A search for supersymmetry in $\sqrt{s} = 13$ TeV proton-proton collisions with the CMS detector at the LHC

Adam Christopher Elwood

Imperial College London
Department of Physics

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

Abstract

An inclusive search for supersymmetry with jets and missing transverse energy is presented. Data from $\sqrt{s} = 13$ TeV pp-collisions with a total integrated luminosity of 12.9 fb^{-1} delivered by the LHC and collected by the CMS detector are analysed. The dominant quantum chromodynamic multijet background is strongly suppressed with several kinematic variables, which are also used to discriminate between Standard Model and supersymmetric processes. The observed events are found to be compatible with the expected contributions from Standard Model processes. This result is interpreted in the context of simplified supersymmetric models of gluino and third-generation squark production. The mass of the gluino, bottom squark and top squark are excluded to 1775, 1025 and 875 GeV respectively.

In preparation for the collection of $\sqrt{s} = 13$ TeV data by CMS, the jet algorithm for the Level-1 trigger is upgraded. The new algorithm allows for dynamic pileup subtraction and takes advantage of hardware upgrades to the trigger. The performance of different types of pileup subtraction are evaluated and the most promising algorithm, *chunky-donut subtraction*, is chosen. The algorithm is found to give a significant performance improvement and has been used to collect data from 2016 onwards.

Declaration

This dissertation is the result of my own work, except where reference is made to the work of others. To produce the results in this thesis I worked in collaboration with other members of the α_T analysis group and the trigger studies group, who were predominantly based at Imperial College London. Figures from CMS preliminary and unpublished results are labelled "CMS Preliminary". Figures from CMS publications are labelled "CMS" or "CMS Simulation", when the data are only taken from simulation. This thesis has not been submitted for another qualification to this or any other university.

Adam Elwood

Acknowledgements

It has been a great privilege to work on the frontier of fundamental physics research with many talented researchers. I would first like to thank Alex Tapper for his excellent supervision and guidance throughout my studies. The help and expertise of the other researchers in the group has also been invaluable. I would like to thank everyone within the α_{T} analysis group, particularly Rob Bainbridge, Stefano Casasso and Bjoern Penning. Also, for their enthusiastic guidance during the development of a new trigger algorithm, I'd like to thank Jad Marrouche and Andrew Rose. Additionally, I'd like to thank all my fellow students and friends for providing support and much needed distraction along the way, including (but not limited to): Mark Baber, Matthew Citron, Louie Corpe, Patrick Dunne, Christian Laner, Lucien Lo, Federico Redi and Adinda de Wit.

I would also like to extend my thanks to everyone in the Imperial College HEP group, who provided a supportive and interesting community in which to carry out research. Also, thanks to the STFC, who provided the funding that allowed me to conduct my research and have a great experience for two years living and working at CERN.

Finally, I would like to thank my family, particularly my parents, for providing the immense support required for me to make it this far. Without the keen interest they displayed in my subject through all stages of my study I wouldn't be in the position I am now. Thanks also to Giuditta for tolerating and supporting me as I jumped the final hurdles.

"Traditional scientific method has always been, at the very best, 20-20 hindsight. It's good for seeing where you've been. It's good for testing the truth of what you think you know, but it can't tell you where you ought to go."

— Robert M. Pirsig, *Zen and the art of motorcycle maintenance*

Contents

Chapter 1

Introduction

Modern physics has now reached a point in which our fundamental understanding can be broken down into two separate arenas. At large length scales, where gravity is dominant, the theory of General Relativity (GR) [1] is incredibly successful in reproducing experimental observations [2]. However, to provide a description of the subatomic constituents of matter and the three other fundamental forces, we rely on a quantum field theory, the Standard Model (SM) of particle physics [3–5]. At very high energy densities, where both theories are relevant, our understanding breaks down [6]. The main aim of fundamental physics research is therefore the reconcilation of quantum field theory with general relativity.

One way to approach this is to explore higher energy scales, above the mass scales of the SM, but before GR becomes relevant. To this end, a giant proton synchrotron, the Large Hadron Collider (LHC), was built at CERN near Geneva, Switzerland [7]. The LHC is designed to collide protons at record breaking energies. Around these collision points are built various detectors that explore the results of these high energy collisions. One of the two general purpose detectors designed for searching for a wide range of phenomena is the Compact Muon Solenoid (CMS) detector [8]. It is the exploration of the new energies beyond the SM with CMS that is the subject of this thesis.

The SM is one of the most successful scientific theories to date. It makes predictions about the physical world that have consistently stood up to experimental scrutiny, culminating in the discovery of a 125 GeV Higgs boson at the LHC in 2012 [9, 10]. Despite its successes, the SM does not provide a description for some significant experimental anomalies. From astronomical observations, it can be inferred that SM particles cannot solely account for the total gravitational behaviour of various objects

in the universe [11–16]. This anomaly can be explained by introducing a new form of weakly interacting particle, known as *dark matter (DM)*. On top of this it is observed that the rate of the expansion of the universe is increasing, implying the existence of *dark energy* [17, 18]. A rough calculation of the scale of dark energy predicted by the SM yields a result that is many orders of magnitude different from the observed value. This implies a signicant lack of understanding of the origin of this phenomenon. Additionally, fine tuning problems are present within the SM itself. Along with the irreconcilability of the SM with GR, these issues point towards the existence of a more fundamental theory that goes beyond the SM.

One popular way to extend the SM is to introduce a new broken spacetime symmetry between fermions and bosons, known as supersymmetry (SUSY) [19]. Initially motivated with purely mathematical arguments, SUSY models can provide a candidate for DM, solve the Higgs hierarchy problem and also unify the strong, weak and electromagnetic forces at the Grand Unified Theory (GUT) scale, which is not possible in the SM. To convincingly solve these problems, SUSY is expected to exhibit itself close to the electroweak scale of the SM. If this is the case, there is a significant chance that supersymmetric particles will be produced at the LHC.

In its first run, *Run 1*, the LHC delivered a large dataset at 7 and 8 TeV centre of mass energies. The analysis of these data, however, has not resulted in any observation of SUSY. With Run 2 of the LHC, that began in 2015, the centre of mass collision energy has been increased to 13 TeV. The data taken in this new run therefore hold the best chance yet for the discovery of electroweak scale SUSY.

This thesis presents the description and results of a search for SUSY in data collected by the CMS detector at the LHC during proton-proton collisions with a centre of mass energy $\sqrt{s} = 13$ TeV. The analysis presented looks for SUSY signatures in a final state with hadronic jets and missing transverse momentum.

Chapter 2

Theory

2.1 The Standard Model of particle physics

The SM describes the interaction of matter through the electromagnetic, weak nuclear and strong forces in the context of a renormalisable quantum field theory [3–5]. Matter particles are represented as spin-$\frac{1}{2}$ fermionic fields and forces are represented as spin-1 bosonic fields. An additional spin-0 Higgs field is included to provide particles with their mass. The SM is built around the concept of local gauge invariance. Taking the fermions and applying the symmetries of the SM local gauge group, $SU(3) \times SU(2) \times U(1)$, implies the existence of the force carrying bosons. This section will briefly explore how this leads to the particle phenomenology of the SM.

The SM is typically considered within a Lagrangian formalism. In a quantum field theory all the relevant fields and their interactions are described by a Lagrangian density. The Lagrangian density of the SM can be divided into four parts:

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}, \tag{2.1}$$

where $\mathcal{L}_{fermion}$ describes the fermion fields and their interactions with the bosons are described in $\mathcal{L}_{gauge}$. The final two terms, $\mathcal{L}_{Higgs}$ and $\mathcal{L}_{Yukawa}$, describe how the particles within the SM obtain mass through interactions with the Higgs field.

Throughout this thesis the convention $c = \hbar = 1$ is used and the Einstein four-vector summation convention is assumed. Four-vector indices are labelled as μ and ν.

2.1.1 The fundamental particles

The fundamental particles of the SM comprise fermions and the force mediating bosons, a summary of them and their relevant electromagnetic, weak and strong force properties can be seen in Table 2.1.

The fermions consist of three generations of charged leptons and their corresponding weak force partners, the neutrinos. There are additionally three generations of up-type quarks and down-type quarks. For all of these twelve fermions there are corresponding antiparticles that have the same mass but opposite quantum numbers. As fermions are spin-$\frac{1}{2}$ particles, they are described by the Dirac equation [20]:

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0, \tag{2.2}$$

where ψ is the wave function of the fermion and γ^{μ} are the Dirac matrices, defined by their anti-commutation relation:

$$\{\gamma^{\mu}, \gamma^{\nu}\} = \gamma^{\mu}\gamma^{\nu} + \gamma^{\nu}\gamma^{\mu} = 2g^{\mu\nu}, \tag{2.3}$$

where $g^{\mu\nu}$ is the Minkowski metric. The covariant derivative is denoted by ∂_{μ} and m is the mass of the particle in question.

There are five types of bosons that arise from the SM gauge symmetries: the photon, gluon, $W^{\pm}$, Z^0 and the Higgs. Their properties will be discussed later in this Chapter.

2.1.2 Gauge symmetries

The insensitivity of the structure of a theory to a specific transformation constitutes a symmetry. This concept is very powerful for gaining insights into fundamental physical theories. For example, the fact that physical laws do not change over time, time-translational symmetry, leads to the conservation of energy. In general, any symmetries have a corresponding conserved quantity, as laid out in Noether's theorem [24]. This concept is used extensively when formulating the SM and allows for the derivation of observed interactions through the imposition of a few, fairly straightforward, symmetries.

Categories	Particle		Mass	Spin	Electric charge	Colour charge	Weak isospin (t_3)
Leptons	electron	e	0.511 MeV				
	muon	μ	106 MeV	$\frac{1}{2}$	-1	0	$-\frac{1}{2}$
	tau	τ	1777 MeV				
Neutrinos	electron	ν_e	<2 eV				
	muon	ν_μ	<0.19 MeV	$\frac{1}{2}$	0	0	$+\frac{1}{2}$
	tau	ν_τ	<18.2 MeV				
Up-type quarks	up	u	2.3 MeV				
	charm	c	1.28 GeV	$\frac{1}{2}$	$+\frac{2}{3}$	r, g, b	$+\frac{1}{2}$
	top	t	173 GeV				
Down-type quarks	down	d	4.8 MeV				
	strange	s	95 MeV	$\frac{1}{2}$	$-\frac{1}{3}$	r, g, b	$-\frac{1}{2}$
	bottom	b	4.18 GeV				
Force mediating bosons	photon	γ	0	1	0	0	0
	gluon	g	0	1	0	$r\bar{g}, r\bar{b}, g\bar{r}, g\bar{b}$ $b\bar{r}, b\bar{g}, \frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g})$ $\frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b})$	0
	W	$W^\pm$	80.4 GeV		± 1		± 1
	Z	Z^0	91.2 GeV	1	0	0	0
	Higgs	h^0	125 GeV		0		$-\frac{1}{2}$

Table 2.1: All the fundamental Standard Model fermions and bosons and their properties [21]. It is worth noting that, despite the fact that the muon and tau neutrinos have less stringent limits on their masses than the electron neutrino, results from oscillation experiments constrain the neutrino mass difference, which suggests that they also have masses < 2 eV [22, 23].

The effect of applying a symmetry within the SM is demonstrated when imposing local $U(1)$ invariance on the Dirac Lagrangian for a fermion, with wavefunction ψ and mass m [20]:

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi - m\bar{\psi}\psi. \tag{2.4}$$

A global $U(1)$ transformation, $\psi \rightarrow e^{iq\theta}\psi$, where the phase θ and q are constant, leaves the Lagrangian invariant. If this $U(1)$ transformation is local, i.e. the phase depends on spacetime position, x, then the Lagrangian is no longer invariant. It now transforms as:

$$\mathcal{L} \rightarrow \mathcal{L} - q(\partial_\mu \theta(x))\bar{\psi}\gamma^\mu\psi. \tag{2.5}$$

However, one can add a vector field, A_μ, that interacts with the fermion field through the Lagrangian term:

$$\mathcal{L}_{int} = q(\bar{\psi}\gamma^\mu\psi)A_\mu. \tag{2.6}$$

This vector field is chosen to transform as $A_\mu \rightarrow A_\mu + \partial_\mu\theta$ and is known as a *gauge field* or *gauge boson*. The interaction Lagrangian term then transforms under a local gauge transformation as:

$$\mathcal{L}_{int} \rightarrow \mathcal{L}_{int} + q(\partial_\mu\theta)\bar{\psi}\gamma^\mu\psi, \tag{2.7}$$

this cancels out the term that violated local gauge invariance in Equation 2.5. The existence of a new gauge field allows the addition of an additional gauge invariant term containing the field strength tensor of the vector field, $F_{\mu\nu}$, which can be written in general as:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f_{abc} A^b_\mu A^c_\nu, \tag{2.8}$$

for a gauge group with the structure constants f^{abc} and self-coupling constant g. For the $U(1)$ group there is only one self-commuting generator so the structure constant is 0. For non-Abelian gauge groups, such as $SU(3)$, the structure constants are non-zero, which introduces self interaction terms within the Lagrangian. In this case the gauge boson is said to carry a *charge* and can interact with itself.

The final Lagrangian for a Dirac fermion can then be written as:

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\mathcal{D}_{\mu}\psi - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.9}$$

where $\mathcal{D}_{\mu} = \partial_{\mu} + iqA_{\mu}$ and is known as the *covariant derivative*. This Lagrangian will be invariant under local $U(1)$ transformations. In this case, the addition of one extra gauge field maintains local invariance. As $U(1)$ transformations have one degree of freedom, this gauge field corresponds to the single generator of the group. To maintain the local gauge invariance of any symmetry, a gauge boson per degree of freedom must be introduced.

The method of obtaining local gauge invariance through the introduction of gauge bosons is applied with great success to the gauge group of the SM. With the choice of an appropriate gauge group, $SU(3) \times SU(2) \times U(1)$, the bosons that describe the strong, weak and electromagnetic forces can all be obtained. In the example demonstrated in this section, the final Lagrangian (Eq. 2.9) describes quantum electrodynamics (QED). It predicts the massless photon field, A_{μ}, from the $U(1)$ local gauge invariance of fermions with a coupling strength corresponding to the electric charge, represented by q.

2.1.3 The strong force

The strong force can be described with the $SU(3)$ gauge group, resulting in the interaction of quark fields via eight massless gauge fields, the gluons. This theory is known as quantum chromodynamics (QCD) in which the quark fields possess a *colour* charge, $C = (r, g, b)$. As the $SU(3)$ group is non-Abelian, the gluons also possess a colour and anti-colour charge. This leads to gluon-self couplings, which results in the short range of the strong force. Additionally, screening effects from virtual gluons leads to the phenomena known as *asymptotic freedom* [25]. It is characterised by the strong coupling constant, α_s, getting weaker over short ranges. This leads to quarks behaving as if they are unbound when they are very close but more strongly coupled as they move apart. The fact that α_s can be large makes it very challenging to calculate QCD perturbatively using the well known techniques that work for the electromagnetic and weak forces. This makes QCD calculations difficult to do and less accurate as a result.

One important property of the strong force is that quarks are confined to exist in *colour-singlet* states. This typically leads to either *mesons* comprising a quark-antiquark

pair or *baryons* that are a triple quark or anti-quark bound state. Four or five quarks can also be bound in *tetra-quark* or *penta-quark* states, although they are typically less stable [21,26]. These different bound states are known collectively as *hadrons*. Despite being very strongly bound, if quarks in these states are given significant energy they can be liberated from their original bound state through the pair production of quark-antiquark pairs, resulting in the production of new hadrons. This process is known as *hadronisation* and occurs when the energy contained within the gluons binding the quarks exceeds the energy contained within the mass of the newly produced quark pair.

Within the environment of a particle collider, the process of hadronisation is a common occurrence. If the hadrons that are produced have significant energy they can also break apart, undergoing further hadronisation, known as *fragmentation*. This leads to a quark or gluon that gains significant momentum in a collision producing a collimated emission of hadrons from the collision point, known as a *jet*.

