conducted using 11.7 fb^{-1} of data, collected with a center-of-mass collision energy of 8 TeV by the CMS detector. The dimensionless kinematic variable α_T is used to select events with genuine missing transverse energy signatures. Standard Model backgrounds are estimated through the use of data driven control samples. No excess over Standard Model expectations is found. Exclusion limits on squark and gluino masses are set at the 95% confidence level in the parameter space of a range of supersymmetric simplified model topologies.

Results of benchmarking the Level-1 (the first line of the CMS trigger system) single jet and hadronic transverse energy trigger efficiencies, before and after the implementation of a change to the Level-1 jet clustering algorithm are presented. Similar performance is observed for all L1 quantities. This change was introduced to negate an increase in trigger cross-section, which can be attributed to soft jets from secondary interactions.

Furthermore, a templated fit method to estimate the Standard Model background distribution of the number of jets originating from a b-quark within a supersymmetric search, is validated in data and simulation. Applicable to searches sensitive to gluino induced third-generation signatures, this technique is utilised as a crosscheck to the results of the α_T analysis. Standard Model background predictions from the template fits are compared to those from the α_T search in the hadronic signal region, where good agreement between the two methods is observed.

Declaration

I, the author of this thesis, declare that the work presented within this document to be my own. The work presented in Chapters 4, 5, 6 and Section 3.4, is a result of the author's own work, or that of which I have been a major contributor unless explicitly stated otherwise, and is carried out within the context of the Imperial College London and CERN SUSY groups, itself a subsection of the greater CMS collaboration. All figures and studies taken from external sources are referenced appropriately throughout this document.

Darren Lee Burton

Acknowledgements

I would like to thank the many people whom I have had the pleasure of working with during the course of the last three and a half years. The opportunity to work as part of the largest scientific collaboration during one of the most exciting times in particle physics for decades, has been a real privilege to be a part of. I could not have achieved the results presented in this thesis without the help of my colleagues who were part of the RA1 team, Edward Laird, Chris Lucas, Henning Flaecher, Yossof Eshaq, Bryn Mathais, Sam Rogerson, Zhaoxia Meng and Georgia Karapostoli whom I worked with on L1 jets. I also thank my supervisor Oliver Buchmuller for his guidance in getting me to this point.

I also feel it important to single out thanks to the postdocs that I have worked with during my PhD. Jad Marrouche from whom I have learnt a great deal and Robert Bainbridge who has been like a second supervisor to me, helping me during my time at Imperial and CERN, especially during those most stressful of times approaching conference deadlines!

My fellow PhD students whom I have lived with and have seen on an almost daily basis for the last few years, Andrew Gilbert, Patrick Owen, Indrek Sepp, Matthew Kenzie and my wonderful girlfriend Hannah. Thanks for putting up with the whinging, complaining and clomping.

Finally my largest thanks go to my Mum and Dad whose patience, encouragement and considerable financial support have allowed me to take the many steps that led me here today.

Contents

List of Figures

List of Tables

Chapter 1.

Introduction

During the 20th century, great advances were made in the human understanding of the universe, its origins, its future and its composition. The Standard Model (SM) first formulated in the 1960s is one of the crowning achievements in science's quest to explain the most fundamental processes and interactions that make up our universe. It has provided a highly successful explanation for a wide range of phenomena in Particle Physics and has stood up to extensive experimental scrutiny [1].

Despite its success it is not a complete theory, with significant questions remaining unanswered. It describes only three of the four known forces with gravity not incorporated within the framework of the SM. Cosmological experiments infer that just $\sim 5\%$ of the observable universe exists as matter, with elusive "Dark Matter" accounting for a further $\sim 27\%$ [2]. However no particle predicted by the SM is able to account for it. At higher energy scales, the (non-)unification of the fundamental forces point to problems with the SM at least at higher energies not yet probed experimentally.

Many theories exist as extensions to the SM, predicting a range of observables that can be detected at the Large Hadron Collider (LHC) of which SUperSYmmetry (SUSY) is one such example. It predicts a new symmetry of nature in which all current particles in the SM would have a corresponding supersymmetric partner. Common to most Supersymmetric theories is a stable, weakly interacting Lightest Supersymmetric Partner (LSP), which has the properties of a possible dark matter candidate. The SM and the main principles of Supersymmetric theories are outlined in Chapter 2, with emphasis placed on how experimental signatures of SUSY may reveal themselves in proton collisions at the LHC.

The experimental goal of the LHC is to further test the framework of the SM, exploring the TeV mass scale for the first time, and to seek a connection between the particles produced in proton collisions and dark matter. The first new discovery by this extraordinary machine was announced on the 4th of July 2012. The long-awaited discovery was the culmination of decades of experimental endeavours in the search for the Higgs boson, which provided an answer to the mechanism of electroweak symmetry breaking within the SM [3][4].

This discovery was made possible through the combination of data taken by the Compact Muon Solenoid (CMS) and A Toroidal LHC ApparatuS (ATLAS), two multipurpose detectors located on the LHC ring. An experimental description of the CMS detector and the LHC is described in Chapter 3, including some of the object reconstruction used by CMS in searches for SUSY signatures.

The performance of the CMS Level-1 single jet and energy sum triggers is also benchmarked within this chapter. The Level-1 trigger is the first line of the CMS trigger system and is of paramount importance to the collection of physics events. A change in the jet clustering algorithm, via the introduction of a jet seed threshold, was introduced approximately half way through the data taking period. The aim of this change, was to reduce the rate at which collisions not of interest to physics analysis were recorded, whilst avoiding impact to the overall performance of the triggers.

Chapter 4, contains a description of the search for direct evidence of the production of supersymmetric particles at the LHC. The main basis of the search centres around the kinematic dimensionless α_T variable; which provides a strong rejection of backgrounds with fake missing transverse energy signatures, whilst maintaining good sensitivity to a variety of SUSY topologies. The author's work (as an integral part of the analysis group) is documented in detail, and has culminated in numerous publications over the past two years, the latest results having been published in the European Physical Journal C (EPJC) [5].

The author in particular has played a major role in the extension of the α_T analysis into additional b-tagged jet (jets identified as originating from a b-quark) and jet multiplicity dimensions, increasing the sensitivity of the analysis to a range of SUSY topologies. Additionally, the author has worked extensively on increasing the statistical precision of the data driven electroweak predictions through analytical techniques. This included work on developing the derivation of data driven systematic uncertainties through the establishment of closure tests within the control samples of the analysis.

The compatibility of the data collected for the α_T search with a SM only hypothesis is documented in Chapter 5. In the absence of an observed excess, interpretations of the data within the framework of a variety of Simplified Model Spectra (SMS), describing an array of possible SUSY event topologies are made.

Finally, a method to search for gluino mediated SUSY signatures rich in top and bottom flavoured jet final states, is introduced in Chapter 6. These particular SUSY topologies are increasingly of interest to physicists in light of the discovery of the Higgs boson. A parametrisation of the b-tagged jet distribution for different electroweak processes is used to establish template shapes, which are then fitted at low b-tagged jet multiplicity, to extrapolate an expected SM background of 3 and 4 b-tagged jet events within an event sample. The α_T control and hadronic signal event selections are used to validate the functionality of this template method in both data and simulation. Background predictions within the hadronic signal region are compared to those presented in Chapter 5, with the intention of serving as a independent crosscheck of the estimated SM backgrounds from the α_T search.

Natural units are used throughout this thesis in which $\hbar = \mathrm{c} = 1$.

Chapter 2.

A Theoretical Overview

Within this chapter, a brief introduction and background to the SM is given. Its success as a rigorously tested and widely accepted theory is discussed as are its deficiencies which lead to the argument that this theory is not a complete description of our universe. The motivations for new physics at the TeV scale and in particular Supersymmetric theories are outlined within Section (2.3). The chapter concludes with how an experimental signature of such theories can be produced and observed at the LHC in Section (2.4).

2.1. The Standard Model

The SM is the name given to the relativistic Quantum Field Theory (QFT), where particles are represented as excitations of fields, which describe the interactions and properties of all the known elementary particles [6][7][8][9]. It is a renormalisable field theory which contains three symmetries: $SU(3)$ for colour charge; SU(2) for weak isospin and; U(1) relating to weak hyper charge, which requires its Lagrangian $\mathcal{L}_{SM}$ to be invariant under local gauge transformation.

Within the SM theory, matter is composed of spin $\frac{1}{2}$ fermions that interact with each other via the exchange of spin-1 gauge bosons. A summary of the known fundamental fermions and bosons is given in Table 2.1.

Fermions are separated into quarks and leptons of which only quarks interact with the strong nuclear force. Quarks unlike leptons are not seen as free particles in nature, but rather exist only within baryons, which are composed of three quarks with an overall integer charge, and quark-anti-quark pairs called mesons. Both leptons and quarks are

Particle	Symbol	Spin	Charge	Mass (GeV)
First Generation Fermions				
Electron Neutrino	ν_e	$\frac{1}{2}$	0	$< 2.2 \times 10^{-6}$
Electron	e	$\frac{1}{2}$	-1	0.51×10^{-3}
Up Quark	u	$\frac{1}{2}$	$\frac{2}{3}$	$2.3^{+0.7}_{-0.5} \times 10^{-3}$
Down Quark	d	$\frac{1}{2}$	$-\frac{1}{3}$	$4.8^{+0.7}_{-0.3} \times 10^{-3}$
Second Generation Fermions				
Muon Neutrino	ν_μ	$\frac{1}{2}$	0	-
Muon	μ	$\frac{1}{2}$	-1	1.05×10^{-3}
Charm Quark	c	$\frac{1}{2}$	$\frac{2}{3}$	1.275 ± 0.025
Strange Quark	s	$\frac{1}{2}$	$-\frac{1}{3}$	$95 \pm 5 \times 10^{-3}$
Third Generation Fermions				
Tau Neutrino	ν_τ	$\frac{1}{2}$	0	-
Tau	τ	$\frac{1}{2}$	-1	1.77
Top Quark	t	$\frac{1}{2}$	$\frac{2}{3}$	173.5 ± 0.8
Bottom Quark	b	$\frac{1}{2}$	$-\frac{1}{3}$	4.65 ± 0.03
Gauge Bosons				
Photon	γ	1	0	0
W Boson	$W^{\pm}$	1	± 1	80.385 ± 0.015
Z Boson	Z	1	0	91.187 ± 0.002
Gluons	g	1	0	0
Higgs Boson	H	0	0	125.3 ± 0.5 [4]

Table 2.1: The fundamental particles of the SM, with spin, charge and mass displayed. Latest mass measurements taken from [1].

grouped into three generations which have the same properties, but with ascending mass in each subsequent generation.

The gauge bosons mediate the interactions between fermions. The field theories of Quantum Electro-Dynamics (QED) and Quantum Chromo-Dynamics (QCD), yield massless mediator bosons, the photon and eight coloured gluons which are consequences of the gauge invariance of those theories (detailed in Section (2.1.1)).

The unification of the electromagnetic and weak-nuclear forces into the current Electroweak theory yield the weak gauge bosons $W^{\pm}$ and Z through the mixing of the associated gauge fields. The force carriers of this theory were experimentally detected by the observation of the weak neutral current. This was first discovered in 1973 by the Gargamelle bubble chamber located at the European Organisation for Nuclear Research (CERN) [10]. The masses of the weak gauge bosons were measured by the UA1 and U2 experiments at the Super Proton Synchrotron (SPS) collider in 1983 [11][12].

2.1.1. Gauge Symmetries of the SM

Symmetries are of fundamental importance in the description of physical phenomena. Noether's theorem states that for a dynamical system, the consequence of any symmetry is an associated conserved quantity [13]. Invariance under translations, rotations, and Lorentz transformations in physical systems lead to the conservation of momentum, energy and angular momentum.

In the SM, a quantum theory described by Lagrangian formalism, the weak, strong and electromagnetic interactions are described in terms of "gauge theories". A gauge theory possesses invariance under a set of "local transformations", which are transformations whose parameters are space-time dependent. The requirement of gauge invariance within the SM necessitates the introduction of force-mediating gauge bosons, and interactions between fermions and the bosons themselves. Given the nature of the topics covered by this thesis, the formulation of Electroweak Sector (EWK) within the SM Lagrangian is reviewed within this section.

The simplest example of the application of the principle of local gauge invariance within the SM is in Quantum Electro-Dynamics (QED), the consequences of which require a massless photon field [14][15].

The free Dirac Lagrangian can be first written as

$$\mathcal{L} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi, \tag{2.1}$$

where ψ represents a free non interacting fermionic field, with the matrices γ^{μ}, $\mu \in 0, 1, 2, 3$ defined by the anti commutator relationship $\gamma^{\mu}\gamma^{\nu} + \gamma^{\mu}\gamma^{\nu} = 2\eta^{\mu\nu}I_4$, with $\eta^{\mu\nu}$ being the flat space-time metric $(+, -, -, -)$, and I_4 the 4×4 identity matrix.

Under a local U(1) abelian gauge transformation, in which ψ transforms as

$$\psi(x) \to \psi^{'}(x) = e^{i\theta(x)}\psi(x) \qquad \bar{\psi}(x) \to \bar{\psi}^{'}(x) = e^{i\theta(x)}\bar{\psi}(x) \tag{2.2}$$

the kinetic term of the Lagrangian will not remain invariant, due to the partial derivative interposed between the $\bar{\psi}$ and ψ yielding

$$\partial_\mu \psi \to e^{i\theta(x)} \partial_\mu \psi + ie^{i\theta(x)} \psi \partial_\mu \theta. \tag{2.3}$$

To ensure that $\mathcal{L}$ remains invariant, a modified derivative, D_μ, that transforms covariantly under phase transformations is introduced. In doing this, a vector field A_μ with transformation properties that cancel out the unwanted term in (2.3) must also be included,

$$D_\mu \equiv \partial_\mu - ieA_\mu, \qquad A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \theta. \tag{2.4}$$

Invariance of the Lagrangian is then achieved by replacing ∂_μ with D_μ:

$$\begin{aligned} \mathcal{L} &= i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi \\ &= \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu \psi A_\mu. \end{aligned} \tag{2.5}$$

An additional interaction term is now present in the Lagrangian, coupling the Dirac particle to this vector field, which is interpreted as the photon in QED. To regard this new field as the physical photon field, a term corresponding to its kinetic energy must be added to the Lagrangian from Equation (2.5). Since this term must also be invariant under the conditions of Equation (2.4), it is defined in the form $F_{\mu\nu} = \partial^\mu A^\nu - \partial_\nu A_\mu$.

This then leads to the Lagrangian of QED,

$$\mathcal{L}_{QED} = \overbrace{i\bar{\psi}\gamma^\mu \partial_\mu \psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}}^{\text{kinetic term}} + \overbrace{m\bar{\psi}\psi}^{\text{mass term}} + \overbrace{e\bar{\psi}\gamma^\mu \psi A_\mu}^{\text{interaction term}} . \tag{2.6}$$

Within the Lagrangian there remains no mass term of the form $m^2 A_\mu A^\mu$, which is prohibited by gauge invariance. This implies that the gauge particle, the photon, must be massless.

2.1.2. The Electroweak Sector and Electroweak Symmetry Breaking

The same application of gauge symmetry and the requirement of local gauge invariance can be used to unify QED and the Weak force in the Electroweak Sector (EWK). The nature of EWK interactions is encompassed within a Lagrangian invariant under transformations of the group $SU(2)_L \times U(1)_Y$.

The weak interactions from experimental observation [16] are known to violate parity and are therefore not symmetric under interchange of left and right helicity fermions. Thus, within the SM the left- and right-handed parts of these fermion fields are treated separately. A fermion field is then split into two left- and right-handed chiral components, $\psi = \psi_L + \psi_R$, where $\psi_{L/R} = (1 \pm \gamma^5)\psi$.

The $SU(2)_L$ group is the special unitary group of 2×2 matrices, U, satisfying $UU^\dagger = I$ and det(U) = 1. It may be written in the form $U = e^{-i\omega_i T_i}$, with the generators of the group written as $T_i = \frac{1}{2}\tau_i$ where τ_i, $i \in$ 1,2,3 are the 2×2 Pauli matrices

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.7}$$

The generators of the group form a non Abelian group obeying the commutation relation $[T^a, T^b] \equiv if^{abc}T^c \neq 0$. The gauge fields that accompany this group are represented by $\hat{W}_\mu = (\hat{W}^1_\mu, \hat{W}^2_\mu, \hat{W}^3_\mu\ )$ and act only on the left handed component of the fermion field ψ_L.

One additional generator, Y, which represents the hypercharge of the particle under consideration is introduced through the $U(1)_Y$ group acting on both components of the fermion field, with an associated vector boson field $\hat{B}_\mu$.

The $SU(2)_L \times U(1)_Y$ transformations of the left- and right-handed components of ψ are summarised by,

$$\begin{aligned} \chi_L \rightarrow \chi_L^{'} &= e^{i\theta(x)\,\cdot\,T + i\theta(x)Y}\chi_L, \\ \psi_R \rightarrow \psi_R^{'} &= e^{i\theta(x)Y}\psi_R, \end{aligned} \tag{2.8}$$

where the left-handed fermions form isospin doubles χ_L and the right handed fermions are isosinglets ψ_R. For the first generation of leptons and quarks this represents

$$\begin{aligned} \chi_L &= \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \qquad \begin{pmatrix} u \\ d \end{pmatrix}_L, \\ \psi_R &= e_R, \qquad\qquad u_R, d_R. \end{aligned} \tag{2.9}$$

Local gauge invariance within $\mathcal{L}_{EWK}$ is once again imposed by modifying the covariant derivative

$$D_\mu = \partial_\mu - \frac{ig}{2}\tau^i W^i_\mu - \frac{ig^{'}}{2} Y B_\mu, \tag{2.10}$$

where g and $g^{'}$ are the coupling constant of the $SU(2)_L$ and $U(1)_Y$ groups respectively. Taking the example of the first generation of fermions defined in Equation (2.9), with input hypercharge values of -1 and -2 for χ_L and e_R respectively, would lead to a Lagrangian $\mathcal{L}_1$ of the form,

$$\begin{aligned} \mathcal{L}_1 =& \bar{\chi}_L \gamma^\mu [i\partial_\mu - g\frac{1}{2}\tau \cdot W_\mu - g^{'}(-\frac{1}{2})B_\mu]\chi_L \\ &+ \bar{e}_R \gamma^\mu [i\partial_\mu - g^{'}(-1)B_\mu]e_R - \frac{1}{4}W_{\mu\nu} \cdot W^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \end{aligned} \tag{2.11}$$

As in QED, these additional gauge fields introduce field strength tensors $B_{\mu\nu}$ and $W_{\mu\nu}$,

$$\hat{B}_{\mu\nu} = \partial_\mu \hat{B}_\nu - \partial_\nu \hat{B}_\mu \tag{2.12}$$

$$\hat{W}_{\mu\nu} = \partial_\mu \hat{W}_\nu - \partial_\nu \hat{W}_\mu - g\hat{W}_\mu \times \hat{W}_\mu \tag{2.13}$$

corresponding to the kinetic energy and self coupling of the W_μ fields and the kinetic energy term of the B_μ field.

None of these gauge bosons are physical particles, and instead linear combinations of these gauge bosons make up γ and the W and Z bosons, defined as

$$W^{\pm} = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right), \qquad \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} cos\theta_W & -sin\theta_W \\ sin\theta_W & cos\theta_W \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}, \tag{2.14}$$

where the mixing angle, $\theta_w = \tan^{-1}\frac{g'}{g}$, relates the coupling of the neutral weak and electromagnetic interactions.

As in the case of the formulation of the QED Lagrangian there remains no mass term for the photon. However contrary to experimental measurement, this is also the case for the W, Z and fermions in the Lagrangian. Any explicit introduction of mass terms would break the symmetry of the Lagrangian, and instead mass terms can be introduced through spontaneous breaking of the EWK symmetry via the Higgs mechanism.

The Higgs mechanism induces spontaneous symmetry breaking through the introduction of a complex scalar SU(2) doublet field ϕ, which attains a non-zero Vacuum Expectation Value (VEV) [17][18][19][20].

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \qquad with \qquad \begin{matrix} \phi^+ \equiv (\phi_1 + i\phi_2)/\sqrt{2} \\ \phi^0 \equiv (\phi_3 + i\phi_4)/\sqrt{2} \end{matrix} \tag{2.15}$$

The Lagrangian defined in Equation (2.11) attains an additional term $\mathcal{L}_{Higgs}$ of the form

$$\begin{aligned} \mathcal{L}_{Higgs} &= \overbrace{(D_\mu\phi)^\dagger(D^\mu\phi)}^{\text{kinetic}} - \overbrace{\mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2}^{\text{potential V}(\phi)} \qquad (\mu^2, \lambda) > 0 \in \mathbb{R}, \\ \mathcal{L}_{SM} &= \mathcal{L}_{EWK} + \mathcal{L}_{Higgs}, \end{aligned} \tag{2.16}$$

where the covariant derivative D_μ is that defined in Equation (2.10). The last two terms of $\mathcal{L}_{Higgs}$ correspond to the Higgs potential, in which real positive values of μ^2 and λ are required to ensure the generation of masses for the bosons and leptons. The minimum of this potential is found at $\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \mu^2/\lambda = v^2$, where v represents the VEV.

The ground state of the ϕ field is defined to be consistent with the $V(\phi)$ minimum. By then expanding around a ground state chosen to maintain an unbroken electromagnetic symmetry. This preserves a zero photon mass [21] and leads to

$$\phi_0 = \sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}, \qquad \phi(x) = e^{i\tau\cdot\theta(x)/v}\sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{2.17}$$

where the fluctuations from the vacuum ϕ_0 are parametrized in terms of four real fields, $\theta_1, \theta_2, \theta_3$ and $h(x)$.

Choosing to gauge away the three massless Goldstone boson fields by setting $\theta(x)$ to zero and substituting $\phi(x)$ back into kinetic term of $\mathcal{L}_{Higgs}$ from Equation (2.16) leads to mass terms for the $W^\pm$ and Z bosons. This is given by,

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}(\partial_\mu h)^2 + \frac{g^2v^2}{2}W_\mu^+W^{-\mu} + \frac{v^2g^2}{8\cos^2\theta_w}Z_\mu Z^\mu + 0A_\mu A^\mu, \tag{2.18}$$

where the relations between the physical and electroweak gauge fields from Equation (2.14) are used. The $W^\pm$ and Z boson masses can then be determined to be

$$M_W = \frac{1}{2}gv \qquad M_Z = \frac{1}{2}\frac{gv}{\cos\theta_w}. \tag{2.19}$$

This mechanism is also used to generate fermion masses by introducing a Yukawa coupling between the fermions and the ϕ field [22], with the coupling strength of a particle to the ϕ field governing its mass. Additionally, a scalar boson h with mass $m_h = v\sqrt{\frac{\lambda}{2}}$, is also predicted as a result of this spontaneous symmetry breaking. This became known as the Higgs boson. Its discovery by the CMS and ATLAS experiments in 2012 is the first direct evidence to support this method of mass generation within the SM.

2.2. Motivation for Physics beyond the Standard Model

As has been described, the SM has proven to be a very successful theory, predicting the existence of the $W^{\pm}$ and Z bosons and the top quark long before they were experimentally observed. However, the theory does not accurately describe all observed phenomena and has some fundamental theoretical flaws that hint at the need for additional extensions to the current theory.

On a theoretical level, the SM is unable to incorporate the gravitational interactions of fundamental particles within the theory. Whilst at the electroweak energy scales the relative strength of gravity is negligible compared to the other three fundamental forces, at much higher energy scales, $\mathrm{M_{planck}} \sim 10^{18}\mathrm{GeV}$, quantum gravitational effects become increasingly dominant. The failure to reconcile gravity within the SM, demonstrates that the SM must become invalid at some higher energy scale.

Other deficiencies with the SM include the fact that the predicted rate of Charge-Parity violation does not account for the matter dominated universe which we inhabit, and that the SM prediction of a massless neutrino conflicts with the observation of neutrino flavour mixing, attributed to mixing between neutrino mass eigenstates [23][24].

Perhaps one of the most glaring gaps in the predictive power of the SM is that there exists no candidate to explain the cosmic dark matter observed in galactic structures through indirect techniques; including gravitational lensing and measurement of the orbital velocity of stars at galactic edges. Any such candidate must be very weakly interacting but must also be stable, owing to the lack of direct detection of the decay products of such a process. Therefore, a predicted stable dark matter candidate is one of the main obstacles to address for any Beyond Standard Model (BSM) physics model.

The recent discovery of the Higgs boson, whilst a significant victory for the predictive power of the SM, brings with it still unresolved questions. This issue is commonly described as the "hierarchy problem".

In the absence of new physics between the TeV and Planck scale, calculating beyond tree-level contributions to the Higgs mass term given by its self interaction, results in divergent terms that push the Higgs mass up to the planck mass M_{planck}.

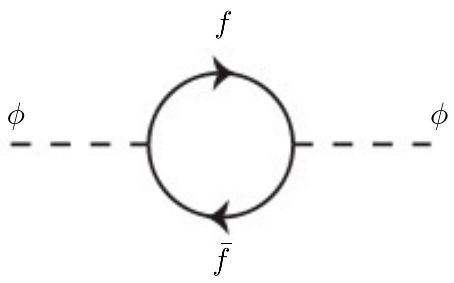


Figure 2.1: One loop quantum corrections to the Higgs squared mass parameter m_h^2 due to a fermion.

This can be demonstrated by considering the one loop quantum correction to the Higgs mass with a fermion f, shown in Figure 2.1 with mass m_f. The Higgs field couples to f with a term in the Lagrangian $-\lambda_f h\bar{f}f$, yielding a correction of the form [25],

$$\delta m_h^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda^2 + .., \tag{2.20}$$

where λ_f represents the coupling strength for each type of fermion $\propto m_f$, and Λ the cutoff energy scale at which the SM ceases to be a valid theory.

To recover the mass of the now discovered Higgs boson would require a fine-tuning of the parameters to cancel out these mass corrections of the Higgs mass, to the scale of 30 orders of magnitude. This appears as an unnatural solution to physicists and it is this hierarchy problem that provides one of the strongest motivations for the theory of SUperSYmmetry (SUSY).

2.3. Supersymmmetry Overview

Supersymmetry provides potential solutions to many of the issues raised in the previous section. It provides a dark matter candidate, can explain baryogenesis in the early universe and also provides an elegant solution to the hierarchy problem [26][27][28][29]. At its heart it represents a new space-time symmetry that relates fermions and bosons. This symmetry converts bosonic states into fermionic states, and vice versa,

$$Q|\text{Boson}\rangle = |\text{Fermion}\rangle \qquad Q|\text{Fermion}\rangle = |\text{Boson}\rangle, \tag{2.21}$$

where the operator Q is the generator of these transformations. Quantum field theories which are invariant under such transformations are called supersymmetric.

This symmetry operator therefore acts upon a particle's spin altering it by a half integer value. The consequences of the application of this additional space-time symmetry introduce a new rich phenomenology. For example, in supersymmetric theories both the left-handed SU(2) doublet and right-handed singlet of fermions will have a spin-0 superpartner containing the same electric charge, weak isospin, and colour as its SM partner. In the case of leptons $(\nu_l, l)_L$, they will have two superpartners, a sneutrino $\widetilde{\nu}_{lL}$ and a slepton $\widetilde{l}_L$, whilst the singlet l_R also has a superpartner slepton $\widetilde{l}_R$.

Each particle in a supersymmetric theory is paired together with their superpartners as a result of these supersymmetric transformations in what is called a supermultiplet. These superpartners will then consequently also contribute to the corrections to the Higgs mass. Bosonic and fermionic loops contributing to the correction appear with opposite signs, and therefore cancellation of these divergent terms will stabilise the Higgs mass, solving the hierarchy problem [30][31].

One of the simplest forms of SUSY, is to simply have a set of SM supersymmetric partners with the same mass and interactions as their counterparts. However, the current lack of any experimental evidence for that predicted sparticle spectrum implies SUSY must be a broken symmetry in which any sparticle masses must be greater than their SM counterparts.

There exists many techniques which can induce supersymmetric breaking [32][33][34]. Of particular interest to experimental physicists are those at which the breaking scale is of an order that is experimentally accessible to the LHC i.e. $\sim$ TeV scale. Whilst there is no requirement for supersymmetric breaking to occur at this energy scale, for supersymmetry to provide a solution to the hierarchy problem, it is necessary for this scale to not differ too drastically from the EWK scale [35][36].

2.3.1. R-Parity

Supersymmetric theories can also present a solution to the dark matter problem. These theories contain a stable Lightest Supersymmetric Partner (LSP), which match the criteria of a Weakly Interacting Massive Particle (WIMP) required by cosmological observation when R-parity is conserved.

Baryon (B) and Lepton (L) number conservation is forbidden in the SM by renormalisability requirements. The violation of Baryon or Lepton number results in a proton lifetime much shorter than those set by experimental limits [37]. Another symmetry called R-parity is then often introduced to SUSY theories to maintain baryon and lepton conservation.

R-parity is described by the equation

$$R_P = (-1)^{3(B-L)+2s}, \tag{2.22}$$

where s represents the spin of the particles. $B = \pm\frac{1}{3}$ for quarks/antiquarks and $B = 0$ for all others, $L = \pm 1$ for leptons/antileptons, $L = 0$ for all others.

R-parity ensures the stability of the proton in SUSY models, and also has other consequences for the production and decay of supersymmetric particles. In particle colliders supersymmetric particles can then only be pair produced. Similarly the decay of any produced supersymmetric particle is restricted to a SM particle and a lighter supersymmetric particle, as allowed by conservation laws. A further implication of R-parity is that once a supersymmetric particle has decayed to the LSP it remains stable, unable to decay into a SM particle.

A LSP will not interact in a detector at a particle collider, leaving behind a missing energy, $\not{E}_T$, signature. The assumption of R-parity and its consequences are used to determine the physical motivation and search strategies for SUSY at the LHC.

2.4. Experimental Signatures of SUSY at the LHC

Should strongly interacting sparticles be within the experimental reach of the LHC, then it is expected that they can be produced in a variety of ways:

- squark/anti-squark and gluino pairs can be produced via both gluon fusion and quark/anti-quark scattering,
- a gluino and squark produced together via quark-gluon scattering,
- squark pairs produced via quark-quark scattering.

Whilst most SUSY searches invoke the requirement of R-parity to explore parameter phase space, there still exist a whole plethora of possible SUSY model topologies, which could yet be discovered at the LHC.

During the 2011 run period at $\sqrt{s} = 7$ TeV, particular models were used to benchmark performance and experimental reach of both CMS searches and previous experiments. The Compressed Minimal SuperSymmetric Model (CMSSM) was initially chosen for a number of reasons [38]. One of the most compelling being the reduction of the up to 105 new parameters that can be introduced by SUSY (in addition to the existing 19 of the SM), to just 5 extra free parameters. It was this simplicity, combined with the theory not requiring any fine tuning of particle masses to produce experimentally verified SM observables, that made it an attractive model to interpret physics results.

However, recent results from the LHC now strongly disfavour large swathes of CMSSM parameter space [39][40][41]. In the face of such results a more pragmatic model independent search strategy is now applied across most SUSY searches at the LHC, see Section (2.4.1).

As previously stated, a stable LSP that exhibits the properties of a dark matter candidate would be weakly interacting and therefore will not be directly detected in a detector environment. Additionally, the cascade decays of supersymmetric particles to this LSP state would also result in significant hadronic activity. These signatures will then be characterised through large amounts of hadronic jets (see Section (3.3.1)), leptons and a significant amount of missing energy all dependent upon the LSP mass and the size of the mass splitting between the LSP and the supersymmetric particle it has decayed from.

The SM contains processes which can exhibit a similar event topology to that described above, with the largest contribution coming from the general QCD multi-jet environment of a hadron collider. A multitude of different analytical techniques are used by experimental physicists to reduce or estimate any reducible or irreducible backgrounds, allowing a possible SUSY signature to be extracted. The techniques employed within this thesis are described in great detail within Section (4.1).

2.4.1. Simplified Models

With such a variety of different ways for a SUSY signal to manifest itself, it is necessary to be able to interpret experimental reach through the masses of gluinos and squarks which can be excluded by experimental searches, rather than on a model specific basis.

This is accomplished through SMS models, which are defined by a set of hypothetical particles and a sequence of their production and decay modes [42][43]. In the SMS models considered within this thesis, only the production process for the two primary particles are considered. Each primary particle can undergo a direct or a cascade decay through an intermediate new particle. At the end of each decay chain there remains a neutral, undetected LSP particle, denoted $\widetilde{\chi}_{LSP}$ which can represent a neutralino or gravitino. Essentially it is easier to consider each SMS with branching ratios set to 100%. The masses of the primary particle and the LSP remain as free parameters, in which the absolute value and relative difference between the primary and LSP particle alter the kinematics of the event.

Different SMS models are denoted with a T-prefix, with a summary of the types interpreted within this thesis listed below [44].

- **T1,T1xxxx**, models represent a simplified version of gluino pair production, with each gluino (superpartner to the gluon) undergoing a three-body decay to a quark-antiquark pair and the LSP (i.e. $\widetilde{g} \rightarrow q\bar{q}\widetilde{\chi}_{LSP}$). The resultant final state from this decay is typically 4 jets + $\not{E}_T$ in the absence of initial/final state radiation and detector effects. xxxx denotes models in which the final state quarks are of a specific flavour, typically t or b quark-antiquarks.

- **T2,T2xx**, models represent a simplified version of squark anti-squark production with each squark undergoing a two-body decay into a light-flavour quark and LSP (i.e. $\widetilde{q} \rightarrow q\widetilde{\chi}_{LSP}$). This results in final states with less jets than gluino mediated production, typically 2 jets + $\not{E}_T$ when again ignoring the effect of initial/final state radiation and detector effects. xx models represent decays in which both the quark and the squark within the final state is of a specific flavour, which in this thesis are again $\widetilde{t}$/t or $\widetilde{b}$/b.

Models rich in b and t quarks are interpreted within this thesis as they remain of particular interest within "Natural SUSY" scenarios [45][46]. The largest contribution to the quadratic divergence in the Higgs mass parameter comes from a loop of top quarks via the Yukawa coupling. Cancellation of these divergences can be achieved in supersymmetric theories by requiring a light right-handed top squark, $\widetilde{t}_R$, and left-handed double $SU(2)_L$ doublet containing top and bottom squarks, $(\frac{\tilde{t}}{\tilde{b}})_L$ [47].

These theories therefore solve the hierarchy problem by predicting light $\sim$EWK scale third generation sleptons, accessible at the LHC. Search strategies involving the requirement of b-tagging (see Section (3.3.2)) are used to give sensitivity to these type of SUSY scenarios and are discussed in greater detail within Chapter 4.

Two example decay chains are shown in Figure 2.2; the pair production of gluinos (`T1`) and the pair production of squarks (`T2`) decaying into SM particles and LSPs.

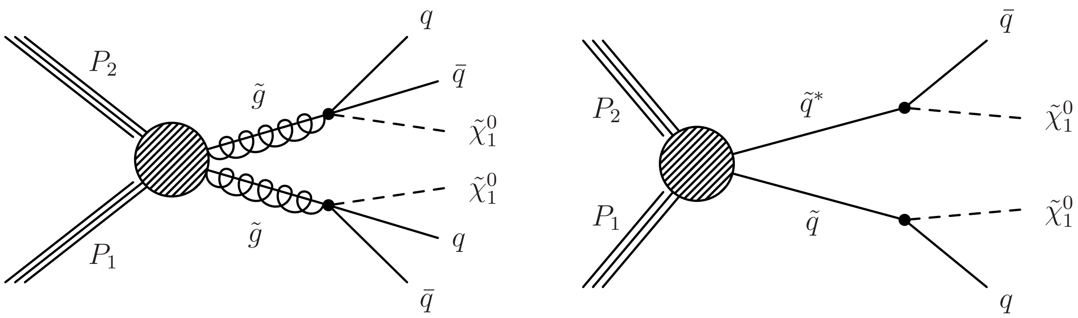


Figure 2.2: Two example SMS model decays (`T1` (left),`T2` (right)), which are used in interpretations of physics reach by CMS.

Chapter 3.

The LHC and the CMS Detector

Probing the SM for signs of new physics would not be possible without the immensely complex electronics and machinery that has made the TeV energy scale accessible to physicists for the first time. This chapter will introduce both the LHC based at European Organisation for Nuclear Research (CERN) and the Compact Muon Solenoid (CMS) detector (of which the author is a member). Section (3.2) serves to present an overview of the different components of the CMS detector, with specific components relevant to the search for supersymmetric particles described in greater detail. Section (3.3) will focus on event and object reconstruction, again, with more emphasis on jet level quantities which are most relevant to the author's analysis research. Finally, Section (3.4) will describe and detail the service work for the CMS Collaboration performed by the author, in measuring the performance of L1 single jet and energy sum triggers in the Global Calorimeter Trigger (GCT) during the 2012-2013 run period.

3.1. The LHC

The LHC is a storage ring, accelerator, and collider of circulating beams of protons or ions. Housed in the tunnel dug for the Large Electron-Positron Collidor (LEP), it is approximately 27km in circumference, 100m underground, and straddles the border between France and Switzerland, outside of Geneva. It is currently the only collider in operation that is able to study physics at the TeV scale. A double-ring circular synchrotron, it was designed to collide both proton-proton (pp) and heavy ion (PbPb) with a centre of mass energy $\sqrt{s} = 14$ TeV at a final design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$.

These counter-circulating beams of protons or Pb ions are merged in four sections around the ring to enable collisions of the beams, with each interaction point being home to one of the four major experiments; A Large Ion Collider Experiment (ALICE) [48], A Toroidal LHC ApparatuS (ATLAS) [49], Compact Muon Solenoid (CMS) [50] and Large Hadron Collider Beauty (LHCb) [51] which record the resultant collisions. The layout of the LHC ring is shown in Figure 3.1. The remaining four sections contain acceleration, collimation and beam dump systems. In the eight arc sections, the beams are steered by magnetic fields of up to 8 T provided by super conduction dipole magnets, which are maintained at temperatures of 2 K using superfluid helium. Additional magnets for focusing and corrections are also present in straight sections within the arcs and near the interaction regions where the detectors are situated.

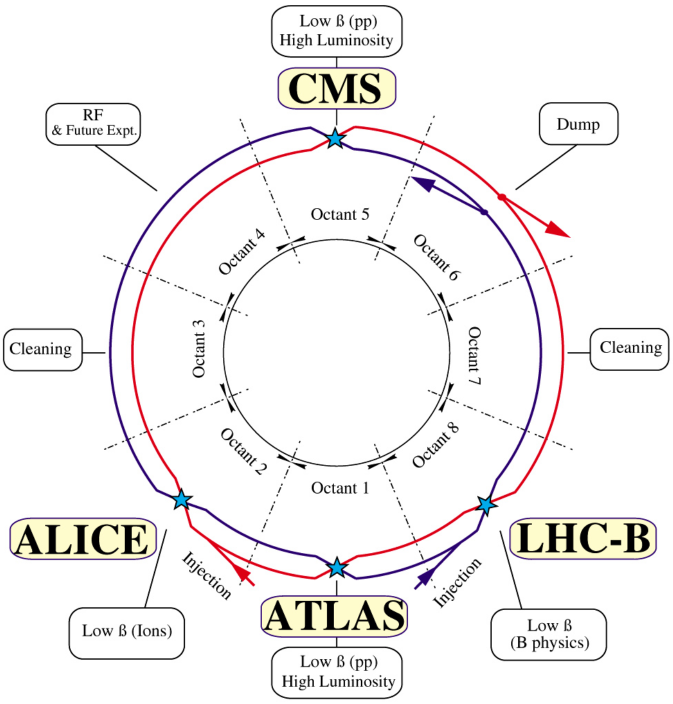


Figure 3.1: A top down layout of the LHC. [52], with the position of the four main detectors labelled.

Proton beams are formed inside the Proton Synchrotron (PS) from bunches of protons 50 ns apart with an energy of 26 GeV. The protons are then accelerated in the Super Proton Synchrotron (SPS) to 450 GeV before being injected into the LHC. These LHC proton beams consists of many "bunches" (i.e. approximately 1.1×10^{11} protons localised into less than 1 ns in the direction of motion). Before collision, the beams are ramped to

4 TeV (2012) per beam, in a process involving increasing the current passing through the dipole magnets. Once the desired $\sqrt{s}$ energy is reached then the beams are allowed to collide at the interaction points. The luminosity falls regularly as the run progresses; protons are lost in collisions, and eventually the beam is dumped before repeating the process again.

Colliding the beams produced an instantaneous luminosity of approximately 5×10^{33} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ during the 2012 run. The high number of protons in each bunch increases the likelihood of multiple interactions with each crossing of the counter-circulating beams. This leads to isotropic energy depositions within the detectors positioned at these interaction points, increasing the energy scale of the underlying event. This is known as *pile-up* and the counteracting of its effects are important to the many measurements performed at the LHC.

In the early phase of prolonged operation, after the initial shutdown, the machine operated in 2010-2011 at 3.5 TeV per beam, $\sqrt{s} = 7$ TeV, delivering 6.13 fb^{-1} of data [53]. During the 2012-2013 run period, data was collected at an increased $\sqrt{s} = 8$ TeV improving the sensitivity of searches for new physics. Over the whole run period 23.3 fb^{-1} of data was delivered, of which 21.8 fb^{-1} was recorded by the CMS detector as shown in Figure 3.2 [53]. A total of 12 fb^{-1} of 8 TeV certified data was collected by October 2012, and it is this data which forms the basis of the results presented within this thesis.

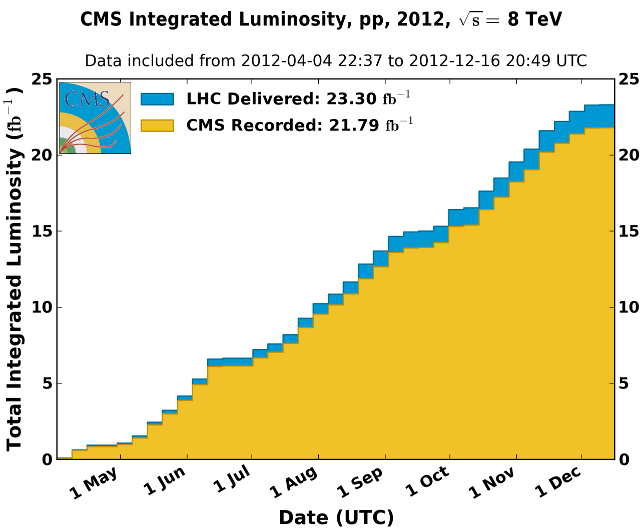


Figure 3.2: The total integrated luminosity delivered to and collected by CMS during the 2012 8 TeV *pp* runs.

3.2. The CMS Detector

The Compact Muon Solenoid (CMS) detector is one of two general purpose detectors at the LHC designed to search for new physics. The detector is designed to provide efficient identification and measurement of many physics objects including photons, electrons, muons, taus, and hadronic showers over wide ranges of transverse momentum and direction. Its nearly 4π coverage in solid angle allows for accurate measurement of global transverse momentum imbalance. These design factors give CMS the ability to search for direct production of SUSY particles at the TeV scale, making the search for Supersymmetric particles one of the highest priorities among the wide range of physics programmes at CMS.

CMS uses a right-handed Cartesian coordinate system with the origin at the interaction point and the z-axis pointing along the beam axis. The x-axis points radially inwards to the centre of the collider ring, with the y-axis pointing vertically upward. The azimuthal angle ϕ, ranging between $[-\pi,\pi]$, is defined in the x-y plane starting from the x-axis. The polar angle θ is measured from the z axis. The common convention in particle physics is to express an out-going particle in terms of ϕ and its pseudorapidity defined as

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

The variable $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ is commonly used to define angular distance between objects within the detector. Additionally, energy and momentum is typically measured in the transverse plane perpendicular to the beam line. These values are calculated from the x and y components of the object and are denoted as $E_T = E \sin\theta$ and $p_T = \sqrt{p_x^2 + p_y^2}$.

3.2.1. Detector Subsystems

As the range of particles produced from pp collisions interact in different ways with matter, CMS is divided into sub-detector systems, which perform complementary roles to identify the identity, the mass, and the momentum of different physics objects present in each event. These detector sub-systems contained within CMS are wrapped in layers around a central 13m long 4 T super conducting solenoid, as shown in Figure 3.3. With the endcaps closed, CMS is a cylinder of length 22m, diameter 15m, and mass 12.5

kilotons. A more detailed complete description of the detector can be found elsewhere [50].

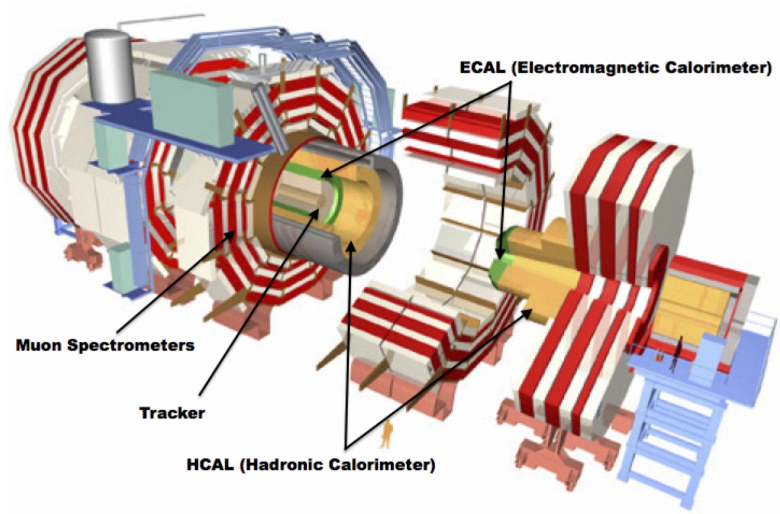


Figure 3.3: A pictorial depiction of the CMS detector with the main detector subsystems labelled [54].

3.2.2. Tracker

The inner-most sub-detector of the barrel is the multi-layer silicon tracker, formed of a pixel detector component encased by layers of silicon strip detectors. The pixel detector consists of three layers of silicon pixel sensors providing measurements of the momentum, position coordinates of the charged particles as they pass, and the location of primary and secondary vertices between 4 cm and 10 cm transverse to the beam. Outside the pixel detector, ten cylindrical layers of silicon strip detectors extend the tracking system out to a radius of 1.20m from the beam line. The tracking system provides efficient and precise determination of the charges, momenta, and impact parameters of charged particles, with the geometry of the tracker extending to cover a rapidity range up to $|\eta| < 2.5$.