2.1.4 Electroweak unification

The electromagnetic and weak forces are described in the SM by the symmetry group $SU(2) \times U(1)$. The requirement of local gauge invariance in the weak sector led to the electromagnetic and weak forces being unified within this group in a landmark achievement in the 1960s [3–5].

The $SU(2)$ group has three generators, $T_i = \tau_i/2$, where $i = 1,2,3$ and τ_i are the Pauli spin matrices. Each of these generators is manifested as a gauge field, labelled W^i_μ. Within the electroweak theory these gauge fields only act on the left handed chiral component of the fermion field, ψ_L, where $\psi_L = \frac{1}{2}(1-\gamma_5)\psi$ and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. This *left-handedness* of the electroweak theory leads to the parity violation that is observed in weak interactions. The charges associated with these gauge fields are known as *weak isospin* and are denoted t_i. As with the $SU(3)$ group the $SU(2)$ group is non-Abelian and the gauge bosons are able to interact with themselves.

The $U(1)$ group has a single generator, with an associated gauge field, B_μ. This field interacts with particles that carry *weak hypercharge*, $y = 2(q - t_3)$. It is worth noting that this is a different charge to that of the $U(1)$ group in QED, which was just the electromagnetic charge, q.

The physical gauge bosons are obtained by mixing the W^i_μ and B_μ gauge fields as follows:

$$\begin{aligned} \mathrm{W}^{\pm}{}_{\mu} &= \frac{1}{\sqrt{2}}\left(\mathrm{W}^1_\mu \mp i\mathrm{W}^2_\mu\right) \\ \mathrm{Z}_\mu &= \cos\left(\theta_W\right)\mathrm{W}^3_\mu - \sin\left(\theta_W\right)B_\mu \\ A_\mu &= \sin\left(\theta_W\right)\mathrm{W}^3_\mu + \cos\left(\theta_W\right)B_\mu, \end{aligned} \tag{2.10}$$

where A_μ is the photon field, Z_μ is the Z boson field and $W^\pm_\mu$ are the W boson fields. The Weinberg angle, θ_W is given by the coupling strengths of the weak hypercharge gauge field, g', and the isospin gauge field, g:

$$\theta_W = \frac{g}{g^2 + g'^2}. \tag{2.11}$$

The $W^\pm$ gauge bosons only couple to the left handed component of the fermion fields. These left-handed components form weak isospin doublets in both the quark, Q_L, and lepton fields e_L. The right-handed components form weak isospin singlets, Q_R and e_R and have $t_3 = 0$. For the first generation of quarks, the u and d and the first generation of leptons, e and ν_e, the left and right handed components of the fields are broken down as follows:

$$e_L = \begin{pmatrix} \nu_{e\,L} \\ e_L \end{pmatrix}, \quad Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad e_R = e_R, \quad Q_R = u_R, d_R, \tag{2.12}$$

where a subscript L denotes the left-handed component and a subscript R denotes the right handed component.

Within the quark doublet the charged current interactions of the $W^\pm$ fields act between up and down type quarks. However, the mass eigenstate of the quarks is not the same as the electroweak eigenstate. The mixing between these two eigenstates is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [27]. The matrix is diagonally dominant, meaning the $W^\pm$ fields are most likely to produce interactions of quarks in the same generation, however this allows for an inter-generational mixing of the quark fields.

2.1.5 Spontaneous symmetry breaking and the Higgs mechanism

Initial iterations of the SM did not provide a way for the fundamental particles to have a gauge invariant mass term in the Lagrangian. This problem was solved through a breaking of the electroweak symmetry that became known as the *Higgs mechanism* [28–33]. This *spontaneous symmetry breaking* allowed the vector bosons of the weak force to obtain mass and provided a way to write gauge invariant mass terms for the fermions.

A symmetry is spontaneously broken if the ground state of the vacuum does not share the symmetry of the Lagrangian [20]. Even though the collection of all states does share the symmetry, when the theory is in its ground state a particular vacuum energy must be chosen. This allows terms which are not gauge invariant to be added to the theory by coupling some of the fields to a new field with a non-zero vacuum expectation value. This is achieved within the SM by introducing a complex scalar $SU(2)$ field with four degrees of freedom called the Higgs field, ϕ:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{2.13}$$

This is implemented into the theory through an additional term in the SM lagrangrian:

$$\mathcal{L}_{Higgs} = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) - V(\phi), \tag{2.14}$$

where the covariant derivative is chosen to keep the Higgs field invariant under $SU(2) \times U(1)$ transformations with a weak hypercharge of $y = \frac{1}{2}$.

To spontaneously break the symmetry, the potential, V, is chosen to take the form:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2, \tag{2.15}$$

where $\mu^2 > 0$ and $\lambda > 0$. This leads to a potential with a non-zero expectation value that forms a circle in phase space. This leads to a continuous set of equivalent minima of which one must be chosen, resulting in the spontaneous symmetry breaking. By

convention a particular minimum is chosen as:

$$\langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ \sqrt{\frac{\mu^2}{2\lambda}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.16}$$

Perturbations about this vacuum expectation value can be parametrised in the form of four real scalar fields. However, with an appropriate choice of gauge, three of these degrees of freedom, known as the *Goldstone bosons*, can be set to zero. This leaves one remaining field, H, and perturbations can be written as:

$$\phi = \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \tag{2.17}$$

This can then be inserted into the Lagrangian to obtain at leading order:

$$\mathcal{L} = \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2}\mu^2 H^2 + \frac{v^2}{8}\left[g_2^2 W_\mu^+ W^{+\mu} + g_2^2 W_\mu^- W^{-\mu} + \left(g_1^2 + g_2^2\right) Z_\mu Z^\mu\right]. \tag{2.18}$$

This provides the weak vector bosons $W_\mu^\pm$ and Z_μ with mass terms $g_2 v/2$ and $\frac{v}{2}\sqrt{g_1^2 + g_2^2}$ respectively. This also introduces a massive scalar field, H, with a mass $\sqrt{2\mu^2}$, which is the Higgs boson. This achieves the aim of providing the weak vector bosons with mass in a gauge invariant way. The Higgs boson has subsequently been discovered by the ATLAS and CMS collaborations with a mass of 125 GeV [9,10].

With the existence of a Higgs field, it is also possible to write gauge invariant mass terms for the fermion fields. These are known as *Yukawa terms* and take the form:

$$\mathcal{L}_{Yuk} = y_f\left(\bar{f}_L \phi f_R + \bar{f}_R \phi^\dagger f_L\right), \tag{2.19}$$

where f_L is the left-handed component of the fermionic field and f_R is the right-handed component. The value y_f is the Yukawa coupling and leads to a fermion mass of $y_f v/\sqrt{2}$. The magnitude of the mass of the fermion is therefore determined by how strongly it couples to the Higgs field.

2.1.6 Beyond the Standard Model

The SM has been incredibly successful in describing the physics we observe up to the scale of electroweak unification, $O(100\ \text{GeV})$. It describes the behaviour of most of the observed fundamental particles very well and all its predictions have been so far verified by experiments throughout the 20th and 21st centuries. However, some inconsistencies between the SM and experimental observations point to issues with the theory that can only be solved with a more fundamental beyond the SM (BSM) theory. Also, the fact that the SM provides no description of gravity is a convincing argument for new physics at the energy scale that gravity becomes relevant. This energy scale is typically referred to as the *Planck scale*.

One of the most obvious experimental issues with the SM is that it predicts neutrinos to be massless. Within the theory, neutrinos are only produced by the weak force in their left-handed state. Without the presence of a right-handed neutrino one cannot write a Yukawa mass term. However, experiments have observed that neutrinos do have a mass, as they undergo flavour oscillations while propagating in their mass eigenstates [22, 23]. This can be included relatively straightforwardly within the SM with the addition of seven new parameters [34]. These include the neutrino masses and the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix that describes how the neutrino mass and flavour eigenstates mix.

As touched upon in Chapter 1, there are other, more major, problems with the SM that are unreconcilable within the theory. A glaring theoretical problem with the SM becomes apparent when calculating corrections to the Higgs mass, m_H, at the loop level [19]. The observable mass of the Higgs boson is very sensitive to loop contributions from fermion and scalar fields. Due to the quartic term proportional to λ in the Higgs potential (Eq. 2.15), the Higgs also interacts with itself at loop level. If it is assumed that the Higgs couples, even indirectly, to any of the new physics that exists at energy scales above the SM then the correction to the Higgs mass, Δm_H, are of the order:

$$\Delta m_H^2 \sim \frac{\lambda}{4\pi^2}\Lambda^2 + \delta M_H^2, \tag{2.20}$$

where Λ is the energy scale of the BSM physics and additional loop corrections are contained within δM_H. If BSM physics does not exist until the Planck scale, then the energy scale of the new physics must be $\Lambda \sim 10^{19}$ GeV. As the Higgs mass is

observed to be at the electroweak scale, $m_H = 125$ GeV, there must be a very precise cancellation provided by the extra corrections to the Higgs mass, over many orders of magnitude. This is not physically forbidden, but results in a *naturalness* problem within the SM, as an unexplained cancellation of this order is deemed to be unnatural. This problem is known as the *hierarchy problem* and guides the development of BSM theories. A potential solution can be obtained by introducing BSM physics at close to the electroweak scale.

Another problem with the SM comes through a significant inconsistency with observational astrophysical data. Studies of cosmological gravitational effects through the study of the rotations of galaxies [11, 12], gravitational lensing [13], the structure of matter distributed through the universe [14] and measurements of the Cosmic Microwave Background (CMB) [15, 16] imply the existence of an additional form of matter named *dark matter (DM)*. The SM provides no viable candidate for DM consistent with these predictions. As DM does not appear to interact strongly with the SM it must be weakly-interacting. Additionally, assuming that dark matter was produced thermally in the early universe, one obtains the correct abundance of DM for a ~ 100 GeV particle that interacts via the weak force, this coincidence is known as the *WIMP miracle* [35]. Attempts to understand the nature of DM typically look for it colliding with matter on earth, signatures from DM annihilation in space or the production of DM from SM collisions in a particle collider. However, no direct observation has yet been made.

The SM successfully unifies the electromagnetic and weak forces, but the strong force is not incorporated. For a full unification, the coupling constants of all the SM forces would unify at a high energy scale known as the *Grand Unified Theory (GUT)* scale. This extra consideration is aesthetically appealing and can also help to motivate future BSM theories.

One of the final major problems of the SM is the fact that it does not account for the observed matter-antimatter asymmetry in the universe. The charge parity (CP) violation within the SM is not significant enough to account for the asymmetry that we observe and should be taken account of in a final BSM theory.

2.2 Supersymmetry

The fundamental group of Minkowski spacetime isometries is known as the *Poincaré group* [36]. The only possible remaining extension of this group is a symmetry between bosons, b, and fermions, f [37]. This symmetry is postulated to exist in nature under the name of *supersymmetry (SUSY)* [19]. This theory predicts the existence of a supersymmetric partner for every SM particle with identical quantum numbers except for a difference in the spin value of $\frac{1}{2}$. As these *superpartners* have not yet been observed, they must have a higher mass than their SM counterparts. This implies that SUSY must be a broken symmetry. As there is considerable freedom in how this symmetry breaking occurs, there exists a series of possible versions of SUSY. The most popular and simplest versions of SUSY are the Minimal Supersymmetric Standard Model, introduced in Sec. 2.2.1.

Along with this theoretical motivation, SUSY can fully or partially solve several of the problems with the SM discussed in Sec. 2.1.6. In SUSY the loop contributions to the mass of the Higgs are opposite in sign for fermions and scalars. If the partners of the SM fermions are SUSY scalars, then the contributions from the SM fermions are cancelled out. This results in Eq. 2.20 taking the form [19]:

$$\Delta m_H^2 = \frac{1}{8\pi^2}(\lambda_S - |\lambda_f|^2)\Lambda^2 + ..., \tag{2.21}$$

where λ_S and λ_f are the respective coupling strengths of the scalars and fermion fields to the Higgs field. Provided the mass of SUSY particles have a value of ~ 1 TeV, the hierarchy problem is resolved with minimal fine-tuning.

Additionally, many versions of SUSY contain a weakly interacting and stable lightest supersymmetric particle (LSP), which provides an excellent candidate for DM. The addition of superpartners also helps to unify the coupling constants of the gauge groups within the SM. Without SUSY the different forces of the SM do not unify at a fixed energy scale, however loop corrections from SUSY particles results in a complete unification, as shown in Fig. 2.1. The fact that the forces can unify in this way points to the possibility of a more complete unification of particle physics with GR at high energy scales. The existence of SUSY is a prerequisite for string theory, for example, a favourite unification candidate.

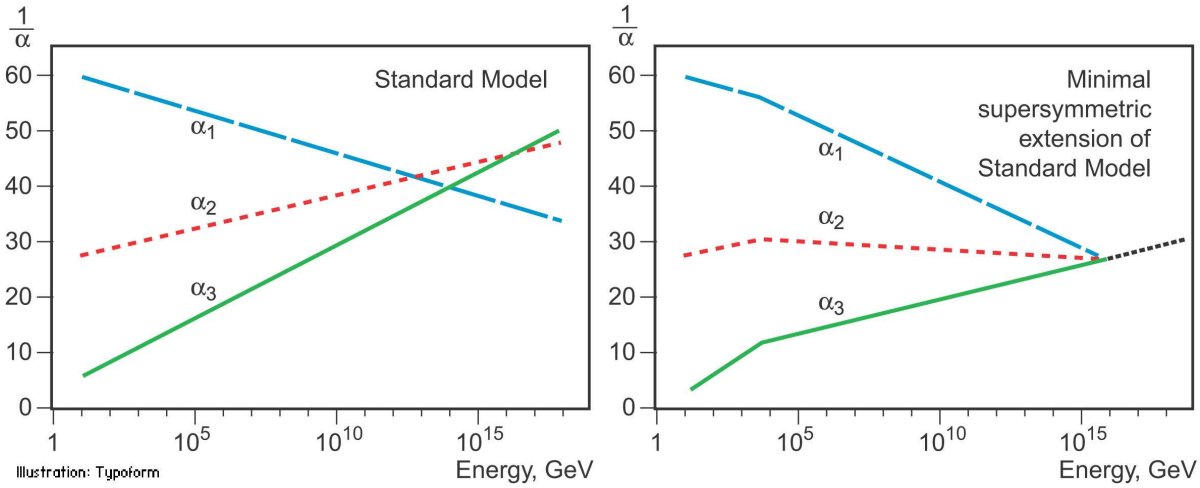


Figure 2.1: Evolution of the inverse of the coupling constants of the electroweak $U(1)$, α_1, electroweak $SU(2)$, α_2, and strong force $SU(3)$, α_3 for the SM and MSSM [38]

2.2.1 The Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) is a model that incorporates SUSY with the SM in a way that adds the minimum number of new particles and interactions [39]. The MSSM has 105 free parameters in total, a significant increase from the 19 in the SM. A summary of the extra particles, *sparticles*, introduced by the MSSM is available in Tab. 2.2. The superpartners to the fermions are denoted with the same symbol as their associated fermion with a tilde. The weak sector is extended within the MSSM and the mixing of the new SUSY fields associated to the electroweak and Higgs bosons leads to the existence of neutralinos, charginos and a total of five Higgs bosons, including two that carry an electromagnetic charge.