The tracking system also plays a crucial part in the identification of jets that originate from b-quarks through the measurement of displaced secondary vertices. The methods in which these b-flavoured jets are identified are discussed within Section (3.3.2). The identification of b-jets is important in many searches for natural SUSY models and forms an important part of the inclusive search strategy described within Section (4.2).

3.2.3. Electromagnetic Calorimeter

Immediately outside of the tracker, but still within the magnet core, sits the Electromagnetic CALorimeter (ECAL). Covering a pseudorapidity up to $|\eta| < 3$ and comprising of over 75×10^3 $PbWO_4$ (lead tungstate) crystals that scintillate as particles deposit energy, the ECAL provides high resolution measurements of the electromagnetic showers from photons and electrons in the detector.

Lead tungstate is used because of its short radiation length ($X_0 \sim 0.9$ cm) and small Molieré radius (~ 2.1 cm) leading to high granularity and resolution. Its fast scintillation time (~ 25 ns) reduces the effects of pile-up, which occurs when energy from previous collisions are still being read out, and its radiation hardness gives it longevity. The crystals are arranged in modules which surround the beam line in a non-projective geometry, angled at 3°, with respect to the interaction point to minimise the risk of particles escaping down the cracks between the crystals.

The ECAL is primarily composed of two sections, the Electromagnetic CALorimeter Barrel (EB) which extends in pseudo-rapidity to $|\eta| < 1.479$ with a crystal front cross section of 22×22 mm and a length of 230 mm corresponding to 25.8 radiation lengths. The Electromagnetic CALorimeter Endcap (EE) covers a rapidity range of $1.479 < |\eta| < 3.0$, which consists of two identical detectors on either side of the EB. A lead-silicon sampling 'pre-shower' detector Electromagnetic CALorimeter pre-Shower (ES) is placed before the endcaps to aid in the identification of neutral pions. Their arrangement is shown in Figure 3.4.

Scintillation photons from the lead tungstate crystals are instrumented with Avalanche Photo-Diodes (APD) and Vacuum Photo-Triodes (VPT), located in the EB and EE respectively. They convert the scintillating light into an electric signal which is consequently used to determine the amount of energy deposited within the crystal. These instruments are chosen for their resistance under operation to the strong magnetic field of CMS. The scintillation of the ECAL crystals, as well as the response of the APDs, vary as a function of temperature; and so cooling systems continually maintain an overall constant ECAL temperature $\pm 0.05°C$.

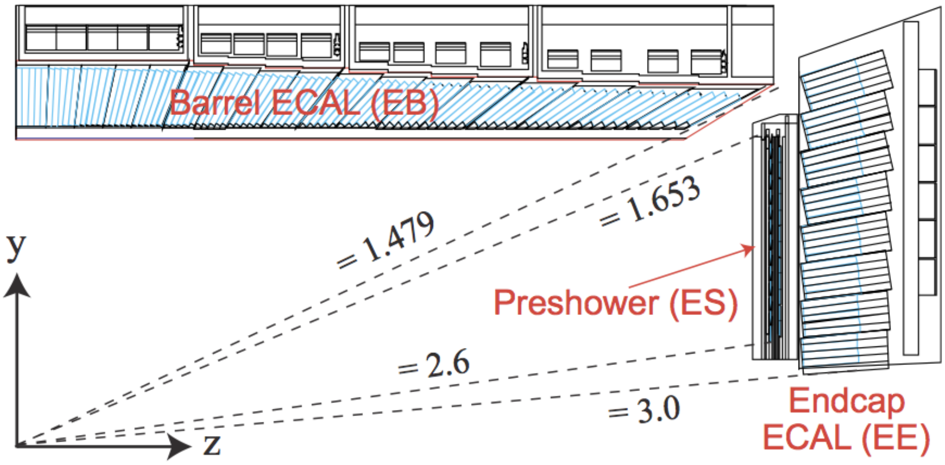


Figure 3.4: Illustration of the CMS ECAL showing the arrangement of the lead tungstate crystals in the EB and EE. The ES is also shown and is located in front of the EE [55].

3.2.4. Hadronic Calorimeter

Beyond the ECAL lies the Hadronic CALorimeter (HCAL) which is responsible for the accurate measurement of hadronic showers, crucial for analyses involving jets or missing energy signatures. The HCAL is a sampling calorimeter which consists of alternating layers of brass absorber and plastic scintillator. The exception being in the hadron forward ($3.0 < |\eta| < 5.0$) region where steel absorbers and quartz fibre scintillators are used because of their increased radiation tolerance. Hadron showers are initiated in the absorber layers inducing scintillation in the plastic scintillator tiles. These scintillation photons are converted by wavelength shifting fibres for read-out by hybrid photodiodes.

The HCAL's size is constrained to a compact size by the presence of the solenoid, requiring the placement of an additional outer calorimeter on the outside of the solenoid to increase the sampling depth of the HCAL. A schematic of the HCAL can be seen in Figure 3.5.

The HCAL covers the range $|\eta| < 5$ and consists of four sub-detectors: the Hadron Barrel (HB) $|\eta| < 1.3$, the Hadron Outer (HO), the Hadron Endcaps (HE) $1.3 < |\eta| < 3.0$ and the Hadron Forward (HF). The HB, contained between the outer edge of the ECAL and the inner edge of the solenoid is formed of 36 azimuthal wedges which are split

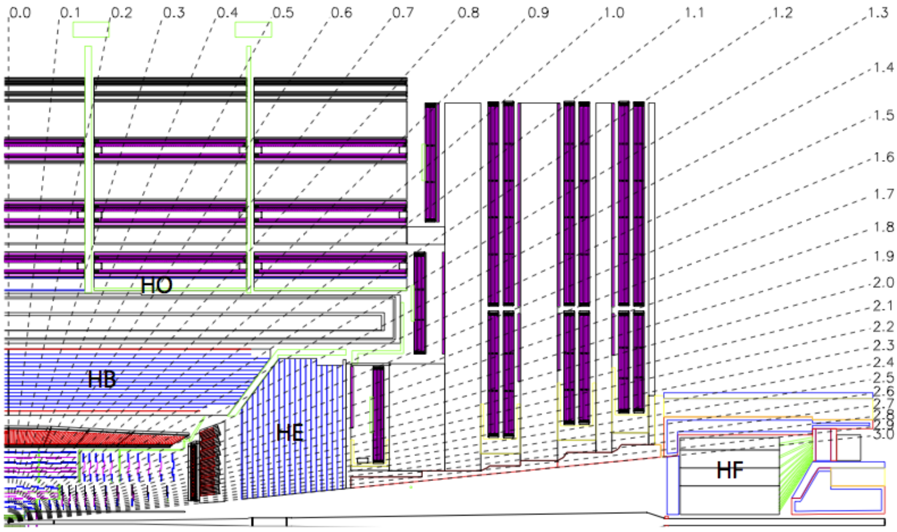


Figure 3.5: Schematic of the hadron calorimeters in the r-z plane, showing the locations of the HCAL components and the HF. [50].

between two half-barrel segments. Each wedge is segmented into four azimuthal angle (ϕ) sectors, and each half-barrel is further segmented into 16 η towers. The electronic readout chain, channels the light from the active scintillator layers from one ϕ-segment and all η-towers of a half-barrel to a Hybrid Photo Diode (HPD).

The relatively short number of interaction lengths (λ_l, the distance a hadron will travel through the absorber material before it has lost $\frac{1}{e}$ of its energy) within the HB, the lowest being $\lambda_l = 5.82$ at $|\eta| = 0$, facilitates the need for the 'tail catching' HO to increase the sampling depth in the central barrel rapidity region $|\eta| < 1.3$ to 11 interaction lengths. Significant fractions of the hadrons energy will also be deposited in the ECAL as it passes through the detector. Therefore, measurements of hadron energies in the central regions $|\eta| < 3.0$ use both the ECAL and HCAL to reconstruct the true energy from showering hadrons.

3.2.5. Muon Systems

Muons being too massive to radiate away energy via Bremsstrahlung, interact little in the calorimeters and mostly pass through the detector until they reach the system of muon detectors which forms the outer most part of the CMS detector.

Outside of the superconducting solenoid are four muon detection layers interleaved with the iron return yokes, which measure the muons energy via ionisation of gas within detector elements. Three types of gaseous chambers are used. The Drift Tube (DT), Cathode Stripe Chamber (CSC), and Resistive Plate Chamber (RPC) systems provide efficient detection of muons with pseudo-rapidity $|\eta| < 2.4$. The best reconstruction performance is obtained when the muon chamber is combined with the inner tracking information to determine muon trajectories and their momenta [56].

3.3. Event Reconstruction and Object Definition

The goal of event reconstruction is to take the raw information recorded by the detector and to compute from it higher-level quantities which can be used at an analysis level. These typically correspond to an individual particle's energy and momenta, groups of particles which shower in a narrow cone, and the overall global energy and momentum balance of the event. The reconstruction of these objects are described in great detail in [57], while covered below are brief descriptions of those which are most relevant to the analysis detailed in Chapter 4.

3.3.1. Jets

Quarks and gluons are produced copiously at the LHC in the hard scattering of partons. As these quarks and gluons fragment, they hadronize and decay into a group of strongly interactive particles and their decay products. These streams of particles travel in the same direction, as they have been "boosted" by the momentum of the primary hadron. These collections of decay products are reconstructed and identified together as a "jet".

At CMS jets are reconstructed from energy deposits in the detector via the anti-kt algorithm [58] with size parameter $\Delta R = 0.5$. The anti-kt jet algorithm clusters jets by defining a distance between hard (high p_T) and soft (low p_T) particles such that soft particles are preferentially clustered with hard particles before being clustered between themselves. This produces jets which are robust to soft particle radiation from the pile-up conditions produced by the LHC.

There are two main types of jet reconstruction used at CMS, Calorimeter (Calo) and Particle Flow (PF) jets [59]. Calorimeter jets are reconstructed using both the ECAL and HCAL cells, combined into calorimeter towers. These calorimeter towers consist of geometrically matched HCAL cells and ECAL crystals. Electronics noise is suppressed by applying a threshold to the calorimeter cells, with pile-up effects reduced by a requirement placed on the tower energy [60]. Calorimeter jets are the jets used within the analysis presented in this thesis.

PF jets are formed from combining information from all of the CMS sub-detectors systems to determine which final state particles are present in the event. Generally, any particle is expected to produce some combination of a track in the silicon tracker, a deposit in the calorimeters, or a track in the muon system. The PF jet momentum and spatial resolutions are greatly improved with respect to calorimeter jets, as the use of the tracking detectors and of the high granularity of ECAL allows resolution and measurement of charged hadrons and photons inside a jet, which together constitute $\sim$ 85% of the jet energy [61].

The jets reconstructed by the clustering algorithm in CMS typically have an energy that differs to the 'true' energy measured by a perfect detector. This stems from the non-linear and nonuniform response of the calorimeters as well as other residual effects including pile-up and underlying events. Therefore, additional corrections are applied to recover a uniform relative response as a function of pseudo-rapidity. These are applied as separate sub corrections [62].

- A pile-up correction is first applied to the jet. It subtracts the average extra energy deposited in the jet that comes from other vertices present in the event and is therefore not part of the hard jet itself.
- p_T and η dependant corrections derived from Monte Carlo simulations are used to account for the non-uniform response of the detector.
- p_T and η residual corrections are applied to data only to correct for difference between data and Monte Carlo. The residual is derived from QCD di-jet samples and the p_T residual from $\gamma+$ jet and $Z+$ jets samples in data.

3.3.2. B-tagging

The decays of b-quarks are suppressed by small CKM matrix elements. As a result, the lifetimes of b-flavoured hadrons, produced in the fragmentation of b-quarks, are relatively long; $\sim$ 1ps. The identification of jets originating from b-quarks is very important for searches for new physics and for measurements of SM processes.

Many different algorithms developed by CMS select b-quark jets based on variables such as; the impact parameters of the charged-particle tracks, the properties of reconstructed decay vertices, and the presence or absence of a lepton, or combinations thereof.

One of the most efficient of which is the Combined Secondary Vertex (CSV) algorithm [63]. This operates based on secondary vertex and track-based lifetime information, benchmarked in 'Loose', 'Medium' and 'Tight' working points, of which the medium point is the tagger used within the α_T search presented in Section (4.1). All figures within this sub-section, demonstrating the performance of this b-tagging algorithm are taken from [64].

Within the CSV tagger, a likelihood-based discriminator distinguishes between jets from b-quarks, and those from charm or light quarks and gluons, shown in Figure 3.6. The minimum thresholds on the discriminator for each working point correspond to the mis-identication probability for light-parton jets of 10%, 1%, and 0.1%, respectively, in jets with an average p_T of about 80 GeV.

The b-tagging performance is evaluated to measure the b-jet tagging efficiency ϵ_b, and the misidentification probability of charm ϵ_c and light-parton jets ϵ_s. The tagging efficiencies for each of these three jet flavours are compared between data and MC simulation, from which a series of p_T and $|\eta|$ dependant jet corrections are determined,

$$SF_{b,c,s} = \frac{\epsilon_{b,c,s}^{data}}{\epsilon_{b,c,s}^{MC}}. \tag{3.2}$$

These are collectively named 'Btag Scale Factors' and allow MC simulation to accurately reflect the running conditions and performance of the tagging algorithm in data. A good understanding of the tagging efficiency for each of the jet flavours is essential in order to

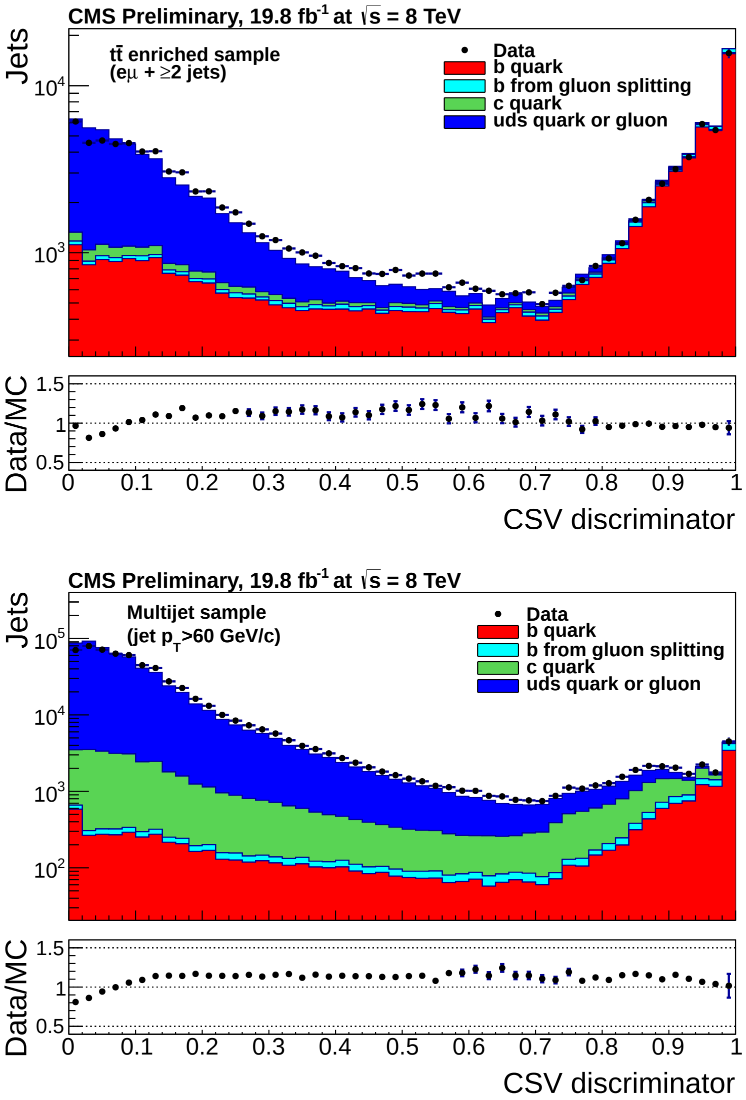


Figure 3.6: CSValgorithm discriminator values in enriched ttbar (top) and inclusive multi-jet samples (bottom) for b,c and light flavoured jets. The discriminator value used for each working points are determined from the misidentification probability for light-parton jets to be tagged as a b-jet, which are given as 0.244 (10%), 0.679 (1%) and 0.898 (0.1%) for the L, M and T working points respectively.

minimise systematic uncertainties in physics analyses that employ b-tagging.

The b-tagging efficiency is measured in data using several methods applied to multi-jet events, primarily based on a sample of jets enriched in heavy flavour content. One method requires the collection of events with a poorly isolated muon within a cone $\Delta R < 0.4$ around the jet axis. Due to the semi-leptonic branching fraction of b hadrons being significantly larger than that for other hadrons, these jets are more likely to arise from b quarks than from another flavour. The resultant momentum component of the muon, transverse to the jet axis, is larger in b-hadron decays than from light or charm flavoured jets.

Additionally, the performance of the tagger can also be benchmarked in $t\bar{t}$ events, where the top quark is expected to decay to a W boson and a b quark about 99.8% of the time [1]. Further selection criteria is applied to these events to further enrich the b quark content of these events. The methods to identify b-jets in data are discussed in great detail at [65]. The jet flavours within simulation are determined using truth level information and are compared to data to determine the appropriate correction scale factors ($SF_{b,c,s}$). The scale factor corrections from simulation to data for b-quark jets with the CSVM tagger are displayed in Figure 3.7.

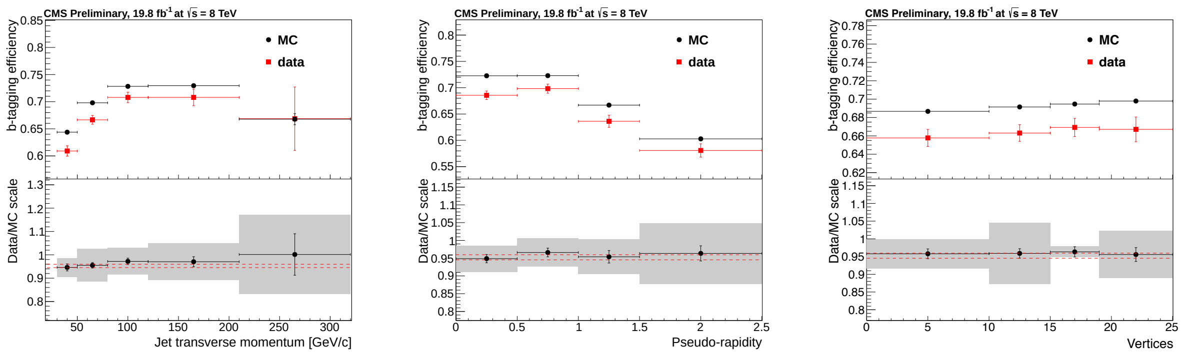


Figure 3.7: Measured in $t\bar{t} \to$ di-lepton events using the CSVM tagger: (upper panels) b-tagging efficiencies and (lower panels) data/MC scale factor SF_b as a function of (left) jet p_T, (middle) jet $|\eta|$ and (right) number of primary vertices. In the lower panels, the grey filled areas represent the total statistical and systematic uncertainties, whereas the dotted lines are the average SF_b values within statistical uncertainties.

The measurement of the misidentication probability for light-parton jets relies on the inversion of tagging algorithms, selecting non-b jets using the same variables and techniques used for benchmarking the b-tagging efficiency. The scale factors (SF_s) to be applied to correct simulation to data are shown in Figure 3.8 for the CSVM tagger.

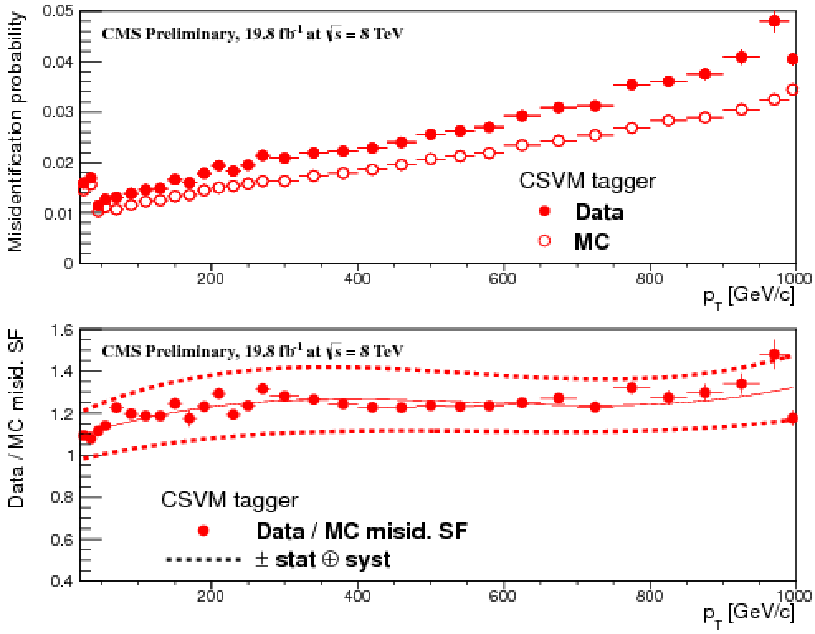


Figure 3.8: For the CSVM tagging criterion: (top) misidentication probability in data (filled circles) and simulation (open circles); (bottom) scale factor for the misidentication probability. The last p_T bin in each plot includes all jets with $p_T > 1000$ GeV. The solid curve is the result of a polynomial fit to the data points. The dashed curves represent the overall statistical and systematic uncertainties on the measurements.

3.4. Triggering System

With bunch crossings separated by just 50 ns, the rate at which data from all collisions would have to be written out and processed would be unfeasible. A two-tiered triggering system is applied at CMS in order to cope with the high collision rate of protons. The CMS trigger is designed to use limited information from each event to determine whether to record the event, reducing the rate of data taking to manageable levels whilst ensuring a high efficiency of interesting physics object events are selected.

The Level 1 Trigger (L1) is a pipelined, dead-timeless system based on custom-built electronics [66], and is a combination of several sub systems which is shown pictorially in Figure 3.9. The L1 system is covered in more detail within the following section, along with a description of the service work undertaken by the author to benchmark the performance of the L1 calorimeter trigger during the 2012 8 TeV run period.

The Higher Level Trigger (HLT) is a large farm of commercial computers [67]. The HLT processes events with software reconstruction algorithms that are more detailed, giving performance more similar to the reconstruction used offline. The HLT reduces the event

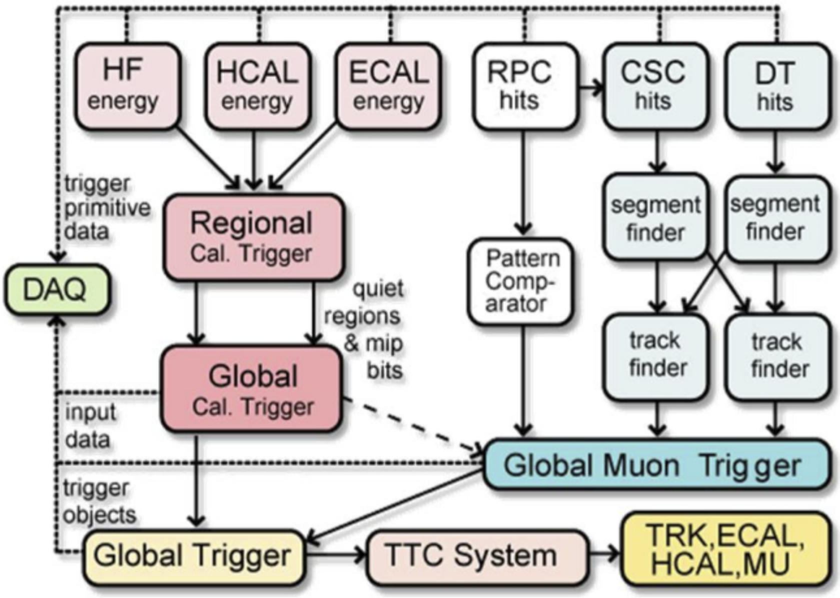


Figure 3.9: An overview of the different components of the CMS L1 trigger system, showing the global calorimeter, muon triggers, and the global trigger.

rate written to disk by a factor of $\sim$ 500 ($\sim$200Hz). The recorded events are transferred from CMS to the CERN computing centre, where event reconstruction is performed, and then distributed to CMS computing sites around the globe for storage and analysis.

3.4.1. The Level-1 Trigger

The L1 trigger reduces the rate of events collected from 20 MHz to $\sim$100 kHz using information from just the calorimeters and muon chambers, but not the tracker. This is due to requirement that data from each and every bunch crossing be analysed with no dead time, drastically reducing time available to process and reconstruct objects in making a trigger decision. This facilitates the need for a pipelined processing architecture, and so a tree system of triggers is used to decide whether to pass on an event to the HLT for further reconstruction.

Calorimeter and muon event information is processed separately by the Regional Calorimeter Trigger (RCT) and Regional Muon Trigger (RMT) systems respectively.

Within the RCT, energy deposits from trigger towers in the ECAL and HCAL calorimeters are summed into coarser calorimeter regions and sent to the Global Calorimeter Trigger (GCT) for jet clustering.

Given that electron and photon are much narrower objects than jets, the RCT is used to identify these candidates but makes no attempt to distinguish between them at this stage given the lack of tracking information. They are first identified by ensuring the energy deposits within the central trigger tower and its surrounding cells are above a certain programmable threshold. To ensure the object is not a hadron, the ratio of HCAL to ECAL in the central tower is calculated and checked to be below 5%. Additional algorithms are employed to ascertain whether the e/γ object is isolated/non-isolated.

In the L1 GCT, coarse measurements of the energy deposited in the electromagnetic and hadronic calorimeters are combined, and by using sophisticated algorithms the following tasks are performed:

- isolated and non-isolated electromagnetic objects are sorted (e and γ), with the four highest ranked (equivalent to highest transverse energy E_T) objects of each type passed onto the Global Trigger (GT),
- energy sums from the calorimeters supplied by the RCT are used in performing jet clustering (described in the following section). The clustered jets are then sub-divided into categories depending on their pseudo-rapidity and the result of τ identication, being classified as either central, forward, or tau (τ). After being sorted by rank, the four highest of each category are passed to the GT for use in trigger decisions,
- total transverse energy (E_T), the scalar sum of the energy deposits measured by L1, and missing transverse energy ($\not{E}_T$), defined as the negative vector sum of the transverse energy deposits measured at L1 are calculated,
- total transverse jet energy (H_T), the scalar sum of the energy of all L1 clustered jet objects, and missing transverse jet energy ($\not{H}_T$), defined as the negative vector sum of the energy from L1 clustered jet objects are calculated and passed to the GT.

In addition, quantities suitable for triggering minimum bias events, forward physics and beam background events are determined. Relevant muon isolation information is also passed on to the Global MuonTrigger (GMT) to be used in decisions involving the muon triggers, where it is combined with information from across the three muon sub-systems.

The resultant final accept/reject decision at L1 is then performed by the GT, based on the objects received from the GCT and GMT (e/γ, μ, jets, E_T, $\not{E}_T$, H_T, $\not{H}_T$).

The L1 trigger is therefore of upmost importance to the functioning of the detector. Without a high-performing, efficient trigger and a good understanding of its performance at ever increasing instantaneous luminosities, the data collected would be useless. Whilst it would be possible to maintain trigger efficiency by increasing the triggering thresholds for different jet or energy sum quantities, this is far from ideal. This could result in the failure to be sensitive to a wide range of new physics signatures, including many types of compressed spectra SUSY models where the mass splitting between squarks/gluinos and the LSP is small.

One such method introduced to help maintain low triggering thresholds, was via the introduction of a jet seed threshold into the L1 jet clustering algorithm. Observations of how the L1 trigger performance is affected by both the jet seed threshold, and changing LHC running conditions over the 2012 run period is presented in the following Sections (3.4.2 - 3.4.6).

3.4.2. The L1 Trigger Jet Algorithm

The L1 jet algorithm clusters jets using the transverse energy sums computed by the calorimeter trigger regions. Each region consists of a 4×4 trigger tower window which within the CMS barrel spans a region of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ in pseudorapidity-azimuth.

A L1 jet is defined by a 3×3 window of calorimeter regions, as shown in Figure 3.10. This corresponds to 12×12 trigger towers in barrel and endcap, or 3×3 larger HF towers in the HF. The ϕ size of the jet window is the same everywhere, whilst the η binning gets somewhat larger at high η due to calorimeter and trigger tower segmentation. The jets are labelled by the (η, ϕ) indices of the central calorimeter region.

A jet candidate is identified if the sum of the transverse ECAL and HCAL energies of a calorimeter region is larger than all of its 8 neighbouring regions $E_{\mathrm{T\ central}} > E_{\mathrm{T\ surround}}$. The jet is then centred at this central calorimeter region.

During the 2012 run period, a minimum threshold of 5 GeV was imposed on the central seeding region to suppress noise from non-collimated pile-up jets. This threshold is applied on the raw, uncorrected energy of the calorimeter regions and affects all clustered

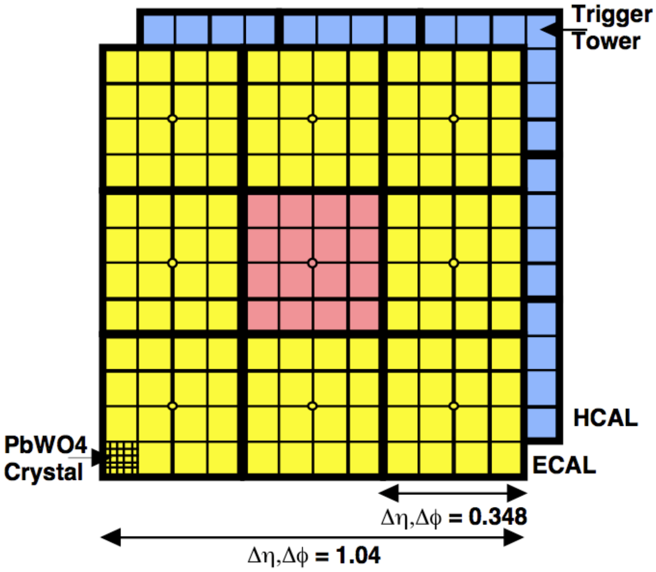


Figure 3.10: Illustration of the dimensions of the Level-1 jet finder window. Each cell represents a trigger tower, which is the sum of the transverse energy contributions from both calorimeter systems.

L1 jets. The effect of such a change to the jet algorithm on the triggering performance of L1 quantities is shown in Section (3.4.4).

To form the jet candidates the GCT utilises a pre-clustering algorithm, which employs 18 jet-finders that operate simultaneously over the whole detector. Each jet-finder spans an area of 11 calorimeter regions in η (half the detector) and two in ϕ ($40°$). Jets are initially created in 2×3 mini-clusters in order to reduce the total amount of data duplicated and shared between the jet-finders. Information is only shared with the two ϕ strips of the neighbouring jet-finders when these mini-clusters jets are found, and is used to form a clustered 3×3 L1 jet object. Additional information on this clustering algorithm can be found at [68].

Within the $|\eta| < 3$ region, the GCT also determines whether a jet is to be classified as a τ or central jet. The hadronic decay modes of the τ typically contain one or three isolated pions, thus leading to more collimated energy deposits with fewer constituents than non-τ jets. Therefore, for a jet candidate to be classed as a τ jet, up to a maximum of one of the eight calorimeter regions neighbouring the central jet seed is permitted to have a transverse energy, E_T, above some programmable isolation threshold.

Jets found between $3.0 < |\eta| < 5.0$ are classified as forward jets, whereas those with $|\eta| < 3.0$ are classified as either a central or τ-jet. The four clustered jets with the highest transverse energy in each category (central, forward and τ-jet) are further passed through Look Up Table (LUT)s, which apply a programmable $\eta-$dependent jet energy scale correction. Finally these jet objects are passed to the GT to make L1 trigger decisions.

The performance of L1 jets within the following sections are evaluated with respect to offline jets, which are taken from the standard Calo jet and the PF jet reconstruction algorithms of CMS. These reconstructed offline jets are corrected for pile-up and detector effects as described in Section (3.3.1). A moderate level of noise rejection is applied to the offline jets by selecting jets passing the "loose" identification criteria for both Calo and PF. These jet criteria are listed in Appendix (A.1).

3.4.3. Measuring L1 Single-Jet Trigger Efficiencies

The efficiency of a L1 single-jet trigger at an offline reconstructed jet E_T is defined as; the fraction of events in a sample containing at least a single reconstructed offline jet, where the leading offline jet is matched with a L1 central or τ jet that also has a measured L1 energy above the trigger threshold being benchmarked.

A match is determined by comparing the L1 and reconstructed offline jets spatially in $\eta - \phi$ space. The ΔR separation between the highest offline reconstructed jet ($E_T > 10$ GeV and $|\eta| < 3$) and each L1 jet in the event is calculated. A match is made to the L1 jet with the minimum ΔR to the reconstructed jet on the condition that it also satisfies $\Delta R < 0.5$.

The matching efficiency for this procedure is found to be close to 100% above an offline jet threshold of 30(45) GeV for the run 2012B(C) data taking period (see Appendix B.1).

Each efficiency curve is fitted with a function which is the cumulative distribution function of an Exponentially Modified Gaussian (EMG) distribution:

$$f(x; \mu, \sigma, \lambda) = \frac{\lambda}{2} \cdot e^{\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2x)} \cdot \mathrm{erfc}\left(\frac{\mu + \lambda\sigma^2 - x}{\sqrt{2}\sigma}\right) \tag{3.3}$$

where erfc is the complementary error function defined as:

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt.$$

In this functional form, the parameter μ determines the point of 50% of the plateau efficiency, and the σ gives the resolution. This parametrisation is used to benchmark the efficiency at the plateau, the turn-on points and resolution for each L1 Jet trigger. The choice of function is purely empirical. Previous studies used the error function alone, which described the data well at high threshold values but could not describe the efficiencies well at lower thresholds [69].

The efficiency turn-on curves for various L1 jet thresholds are evaluated as a function of the offline reconstructed jet E_T for central jets with $|\eta| < 3$. These are measured using single isolated μ triggers which have high statistics, and are orthogonal and therefore unbiased to the hadronic triggers under study. Events are selected with some loose detector based isolation requirements to make sure the muon does not overlap with a jet, causing a discrepancy in the measurement of the calorimetric energy.

The efficiency is calculated with respect to offline Calo and PF Jets in Figure 3.11. Table 3.1 shows the values of these parameters, calculated for three example L1 single jet triggers taken from 2012 8 TeV data.

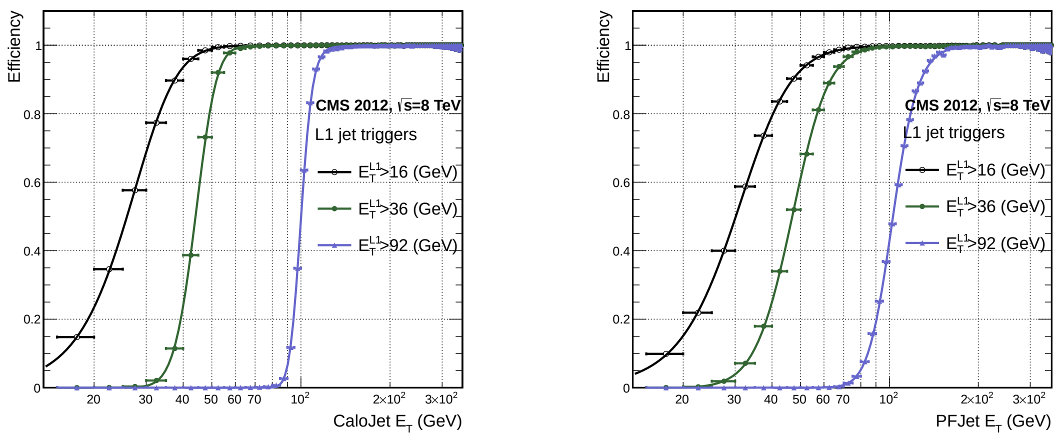


Figure 3.11: L1 jet efficiency turn-on curves as a function of the offline CaloJet E_T (left) and PFJet E_T (right), measured in 2012 Run Period C data and collected with an isolated single μ data sample.

The results from the L1 single jet triggers shows good performance for both Calo and PF jets. A better resolution is observed for Calo jets with respect to L1 single-jet quantities. This effect is due to Calo jet reconstruction using the same detector subsystems as the L1 jets. In contrast the PF jet reconstruction algorithm additionally utilises tracker and muon information, resulting in a poorer resolution when directly compared to L1 jet objects.

	Calo		PF	
Trigger	μ	σ	μ	σ
L1_SingleJet16	21.09 ± 0.03	7.01 ± 0.02	22.17 ± 0.04	7.83 ± 0.03
L1_SingleJet36	41.15 ± 0.05	5.11 ± 0.02	39.16 ± 0.06	8.04 ± 0.03
L1_SingleJet92	95.36 ± 0.13	5.62 ± 0.03	90.85 ± 0.19	11.30 ± 0.10

Table 3.1: Results of a cumulative EMG function fit to the turn-on curves for L1 single jet triggers in run 2012 Run Period C, measured in an isolated μ data sample. The turn-on point, μ, and resolution, σ, of the L1 jet triggers are measured with respect to offline Calo Jets (left) and PF Jets (right).

3.4.4. Effects of the L1 Jet Seed

Between run period B and C of the 2012 data taking period, a jet seed threshold was introduced into the L1 jet clustering algorithm. There was previously no direct requirement made on the energy deposited in the central region.

The introduction of a jet seed threshold required that the central region have an uncorrected energy deposit of $E_T \geq 5$ GeV. This value was motivated by studies of the effect that different jet seed thresholds had upon the trigger cross-sections and efficiencies of various H_T, single jet and multi-jet triggers. It was found that the 5 GeV threshold gave large reductions in trigger cross-sections particularly in the case of multi-jet and H_T triggers, whilst having a small impact on the measured efficiencies of these triggers [70].

Its main purpose was to counteract the effects of high pile up running conditions which create a large number of soft non-collimated jets, that are then added to the jets from the primary interaction or other soft jets from other secondary interactions [71]. This in turn causes a large increase in trigger rate, due to the increase in the likelihood that the event causes the L1 trigger to fire.

The effect of the introduction of this jet seed threshold between these two run periods is benchmarked through a comparison of the efficiency of the L1 jet triggers with respect to offline Calo jets and is shown in Figure 3.12.

The L1 H_T trigger efficiency is also benchmarked at two values, which is shown in Figure 3.13. The L1 H_T sum is compared against the offline H_T constructed from Calo jets with $E_T \geq 40$ GeV. This requirement is imposed to account for the relative difference between uncorrected jet energy deposits within the GCT used to calculate the L1 H_T sum, and those same deposits after full object reconstruction has occurred.

To negate any effects from different pile-up conditions in the run periods, the efficiencies are measured in events which contain between 15 and 20 primary vertices, as defined in Appendix (A.2).

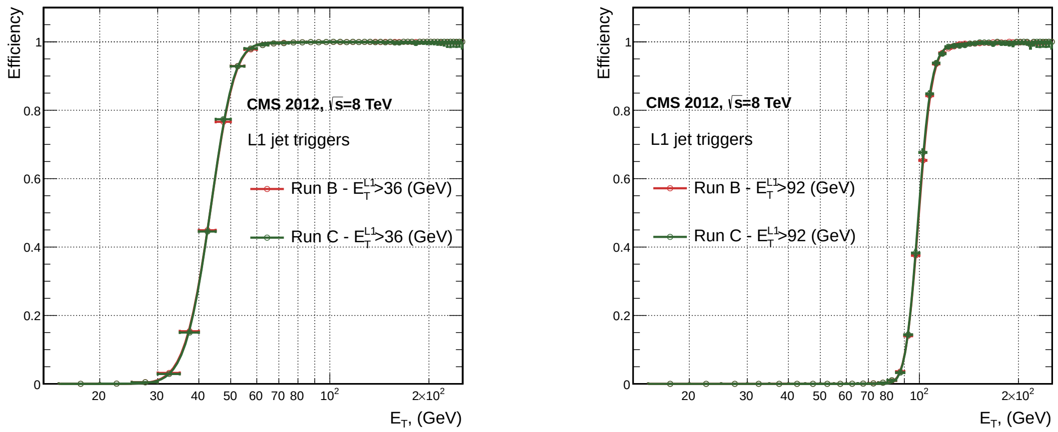


Figure 3.12: L1 jet efficiency turn-on curves as a function of the offline CaloJet E_T, measured for the L1 SingleJet 36 and 92 trigger in 2012 run period B and C collected with an isolated single μ sample.

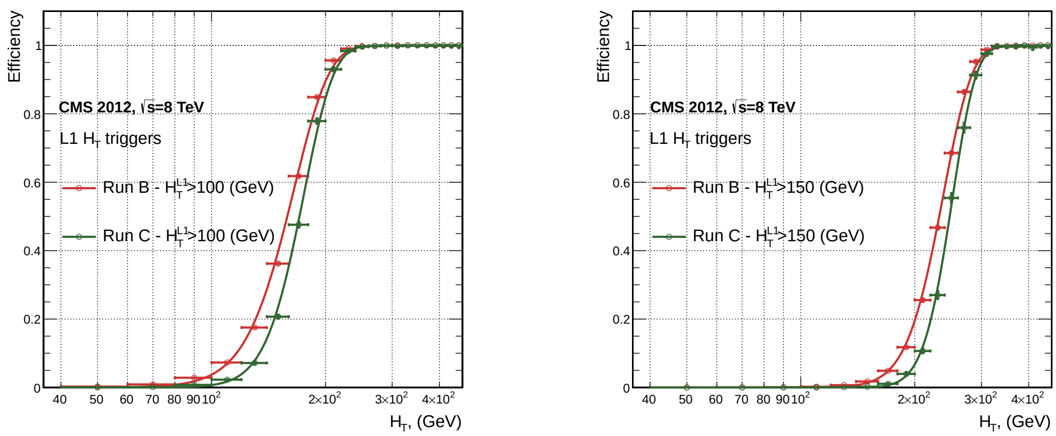


Figure 3.13: L1 H_T efficiency turn-on curves as a function of the offline CaloJet H_T, measured for the L1 H_T 100 and 150 trigger during the run 2012 B and C, collected using an isolated single μ triggered sample.

It can be seen that the performance of the $E_T > 36, 92$ single jet triggers are almost identical, with the jet seed having no measurable effect on these triggers as shown in Table 3.2.

In the case of the H_T triggers, without the jet seed threshold a large increase in the trigger cross-section during high luminosity collisions will occur. The low energy threshold

	2012B		2012C	
Trigger	μ	σ	μ	σ
L1_SingleJet36	40.29 ± 0.04	5.34 ± 0.02	40.29 ± 0.11	5.21 ± 0.05
L1_SingleJet92	94.99 ± 0.09	5.93 ± 0.06	94.82 ± 0.29	5.74 ± 0.18

Table 3.2: Results of a cumulative EMG function fit to the turn-on curves for L1 single jet triggers in the 2012 run period B and C, preselected on an isolated muon trigger. The turn-on point μ and resolution σ of the L1 jet triggers are measured with respect to offline Calo Jets in run B (left) and run C (right).

requirement for a jet to be clustered and added to the L1 H_T sum, will allow many soft jets from other secondary interactions to enter the calculation. The introduction of the jet seed threshold prevents the clustering of many of these diffuse low E_T pile-up jets, thus lowering the L1 GCT H_T calculation. Resultantly, different behaviours for the trigger turn-ons after the introduction of the jet seed threshold are expected for these triggers.

The mean, μ, values are measured to reside at higher H_T for both benchmarked H_T triggers, whilst a better resolution is observed after the introduction of the jet seed threshold. These values can be found within Table 3.3.

	2012B		2012C	
Trigger	μ	σ	μ	σ
L1 HT-100	157.5 ± 0.08	32.9 ± 0.08	169.8 ± 0.08	28.7 ± 0.03
L1 H1-150	230.9 ± 0.02	37.3 ± 0.01	246.4 ± 0.16	31.8 ± 0.05

Table 3.3: Results of a cumulative EMG function fit to the turn-on curves for H_T in run 2012 B and C, preselected on an isolated single μ trigger. The turn-on point μ, resolution σ of the L1 H_T triggers are measured with respect to offline H_T, formed from CaloJets with a $E_T \geq 40$ in run period B (left) and C (right).

Despite this slight increase in the turn-on point of the H_T triggers, a large reduction in the trigger cross-section is achieved for all H_T triggers. As an example, the expected trigger cross-section for the `L1HTT150` trigger as a function of instantaneous luminosity is shown in Figure 3.14.

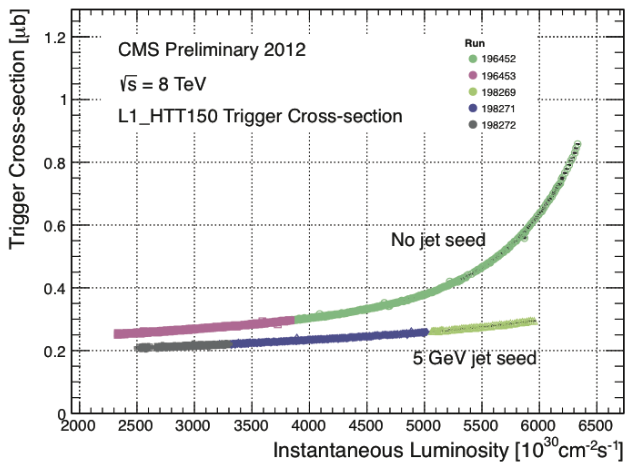


Figure 3.14: Trigger cross section for the `L1HTT150` trigger path. Showing that a 5 GeV jet seed threshold dramatically reduces the dependance of cross section on the instantaneous luminosity for L1 H_T triggers [72].

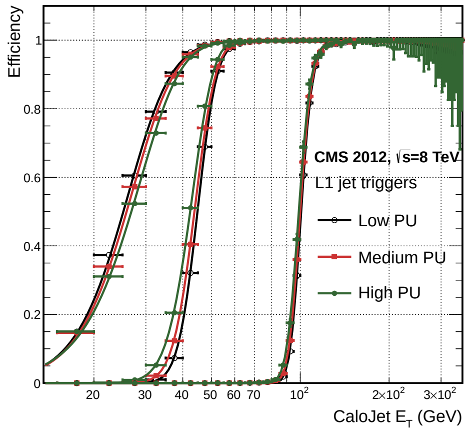
It can be seen that this slight degradation in the offline value at which these H_T triggers become fully efficient due to the jet seed threshold, can be justified from the large reduction in the trigger cross-section rate. Any inefficiencies can then if necessary be compensated through a reduction in the H_T trigger threshold of the L1 seed.

3.4.5. Robustness of L1 Jet Performance against Pile-up

The performance of the L1 single jet triggers is evaluated in different pile-up conditions to determine any dependence on pile-up. Three different pile-up categories of 0-10, 10-20 and >20 vertices are defined, reflecting the low, medium and high pile-up running conditions at CMS in 2012.