Unlike the SM, the MSSM permits lepton and baryon number violation, the total numbers of leptons and baryons do not need to be conserved. This could lead to interactions in which quarks can produce leptons, such interactions would allow protons to spontaneously decay. However, stringent limits have been set on the lifetime of a proton, $\tau > 10^{33}$ years [40], that should exclude or heavily suppress this kind of behaviour. This is dealt with in the MSSM by introducing a new conserved quantity known as *R-parity*, R:

$$R \equiv (-1)^{3(B-L)+2s} \tag{2.22}$$

Name	Sparticle	Spin	Charge, q
Up-type squarks	$\tilde{u}, \tilde{c}, \tilde{t}$	0	$+\frac{2}{3}$
Down-type squarks	$\tilde{d}, \tilde{s}, \tilde{b}$	0	$-\frac{1}{3}$
Charged sleptons	$\tilde{e}, \tilde{\mu}, \tilde{\tau}$	0	± 1
Sneutrinos	$\tilde{\nu_e}, \tilde{\nu_\mu}, \tilde{\nu_\tau}$	0	0
Gluino	$\tilde{g}$	$\frac{1}{2}$	0
Neutralinos	$\tilde{\chi_1}^0, \tilde{\chi_2}^0, \tilde{\chi_3}^0, \tilde{\chi_4}^0$	$\frac{1}{2}$	0
Charginos	$\tilde{\chi_1}^\pm, \tilde{\chi_2}^\pm$	$\frac{1}{2}$	± 1
Neutral Higgs bosons	h^0, H^0, A^0	0	0
Charged higgs bosons	$H^\pm$	0	± 1

Table 2.2: The extra particles introduced by the MSSM. The symbol h^0 is typically used to denote the SM Higgs boson. [19]

where B is the baryon number, L the lepton number and s the spin of the particle. This leads to SUSY particles having $R = -1$ and SM particles having $R = +1$. The consequence of this is that SUSY particles always decay to at least one other SUSY particle. Additionally, any SUSY particles produced by SM interactions must be produced in pairs and vice versa. With R-parity conservation the lightest supersymmetric particle (LSP) is stable. This allows it to fulfil the role of DM in the case that the LSP is neutral. In most favoured SUSY models the LSP is therefore typically taken to be a neutralino [41].

For the MSSM to successfully solve the hierarchy problem, an additional constraint is applied to the SUSY particles. As there is no explicit mass hierarchy in the squark generations predicted in SUSY, the sparticle of the top-quark, the *stop*, is typically required to be the lightest of the squarks. This ensures maximal cancellation of the contributions to the Higgs mass from the top-quark, which couples most strongly to the Higgs in the SM.

2.2.2 Signatures of supersymmetry at the LHC

Despite the fact that SUSY presents a promising extension to the SM there have been no significant experimental hints of its existence. For SUSY to convincingly solve the hierarchy problem, it is expected to present itself at the TeV energy scale, which is starting to be explored by the LHC.

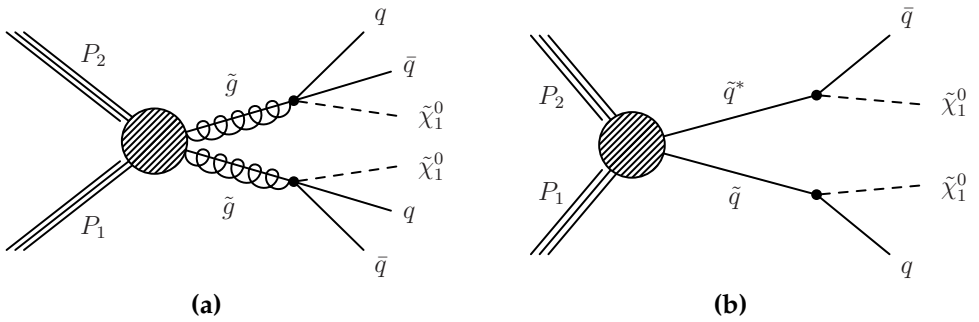


Figure 2.2: Representative SUSY production and decay of gluinos, (a), or squarks, (b), in proton-proton collisions. The SUSY particles decay to a weakly interacting neutralino, $\tilde{\chi_1}^0$, via SM quarks [43].

The fact that there are a lot of different free parameters in the MSSM leads to a wide range of possible SUSY phenomenologies. However, constraints made by R-parity and naturalness considerations help to limit the possible signatures. These constraints typically lead to the pair production of SUSY particles that rapidly decay via SM particles to a weakly interacting neutral LSP. As the produced SUSY particles have a reasonably high mass, this leaves a significant missing momentum signature in any detector that is looking for SUSY signatures.

As the LHC is a hadron collider, the highest cross section SUSY production processes occur via the strong force [19] [42]. These processes result in the production of pairs of squarks or gluinos, the SUSY particles with colour charge, that predominantly decay via an even number of SM quarks to an LSP. A Feynman diagram of typical squark and gluino production and decay is shown in Fig. 2.2. The resulting signature will usually be characterised by high energy hadronic jets of SM particles recoiling from invisible particles.

2.2.3 Simplified models

As there are a large number of parameters in the MSSM, there are many different versions of the theory that can all result in different phenomenologies. One of the things that has the biggest effect on searches for SUSY production is the mass hierarchy. To help to rationalise this the results of SUSY searches are typically interpreted within the context of *simplified models* [44, 45]. These models utilise a simplified model spectra (SMS) framework that typically consider only one possible SUSY decay per model. The simulated *parent* SUSY particles and the *daughters* that they decay to are kept at a

reasonable mass scale. All other SUSY particles are assumed to be at a much higher scale, such that they have no effect on the decay in question. An example of a decay targeted by a simplified model with gluino parents and neutralino daughters is in Fig. 2.2a.

Searches for SUSY are typically interpreted in different simplified models with a wide range of parent and daughter masses. In the case that no SUSY signature is observed, the simplified models that are excluded allow limits to be set on the mass of different SUSY particles. As the models separate out different possible SUSY production and decay mechanisms, searches will typically interpret their results in different models. Searches that look in a purely hadronic final state, for example, will consider simplified models that do not have leptons in the final state.

When the parent and daughter particles are very close in mass, the model is described to have a *compressed spectrum*. In these models the SM decay products usually carry very little energy. Searches for these types of models therefore rely on initial state radiation (ISR) or final state radiation (FSR) recoiling from the invisible SUSY system.

2.2.4 Status of experimental searches for supersymmetry

Signatures of SUSY have been extensively searched for at previous particle collider experiments including HERA (ep) [46, 47], LEP ($e\bar{e}$) [48] and the Tevatron ($p\bar{p}$) [49]. These were then surpassed in most regions of phase space by searches with the CMS and ATLAS experiments with the LHC during Run 1. The degree to which the 8 TeV Run 1 results have excluded the mass scales of different SUSY particles within a simplified model context can be seen in Fig. 2.3.

For SUSY at the electroweak scale to exist in a sensible form that convincingly solves the problems described in Sec. 2.1.6 it should be observable during Run 2 of the LHC, which probes even higher energies with 13 TeV proton collisions. If no evidence presents itself, the typical assumptions made when building SUSY models will have to be reassessed.

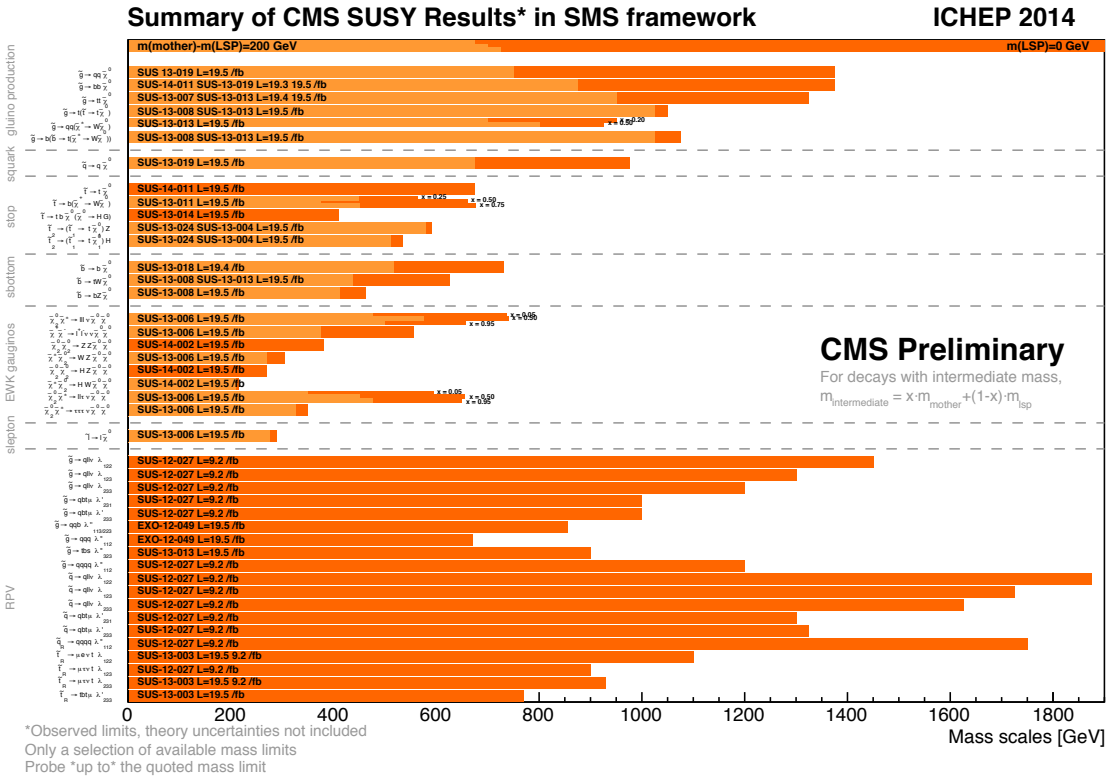


Figure 2.3: Maximum exclusion limits for the mass of SUSY parent particles for simplified models with an LSP of 0 GeV mass (dark shades) or parent mass minus LSP mass of 200 GeV (light shades) for a representative selection of simplified model topologies. The results were made with data collected by CMS with 8 TeV centre of mass proton-proton collisions [43].

Chapter 3

The CMS experiment at the LHC

The LHC is one of the largest machines ever built. It exists within a 27 km circumference tunnel $\sim$100 m underground on the border between France and Switzerland. Around the ring of the LHC are built a series of detectors that can record in a high level of detail the result of particle collisions produced by the collider. This chapter will focus on the details and performance of CMS, a multi-purpose detector optimised to search for new, as yet undiscovered, particles.

3.1 The LHC

The LHC is a hadron collider designed to collide protons and lead ions at centre of mass energies up to 14 TeV, the highest ever achieved by such a machine [7, 50–52]. The proton-proton collisions are most useful in direct searches for new physics and therefore take up the vast majority of the running time of the LHC.

To bring protons up to the 6.5 TeV required for $\sqrt{s} = 13$ TeV collisions, they are accelerated through a series of stages. Hydrogen atoms are initially stripped of their electrons and accelerated to 50 MeV by Linear Accelerator 2 (LINAC2). The energy is then increased to 1.4 GeV by the Proton Synchrotron Booster (PSB) before being injected into the Proton Synchrotron (PS) which boosts the energy up to 26 GeV. A final kick up to 450 GeV is provided by the Super Proton Synchrotron (SPS). This chain of accelerators also collects the protons into bunches that are either 25 ns (from Run 2 onwards) or 50 ns apart (during Run 1 and early stages of Run 2). These bunches are then injected into the LHC, in which they are steered by around 1200 superconducting dipole magnets while being accelerated up to 6.5TeV with radio frequency (RF) cavities.

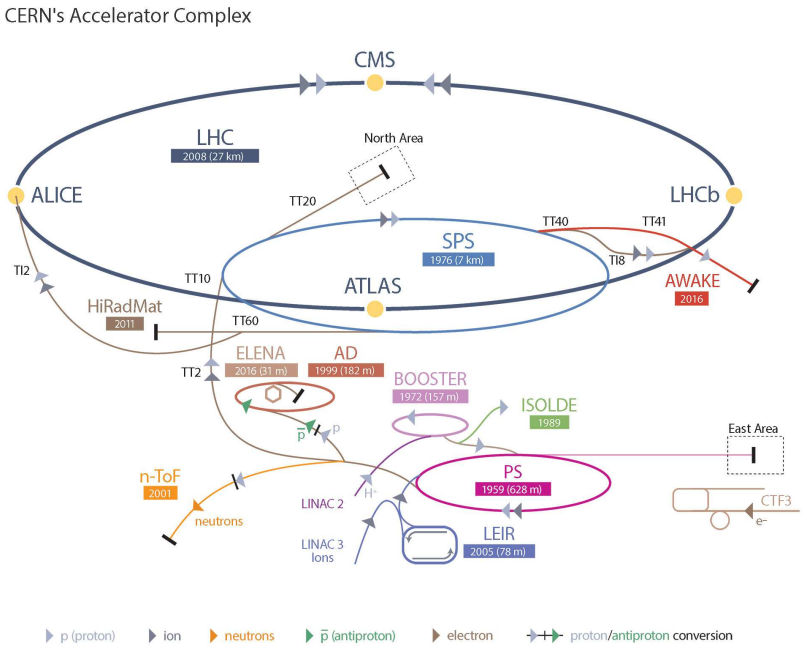


Figure 3.1: A representation of the CERN accelerator complex that accelerates hadrons to high energies within the LHC [56]

Once the beam has reached the intended energy and is stable, protons are collided at four different points on the ring, around which are built the four major LHC detectors, ALICE [53], ATLAS [54], LHCb [55] and CMS [8]. A representation of this accelerator complex and the location of the detectors can be seen in Fig. 3.1.

As well as attaining record breaking energies, the LHC is designed to collide hadrons at a very high luminosity, with a bunch collision rate of up to 40 MHz [7]. This is necessitated by the fact that the rate at which electroweak scale processes occur in proton collisions is significantly lower than their associated backgrounds, demonstrated in Fig. 3.2. The LHC was therefore designed to run at an instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ to maximise the occurrence of these rare processes. Along with the high collision rate, this luminosity is achieved by squeezing the proton bunches to increase the number of simultaneous collisions per bunch crossing, the extra simultaneous collisions are known as pileup (PU). The LHC has typically operated with a PU of $\sim$10-20, however to increase the luminosity in the future this value will be increased up to a PU of O(100).

During Run 1 of the LHC, from 2010-2013, a total of 23.3 fb^{-1} of data were collected at centre of mass energies of $\sqrt{s} = 7$ TeV and 8 TeV. After this there was a period

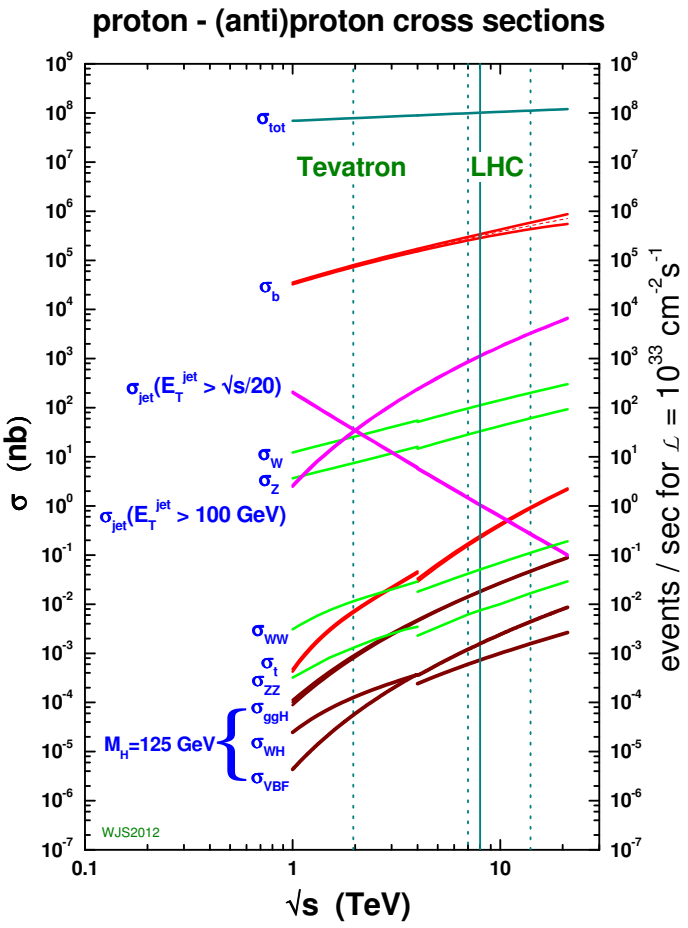


Figure 3.2: The cross sections for various standard model processes as a function of proton collider energy, demonstrating the importance of high luminosities when observing electroweak scale processes [57].

of shutdown in which the LHC and the detectors underwent a series of upgrades. Run 2 then began in 2015 with the collision of protons at $\sqrt{s} = 13$ TeV. During 2015 a total integrated luminosity of 4.3 fb^{-1} was collected at this energy. In 2016 the LHC delivered 34.6 fb^{-1}, a record breaking number of collisions at the highest energy ever recorded.