The L1 triggers are benchmarked relative to Calo and PF jets in the run period where the jet seed threshold *is* applied, for the L1 single jet thresholds of 16, 36 and 92 GeV, shown in Figure 3.15. The results of fitting an EMG function to these efficiency turn-on curves are given in Table 3.4 and Table 3.5 for Calo and PF jets respectively.

No significant drop in efficiency is observed in the presence of a high number of primary vertices. The increase in hadronic activity in higher pile-up conditions, combined with the absence of pile-up subtraction for L1 jets, results in the expected observation of a decrease in the μ value of the efficiency turn-ons as a function of pile-up. Similarly, the resolution, σ, of the turn-ons are found to worsen at a higher number of primary



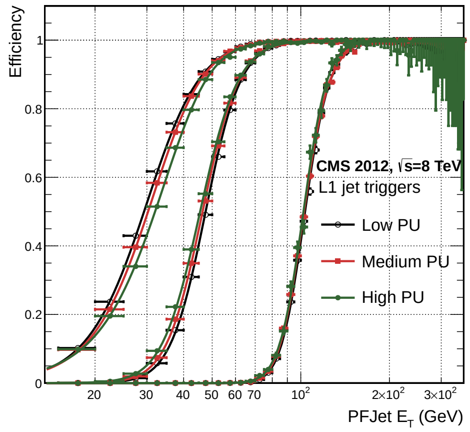


Figure 3.15: L1 jet efficiency turn-on curves as a function of the leading offline E_TCalo (left) and PF (right) jet, for low, medium and high pile-up conditions.

Vertices	0-10		11-20		> 20	
	μ	σ	μ	σ	μ	σ
L1_SingleJet16	19.9 $\pm$ 0.1	6.1 $\pm$ 0.3	20.8 $\pm$ 0.1	6.5 $\pm$ 0.1	22.3 $\pm$ 0.2	7.5 $\pm$ 0.1
L1_SingleJet36	41.8 $\pm$ 0.1	4.6 $\pm$ 0.1	40.9 $\pm$ 0.1	5.1 $\pm$ 0.1	40.6 $\pm$ 0.6	5.9 $\pm$ 0.2
L1_SingleJet92	95.9 $\pm$ 0.2	5.4 $\pm$ 0.1	95.2 $\pm$ 0.2	5.6 $\pm$ 0.1	94.5 $\pm$ 0.6	6.2 $\pm$ 0.3

Table 3.4: Results of a cumulative EMG function fit to the efficiency turn-on curves for L1 single jet triggers in the 2012 run period C, measured from isolated μ triggered data. The turn-on point, μ, and resolution, σ, of the L1 jet triggers are measured with respect to offline Calo jets in low (left), medium (middle) and high (right) pile-up conditions.

vertices due to the increasing size of the pile-up corrections being applied to the offline reconstructed jets.

These features are further emphasised when shown as a function of

$$\frac{(\text{L1 E}_T - \text{Offline E}_T)}{\text{Offline E}_T} \tag{3.4}$$

in bins of matched leading offline jet E_T. The results of these individual fits categorised as a function of matched leading offline jet E_T can be found in Appendix (B.2), where each of the distributions are fitted with an EMG function as defined in Equation (3.3).

The μ, σ and λ values extracted for the low, medium and high pile-up conditions are shown for Calo and PF jets in Figure 3.16 and Figure 3.17 respectively. The central value

Vertices	0-10		11-20		> 20	
	μ	σ	μ	σ	μ	σ
L1_SingleJet16	21.1 $\pm$ 0.1	7.16 $\pm$ 0.05	22.34 $\pm$ 0.1	7.9 $\pm$ 0.1	24.6 $\pm$ 0.2	9.5 $\pm$ 0.1
L1_SingleJet36	39.6 $\pm$ 0.1	7.4 $\pm$ 0.1	38.4 $\pm$ 0.1	7.4 $\pm$ 0.1	37.1 $\pm$ 0.2	7.5 $\pm$ 0.1
L1_SingleJet92	91.6 $\pm$ 0.3	11.3 $\pm$ 0.2	91.4 $\pm$ 0.3	11.2 $\pm$ 0.1	90.0 $\pm$ 0.9	12.1 $\pm$ 0.4

Table 3.5: Results of a cumulative EMG function fit to the efficiency turn-on curves for Level-1 single jet triggers in the 2012 run period C, measured from isolated μ triggered data. The turn-on point, μ, and resolution, σ, of the L1 jet triggers are measured with respect to offline PF jets in low (left), medium (middle) and high (right) pile-up conditions.

of $\frac{(\text{L1 E}_T - \text{Offline E}_T)}{\text{Offline E}_T}$ is observed as expected, to increase as a function of jet E_T, whilst the resolution also improves as a function of increasing offline jet E_T for all pile-up categories.

When comparisons are made between the individual pile-up scenarios, it can be seen that in the presence of higher pile-up , μ is seen to shift to larger values and a poorer resolution, σ, observed. This is particularly evident at low lead jet transverse energy values. These differences between the different pile-up scenarios, can once again be attributed to an increasing number of soft pile-up jets that add to the transverse energy of the lead jet from the primary interaction within each successive pile-up category. However, when comparisons of the trigger performance at larger lead jet transverse energy values (> 100 GeV) are made, similar performance is observed between the separate pile-up categories.

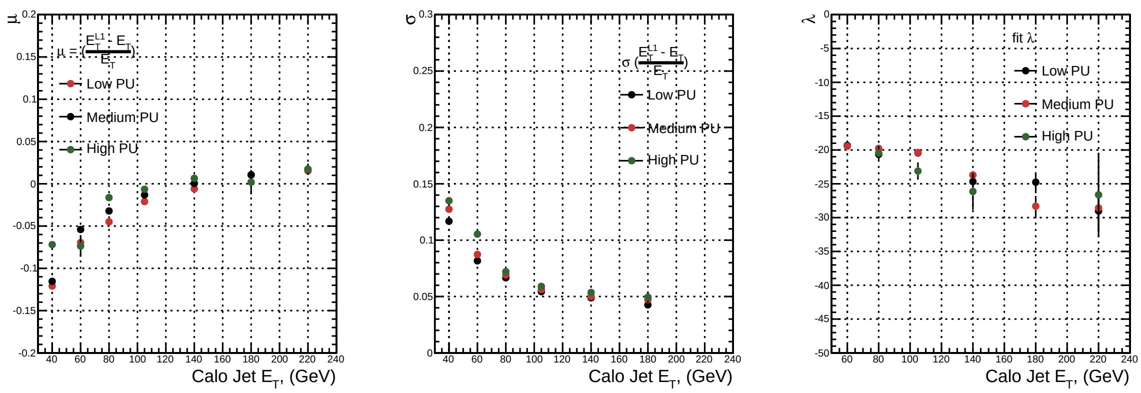


Figure 3.16: Fit values from an EMG function fitted to the resolution plots of leading Calo jet E_T measured as a function of $\frac{(\text{L1 E}_T - \text{Offline E}_T)}{\text{Offline E}_T}$ for low, medium and high pile-up conditions. The plots show the mean μ (left), resolution σ (middle) of the Gaussian as well as the decay term λ (right) of the exponential.

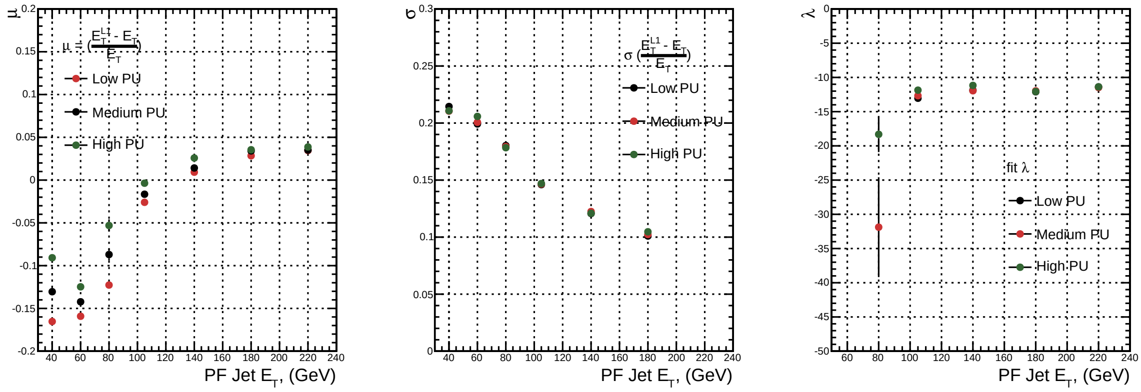


Figure 3.17: Fit values from an EMG function fitted to the resolution plots of leading PF jet E_T measured as a function of $\frac{(\text{L1 E}_T - \text{Offline E}_T)}{\text{Offline E}_T}$ for low and medium pile-up conditions. The plots show the mean μ (left), resolution σ (middle) of the Gaussian, as well as the decay term λ (right) of the exponential.

The resolution of the L1 jet based energy sum quantities, $\not\!\!H_T$ and H_T parameterised as in Equation (3.4), can be found in Appendix (B.3).

3.4.6. Summary

The performance of the CMS Level-1 Trigger has been studied and evaluated for jets and jet energy sum quantities using data collected during the 2012 LHC 8 TeV run. These studies include the effect of the introduction of a 5 GeV jet seed threshold into the jet clustering algorithm. The purpose of this change was to mitigate the increase in L1 trigger cross-sections, due to larger isotropic energy deposits from an increased number of secondary interactions, whilst not adversely affecting the efficiency of these triggers. Measurements are made for a range of L1 jet quantities and thresholds, where no significant change is observed in the measured efficiencies that would indicate a noticeable effect on the overall triggering performance of the detector.

Chapter 4.

SUSY Searches in Hadronic Final States

In this chapter a model independent search for SUSY, in hadronic final states with $\not{E}_T$ using the α_T variable is introduced and described in detail. The results presented are based on a data sample of pp collisions collected in 2012 at $\sqrt{s}$ =8 TeV, corresponding to an integrated luminosity of 11.7 ± 0.5 fb^{-1} [5].

The kinematic variable α_T is motivated as a variable to provide strong rejection of the overwhelming QCD multi-jet background, which is prevalent to jets $+\not{E}_T$ final states at the LHC. This is achieved whilst maintaining sensitivity to a range of possible SUSY signals and is described in Section (4.1). The search and trigger strategy in addition to the event reconstruction and selection are outlined within Sections (4.2 - 4.3).

The method in which the SM background is estimated using data driven control samples and an analytical technique to improve statistical precision at higher b-tagged jet multiplicities is detailed within Section (4.5). Included in this section is a discussion on the impact of b-tagging and mis-tagging scale factors between data and simulation on any background predictions. Improved precision in estimating background yields at large number of b-tagged jets, is important in the context of sensitivity to third generation SUSY models, first outlined in Section (2.4.1).

A description of the formulation of appropriate systematic uncertainties to be applied to the background predictions to account for theoretical uncertainties, limitations in the modelling of event kinematics and instrumental effects is covered in Section (4.6). Similarly the systematic determination for the SMS signal samples used to interpret the physics reach of the analysis are examined in Section (4.7).

Finally the statistical likelihood model to test the compatibility of the data with a SM only hypothesis, and to interpret the observations within the context of SMS models is described in Section (4.8). The experimental reach of the analysis discussed within this thesis is interpreted in two classes of SMS models, both introduced in Section (2.4.1). The SMS models considered in this analysis are summarised in Table 4.1. For each model, the LSP is assumed to be the lightest neutralino.

Within the table are also defined reference points, parameterised in terms of parent gluino/squark and LSP sparticle masses, m_{parent} and m_{LSP}, respectively. These are used within the following two chapters to demonstrate potential signal yields within the hadronic search region of the analysis. The masses of each signal topology are chosen to reflect parameter space which is within the expected sensitivity reach of the search.

Model	Production/decay mode	Reference model	
		m_{parent}	m_{LSP}
G1 (`T1`)	$pp \rightarrow \widetilde{g}\widetilde{g}^* \rightarrow q\bar{q}\widetilde{\chi}_1^0 q\bar{q}\widetilde{\chi}_1^0$	700	300
G2 (`T1bbbb`)	$pp \rightarrow \widetilde{g}\widetilde{g}^* \rightarrow b\bar{b}\widetilde{\chi}_1^0 b\bar{b}\widetilde{\chi}_1^0$	900	500
G3 (`T1tttt`)	$pp \rightarrow \widetilde{g}\widetilde{g}^* \rightarrow t\bar{t}\widetilde{\chi}_1^0 t\bar{t}\widetilde{\chi}_1^0$	850	250
D1 (`T2`)	$pp \rightarrow \widetilde{q}\widetilde{q}^* \rightarrow q\widetilde{\chi}_1^0 \bar{q}\widetilde{\chi}_1^0$	600	250
D2 (`T2bb`)	$pp \rightarrow \widetilde{b}\widetilde{b}^* \rightarrow b\widetilde{\chi}_1^0 \bar{b}\widetilde{\chi}_1^0$	500	150

Table 4.1: A summary of the SMS models interpreted in this analysis, involving both direct (D) and glunio-induced (G) production of squarks and their decays. Reference models are also defined in terms of parent and LSP sparticle mass.

4.1. An Introduction to the α_T Search

A proton-proton collision resulting in the production and decay of supersymmetric particles, would manifest as a final state containing energetic jets and $\not{E}_T$ in the hadronic channel. The search focuses on topologies where new heavy supersymmetric, R-parity conserving particles are pair-produced in pp collisions. These particles decaying to a LSP escape the detector undetected, leading to significant missing energy and missing hadronic transverse energy,

$$\not{H}_T = |\sum_{i=1}^{n} \vec{p_T}^{jet_i}|. \tag{4.1}$$

This is defined as the vector sum of the transverse energies of jets selected in an event. Energetic jets produced in the decay of these supersymmetric particles also can produce significant visible transverse energy,

$$H_T = \sum_{i=1}^{n} E_T^{jet_i}, \tag{4.2}$$

defined as the scalar sum of the transverse energies of jets selected in an event.

A search within this channel is greatly complicated in a hadron collider environment; where the overwhelming background comes from inherently balanced multi-jet ("QCD") events, which are produced with an extremely large cross-section as demonstrated within Figure 4.1. $\not{E}_T$ can appear in such events due to a substantial detector mis-measurement, stochastic fluctuations of jet energy, or missed objects due to detector mis-calibration or noise effects.

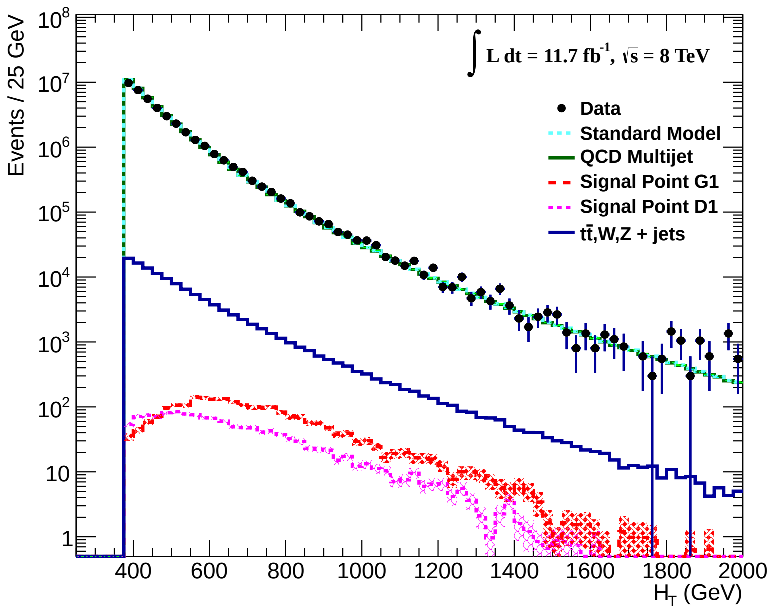


Figure 4.1: Reconstructed offline H_T distribution in the hadronic signal selection (detailed in the following section), from 11.7fb^{-1} of data, in which no α_T requirement was made. The sample is collected from prescaled H_T triggers. Overlaid are expectations from simulation of EWK processes as well as two reference signal models (labelled G1 and D1 from Table 4.1).

Additional SM background from EWK processes with genuine $\not\!\!E_T$ from escaping neutrinos comprise the irreducible background within this search and come mainly from:

- $Z \to \nu\bar{\nu}$ + jets,
- $W \to l\nu$ + jets in which a lepton falls outside of detector acceptance, is not reconstructed, is mis-identified, or the lepton decays hadronically $\tau \to$ had,
- $t\bar{t}$ with at least one leptonically decaying W, which is missed in the detector as detailed above,
- small background contributions from DY, single top and Diboson (WW,ZZ,WZ) processes.

The search is designed to have a strong separation between events with genuine and "fake" $\not\!\!E_T$ which is achieved primarily though the dimensionless kinematic variable, α_T [73][74].

4.1.1. The α_T Variable

For a perfectly measured di-jet QCD event, conservation laws dictate that both jets must be of equal magnitude and produced in opposite directions. However, in the case of di-jet events with genuine $\not\!\!E_T$ (as detailed above), no such requirement is made of the two jets, as depicted in Figure 4.2.

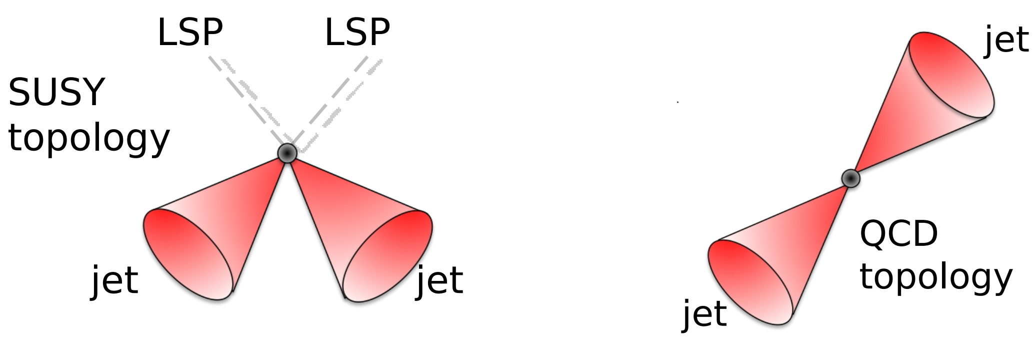


Figure 4.2: The event topologies of background QCD diet events (right) and a generic SUSY signature with genuine $\not\!\!E_T$ (left).

Exploiting this feature leads to the formulation of α_T (first inspired by [75]) in di-jet systems defined as,

$$\alpha_T = \frac{E_T^{j_2}}{M_T}, \tag{4.3}$$

where $E_T^{j_2}$ is the transverse energy of the least energetic of the two jets and M_T defined as:

$$M_T = \sqrt{\left(\sum_{i=1}^{2} E_T^{j_i}\right)^2 - \left(\sum_{i=1}^{2} p_x^{j_i}\right)^2 - \left(\sum_{i=1}^{2} p_y^{j_i}\right)^2} \equiv \sqrt{H_T^2 - \not{H}_T{}^2}. \tag{4.4}$$

A perfectly balanced di-jet event i.e. $E_T^{j_1} = E_T^{j_2}$ would yield an α_T value of 0.5. In processes where a W or Z recoils off a system of jets, these jets will not necessarily be perfectly balanced and α_T can then achieve values in excess of 0.5. Most importantly, balanced multi-jet events in which jets *are* mis-measured, will generally result in an α_T value of less than 0.5, thus giving the α_T variable discriminating power between these processes.

α_T can be further extended to apply to any arbitrary number of jets. This is undertaken by modelling a system of n jets as a di-jet system, through the formation of two pseudo-jets [76]. The two pseudo-jets are built by merging the jets present, and are chosen to be as balanced as possible, i.e the $\Delta H_T \equiv |E_T^{pj_1} - E_T^{pj_2}|$ is minimised between the two pseudo jets. Using Equation (4.4), α_T can be rewritten as,

$$\alpha_T = \frac{1}{2}\frac{H_T - \Delta H_T}{\sqrt{H_T{}^2 - \not{H}_T{}^2}} = \frac{1}{2}\frac{1 - \Delta H_T/H_T}{\sqrt{1 - (\not{H}_T/H_T)^2}}. \tag{4.5}$$

The distribution of α_T for the two jet multiplicity categories used within this analysis, $2 \leq n_{jet} \leq 3$ and $n_{jet} \geq 4$ jets, is shown in the Figure 4.3. It can be seen that the distributions peak at an α_T value of 0.5, before falling away sharply and being free of a simulated multi-jet background at larger α_T values. These distributions serve to demonstrate the ability of the α_T variable to discriminate between multi-jet events and EWK processes with genuine $\not{E}_T$ in the final state.

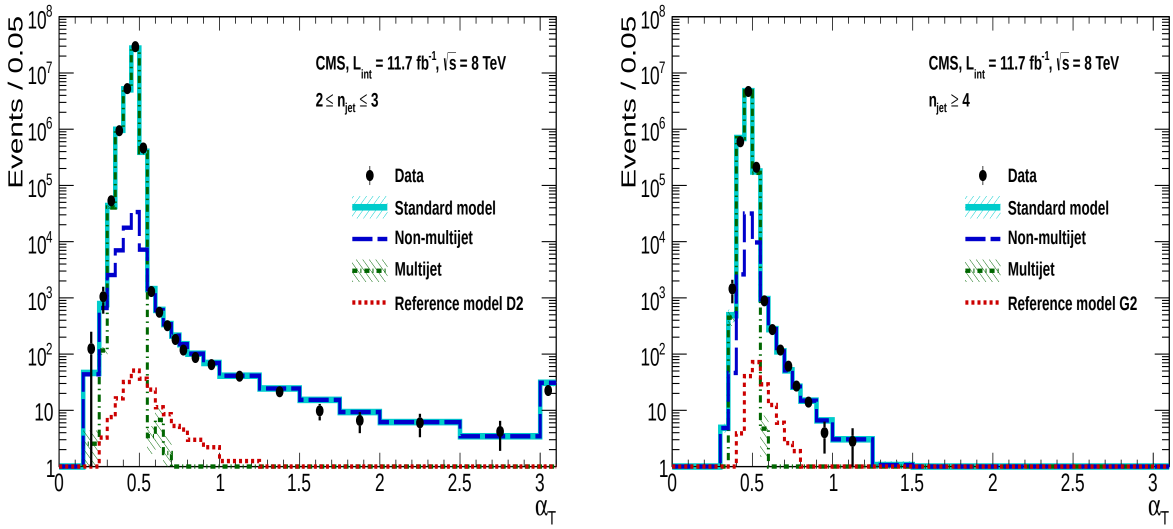


Figure 4.3: The α_T distributions for the low 2-3 (left) and high ≥ 4 (right) jet multiplicities after a full analysis selection and $H_T > 375$ requirement. Data is collected using both prescaled H_T triggers and dedicated α_T triggers for below and above $\alpha_T = 0.55$ respectively. Expected yields as given by simulation are also shown for multi-jet events (green dash-dotted line), EWK backgrounds with genuine $\not{E}_T$ (blue long-dashed line), the sum of all SM processes (cyan solid line) and the reference signal model D2 (left, red dotted line) or G2 (right, red dotted line).

The α_T requirement used within the search is chosen to be $\alpha_T > 0.55$ to ensure that the QCD multi-jet background is negligible even in the presence of moderate jet mis-measurement. There still remain other effects which can cause multi-jet events to artificially have a large α_T value, methods to combat them are discussed in detail in Section (4.2.2).

4.2. Search Strategy

The aim of the analysis presented in this thesis is to identify an excess of events in data over the SM background expectation in multi-jet final states and significant $\not{E}_T$. The essential suppression of the dominant multi-jet background for such a search is addressed by the α_T variable, described in the previous section. For estimation of the remaining EWK backgrounds, three independent data control samples are used to predict the different processes that compose the background :

- μ + jets control sample to determine W + jets, $t\bar{t}$ and single top backgrounds,
- γ + jets control sample to determine the irreducible Z $\rightarrow \nu\bar{\nu}$ + jets background,

- $\mu\mu$ + jets control sample to also determine the irreducible $Z \rightarrow \nu\bar{\nu}$ + jets background.

These control samples are chosen to be rich in specific EWK processes, free of QCD multi-jet events and to also be kinematically similar to the hadronic signal region that they are estimating the backgrounds of, see Section (4.2.3). The redundancy of using the γ + jets and $\mu\mu$ + jets sample to predict the same background within the signal region, brings an opportunity to reliably crosscheck and validate the background estimation method, and is utilised in both the determination of background estimation systematics (Section(4.6)) and in the maximum likelihood fit (Section(4.8)).

To remain inclusive to a large range of possible SUSY models, the signal region is split into the following categories to allow for increased sensitivity to different SUSY topologies:

Sensitivity to a range of SUSY mass splittings

The hadronic signal region is defined by $H_T > 275$, divided into eight bins in H_T.

- Two bins of width 50 GeV in the range $275 < H_T < 375$ GeV,
- five bins of width 100 GeV in the range $375 < H_T < 875$ GeV,
- and a final open bin, $H_T > 875$ GeV.

The choice of the lowest H_T bin in the analysis is driven primarily by trigger constraints. The mass difference between the LSP and the particle that it decays from is an important factor in the amount of hadronic activity in the event.

A large mass splitting will lead to hard high p_T jets which contribute to the H_T sum. From Figure 4.1 it can be seen that the SM background falls sharply at high H_T values, therefore many H_T categories will lead to easier identification of such signals. Conversely, smaller mass splittings lead to softer jet p_T's which will subsequently fall into the lower H_T range.

Sensitivity to production method of SUSY particles

The production mechanism of any potential SUSY signal can lead to different event topologies. One such way to discriminate between gluino ($g\widetilde{g}$ - "high multiplicity"), and direct squark ($q\widetilde{q}$ - "low multiplicity") induced production of SUSY particles is realised through the number of reconstructed jets in the final state.

The analysis is thus split into two jet categories: $2 \leq n_{\mathrm{jet}} \leq 3$ jets, $n_{\mathrm{jet}} \geq 4$ jets to give sensitivity to both of these mechanisms.

Sensitivity to "Natural SUSY" via tagging jets from b-quarks

Jets originating from the hadronisation of bottom quarks (b-jets) are identified through vertices that are displaced with respect to the primary interaction. The algorithm used in the analysis to identify b-jets is the Combined Secondary Vertex Medium Working Point (CSVM) tagger, described within Section (3.3.2). A cut is placed on the discriminator variable of > 0.679, leading to a gluon/light-quark tagging efficiency of $\sim$ 1%, a c-quark tagging efficiency of $\sim$ 20% and a jet $\mathrm{p_T}$ dependant b-tagging efficiency of 60-70% [77].

Natural SUSY models would be characterised through final-state signatures rich in bottom quarks. A search relying on methods to identify jets originating from bottom quarks through b-tagging, will significantly improve the sensitivity to this class of signature. This gain in sensitivity stems from a vast reduction in the vector boson + jet backgrounds (W, Z) at higher b-tag jet multiplicities, which typically have no b-flavoured quarks in their decays.

Therefore, events are categorised according to the number of offline reconstructed b-tagged jets, $n_{\mathrm{b}}^{\mathrm{reco}}$, identified within each event. The following five categories are used; $n_{\mathrm{b}}^{\mathrm{reco}}$ =0, =1, =2, =3 and ≥ 4. In the $n_{\mathrm{b}}^{\mathrm{reco}} \geq 4$ category, due to a limited number of expected signal and background events just three H_T bins are employed: 275-325 GeV, 325-375 GeV, $\geq$ 375 GeV.

This characterisation is identically mirrored in all control samples, with the information from all samples and b-tag categories used simultaneously in the likelihood model, see Section (4.8).

The combination of the H_T, jet multiplicity and b-tag categorisation of the signal region as described above, results in 67 different bins in which the analysis is interpreted in. A visualisation of the analysis categorisation is depicted in Figure 4.4.

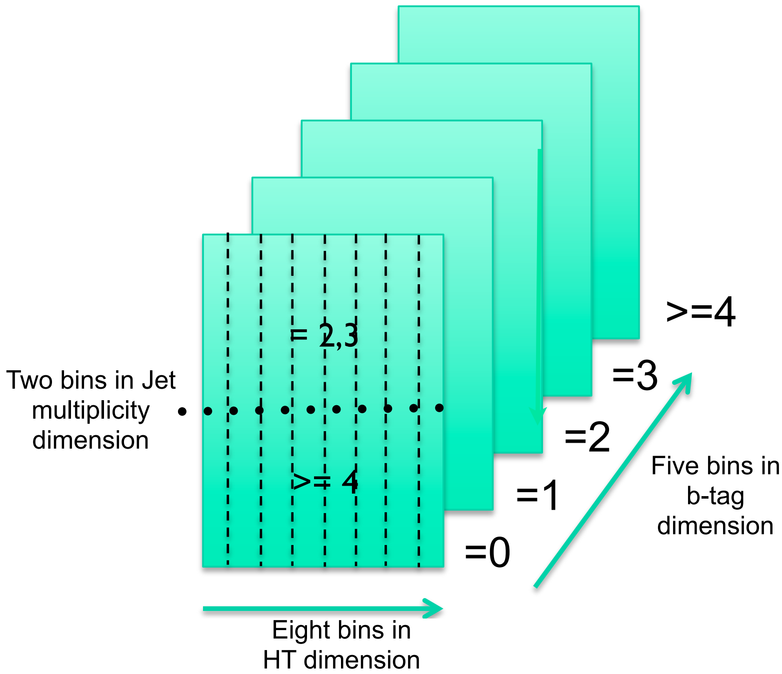


Figure 4.4: Pictorial depiction of the analysis strategy employed by the α_T search to increase sensitivity to a wide spectra of SUSY models.

4.2.1. Physics Objects

The physics objects used in the analysis are defined below, and follow the recommendation of the various CMS Physics Object Groups (POGs).

- **Jets**

  The jets used in this analysis are CaloJets, reconstructed as described in Section (3.3.1) using the anti-k$_T$ jet clustering algorithm.

  To ensure the jet object falls within the calorimeter systems a pseudo-rapidity requirement of $|\eta| < 3$ is applied. Each jet must pass a "loose" identification criteria to reject jets resulting from unphysical energy, the criteria of which are detailed in Table A.1 [78].

- **Muons**

  Muons are selected in the μ + jets and $\mu\mu$ + jets control samples, and vetoed in the signal region. The same cut based identification criteria is applied to muons in both search regions and is summarised in Table 4.2 [79].

Variable	Definition
Is Global Muon	Muon contains both a hit in the muon chamber and a matched track in the inner tracking system.
$\chi^2 < 10$	χ^2 of global muon track fit. Used to suppress hadronic punch-through and muons from decays in flight.
Muon chamber hits > 0	At least one muon chamber hit included in global muon track fit.
Muon station hits > 1	Muon segment hits in at least two muon stations, which suppresses hadronic punch-through and accidental track-to-segment matches.
$d_{xy} < 0.2$mm	The tracker track transverse impact parameter w.r.t the primary vertex. Suppresses cosmic muons and muons from decays in flight.
$d_z < 0.5$mm	The longitudinal distance of the tracker track w.r.t the primary vertex. Loose selection requirement to further suppress cosmic muons, muons from decays in flight and tracks from pileup.
Pixel hits > 0	Suppresses muons from decays in flight by requiring at least one pixel hit in the tracker.
Track layer hits > 5	Number of tracker layers with hits, to guarantee a good p_T measurement. Also suppresses muons from decays in flight.
PF Iso <0.12	Isolation based upon the sum of the charged and neutral hadrons and photon objects within a ΔR 0.4 cone of the muon object, corrected for pile up effects on the isolation sum.

Table 4.2: Muon identification criteria used within the analysis for selection/veto purposes in the muon control/signal selections.

Additionally muons are required to be within the acceptance of the muon tracking systems. For the muon control samples, trigger requirements necessitate a $|\eta| < 2.1$ for the selection of muons. In the signal region where muons are vetoed, these conditions are relaxed to $|\eta| < 2.5$ and a minimum threshold of $p_T > 10$ GeV is required of muon objects.

- **Photons**

Photons are selected within the γ + jets control sample and vetoed in all other selections. Photons are identified in both cases according to the cut based criteria listed in Table 4.3 [80].

Variable	Definition
H/E $<$ 0.05	The ratio of hadronic energy in the HCAL tower directly behind the ECAL super-cluster and the ECAL super-cluster itself.
$\sigma_{i\eta i\eta} < 0.011$	The log energy weighted width (σ), of the extent of the shower in the η dimension.

Continued on next page

R9 < 1.0	The ratio of the energy of the 3×3 crystal core of the super-cluster compared to the total energy stored in the 5×5 super-cluster.
Combined Isolation < 6 GeV	The photons are required to be isolated with no electromagnetic or hadronic activity within a radius $\Delta R = 0.3$ of the photon object. A combination of the pileup subtracted [81], ECAL, HCAL and tracking isolation sums are used to determine the combined total isolation value.

Table 4.3: Photon identification criteria used within the analysis for selection/veto purposes in the γ + jets control/signal selections.

Photon objects are also required to have a minimum momentum of $p_T > 25$ GeV.

- **Electrons**

Electron identification is defined for veto purposes. They are selected according to the following cut-based criteria listed in Table 4.4, utilising PF-based isolation.

Variable	Barrel	EndCap	Definition
$\Delta\eta_{In}$	<0.007	<0.009	$\Delta\eta$ between SuperCluster position and the coordinate of the associated track at the interaction vertex, assuming no radiation.
$\Delta\phi_{In}$	<0.15	<0.10	$\Delta\phi$ between SuperCluster position and track direction at interaction vertex extrapolated to ECAL assuming no radiation.
$\sigma_{i\eta i\eta}$	<0.01	<0.03	Cluster shape covariance, measure the η dispersion of the electrons electromagnetic shower over the ECAL supercluster.
H/E	<0.12	<0.10	The ratio of hadronic energy in the HCAL tower directly behind the ECAL super-cluster and the ECAL super-cluster itself.
d0 (vtx)	<0.02	<0.02	The tracker track transverse impact parameter w.r.t the primary vertex.
dZ (vtx)	<0.20	<0.20	The longitudinal distance of the tracker track w.r.t the primary vertex.
$\lvert(\frac{1}{E_{ECAL}} - \frac{1}{p_{track}})\rvert$	<0.05	<0.05	Comparison of energy at supercluster $1/E_{ECAL}$ and that of the track momentum at the vertex $1/p_{track}$. Causes suppression of fake electrons at low p_T.
PF Iso	<0.15	<0.15	Combined PF isolation of charged hadrons, photons, neutral hadrons within a $\Delta R < 0.3$ cone size. Isolation sum is corrected for pileup using effective area corrections for neutral particles.

Table 4.4: Electron identification criteria used within the analysis for veto purposes.

Electrons are required to be identified at $|\eta| < 2.5$, with a minimum $p_T > 10$ GeV threshold to ensure that the electrons fall within the tracking system of the detector.

- **Noise and $\not{E}_T$ Filters**

  A series of noise filters are applied to veto events which contain spurious non-physical jets that are not picked up by the jet id, and events which give large unphysical $\not{E}_T$ values. These filters are listed within Table 4.5.

Noise Filters Variable	Definition
CSC tight beam halo filter	As proton beams circle the LHC, proton interactions with the residual gas particles or the beam collimators can occur, producing showers of secondary particles which can interact with the CMS detector.
HBHE noise filter with isolated noise rejection	Anomalous noise in the HCAL not due to electronics noise, but rather due to instrumentation issues associated with the HPD's and Readout Boxes (RBXs).
HCAL laser filter	The HCAL uses laser pulses for monitoring the detector response. Some laser pulses have accidentally been fired in the physics orbit, and ended up polluting events recorded for physics analysis.
ECAL dead cell trigger primitive (TP) filter	EB and EE have single noisy crystals which are masked in reconstruction. Use the Trigger Primative (TP) information to assess how much energy was lost in masked cells.
Bad EE Supercrystal filter	Two supercrystals in EE are found to occasionally produce high amplitude anomalous pulses in several channels at once, causing a large $\not{E}_T$ spike.
ECAL Laser correction filter	A laser calibration multiplicative factor is applied to correct for transparency loss in each crystal during irradiation. A small number of crystals receive unphysically large values of this correction and become very energetic, resulting in $\not{E}_T$.

Table 4.5: Noise filters that are applied to remove spurious and non-physical $\not{E}_T$ signatures within the CMS detector.

4.2.2. Event Selection

The selection criteria for events within the analysis are detailed below. A set of common cuts are applied to both signal (maximise acceptance to a range of SUSY signatures), and control samples (retain similar jet kinematics for background predictions), with additional selection cuts applied to each control sample to enrich the sample in a particular EWK processes, see Section (4.2.3).

The jets considered in the analysis are required to have a transverse momentum $p_T > 50$ GeV, with a minimum of two jets required in the event. The highest E_T jet is required to lie within the central tracker acceptance $|\eta| < 2.5$, and the two leading p_T jets must each have $p_T > 100$ GeV. Any event which has a jet with $p_T > 50$ GeV that either fails the "loose" identification criteria described in Section(4.2.1) or has $|\eta| > 3.0$, is rejected. Similarly events in which an electron, muon or photon fails object identification but passes η and p_T restrictions, are identified as an "odd" lepton/photon and the event is vetoed.

At low H_T, the jet p_T threshold requirements required to be considered as part of the analysis and enter the H_T sum are scaled downwards. These are scaled down in order to extend phase space at low H_T, preserving similar jet multiplicities and background admixture seen at higher H_T, as listed in Table 4.6.

H_T bin	minimum jet p_T	second leading jet p_T
$275 < H_T < 325$	36.7	73.3
$325 < H_T < 375$	43.3	86.6
$375 < H_T$	50.0	100.0

Table 4.6: Jet thresholds used in the three H_T regions of the analysis.

Within the signal region, to suppress SM processes with genuine $\not\!\!E_T$ from neutrinos, events containing isolated electrons or muons are vetoed. Furthermore to ensure a pure multi-jet topology, events are vetoed if an isolated photon is found with $p_T > 25$ GeV.

An α_T requirement of > 0.55 is required to reduce the QCD multi-jet background to a negligible amount. Finally, additional cleaning cuts are applied to protect against pathological deficiencies such as reconstruction failures or severe energy mis-measurements due to detector inefficiencies:

- Significant $\not\!\!H_T$ can arise in events with no real $\not\!\!E_T$ due to multiple jets falling below the p_T threshold for selecting jets. This in turn leads to events which can then incorrectly pass the α_T requirements of the analysis. This effect can be negated by requiring that the missing transverse momentum reconstructed from jets alone does not greatly exceed the missing transverse momentum reconstructed from all of the detector's calorimeter towers,

$$R_{miss} = \not\!\!H_T / \not\!\!E_T < 1.25.$$

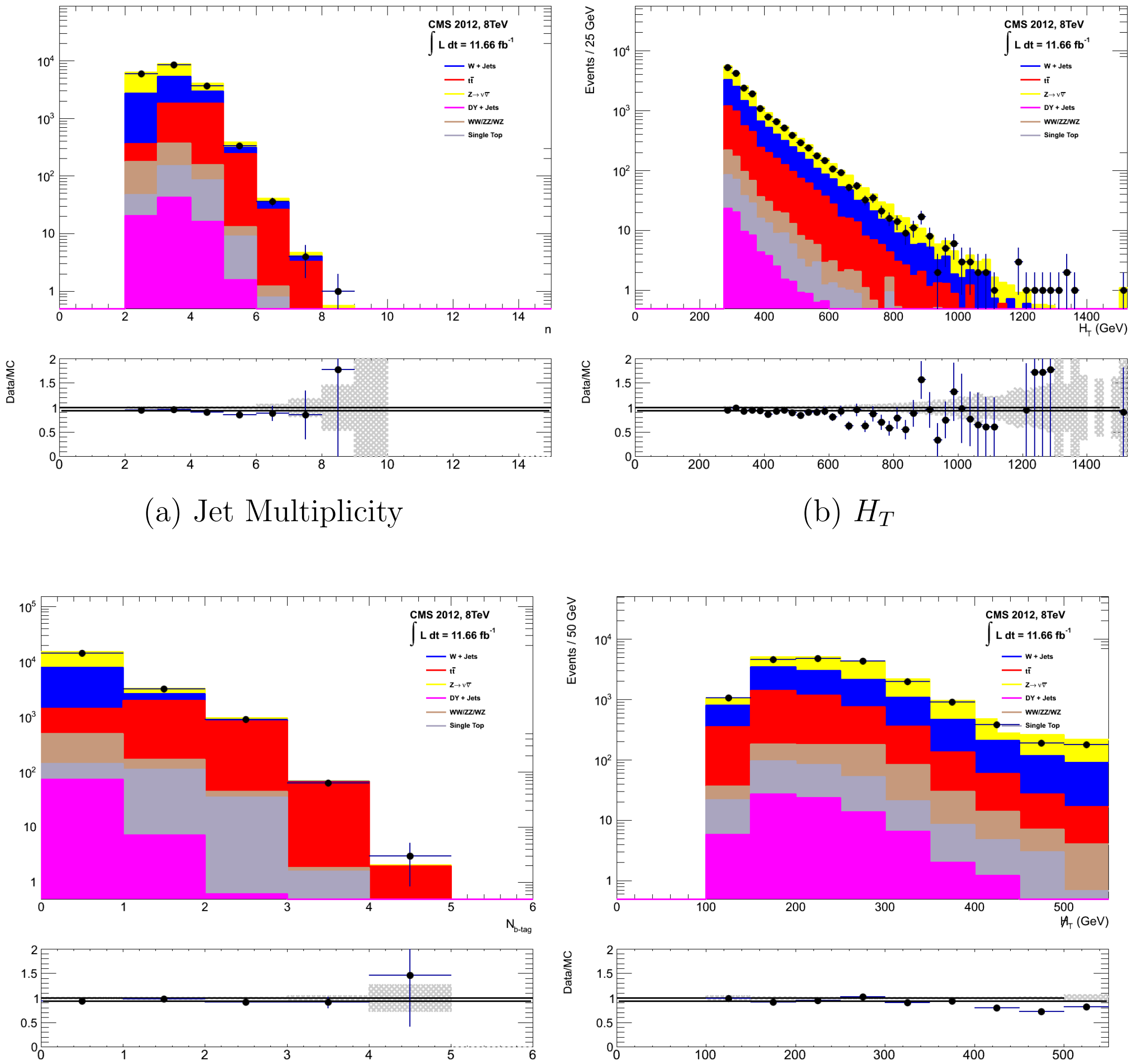
- Fake $\not{E}_T$ and $\not{H}_T$ can arise due to significant jet mis-measurements caused by a small number of non-functioning ECAL regions. These regions absorb electromagnetic showers which are subsequently not added to the jet energy sum. To circumvent this problem the following procedure is employed: For each jet in the event, the angular separation

$$\Delta\phi_j^* \equiv \Delta\phi(\vec{p_j} - \sum_{i \neq j} \vec{p_i}), \tag{4.6}$$

  is calculated where that jet is itself removed from the event. Here $\Delta\phi^*$ is a measure of how aligned the $\not{H}_T$ of an event is with a jet. A small value (i.e. the $\not{H}_T$ vector lies along the jet axis) is indicative of an inherently balanced event in which a jet has been mis-measured. For every jet in an event with $\Delta\phi^* < 0.5$, if the ΔR distance between the selected jet and the closest dead ECAL region is also < 0.3, then the event is rejected. Similarly events are rejected if the jet points within $\Delta R < 0.3$ of the ECAL barrel-endcap gap at $|\eta| = 1.5$.

Some of the key distributions of the analysis are compared to simulated SM processes, shown in Figure 4.5. The simulated samples are normalised to a luminosity of 11.7 fb^{-1}, with no requirement placed upon the number of b-tagged jets or number of jets in the distributions shown. In the case of this inclusive selection, the dominant backgrounds in the signal regions are, Z $\rightarrow \nu\bar{\nu}$ and W + jet processes, with a smaller $t\bar{t}$ background accompanied by other residual backgrounds.

The distributions shown are presented for purely illustrative purposes, with the simulation not used in absolute terms for the estimation of background processes within the signal region, see Sections (4.2.3, 4.5). However it is nevertheless important to demonstrate that good agreement exists between the modelling of key variables in simulation and data.



(a) Jet Multiplicity

(b) H_T

(c) Btag Multiplicity

(d) $\not{H}_T$

Figure 4.5: Data/MC comparisons of key variables for the hadronic signal region, following the application of the hadronic selection criteria and the requirements of $H_T >$ 275 GeV and $\alpha_T > 0.55$. Bands represent the uncertainties due to the statistical size of the MC samples. No requirement is made upon the number of b-tagged jets or jet multiplicity in these distributions.

4.2.3. Control Sample Definition and Background Estimation

The method used to estimate the background contributions in the hadronic signal region relies on the use of a Transfer Factor (TF). This is determined from simulation in both the control, $\mathrm{N}_{\mathrm{MC}}^{\mathrm{control}}$, and signal, $\mathrm{N}_{\mathrm{MC}}^{\mathrm{signal}}$, region to transform the observed yield measured in data for a control sample, $\mathrm{N}_{\mathrm{obs}}^{\mathrm{control}}$, into a background prediction, $\mathrm{N}_{\mathrm{pred}}^{\mathrm{signal}}$, via Equation (4.7),

$$\mathrm{N}_{\mathrm{pred}}^{\mathrm{signal}} = \frac{\mathrm{N}_{\mathrm{MC}}^{\mathrm{signal}}}{\mathrm{N}_{\mathrm{MC}}^{\mathrm{control}}} \times \mathrm{N}_{\mathrm{obs}}^{\mathrm{control}}. \tag{4.7}$$

All simulation samples are normalised to the luminosity of the data samples of the relevant selection they are being applied to. Through this method, “vanilla” predictions for the SM background in the signal region can be made by considering separately the sum of the prediction from the combination of either the μ + jets and γ + jets, or μ + jets and $\mu\mu$ + jets samples.

It must be noted that the final background estimation from which results are interpreted, is calculated via a fitting procedure defined formally by the likelihood model described in Section (4.8).

The sum of the expected yields from all simulated processes listed in Section (4.1), enter the denominator, $\mathrm{N}_{\mathrm{MC}}^{\mathrm{control}}$, of the TF defined in Equation (4.7) for each control sample. However, only the specific processes being estimated by the control sample enter the numerator, $\mathrm{N}_{\mathrm{MC}}^{\mathrm{signal}}$.