3.2 The CMS detector

The CMS detector is one of two multipurpose detectors built around proton beam collision points, the other being ATLAS. It is situated at Point 5 on the LHC, as visibile in Fig. 3.1. The key goals of the CMS detector at its conception were the discovery of the SM Higgs boson and searches for generic signatures of BSM physics. In CMS the results of collisions are measured with a series of subdetectors, built within and around a 3.8 T superconducting solenoid. They are designed to track, identify and record the energy of all non-neutrino SM particles [58]. With its comprehensive solid angle coverage, CMS is well suited to inferring the existence of weakly interacting particles through the momentum imbalance of visible particles. This is particularly relevant when searching for BSM physics.

A representative view of CMS and its components can be seen in Fig. 3.3. The detector is designed in a series of cylindrical layers of subdetectors working out from the central point, where the proton collisions occur. The first layer consists of the silicon tracking system. This tracker is designed to allow the reconstruction of the trajectory of charged particles produced in the collision point as they move through the magnetic field. The degree to which the path of these particles is bent allows for an accurate determination of their momenta. The next layer beyond the silicon tracker is the electromagnetic calorimeter (ECAL), which is designed to absorb and measure the energy of electrons and photons. Surrounding this is the hadron calorimeter (HCAL) that absorbs the remaining hadronic particles that have punched through the ECAL. Built around the tracker and calorimeters is the superconducting solenoid. In the final layer are the muon chambers and iron return yoke. The chambers are designed to detect the presence of muons, which will not be absorbed by the central components of the detector. The data from all these subdetectors are read out by dedicated front end electronics and passed through the CMS trigger system, which selects the most promising data to be kept and stored for offline processing.

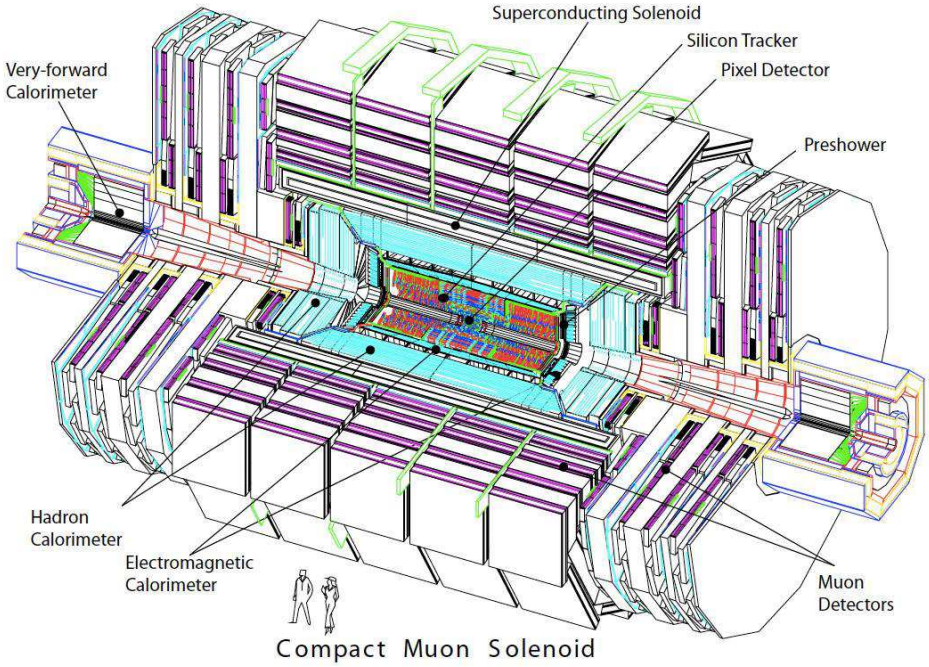


Figure 3.3: An internal view of the CMS detector highlighting the key detecting components [58]

Measurements of physical quantities made by CMS are typically interpreted in a three dimensional coordinate system that originates from the centre of the detector. The x-axis points to the centre of the LHC ring, the y-axis points vertically upwards and the z-axis points along the direction of the LHC beam pipe. It is then helpful to define the azimuthal angle, ϕ, which is in the x-y plane and measured relative to the x-axis. Measurements of momentum and energy in this plane are described as transverse and known as p_{T} and E_{T} respectively. The polar angle, θ, is then defined as relative to the z-axis. This angle is used to construct the pseudorapidity, defined as $\eta = -\ln(\tan(\theta/2))$. Distances in the η-ϕ plane are then given as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

3.2.1 The tracker

The CMS inner tracking system is designed to accurately determine the trajectories of charged particles produced in hadron collision events [59]. In the presence of the strong magnetic field provided by the CMS solenoid, the curvature of these tracks can be used to reconstruct momenta with a resolution between 1.5% and 3% for $p_T \sim 100$ GeV charged particles. The tracker is also capable of tracking $p_T > 1$ GeV charged particles with an efficiency greater than 99% [58]. Along with this, the spatial

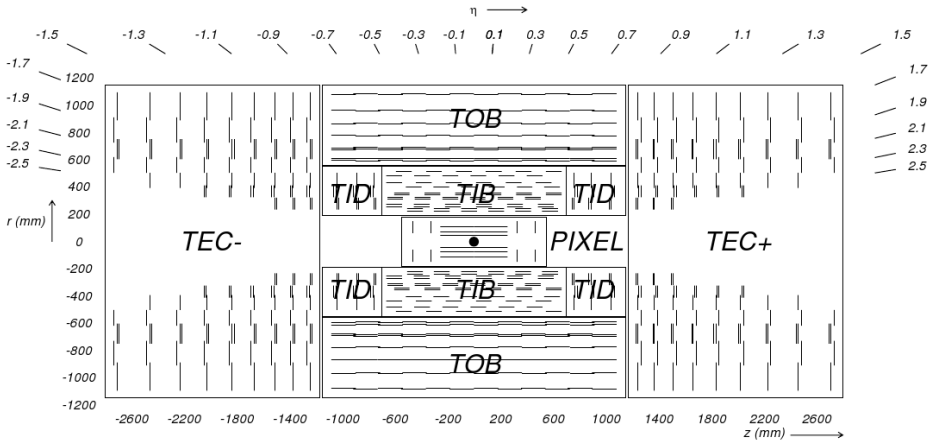


Figure 3.4: A schematic of a cross section through the CMS tracker. Detector modules are represented by the isolated black lines [8]

resolution of the tracker is such that the points of origin of event decay products can be inferred within 10 μm. This allows for the independent identification of individual proton collisions with a high level of efficiency, meaning the high performance of CMS can extend up to high levels of PU.

The tracker is required to operate in a challenging, high radiation, environment. Additionally, in an ideal detector the particles produced in collisions are solely absorbed by the calorimeters. The tracker must therefore consist of as little material as possible. To achieve the high level of precision and fast response time required given these conditions, the tracker makes use of silicon technology. As charged particles pass through doped silicon an electron-hole pair is produced. In the presence of an electric field this gives rise to a pulse of electrical current in the previously resistive silicon. This behaviour is utilised by the tracker in a series of silicon pixel and strip detectors covering all angles in ϕ and extending up to $|\eta| < 2.5$. The layout of the tracker is shown in Fig. 3.4.

The pixel detector is the high granularity component of the tracking system that sits closest to the interaction point, covering the pseudorapidity region $|\eta| < 2.5$. It consists of three cylindrical layers of hybrid pixel detector modules that are complemented by two disks of pixel modules on each side. The pixel detector makes use of 66 million pixels covering an area of ~ 1 m^2 to give the tracker its excellent spatial resolution of 15-20 μm in both the r-ϕ and z direction. This resolution is essential for a precise determination of the position of collision vertices and for the observation of vertices

displaced from this origin that can be used to identify particles such as hadrons containing b-quarks.

Surrounding the pixel detector is the silicon strip tracker that covers the region up to $|\eta| = 2.4$. It consists of three different subsystems built from 9.6 million silicon strips that cover an area of 198 m^2. Each of these strips are 10-20 cm long and 80-180 μm wide. Working out from the centre the subsystems are the Tracker Inner Barrel and Disks (TIB/TID), the Tracker Outer Barrel (TOB) and the Tracker End Caps (TEC). They are arranged in a geometry that maintains a good degree of coverage across all angles and can be seen in detail in Fig. 3.4. Along with the spatial resolution provided by the pixel detector the silicon strip tracker adds enough modules to reconstruct the trajectory of particles to the required high level of precision.

3.2.2 The electromagnetic calorimeter

The ECAL is constructed from $\sim$75 848 lead tungstate ($PbWO_4$) scintillating crystals covering the region $|\eta| < 3$ [60]. It is designed to absorb electrons and photons and emit light proportional to the energy deposited. The light is then detected by custom photodiodes that perform well in high magnetic fields. This is achieved in a way which is fast, radiation resistant and with a high granularity.

The ECAL is divided into the ECAL barrel (EB) which covers the region $|\eta| < 1.479$ and the ECAL endcaps (EE) which cover the region $1.479 < |\eta| < 3.0$. Additionally, built just before the EE in the region $1.653 < |\eta| < 2.6$ is the Preshower. Unlike the other components, which are predominantly $PbWO_4$ crystals, the Preshower is a lead and silicon sampling calorimeter. Its main aim is to improve the position resolution of particles in the forward direction and help to distinguish collinear π^0 decays from high energy photons [8]. The layout of the ECAL can be seen in Fig. 3.5.

As high-energy electrons or photons enter one of the crystals in the ECAL they initiate an electromagnetic shower. This results in a cascade of lower energy particles that undergo bremsstrahlung and pair production. These charged particles ionise atoms in the crystal which then emit scintillation light as they de-excite. As the crystals are transparent, this light can be measured by avalanche photodiodes and vacuum phototriodes which convert it into an electronic current. The magnitude of this current is proportional to the energy deposited in the crystal and can be used to accurately infer the total energy deposited. Irradiation of the crystals decreases their transparency

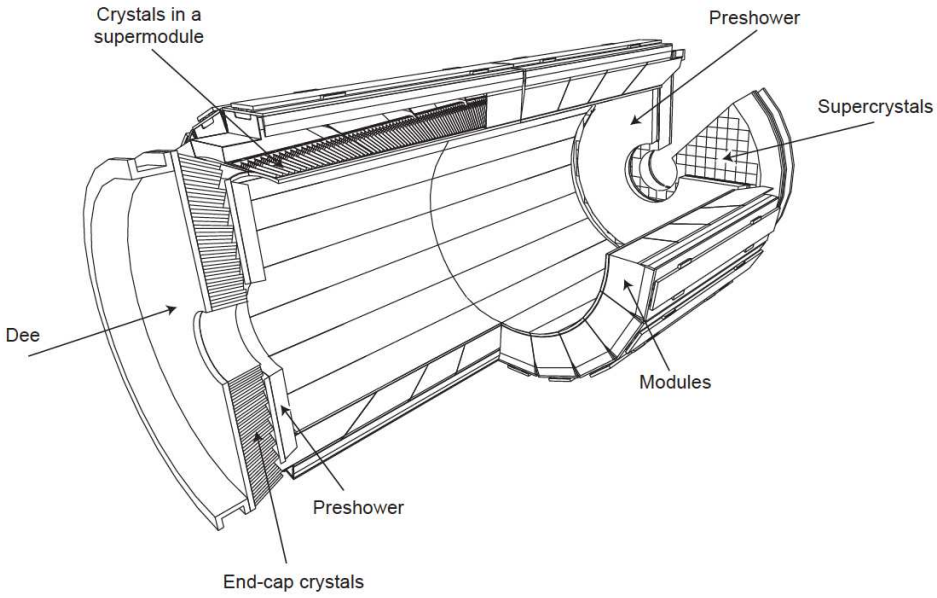


Figure 3.5: A cutaway diagram of the CMS ECAL. All the key components, including the barrel and endcap crystal layouts, are displayed [8]

over time. To counteract this a time dependent calibration is carried out for each of the crystals with a laser of wavelength $\lambda = 440$ nm.

Measurements in a test beam in the absence of a magnetic field have measured the $PbWO_4$ crystals to have a resolution, σ_E, given by the following formula [61]:

$$\left(\frac{\sigma_E}{E[\text{GeV}]}\right)^2 = \left(\frac{2.8\%}{\sqrt{E[\text{GeV}]}}\right)^2 + \left(\frac{12\%}{E[\text{GeV}]}\right)^2 + (0.30\%)^2. \tag{3.1}$$

In this equation E denotes the energy of the incident particle. The first term encapsulates uncertainties from fluctuations in the scintillation light. The second term takes account of noise in the electronics and digitisation. The final term then covers any non-uniform longitudinal response or inter-calibration errors.

3.2.3 The hadronic calorimeter

The final layer of calorimetry in CMS is the HCAL [62], designed to absorb hadrons that have passed through the ECAL. It is a sampling calorimeter that is constructed from brass absorbers interleaved with scintillating plastic tiles covering $|\eta| < 3$. The scintillation light is read out with hybrid photodiodes via wavelength shifting fibres.

Additionally, the hadronic calorimetry is extended up to $|\eta| = 5.2$ with the hadron forward (HF), made from steel absorber with quartz scintillating fibre.

The brass absorbers in the HCAL are arranged in plates that are interspersed with plastic tiles. Incident particles induce hadronic showers in the brass layers that produce scintillating light in the plastic. This light is collected by wavelength-shifting fibres which transfer the signal to on-detector amplifiers for read-out. Brass has the advantage of being non-magnetic and having a short nuclear interaction length of 16.42 cm. In the HF, steel replaces the brass and quartz fibre replaces the plastic due to the very high radiation environment present in the forward region of the detector. The total time to collect a HCAL signal pulse is large with respect to the collision rate of the LHC, only 68% of the pulse is collected within 25 ns. This leads to cases of out of time pileup (OOTPU), where signal from a bunch crossing can influence the read out of future bunch crossings. This is also an issue for the read-out from the ECAL.

The subdetectors that make up the HCAL along with the HF can be seen in Fig. 3.6. Within the solenoid is the hadron barrel (HB), which extends up to $|\eta| = 1.3$, and the hadron endcaps (HE), covering $1.3 < |\eta| < 3.0$. They are segmented in η-ϕ towers with a size of 0.087×0.087 in the HB varying up to 0.17×0.17 in some areas of the HE. The HB towers are lined up with 5×5 arrays of ECAL crystals and each read-out individually. On its own the HB provides between 5.8 and 10.6 interaction lengths of absorber while the HE provides ~ 10 interaction lengths. Beyond the magnet coil is the hadron outer (HO) that increases the interaction length in line with the HB to a minimum of 11.8. The magnet coil acts as an additional absorber for the scintillators in the HO that absorbs any late-starting or highly-penetrating showers.

The combined resolution, σ_E, of the ECAL and HCAL when considered together have been measured in a test beam as [63]:

$$\left(\frac{\sigma_E}{E[\,\mathrm{GeV}]}\right)^2 = \left(\frac{84.7 \pm 1.6\%}{\sqrt{E[\,\mathrm{GeV}]}}\right)^2 + (7.4 \pm 0.8\%)^2. \tag{3.2}$$

In this equation E is the energy of the incident particle. An in-situ calibration is also performed with a UV laser and $^{137}\mathrm{Cs}/^{60}\mathrm{Co}$ sources that can be inserted into the scintillation tiles in the HB and HE.

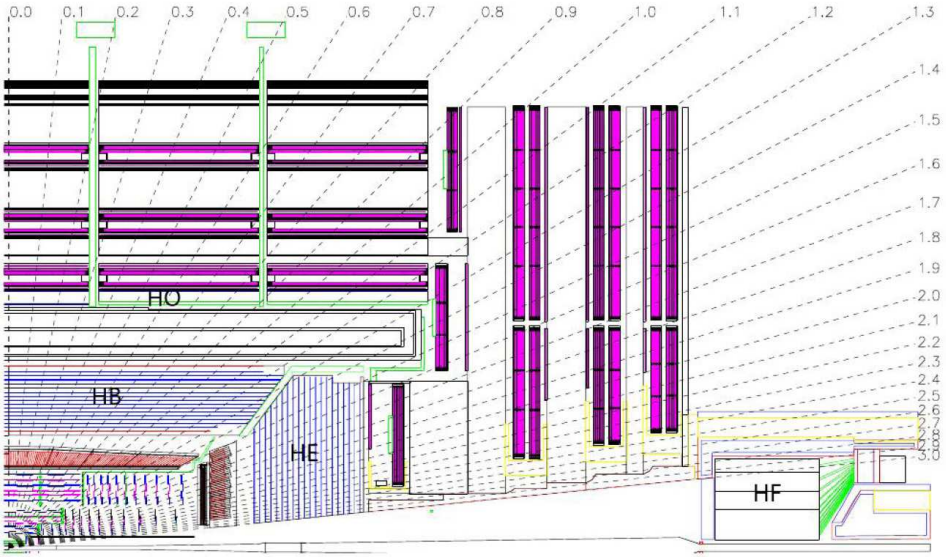


Figure 3.6: A schematic of a quadrant of the CMS HCAL. The locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters are displayed [8]

3.2.4 The muon system

As muons are heavier than electrons, they are minimally ionising and lose little energy through bremsstrahlung. They therefore mostly pass through the ECAL and HCAL. As muons are a key component of many electroweak decays, CMS has a dedicated muon system interleaved with the iron return yoke surrounding the solenoid. This muon system consists of wire chambers containing ionising gas that, in combination with the tracker, allows the measurement of muon momenta with a better than 1% precision [64].

The layout of the muon system can be seen in Fig. 3.7. It consists of the drift tube (DT), resistive plate chamber (RPC) and cathode strip chamber (CSC) subsystems. They each utilise different gaseous chamber technologies to perform measurements of muons in different operating regions of CMS. In the barrel region the DT chambers cover the range $|\eta| < 1.2$ and are filled with a mixture of Ar and CO_2 gas. As a muon passes through the chamber this gas is ionised. Within the chambers are $\sim 172\,000$ wires, each 2.4 m long, with a potential difference applied across them. Free electrons produced in the ionisation drift towards the anode of the wires and induce an electrical signal which is read-out.

Within the endcap region the CSCs cover the range $0.9 < |\eta| < 2.4$. The CSCs are required to have a fast response time and be radiation hard due to the higher

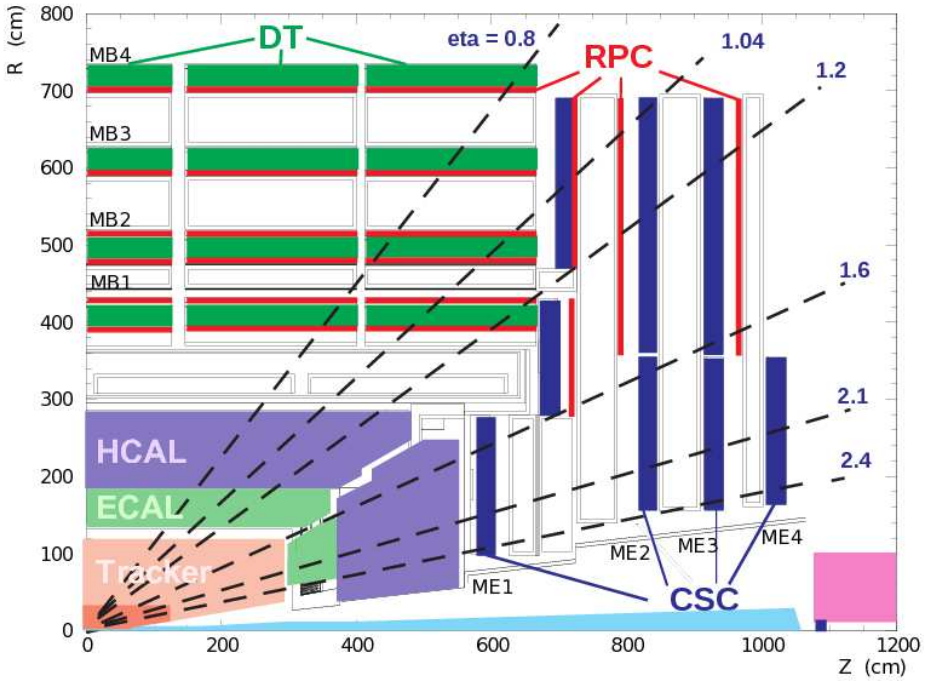


Figure 3.7: A schematic of a quadrant of the CMS muon system. The locations of the Drift Tube (DT), Resistive Plate Chamber (RPC) and Cathode Strip Chamber (CSC) subsystems are displayed [65].

muon and background rates in the forward region of CMS. They consist of chambers containing an Ar, CO_2 and CF_4 gas mix around anode wires sandwiched by copper cathode strips.

The RPCs cover the range $|\eta| < 2.1$ and augment the behaviour of the DTs and CSCs. They consist of oppositely charged bakelite plate electrodes separated by a 2 mm gas gap. They have a much poorer position resolution than the other subsystems, 0.8 - 1.2 cm as opposed to 40 - 150 μm. However, they have a much better temporal resolution of $\sim$3 ns. This allows them to be used as an independent muon trigger that can identify the bunch crossing in which the muon originates.

The muon systems on their own provide an energy resolution of 9-11% for muons with $p_T < 200\,\text{GeV}$ and $|\eta| < 2.4$. However, better performance is obtained by combining muon chamber hits with the tracker as described in Chapter 4. The muon system also has excellent charge identification with a misassignment rate of $< 0.1\%$ for $p_T < 100\,\text{GeV}$ muons.

3.2.5 The trigger and data acquisition system

At 40 MHz, the rate of collisions at the LHC is so high that it would be too computationally expensive to reconstruct and store the ~ 1 MB detector read-out for every bunch crossing. As the majority of proton collisions at the LHC are soft scattering QCD processes, they are not useful in the search for new physics at the electroweak energy scale. This necessitates a multi-level trigger system that is designed to pick out and store only "interesting" high centre-of-mass physics processes [66, 67]. Once events have been selected by the trigger system the chosen bunch-crossings can be stored on tape for offline analysis. They are then centrally reconstructed through the GRID computing infrastructure and used for high-level physics analysis [68].

The Level-1 Trigger (L1T) is the first component of the two layer trigger system and is made from custom field programmable gate array (FPGA) computational boards situated close to the detector. As the data from a bunch crossing comes out of CMS it can only be stored in the read-out pipelines a total of < 4 μs, of which ~ 2 μs is required for transmitting data to and from the detector. The L1T must therefore make a decision on this time scale using coarse information from the calorimeters and muon system, with the aim to reduce the event rate to ~ 100 kHz. The FPGA technology allows for these decisions to be made with very low latency times.

A flowchart showing how the data from the different detector subsystems passes through the L1T can be seen in Fig. 3.8. At the Global Trigger (GT) a decision is made whether to pass the event to the next triggering stage. The calorimeter trigger takes calorimeter deposits at the tower level and coarsely reconstructions physics objects. Events with high total transverse or missing transverse energy are typically accepted. More details of the calorimeter trigger utilised in Run 2 are given in Chapter 5. The muon trigger looks for the presence of muons with a reasonably high energy to aid the trigger decision.

After an event has been accepted by the L1T it is passed to the high-level trigger (HLT), which uses full detector information to reconstruct the events and reduce the data rate to $\sim$1kHz on a high-performance computing farm [69]. With the reduced event rate provided by the L1T, the HLT has a larger latency budget but the algorithms used are still limited in their complexity compared to those used for offline analysis. The HLT can be tailored towards targeting particular physics processes in a way that would not be possible at the L1T.

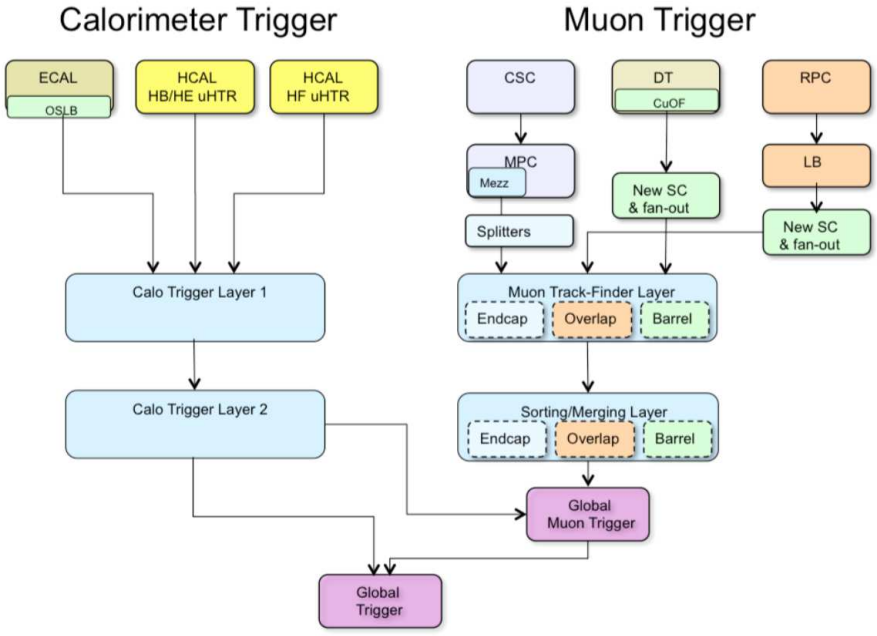


Figure 3.8: Data-flow of the Level-1 trigger used to collect data in Run 2 of the LHC [66].

Events that are finally accepted by the HLT are then transmitted to the Tier-0 data centre, located at CERN, for permanent storage and event reconstruction. These data are distributed and stored around the world through a GRID computing infrastructure [68]. This infrastructure has a four tier structure that represents the importance and accessibility of the user at each site. From the Tier-0, data are transferred to 13 Tier-1 computing centres. These data are then available to Tier-2 and Tier-3 sites where offline data analysis can be performed by different members of the CMS collaboration anywhere on the planet.

Chapter 4

Event reconstruction and simulation

After the results of proton collisions have been observed in the various subdetectors that comprise CMS, the physical objects produced in each collision event are reconstructed. Through this reconstruction, an overall analysis of the underlying physical process can be carried out. To aid in classifying the types of collision events, Monte Carlo (MC) simulations are also carried out. The reconstruction algorithms and simulations that are relevant to the search for supersymmetry presented in this thesis are described in this chapter.

4.1 Tracks and vertices

As charged particles pass through the CMS tracker they leave energy deposits, known as *hits*, in each layer. These hits are reconstructed as tracks with the combinatorial track finder (CTF) algorithm [70], which tries to associate hits that belong to a single charged particle. This allows for the determination of the path taken by the charged particle, its *track*. The curvature of the particle as it passes through the magnetic field is then used to determine its momentum. The steps of the algorithm are as follows:

- Initially two or three hits in the inner layers of the tracker are taken as seeds for initial track candidates. Quality criteria are applied on the selected hits, which retain only promising seeds.
- Each seed is extrapolated along the expected trajectory using a Kalman filter [71] with a helical tracking hypothesis. This allows the seeds to be associated with hits in an outer tracker layer while taking account of uncertainties in the measurement.

- The extrapolation is carried out recursively into the subsequent tracker layers until the outer-most layer is reached.
- Once the tracking candidates are found, additional quality criteria are required to reject fake tracks.
- This series of steps is repeated up to six times and the hits associated with identified tracks are removed after each iteration.

Track reconstruction efficiencies for a variety of charged particles are shown in Fig. 4.1 as a function of p_{T} and η. In the central region of the detector all particles with a p_{T} from 10 to 100 GeV are reconstructed with a 90-100% efficiency. In the forward detector region this efficiency remains above $\sim 80\%$.

After charged particle tracks are reconstructed, they can be used to infer the positions of the different proton-proton collisions in the event, known as interaction vertices. Tracks are required to originate from a region that is compatible with the LHC *beamspot*, the area in which the proton beams cross and collisions occur. The z-coordinates of tracks at the closest point of approach to the beamspot are taken as an input for the deterministic annealing clustering algorithm [72]. This finds the most probable vertex positions and assigns each track to a vertex. The final x,y,z position of these vertices is then found using the adaptive vertex fitter [73]. This assigns the most probable position of the vertex for the given set of input tracks. Quality criteria are then applied to reject fake vertices, they are chosen in such a way as to remain efficient for real vertices that typically have a large number of tracks compatible with them. Finally, the *primary vertex (PV)* is determined as the vertex with tracks that have the greatest scalar sum of p_{T}. Other vertices are then initially attributed to PU. However, vertices that are displaced from the initial proton collision are common signatures of unstable particles that decay within the detector, such as b-hadrons. These can be found in subsequent levels of reconstruction.

The efficiency for finding the PV as a function of the number of tracks in a cluster can be seen in Fig. 4.2. The efficiency increases to close to 99.9% for events with ≥ 3 tracks. The vertex resolution reaches ~ 10 μm in x, y and ~ 15 μm in z with > 40 reconstructed tracks.

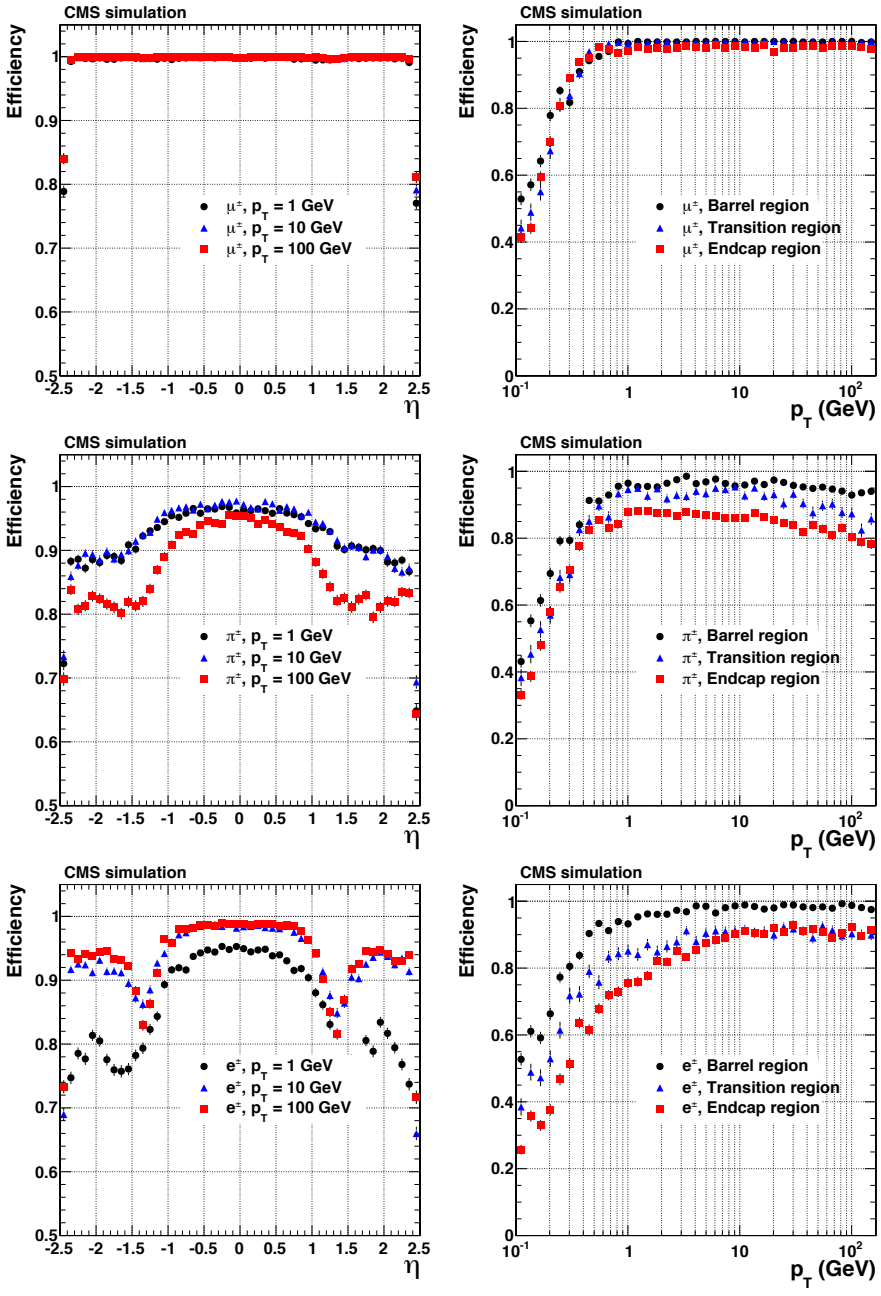


Figure 4.1: Efficiencies of track reconstruction for different charged particles as a function of p_{T} and η. Muons are shown at the top, pions in the middle and electrons at the bottom. The barrel, transition and endcap regions are defined by the η intervals of 0-0.9, 0.9-1.4 and 1.4-2.5 respectively. For all the tracks *high-purity* quality requirements are made [70]