For the μ + jets sample the processes entering the numerator are,

$$\mathrm{N}_{\mathrm{MC}}^{\mathrm{signal}}(H_T, n_{\mathrm{jet}}) = N_W + N_{t\bar{t}} + N_{DY} + N_t + N_{di-boson}, \tag{4.8}$$

whilst for both the $\mu\mu$ + jets and γ + jets samples the only simulated processes used in the numerator is,

$$\mathrm{N}_{\mathrm{MC}}^{\mathrm{signal}}(H_T, n_{\mathrm{jet}}) = N_{Z\rightarrow\nu\bar{\nu}}. \tag{4.9}$$

The control samples and the EWK processes they are specifically tuned to select are defined below, with distributions of key variables for each of the control samples shown for illustrative purposes in Figures 4.6, 4.7 and 4.8. No requirement is placed upon the number of b-tagged jets or jet multiplicity in the distributions shown. The distributions highlight the background compositions of each control sample, where in general, good agreement is observed between data and simulation, giving confidence that the samples are well understood. The contribution from QCD multi-jet events is expected to be negligible:

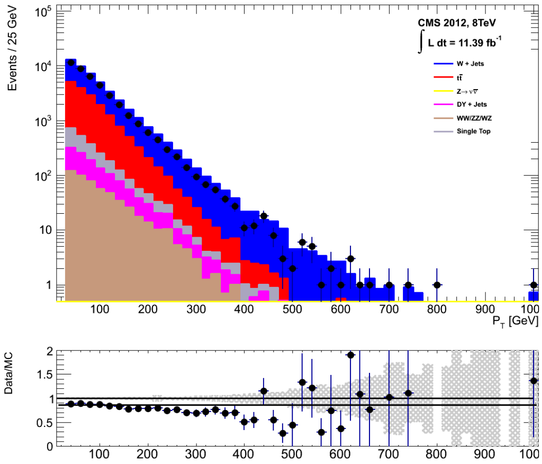
The μ + jets control sample

Events from W + jets and $t\bar{t}$ processes enter into the hadronic signal sample due to unidentified leptons from acceptance effects or reconstruction inefficiencies and hadronic tau decays. These leptons originate from the decay of high p_T W bosons.

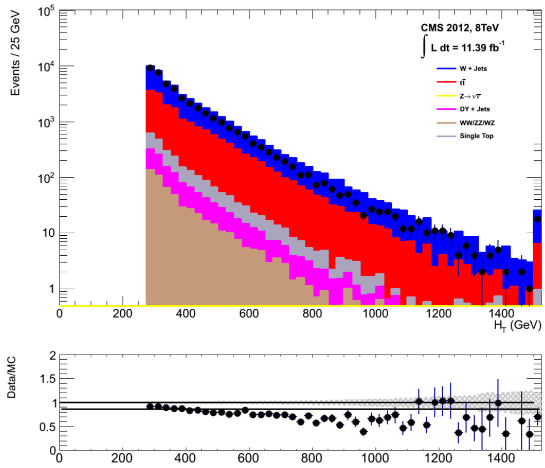
The control sample specifically identify $W \to \mu\bar{\nu}$ decays within a similar phase-space of the signal region, where the muon is subsequently ignored in the calculation of event level variables, i.e. H_T, $\not{H}_T$, α_T.

All kinematic jet-based selection criteria are identical to those applied in the hadronic search region (with the exception of an α_T, requirement discussed below) detailed in Section (4.2.2), with the same H_T, jet multiplicity and b-jet multiplicity categorisation described previously. Furthermore, the following selection criteria are also required:

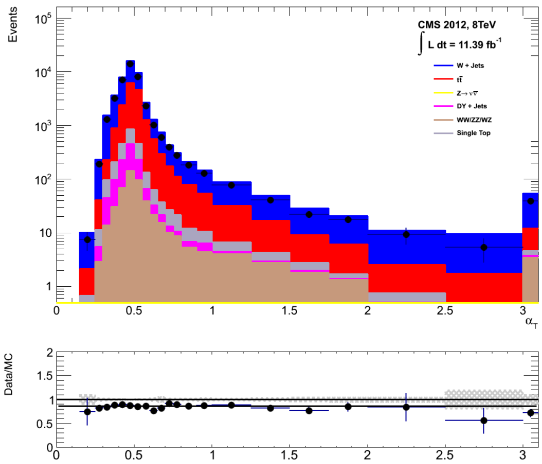
- Muons originating from W boson decays are selected by requiring one tightly isolated muon defined in Table 4.2, with a $p_T > 30$ GeV and $|\eta| < 2.1$. Both of these thresholds arise from trigger restrictions.
- The transverse mass of the W candidate must satisfy M_T $(\mu, \not{E}_T) > 30$ GeV (to suppress QCD multi-jet events).
- Events which contain a jet overlapping with a muon $\Delta\mathrm{R}(\mu, \mathrm{jet}) < 0.5$ are vetoed to remove events from muons produced as part of a jet's hadronisation process.
- Events containing a second muon candidate which has failed id, but passing p_T and $|\eta|$ requirements, are checked to have an invariant mass that satisfies $|M_{\mu\mu} - m_Z| > 25$, thus removing $Z \to \mu\mu$ contamination.



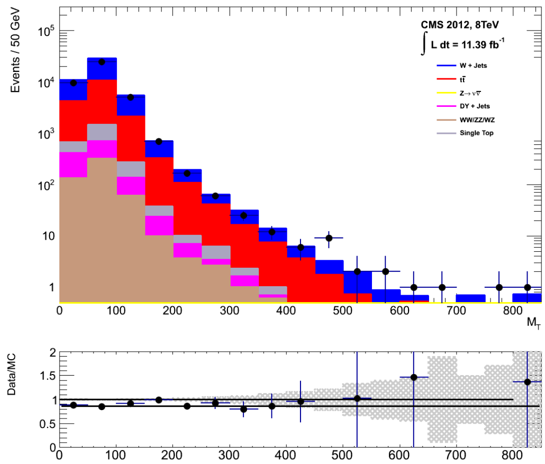
(a) Lead Muon p_T



(b) H_T



(c) α_T



(d) Transverse mass M_T

Figure 4.6: Data/MC comparisons of key variables for the μ + jets selection,following the application of selection criteria and the requirements that $H_T > 275$ GeV. Bands represent the uncertainties due to the statistical size of the MC samples. No requirement is made upon the number of b-tagged jets or jet multiplicity in these distributions.

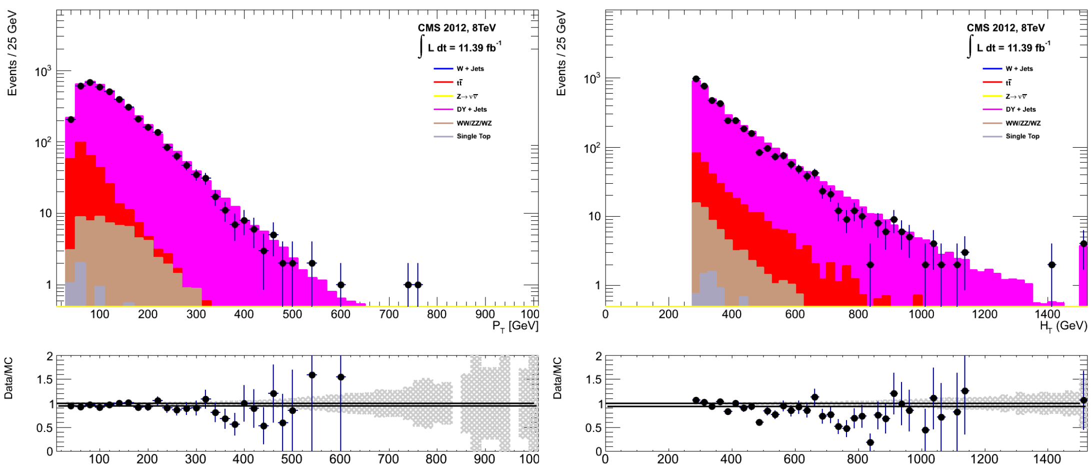
The $\mu\mu$ + jets control sample

An irreducible Z $\rightarrow \nu\bar{\nu}$ + jets background enters into the signal region from genuine $\not{E}_T$ from the escaping neutrinos. This background is estimated using two control samples, the first of which is the Z$\rightarrow$ $\mu\bar{\mu}$ + jets process, which posses identical kinematic properties, but with a different acceptance and branching ratio [1].

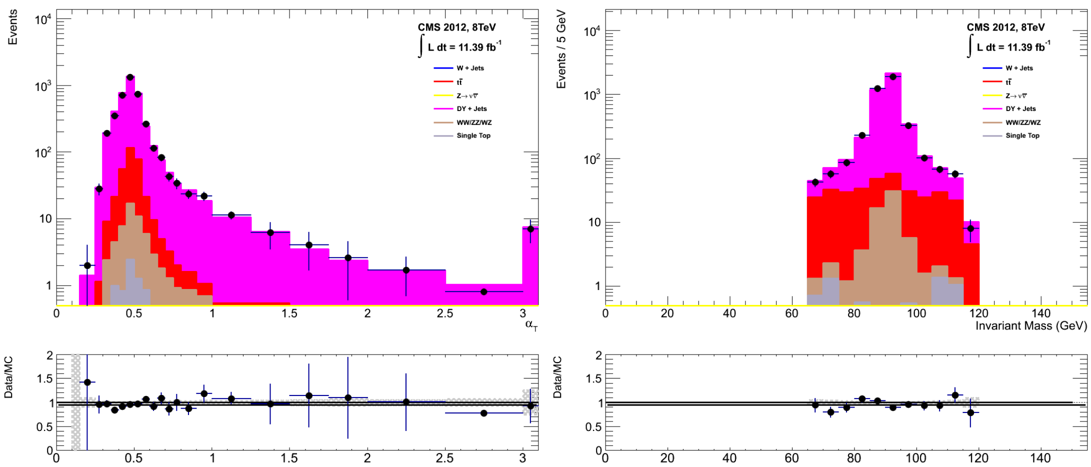
The same acceptance requirements as the μ + jets selection for muons is applied, as defined in Table 4.2. Muons in the event are ignored for the purpose of the calculation of event level variables. In addition to kinematic jet-based selection criteria (with the exception of an α_T requirement, discussed below) and event categorisation, which are identical to the hadronic search region, the following selection criteria are also specified:

- Muons originating from a Z boson decay are selected, requiring exactly two tightly isolated muons. Due to trigger requirements the leading muon is required to have $p_T > 30$ GeV and $|\eta| < 2.1$. The requirement of the p_T on the second muon is relaxed to 10 GeV.
- Events are vetoed if containing a jet overlapping with a muon $\Delta R(\mu, \text{jet}) < 0.5$.
- In order to specifically select two muons both originating from a single Z boson decay, the invariant mass of the two muons must satisfy $|M_{\mu\mu} - m_Z| < 25$.

The $\mu\mu$ + jets sample is able to make predictions in the signal region of the two lowest H_T bins, providing coverage where the γ + jets sample is unable to, due to trigger requirements. In higher H_T bins, the higher event statistics of the γ + jets sample is also used in determining the Z $\rightarrow \nu\bar{\nu}$ estimation.



(a) Lead Muon p_T (b) H_T



(c) α_T (d) $\mu\mu$ invariant mass

Figure 4.7: Data/MC comparisons of key variables for the $\mu\mu$ + jets selection,following the application of selection criteria and the requirements that $H_T > 275$ GeV. Bands represent the uncertainties due to the statistical size of the MC samples. No requirement is made upon the number of b-tagged jets or jet multiplicity in these distributions.

The γ + jets control sample

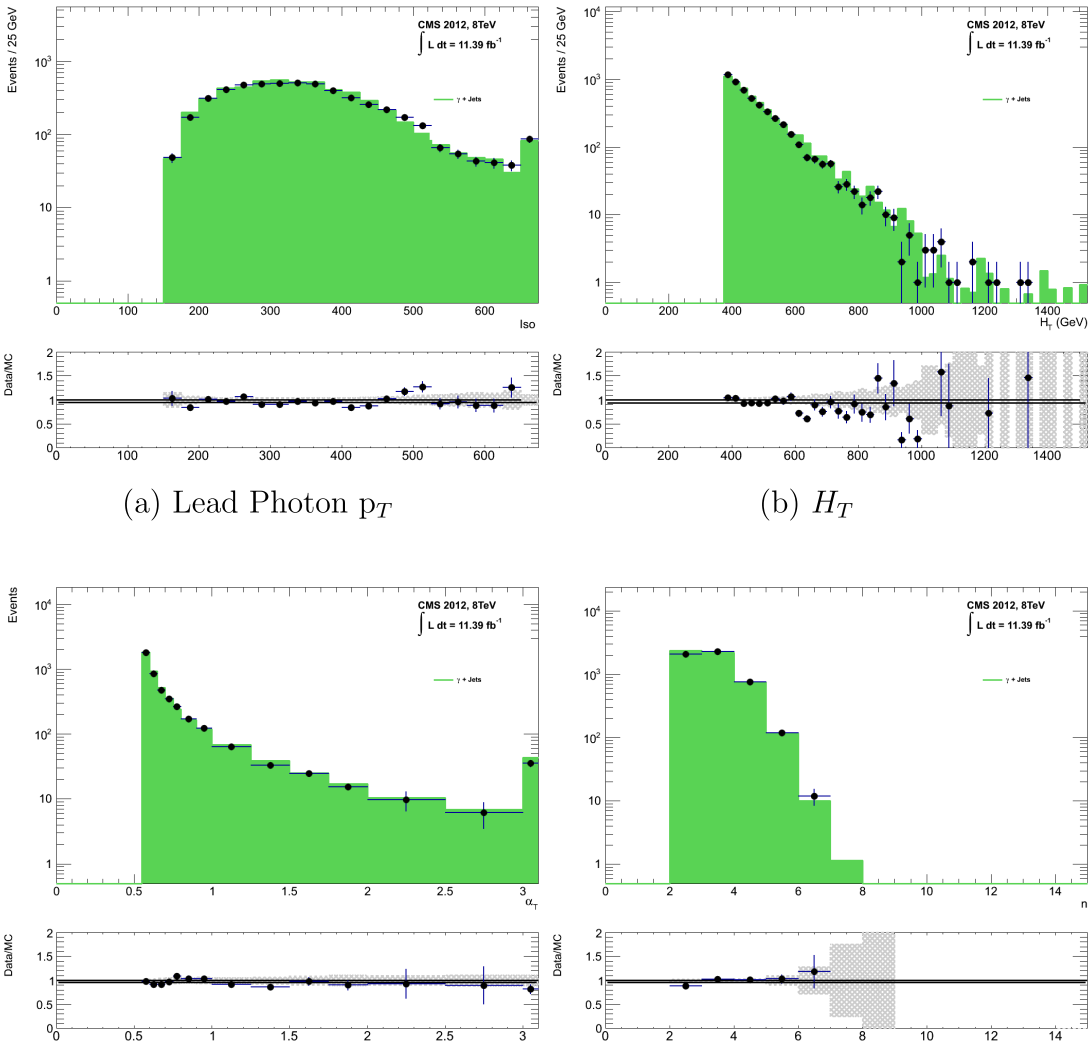
The Z $\rightarrow \nu\bar{\nu}$ + jets background is also estimated from a γ + jets control sample. When the E_T of the photon is greater than the mass of the Z, it possesses a larger cross-section and has kinematic properties similar to those of Z $\rightarrow \nu\bar{\nu}$ events if the photon is ignored [82].

Within the control channel, the photon is ignored for the purpose of the calculation of event level variables, and identical selection criteria to the hadronic signal region are applied. In addition the follow requirements are also made:

– Exactly one photon is selected, satisfying identification criteria as detailed in Table 4.3, with a minimum p_T >165 GeV to satisfy trigger thresholds and $|\eta|$ < 1.45 to ensure the photon remains in the barrel of the detector.

– A selection criteria of $\Delta R(\gamma, jet) < 1.0$, between the photon and all jets is applied to ensure the acceptance of only well isolated γ + jets events.

– Given that the photon is ignored, this control sample can only be applied in the H_T region > 375 GeV, due to the trigger thresholds on the minimum p_T of the photon, and the $\not{H}_T$ requirement of an $\alpha_T > 0.55$ cut from Equation (4.5). This is maintained in this control sample due to contamination from QCD processes in the absence of an α_T cut.



(a) Lead Photon p_T (b) H_T

(c) α_T (d) Jet multiplicity

Figure 4.8: Data/MC comparisons of key variables for the γ + jets selection, following the application of selection criteria and the requirements that $H_T > 375$ GeV and $\alpha_T > 0.55$. Bands represent the uncertainties due to the statistical size of the MC samples. No requirement is made upon the number of b-tagged jets or jet multiplicity in these distributions.

The selection criteria of the three control samples are defined to ensure background composition and event kinematics mirror closely the signal region. This is done in order to minimise the reliance on simulation to model correctly the backgrounds and event kinematics in the control and signal samples.

However in the case of the μ + jets and $\mu\mu$ + jets samples, the α_T requirement is relaxed in the selection criteria of these samples. This is made possible as contamination from QCD multi-jet events is suppressed to a negligible level by the other kinematic selection criteria within the two control samples, selecting pure EWK processes. Thus in this way, the acceptance of the two muon control samples can be significantly increased, which simultaneously improves their statistical and predictive power and also dilutes the effect of any potential signal contamination.

The modelling of the α_T variable is probed through a dedicated set of closure tests, described in Section (4.6), which demonstrate that the different α_T acceptances for the control and signal samples have no significant systematic bias on the background predictions.

4.2.4. Estimating the QCD Multi-jet Background

A negligible background from QCD multi-jet events within the hadronic signal region is expected due to a combination of selection requirements, and additional applied cleaning filters. However a conservative approach is still adopted and the likelihood model, see Section (4.8.2), is given the freedom to accommodate any potential QCD multi-jet contamination.

Any potential contamination can be identified through the variable R_{α_T}, defined as the ratio of events above and below the α_T threshold value used in the analysis. This is modelled by a H_T dependant falling exponential function which takes the form,

$$R_{\alpha_T}(H_T) = A_{\mathrm{QCD}} \exp^{-k_{QCD} H_T}, \tag{4.10}$$

where the parameters $\mathrm{A_{QCD}}$ and k_{QCD} are the normalisation and exponential decay constants, respectively.

For QCD multi-jet event topologies, this exponential behaviour as a function of H_T is expected for several reasons. The improvement of jet energy resolution at higher H_T due to higher p_T jets leads to a narrower peaked α_T distribution, causing R_{α_T} to fall. Similarly at higher H_T values $>$ 375 GeV, the jet multiplicity rises slowly with H_T. As shown in Figure 4.3, at higher jet multiplicities the result of the combinatorics used in the determination of α_T lead to more conservative α_T values, also resulting in a narrower distribution.

The value of the decay constant k_{QCD} is constrained via measurements within data sidebands to the signal region. This is also done to validate the falling exponential assumption for QCD multi-jet topologies. The sidebands are enriched in QCD multi-jet background and defined as regions where either α_T is relaxed or that the R_{miss} cut is inverted. Figure 4.9 depicts the definition of these data sidebands used to constrain the value of k_{QCD}.

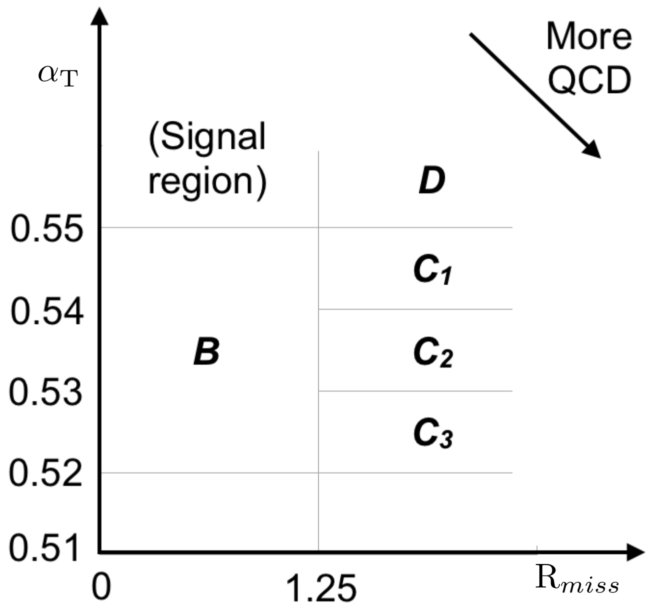


Figure 4.9: QCD sideband regions, used for determination of k_{QCD}.

The fit results used to determine the value of k_{QCD} are shown in Appendix (C.1), for which the best fit parameter value obtained from sideband region B is determined to be $k_{\mathrm{QCD}} = 2.96 \pm 0.64 \times 10^{-2}$ GeV $^{-1}$.

The best fit values of the remaining three C sideband regions are used to estimate the systematic uncertainty on the central value obtained from sideband region B. The variation of these measured values is used to determine the error on the determined central value, and is calculated to be $1.31 \pm 0.26 \times 10^{-2}$ GeV^{-1}. This relative error of $\sim$ 20% gives an estimate of the systematic uncertainty of the measurement to be applied to k_{QCD}.

Finally the same procedure is performed for sideband region D as an independent crosscheck, to establish that the value of k_{QCD} extracted from a lower α_T slice, can be applied to the signal region $\alpha_T > 0.55$. The likelihood fit is performed across all H_T bins within the QCD enriched region with no constraint applied to k_{QCD}. The resulting best fit value for k_{QCD} shows good agreement between that and the weighted mean, determined from the three C sideband regions. This demonstrates that the assumption of using the central value determined from sideband region B, to provide an unbiased estimator for k_{QCD} in the signal region ($\alpha_T > 0.55$) is valid.

Table 4.7 summarises the best fit k_{QCD} values determined for each of the sideband regions to the signal region.

Sideband region	$k_{\mathrm{QCD}}(\times 10^{-2} GeV^{-1})$	$p-$value
B	2.96 $\pm$ 0.64	0.24
C_1	1.19 $\pm$ 0.45	0.93
C_2	1.47 $\pm$ 0.37	0.42
C_3	1.17 $\pm$ 0.55	0.98
C(weighted mean)	1.31 $\pm$ 0.26	-
D(likelihood fit)	1.31 $\pm$ 0.09	0.57

Table 4.7: Best fit values for the parameters k_{QCD} obtained from sideband regions B,C_1,C_2,C_3. The weighted mean is determined from the three measurements made within sideband region C. The maximum likelihood value of k_{QCD} given by the simultaneous fit using sideband region D. Quotes errors are statistical only.

4.3. Trigger Strategy

A cross trigger based on the H_T and α_T values of an event, is used with varying thresholds across H_T bins to record the events used in the hadronic signal region. The α_T legs of the `HT_alphaT` triggers used in the analysis, are chosen to suppress QCD multi-jet events and control trigger rate, whilst maintaining signal acceptance. To maintain an acceptable rate for these analysis triggers, only calorimeter information is used in the reconstruction of the H_T sum, leading to the necessity for Calo jets to be used within the analysis.

A single object prescaled `HT` trigger is used to collect events for the hadronic control region, described above in Section (4.2.4).

The performance of the α_T and H_T triggers used to collect data for the signal and hadronic control region is measured with respect to a reference sample collected using the muon system. This allows measurement of both the Level 1 seed and higher level triggers simultaneously, as the reference sample is collected independently of any jet requirements.

The selection for the trigger efficiency measurement is identical to that described in Section (4.2.2), with the requirement of exactly one well identified muon with $p_T > 30$ GeV. This muon is then subsequently ignored.

The efficiencies measured for the `HT_alphaT` triggers in each individual H_T and α_T leg, is summarised in Table 4.8 for each H_T category of the analysis.

H_T range (GeV)	ϵ on H_T leg (%)	ϵ on α_T leg (%)
275-325	$87.7^{+1.9}_{-1.9}$	$82.8^{+1.0}_{-1.1}$
325-375	$90.6^{+2.9}_{-2.9}$	$95.9^{+0.7}_{-0.9}$
375-475	$95.7^{+0.1}_{-0.1}$	$98.5^{+0.5}_{-0.9}$
475-∞	$100.0^{+0.0}_{-0.0}$	$100.0^{+0.0}_{-4.8}$

Table 4.8: Measured efficiencies of the H_T and α_T legs of the `HT` and `HT_alphaT` triggers in independent analysis bins. The product of the two legs gives the total efficiency of the trigger in a given offline H_T bin.

Data for the control samples of the analysis, detailed in Section (4.2.3), are collected using a single object photon trigger for the γ + jets sample, and a single object muon trigger for both the μ + jets and $\mu\mu$ + jets control samples.

The photon trigger is measured to be fully efficient for the threshold ${p_T}^{\text{photon}} > 150$ GeV, whilst the single muon efficiency satisfying ${p_T}^{\text{muon}} > 30$ GeV is measured to have an efficiency of $(88 \pm 2)\%$ that is independent of H_T. In the case of the $\mu\mu$ + jets control sample, the efficiency is measured to be $(95 \pm 2)\%$ for the lowest H_T bin, rising (due to the average p_T of the second muon in the event increasing at larger H_T) to $(98 \pm 2)\%$ in the highest H_T category.

4.4. Measuring Standard Model Process Normalisation Factors via H_T Sidebands

The theoretical cross-sections of different SM processes at Next to Next Leading Order (NNLO) and the number of available simulated events generated for a particular process, is typically used to determine the appropriate normalisation for a simulation sample. However within the particular high-H_T and high-$\not{E}_T$ corners of kinematic phase space probed within this search, the theoretical cross sections for various processes are far less well understood.

To mitigate the problem of theoretical uncertainties and arbitrary choices of cross sections, the normalisation of the simulation samples are determined through the use of data sidebands. The sidebands are used to calculate sample specific correction factors (k-factors), that are appropriate for the H_T-$\not{E}_T$ phase space covered by this analysis.

They are defined within the μ + jets and $\mu\mu$ + jets control sample, by the region 200< H_T <275, using the same jet p_T thresholds as the adjacent first analysis bin. Individual EWK processes are isolated within each of these control samples via requirements on jet multiplicity and the requirement on b-tag multiplicity, summarised in Table 4.9. The purity of the samples are typically > 90% with any residual contamination subtracted prior to determination of the correction factors. The resultant k-factor for each process is determined by then taking ratio of the data yield over the expectation from simulation in the sideband. Subsequently these k-factors are then applied to the processes within the phase space of the analysis.

Process	Selection	Observation	MC expectation	k-factor
W + jets	μ + jets, n_b=0, n_{jet} = 2,3	26950	29993.2 $\pm$ 650.1	0.90 $\pm$ 0.02
$Z \to \mu\mu$ + jets	$\mu\mu$ + jets, n_b=0, n_{jet} = 2,3	3141	3402.0 $\pm$ 43.9	0.92 $\pm$ 0.02
$t\bar{t}$	μ + jets, n_b=2, n_{jet} = $\geq$4	2190	1967.8 $\pm$ 25.1	1.11 $\pm$ 0.02

Table 4.9: k-factors calculated for different EWK processes. All k-factors are derived relative to theoretical cross-sections calculated in NNLO. The k-factors measured for the Z$\to$ $\mu\mu$ + jets processes, are also applied to the Z $\to \nu\bar{\nu}$ + jets and γ + jets simulation samples.

It is worth pointing out that these correction factors have a negligible effect when providing a background estimation for the signal region. The TFs used in the analysis are found to be unaffected by application of these k-factors due to the similarity in the background composition of the control and signal regions. However when systematic uncertainties are determined in Section (4.6), the closure tests performed are sensitive to these corrections when extrapolations between different n_b^{reco} and $n_{\rm jet}$ categories are performed.

4.5. Determining Monte Carlo Simulation Yields with Higher Statistical Precision

Reconstructing events from EWK processes with many b-tagged jets (≥3), $n_{\mathrm{b}}^{\mathrm{reco}}$, is largely driven by the mis-tagging of light jets within the event. This is clear when considering the main EWK backgrounds in the analysis, such as $t\bar{t}$ + jets events, which typically contain two underlying b-quarks in the final state from the decay of the top quarks. Conversely W + jets and Z→ $\mu\mu$ + jets events will typically contain no underlying b-quarks in its final state.

When the expectation for the number of $n_{\mathrm{b}}^{\mathrm{reco}}$ jets is taken directly from simulation, the statistical uncertainty at large reconstructed b-tagged jet multiplicities becomes relatively large. One approach to reduce this uncertainty is to use the information encoded throughout all events in the simulation sample, to measure each of the following four ingredients:

1. the averaged b-tagging efficiency in the event selection,
2. the averaged charm-tagging efficiency in the event selection,
3. the averaged mis-tagging efficiency in the event selection,
4. the underlying flavour distribution of the jets in the event sample.

Together they can be used to determine the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution of the process being measured. This method allows the determination of higher b-tagged jet multiplicities to a higher degree of accuracy, reducing the statistical uncertainties of the simulation yields which enter into the TF's. For the discussion that follows, this approach will be known as the formula method.

4.5.1. The Formula Method

The assigning of jet flavours to reconstruction level jets in simulation is achieved via an algorithmic method defined as:

- attempt to find the parton that most likely determines the properties of the jet and assign that flavour as the true flavour,

- "final state" partons (after showering, radiation) are analysed (also within $\Delta R < 0.3$ of reconstructed jet cone),
- if there is a b/c flavoured parton within jet cone: label jet as a b/c flavoured jet,
- otherwise: assign flavour of the hardest (highest p_T) parton within the jet cone.

This process is employed within each individual simulation sample and independently for each H_T- $n_{\rm jet}$ category in the analysis.

Let $N(n_b^{gen}, n_c^{gen}, n_q^{gen})$ represent the 3-dimensional underlying jet flavour distribution in simulation, with b underlying b-quarks, c underlying c-quarks and q underlying light quarks which are matched to reconstructed jets as detailed above. Light quarks defined as those which originate from a u, d, s, g and τ jets, which having similar mis-tagging rates are grouped together.

The $n_{\rm b}^{\rm reco}$ distribution within each H_T- $n_{\rm jet}$ category of the analysis can be constructed for each process in turn in an analytical way using the formula:

$$\begin{aligned} N(n) = & \sum_{n_b^{gen}+n_c^{gen}+n_q^{gen}=n_{\rm jet}^{cat}} \; \sum_{n_b^{tag}+n_c^{tag}+n_q^{tag}=n} N(n_b^{gen}, n_c^{gen}, n_q^{gen}) \times P(n_b^{tag}, n_b^{gen}, \epsilon) \times \\ & P(n_c^{tag}, n_c^{gen}, \beta) \times P(n_q^{tag}, n_q^{gen}, m), \end{aligned} \tag{4.11}$$

with N(n) representing the yield of n b-tagged jet events of a simulated process as calculated by the formula method.

The variables $P(n_b^{tag}, n_b^{gen}, \epsilon)$, $P(n_c^{tag}, n_c^{gen}, \beta)$ and $P(n_q^{tag}, n_q^{gen}, m)$ correspond to the binomial probabilities for the tagging of a jet flavour to occur, based on its measured tagging efficiency (ϵ, β, m). These efficiencies are measured individually for each analysis category from simulation, using all simulated process events passing selection criteria. Thus the tagging efficiencies used within the above formula, represent the averaged tagging efficiency of each jet flavour within the phase space of the analysis category.

Finally, the constraints $n_{b/c/q}^{tag}$ signify the number of tagged jets of a particular jet flavour, of which the sum of the three terms must equal the number of n tagged jets being calculated. Similarly each $n_{b/c/q}^{gen}$ term represents the identified jet flavour of each jet in the event, and is required by definition for the sum of the three terms to fall within the $n_{\rm jet}$ category being analysed.

This approach ultimately results in a more precise $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution prediction, due to the utilisation of all events in the simulation sample which pass selection in extracting the overall $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution.

4.5.2. Establishing Proof of Principle

In order to validate the procedure, the predictions determined from the formula method summarised in Equation (4.11), are compared directly with those obtained directly from simulation. Resultantly no simulation to data correction factors are applied.

This sanity check for the μ + jets control sample is presented in Table 4.10, for all n_b^{reco} and H_T categories with no requirement placed upon the jet multiplicity of the events.

H_{T} Bin (GeV)	275–325	325–375	375–475	475–575
Formula $n_b = 0$	12632.66 ± 195.48	6696.08 ± 82.59	6368.96 ± 75.34	2906.27 ± 39.65
Vanilla $n_b = 0$	12612.95 ± 198.68	6687.97 ± 83.78	6359.27 ± 76.50	2898.27 ± 36.89
Formula $n_b = 1$	4068.09 ± 45.71	2272.76 ± 26.14	2181.32 ± 25.07	1089.14 ± 13.82
Vanilla $n_b = 1$	4067.73 ± 60.30	2268.02 ± 30.20	2180.69 ± 28.73	1094.37 ± 24.14
Formula $n_b = 2$	1963.71 ± 22.44	1087.55 ± 13.57	1055.57 ± 13.25	554.96 ± 7.95
Vanilla $n_b = 2$	1984.53 ± 26.19	1094.43 ± 16.67	1068.96 ± 16.36	558.14 ± 10.51
Formula $n_b = 3$	146.94 ± 2.07	79.97 ± 1.37	78.05 ± 1.35	49.84 ± 1.03
Vanilla $n_b = 3$	149.52 ± 4.84	85.98 ± 3.64	74.45 ± 3.29	49.54 ± 2.68
Formula $n_b \geq 4$	2.26 ± 0.12	1.29 ± 0.10	5.32 ± 0.20	-
Vanilla $n_b \geq 4$	1.84 ± 0.50	1.02 ± 0.39	4.86 ± 0.83	-

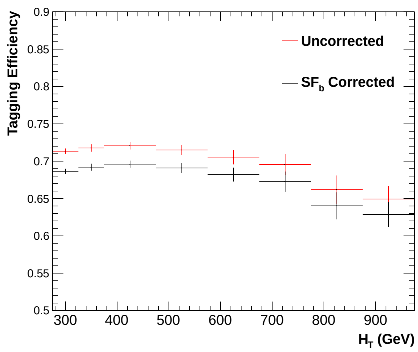
H_{T} Bin (GeV)	575–675	675–775	775–875	>875
Formula $n_b = 0$	1315.68 ± 19.49	640.49 ± 11.90	327.81 ± 7.91	424.27 ± 9.27
Vanilla $n_b = 0$	1315.23 ± 20.20	641.96 ± 12.48	329.09 ± 8.36	424.02 ± 9.73
Formula $n_b = 1$	490.41 ± 7.45	226.95 ± 4.42	109.91 ± 2.84	129.97 ± 3.07
Vanilla $n_b = 1$	490.52 ± 9.92	222.22 ± 6.21	107.46 ± 4.15	129.64 ± 4.64
Formula $n_b = 2$	256.75 ± 4.58	113.45 ± 2.70	52.10 ± 1.69	59.29 ± 1.78
Vanilla $n_b = 2$	253.43 ± 6.52	117.17 ± 4.27	52.70 ± 2.80	59.45 ± 3.00
Formula $n_b = 3$	25.66 ± 0.69	12.48 ± 0.46	5.52 ± 0.31	6.83 ± 0.33
Vanilla $n_b = 3$	29.18 ± 2.06	11.77 ± 1.26	6.18 ± 0.95	7.53 ± 1.05

Table 4.10: Comparing yields in simulation within the μ + jets selection determined from the formula method described in Equation (4.11), and that taken directly from simulation. The numbers are normalised to 11.4fb^{-1}. No simulation to data corrections are applied.

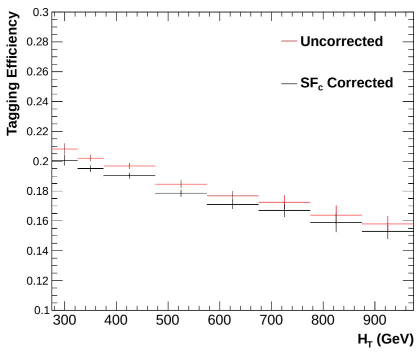
It can be seen as expected, that there is good consistency between the results determined via the formula method and 'raw' simulation yields. Similarly the power of this approach can be seen in the reduction of this statistical error in the prediction across all H_T and n_b^{reco} categories. In particular the statistical uncertainty is reduced by several factors in the highest $n_b^{reco} \geq 4$ category.

4.5.3. Correcting Measured Efficiencies in Simulation to Data

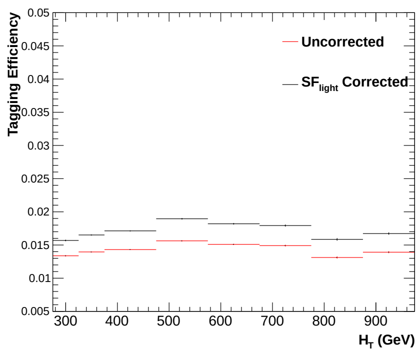
As detailed in Section (3.3.2), it is necessary for certain p$_T$ and η dependant corrections, to be applied to both the b-tagging efficiency and mis-tagging rates in order to correct the efficiencies from simulation to the efficiencies measured in data. These correction factors are considered when determining the simulation yields for each selection, which are used to construct the TFs of the analysis. The magnitude of this correction are measured individually for each H_T category and are shown in Figure 4.10.



(a) b-jets



(b) c-jets



(c) light-jets

Figure 4.10: Tagging efficiencies of (a) b-jets, (b) c-jets, and (c) light-jets as a function all jets within each H_T category. Efficiencies measured directly from simulation (black) and with data to simulation $SF_{b,c,light}$ correction factors (red) are applied.

Each of the correction factors for the b, c and light flavoured jets come with an associated systematic uncertainty. The uncertainties across different jet p_T and η categories, are considered as fully correlated. When computing the magnitude of the effect of this systematic uncertainty on the TFs of the analysis, the measured tagging efficiencies for each jet flavour are scaled up/down simultaneously within each H_T and n_{jet} category by the systematic uncertainty of the $SF_{\mathrm{b,\ c,\ light}}$ scale factors.

Varying the scale factor corrections by their systematic uncertainty will change the absolute yields within each n_b^{reco} bin of all selections. However, ultimately it is the change in the TFs which influences the final background prediction from each of the control samples. The magnitude of the absolute change in each TF, constructed from when the μ + jets control sample is used to predict the entire hadronic signal region background, is shown in Table 4.11.

n_b^{reco}	275–325	325–375	375–475	475–575
$= 0$	$0.557\ ^{+0.001}_{-0.001} \pm 0.012$	$0.495\ ^{+0.001}_{-0.001} \pm 0.009$	$0.383\ ^{+0.001}_{-0.001} \pm 0.005$	$0.307\ ^{+0.001}_{-0.002} \pm 0.006$
$= 1$	$0.374\ ^{+0.006}_{-0.006} \pm 0.006$	$0.320\ ^{+0.006}_{-0.005} \pm 0.005$	$0.251\ ^{+0.005}_{-0.005} \pm 0.004$	$0.185\ ^{+0.003}_{-0.003} \pm 0.004$
$= 2$	$0.226\ ^{+0.002}_{-0.002} \pm 0.004$	$0.201\ ^{+0.001}_{-0.002} \pm 0.004$	$0.159\ ^{+0.001}_{-0.001} \pm 0.004$	$0.134\ ^{+0.000}_{-0.001} \pm 0.004$
$= 3$	$0.221\ ^{+0.002}_{-0.002} \pm 0.005$	$0.208\ ^{+0.002}_{-0.001} \pm 0.007$	$0.164\ ^{+0.001}_{-0.000} \pm 0.006$	$0.144\ ^{+0.001}_{-0.001} \pm 0.007$
≥ 4	$0.222\ ^{+0.004}_{-0.005} \pm 0.015$	$0.248\ ^{+0.003}_{-0.003} \pm 0.035$	$0.123\ ^{+0.002}_{-0.003} \pm 0.009$	-
	575–675	675–775	775–875	≥875
$= 0$	$0.263\ ^{+0.001}_{-0.002} \pm 0.006$	$0.215\ ^{+0.000}_{-0.001} \pm 0.007$	$0.171\ ^{+0.000}_{-0.001} \pm 0.009$	$0.111\ ^{+0.000}_{-0.001} \pm 0.006$
$= 1$	$0.154\ ^{+0.003}_{-0.003} \pm 0.005$	$0.138\ ^{+0.003}_{-0.004} \pm 0.006$	$0.121\ ^{+0.005}_{-0.005} \pm 0.007$	$0.091\ ^{+0.002}_{-0.002} \pm 0.006$
$= 2$	$0.104\ ^{+0.000}_{-0.001} \pm 0.005$	$0.079\ ^{+0.001}_{-0.001} \pm 0.006$	$0.063\ ^{+0.001}_{-0.002} \pm 0.007$	$0.071\ ^{+0.000}_{-0.000} \pm 0.008$
$= 3$	$0.116\ ^{+0.001}_{-0.001} \pm 0.009$	$0.069\ ^{+0.001}_{-0.001} \pm 0.007$	$0.079\ ^{+0.001}_{-0.001} \pm 0.017$	$0.095\ ^{+0.003}_{-0.002} \pm 0.020$

Table 4.11: The absolute change in the TFs used to predict the entire signal region SM background, using the μ + jets control sample when the systematic uncertainties of the data to simulation scale factors are varied by $\pm 1\sigma$. The impact of the change is shown for each H_T and n_b^{reco} category with no requirement made on the jet multiplicity of the events. (Also quoted are the statistical uncertainties)

It can be seen that the TFs are found to be relatively insensitive to the systematic uncertainty of the b-tag scale factors (showing typically less than $\sim 2\%$ change). This can be accounted for by the similar composition of the signal and control sample backgrounds, such that any change in the underlying n_b^{reco} distribution will be reflected in both signal and control regions and cancel out in the TF.

Any overall systematic effect on the overall background prediction of the analysis from these b-tag scale factor uncertainties is incorporated within the data driven systematics introduced in the following section.

4.6. Systematic Uncertainties on Transfer Factors

Since the TFs used to establish the background prediction are obtained from simulation, an appropriate systematic uncertainty is assigned to account for theoretical uncertainties [83] and limitations in the simulation modelling of event kinematics and instrumental effects.

The magnitudes of these systematic uncertainties are established through a data driven method, in which the three independent control samples of the analysis (μ + jets, $\mu\mu$ + jets, γ + jets) are used to in a series of closure tests. The yields from one of these control samples, along with the corresponding TF obtained from simulation, are used to predict the expected yields in another control sample. This procedure therefore utilises the same method used in determining a background prediction for the signal region as already established in Section (4.2.3).

The level of agreement between the predicted and observed yields is expressed as the ratio

$$\frac{(N_{\text{obs}} - N_{\text{pred}})}{N_{\text{pred}}}, \tag{4.12}$$

while considering only the statistical uncertainties on the prediction, N_{pred}, and the observation, N_{obs}. No systematic uncertainty is assigned to the prediction, and resultantly the level of closure is defined by the statistical significance of a deviation from the ratio from zero.

This ratio is measured for each H_T category in the analysis, allowing these closure tests to be sensitive to both the presence of any significant biases or any possible H_T dependence to the level of closure.

Eight sets of closure tests are defined between the three data control samples, conducted independently between the two jet multiplicity ($2 \leq n_{\text{jet}} \leq 3$, $n_{\text{jet}} \geq 4$) categories. Each of these tests are specifically chosen to probe each of the different key ingredients of the simulation modelling that can affect the background prediction.

Each of the different modelling components and the relevant closure tests are described below:

α_T modelling

The modelling of the α_T distribution in genuine $\not{E}_T$ events is probed with the μ + jets control sample. This test is important to verify the approach of removing the $\alpha_T > 0.55$ requirement from the μ + jets and $\mu\mu$ + jets samples to increase the precision of the background prediction. The test uses the μ + jets sample without an α_T cut to make a prediction into the μ + jets sample defined with the requirement $\alpha_T > 0.55$.

Background admixture

The sensitivity of the translation factors to the relative admixture of events from $W+$ jets and $t\bar{t}$ processes is probed by two closure tests.

Within the μ + jets sample, a W boson enriched sub-sample ($n_b = 0$) is used to predict yields in a $t\bar{t}$ enriched sub-sample ($n_b = 1$). Similarly, the $t\bar{t}$ enriched sub-sample (n_b =1) is also used to predict yields for a further enriched $t\bar{t}$ sub-sample ($n_b = 2$), further probing the modelling of the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution.

Similarly a further closure test probes the relative contribution of $Z+$ jets to W+jets and $t\bar{t}$ events, through the use of the μ + jets sample to predict yields for the $\mu\mu$ + jets control sample. This closure test, also at some level probes the muon trigger and reconstruction efficiencies, given that exactly one or two muons are required by the different selections.

These tests represent an extremely conservative approach as the admixture of the two backgrounds remains similar when a prediction is made between the control samples and the signal region. This is contrary to the closure tests defined above which make predictions between two very different admixtures of $W+$ jets and $t\bar{t}$ events.

Consistency check between Z $\rightarrow \nu\bar{\nu}$ predictions

This is an important consistency check between the $\mu\mu$ + jets and γ + jets, which are both used in the prediction of the Z $\rightarrow \nu\bar{\nu}$ in the signal region. This is conducted by using the γ + jets sample to predict yields for the $\mu\mu$ + jets control sample. Using γ + jets processes as a method to predict Z + jet processes is subject to theory uncertainties [84], which can be probed by this data driven closure test within a $Z \to \mu\mu$ control sample.

Modelling of jet multiplicity

The simulation modelling of the jet multiplicity within each control sample is important due to the exclusive jet multiplicity categorisation within the analysis. This is probed via the use of each of the three control samples to independently predict from the lower jet multiplicity category $2 \leq n_{\mathrm{jet}} \leq 3$, to the high jet category $n_{\mathrm{jet}} \geq 4$.

For the case of the μ + jets and $\mu\mu$ + jets control samples, this test also serves as a further probe of the admixture between W+ jets/Z+ jets and $t\bar{t}$.

To test for the assumption that no H_T dependencies exist within the background predictions of the analysis, the first five closure tests defined above are used, with zeroeth and first order polynomial fits applied to each test individually. This is summarised in Table 4.12 and Table 4.13 which show the results for both the $2 \leq n_{jet} \leq 3$ and ≥ 4 jet multiplicity bins respectively.

		Constant fit		Linear fit	
Closure test	Symbol	Best fit value	p-value	Slope (10^{-4})	p-value
$\alpha_T < 0.55 \to \alpha_T > 0.55$ (μ + jets)	Circle	-0.06 ± 0.02	0.93	-1.3 ± 2.2	0.91
0 b-jets $\to$ 1 b-jet (μ + jets)	Square	0.07 ± 0.02	0.98	-1.6 ± 1.6	1.00
1 b-jets $\to$ 2 b-jet (μ + jets)	Triangle	-0.07 ± 0.03	0.76	-2.7 ± 3.0	0.76
μ + jets $\to$ $\mu\mu$ + jets	Cross	0.10 ± 0.03	0.58	-1.1 ± 2.3	0.49
$\mu\mu$ + jets $\to$ γ + jets	Star	-0.06 ± 0.04	0.31	4.2 ± 4.3	0.29

Table 4.12: A summary of the results obtained from zeroeth order polynomial (i.e. a constant) and linear fits to five sets of closure tests performed in the $2 \geq n_{\mathrm{jet}} \geq 3$ category. The two columns show the best fit value for the slope obtained when performing a constant (left) and linear (right) fit and the p-value for that fit.

Table 4.14 shows the same fits applied to the three closure tests that probe the modelling between the two n_{jet} categories. The best fit value and its uncertainty is listed for each set of closure tests in all three tables, along with the p-value of the constant and linear fits applied.

The best fit value for the constant parameter is indicative of the level of closure, averaged across the full H_T range of the analysis, and the p-value an indicator of any significant dependence on H_T within the closure tests. The best fit values of all the tests are either

		Constant fit		Linear fit	
Closure test	Symbol	Best fit value	p-value	Slope (10^{-4})	p-value
$\alpha_T < 0.55 \rightarrow \alpha_T > 0.55$ (μ + jets)	Circle	-0.05 ± 0.03	0.21	3.0 ± 2.9	0.21
0 b-jets $\rightarrow$ 1 b-jet (μ + jets)	Square	-0.03 ± 0.03	0.55	-1.0 ± 1.9	0.47
1 b-jets $\rightarrow$ 2 b-jet (μ + jets)	Triangle	-0.02 ± 0.03	0.39	1.1 ± 2.2	0.31
μ + jets $\rightarrow \mu\mu$ + jets	Cross	0.08 ± 0.07	0.08	4.8 ± 4.3	0.07
$\mu\mu$ + jets $\rightarrow \gamma$ + jets	Star	-0.03 ± 0.10	0.72	-4.0 ± 7.0	0.64

Table 4.13: A summary of the results obtained from zeroeth order polynomial (i.e. a constant) and linear fits to five sets of closure tests performed in the $n_{\text{jet}} \geq 4$ category. The two columns show the best fit value for the slope obtained when performing a constant (left) and linear (right) fit and the p-value for that fit.

		Constant fit		Linear fit	
Closure test	Symbol	Best fit value	p-value	Slope (10^{-4})	p-value
μ + jets	Inverted triangle	-0.03 ± 0.02	0.02	0.0 ± 1.0	0.01
μ + jets (outlier removed)	Inverted triangle	-0.04 ± 0.01	0.42	-1.4 ± 1.1	0.49
γ + jets	Diamond	0.12 ± 0.05	0.79	6.0 ± 4.7	0.94
$\mu\mu$ + jets	Asterisk	-0.04 ± 0.07	0.20	4.9 ± 4.4	0.20

Table 4.14: A summary of the results obtained from zeroeth order polynomial (i.e. a constant) and linear fits to three sets of closure tests performed between the $2 \leq n_{\text{jet}} \leq 3$ and $n_{\text{jet}} \geq 4$ categories. The two columns show the best fit value for the slope obtained when performing a constant (left) and linear (right) fit and the p-value for that fit.

statistically compatible with zero bias (i.e. less than 2σ from zero) or at the level of 10% or less, with the exception of one closure test discussed below.

Within Table 4.14, there exists one test that does not satisfy the above statement, which is the $2 \leq n_{jet} \leq 3 \rightarrow n_{jet} \geq 4$ test using the μ + jets control sample. The low p-value can be largely attributed to an outlier between $675 < H_T < 775$ GeV, rather than any significant trend in H_T. Removing this single outlier from the constant fit performed, gives a best fit value of -0.04 ± 0.01, $\chi^2/$ d.o.f $= 6.07/6$. and a p-value of 0.42. These modified fit results are also included in Table 4.14.

Additionally, it is found that the best fit values for the slope terms of the linear fits in all three tables are of the order 10^{-4}, which corresponds to a percent level change per 100 GeV. However in all cases, the best fit values are fully compatible with zero (within 1σ) once again with the exception detailed above, indicating that the level of closure is indeed H_T independent.

4.6.1. Determining Systematic Uncertainties from Closure Tests

Once it has been established that no significant bias or trend exists within the closure tests, systematic uncertainties are determined. The statistical precision of the closure tests is considered a suitable benchmark for determining the systematic uncertainties that are assigned to the TFs, which are propagated through to the likelihood fit.

The systematic uncertainty band is split into five separate regions of H_T. Within each region the square root of the sample variance, σ^2, is taken over the eight closure tests to determine the systematic uncertainties to be applied within that region.

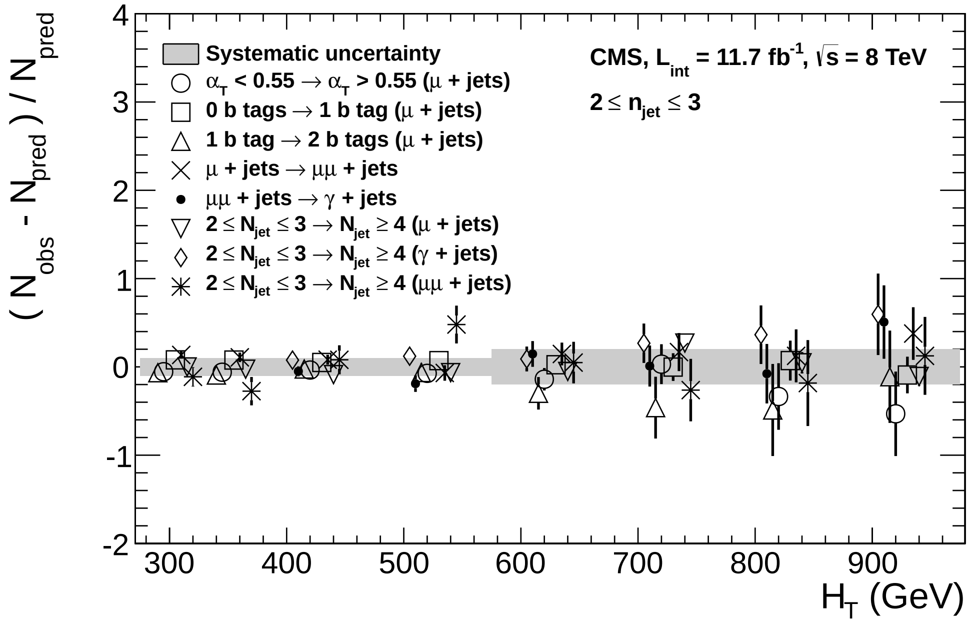
Using this procedure the systematic uncertainties for each region are calculated and are shown in Table 4.15, with the systematic uncertainty to be used in the likelihood model conservatively rounded up to the nearest decile and applied across all $n_{\mathrm{b}}^{\mathrm{reco}}$ categories.

H_T band (GeV)	$2 \leq n_{jet} \leq 3$	$n_{jet} \geq 4$
$275 < H_T < 325$	10%	10%
$325 < H_T < 375$	10%	10%
$375 < H_T < 575$	10%	10%
$575 < H_T < 775$	20%	20%
$H_T > 775$	20%	30%

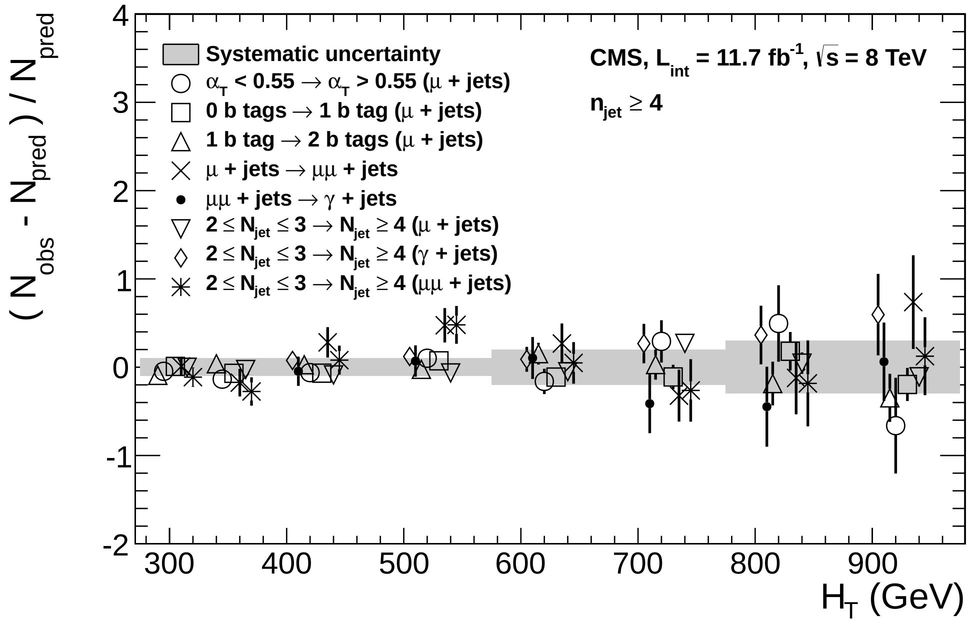
Table 4.15: Calculated systematic uncertainties for the five H_T regions, determined from the closure tests. Uncertainties shown for both jet multiplicity categories. Values used within the likelihood model are conservatively rounded up to the nearest decile.

Figure 4.11 shows the sets of closure tests overlaid on top of grey bands that represent the H_T dependent systematic uncertainties. These systematic uncertainties are assumed to be fully uncorrelated between the different n_b multiplicity categories and across the five H_T regions. This can be considered a more conservative approach given that some correlations between adjacent H_T categories could be expected due to comparable kinematics.

These closure tests represent a conservative estimate of the systematic uncertainty in making a background prediction for the signal region. This is due to significant differences in the background composition and event kinematics between the two sub-samples used in the closure tests. This is not the case when a signal region prediction is made, due to the two sub-samples both having a comparable background admixture and similar kinematics owing to the fact that the TFs are always constructed using the same (n_{jet}, $n_{\mathrm{b}}^{\mathrm{reco}}$, H_T) category.



(a)



(b)

Figure 4.11: Sets of closure tests (open symbols) overlaid on top of the systematic uncertainty used for each of the five H_T regions (shaded bands) and for the two different jet multiplicity categories: (a) $2 \leq n_{\mathrm{jet}} \leq 3$ and (b) $n_{\mathrm{jet}} \geq 4$.

This point is emphasised when we examine the sensitivity of the TFs to a change in the admixture of W + jets and $t\bar{t}$ with the control and signal samples. This is accomplished by varying the cross-sections of the W +jets and $t\bar{t}$ by +20% and -20%, respectively.

Figures C.2 and C.3 within Appendix C, show the effect upon the closure tests for both jet multiplicity categories. Given these variations in cross-sections, the level of closure is found to be significantly worse, with biases as large as $\sim$ 30%, most apparent in the lowest H_T bins. However, the TFs used to extrapolate from control to signal are seen to change only at the percent level by this large change in cross-section, shown in Table C.1.

Given the robust behaviour of the translation factors with respect to large (and opposite) variations in the W + jets and $t\bar{t}$ cross-sections, one can assume with confidence that any bias in the translation factors is adequately (and conservatively) covered by the systematic uncertainties used in the analysis.

4.7. Simplified Models, Efficiencies and Systematic Uncertainties

The results of the analysis are interpreted using various SMS signal models, which as already introduced in Section (2.4.1) offer a natural starting point for quantifying and characterising SUSY signals, and a means to identify the boundaries of search sensitivity for different mass splittings, kinematic ranges, and final states.

Each model is parameterised in a two dimensional parameter space, $(m_{\widetilde{q}/\widetilde{g}}, m_{\mathrm{LSP}})$, from which upper limits on the production cross-sections of the various SMS models can be set.

Each signal sample is generated at Leading Order (LO) with Pythia [85], and cross-sections calculated for Next to Leading Order (NLO) and Next to Leading Logorithmic Order (NLL) [86], with events simulated using the `Fastsim` framework. This framework represents a simplified simulation of the CMS detector, but allows for faster production of various signal topologies with different mass parameters.

A series of correction factors are applied to account for differences between `Fastsim` [87] and `Fullsim` [88] simulation, which can affect the resultant $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution and which are detailed in Section (4.7.2).

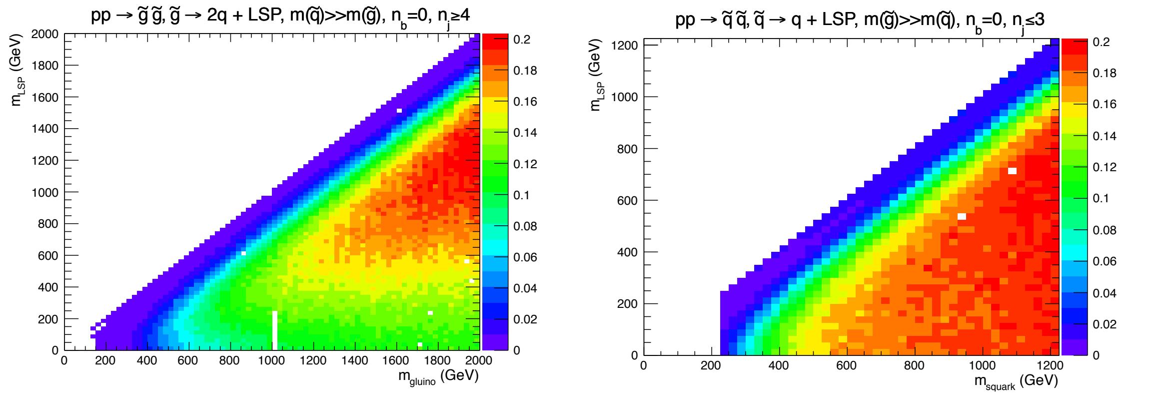
4.7.1. Signal Efficiency

The analysis selection efficiency, ϵ, is measured for each mass point of the interpreted model. This serves as a measure of the sensitivity of the signal selection for that particular sparticle, LSP mass and final state topology. The signal yield is then given by

$$Y(m_{\widetilde{q}/\widetilde{g}}, m_{LSP}) = \epsilon \times \sigma \times \mathcal{L}, \tag{4.13}$$

where σ represents the model's cross-section and $\mathcal{L}$ the luminosity. An upper limit on σ taken from theory can then allow for the setting of limits in terms of the particle mass.

Figure 4.12 shows the expected signal efficiency of the signal selection for the `T1` and `T2` SMS models interpreted in this analysis. The efficiency maps are produced with the requirement $H_T > 275$ GeV (i.e. no H_T categorisation) and requirements on n_{jet} and n_b^{reco} are the most sensitive to the model in question.



(a) Model `T1`, $n_{jet} \geq 4$, $n_b^{reco} = 0$ (b) Model `T2`, $n_{jet} \leq 3$, $n_b^{reco} = 0$

Figure 4.12: Signal efficiencies for the SMS models (a) `T1` ($\widetilde{g}\widetilde{g}^* \rightarrow q\bar{q}\widetilde{\chi}_1^0 q\bar{q}\widetilde{\chi}_1^0$) and (b) `T2` ($\widetilde{q}\widetilde{q}^* \rightarrow q\widetilde{\chi}_1^0 \bar{q}\widetilde{\chi}_1^0$) when requiring $n_{jet} \geq 4$ and ≤ 3 respectively, and $n_b^{reco} = 0$.

The same procedure is conducted in the analysis control samples. It is found in the μ + jets control samples, that the signal-to-background ratios for the expected signal yields in each of the SMS models are many times smaller than in the hadronic signal region. The relative contamination for the $\mu\mu$ + jets sample is smaller still due to the requirement of a second muon. The relative contamination for the γ + jets sample is expected to be zero for the models under consideration. These small, relative levels of contamination are accounted for in the fitting procedure, as described in Section (4.8.4).

4.7.2. Applying B-tagging Scale Factor Corrections in Signal Samples

High-statistic `FastSim` signal simulation samples are unavailable for each signal point, which means that a different procedure to the formula method described in Section (4.5) is employed. Furthermore, the use of the `FastSim` framework in the reconstruction introduces an extra set of scale-factor corrections, to be applied simultaneously with those correcting `FullSim` to the data.

For these signal models, an event-by-event re-weighting procedure is applied. This applied weight depends on both the flavour content and the b-tagging status of the reconstruction level jets in the event.

The re-weighting procedure can be described by first considering a single jet within a signal event. The flavour of the jet is determined using the method described in Section (4.5.1).

Maps of the tagging efficiencies, parameterised as a function of jet p_T and η are produced from `FullSim` simulation samples for each of the b, c and light jet flavours. These efficiencies are calculated from simulation events which pass the hadronic signal selection. The p_T and η binning of each map is chosen to match the correction maps of `FullSim` to data defined in [77].

The actual tagging efficiency of the `FastSim` jet, $\epsilon_{\texttt{FastSim}}(p_T, \eta, f)$, differs from that measured in `FullSim`, $\epsilon_{MC}(p_T, \eta, f)$, as detailed above and is related via an additional correction factor,

$$\epsilon_{\texttt{FastSim}}(p_T, \eta, f) = \frac{\epsilon_{MC}(p_T, \eta, f)}{SF_{\texttt{Fast}\rightarrow\texttt{Full}}(p_T, \eta, f)}. \tag{4.14}$$

$SF_{\texttt{Fast}\rightarrow\texttt{Full}}(p_T, \eta, f)$ represents a set of p_T and η dependant corrections, that are specific for each SMS model. These corrections are calculated from the ratio of tagging rates between a `FullSim` $t\bar{t}$ sample, and a selection of mass points from each `FastSim` SMS model, again measured individually for b, c and light-flavoured jets.

The tagging efficiencies measured in data [77], $\epsilon_{Data}(p_T, \eta, f)$, can then be related to $\epsilon_{\texttt{FastSim}}(p_T, \eta, f)$ by the equation,

$$\begin{aligned}\epsilon_{Data}(p_T,\eta,f) &= \epsilon_{MC}(p_T,\eta,f) \times SF_{MC\to Data}(p_T,\eta,f)\\ &= \epsilon_{\texttt{FastSim}}(p_T,\eta,f) \times \underbrace{SF_{\texttt{Fast}\to\texttt{Full}}(p_T,\eta,f) \times SF_{MC\to Data}(p_T,\eta,f)}_{\mathrm{SF}_{\texttt{Fast}}\,\to\,Data}.\end{aligned} \tag{4.15}$$

For each jet, the weight of the event is re-weighted according to whether the jet fires the tagger. In the instance that the jet *is* tagged, the event weight will be modified by,

$$\text{weight} = SF_{\texttt{Fast}\to Data} \times \text{weight}, \tag{4.16}$$

and in the case that the jet does *not* fire the tagger,

$$\text{weight} = \frac{1-\epsilon_{Data}(p_T,\eta,f)}{1-\epsilon_{\texttt{FastSim}}(p_T,\eta,f)} \times \text{weight}. \tag{4.17}$$

This procedure is applied to all events that pass the selection criteria, thus correcting the `FastSim` $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution to data.

4.7.3. Experimental Uncertainties

The systematic uncertainty on the expected signal acceptance $\times$ analysis efficiency is determined independently for the each SMS model considered. These systematics stem from uncertainties on the parton distribution functions, the luminosity measurement, jet energy scale, b-tag scale factor measurements and the efficiencies of various selection criteria used in the signal selection, including the $\not{H}_T/\not{E}_T$, dead ECAL cleaning filter and lepton / photon event vetoes.

Rather than trying to estimate the level of systematic that is applicable point-by-point in a model space, general behaviours are considered; and instead constant systematics are estimated in two regions of the SMS models parameter space.

These two regions are defined as, near (small mass splittings) and far (large mass splittings) from the mass degenerate diagonal, where the far region is bounded by the condition

$$m_{\widetilde{q}/\widetilde{g}} - m_{LSP} > 350GeV \qquad\qquad m_{\widetilde{q}/\widetilde{g}} > 475GeV.$$

The total systematics in each region are evaluated in the following ways:

Jet energy scale: The relative change in the signal efficiency is gauged by varying the energy of all jets in an event up or down according to a p_T and η dependent jet energy scale uncertainty. Within the two systematic regions, the resulting systematic uncertainties for each SMS model are determined by taking the value of the 68$^{\text{th}}$ percentile for the distributions of the relative change in the signal efficiency.

Luminosity measurement: The uncertainty on the measurement of the luminosity collected propagates through to an uncertainty on the signal event yield when considering any new physics model, which is currently 4.4% [89].

Parton density function : Each signal sample is produced using the CTEQ6L1 parton density function. The effect on the signal acceptance when re-weighting to the central value of three different parton distribution functions, CT10, MSTW08 and NNPDF2.1 are examined [90]. It is found that the change of the signal efficiency in different SMS models, due to the alternate PDF sets are typically a few percent, and approaches 10% at higher squark/gluon and LSP masses.

$\not{H}_T/\not{E}_T$ cleaning filter: The ratio of the efficiencies of the cleaning filter are compared in simulation and data after application of the μ + jets control sample selection. No α_T requirement or further event cleaning filters are applied. The ratio of the efficiencies observed in data and simulation for a cut value of $\not{H}_T/\not{E}_T <$ 1.25 and the two jet multiplicity categories, $2 \leq n_{\text{jet}} \leq 3$ and $n_{\text{jet}} \geq 4$ are 1.028 ± 0.007 and 1.038 ± 0.015 respectively. These deviations are taken to represent the systematic uncertainty on the simulation modelling of this variable.

Dead ECAL cleaning filter: The ratio of the efficiencies observed in data and simulation for this filter in the two jet multiplicity categories, $2 \leq n_{\text{jet}} \leq 3$ and $n_{\text{jet}} \geq 4$, are 0.961 ± 0.008 and 0.961 ± 0.009, respectively. These deviations from unity are taken to represent the systematic uncertainties in the modelling in simulation of this filter.

Lepton and photon vetoes: The uncertainty on the efficiency of the lepton and photon vetoes is determined by considering truth information. The efficiency of the vetoes is measured after applying relevant object filters with identical logic, but based on truth instead of reconstructed objects. Where the efficiency is found to not be 100%, it is taken to represent the fraction of signal events that are incorrectly vetoed. This deviation is taken directly as the systematic uncertainty on the efficiency. The systematic uncertainty is only non-zero for models which contain third-generation quarks in the final state, where the uncertainties are at the order of 1% level.

B-tag scale factor uncertainties: The relative change in the signal efficiency is observed when relevant flavour, p_T and η dependant b-tag correction factors, are varied up or down by their systematic uncertainty. Within the two systematic regions, the resulting systematic uncertainties for each SMS model are determined by taking the value of the 68^{th} percentile for the distributions of the relative change in the signal efficiency, over all mass points.

Tables 4.16 and 4.17 summarise all the aforementioned systematic uncertainties on the signal efficiencies for each individual SMS model interpreted in the analysis. In the case of the `T1tttt` model, in which pair produced gluinos decay to $t\bar{t}$ pairs and the LSP, the near region of SMS space is not considered, and so no systematic uncertainties are included.

In both of the defined regions it is found that the systematic uncertainties are relatively flat justifying the approach taken. The systematic uncertainties applied to the region near to the diagonal fall in the range 13-15%; similarly, for the region far from the diagonal the determined uncertainties are in the range of 12-23%. These uncertainties are all propagated through to the limit calculation.

Model	Luminosity	p.d.f	JES	$\not{H}_T/\not{E}_T$	Dead ECAL	Lepton Vetoes	b-tagging	Total
`T1`	4.4	10.0	5.6	3.8	4.1	n/a	3.1	13.9
`T2`	4.4	10.0	4.1	2.8	4.1	n/a	2.4	12.9
`T2bb`	4.4	10.0	4.8	2.8	4.1	0.3	2.2	13.1
`T1tttt`	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
`T1bbbb`	4.4	10.0	7.3	3.8	4.1	0.5	2.7	14.5

Table 4.16: Estimates of systematic uncertainties on the signal efficiency (%) for various SMS models when considering points in the region near to the diagonal (i.e. small mass splitting and compressed spectra). The uncertainties are added in quadrature to obtain the total.

Model	Luminosity	p.d.f	JES	$\not{H}_T/\not{E}_T$	Dead ECAL	Lepton Vetoes	b-tagging	Total
`T1`	4.4	10.0	0.8	3.8	4.1	n/a	6.6	14.0
`T2`	4.4	10.0	1.1	2.8	4.1	n/a	5.8	13.4
`T2bb`	4.4	10.0	0.9	2.8	4.1	0.3	2.7	12.3
`T1tttt`	4.4	10.0	0.5	3.8	4.1	1.4	19.4	23.0
`T1bbbb`	4.4	10.0	1.5	3.8	4.1	0.4	10.1	16.0

Table 4.17: Estimates of systematic uncertainties on the signal efficiency (%) for various SMS models when considering points in the region far from the diagonal (i.e. large mass splitting). The uncertainties are added in quadrature to obtain the total.

4.8. Statistical Interpretation

For a given category of events satisfying requirements on both n_{jet} and n_b^{reco}, a likelihood model of the observations in multiple data samples is used to gauge agreement between the observed yields in the hadronic signal region, and the predicted yields obtained from the control samples. In addition to checking whether the predictions are compatible with a SM only hypothesis, the likelihood model is also used to test for the presence of a variety of signal models. The statistical framework outlined within this section is presented in greater detail within [91].

4.8.1. Hadronic Sample

Let N be the number of bins in H_T, with n^i the number of events observed satisfying all selection requirements in each H_T bin i. The likelihood of the observations can then be written:

$$L_{had} = \prod_i \mathrm{Pois}(n^i|b^i + s^i), \tag{4.18}$$

where b^i represents the expected SM background

$$b^i = EWK_i + QCD_i, \tag{4.19}$$

and s^i the expected number of signal events (see Section(4.8.4)) from the different SMS models interpreted. Pois refers to the Poisson distribution of these values and is defined as:

$$\mathrm{Pois}(\chi|\lambda) = \frac{\lambda^{\chi} \exp^{-\lambda}}{k!}. \tag{4.20}$$

4.8.2. H_T Evolution Model

The hypothesis, that for a process the α_T ratio falls exponentially (see Section (4.2.4)) in H_T is defined by Equation (4.10), where k_{QCD} is constrained by measurements in a signal sideband region.

The expected QCD background, QCD^i, within a bin i is then modelled as,

$$QCD^i = m^i A_{QCD} e^{-k_{QCD}\langle H_T \rangle}, \tag{4.21}$$

where m_i represent the number of events observed with $\alpha_T \leq 0.55$ in each H_T bin i, and $\langle H_T \rangle$ represent the mean H_T of each bin. Expressed as functions of just the zeroth bin, QCD^0, and k_{QCD}, the QCD expectation is given by

$$QCD^i = QCD^0(\frac{m^i}{m^0})e^{-k_{QCD}(\langle H_T \rangle^i - \langle H_T \rangle^0}. \tag{4.22}$$

4.8.3. EWK Control Samples

The EWK background estimation within each bin, i, is broken into two components, the expected yield from Z $\rightarrow \nu\bar{\nu}$ and $t\bar{t}$-W (plus other residual backgrounds) events. This is written as, Z^i_{inv} and $t\bar{t}W^i$, and it follows that

$$EWK^i = Z^i_{inv} + t\bar{t}W^i. \tag{4.23}$$

This can be further expressed as

$$Z^i_{inv} \equiv f^i_{Zinv} \times EWK^i, \tag{4.24}$$
$$t\bar{t}W^i \equiv (1 - f^i_{Zinv}) \times EWK^i, \tag{4.25}$$

where f^i_{Zinv} represents the expected yield from Z $\rightarrow \nu\bar{\nu}$ in bin i divided by the expected EWK background EWK^i. This fraction is modelled as a linear component

$$f^i_{Zinv} = f^0_{Zinv} + \frac{\langle H_T\rangle^i - \langle H_T\rangle^0}{\langle H_T\rangle^{N-1} - \langle H_T\rangle^0}(f^{N-1}_{Zinv} - f^0_{Zinv}), \tag{4.26}$$

where N again represents the number of H_T bins, and f^0_{Zinv} and f^{N-1}_{Zinv} are float parameters whose final values are limited between zero and one.

Within each H_T bin there are three background measurements for the different control samples, n^i_γ, n^i_μ and $n^i_{\mu\mu}$, representing the event yields from the γ + jets, μ + jets and $\mu\mu$ + jets control samples respectively. Each of these have a corresponding yield in simulation, MC^i_γ, MC^i_μ and $MC^i_{\mu\mu}$. Within the hadronic signal region there are also corresponding simulated yields for Z $\rightarrow \nu\bar{\nu}$ (MC^i_{Zinv}) and $t\bar{t}$ + W ($MC^i_{t\bar{t}+W}$), which are used to define

$$r^i_\gamma = \frac{MC^i_\gamma}{MC^i_{Zinv}}; \qquad r^i_{\mu\mu} = \frac{MC^i_{\mu\mu}}{MC^i_{Zinv}}; \qquad r^i_\mu = \frac{MC^i_\mu}{MC^i_{t\bar{t}+W}} \tag{4.27}$$

where r_p^i represents the inverse of the TFs used to extrapolate the yield of each background process.

The likelihoods regarding the three measured yields n_γ^i, $n_{\mu\mu}^i$, n_μ^i can then be fully expressed as

$$L_\gamma = \prod_i \text{Pois}(n_\gamma^i | \rho_{\gamma Z}^j . r_\gamma^i . Z_{inv}^i), \tag{4.28}$$

$$L_{\mu\mu} = \prod_i \text{Pois}(n_{\mu\mu}^i | \rho_{\mu\mu Z}^j . r_{\mu\mu}^i . Z_{inv}^i), \tag{4.29}$$

$$L_\mu = \prod_i \text{Pois}(n_\mu^i | \rho_{\mu Y}^j . r_\mu^i . Y^i + s_\mu^i), \tag{4.30}$$

$$\tag{4.31}$$

which contain an additional term s_μ^i, which represents the signal contamination in the μ + jets sample. The parameters $\rho_{\gamma Z}^j$, $\rho_{\mu\mu}^j$ and ρ_μ^j represent "correction factors" that accommodate the data driven systematic uncertainties derived from the control samples in Section (4.12).

Each of these equations are used to estimate the maximum likelihood value for relevant background in the signal region given the observations n_p^i in each of the control samples (see Section (4.2.3)).

The measurements in each of the control samples and the hadronic signal region, along with the ratios r_γ^i, $r_{\mu\mu}^i$, and r_μ^i, are all considered simultaneously through the relationships defined by Equations (4.19),(4.24) and (4.25).

In addition to the Poisson product, an additional log-normal term is introduced to accommodate the systematic uncertainties given by,

$$L_{EWK\ syst} = \prod_j \text{Logn}(1.0 | \rho_{\mu W}^j, \sigma_{\mu W}^j) \times \text{Logn}(1.0 | \rho_{\mu\mu Z}^j, \sigma_{\mu\mu Z}^j) \times \text{Logn}(1.0 | \rho_{\gamma Z}^j, \sigma_{\gamma Z}^j). \tag{4.32}$$

The parameters ρ^j, ρ^j and ρ^j represent the already introduced "correction factors" that accommodate the systematic uncertainties, while the quantities $\sigma_{\gamma Z}^j$, $\sigma_{\mu\mu Z}^j$ and

$\sigma^j_{\mu W}$ represent the relative systematic uncertainties for the respective control sample constraints. Logn represents the log-normal distribution [92],

$$\mathrm{Logn}(x \mid \mu, \sigma_{rel}) = \frac{1}{x\sqrt{2\pi}\mathrm{ln\ k}} \exp\left(\frac{\mathrm{ln}^2(\frac{x}{\mu})}{2\mathrm{ln}^2\mathrm{k}}\right); \quad k = 1 + \sigma_{rel}. \tag{4.33}$$

Five parameters per control sample are used to span the eight H_T categories, with just one used for the three H_T in the $n_b^{reco} \geq 4$ category. These parameters span the same H_T ranges described in Section (4.6) and is shown in Table 4.18.

H_T bin (i)	0	1	2	3	4	5	6	7
syst. parameter (j)	0	1	2	2	3	3	4	4

H_T bin (i)	0	1	2
syst. parameter (j)	0	0	0

Table 4.18: The systematic parameters used in H_T bins. Left: categories with eight bins; right: category with three bins.

Alternatively, in the higher n_b^{reco} categories ($n_b^{reco} = 2$ and above), only the single muon sample is used to constrain the total EWK background. This is due to a lack of statistics in the $\mu\mu$ + jets and γ + jets at these $n_{\mathrm{b}}^{\mathrm{reco}}$ multiplicities. Therefore the likelihood functions for the control samples are reduced and simply represented by

$$L'_\mu = \prod_i \mathrm{Pois}(n^i_\mu | \rho^j_{\mu Y}.r'^i_\mu.EWK^i + s^i_\mu), \tag{4.34}$$

where

$$r'^i_\mu = \frac{MC^i_\mu}{MC^i_{t\bar{t}+W} + MC^i_{Zinv}}. \tag{4.35}$$

4.8.4. Contributions from Signal

The cross-section for each model is represented by x, while l represents the total recorded luminosity considered by the analysis in the signal region. Let ϵ^i_{had} and ϵ^i_μ represent the analysis selection efficiency for that particular signal model in H_T bin i of the hadronic and μ + jets control sample respectively. Letting δ represent the relative uncertainty on the signal yield, assumed to be fully correlated across all bins, and ρ_{sig} the "correction factor" to the signal yield which accommodates this uncertainty. f represents an unknown multiplicative factor on the signal cross section, for which an allowed interval is computed. The expected signal yield s^i is thus given by

$$s^i \equiv f \rho_{sig} x l \epsilon^i_{had} \tag{4.36}$$

and signal contamination with the μ + jets control sample by

$$s^i_\mu \equiv f \rho_{sig} x l \epsilon^i_\mu . \tag{4.37}$$

The systematic uncertainty on the signal is additionally incorporated by the term

$$L_{\text{sig}} = \text{Logn}(1.0 | \rho_{sig}, \delta). \tag{4.38}$$

A discussion of the SMS signal models through which the analysis is interpreted can be found in the following chapter.

4.8.5. Total Likelihood

The total likelihood function for a given signal category k(n_b^{reco}, n_{jet}) is then given by the product of the likelihood functions introduced within the previous sections:

$$\begin{aligned} L^k_{\text{Tot}} &= L^k_{had} \times L^k_\mu \times L^k_\gamma \times L^k_{\mu\mu} \times L^k_{EWKsyst} \times L^k_{QCD} && (0 \leq n_{\text{b}}^{\text{reco}} \leq 1), \\ L^k_{\text{Tot}} &= L^k_{had} \times L'^k_\mu \times L^k_{\mu\ syst} \times L^k_{QCD} && (n_{\text{b}}^{\text{reco}} \geq 2). \end{aligned} \tag{4.39}$$

In categories containing eight H_T bins and utilising the three control samples (μ + jets, $\mu\mu$ + jets, γ + jets), there are 25 nuisance parameters. When just one control sample is used to estimate the EWK background, this is reduced to 15 nuisance parameters. In the $n_{\rm b}^{\rm reco} \geq 4$ category where only three H_T bins are used, there are just 6 nuisance parameters. This information is summarised within Table 4.19.

Nuisance parameter	Total
$(EWK^i)_{\ i\ :0-7(2)}$	8 (3)
f^0_{Zinv}	1*
f^7_{Zinv}	1*
QCD^0	1
k_{QCD}	1
$(\rho^j_{\gamma Z})_{\ j\ :2-4}$	3 *
$(\rho^j_{\mu\mu Z})_{\ j\ :0-4}$	5 *
$(\rho^j_{\mu W})_{\ j\ :0-4(0)}$	5 (1)

Table 4.19: Nuisance parameters used within the different hadronic signal bins of the analysis. Parameters denoted by a * are not considered in the case of a single control sample being used to predict the EWK background. Numbers within brackets highlight the number of nuisance parameters in the case of three H_T bins being used.

When considering SUSY signal models within the likelihood, the additional $L_{\rm sig}$ term is included and therefore when multiple categories are fitted simultaneously the total likelihood is then represented by

$$L^{\rm signal}_{\rm Tot} = L_{sig} \times \prod_k L^k_{\rm Tot}. \tag{4.40}$$

Chapter 5.

Results and Interpretation

Using the statistical framework outlined in the previous chapter, results are shown for the compatibility of the collected data with a SM-only hypothesis in Section (5.1). The data is further interpreted within the context of various SMS models within Section (5.2).

5.1. Compatibility with the Standard Model Hypothesis

The SM background only hypothesis is tested by removing any signal contributions within the signal and control samples, and the likelihood function defined in Equation (4.39) maximised over all parameters using Rootfit [93] and MINUIT [94]. The results of the search consist of the observed yields in the hadronic signal sample, and the μ + jets, $\mu\mu$ + jets and γ + jets control samples.

These observed yields along with the expectations and uncertainties given by the simultaneous fit for the hadronic signal region are displayed in Table 5.2. The results obtained from the simultaneous fits, including that of the three control samples, are shown in Figure 5.1-5.8, and which is summarised in Table 5.1.

The figures show a comparison between the observed yields and the SM expectations as given by the fit across all H_T bins, and in all n_{jet} and n_b^{reco} multiplicity categories. In all categories the samples are well described by the SM only hypothesis. In particular no significant excess is observed above SM expectation within the hadronic signal region.

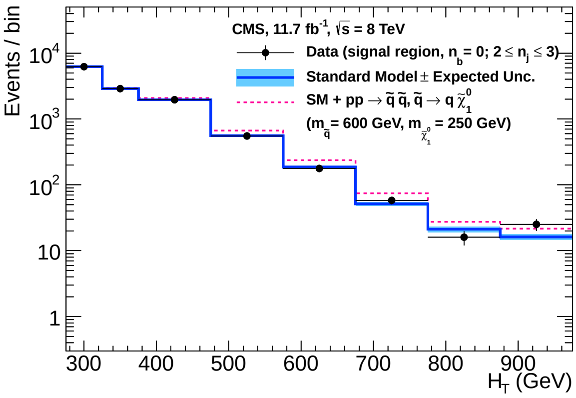
n_{jet}	n_b^{reco}	Control samples fitted	Figure
2-3	0	μ + jets, $\mu\mu$ + jets, γ + jets	5.1
2-3	1	μ + jets, $\mu\mu$ + jets, γ + jets	5.2
2-3	2	μ + jets	5.3
$\geq$4	0	μ + jets, $\mu\mu$ + jets, γ + jets	5.4
$\geq$4	1	μ + jets, $\mu\mu$ + jets, γ + jets	5.5
$\geq$4	2	μ + jets	5.6
$\geq$4	3	μ + jets	5.7
$\geq$4	4	μ + jets	5.8

Table 5.1: Summary of control samples used by each fit results, and the Figures in which they are displayed.

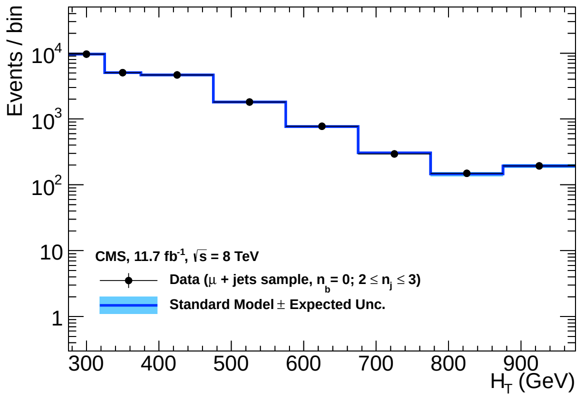
			H_T bin (GeV)							
Cat	n_b^{reco}	n_{jet}	275-325	325-375	375-475	474-575	575-675	675-775	775-875	875-∞
SM	0	≤ 3	6235^{+100}_{-67}	2900^{+60}_{-54}	1955^{+34}_{-39}	558^{+14}_{-15}	186^{+11}_{-10}	$51.3^{+3.4}_{-3.8}$	$21.2^{+2.3}_{-2.2}$	$16.1^{+1.7}_{-1.7}$
Data			6232	2904	1965	552	177	58	16	25
SM	0	≥ 4	1010^{+34}_{-24}	447^{+19}_{-16}	390^{+19}_{-15}	250^{+12}_{-11}	111^{+9}_{-7}	$53.3^{+4.3}_{-4.3}$	$18.5^{+2.4}_{-2.4}$	$19.4^{+2.5}_{-2.7}$
Data			1009	452	375	274	113	56	16	27
SM	1	≤ 3	1162^{+37}_{-29}	481^{+18}_{-19}	341^{+15}_{-16}	$86.7^{+4.2}_{-5.6}$	$24.8^{+2.8}_{-2.7}$	$7.2^{+1.1}_{-1.0}$	$3.3^{+0.7}_{-0.7}$	$2.1^{+0.5}_{-0.5}$
Data			1164	473	329	95	23	8	4	1
SM	1	≥ 4	521^{+25}_{-17}	232^{+15}_{-12}	188^{+12}_{-11}	106^{+6}_{-6}	$42.1^{+4.1}_{-4.4}$	$17.9^{+2.2}_{-2.0}$	$9.8^{+1.5}_{-1.4}$	$6.8^{+1.2}_{-1.1}$
Data			515	236	204	92	51	13	13	6
SM	2	≤ 3	224^{+15}_{-14}	$98.2^{+8.4}_{-6.4}$	$59.0^{+5.2}_{-6.0}$	$12.8^{+1.6}_{-1.6}$	$3.0^{+0.9}_{-0.7}$	$0.5^{+0.2}_{-0.2}$	$0.1^{+0.1}_{-0.1}$	$0.1^{+0.1}_{-0.1}$
Data			222	107	58	12	5	1	0	0
SM	2	≥ 4	208^{+17}_{-9}	103^{+9}_{-7}	$85.9^{+7.2}_{-6.9}$	$51.7^{+4.6}_{-4.7}$	$19.9^{+3.4}_{-3.0}$	$6.8^{+1.2}_{-1.3}$	$1.7^{+0.7}_{-0.4}$	$1.3^{+0.4}_{-0.3}$
Data			204	107	84	59	24	5	1	2
SM	3	≤ 3	$8.6^{+2.8}_{-0.8}$	$4.6^{+1.0}_{-0.9}$	$2.7^{+0.7}_{-0.7}$	$0.3^{+0.2}_{-0.1}$	$0.^{+0.0}_{-0.0}$	$0.^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$
Data			8	3	2	0	1	0	0	0
SM	3	≥ 4	$25.3^{+5.0}_{-4.2}$	$11.7^{+1.7}_{-1.8}$	$6.7^{+1.4}_{-1.2}$	$3.9^{+0.8}_{-0.8}$	$2.3^{+0.6}_{-0.6}$	$1.2^{+0.3}_{-0.4}$	$0.3^{+0.2}_{-0.1}$	$0.1^{+0.1}_{-0.1}$
Data			25	13	4	2	2	3	0	0
SM	4	≥ 4	$0.9^{+0.4}_{-0.7}$	$0.3^{+0.2}_{-0.2}$	$0.6^{+0.3}_{-0.3}$					
Data			1	0	2					

Table 5.2: Comparison of the measured yields in each H_T, n_{jet} and n_b^{reco} jet multiplicity bins for the hadronic sample with the SM expectations and combined statistical and systematic uncertainties given by the simultaneous fit. Note that the $n_{\mathrm{b}}^{\mathrm{reco}} = 3$, $n_{\mathrm{jet}} \leq 3$ category is not used in any interpretations within this section but is included for completeness.

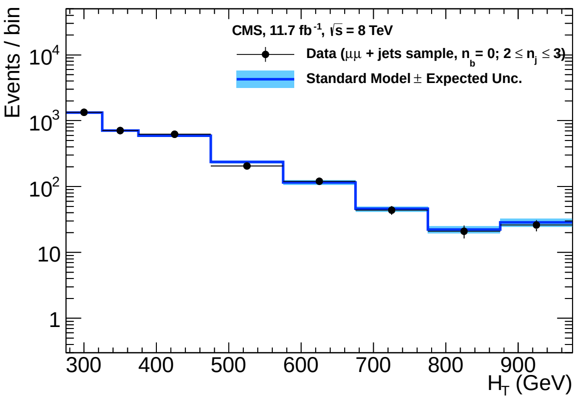
Given the lack of an excess in data hinting at a possible supersymmetric signature within the data, interpretations are made on the production masses and cross-section of a range of SUSY decay topologies within the following section.



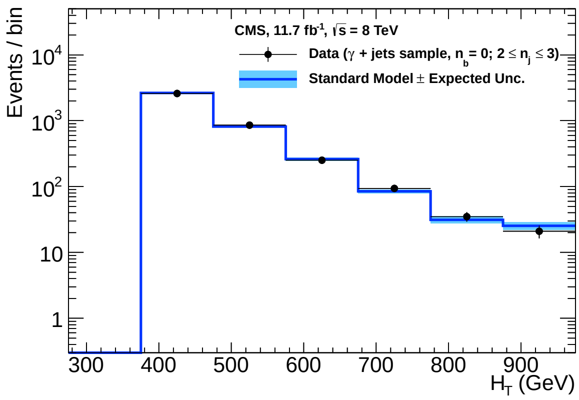
(a) Hadronic sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 0$



(b) μ + jets sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 0$

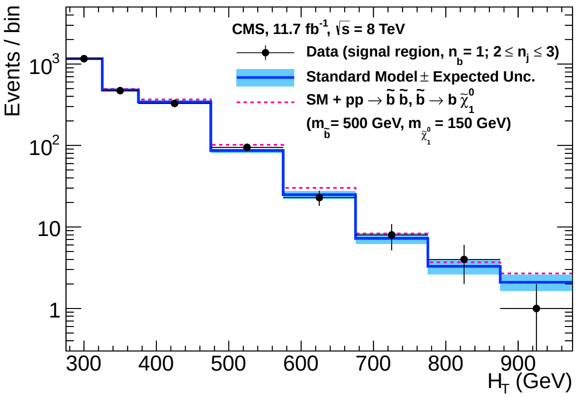


(c) $\mu\mu$ + jets sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 0$

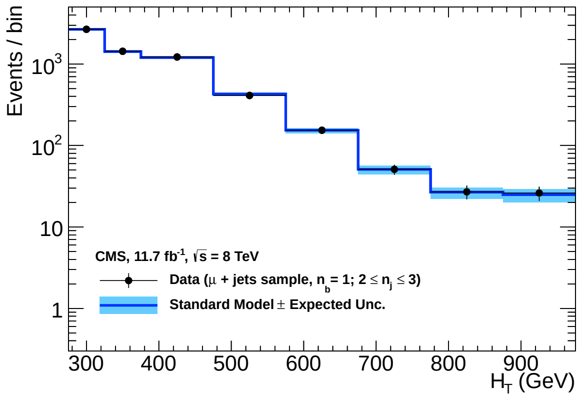


(d) γ + jets sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 0$

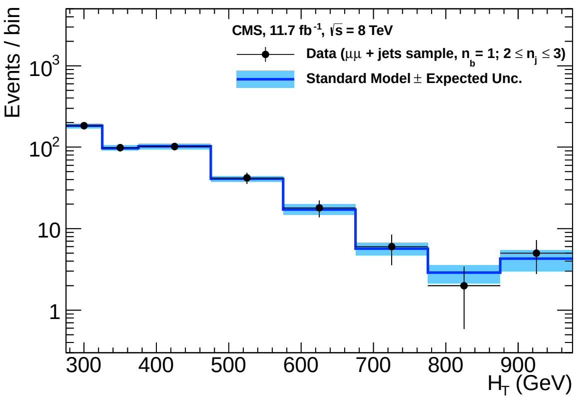
Figure 5.1: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} = 0$ and $n_{jet} \leq 3$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown. An example signal expectation (red solid line) for the $D1$ SMS signal point from Table 4.1 is superimposed on the SM background expectation.