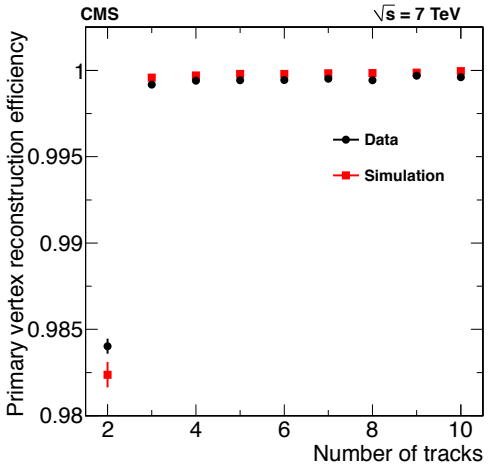


Figure 4.2: The vertex reconstruction efficiency as a function of the number of tracks originating from the vertex. Measured in data and simulation for $\sqrt{s} = 7$ TeV proton collisions. [70]

4.2 Particle flow

Each of the subdetectors of CMS provide complimentary information about the different types of particles that pass through the detector. This is exploited to identify the different types of particle with the particle flow (PF) algorithm [74–76]. As CMS has accurate momentum resolution in the tracker and a high granularity ECAL this algorithm allows both to augment the measurement of objects in the HCAL. This then allows calibrations that are specific to both charged and neutral hadrons to be applied.

The PF algorithm searches for a set of individual particles that are known as PF *candidates*. They are then classified as charged or neutral hadrons, photons, muons or electrons. The set of PF candidates can then be utilised to calculate other event level variables, such as for jet reconstruction described in Sec. 4.5.

The algorithm starts by taking the tracks, which are reconstructed as in Sec. 4.1, along with clusters of energy in the calorimeters, which are reconstructed separately in the ECAL and HCAL. Clusters are paired with tracks if the track trajectory is compatible with the cluster position. These pairs are then used to identify charged particles. Electrons and hadrons are typically differentiated based on the proportion of energy they deposit in the ECAL or HCAL. Electrons will deposit nearly all their energy in the ECAL whereas hadrons will deposit much more in the HCAL. A similar pairing between tracks and hits in the muon system is used to identify muons.

Neutral particles are then identified through calorimeter clusters that do not have a compatible track associated to them. Photons, for example, will leave a significant ECAL deposit with no track, which would be present for electrons. Similarly, neutral hadrons will leave significant deposits in the HCAL with no associated track.

4.3 Electrons and photons

Electrons and photons interact with the ECAL in a similar way. The reconstruction techniques used for both are therefore very similar, with the main difference being the lack of a track in photon reconstruction. As electrons interact with the tracker, they lose on average 33% of their energy before reaching the ECAL [77]. Most of this energy loss occurs through bremsstrahlung, which has a non-Gaussian loss distribution. As the Kalman filter that is used in track reconstruction assumes Gaussian energy losses, another specialist track reconstruction for electrons is employed, known as the Gaussian Sum Filter algorithm [78]. To maintain a good energy resolution, the photons that are produced during bremsstrahlung must also be properly associated with the electron when clustering in the calorimeter. As the electrons bend in the magnetic field but the photons they emit do not, the calorimeter clusters are allowed to extend along the azimuthal direction. These rectangular ECAL windows are know as *superclusters*.

To form the superclusters in the barrel the *hybrid* clustering algorithm is used. A single seed crystal with a local maximum transverse energy $E_T > 1$ GeV is identified and a rectangular configuration of 3×1 or 5×1 crystals in η-ϕ is formed around it. The algorithm then looks in the ϕ region adjacent to this rectangle up to $\Delta\phi \pm 0.3$. Additional rectangular regions that have $E_T > 100$ MeV are kept and grouped to form the supercluster. In the endcaps the *multi* 5×5 algorithm is used instead, which aggregates 5×5 arrays of crystals within $\Delta\eta < 0.07$ and $\Delta\phi < 0.3$ of the seed.

Electrons are identified through a supercluster matched to a track, or photons are identified through an unmatched supercluster. After this, additional selection criteria are applied to suppress backgrounds. The major backgrounds come from misreconstructed hadronic jets or semi-leptonic decays of heavy quarks. To suppress these backgrounds, cuts are made on the ratio of energy deposits in the HCAL and the ECAL, the difference in direction between the track and supercluster (in the case of electrons) and the width of the cluster, which is larger for hadrons.

4.4 Muons

Unlike electrons and photons, muons are minimally ionising so do not lose much of their energy to the tracker or calorimeters [8]. Muon reconstruction therefore uses a combination of tracks and hits in the muon system. To maintain high efficiencies across a wide range of muon energies, two algorithms are utilised [79]. The *global* muon algorithm functions in the same way as the PF algorithm, mentioned in Sec. 4.2. To maintain efficiency for low energy muons that have a lower probability of traversing the entire muon system the *tracker* muon algorithm is additionally used.

The global muon algorithm starts with hits in each layer of the muon system that give an initial estimate of the position, energy and direction of candidate muons. It then searches for tracks in the inner detector that are compatible with this candidate. If this is the case, the track is extrapolated to the hits in the muon system using a Kalman filter, similar to the way tracks are reconstructed in Sec. 4.1.

The tracker muon algorithm starts with tracks with $p_{\mathrm{T}} > 0.5$ GeV and extrapolates them into the muon system, again with a Kalman filter. If any of these tracks can be matched to at least one muon chamber hit it is considered to be a muon. For muons with $p_{\mathrm{T}} < 200$ GeV the tracker provides the highest resolution energy measurement, while the muon system provides a better resolution for the straighter, higher p_{T}, tracks.

To reduce the background from hadrons that have punched through the HCAL or muons that do not originate from the primary vertex, such as those from cosmic rays, a series of selections are applied to the reconstructed muon candidates. These include a quality requirement on the fit of the tracks, a minimum number of hits in the muon systems and a track compatible with originating from the beamspot. A different set of criteria can be applied to trade off the efficiency and fake rate of the muons. After the application of a *tight* set of criteria, the efficiency of muons with $p_{\mathrm{T}} > 10$ GeV is measured to be $> 96\%$, shown in Fig. 4.3.

4.5 Jets

As there is a very high probability that proton collisions will produce quarks and gluons, identifying and measuring them is a crucial part of object reconstruction in CMS. This is particularly relevant when searching for strongly produced SUSY

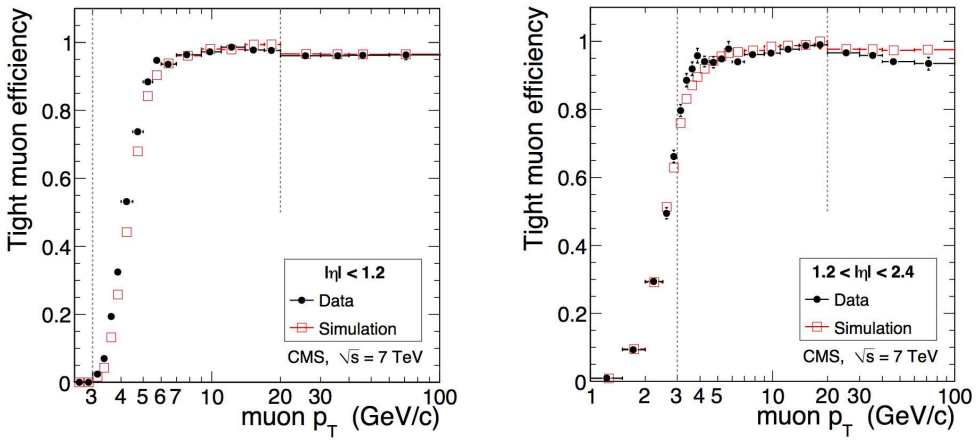


Figure 4.3: The tight muon reconstruction efficiency as a function of the p_{T} of the muon in $\sqrt{s} = 7$ TeV proton collisions. The efficiency is measured separately in the barrel (a) and endcap (b) regions. [79]

particles as they typically decay via hadrons to the LSP, as described in Chapter 2. Due to the nature of the strong force, high p_{T} quarks and gluons undergo immediate hadronisation. The result of this is a collimated shower of, predominantly hadronic, particles. To reconstruct the quarks and gluons produced in an event these collimated showers of particles are clustered and reconstructed as *jets* [80].

4.5.1 The anti-k_T clustering algorithm

When determining the clustering algorithm to be used, the theoretical behaviour of hadronisation must be taken into account [80]. Specifically, jets can undergo soft gluon radiation of arbitrarily low energy gluons with a high probability. Similarly, a gluon can split into two, almost collinear, gluons that share its energy. The final result of the clustering algorithm cannot depend on either of these things happening. It must be both *infrared safe* and *collinear safe*. The jet algorithms used in CMS are typically some form of a *sequential recombination algorithm* that fulfills the above criteria. After defining a *distance* between all pairs of particles in the event, d_{ij}, and a distance from each particle to the beamline, d_{iB}, the algorithm undertakes the following steps:

- Calculate d_{ij} for all pairs of particles in the event and d_{iB} for each particle.

- If a value of d_{ij} is smallest, combine the pair of particles with this distance into a single new particle and start again.
- If a value of d_{iB} is smallest remove this particle from the list of particles, classify it as a cluster and start again.
- Stop when there are no more particles remaining.

For the sequential recombination algorithms used by CMS the distance parameters used are defined as:

$$d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{Ti}^{2k} \tag{4.1}$$

for particles i and j where $k = -1, 0, 1$, ΔR is the separation in the η-ϕ plane and R is a fixed parameter that sets the jet size. The most commonly used of the CMS algorithms has $k = -1$ and is known as the *anti-k_T* algorithm [81]. The anti-k_T algorithm tends to cluster circular jets built around the hardest particles in a particular event. An area of $R = 0.4$ is often chosen as it contains the hadronic showers of most quarks and gluons produced in $13 - 14$ TeV proton collisions, while remaining insensitive to contamination from PU.

4.5.2 Jet identification

For the results presented in this thesis the FASTJET [82] package is used to cluster the PF candidates, that are described in Sec. 4.2. The anti-k_T algorithm is used with $R = 0.4$. The PF candidates add information from the tracker when calculating the momentum contribution to each jet from charged particles. As jets typically consist of 65% charged hadrons, 25% photons and 10% neutral hadrons, this gives a significant improvement over clustering the calorimeter deposits on their own.

To reject fake jets from background sources or detector noise, additional selection is applied to jet candidates. The *loose* set of criteria for jets provides an 84% suppression of fake jets while maintaining a $> 99\%$ efficiency for real jets. It requires at least two PF candidates; $< 99\%$ of the jet to come from only hadrons, photons or electrons; and at least one charged track.

The energy resolution of jets produced in 13 TeV proton collisions for different η ranges can be seen in Fig. 4.4. For jets with a $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ the resolution is

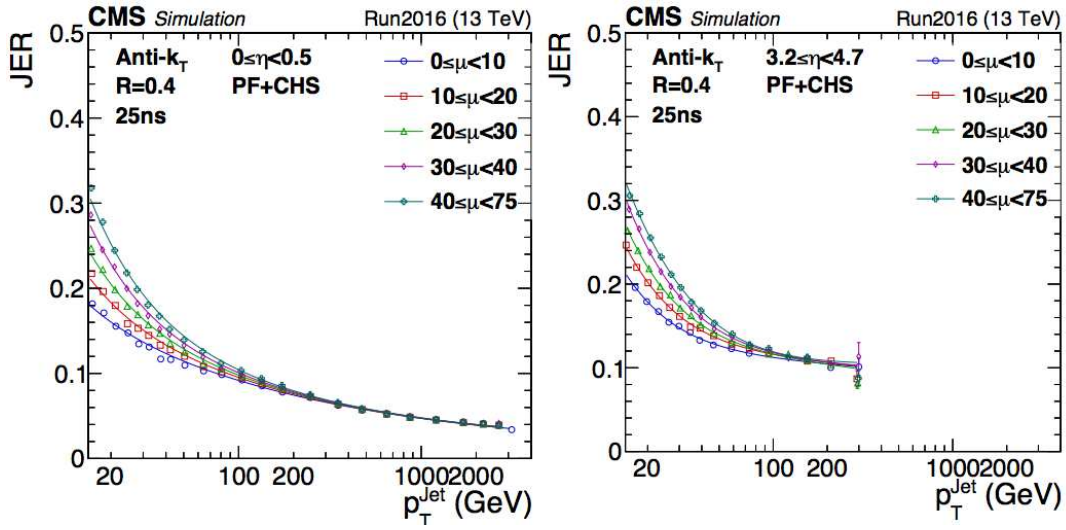


Figure 4.4: The jet energy resolution (JER) as a function of jet p_{T} for the central (left) and forward (right) detector regions [86].

typically better than 10%. Further information on the jet performance can be found in references [83–85].

4.5.3 Jet energy corrections

Due to the imperfect resolution of the CMS detector, the initial measured transverse momentum, p_T^{raw}, of a jet does not necessarily match the energy of the particle that initiated the jet. The difference between the *raw* p_{T} and the *true* p_{T} typically depends on the p_{T} of the jet and its position in the detector, along with components of the jet that are measured by different subdetectors. To correct the raw p_{T} of the jet, a correction is applied with the functional form [87]:

$$p_T^{cor} = C^{Off}(p_T^{raw}, \rho, A_j) \cdot C^{Rel}(\eta) \cdot C^{Abs}(p_T^{Rel}) \cdot C^{Res}(p_T^{Abs}) \cdot p_T^{raw}. \tag{4.2}$$

The first step is the offset calibration, C^{Off}, which performs a correction that removes the effect of other jets originating from PU in the jet cone. Initially, the contribution made by charged particles that do not originate from the primary vertex is removed. This process is known as charged hadron subtraction (CHS). The remaining contribution from neutral PU particles is corrected using jet area subtraction [88]. In this case the jet area, A_j is multiplied by the average pileup energy density, ρ. The value of ρ is determined from the energy density of PF candidates across the entire

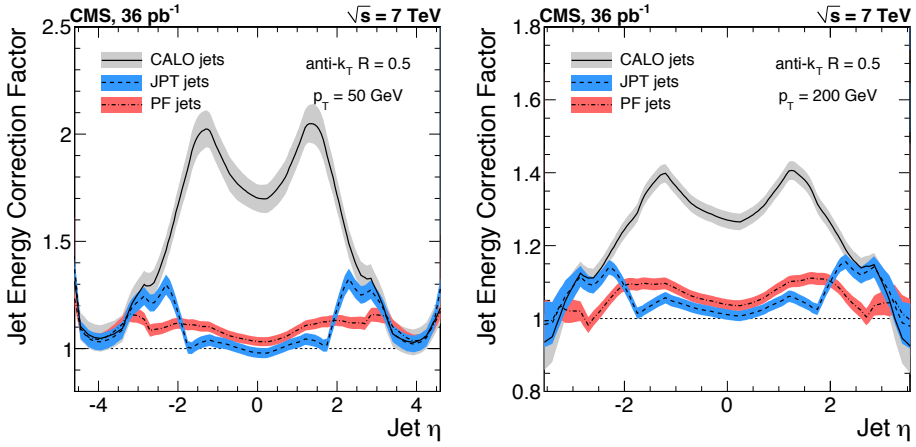


Figure 4.5: The jet energy correction factors and their corresponding uncertainty as a function of jet η for jets with $p_{\mathrm{T}} = 50$ GeV (left) and jets with $p_{\mathrm{T}} = 200$ GeV (right) for different types of jet reconstruction. The correction for jets reconstructed with PF candidates is shown in the red line, the CALO label indicates jets reconstructed purely with calorimeter deposits and no tracker information [87]

detector. Next, a relative correction, C^{Rel}, is applied as a function of the η of the jet. This compensates for the differing response of different parts of the detector across different pseudorapidity ranges. An absolute correction, C^{Abs} is then applied on the output of the first two calibration steps, p_T^{Rel}. This correction is designed to correct the dependence of the jet p_{T} on the magnitude of the measured p_{T}. These two corrections are calculated using information taken from both simulation and data. The final correction, C^{Res}, is applied only to data and not simulation. This is designed to correct the residual differences in the p_{T} and η response between the data and simulation. The magnitude of the corrections as a function of η for two representative values of p_{T} can be seen on Fig. 4.5.