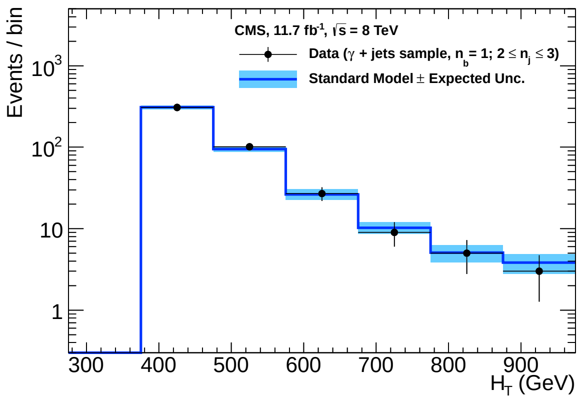
(a) Hadronic sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 1$



(b) μ + jets sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 1$

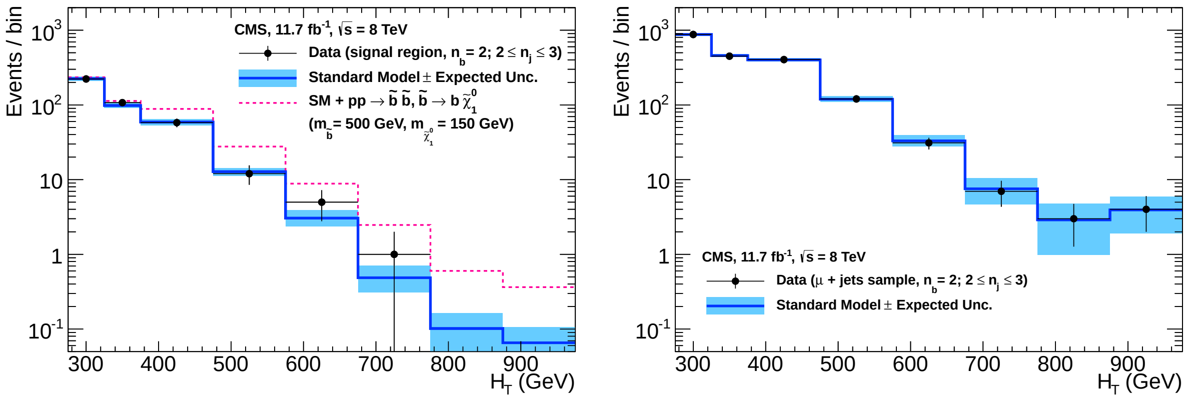


(c) $\mu\mu$ + jets sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 1$



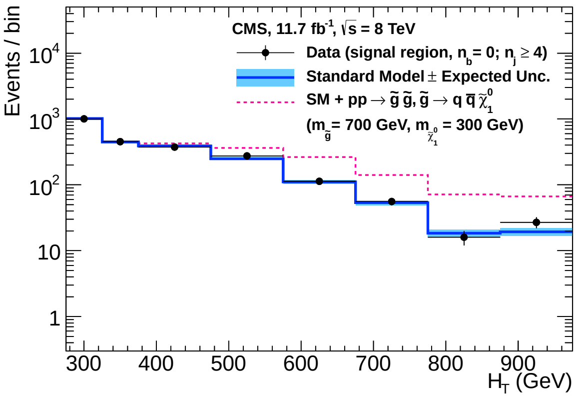
(d) γ + jets sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 1$

Figure 5.2: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} = 1$ and $n_{jet} \leq 3$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown. An example signal expectation (red solid line) for the $D2$ SMS signal point from Table 4.1 is superimposed on the SM background expectation.

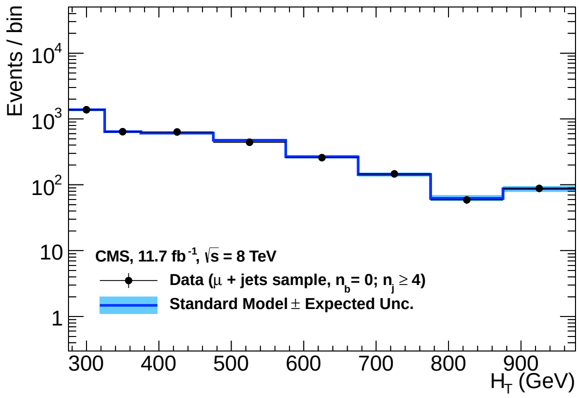


(a) Hadronic sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 2$ (b) μ + jets sample, $2 \leq n_{jet} \leq 3$ and $n_b^{reco} = 2$

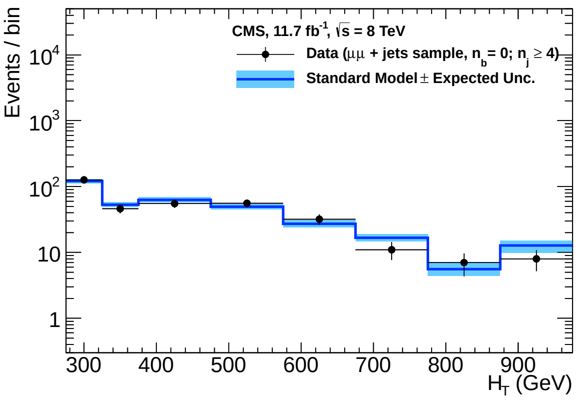
Figure 5.3: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} = 2$ and $n_{jet} \leq 3$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown. An example signal expectation (red solid line) for the $D2$ SMS signal point from Table 4.1 is superimposed on the SM background expectation.



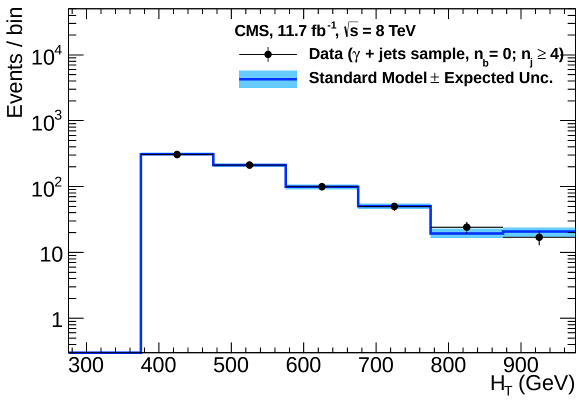
(a) Hadronic sample, $n_{jet} \geq 4$ and $n_b^{reco} = 0$



(b) μ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 0$

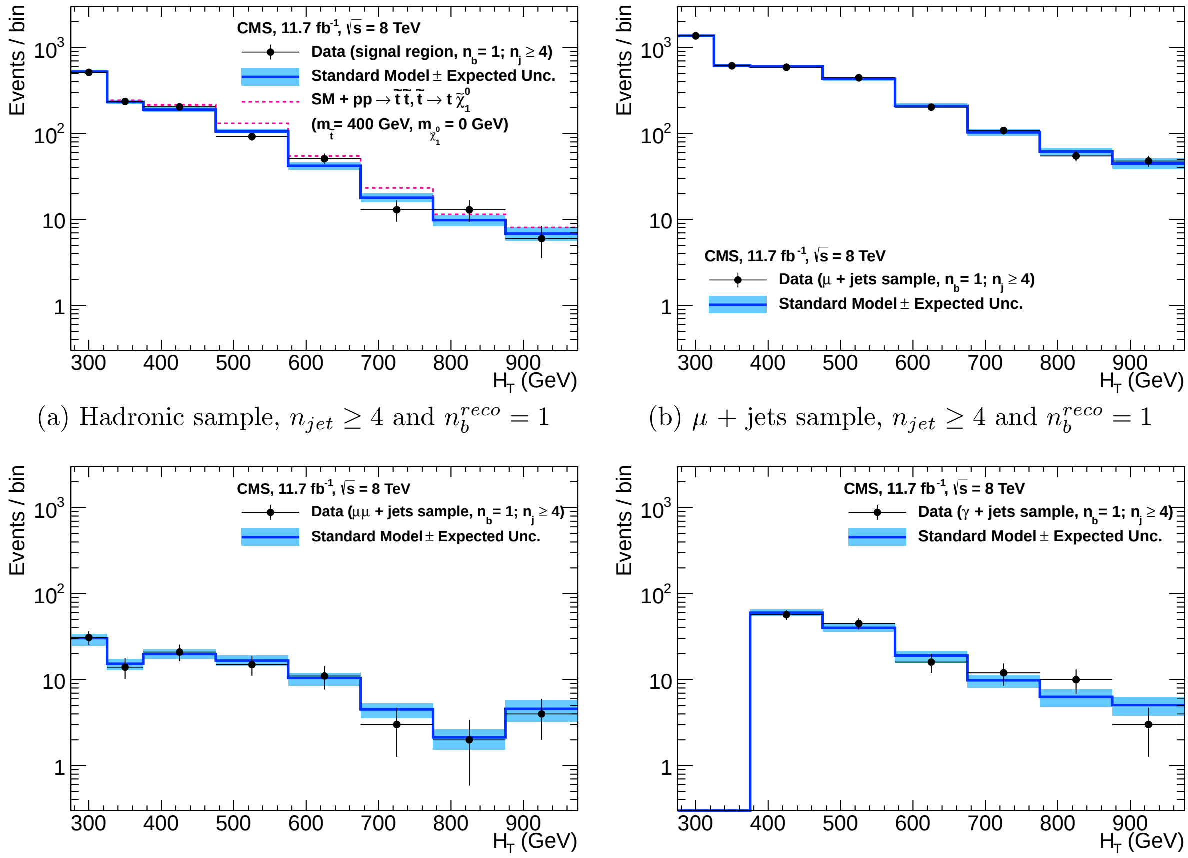


(c) $\mu\mu$ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 0$



(d) γ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 0$

Figure 5.4: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} = 0$ and $n_{jet} \geq 4$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown. An example signal expectation (red solid line) for the $D2$ SMS signal point from Table 4.1 is superimposed on the SM background expectation.



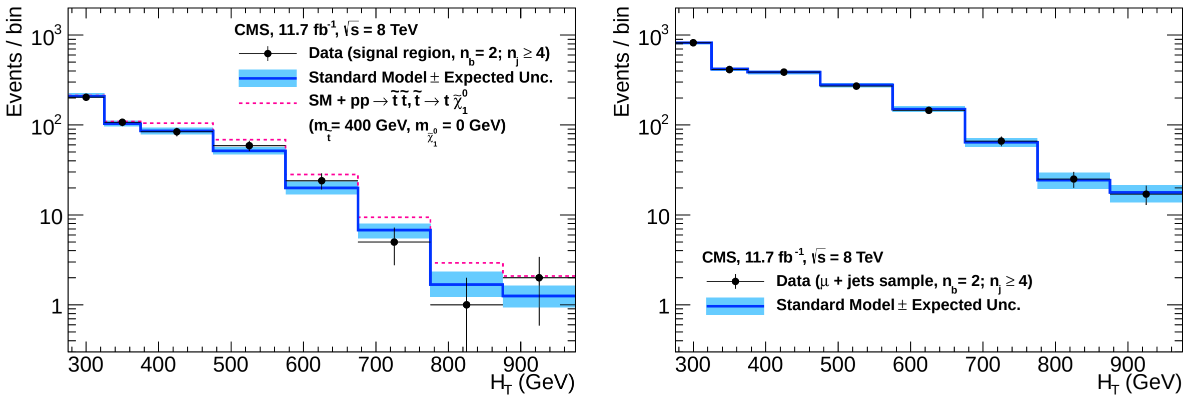
(a) Hadronic sample, $n_{jet} \geq 4$ and $n_b^{reco} = 1$

(b) μ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 1$

(c) $\mu\mu$ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 1$

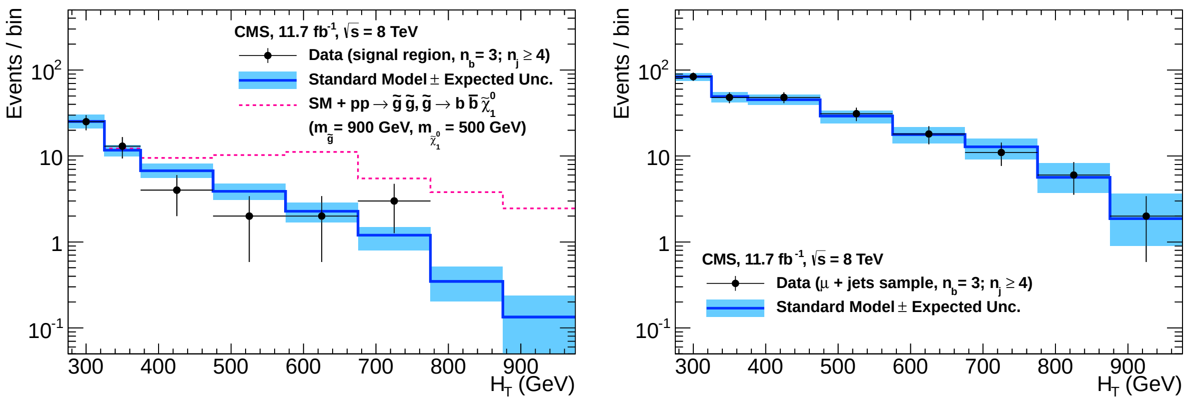
(d) γ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 1$

Figure 5.5: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} = 1$ and $n_{jet} \geq 4$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown.



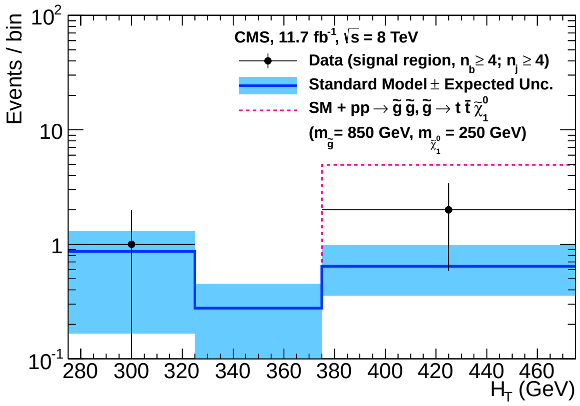
(a) Hadronic sample, $n_{jet} \geq 4$ and $n_b^{reco} = 2$ (b) μ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 2$

Figure 5.6: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} = 2$ and $n_{jet} \geq 4$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown. An example signal expectation (red solid line) for the $D3$ SMS signal point from Table 4.1 is superimposed on the SM background expectation.

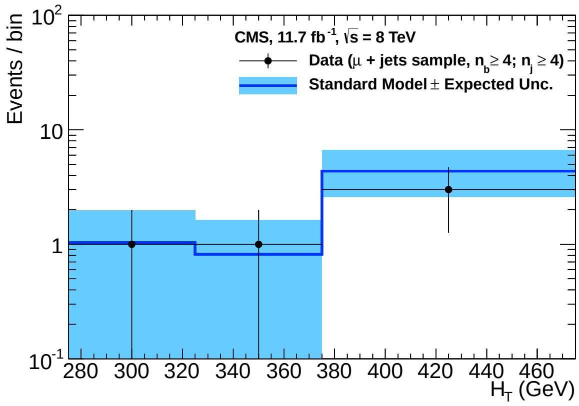


(a) Hadronic sample, $n_{jet} \geq 4$ and $n_b^{reco} = 3$ (b) μ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} = 3$

Figure 5.7: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} = 3$ and $n_{jet} \geq 4$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown. An example signal expectation (red solid line) for the $G2$ SMS signal point from Table 4.1 is superimposed on the SM background expectation.



(a) Hadronic sample, $n_{jet} \geq 4$ and $n_b^{reco} \geq 4$



(b) μ + jets sample, $n_{jet} \geq 4$ and $n_b^{reco} \geq 4$

Figure 5.8: Comparison of the observed yields and SM expectations given by the simultaneous fit in bins of H_T for the (a) hadronic, (b) μ + jets, (c) $\mu\mu$ + jets and (d) γ + jets samples when requiring $n_b^{reco} \geq 4$ and $n_{jet} \geq 4$. The observed event yields in data (black dots) and the expectations and their uncertainties for all SM processes (blue line with light blue bands) are shown. An example signal expectation (red solid line) for the $G3$ SMS signal point from Table 4.1 is superimposed on the SM background expectation.

5.2. SUSY

Limits are set on sparticle and LSP masses in the parameter space of a set of SMS models that characterise supersymmetric final states resulting from; direct third generation or light squark production, and gluino induced production of supersymmetric particles. However as detailed in Section (2.4.1), the individual models are not representative of a real physical SUSY model as only one decay process is considered. Instead these models represent a way to test for signs of specific signatures indicating new physics.

5.2.1. The $\mathbf{CL_s}$ Method

The CLs method [95][96][97] is used to compute the limits for signal models, with the one-sided profile likelihood ratio as the test statistic [98].

The test statistic is defined as

$$q(\mu) = \begin{cases} -2\log\lambda(\mu) & \text{when } \mu \geq \hat{\mu}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

where

$$\lambda(\mu) = \frac{L(\mu, \theta_\mu)}{L(\hat{\mu}, \hat{\theta})} \tag{5.2}$$

represents the profile likelihood ratio, in which $\mu \equiv f$ from Section (4.8.4), is the parameter characterising the signal strength. $\hat{\mu}$ is defined as the maximum likelihood value, $\hat{\theta}$ the set of maximum likelihood values of the nuisance parameters and θ_μ the set of maximum values of the nuisance parameters for a given value of μ.

When $\mu \equiv f = 1$, the signal model is considered at its nominal production cross section. The distribution of q_μ is built up via the generation of pseudo experiments in order to obtain two distributions for the background (B) and signal plus background (S+B) cases.

The compatibility of a signal model with observations in data is determined by the parameter CL_s,

$$\mathrm{CL}_S = \frac{\mathrm{CL}_{S+B}}{\mathrm{CL}_B}, \tag{5.3}$$

with CL_B and CL_{S+B} defined as one minus the quantiles of the observed value in the data of the two distributions. A model is considered to be excluded at 95% confidence level when $\mathrm{CL}_s \leq 0.05$ [99].

5.2.2. Interpretation in Simplified Signal Models

Different n_{jet} and $n_{\mathrm{b}}^{\mathrm{reco}}$ bins are used in the interpretation of different SMS models. The choice of categories used, are made such that the signal to background ratio will be maximised for the model in question, increasing sensitivity to that particular type of final state signature. The production and decay modes of the SMS models under consideration are summarised in Table 5.3, with limit plots of the experimental reach in these models shown in Figure 5.10.

The models `T1` and `T2` are used to characterise the pair production of gluinos and first or second generation squarks respectively. The low number of third generation quarks produced from this decay topology makes choosing to interpret within the $n_{\mathrm{b}}^{\mathrm{reco}} = 0$ category beneficial to improving sensitivity to these models. In the case of the `T2` model, two sets of exclusion contours are shown. These correspond to the production of eight first- and second-generation (left-/right-handed) squarks with degenerate masses and the case of just a single light squark with all other squarks decoupled at much higher masses.

Conversely the `T2bb`, `T1tttt`, and `T1bbbb` SMS models describe various production and decay mechanisms in the context of third-generation squarks. In this situation considering higher $n_{\mathrm{b}}^{\mathrm{reco}}$ categories bring significant improvements to the sensitivity to these types of final state signature.

Finally the choice of jet category is made dependant upon the production mechanism, where gluino induced and direct squark production results in a large or small number of final state jets respectively.

Experimental uncertainties on the SM background predictions ($10 - 30\%$, described in Section (4.6.1)), the luminosity measurement (4.4%), and the total acceptance times

Model	Production/decay	n_{jet}	n_b^{reco}	Process	Limit	$\mathrm{m}^{\mathrm{best}}_{\widetilde{q}(\widetilde{g})}$ (GeV)	$\mathrm{m}^{\mathrm{best}}_{\mathrm{LSP}}$ (GeV)
`T1`	$pp \to \widetilde{g}\widetilde{g}^* \to q\bar{q}\widetilde{\chi}^0_1 q\bar{q}\widetilde{\chi}^0_1$	≥ 4	0	5.9(a)	5.10(a)	~ 950	~ 450
`T2`	$pp \to \widetilde{q}\widetilde{q}^* \to q\widetilde{\chi}^0_1 \bar{q}\widetilde{\chi}^0_1$	≤ 3	0	5.9(b)	5.10(b)	~ 775	~ 325
`T2bb`	$pp \to \widetilde{b}\widetilde{b}^* \to b\widetilde{\chi}^0_1 \bar{b}\widetilde{\chi}^0_1$	≤ 3	1,2	5.9(c)	5.10(c)	~ 600	~ 200
`T1tttt`	$pp \to \widetilde{g}\widetilde{g}^* \to t\bar{t}\widetilde{\chi}^0_1 t\bar{t}\widetilde{\chi}^0_1$	≥ 4	2,3,≥ 4	5.9(d)	5.10(d)	~ 975	~ 325
`T1bbbb`	$pp \to \widetilde{g}\widetilde{g}^* \to b\bar{b}\widetilde{\chi}^0_1 b\bar{b}\widetilde{\chi}^0_1$	≥ 4	2,3,≥ 4	5.9(e)	5.10(e)	~ 1125	~ 650

Table 5.3: A table representing the SMS models interpreted within the analysis. The model name and production and decay chain is specified in the first two columns. Each SMS model is interpreted in specific n_{jet} and n_b^{reco} categories which are detailed in the third and fourth columns. The last two columns indicate the search sensitivity for each model, representing the largest $m_{\widetilde{q}/\widetilde{g}}$ mass beyond which no limit can be set for this particular decay topology. The quoted values are conservatively determined from the observed exclusion based on the theoretical production cross section minus 1σ uncertainty.

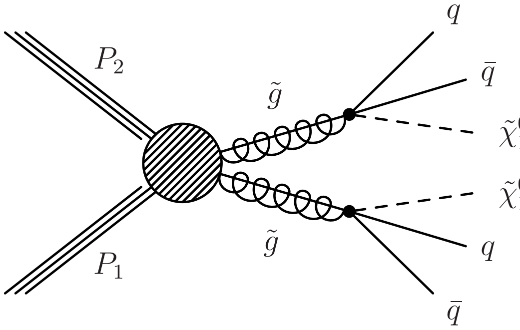
efficiency of the selection for the considered signal model ($12 - 18\%$, from Section (4.7)) are included in the calculation of the limit.

Signal efficiency in the kinematic region defined by $0 < m_{\widetilde{g}(\widetilde{q})} - m_{LSP} < 175$ GeV or $m_{\widetilde{g}(\widetilde{q})} < 300$ GeV is strongly affected by the presence of Initial State Radiation (ISR). This is a region in which direct (i.e. non-ISR induced) production is kinematically forbidden due to the $H_T > 275$ GeV requirement, therefore a large percentage of signal acceptance is due to the effect of ISR jets. Given the large associated uncertainties, no interpretation is provided for this kinematic region.

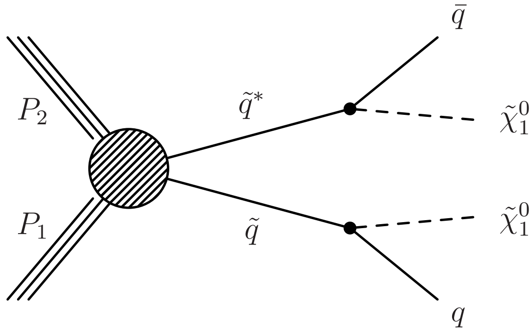
The estimates on mass limits shown in Table 5.3, are determined conservatively from the observed exclusion based on the theoretical production cross section, minus 1σ uncertainty. The most stringent mass limits on pair-produced sparticles are obtained at low LSP masses and larger squark and gluino masses due to the high p_T jets and consequently high H_T of such signal topologies. The limits are seen to weaken for compressed spectra points closer to the diagonal, where the signal populates the lower H_T bins in which more background resides. For all of the considered models, there is an LSP mass beyond which no limit can be set, which can be observed from the figures referenced in the table.

Two small upwards fluctuations are observed within the data, and are seen at high H_T within the $n_{\mathrm{b}}^{\mathrm{reco}} = 0$ category and at mid-H_T in the $n_{\mathrm{b}}^{\mathrm{reco}} = 1, 2$ categories (see Table 5.2). As each of these fluctuations occur within at least one of the analysis categories that each SMS model interpretation is made, the observed exclusions within all SMS models are generally found to be weaker than the expected limits in the region of 1-2 standard deviations. In isolation these fluctuations are not significant and additional data would be necessary to make any further conclusions.

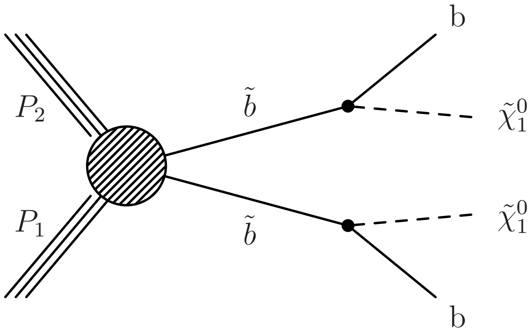
Despite these fluctuations, the range of parameter space that can be excluded has been extended with respect to analysis based upon the $\sqrt{s} = 7$ TeV dataset [100], by up to 225 and 150 GeV for $m^{best}_{\tilde{q}(\tilde{g})}$ and m^{best}_{LSP} respectively. The parameter space for third generation signatures is increasingly squeezed for larger mass splitting, with exclusions in the region of 1 TeV in these topologies.



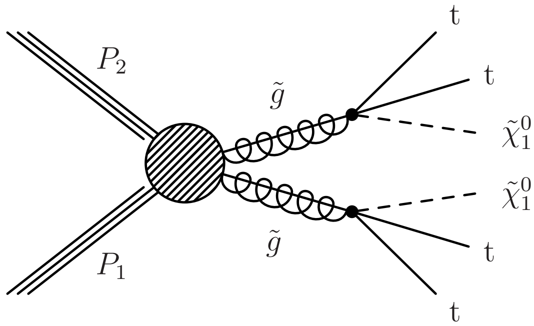
(a) $\widetilde{g}\widetilde{g}^* \rightarrow q\bar{q}\widetilde{\chi}^0_1 q\bar{q}\widetilde{\chi}^0_1$ (`T1`)



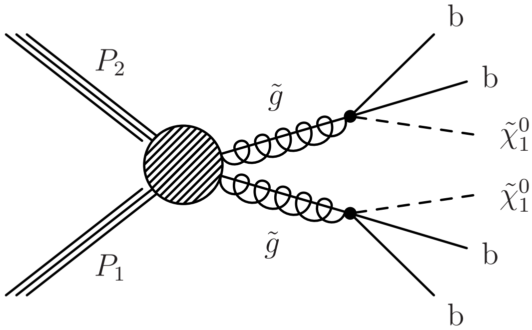
(b) $\widetilde{q}\widetilde{q}^* \rightarrow q\widetilde{\chi}^0_1\bar{q}\widetilde{\chi}^0_1$ (`T2`)



(c) $\widetilde{b}\widetilde{b}^* \rightarrow b\widetilde{\chi}^0_1\bar{b}\widetilde{\chi}^0_1$ (`T2bb`)

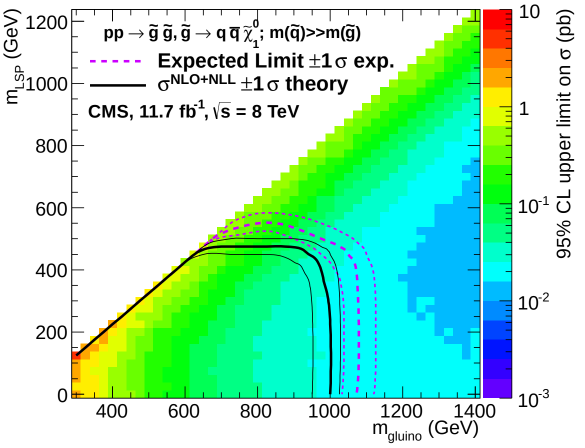


(d) $\widetilde{g}\widetilde{g}^* \rightarrow t\bar{t}\widetilde{\chi}^0_1 t\bar{t}\widetilde{\chi}^0_1$ (`T1tttt`)

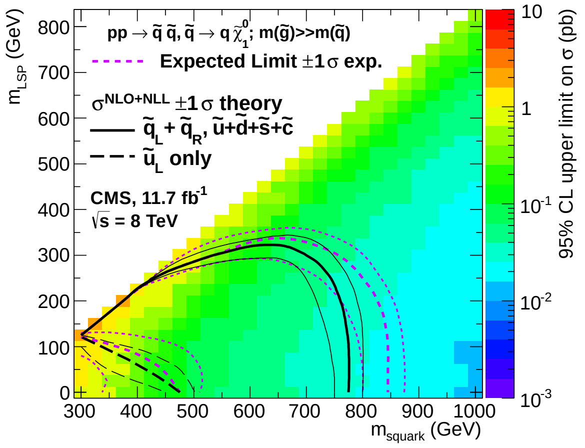


(e) $\widetilde{g}\widetilde{g}^* \rightarrow b\bar{b}\widetilde{\chi}^0_1 b\bar{b}\widetilde{\chi}^0_1$ (`T1bbbb`)

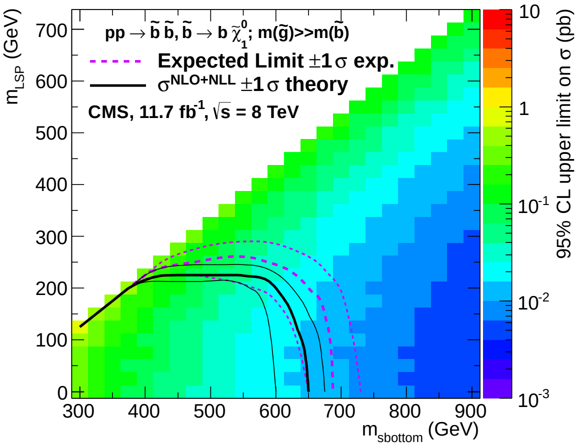
Figure 5.9: Production and decay modes for the various SMS models interpreted within the analysis.



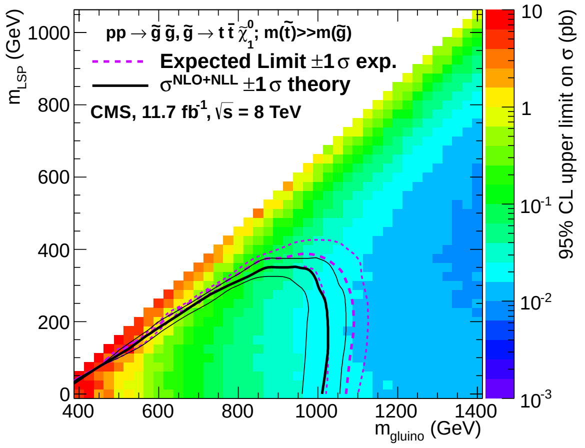
(a) $\widetilde{g}\widetilde{g}^{*} \to q\bar{q}\widetilde{\chi}_1^0 q\bar{q}\widetilde{\chi}_1^0$ (T1)



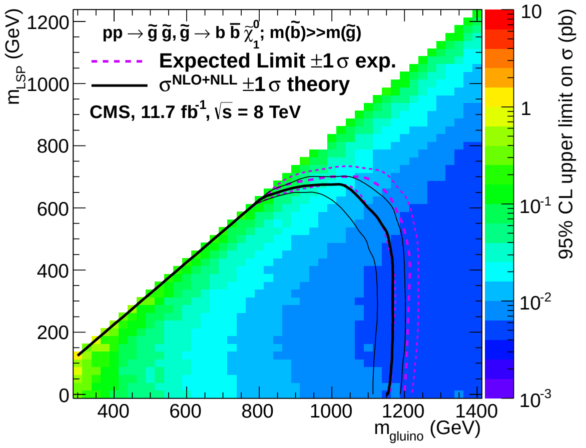
(b) $\widetilde{q}\widetilde{q}^{*} \to q\widetilde{\chi}_1^0 \bar{q}\widetilde{\chi}_1^0$ (T2)



(c) $\widetilde{b}\widetilde{b}^{*} \to b\widetilde{\chi}_1^0 \bar{b}\widetilde{\chi}_1^0$ (T2bb)



(d) $\widetilde{g}\widetilde{g}^{*} \to t\bar{t}\widetilde{\chi}_1^0 t\bar{t}\widetilde{\chi}_1^0$ (T1tttt)



(e) $\widetilde{g}\widetilde{g}^{*} \to b\bar{b}\widetilde{\chi}_1^0 b\bar{b}\widetilde{\chi}_1^0$ (T1bbbb)

Figure 5.10: Upper limit of cross section at 95% CL as a function of $m_{\widetilde{q}/\widetilde{g}}$ and m_{LSP} for various SMS models. The solid thick black line indicates the observed exclusion region assuming NLO and NLL SUSY production cross section. The analysis selection efficiency is measured for each interpreted model, with the signal yield per point given by $\epsilon \times \sigma$. The thin black lines represent the observed excluded region when varying the cross section by its theoretical uncertainty. The dashed purple lines indicate the median (thick line) 1σ (thin lines) expected exclusion regions.

Chapter 6.

SUSY Searches with B-tag Templates

Within this chapter a complementary technique is discussed as a means to predict the distribution of three and four reconstructed b-tagged ($n_{\mathrm{b}}^{\mathrm{reco}}$ = 3, 4), jets in an event sample. The recent discovery of the Higgs boson has made "Natural SUSY" models attractive, given that light top and bottom squarks are a candidate to stabilise divergent loop corrections to the Higgs boson mass. A light gluino which subsequently decays to third generation sparticle pairs, will give rise to many events with a large number of final state b-tagged jets.

The method described within this chapter is used to estimate the SM background at high b-tagged jet multiplicities (3-4), from a templated fit conducted in a low b-tagged jet (0-2) control region of an event sample. This approach can hypothetically be applied to generic supersymmetric searches, to gain sensitivity to signals which contain a higher number of b-tagged jets than the search's dominant SM backgrounds.

As a proof-of-concept, the procedure is applied to the SM enriched μ + jets control sample of the α_T search detailed in Chapter 4, and validated in both data and simulation. This method is then further utilised to provide an independent crosscheck of the SM background estimations determined by the α_T search within its hadronic signal region at high b-tagged jet multiplicities.

To highlight the relative insensitivity of this method to the choice of b-tagging algorithm working point, results are presented using the CSV tagger (introduced in Section (3.3.2)) for the "Loose", "Medium" and "Tight" working points.

6.1. Concept

The dominant SM backgrounds of most SUSY searches are typically $t\bar{t}$ + jets, W + jets, Z $\rightarrow \nu\bar{\nu}$ + jets or other rare processes (e.g. Diboson, $t\bar{t}W+$ jets production in the case of hadronic searches) with neutrinos in the final state. These processes are characterised by typically having zero or two underlying b-quarks per event as shown in Table 6.1. This ultimately means that the resultant shape of the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution for these two types of event topologies will differ significantly due to the varying tagging probabilities of the different jet flavours present in the final state of these processes.

Similarly, SMS models comprising the gluino-mediated production of third generation squarks, such at the `T1tttt` and `T1bbbb` models described in the previous chapter, will contain four underlying b-quarks in its decay. Therefore the resultant shape of the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution from such a signal will be further skewed towards a higher number of b-tagged jets. As SM processes with a similarly large number of underlying b-quarks are rare, a signal indicative of natural SUSY can potentially be easily identified, via an observed excess of $n_{\mathrm{b}}^{\mathrm{reco}} = 3, \geq 4$ events with respect to the expected yields from SM processes.

Typical underlying b-quark content	Process
$= 0$	$W \rightarrow l\nu$ + jets
	Z $\rightarrow \nu\bar{\nu}$ + jets
	$Z/\gamma^* \rightarrow \mu\mu$ + jets
$= 1$	t + jets
$= 2$	$t\bar{t}$ + jets

Table 6.1: Typical underlying b-quark content of different SM processes which are common to many SUSY searches.

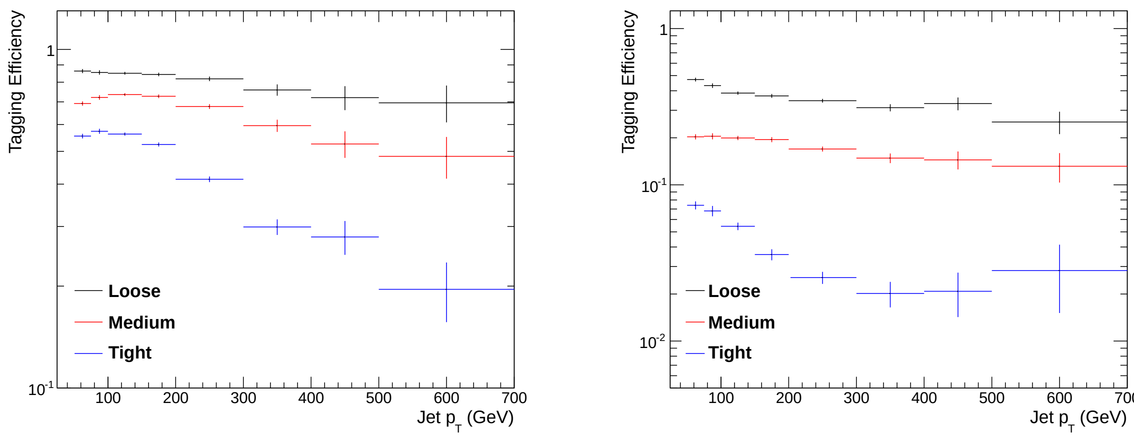
Within a supersymmetric or indeed any search for new physics, the compatibility of the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution in data with SM expectations can be tested, via the shape parameterisation of the SM background $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution, grouped in terms of these two most common underlying b-quark topologies.

6.1.1. Fitting Procedure

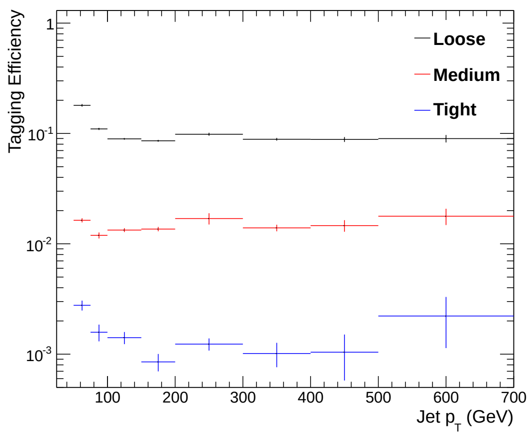
Two templates, representing processes which have an underlying b-quark content of zero or two are defined as Z0 and Z2 respectively (single top processes are a negligible background, $< 1\%$, within the α_T search to which this method is applied in the following section, and are thus incorporated within the Z2 template). SM background estimates at high $n_{\mathrm{b}}^{\mathrm{reco}}$ multiplicities can then be extrapolated from the fitting of these two template shapes in a low $n_{\mathrm{b}}^{\mathrm{reco}}$ control region (0-2) under the assumption of negligible signal contamination.

The simplest way to determine the shapes of the $n_{\mathrm{b}}^{\mathrm{reco}}$ distributions for both templates would be, after the application of the relevant event selection, to take the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution as given directly from simulation. However as discussed within Section (4.5), there are large statistical uncertainties in simulation at high n_b^{reco} multiplicities (which is the region in which we wish to use the templates to estimate the SM backgrounds). This statistical uncertainty is particularly pronounced for processes incorporated within the Z0 templates, where events with a large number reconstructed b-tagged jets stem largely from the mis-tagging of all the light-flavoured jets in the final state. Therefore to improve the statistical precision of the final background prediction at high b-tagged jet multiplicities, the formula method first introduced in Section (4.5.1) is utilised to generate the template shapes.

The template shapes of each analysis category (H_T and n_{jet} in the case of the α_T analysis) are dependant upon the jet-flavour content and their tagging efficiencies within the phase space of interest, with the tagging efficiency of a jet being a function of the jet p$_T$, the pseudo-rapidity $|\eta|$, and jet-flavour. This can be seen in Figure 6.1, where the tagging efficiency of jets identified as stemming from the hadronisation of a b, c or light quark from truth information in simulation, is shown for the three working points of the CSV tagger as a function of jet p$_T$.



(a) b-jets (b) c-jets



(c) light-jets

Figure 6.1: The b-quark (a), c-quark (b), and light-quark (c) tagging efficiency as a function of jet p_T, measured in simulation after the application of α_T analysis μ + jets control sample selection, in the region $H_T > 375$. Efficiencies are measured for the three CSV working points.

Therefore, before the template shapes are generated via the formula method, the jet p_T and η averaged tagging efficiencies of each jet flavour are determined within each individual analysis category. Additionally, the relevant jet p_T and η corrections are applied to correct the measured b-tagging rate in simulation to that of data, as specified in Section (4.5.3). These corrections propagate through to the average determined tagging efficiency for each jet flavour, consequently affecting the final Z0 and Z2 template shape of the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution, determined within each analysis category.

Using the truth-level flavour information of each of the defined Z0 and Z2 templates and the measured tagging efficiencies of each jet flavour, the template shapes are constructed from simulation via the formula method. These two shapes are then fitted to data in a low n_b^{reco} control region (0-2), by allowing the normalisation constants θ_{Z0} and θ_{Z2} of the two templates to float. The fits are performed independently within each of the defined analysis category to remove any dependence on the modelling of jet multiplicity between simulation and data. Best fit values of θ_{Z0} and θ_{Z2} are used, along with the fixed shape of each template, to extrapolate a SM background estimation within the high n_b^{reco} signal region (3,4) as shown in Figure 6.2.

In deriving the uncertainty on the background prediction the following statistical uncertainties are considered;

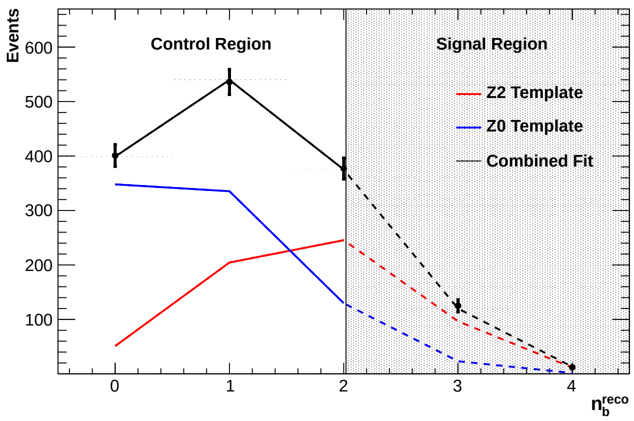


Figure 6.2: An example of a template fit with the defined Z0 (blue) and Z2 (red) templates to data within the low $n_{\mathrm{b}}^{\mathrm{reco}}$ control region (left). The shape of the two templates are fixed but the normalisations θ_{Z0} and θ_{Z2} are allowed to vary. The best fit values are then applied to extrapolate a combined background prediction from the shaded signal region (right), represented by the dashed black line. Statistical and systemic uncertainties are not shown within this figure.

Fit uncertainty: The statistical uncertainty on the normalisation factors θ_{Z0} and θ_{Z2} as determined by the fit to data.

Measured tagging efficiency uncertainty: The uncertainty of the template shapes due to the uncertainty on the measured average tagging efficiencies of each jet flavour from simulation. This uncertainty is propagated through to the template prediction for each $n_{\mathrm{b}}^{\mathrm{reco}}$ multiplicity by profiling the distribution of the θ_{Z0} and θ_{Z2} best-fit values from multiple pseudo-experiments.

For each pseudo-experiment, a Z0 and Z2 template shape is generated and fitted to data. The tagging efficiencies of each jet flavour used by the formula method to generate the template shape, are determined from a Gaussian distribution centred on the nominal measured efficiency with a width equal to its measured statistical uncertainty. The uncertainties on the nominal θ_{Z0} and θ_{Z2} normalisation factors are then determined from the value of the 68th percentile in the best fit θ_{Z0} and θ_{Z2} distributions constructed from all of the pseudo-experiments.

Formula method statistical error: The statistical uncertainties of the two templates Z0 and Z2 at each $n_{\mathrm{b}}^{\mathrm{reco}}$ multiplicity are propagated through to the overall uncertainty. This is due to the finite amount of simulated events used in the formula method to generate the template shapes.

B-tag scale factor systematic error: When this procedure is applied to data, an additional systematic error is also incorporated into the template uncertainty. This takes into account the uncertainty in correcting the tagging efficiencies measured in simulation to data as first shown in Figures 3.7 and 3.8. The systematic uncertainty for each template is determined by varying these simulation to data scale factors ($SF_{\mathrm{b,\ c,\ light}}$), up and down by their systematic uncertainties. These scale factor uncertainties are conservatively taken as fully correlated across all jet flavours [77]. The resultant relative difference due to these variations in the template shape at each $n_{\mathrm{b}}^{\mathrm{reco}}$ multiplicity of the template, is taken as the systematic uncertainty on the nominal best fit template value.

All statistical and systematic errors are added in quadrature to determine an overall template fit uncertainty at each $n_{\mathrm{b}}^{\mathrm{reco}}$ multiplicity in the control and signal regions. These are represented in all figures by a shaded grey band.

Any large excess in data is an indication that the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution is not adequately described by the SM backgrounds encapsulated by the templates. This could mean there are additional SM backgrounds that fall within the selection of the analysis that need to be considered, or that there is signal present within the data. This method relies solely on fitting to the shape of the n_b^{reco} distribution, and can in principle, be applied to any analysis where the signal hypothesis has a larger underlying b-quark spectra than the SM backgrounds.

However, in the scenario where a SUSY signal sits at a low number of underlying b-quarks, the template would be unable to discriminate between this signal and background during the fit in the control region. This will be the case unless the jet p_T distribution of the signal and background were drastically different, in which case there would anyway be many more sensitive and practical ways to establish the presence of a signal in the data than this method. Indeed the template method is only really applicable to the hypothesis that any signal resides at high $n_{\mathrm{b}}^{\mathrm{reco}}$ and that the control region $0 \leq n_{\mathrm{b}}^{\mathrm{reco}} \leq 2$ has negligible signal contamination.

6.2. Application to the α_T Search

As detailed in the previous chapter, the α_T analysis is a search for supersymmetric particles in all hadronic final states, utilising the kinematic variable α_T to suppress QCD to a negligible level. SM enriched control samples are used to estimate the background within a hadronic signal region.

The selection for the μ + jets control samples defined in Section (4.2.3) is used to demonstrate the template fitting procedure both conceptually in simulation, and also when applied in data. This is chosen, as such a selection is dominated by events stemming from the SM processes with little or no signal contamination from potential new physics. Contributions from rare SM processes with a higher underlying b-quark content (e.g. $t\bar{t}b\bar{b}$) are also found to be negligible from studies in simulation. For these reasons, there is a degree of confidence that the procedure should adequately describe the observations in data when extrapolated to the signal region.

As a departure from the α_T search strategy described in the previous section, events are categorised according to jet multiplicity categories of 3, 4 and ≥ 5 reconstructed jets per event (di-jet events are not included as there is no contribution to the high n_b^{reco} region (3,4)), in order to reduce the kinematic range of the jet p_T's within each category. Furthermore the analysis is split into just three H_T regions, for the purpose of increasing statistics within the control region,

- 275-325 GeV
- 325-375 GeV
- $>$ 375 GeV

contrary to the eight used within the α_T analysis. Templates for both underlying b-quark content hypotheses are then generated for the nine defined event categories.

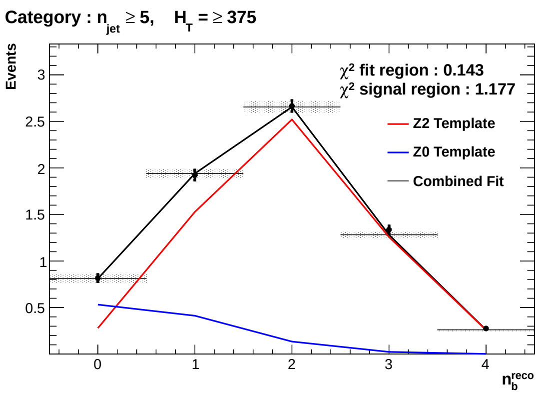
6.2.1. Proof of Principle in Simulation

This template procedure must be first demonstrated to work within simulated events free from any potential signal contamination before it can be applied to data. By combining the relevant ingredients necessary to employ the formula method, $n_{\mathrm{b}}^{\mathrm{reco}}$ shape templates are generated individually for each n_{jet} and H_T category using one half of the available simulated events for each SM process. In this case, as the template shapes are being fitted to simulation, it is *not* necessary to apply the relevant corrections of the b-tagging rates between data and simulation.

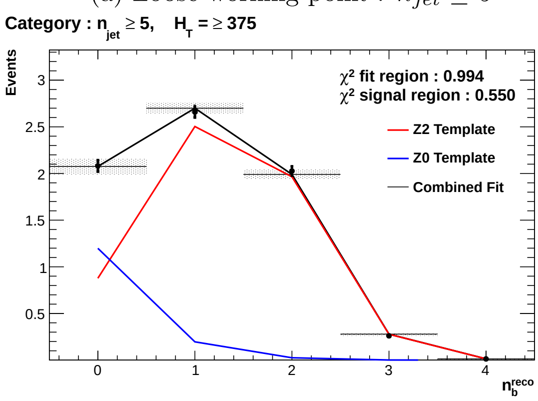
The other half of simulated events is utilised to provide a statistically independent sample from which the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution is taken directly. The two generated templates are then fit within the low $n_{\mathrm{b}}^{\mathrm{reco}}$ (0-2) control region to this pseudo-data, from which a signal region prediction is then extrapolated from the template best fit values.

The aim of this procedure is to ensure that the template fit can accurately extrapolate the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution within the defined signal region from two independent but kinematically identical samples. Furthermore, as the pseudo-data of the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution is taken directly from simulation, observation of good closure for both the initial fit of the two templates within the control region and after extrapolation to the signal region will serve as a validation of the formula method in recovering the original n_b^{reco} distribution itself.

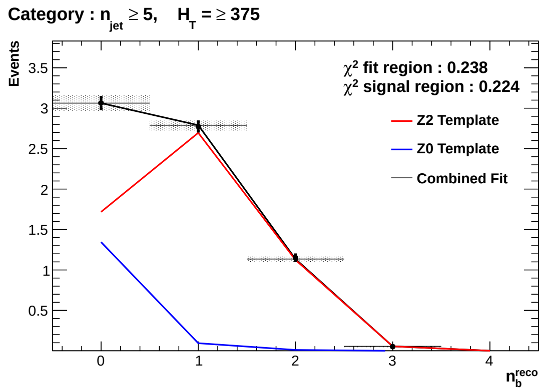
Results are presented in Figure 6.3 for each CSV working point in the $n_{\mathrm{jet}} \geq 5$ category, using the μ + jets control sample selection and the inclusive $H_T > 375$ GeV analysis bin. Additional fit results for other n_{jet} categories which show a similar level of closure can be found within Appendix D.1. The grey bands represent the statistical uncertainty of the template prediction at each $n_{\mathrm{b}}^{\mathrm{reco}}$ multiplicity derived from adding in quadrature the statistical uncertainties introduced in the previous section.



(a) Loose working point : $n_{jet} \geq 5$



(b) Medium working point : $n_{jet} \geq = 5$



(c) Tight working point : $n_{jet} \geq 5$

Figure 6.3: Results of fitting the Z = 0 and Z = 2 templates in the n_b^{reco} = 0-2 control region to yields from simulation in the μ + jets control sample for the $H_T > 375$ GeV, $n_{jet} \geq 5$ category for all CSV working points. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameters represent the goodness of fit to the control and signal region.

The extrapolated fit predictions summed over all n_{jet} multiplicities within the high n_b^{reco} signal region, are summarised for all H_T bins and working points in Table 6.2.

H_T		275-325	325-375	>375
		Loose working point		
Simulation	$n_b = 3$	786.4 ± 14.7	392.7 ± 10.3	802.2 ± 14.4
Template		789.6 ± 27.5	375.6 ± 16.6	770.1 ± 22.9
Simulation	$n_b = 4$	67.4 ± 3.9	28.2 ± 2.7	93.7 ± 4.9
Template		64.5 ± 5.9	26.4 ± 3.3	82.3 ± 5.8
		Medium working point		
Simulation	$n_b = 3$	134.2 ± 5.8	74.4 ± 4.5	161.9 ± 6.3
Template		129.9 ± 6.6	68.3 ± 4.8	159.9 ± 7.7
Simulation	$n_b = 4$	1.5 ± 0.4	0.6 ± 0.4	3.1 ± 0.6
Template		1.7 ± 0.3	0.9 ± 0.3	3.9 ± 0.6
		Tight working point		
Simulation	$n_b = 3$	28.1 ± 2.7	13.9 ± 1.9	29.2 ± 2.7
Template		25.9 ± 2.0	12.2 ± 1.5	28.3 ± 2.4
Simulation	$n_b = 4$	0.5 ± 0.4	-	0.2 ± 0.2
Template		0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.1

Table 6.2: Summary of the fit predictions in the n_b^{reco} signal region after combination of the $n_{jet} = 3, = 4, \geq 5$ categories compared against yields taken directly from simulation. The fit predictions are extrapolated from a $n_b^{reco} = 0$, 1, 2 control region and simulation yields are normalised to an integrated luminosity of 10 fb^{-1}. The uncertainties quoted on the template yields are purely statistical.

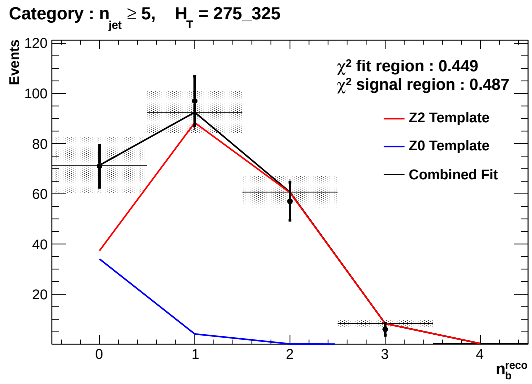
The pull distributions for all the fits performed can be found in Appendix D.2, and are compatible with a mean of zero and standard deviation of one, showing no obvious bias to the fitting procedure. Each of the fits performed show good compatibility between the template shapes and data from simulation within the defined control region, with additional good overall agreement also observed for extrapolation to the signal region as shown in Table 6.2. This validates both the formula method used in the generation of the template shapes as well as the method of predicting the SM background in the high $n_{\mathrm{b}}^{\mathrm{reco}}$ signal region.

The application of this method to the same selection in a data control sample is now used to demonstrate necessary control over the efficiency and mis-tagging rates when b-tagging scale factors are applied, and to test the assumption of no signal contamination with the μ + jets control sample.

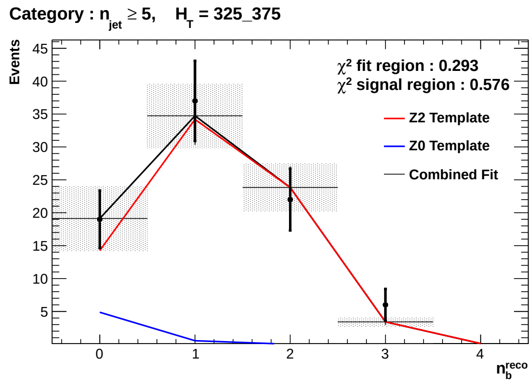
6.2.2. Results in a Data Control Sample

The procedure is now applied to the 2012 8 TeV dataset in the μ + jets control sample, to establish the validity of this method in data. The relevant data to simulation b-tagging scale factors are applied to produce corrected values of the efficiency and mis-tagging rates within each analysis category [77].

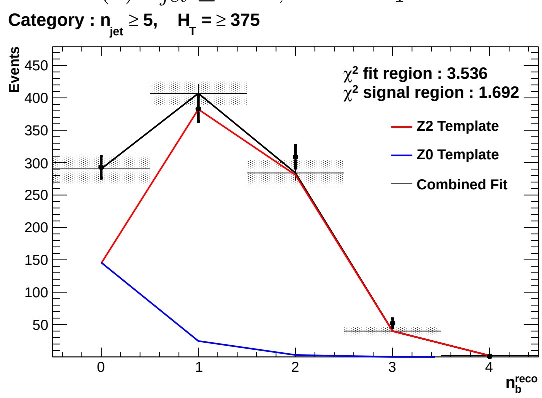
Figure 6.4 shows the results of the templates derived from simulation to each of the three defined H_T bins, in the $n_{jet} \geq 5$ category for the medium working point CSV tagger (the same working point used within the α_T analysis). Grey bands represent the previously detailed statistical uncertainty of the fit combined in quadrature with the systematic uncertainties of varying up and down the simulation to data scale factors by their b-tag scale factor systematic uncertainties. Additional fit results for other jet multiplicities are found in Appendix D.3.



(a) $n_{jet} \geq 5$, $275 < H_T < 325$



(b) $n_{jet} \geq = 5$, $325 < H_T < 375$



(c) $n_{jet} \geq 5$, $H_T > 375$

Figure 6.4: Results of fitting the Z = 0 and Z = 2 templates in the n_b^{reco} = 0-2 control region to data from the μ + jets control sample, for the CSV medium working point, with $n_{jet} \geq 5$ in each H_T category. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameters represent the goodness of fit to the control and signal region.

The numerical results and extrapolation to the n_b^{reco} =3, 4 bins for all H_T and working points, is shown in Table 6.3.

H_T		275-325	325-375	>375
		Loose working point		
Data	$n_b = 3$	838	394	717
Template		871.8 ± 46.9	369.9 ± 23.7	678.5 ± 42.5
Data	$n_b = 4$	81	43	81
Template		79.4 ± 9.9	32.9 ± 4.2	74.4 ± 10.0
		Medium working point		
Data	$n_b = 3$	137	79	152
Template		132.6 ± 9.3	69.8 ± 5.4	133.1 ± 10.8
Data	$n_b = 4$	1	1	3
Template		1.9 ± 0.4	0.9 ± 0.3	3.2 ± 0.6
		Tight working point		
Data	$n_b = 3$	24	15	25
Template		22.3 ± 1.9	12.1 ± 1.2	20.3 ± 2.4
Data	$n_b = 4$	0	0	1
Template		0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.1

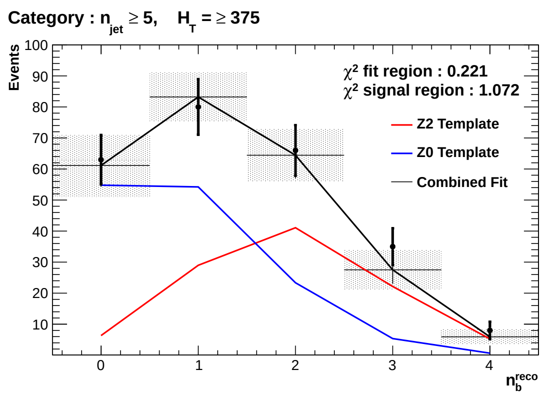
Table 6.3: Summary of the fit predictions in the n_b^{reco} signal region of the μ + jets control sample, after combination of the $n_{jet} = 3, = 4, \geq 5$ categories. The predictions are extrapolated from a $n_b^{reco} = 0$, 1, 2 control region using 11.4 fb^{-1} of $\sqrt{s} = 8$TeV data. The uncertainties quoted on the template yields are a combination of statistical and systematic uncertainties.

When this method is applied to the μ + jets control sample, it is expected that good agreement would be observed between the template predictions and observation in the absence of signal contamination. The good compatibility for all working points as shown in the table, demonstrate that this is the case and that the method is able to accurately predict the background yields. However the assumption of negligible signal contamination can no longer made when applied to the hadronic signal region of the α_T search, where agreement between estimated backgrounds and observations in data is now not necessarily expected.

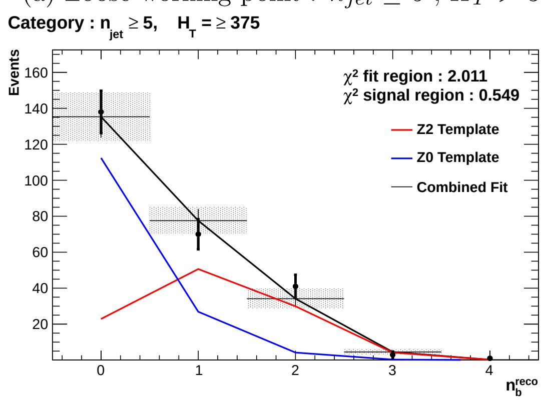
6.2.3. Application to the α_T Hadronic Search Region

As an accompaniment to the background estimation methods outlined in the α_T search, the b-tag template method offers a complementary way of testing the SM only background hypothesis within the hadronic signal region of the search. In the presence of a natural SUSY signature mediated by a light gluino and containing four underlying $\widetilde{b}$ or $\widetilde{t}$ squarks, which subsequently decay to t or b quarks, the number of reconstructed $n_{\mathrm{b}}^{\mathrm{reco}} = 3, \geq 4$ events will be enhanced.

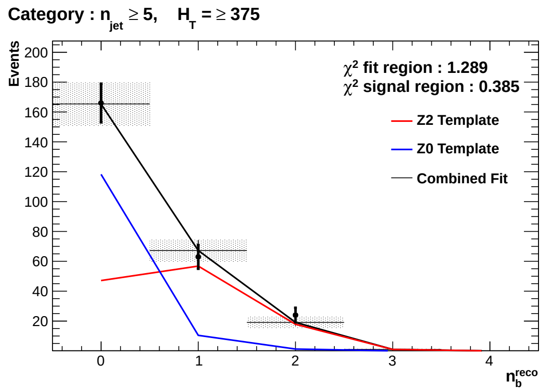
Figure 6.5 shows the the results of the template shapes derived from simulation and fitted to data for each of the three CSV working points, in the $n_{jet} \geq 5$, $H_T > 375$ GeV category. Grey bands represent the statistical uncertainty of the fit combined in quadrature with the systematic uncertainties of varying the simulation to data scale factors up and down by their measured systematic uncertainties. Additional fit results for other jet multiplicities are found in Appendix D.4.



(a) Loose working point : $n_{jet} \geq 5$, $H_T > 375$



(b) Medium working point : $n_{jet} \geq 5$, $H_T > 375$



(c) Tight working point : $n_{jet} \geq 5$, $H_T > 375$

Figure 6.5: Results of fitting the Z = 0 and Z = 2 templates in the n_b^{reco} = 0-2 control region to data from the hadronic signal selection, in the $n_{jet} \geq 5$ and $H_T > 375$ category for all CSV working points. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameters represent the goodness of fit to the control and signal region.

H_T		275-325	325-375	>375
		Loose working point		
Data	$n_b = 3$	198	85	126
Template		207.1 ± 28.7	103.4 ± 12.2	124.98 ± 14.4
Data	$n_b = 4$	15	9	16
Template		15.9 ± 5.4	8.05 ± 2.1	13.1 ± 3.2
		Medium working point		
Data		33	16	14
Template	$n_b = 3$	25.4 ± 4.0	12.7 ± 2.2	19.9 ± 2.9
α_T ML Fit		$33.9^{+5.7}_{-4.3}$	$16.3^{+1.9}_{-2.0}$	$17.5^{+1.4}_{-1.4}$
Data		1	0	2
Template	$n_b = 4$	0.3 ± 0.2	0.3 ± 0.1	0.5 ± 0.2
α_T ML Fit		$0.9^{+0.4}_{-0.7}$	$0.3^{+0.2}_{-0.2}$	$0.6^{+0.3}_{-0.3}$
		Tight working point		
Data	$n_b = 3$	5	2	0
Template		4.03 ± 0.8	2.4 ± 0.5	3.1 ± 0.6
Data	$n_b = 4$	1	0	0
Template		0.1 ± 0.1	0.1 ± 0.1	0.0 ± 0.1

Table 6.4: Summary of the fit predictions in the n_b^{reco} signal region of the α_T hadronic signal selection, after combination of the $n_{jet} = 3, = 4, \geq 5$ categories. The predictions are extrapolated from a n_b^{reco} = 0, 1, 2 control region using 11.7 fb^{-1} of $\sqrt{s} = 8$TeV data. The uncertainties quoted on the template yields are a combination of statistical and systematic uncertainties.

The numerical results and extrapolation to the n_b^{reco} =3, 4 bins for all H_T and working points are shown in Table 6.4. Included within the table are the combined SM background predictions as determined by the maximum likelihood fit for both jet multiplicity categories of the α_T analysis using the CSVM tagger. No notable discrepancy is found in any of the three CSV working points between the data and the background expectations as determined by this method. The template predictions within the hadronic signal region are additionally found to be statistically compatible with the background predictions determined by the α_T maximum likelihood fit (introduced in Table 5.2).

6.3. Summary

A SUSY signature such as one from gluino-induced third-generation squark production, would result in a final state with an underlying b-quark content greater than two. In order to be able to discriminate such signatures from the SM background, templates are generated based on a parameterisation of SM processes, where the underlying b-quarks per event is typically zero or two. These templates are then fit to data in a low n_b^{reco} (0-2) control region in order to extrapolate a prediction within a high n_b^{reco} (3-4) signal region. This approach is built upon the assumptions that the defined control region is almost entirely free of any possible signal contamination from possible signal topologies with a small number of b quarks in the final state.

The method was demonstrated both in simulation and also in data, using the SM enriched μ + jets selection from the α_T search. This was conducted to prove conceptually and experimentally that the method is valid and that there is adequate control over the measurement of the efficiency of each jet flavour for all working points of the CSV tagger. Additionally this method was further applied to the hadronic signal region of the α_T analysis, where good agreement is observed between the SM background predictions from the template method, observations in data and also the background estimation procedure of the α_T analysis.

Chapter 7.

Conclusions

A search for supersymmetry has been presented based on a data sample of pp collisions collected at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 11.7 ± 0.5 fb^{-1}. Final states with two or more jets and significant missing transverse energy, a typical final state topology of R-parity conserving SUSY models have been analysed. The α_T variable is utilised as the main discriminator between balanced multi-jet backgrounds and those with real missing transverse energy.

Within the search presented, Standard Model (SM) backgrounds are estimated from a simultaneous binned likelihood fit to a hadronic signal selection as well as three SM process enriched control samples. The search is split into total transverse hadronic energy (H_T), jets identified as originating for a b-quark ($n_{\mathrm{b}}^{\mathrm{reco}}$), and jet multiplicity (n_{jet}) categories to improve sensitivity to a range of possible supersymmetric final states. Systematic errors due to theory, detector effects and simulation deficiencies are quantified through the use of data driven closure tests and accounted for in the final interpretation. Observations in data are found to be compatible with a SM only hypothesis.

In the absence of a signal like excess the analysis is further interpreted in a set of Simplified Model Spectra (SMS) models, representing a set of model independent decay topologies parameterised only by the production process and the masses of their parent sparticle and Lightest Supersymmetric Partner (LSP). In models mediated by gluino pair production and containing a large mass difference between the gluino and LSP, exclusion limits of the gluino mass are set in the range 950-1125 GeV. For SMS models describing direct squark pair production, first or second generation squarks are excluded up to 775 GeV, with direct bottom squarks production excluded up to masses of 600 GeV.

In the case of gluino mediated third generation signatures containing many jets originating from b-quarks in the final state, mass limits are set in the range of 975-1125 GeV for large mass splittings between the gluino and the LSP. The experimental sensitivity to these models is attributed to the $n_{\mathrm{b}}^{\mathrm{reco}}$ categorisation of the analysis, where the signal-to-background is enhanced within the phase space of the search at high $n_{\mathrm{b}}^{\mathrm{reco}}$.

Furthermore, a measurement of the performance of the Level-1 trigger for jets and jet energy sum quantities has also been presented. These studies quantify any change in Level-1 performance after the introduction of a 5 GeV jet seed threshold into the jet clustering algorithm. No significant change in single jet trigger efficiencies is observed and good performance is observed for a range of Level-1 jet energy sum quantities.

This change was introduced to facilitate a reduction in the rate of events triggered by energy deposits due to soft non-collimated jets from secondary interactions, and which are not of interest to physics analyses. This was necessary to ensure, that trigger thresholds can be maintained at low values in the presence of an ever increasing number of bunch crossings per proton interaction. In the context of SUSY, this is a necessity to keep CMS sensitive to types of compressed spectra signatures characterised by low transverse energy jets and small missing transverse energy signatures.

Finally, an approach that uses a template fit method to the $n_{\mathrm{b}}^{\mathrm{reco}}$ distribution of SM processes within a supersymmetric search is introduced and then validated in simulation and data. The approach can be used to identify any excess in data arising from gluino mediated third generation supersymmetric signatures. It is utilised within this thesis as a crosscheck to the α_T background prediction at high b-tagged jet multiplicities. This method is found to give a SM background estimation that is in good agreement with the α_T search within the hadronic signal region.

The continued absence of a supersymmetric signal in the α_T search or other analyses at CMS [101][102][103], puts pressure on the parameter space in which SUSY can reside. Indeed the smoking gun that many theorists and experimentalists hoped to see at the LHC has not materialised. Instead identifying a SUSY signal may now only result from many years of data taking and the incorporation of increasingly advanced analysis techniques. An unenviable task considering the difficulties of not knowing where SUSY may reside, but perhaps solace can be taken in remembering that nothing worth having ever comes easy.

Appendix A.

Miscellaneous

A.1. Jet Identification Criteria

For Calo jets the following identification criteria were applied:

Loose CaloJet Id	
Variable	Definition
$f_{HPD} < 0.98$	Fraction of jet energy contributed from "hottest" HPD, which rejects HCAL noise.
$f_{EM} > 0.01$	Noise from the HCAL is further suppressed by requiring a minimal electromagnetic component to the jet f_{EM}.
$N^{90}_{hits} \geq 2$	Jets that have > 90% of its energy from a single channel are rejected, to serve as a safety net that catches jets arising from undiagnosed noisy channels.

Table A.1: Criteria for a reconstructed jet to pass the loose calorimeter jet id.

For PF jets the following identification criteria were applied:

Loose PF jet Id	
Variable	Definition
nfhJet < 0.99	Fraction of jet composed of neutral hadrons. HCAL noise tends to populate high values of neutral hadron fraction.
nemfJet < 0.99	Fraction of jet composed of neutral electromagnetic energy. ECAL noise tends to populate high values of neutral EM fraction.
nmultiJet > 1	Number of constituents that jet is composed from.
chfJet > 0	Fraction of jet composed of charged hadrons.
cmultiJet > 0	Number of charged particles that compose jet.
cemfJet < 0.99	Fraction of jet composed of charged electromagnetic energy.

Table A.2: Criteria for a reconstructed jet to pass the loose PF jet id.

A.2. Primary Vertices

The pile-up per event is defined by the number of 'good' reconstructed primary vertices in the event, with each vertex satisfying the following requirements:

Good primary vertex requirement	
Variable	Definition
$N_{dof} > 4$	The number of degree of freedom, from the vertex fit to compute the best estimate of the vertex parameters.
$\|\Delta z_{vtx}\| < 24$cm	The distance, $\|\Delta z_{vex}\|$, to the position of the closest HLT primary vertex.
$\rho < 2$cm	The perpendicular distance of track position to the beam spot.

Table A.3: Criteria for a vertex in an event to be classified as a 'good' reconstructed primary vertex.

Appendix B.

Additional Material for L1 Jet Performance Studies

B.1. Jet Matching Efficiencies

The single jet turn-on curves are derived from events independent of whether the leading jet in an event is matched to a Level 1 jet using ΔR matching detailed in Section (3.4.3). These turn-ons are produced from events which are not triggered on jet quantities and therefore it is not guaranteed that the lead jet of an event will be seeded by a Level 1 jet. Figure B.1 shows the particular matching efficiency of a lead jet to a L1 jet.

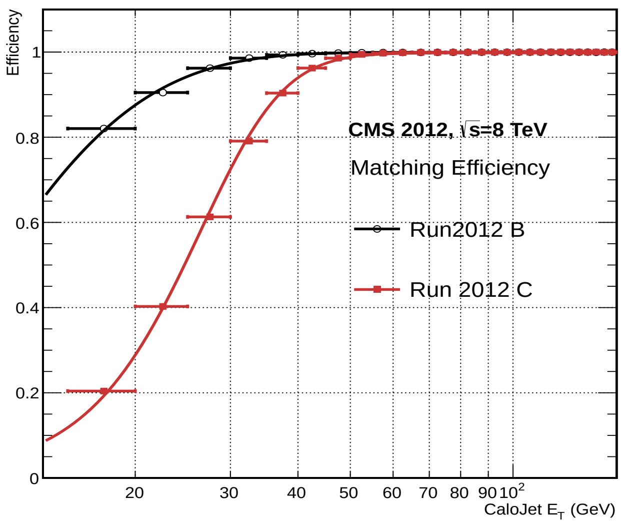


Figure B.1: Leading jet matching efficiency as a function of the offline CaloJet E_T, measured in an isolated muon triggered dataset in the 2012B and 2012C run periods.

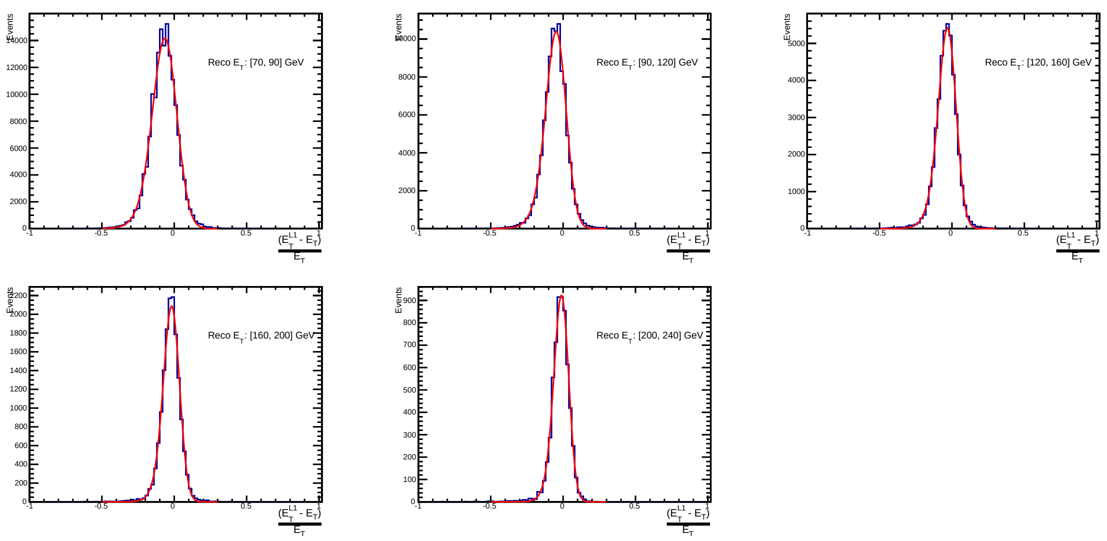
Run Period	μ	σ
2012B	6.62 $\pm$ 0.01	0.79 $\pm$ 0.03
2012C	19.51 $\pm$ 0.03	7.14 $\pm$ 0.02

Table B.1: Results of a cumulative EMG function fit to the turn-on curves for the matching efficiency of the leading jet in an event to a Level-1 jet in run 2012C and 2012B data, measured in an isolated muon triggered sample. The turn-on point, μ, and resolution, σ, are measured with respect to offline Calo Jet E_T.

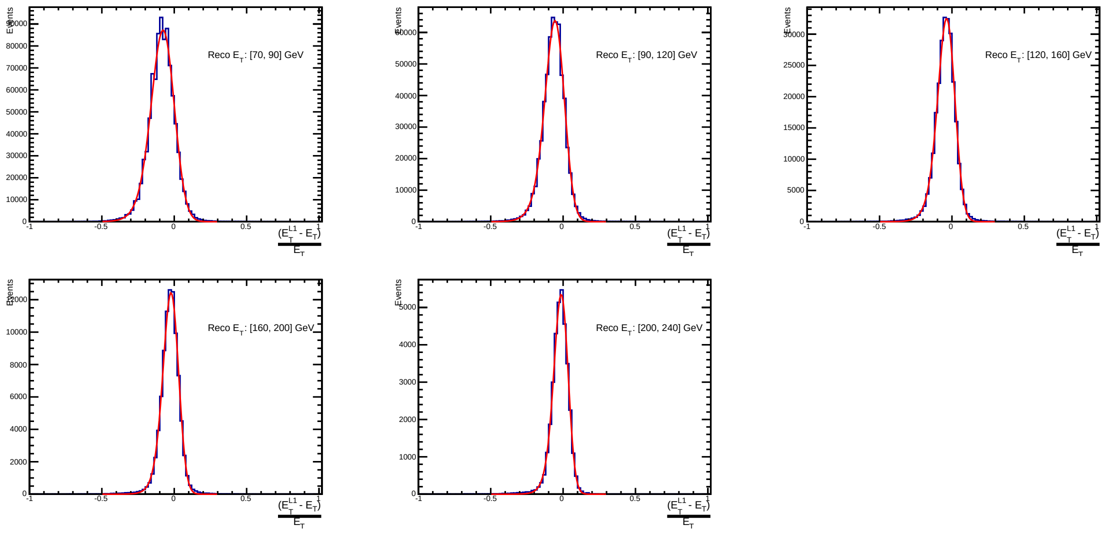
It can be seen that the turn-on occurs at a lower E_T during the 2012B run period. The seed threshold requirement of a 5 GeV jet seed in run 2012C result in more events in which the lead offline jet does not have an associated L1 jet. This can be attributed to events with soft non-collimated jets in which the energy deposits are not centralised in a calorimeter region. However, for larger jet E_T thresholds typical of those used by physics analyses, 100% efficiency is observed, and therefore this effect has no impact to overall physics performance.

A fit of an EMG function to the matching efficiencies find mean, μ, values of 6.62 GeV and 19.51 GeV for Run 2012B and 2012C respectively and is shown in Table B.1.

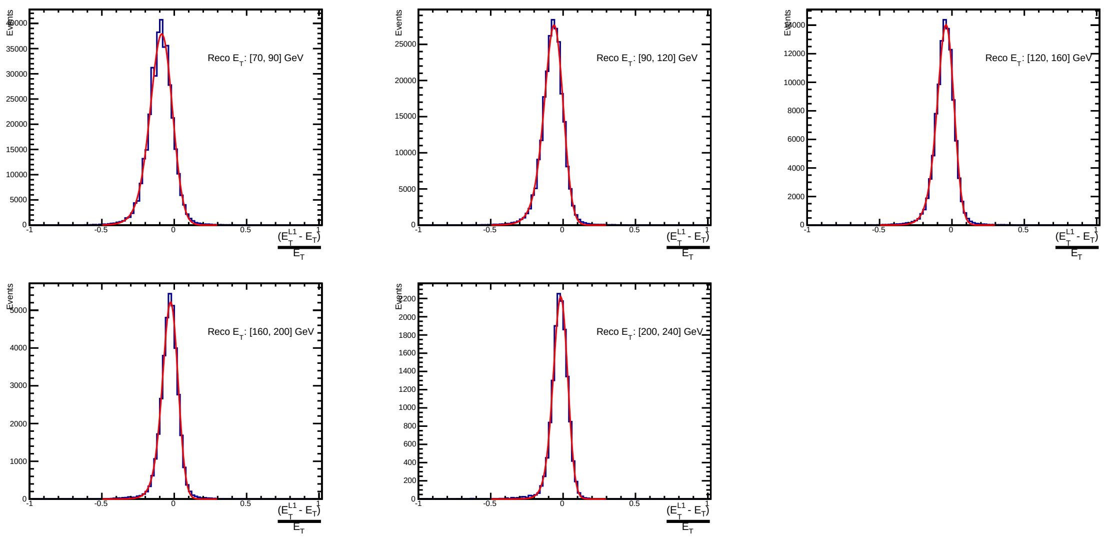
B.2. Leading Jet Energy Resolution



(a)

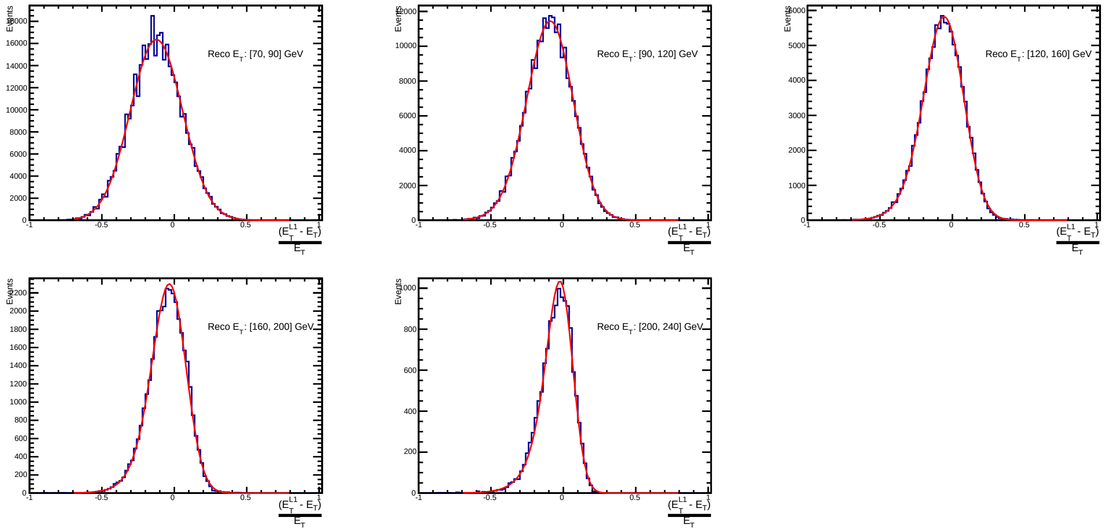


(b)

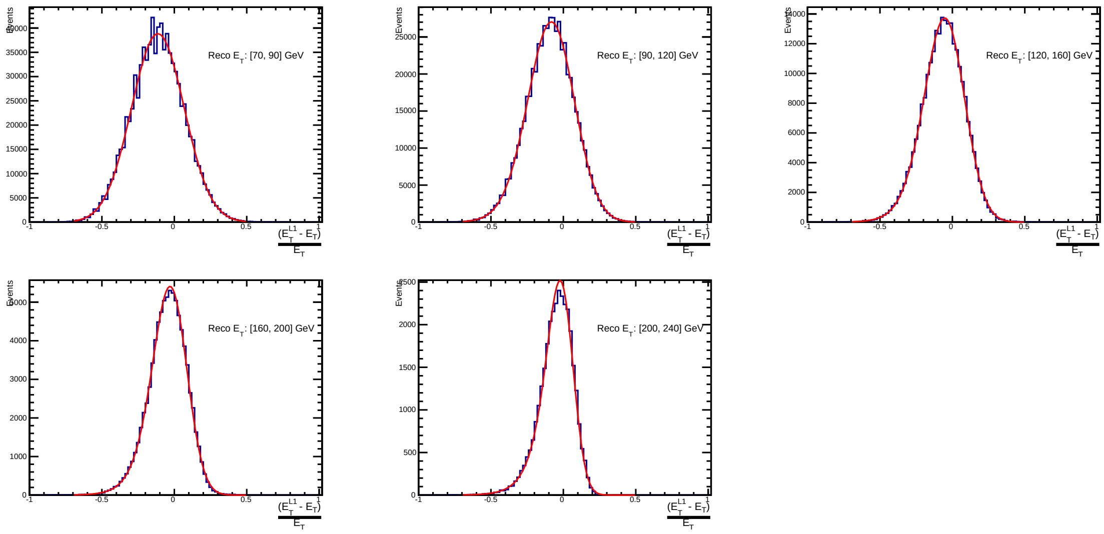


(c)

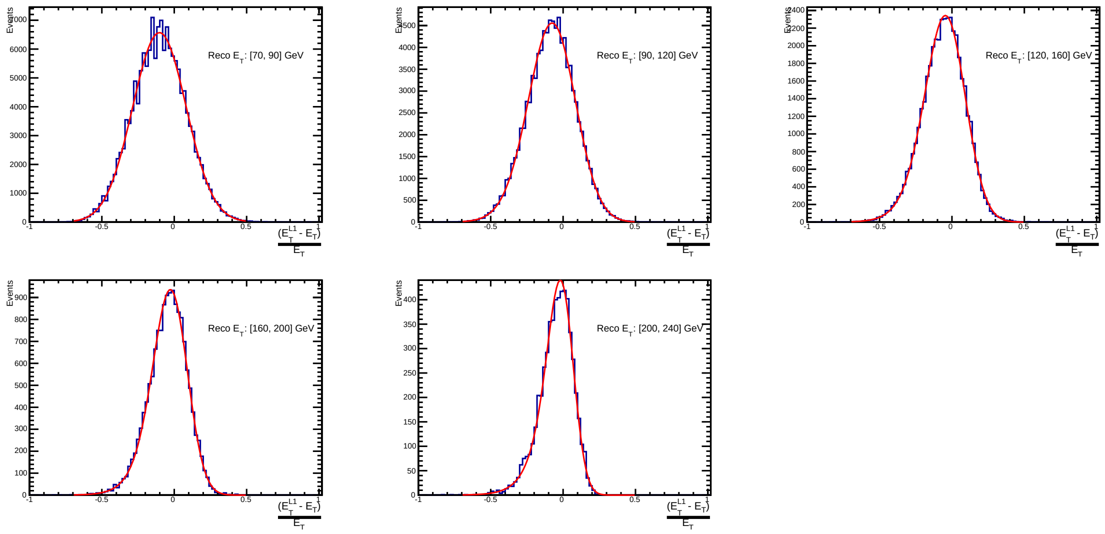
Figure B.2: Resolution plots of the leading offline jet Calo E_T measured as a function of $\frac{(\text{L1 E}_T - \text{Offline E}_T)}{\text{Offline E}_T}$ for (a) low, (b) medium, and (c) high pile-up conditions as defined in Section (3.4.4).


Events
Reco E_T: [70, 90] GeV
Reco E_T: [90, 120] GeV
Reco E_T: [120, 160] GeV
Reco E_T: [160, 200] GeV
Reco E_T: [200, 240] GeV
(E_T^L1 - E_T) / E_T

(a)


Events
Reco E_T: [70, 90] GeV
Reco E_T: [90, 120] GeV
Reco E_T: [120, 160] GeV
Reco E_T: [160, 200] GeV
Reco E_T: [200, 240] GeV
(E_T^L1 - E_T) / E_T

(b)



(c)

Figure B.3: Resolution plots of the leading offline jet PF E_T measured as a function of $\frac{(\text{L1 E}_T - \text{Offline E}_T)}{\text{Offline E}_T}$ for (a) low, (b) medium, and (c) high pile-up conditions as defined in Section (3.4.4).

B.3. Resolution for Energy Sum Quantities

The following plots show the resolution parameters for energy sum quantities as a function of the quantity (q) itself. In this case, the μ, σ and λ fit values to an EMG function defined by Equation (3.3) for each of the individual $\frac{(\text{L1 q} - \text{Offline q})}{\text{Offline q}}$ distributions, in bins of the quantity (q) is displayed.

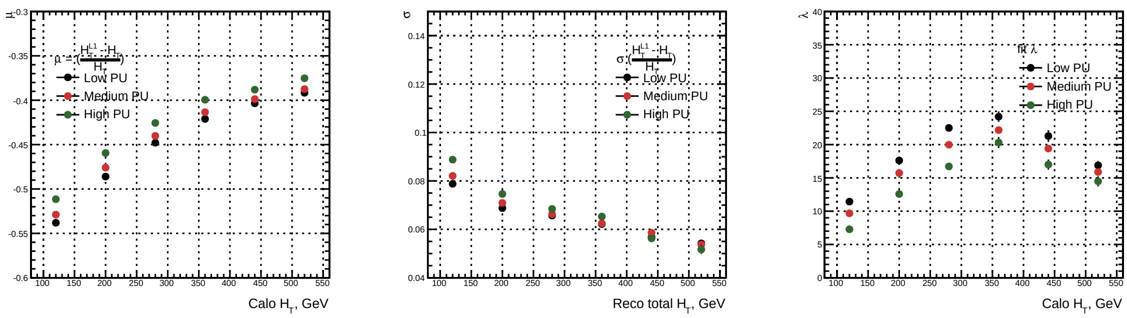


Figure B.4: H_T resolution parameters in bins of Calo H_T measured for the defined low, medium and high pile-up conditions. Shown are the mean μ (left), resolution σ (middle) and λ (right) fit values to an EMG function for the $\frac{(L1 H_T - H_T)}{H_T}$ distributions.

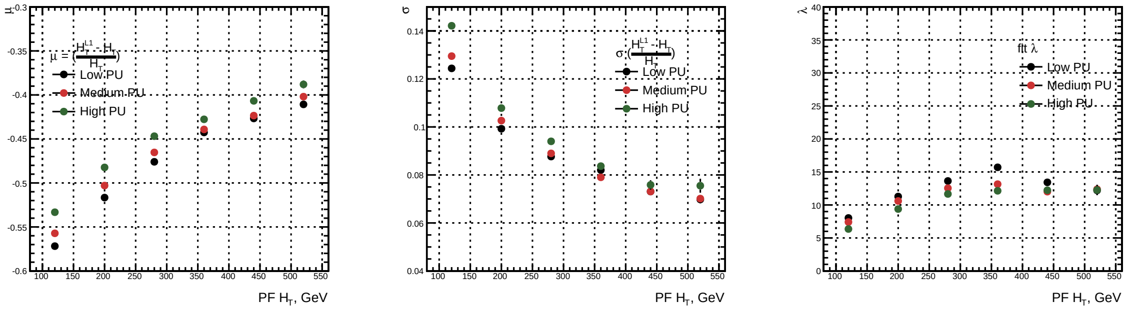


Figure B.5: H_T resolution parameters in bins of PF H_T measured for the defined low, medium and high pile-up conditions. Shown are the mean μ (left), resolution σ (middle) and λ (right) fit values to an EMG function for the $\frac{(L1H_T - H_T)}{H_T}$ distributions.

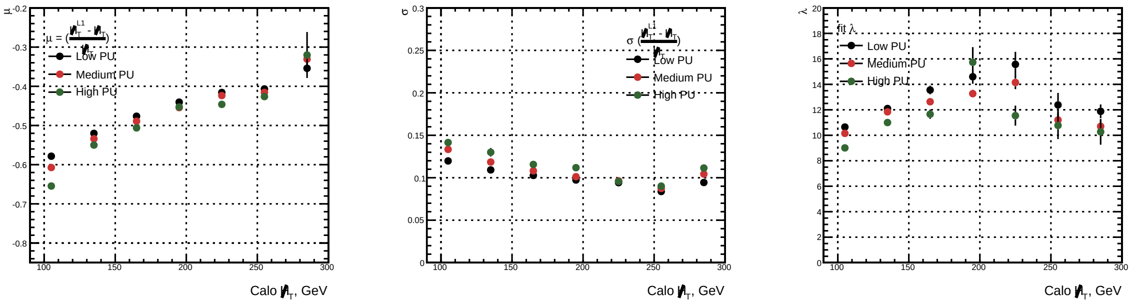


Figure B.6: $\not{H}_T$ resolution parameters in bins of Calo $\not{H}_T$ measured for the defined low, medium and high pile-up conditions. Shown are the mean μ (left), resolution σ (middle) and λ (right) fit values to an EMG function for the $\frac{(L1\not{H}_T - \not{H}_T)}{\not{H}_T}$ distributions.