4.5.4 Tagging *b*-jets

The b-quark is of particular interest when searching for indications of new physics. The production of these quarks is relatively rare in SM interactions and often indicates the presence of a top quark that has decayed to a b. In a wide range of SUSY models the decay via top quarks is favoured over other, lighter quarks. It is therefore very

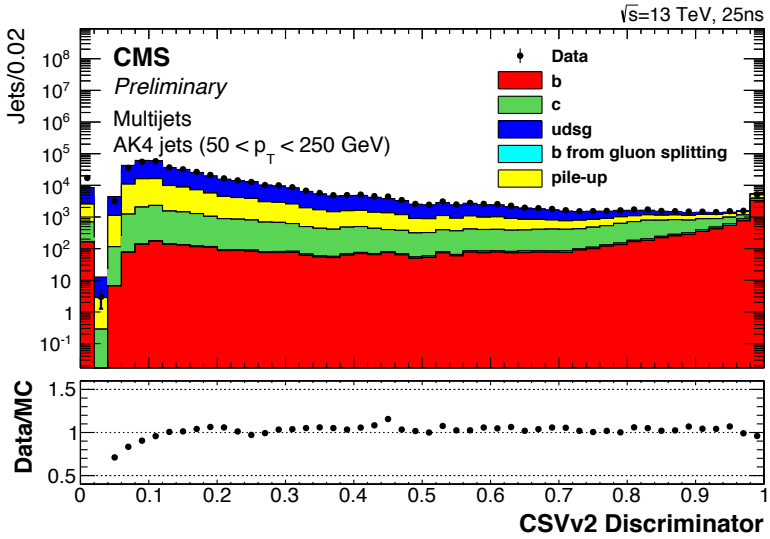


Figure 4.6: The distribution of the discriminator CSVv2 algorithm for b-tagging in multijet events. Tagged jets are clustered with the anti-k_T algorithm with $R = 0.4$ and span $50 < p_{\mathrm{T}} < 250$ GeV. A working point is chosen to trade off b-tagging efficiency for mistag rate by making a cut on the discriminator [90].

advantageous to have a way of identifying jets as originating from b-quarks, known as *b-tagging*.

As b-hadrons have a relatively long lifetime, $\tau \sim 1.5$ ps [89], they typically travel $c\tau \sim 450$ μm from the primary vertex before decaying. The CMS tracker has a spatial resolution of ~ 30 μm for ~ 5 GeV tracks [8], which means the displaced vertex created by the ~ 5 GeV mass b-quarks can be resolved. This is exploited by the Combined Secondary Vertex version 2 (CSVv2) algorithm used for b-tagging in Run 2 [90]. This algorithm uses a neural net with detector inputs, including vertices identified by the tracker that are displaced from the primary vertex and associated to the jet and information about the tracks originating from this vertex.

As with the other physics object definitions in this chapter, different selection criteria can be chosen to trade-off the mistag rate and efficiencies for b-tagging. For each working point a different discrete cut on the CSVv2 discriminator is chosen, its distribution is shown in Fig. 4.6. A *medium* working point cut of 0.8 corresponds to a tagging efficiency of $\sim 69\%$ and a mistag rate of $\sim 1\%$ [90].

4.6 Isolation and jet cross-cleaning

It is often important to differentiate *prompt* leptons produced directly in the primary vertex from those that are the result of decays of other particles, such as those in hadronic jets. To do this an *isolation* variable, I^{rel}, is defined for each lepton. To construct it the p_{T} of the PF candidates within a cone around the lepton are summed and the estimated neutral-charged contribution of PU within that cone is subtracted. This is then divided by the p_{T} of the lepton in question, ${p_{\text{T}}}^{l}$. In the case that the lepton is part of a jet I^{rel} will have a large value, this will not typically be the case if it is prompt.

In the work presented in this thesis, two variants of cone size are used. In the standard case, denoted I^{rel}, a cone size of $\Delta R = 0.3$ is used. To maintain acceptance to leptons produced in the decays of boosted objects, such as top-quarks, a *mini-isolation* can be defined that has a variable cone that depends on the p_{T} of the object. A value of $\Delta R = 0.2$ is chosen for ${p_{\text{T}}}^{l} < 50$ GeV, but this is scaled down to a minimum $\Delta R = 0.05$ for ${p_{\text{T}}}^{l} > 200$ GeV.

The contribution from PU is typically estimated in two different ways. For *effective-area* correction the average PU energy density of neutral particles, ρ^{neutral}, is multiplied by the area of the isolation cone, similar to the PU correction carried out in jets as described in Sec. 4.5.3. For the $\Delta\beta$ correction, the energy of neutral particles is estimated as half the energy from charged particles that originate from PU vertices in the cone. This ratio of charged to neutral particles is determined from simulation.

To avoid isolated leptons being classified as jets, a *cross-cleaning* can be performed on the result of the jet clustering. In this case, all jets that are within $\Delta R = 0.4$ of an isolated lepton are removed from the event.

4.7 Missing transverse energy ($\not{E}_{\text{T}}$) and energy sums

Weakly interacting particles, such as neutrinos or neutralinos, will pass through all components of the CMS detector. Their production can therefore only be inferred from the imbalance of momentum in the decay products of a proton collision that can be observed by CMS. To measure the magnitude and direction of the missing momentum the missing transverse energy variable, $\not{E}_{\text{T}}$, is defined as the negative vector sum of

the momentum, $\vec{p}_T$, of all the particles in an event:

$$\not{E}_{\mathrm{T}} = -\sum \vec{p}_T. \tag{4.3}$$

This is reconstructed taking the PF candidates up to $|\eta| = 5$ as input [91].

A correction is applied to $\not{E}_{\mathrm{T}}$ based on the jet energy correction described in Sec. 4.5.3. This is known as the *Type-I* correction and applies the jet energy correction to the PF candidates that are clustered into jets. This method takes jets with $p_{\mathrm{T}} > 15$ GeV and helps to signficantly improve the $\not{E}_{\mathrm{T}}$ resolution.

It is also advantageous to define other energy sums that take only jets as input, which are typically better understood and calibrated than each unclustered PF candidate. To gain a measure of the scale of hadronic energy in an event, the H_{T} variable is defined as the scalar sum of jet momenta, ${p_{\mathrm{T}}}^{\mathrm{jet}}$. To gain an alternative measure of the missing energy the negative vector sum of jet momenta, ${\vec{p_{\mathrm{T}}}}^{\mathrm{jet}}$, denoted $\not{H}_{\mathrm{T}}$ is defined.

$$H_{\mathrm{T}} = \sum {p_{\mathrm{T}}}^{\mathrm{jet}}, \quad \not{H}_{\mathrm{T}} = -\sum {\vec{p_{\mathrm{T}}}}^{\mathrm{jet}}. \tag{4.4}$$

4.8 Monte Carlo (MC) simulation

A major part of understanding the data collected by CMS involves the use of simulated events for different types of SM and BSM processes. With a good simulation, it is possible to classify the events observed in data as demonstrating particular physical processes that are predicted by theory. The simulation of proton collisions is particularly challenging, as it requires an accurate modelling of a wide range of SM phenomena as well as a very good understanding of the CMS detector. The simulation is split up into a series of stages that split up the theoretical simulation and detector modelling [92].

In the first stage, the hard scattering of the colliding constituents, *partons*, of the incoming protons through electroweak or QCD interactions is modelled. The fraction of the proton momentum carried by each parton is sampled from a parton distribution function (PDF). This modelling involves a calculation with perturbation theory to a fixed order, typically leading order (LO) or next-to-leading order (NLO), with a dedicated MC generator such as MADGRAPH [93] or PYTHIA [94].

In the next stage an iterative process of *parton showering* is carried out. Simulated QCD radiation from the hard scatter decay products is carried out until the particles in the shower reach the QCD cut-off scale, $\Lambda_{\text{QCD}} \sim 1$ GeV, where perturbation theory breaks down. After the parton showering the remaining particles undergo hadronisation, where colourless hadrons are formed and allowed to decay, resulting in *generator-level* particles with well defined four-momenta. The hadronisation is carried out with MC generators such as PYTHIA [94, 95] and HERWIG [96] that each use different techniques to carry the connection of the colour-flow from the initial state to final state particles. Remaining unstable particles have their decays simulated with dedicated MC algorithms.

The output of the hadronisation stage is then passed through a simulation of CMS with software such as GEANT4 [97]. The individual components and their responses are faithfully reproduced to output a simulated event that is comparable to one measured in real proton collisions. The standard reconstruction, such as that described in this chapter, can then be performed.

Finally, the PU must be simulated in MC. A large sample of minimum bias, typically soft, QCD events are simulated. Individual events from this sample are taken and overlaid on the simulated event of interest to imitate the effects of PU. The effects of out of time pileup (OOTPU) are also taken into account by simulating other minimum bias MC events within a 12 bunch crossing window.

Chapter 5

The Level-1 trigger upgrade jet algorithm

5.1 The Calorimeter Trigger upgrade for Run 2

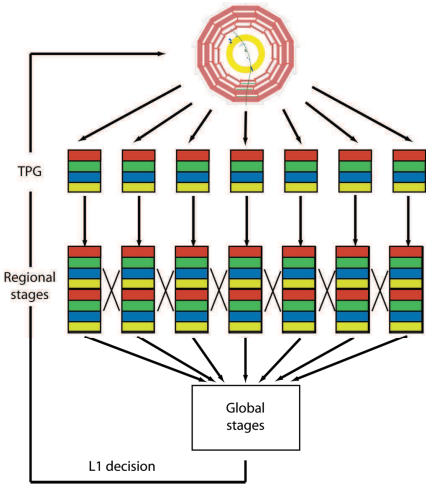
With the advent of Run 2 of the LHC, the demands placed on the CMS trigger greatly increased. After a long period of shutdown (known as Long Shutdown 1 (LS1)) the energy of proton collisions was increased to 13 TeV and the bunch crossing rate decreased to 25 ns. With this configuration the design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ can be reached with an average PU of around 25 events per bunch crossing. However, the potential luminosity of the LHC after the LS1 upgrade exceeds the original design value, allowing a PU of up to 40 simultaneous collisions. The original Level-1 trigger was not designed to deal with this amount of PU. The presence of extra collisions in each event results in a general increase in the average energy deposited in the calorimeters of the detector. This extra energy smears out the measurement of energy deposited by particles from the hard scatter, reducing the resolution with which it is measured. This acts to increase the rate at which the trigger accepts events and decrease the accuracy with which it identifies interesting events. As this rate is limited by the readout capability, a trigger that is not upgraded will have a reduced efficiency of acceptance of interesting physics events. It is therefore very advantageous to upgrade the Level-1 trigger in a way that allows it to deal with the extra PU [66].

Along with maintaining a low output rate for a high efficiency, another key consideration for the Level-1 trigger is the time taken for the algorithms to run on the boards that carry out the trigger processing, their *latency*. This must be kept low to ensure

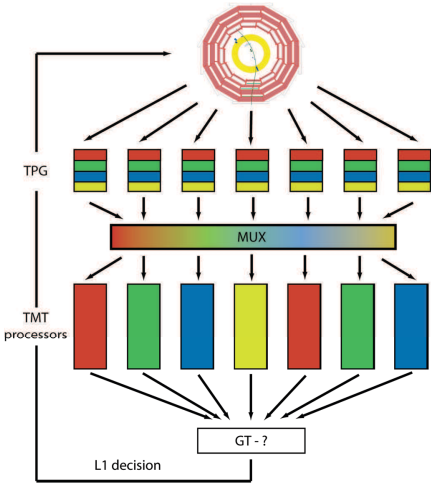
that every bunch crossing is analysed in as much detail as possible. To be able to carry out this sort of processing, the Level-1 trigger makes use of custom FPGA boards, as described in Sec. 3.2.5. The complexity and structure of the trigger algorithms is limited to ensure they can run quickly enough on the FPGA boards. An upgrade to the trigger hardware therefore allows more sophisticated algorithms to be carried out, while still being able to process every bunch crossing.

The amount of calorimeter information that can be processed by a single FPGA is also limited by its input and output capacity. To pass the full calorimeter data from a single bunch crossing through an FPGA in the Level-1 trigger it takes the equivalent of ~ 10 bunch crossings of time. In Run 1 of the LHC, the calculations for triggering on the calorimeters were performed with the Global Calorimeter Trigger (GCT) [98]. The GCT dealt with this delay by processing different sections of the calorimeter in separate boards, necessitating duplication of information at the detector section boundaries. However, by time multiplexing the calorimeter data, it is possible to have individual FPGA boards that each process one bunch crossing with minimal information duplication. To achieve this, 10 FPGA boards are utilised in a round-robin configuration. The data collected from the first bunch crossing is processed by one board, the data from the second bunch crossing by another board and so on. Each bunch crossing is processed by another board until the first board has finished and can subsequently process the next available bunch crossing. This would not be possible without the timing information provided by time multiplexing the data. As well as increasing the total possible information bandwidth, this allows greater flexibility in the algorithms used as each FPGA has access to the full event information [99, 100]. The upgrade architecture of the Level-1 trigger is therefore that of a Time-Multiplexed Trigger (TMT), a representation of the difference between this architecture and that of the GCT can be seen in Fig. 5.1. The hardware is also upgraded, exploiting the recent technological improvement in the performance of FPGA processors and high-speed optical links [101].

The inputs to the Calorimeter Trigger from the various ECAL and HCAL components are split into a 56×72 grid of Trigger Towers (TTs) in η and ϕ within the barrel and endcap region ($|\eta| < 3$), with each tower covering the area of 5×5 ECAL crystals. As the processing power of the GCT was limited, for trigger calculations the TTs were grouped into 4×4 blocks, known as *Regional Calorimeter Trigger (RCT) regions* [98]. These different regions are visually represented in Fig. 5.2. Due to the advantages brought by the upgrade, the new algorithms have access to the full granularity in-



(a) Run 1 trigger architecture.



(b) Run 2 upgrade TMT trigger architecture.

Figure 5.1: A representation of the time multiplexed trigger (TMT) architecture (b) as opposed to the pre-upgrade trigger architecture (a). The entire information for one event is processed by one board rather than just parts of the detector for each board [99]

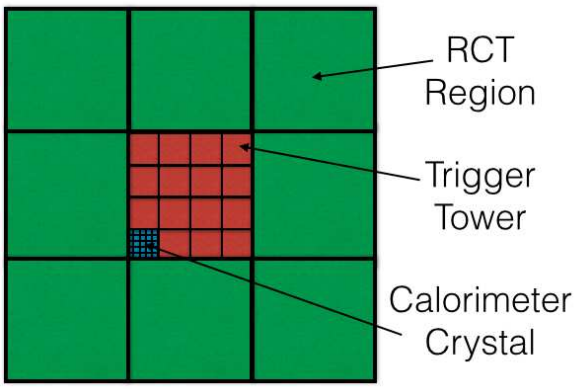


Figure 5.2: The breakdown of the different calorimeter regions, the RCT regions and trigger towers are both shown within the context of the ECAL calorimeter crystals.

formation of all the TTs. A representation of the upgrade to the spatial resolution afforded by this change is in Fig. 5.3.