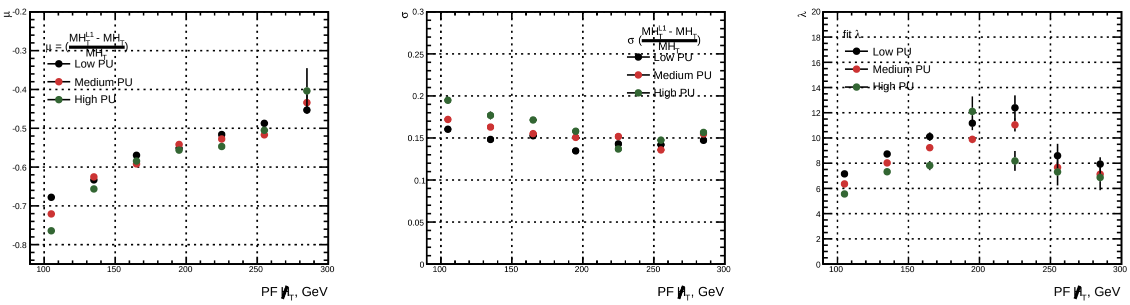
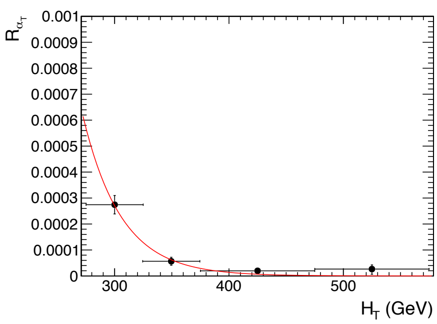


Figure B.7: $\not{H}_T$ resolution parameters in bins of PF $\not{H}_T$ measured for the defined low, medium and high pile-up conditions. Shown are the mean μ (left), resolution σ (middle) and λ (right) fit values to an EMG function for the $\frac{(L1\not{H}_T - \not{H}_T)}{\not{H}_T}$ distributions.

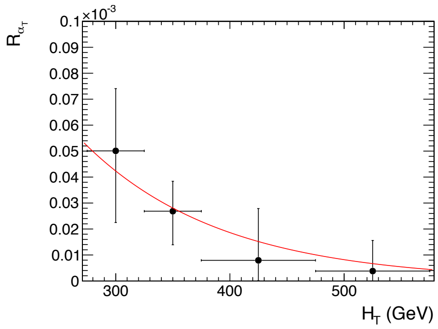
Appendix C.

Additional Material on Background Estimation Methods

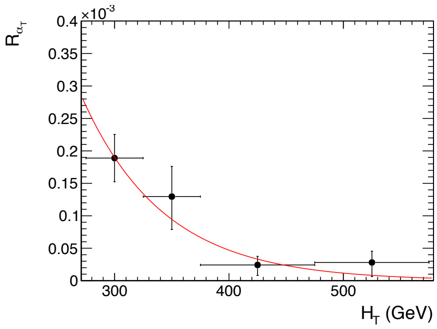
C.1. Determination of k_{QCD}



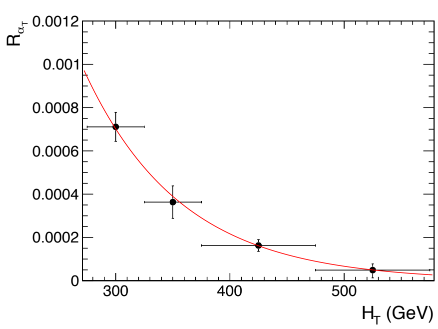
(a) Side-band region B.



(b) Side-band region C_1.



(c) Side-band region C_2.

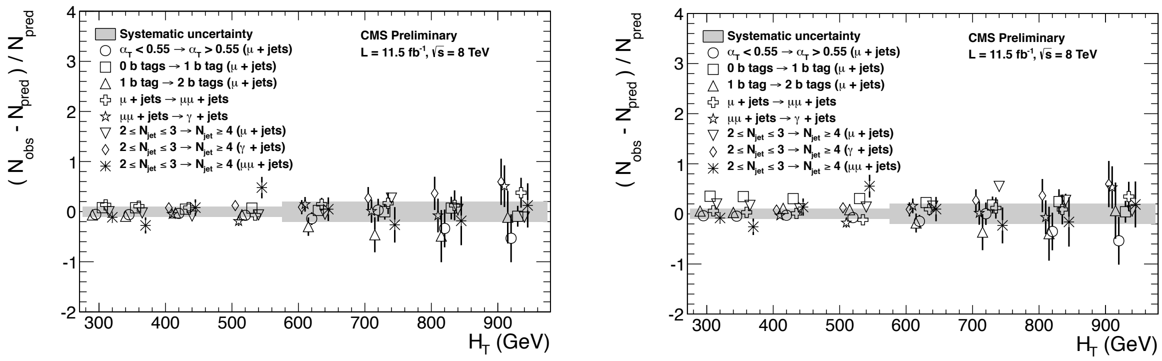


(d) Side-band region C_3.

Figure C.1: $R_{\alpha_T}(H_T)$ and exponential fits for each of the data sideband regions. Fit is conducted between the H_T region $275 < H_T < 575$.

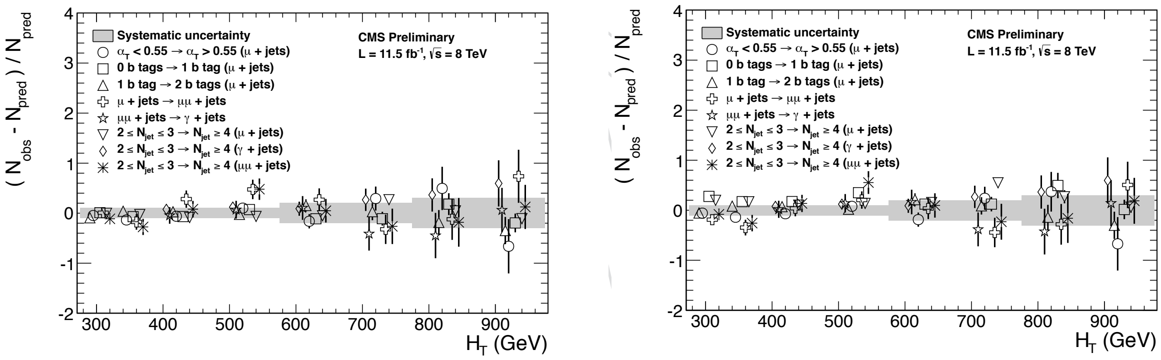
C.2. Effect of Varying Background Cross-sections on Closure Tests

Closure tests with cross section variations of +20% and -20% applied to W + jets and $t\bar{t}$ processes respectively.



(a) (b)

Figure C.2: Sets of closure tests (open symbols) overlaid on top of the systematic uncertainty used for each of the five H_T regions (shaded bands) in the $2 \leq n_{jet} \leq 3$ jet multiplicity category for nominal and varied cross-sections; (a) Nominal and (b) Varied $\pm 20\%$.



(a) (b)

Figure C.3: Sets of closure tests (open symbols) overlaid on top of the systematic uncertainty used for each of the five H_T regions (shaded bands) in the $n_{jet} \geq 4$ jet multiplicity category for nominal and varied cross-sections; (a) Nominal (b) Varied $\pm 20\%$.

		H_T (GeV)			
n_b^{reco}	Cross Section	275–325	325–375	375–475	475–575
0	Nominal	0.303 ± 0.010	0.258 ± 0.007	0.192 ± 0.003	0.148 ± 0.004
0	Varied	0.300 ± 0.010	0.256 ± 0.007	0.191 ± 0.003	0.147 ± 0.004
1	Nominal	0.294 ± 0.005	0.246 ± 0.004	0.189 ± 0.003	0.139 ± 0.003
1	Varied	0.295 ± 0.006	0.248 ± 0.004	0.191 ± 0.003	0.140 ± 0.003
2	Nominal	0.208 ± 0.003	0.183 ± 0.004	0.145 ± 0.003	0.123 ± 0.004
2	Varied	0.211 ± 0.004	0.185 ± 0.004	0.147 ± 0.003	0.124 ± 0.004
3	Nominal	0.214 ± 0.005	0.202 ± 0.007	0.159 ± 0.006	0.140 ± 0.007
3	Varied	0.215 ± 0.005	0.203 ± 0.007	0.159 ± 0.006	0.140 ± 0.007
≥4	Nominal	0.220 ± 0.015	0.245 ± 0.035	0.119 ± 0.009	-
≥4	Varied	0.220 ± 0.015	0.245 ± 0.035	0.119 ± 0.009	-
n_b^{reco}	Cross Section	575–675	675–775	775–875	875–∞
0	Nominal	0.119 ± 0.004	0.098 ± 0.005	0.077 ± 0.006	0.049 ± 0.005
0	Varied	0.120 ± 0.005	0.098 ± 0.006	0.077 ± 0.007	0.049 ± 0.005
1	Nominal	0.115 ± 0.004	0.093 ± 0.005	0.075 ± 0.007	0.063 ± 0.006
1	Varied	0.116 ± 0.004	0.098 ± 0.005	0.081 ± 0.007	0.065 ± 0.006
2	Nominal	0.096 ± 0.005	0.070 ± 0.006	0.051 ± 0.007	0.063 ± 0.008
2	Varied	0.098 ± 0.005	0.073 ± 0.006	0.053 ± 0.007	0.064 ± 0.008
3	Nominal	0.114 ± 0.009	0.065 ± 0.007	0.070 ± 0.017	0.092 ± 0.020
3	Varied	0.114 ± 0.009	0.066 ± 0.007	0.070 ± 0.016	0.093 ± 0.020

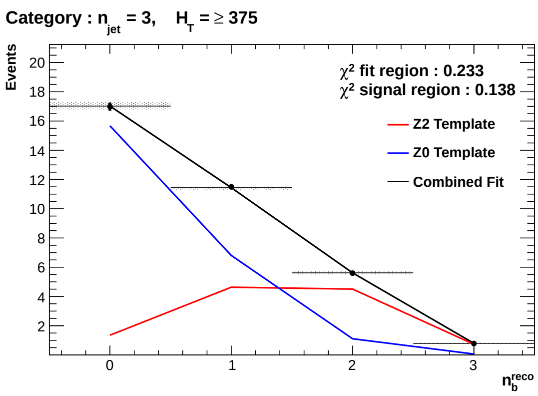
Table C.1: Translation factors constructed from the μ + jets control sample and signal selection MC, to predict yields for the W + jets and $t\bar{t}$ back-grounds in the signal region with (a) NNLO cross sections corrected by k-factors determined from a data sideband see Section (4.4), marked as Nominal, and (b) the same cross sections but with those for W + jets and $t\bar{t}$ varied up and down by 20%, respectively, marked as Varied. No requirement is placed on the jet multiplicity of events within this table.

Appendix D.

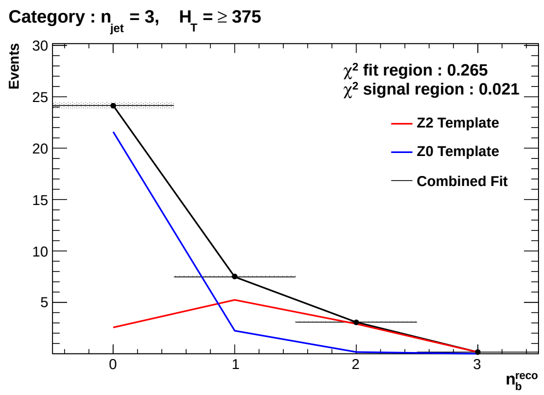
Additional Material for B-tag Template Method

D.1. Templates Fits in Simulation

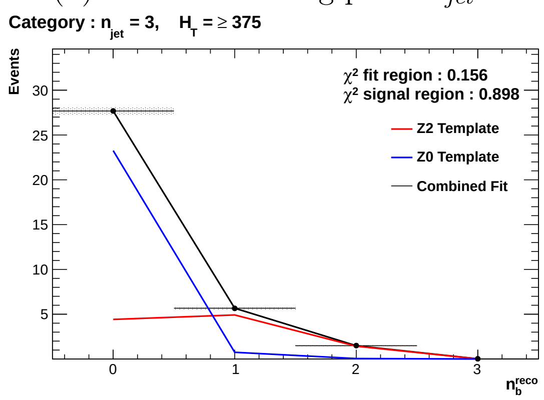
The result of template fits for the three CSV working points in the $n_{jet} = 3$, $H_T > 375$ category:



(a) Loose working point $n_{jet} = 3$



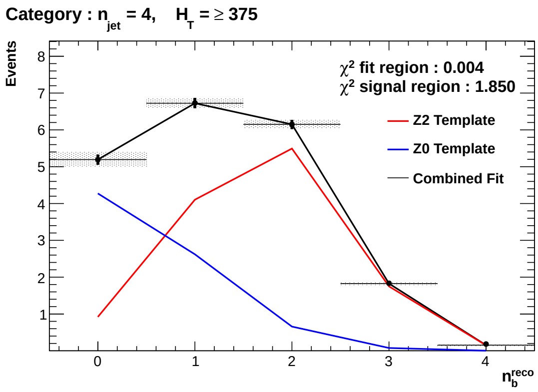
(b) Medium working point $n_{jet} = 3$



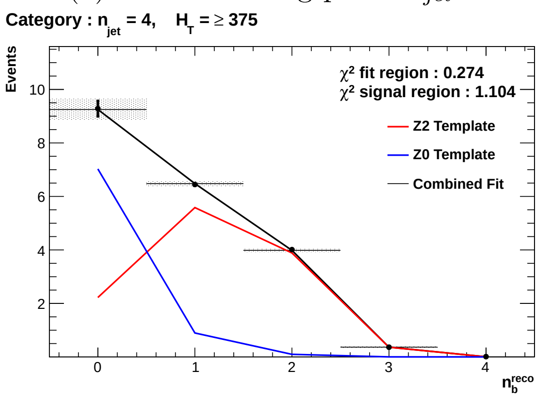
(c) Tight working point $n_{jet} = 3$

Figure D.1: Results of fitting the Z = 0 and Z = 2 templates in the n_b^{reco} = 0-2 control region to yields from simulation in the μ + jets control sample for the $H_T > 375$ GeV, $n_{jet} = 3$ category. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameter represent the goodness of fit to the control and signal region.

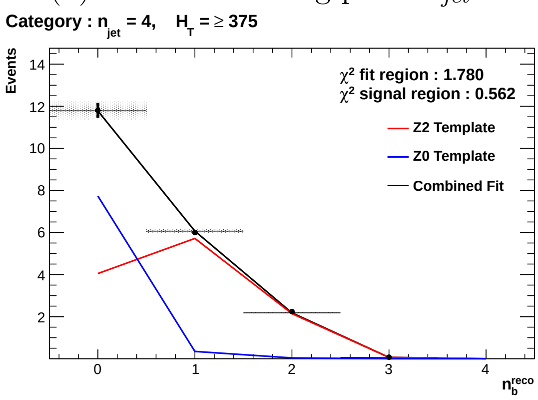
Template fits for the three CSV working points in the $n_{jet} = 4$, $H_T > 375$ category:



(a) Loose working point $n_{jet} = 4$



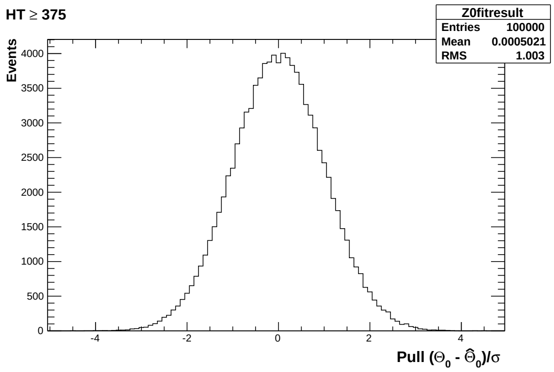
(b) Medium working point $n_{jet} = 4$



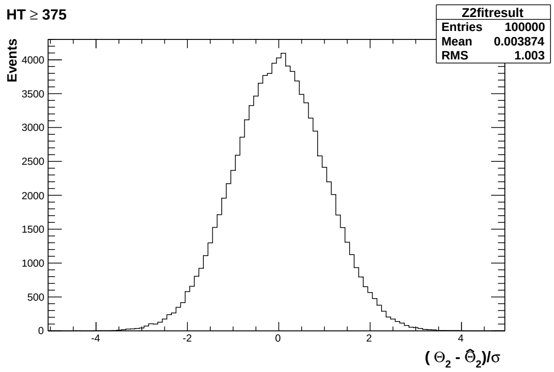
(c) Tight working point $n_{jet} = 4$

Figure D.2: Results of fitting the Z = 0 and Z = 2 templates in the $n_b^{reco} = 0\text{-}2$ control region to yields from simulation in the μ + jets control sample for the $H_T > 375$ GeV, $n_{jet} = 4$ category. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the statistical uncertainty of the fit. The χ^2 parameters represent the goodness of fit to the control and signal region.

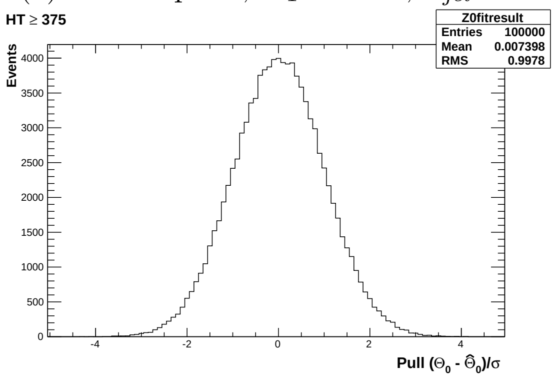
D.2. Pull Distributions for Template Fits



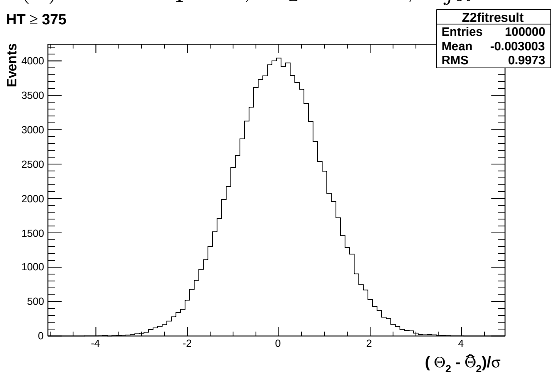
(a) Z0 Template, $H_T > 375$, $n_{jet} = 3$



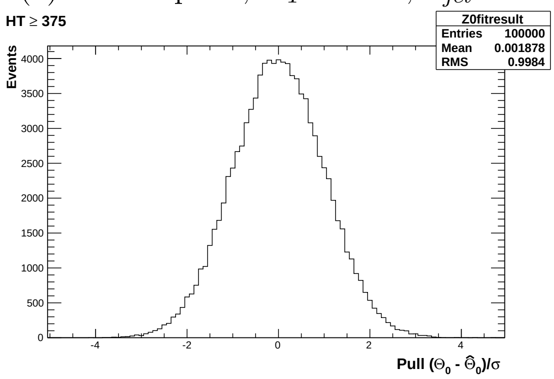
(b) Z2 Template, $H_T > 375$, $n_{jet} = 3$



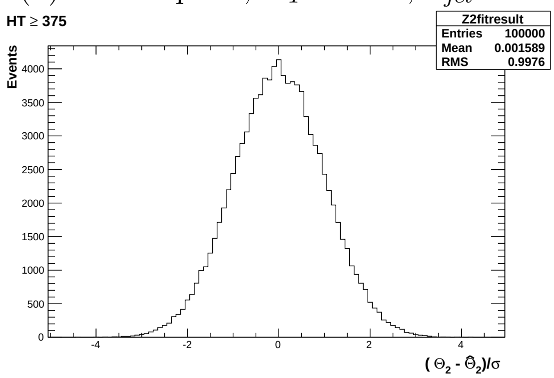
(a) Z0 Template, $H_T > 375$, $n_{jet} = 4$



(b) Z2 Template, $H_T > 375$, $n_{jet} = 4$



(a) Z0 Template, $H_T > 375$, $n_{jet} \geq 5$

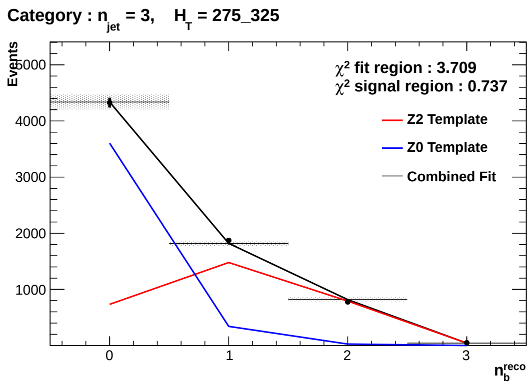


(b) Z2 Template, $H_T > 375$, $n_{jet} \geq 5$

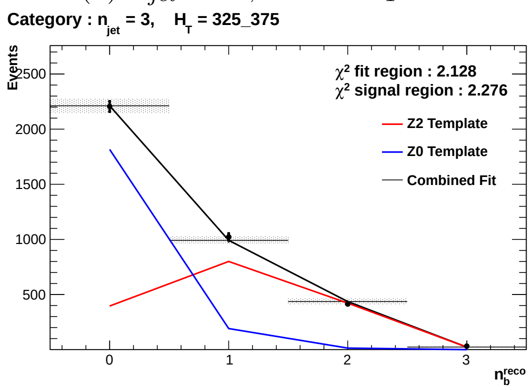
Figure D.3: Pull distributions of the normalisation parameter of each template, $\frac{(\theta-\hat{\theta})}{\sigma}$. Distributions are constructed from 10^4 pseudo-experiments generated by a gaussian distribution with width σ, centred on the nominal template value of each point within the low $n_{\mathrm{b}}^{\mathrm{reco}}$ control region. Distributions are shown for both Z0 and Z2 templates for the medium CSV working point.

D.3. Templates Fits in Data Control Sample

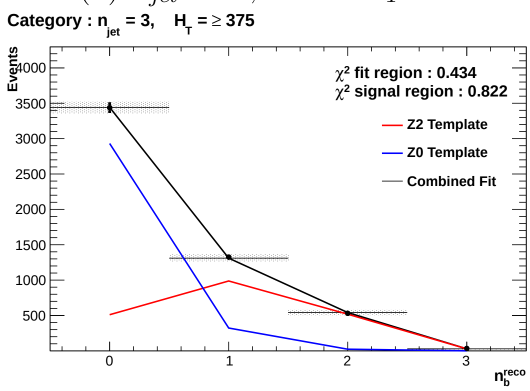
Template fits for the three H_T bins, in the $n_{jet} = 3$, medium CSV working point:



(a) $n_{jet} = 3$, $275 < H_T < 325$



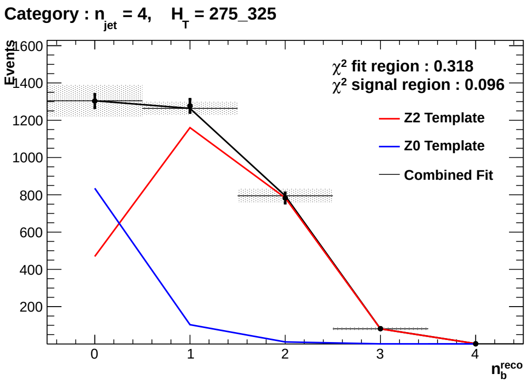
(b) $n_{jet} = 3$, $325 < H_T < 375$



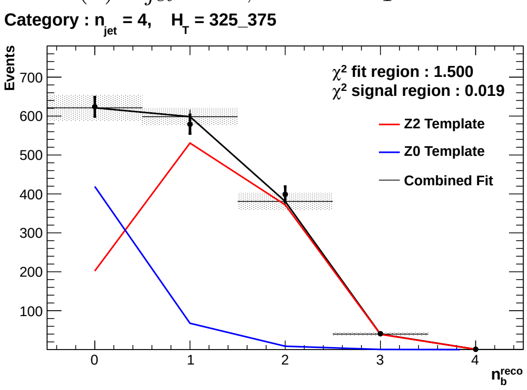
(c) $n_{jet} = 3$, $H_T > 375$

Figure D.4: Results of fitting the Z = 0 and Z = 2 templates in the n_b^{reco} = 0-2 control region to data from the μ + jets control sample, for the CSVM working point, with $n_{jet} = 3$ in each H_T category. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameters represent the goodness of fit to the control and signal region.

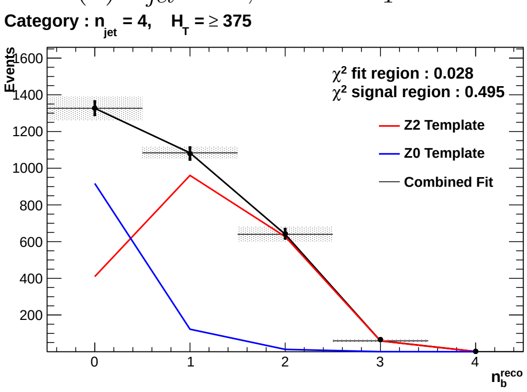
Template fits for the three H_T bins, in the $n_{jet} = 4$, medium CSV working point:



(a) $n_{jet} = 4$, $275 < H_T < 325$



(b) $n_{jet} = 4$, $325 < H_T < 375$

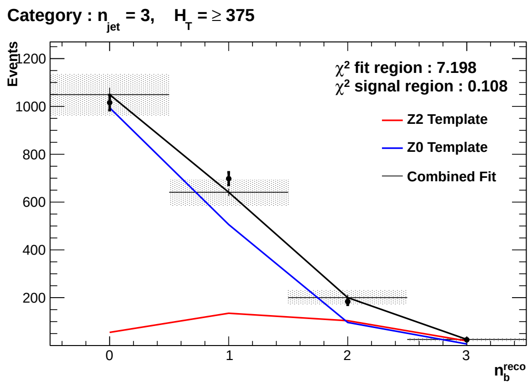


(c) $n_{jet} = 4$, $H_T > 375$

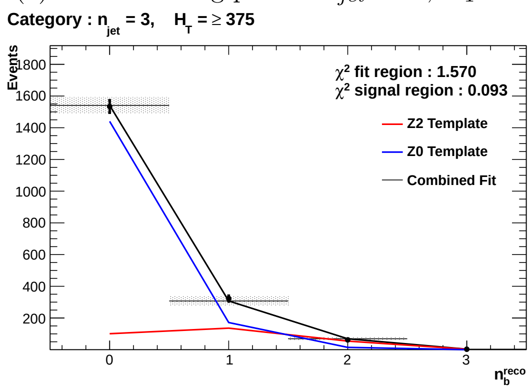
Figure D.5: Results of fitting the Z = 0 and Z = 2 templates in the n_b^{reco} = 0-2 control region to data from the μ + jets control sample, for the CSVM working point, with $n_{jet} = 4$ in each H_T category. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameters represents the goodness of fit to the control and signal region.

D.4. Templates Fits in Data Signal Region

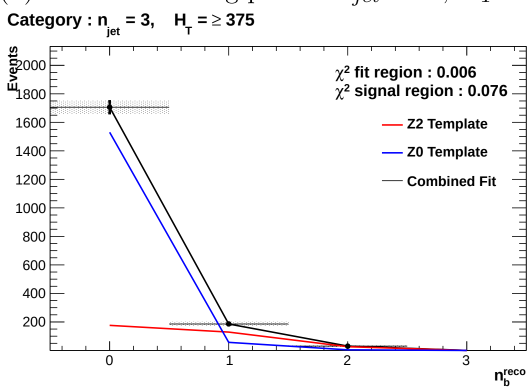
Template fits for the three CSV working points, in the $n_{jet} = 3$, $H_T > 375$ category :



(a) Loose working point : $n_{jet} = 3$, $H_T > 375$



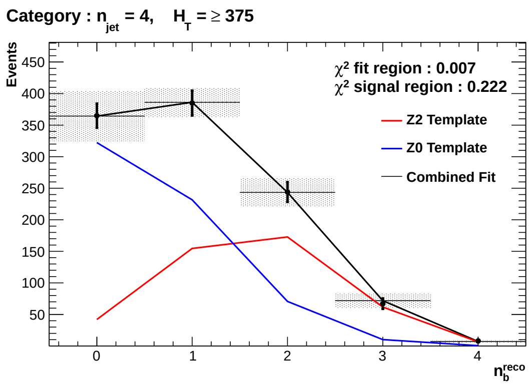
(b) Medium working point : $n_{jet} = 3$, $H_T > 375$



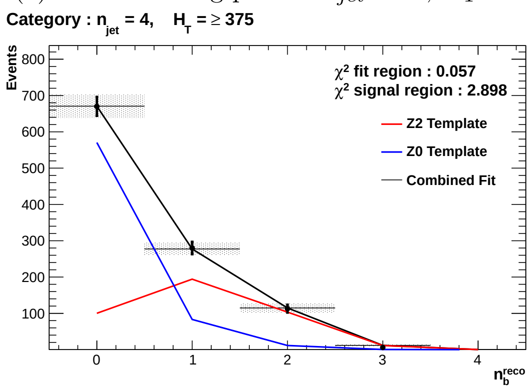
(c) Tight working point : $n_{jet} = 3$, $H_T > 375$

Figure D.6: Results of fitting the Z = 0 and Z = 2 templates in the $n_b^{reco} = 0\text{-}2$ control region to data from the hadronic signal selection, in the $n_{jet} = 3$ and $H_T > 375$ category for all CSV working points. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameters represent the goodness of fit to the control and signal region.

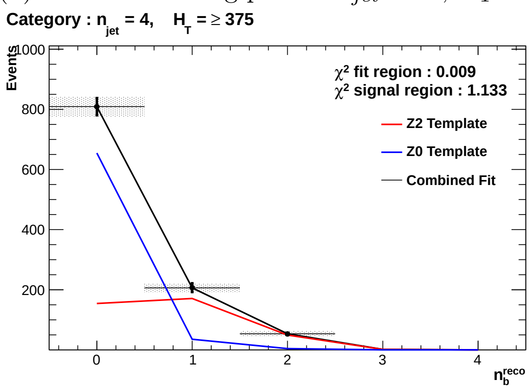
Template fits for the three CSV working points, in the $n_{jet} = 4$, $H_T > 375$ category :



(a) Loose working point : $n_{jet} = 4$, $H_T > 375$



(b) Medium working point : $n_{jet} = 4$, $H_T > 375$



(c) Tight working point : $n_{jet} = 4$, $H_T > 375$

Figure D.7: Results of fitting the Z = 0 and Z = 2 templates in the $n_b^{reco} = $ 0-2 control region to data from the hadronic signal selection, in the $n_{jet} = 4$ and $H_T > 375$ category for all CSV working points. Data is represented by the black circles with the blue, red and black lines representing the Z=0, Z=2 and combination of both templates respectively. Grey bands represent the uncertainty of the fit. The χ^2 parameters represent the goodness of fit to the control and signal region.

Bibliography

[1] Particle Data Group Collaboration, "Review of Particle Physics (RPP)", *Phys.Rev.* **D86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[2] Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters", arXiv:1303.5076.

[3] 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, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[4] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys.Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[5] CMS Collaboration, "Search for supersymmetry in hadronic final states with missing transverse energy using the variables α_T and b-quark multiplicity in pp collisions at 8 TeV", *Eur.Phys.J.* **C73** (2013) 2568, doi:10.1140/epjc/s10052-013-2568-6, arXiv:1303.2985.

[6] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) doi:10.1103/PhysRevLett.19.1264.

[7] S. Glashow, "Partial Symmetries of Weak Interactions", *Nucl.Phys.* **22** (1961) doi:10.1016/0029-5582(61)90469-2.

[8] A. Salam, "Weak and Electromagnetic Interactions", *Conf.Proc.* **C680519** (1968).

[9] G. Hooft, "Renormalizable Lagrangians for massive Yang-Mills fields", *Nuclear Physics B* **35** (1971) doi:http://dx.doi.org/10.1016/0550-3213(71)90139-8.

[10] Gargamelle Neutrino Collaboration, "Observation of Neutrino Like Interactions

Without Muon Or Electron in the Gargamelle Neutrino Experiment", *Phys.Lett.* **B46** (1973) 138–140, doi:10.1016/0370-2693(73)90499-1.

[11] UA1 Collaboration, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV at the CERN SPS Collider", *Phys.Lett.* **B126** (1983) 398–410, doi:10.1016/0370-2693(83)90188-0.

[12] UA2 Collaboration, "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $\bar{p}p$ Collider", *Phys.Lett.* **B122** (1983) doi:10.1016/0370-2693(83)91605-2.

[13] E. Noether, "Invariante Variationsprobleme", *Nachrichten von der Gesellschaft der Wissenschaften zu Gttingen, Mathematisch-Physikalische Klasse* **1918** (1918).

[14] F. Halzen and A. D. Martin, "Quarks and Leptons". 1985.

[15] "Introduction to Elementary Particles". Wiley-VCH, 2nd edition, October, 2008.

[16] C. S. Wu et al., "Experimental Test of Parity Conservation in Beta Decay", *Physical Review* **105** (February, 1957) doi:10.1103/PhysRev.105.1413.

[17] P. Higgs, "Broken symmetries, massless particles and gauge fields", *Physics Letters* **12** (1964), no. 2, doi:http://dx.doi.org/10.1016/0031-9163(64)91136-9.

[18] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) doi:10.1103/PhysRevLett.13.321.

[19] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) doi:10.1103/PhysRevLett.13.508.

[20] G. S. Guralnik, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (Nov, 1964) doi:10.1103/PhysRevLett.13.585.

[21] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[22] H. Yukawa, "On the Interaction of Elementary Particles. I", *Progress of Theoretical Physics Supplement* **1** (1955) doi:10.1143/PTPS.1.1.

[23] Super-Kamiokande Collaboration, "Evidence for Oscillation of Atmospheric Neutrinos", *Phys. Rev. Lett.* **81** (Aug, 1998) doi:10.1103/PhysRevLett.81.1562.

[24] R. Becker-Szendy et al., "A Search for muon-neutrino oscillations with the IMB detector", *Phys.Rev.Lett.* **69** (1992) doi:10.1103/PhysRevLett.69.1010.

[25] S. P. Martin, "A Supersymmetry primer", arXiv:hep-ph/9709356.

[26] H. Nilles, "Supersymmetry, Supergravity and Particle Physics". Physics reports. North-Holland Physics Publ., 1984.

[27] H. E. Haber and G. L. Kane, "The Search for Supersymmetry: Probing Physics Beyond the Standard Model", *Phys.Rept.* **117** (1985) doi:10.1016/0370-1573(85)90051-1.

[28] E. Witten, "Dynamical Breaking of Supersymmetry", *Nucl.Phys.* **B188** (1981) doi:10.1016/0550-3213(81)90006-7.

[29] J. Wess and B. Zumino, "Supergauge transformations in four dimensions", *Nuclear Physics B* **70** (1974), no. 1, doi:http://dx.doi.org/10.1016/0550-3213(74)90355-1.

[30] H. Muller-Kirsten and A. Wiedemann, "Introduction to Supersymmetry". World Scientific lecture notes in physics. World Scientific, 2010.

[31] I. Aitchison, "Supersymmetry in Particle Physics: An Elementary Introduction". Cambridge University Press, 2007.

[32] K. A. Intriligator and N. Seiberg, "Lectures on Supersymmetry Breaking", *Class.Quant.Grav.* **24** (2007) arXiv:hep-ph/0702069.

[33] Y. Shadmi, "Supersymmetry breaking", arXiv:hep-th/0601076.

[34] C. Burgess et al., "Warped Supersymmetry Breaking", *JHEP* **0804** (2008) doi:10.1088/1126-6708/2008/04/053, arXiv:hep-th/0610255.

[35] H. Murayama, "Supersymmetry breaking made easy, viable, and generic", arXiv:0709.3041.

[36] H. Baer and X. Tata, "Weak Scale Supersymmetry: From Superfields to Scattering Events". Cambridge University Press, 2006.

[37] S. P. Martin, "Implications of supersymmetric models with natural R-parity conservation", doi:10.1103/PhysRevD.54.2340, arXiv:hep-ph/9602349.

[38] G. L. Kane et al., "Study of constrained minimal supersymmetry", *Phys.Rev.*

D49 (1994) doi:10.1103/PhysRevD.49.6173, arXiv:hep-ph/9312272.

[39] C. Strege et al., "Updated global fits of the cMSSM including the latest LHC SUSY and Higgs searches and XENON100 data", *JCAP* **1203** (2012) doi:10.1088/1475-7516/2012/03/030, arXiv:1112.4192.

[40] M. Citron et al., "The End of the CMSSM Coannihilation Strip is Nigh", *Phys.Rev.* **D87** (2013) doi:10.1103/PhysRevD.87.036012, arXiv:1212.2886.

[41] D. Ghosh et al., "How Constrained is the cMSSM?", *Phys.Rev.* **D86** (2012) doi:10.1103/PhysRevD.86.055007, arXiv:1205.2283.

[42] LHC New Physics Working Group Collaboration, "Simplified Models for LHC New Physics Searches", *J.Phys.* **G39** (2012) 105005, doi:10.1088/0954-3899/39/10/105005, arXiv:1105.2838.

[43] J. Alwall et al., "Simplified Models for a First Characterization of New Physics at the LHC", *Phys.Rev.* **D79** (2009) 075020, doi:10.1103/PhysRevD.79.075020, arXiv:0810.3921.

[44] CMS Collaboration, "Interpretation of Searches for Supersymmetry with simplified Models", *Phys.Rev.* **D88** (2013) 052017, doi:10.1103/PhysRevD.88.052017, arXiv:1301.2175.

[45] J. Hisano et al., "Natural effective supersymmetry", *Nucl.Phys.* **B584** (2000) 3–45, doi:10.1016/S0550-3213(00)00343-6, arXiv:hep-ph/0002286.

[46] M. Papucci et al., "Natural SUSY Endures", *JHEP* **1209** (2012) 035, doi:10.1007/JHEP09(2012)035, arXiv:1110.6926.

[47] B. Allanach and B. Gripaios, "Hide and Seek With Natural Supersymmetry at the LHC", *JHEP* **1205** (2012) 062, doi:10.1007/JHEP05(2012)062, arXiv:1202.6616.

[48] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[49] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) doi:10.1088/1748-0221/3/08/S08003.

[50] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **0803** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[51] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[52] J.-L. Caron, "LHC Layout. Schema general du LHC.", (Sep, 1997).

[53] CMS Collaboration, "CMS Luminosity - Public Results", , (2011). http://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[54] CERN, "CMS Compact Muon Solenoid.", (Feb, 2010). http://public.web.cern.ch/public/Objects/LHC/CMSnc.jpg.

[55] "The CMS Electromagnetic Calorimeter Project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[56] "The CMS Muon Project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[57] CMS Collaboration, "The CMS Physics Technical Design Report, Volume 1", *CERN/LHCC* **2006-001** (2006).

[58] M. Cacciari et al., "The anti-k$_t$ jet clustering algorithm", *Journal of High Energy Physics* **2008** (2008), no. 04, 063.

[59] "Jet Performance in pp Collisions at 7 TeV", CMS-PAS-JME-10-003, CERN, Geneva, (2010).

[60] X. Janssen, "Underlying event and jet reconstruction in CMS", CMS-CR-2011-012, CERN, Geneva, (Jan, 2011).

[61] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation* **6** (2011), no. 11,.

[62] R. Eusebi, "Jet energy corrections and uncertainties in CMS: reducing their impact on physics measurements", *Journal of Physics: Conference Series* **404** (2012).

[63] CMS Collaboration, "Algorithms for b Jet identification in CMS", CMS-PAS-BTV-09-001, CERN, 2009. Geneva, (Jul, 2009).

[64] CMS Collaboration, "Performance of b-tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events", CMS-PAS-BTV-13-001, CERN, Geneva, (2013).

[65] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *Journal of Instrumentation* **8** (2013), no. 04,.

[66] CMS Collaboration, “CMS. The TriDAS Project. Technical design report, Vol. 1: The trigger systems”,.

[67] CMS Collaboration, “CMS: The TriDAS Project. Technical design report, Vol. 2: Data acquisition and high-level trigger”,.

[68] G. Iles, Brooke et al., “Revised CMS Global Calorimeter Trigger Functionality & Algorithms”,.

[69] CMS Collaboration, “ Calibration and Performance of the Jets and Energy Sums in the Level-1 Trigger”, CMS IN 2013/006 (2013), CERN, Geneva, (2013).

[70] B. Shorney-Mathias, “Search for supersymmetry in pp collisions with all-hadronic final states using the α_T variable with the CMS detector at the LHC”. PhD thesis, 2014.

[71] B. Mathias et al., “Study of Level-1 Trigger Jet Performance in High Pile-up Running Conditions”, CMS-IN-2013-007, CERN, Geneva, (2013).

[72] J. J. Brooke, “Performance of the CMS Level-1 Trigger”, CMS-CR-2012-322, CERN, Geneva, (Nov, 2012).

[73] CMS Collaboration, “Search for supersymmetry in final states with missing transverse energy and 0, 1, 2, or at least 3 b-quark jets in 7 TeV pp collisions using the variable α_T”, *JHEP* **1301** (2013) 077, `doi:10.1007/JHEP01(2013)077`, `arXiv:1210.8115`.

[74] CMS Collaboration, “SUSY searches with dijet events”, CMS-PAS-SUS-08-005, (2008).

[75] L. Randall and D. Tucker-Smith, “Dijet Searches for Supersymmetry at the Large Hadron Collider”, *Phys. Rev. Lett.* **101** (Nov, 2008) 221803, `doi:10.1103/PhysRevLett.101.221803`.

[76] CMS Collaboration, “Search strategy for exclusive multi-jet events from supersymmetry at CMS”, CMS-PAS-SUS-09-001, CERN, (Jul, 2009).

[77] CMS Collaboration, “CMS Btag POG : CMS b-tagging performance database”, , (2013). https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG.

[78] CMS Collaboration, “Calorimeter Jet Quality Criteria for the First CMS Collision Data”, CMS-PAS-JME-09-008, CERN, (Apr, 2010).

[79] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at 7 TeV”, *Journal of Instrumentation* **7** (October, 2012) 2P, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[80] CMS Collaboration, “Search for supersymmetry in events with photons and missing energy”, CMS-PAS-SUS-12-018, CERN, Geneva, (2012).

[81] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys.Lett.* **B659** (2008) doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[82] Z. Bern et al., “Driving missing data at next-to-leading order”, *Phys. Rev. D* **84** (Dec, 2011) 114002, doi:10.1103/PhysRevD.84.114002.

[83] Z. Bern et al., “Driving Missing Data at Next-to-Leading Order”, *Phys.Rev.* **D84** (2011) 114002, doi:10.1103/PhysRevD.84.114002, arXiv:1106.1423.

[84] CMS Collaboration, “Data-Driven Estimation of the Invisible Z Background to the SUSY MET Plus Jets Search”, CMS-PAS-SUS-08-002, CERN, (Jan, 2009).

[85] T. Sjostrand et al., “PYTHIA 6.4 Physics and Manual”, *JHEP* **0605** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[86] W. Beenakker et al., “Squark and gluino production at hadron colliders”, *Nucl.Phys.* **B492** (1997) 51–103, doi:10.1016/S0550-3213(97)80027-2, arXiv:hep-ph/9610490.

[87] S. Abdullin et al., “The Fast Simulation of the CMS Detector at LHC”, *Journal of Physics: Conference Series* **331** (2011), no. 3,.

[88] S. Banerjee, M. D. Hildreth, and the CMS Collaboration, “Validation and Tuning of the CMS Full Simulation”, *Journal of Physics: Conference Series* **331** (2011).

[89] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2012 Update”, CMS-PAS-LUM-12-001, CERN, Geneva, (2012).

[90] M. Botje et al., “The PDF4LHC Working Group Interim Recommendations”, arXiv:1101.0538.

[91] E. M. Laird, “A Search for Squarks and Gluinos with the CMS Detector”. PhD thesis, 2012.

[92] R. Cousings, “Probability Density Functions for Positive Nuisance Parameters”. 2012. http://www.physics.ucla.edu/cousins/stats/cousinslognormalprior.pdf.

[93] L. Moneta et al., "The RooStats project", 2010. arXiv:1009.1003.

[94] 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, doi:10.1016/0010-4655(75)90039-9.

[95] A. L. Read, "Presentation of search results: the CL s technique", *Journal of Physics G: Nuclear and Particle Physics* **28** (2002).

[96] T. Junk, "Confidence level computation for combining searches with small statistics", *Nuclear Instruments and Methods in Physics Research Section A* **434** (1999), no. 23, doi:http://dx.doi.org/10.1016/S0168-9002(99)00498-2.

[97] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)", CERN-OPEN-2000-205, (2000).

[98] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *The European Physical Journal C* **71** (2011) doi:10.1140/epjc/s10052-011-1554-0.

[99] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *European Physical Journal C* **71** (February, 2011) doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727.

[100] CMS Collaboration, "Search for supersymmetry in final states with missing transverse energy and 0, 1, 2, or at least 3 b-quark jets in 7 TeV pp collisions using the variable α_T", *JHEP* **1301** (2013) 077, doi:10.1007/JHEP01(2013)077, arXiv:1210.8115.

[101] CMS Collaboration, "Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 8$ TeV", arXiv:1402.4770.

[102] CMS Collaboration, "Search for new physics in events with same-sign dileptons and jets in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **1401** (2014) 163, doi:10.1007/JHEP01(2014)163, arXiv:1311.6736.

[103] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 8$ TeV in events with a single lepton, large jet multiplicity, and multiple b jets", arXiv:1311.4937.

Acronyms

ALICE A Large Ion Collider Experiment

ATLAS A Toroidal LHC ApparatuS

APD Avalanche Photo-Diodes

BSM Beyond Standard Model

CERN European Organisation for Nuclear Research

CMS Compact Muon Solenoid

CMSSM Compressed Minimal SuperSymmetric Model

CSC Cathode Stripe Chamber

CSV Combined Secondary Vertex

CSVM Combined Secondary Vertex Medium Working Point

DT Drift Tube

ECAL Electromagnetic CALorimeter

EB Electromagnetic CALorimeter Barrel

EE Electromagnetic CALorimeter Endcap

ES Electromagnetic CALorimeter pre-Shower

EMG Exponentially Modified Gaussian

EPJC European Physical Journal C

EWK Electroweak Sector

GCT Global Calorimeter Trigger

GMT Global MuonTrigger

GT Global Trigger

HB Hadron Barrel

HCAL Hadronic CALorimeter

HE Hadron Endcaps

HF Hadron Forward

HLT Higher Level Trigger

HO Hadron Outer

HPD Hybrid Photo Diode

ISR Initial State Radiation

LUT Look Up Table

L1 Level 1 Trigger

LEP Large Electron-Positron Collidor

LHC Large Hadron Collider

LHCb Large Hadron Collider Beauty

LO Leading Order

LSP Lightest Supersymmetric Partner

NLL Next to Leading Logorithmic Order

NLO Next to Leading Order

NNLO Next to Next Leading Order

POGs Physics Object Groups

PS Proton Synchrotron

QED Quantum Electro-Dynamics

QCD Quantum Chromo-Dynamics

QFT Quantum Field Theory

RBXs Readout Boxes

RPC Resistive Plate Chamber

RCT Regional Calorimeter Trigger

RMT Regional Muon Trigger

SUSY SUperSYmmetry

SM Standard Model

SMS Simplified Model Spectra

SPS Super Proton Synchrotron

TF Transfer Factor

TP Trigger Primative

VEV Vacuum Expectation Value

VPT Vacuum Photo-Triodes

WIMP Weakly Interacting Massive Particle