5.2 The jet finder and energy sums

To make use of the potential brought by the Level-1 trigger upgrade, new trigger algorithms are developed to improve the selection of interesting physics processes.

(a) Run 1 calorimeter trigger resolution.

(b) Run 2 upgrade calorimeter trigger resolution.

Figure 5.3: A representation of the spatial resolution of the trigger inputs to the calorimeter trigger for the Level-1 trigger before and after the upgrade.

With a better view of the substructure of energy deposits within the calorimeters, the identification of three pronged tau jets and isolated electrons can be significantly improved, for example [102, 103].

One area in which algorithm improvements can have a significant impact is in the jet finding algorithm. As the Level-1 trigger must be of a fixed latency, it cannot make use of an iterative algorithm such as those used during offline jet-finding (described in Sec. 4.5). In the algorithm used during Run 1, the jet candidates are created from a 3×3 RCT region sliding window. Candidates in which the central region has the maximum energy are kept as jets [8]. These jets cover an area of $\Delta\eta \times \Delta\phi = 1.04 \times 1.04$, which matches reasonably well the area covered by circular jets with $R = 0.5$, the maximum radius for Run 1 offline jets. To help mitigate the reconstruction of jets originating from PU, the central region of the jet finder is also required to exceed a minimum energy, known as the *seed threshold*. Due to latency and processing power constraints, there is no way in which contribution to the energy of the jet from PU can be subtracted dynamically. Along with the relatively large jet size, this means the performance of the algorithm will be severely reduced by the conditions present during Run 2. An upgrade provides the extra processing power to carry out a dynamic pileup subtraction (PUS), correcting the energy of each jet from the influence of PU. The increased position resolution will also allow for a more accurate determination of the direction of the jets.

Once a jet algorithm is chosen, the jets can be used to make Level-1 trigger decisions. This typically involves selecting events that are observed to have jets, or sums of jets, with an energy above a certain threshold. The selection made within the Level-1 trigger

is designed to pick out interesting physical processes, while removing background events. The particular configuration of jets that lead to an acceptance of an event is known as a *trigger*. For example, a single-jet trigger would require at least one jet above a certain p_{T} threshold, or a quad-jet trigger would require at least four jets above a certain threshold. The performance of the Level-1 trigger algorithms are therefore typically measured within this context, as discussed in Sec. 5.5. When designing these triggers, the total rate of acceptance of the Level-1 trigger cannot exceed a predefined value $\sim$100 kHz. The sum of the rates of all the trigger decisions must be kept below this number. A trade off between all the potential triggers must be made, making it of utmost importance to design triggers with manageable rates.

5.2.1 The upgraded jet algorithm

The jet algorithm for the Run 2 CMS trigger upgrade operates on the sum of the ECAL and HCAL energy deposits in the TTs. The studies in this Chapter consider the TTs within $|\eta| < 3$.

Starting with a similar idea as in Run 1, the upgrade makes use of a sliding window algorithm. In the upgrade, however, the window considered is a 9×9 square of TTs, covering $\Delta\eta \times \Delta\phi = 0.78 \times 0.78$. This odd number of trigger towers results in an easily defined central tower as well as matching the area covered by the $R = 0.4$ offline jets that are used in Run 2. Within a window the central tower is considered as the direction of a candidate jet. The candidate is vetoed if any of the other TTs in the square have an energy deposit of either greater than or, in some cases, equal to it. The veto condition is antisymmetric along the diagonal of the square to prevent TTs with the same energy from vetoing one another. A representation of the window considered can be seen in Fig. 5.4. Any TTs that pass this criterion are considered as jet centres, where the jet energy is equal to the sum of all the towers within the 9×9 square.

The veto conditions applied on the central TT ensure that no two overlapping jets are reconstructed, avoiding the duplication of energy deposits. However, the algorithm can introduce inefficiencies in very specific jet topologies. In these cases a high energy TT vetoes a medium energy TT which then vetoes a lower energy TT. The medium energy TT is included in the jet constructed by the high energy TT, however the lower energy TT is lost. This is not usually a problem when making decisions

Figure 5.4: The consideration of a Trigger Tower candidate for the upgrade Level-1 Trigger jet algorithm. The candidate (green) is vetoed if the energy of the other towers meets the condition shown in the blue and purple towers.

with the Level-1 trigger, as the high energy TT will usually ensure that the event is triggered, despite energy being lost.

5.2.2 A comparison with the offline jet algorithm

In order to test the performance of the upgrade algorithm it is compared to jet finding with anti-k_T clustering, the most popular algorithm for offline jet reconstruction. As the Level-1 algorithm has less sophisticated inputs than the PF candidates used offline, the anti-k_T algorithm is used to find jets with the TTs as input. The test was carried out on a $t\bar{t}$ MC simulation, which contains a high multiplicity of jets that are produced in top-quark decays. The performance of the algorithm compared to anti-k_T jet clustering with $R = 0.4$ can be seen in Fig. 5.5. For the jet with the highest p_{T}, the leading jet, the distributions of p_{T} and η are very similar for both the jet finding algorithms. For the fourth-leading jet more differences emerge at low p_{T} and the edges of the η distribution. This is probably due to the ability of the anti-k_T algorithm to adaptively fit smaller radius jets and those at the edge of the detector acceptance.

5.2.3 Energy sums

Along with finding jets in an event, the Level-1 trigger is required to calculate energy sums, analogous to those discussed in Sec. 4.7. These energy sums can take either the jets or TTs as input. They include:

- total E_T, the scalar sum of the transverse energy in all the TTs

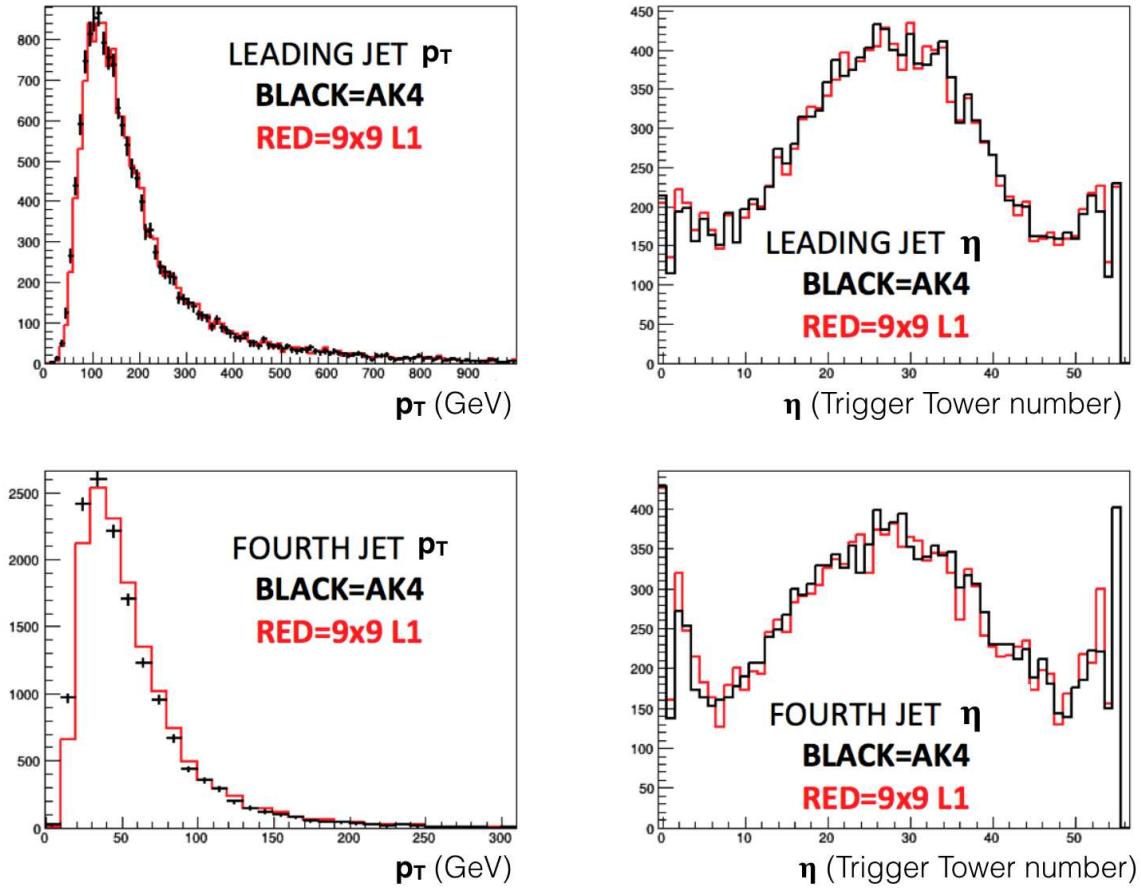


Figure 5.5: A comparison between the upgrade Level-1 trigger (9×9L1) jet algorithm and the anti-k_T offline algorithm with R=0.4 (AK4), taking the TTs as input. These plots are produced from a $t\bar{t}$ simulation with 13 TeV proton collisions and PU ~ 40. The units for the pseudorapidity are in trigger tower units, which range from 0 at $\eta = -3$ to 56 at $\eta = 3$.

- $\not{E}_{\mathrm{T}}$, the negative vector sum of the transverse energy in all the TTs
- H_{T}, the scalar sum of all the Level-1 jets
- $\not{H}_{\mathrm{T}}$, the negative vector sum of all the Level-1 jets

5.3 Pileup subtraction

5.3.1 Characterising pileup

On average, the decays from PU are expected to be isotropically distributed around the detector. However, there will be significant event-by-event differences due to Poisson fluctuations in the number of PU collisions. While the number of simultaneous interactions is such that these fluctuations are significant, it is advantageous to correct for the PU on a per-event basis. To be able to remove the effects of PU in the Level-1 trigger there needs to be a way to remove the objects that originate from PU. Additionally, an estimate of the average energy deposited in each event should be used to correct the E_T of the remaining objects. These procedures are known collectively as *pileup subtraction (PUS)*.

It is observed that the density of decays from PU depend on their position in η [104]. This is most likely caused by the differing response of the detector along its length. The fact that the trigger only has access to coarse calorimeter information and no tracks means these detector effects are likely to be enhanced. It is therefore desirable for PUS algorithms to measure the PU energy density near to the objects that are corrected. This *local* PUS then requires a trade off between the proximity to the object and the effects of the larger statistical fluctuations from sampling a smaller area.

A *global* subtraction can be used to determine the average energy density across the whole detector. This method is much less susceptible to fluctuations and contamination from particles from the hard-interaction, but does not take account of local fluctuations or detector effects.

The ideal form of PUS will remove the dependence of the Level-1 trigger rate and efficiency on the number of simultaneous interactions in an event. It should do this in a way that improves the resolution of object energy measurements, increasing the acceptance of interesting physics events while allowing for the rejection of soft

QCD multijet processes. It is also important to take account of the Level-1 trigger architecture when designing the algorithm to ensure that it can be performed with a fixed latency. Different forms of PUS in the Level-1 trigger are outlined in this section and their performance subject to these criteria is examined in Sec. 5.5.

As well as mitigating PU from simultaneous collisions, the effects of OOTPU must be removed. This is predominantly achieved during the generation of the TT inputs for the calorimeter trigger. When reading out the energy deposited in the ECAL and HCAL the average extra energy deemed to have come from a previous bunch crossing is subtracted and a correction factor is applied. This mitigation performs well enough to make the effects of OOTPU subdominant with respect to in-time PU.

5.3.2 Global pileup subtraction

A prominent method of PUS is ρ-area correction [88, 105]. It works by finding the average energy per unit area in the calorimeter due to PU, ρ, and subtracting it from each reconstructed jet. A favoured estimator for this quantity is:

$$\rho \equiv median(\frac{p_{Tj}}{A_j}), \tag{5.1}$$

where p_{Tj} is the transverse momentum of a jet j, A_j is its area and the median is taken over all reconstructed jets in an event. In the plots in this chapter, this form of PUS is known as *global*. It acts to remove half the jets in an event, the low energy half, and correct down the energy of the remaining half. This works particularly well in high PU cases as ρ is insensitive to fluctuations in the energy of interesting physics events.

The mean value of ρ, as a function of the number of interactions, is shown in Fig. 5.6. This is taken from a minimum bias MC simulation in which there is no visible hard interaction but only overlaid PU collisions. There is a good correlation between the number of simultaneous collisions and ρ, indicating that it is indeed a reasonable measure of the PU in the event. The correlation also passes through the origin, indicating that OOTPU is not a problem with a 50 ns bunch spacing.

Subtraction with the ρ-area method has a latency penalty when implemented in hardware as the jets must be found before ρ can be calculated and subtracted from their energies. Depending on the final firmware designs this can present a problem and makes global ρ subtraction harder to perform in hardware. Despite this fact,

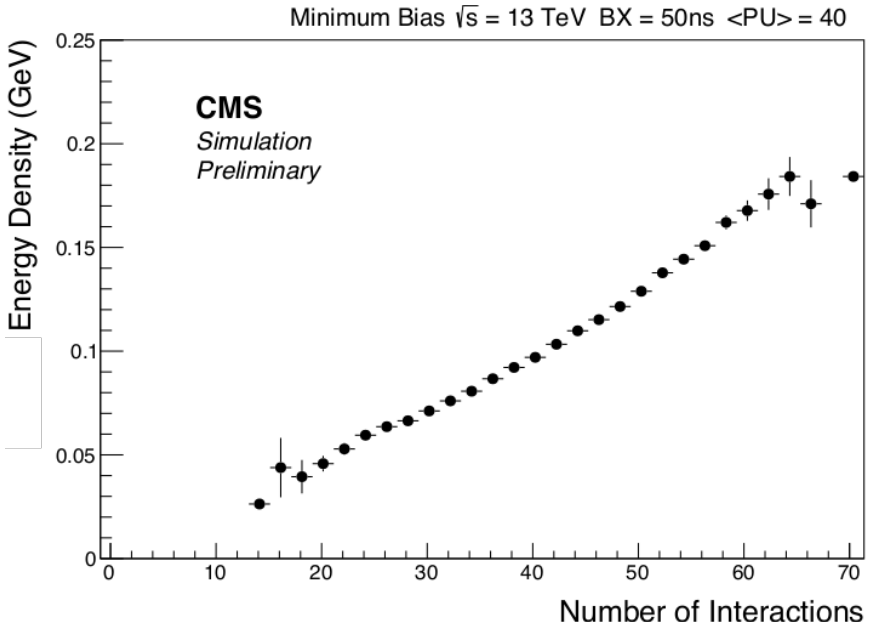


Figure 5.6: The median energy density, ρ, of Level-1 jets in the CMS calorimeters as a function of the number of simultaneous collisions. This is taken from minimum bias MC simulation at 13 TeV with a 50 ns bunch crossing time

as global ρ is a popular and well understood form of offline PUS it acts as a good benchmark against which to test other algorithms.

5.3.3 Donut subtraction

As jets from the hard scatter are typically boosted objects, most of their energy is deposited very close to the central TT of the jet algorithm [106]. The p_{T} weighted distance from the centre of the jet axis of particles that make up a jet is shown in Fig. 5.7. In the case of isolated jets, the ring five TTs from the centre of the jet can be assumed to contain only contamination from PU jets external to the jet in question. The *donut subtraction* algorithm therefore takes the energy per unit area in the ring of TTs surrounding the jet (shown in Fig. 5.8a) and scales it up by the area of the jet. The resulting energy is then subtracted from the jet to correct any PU contamination. This kind of PUS has been applied in the analysis of heavy ion collisions at the LHC [107].

This approach only works for correcting isolated jets, if one jet from the hard-scatter is in the vicinity of another, the energy in the donut can be increased to above that of PU. To mitigate this, only the median two 4×1 TT strips of the four that make up the

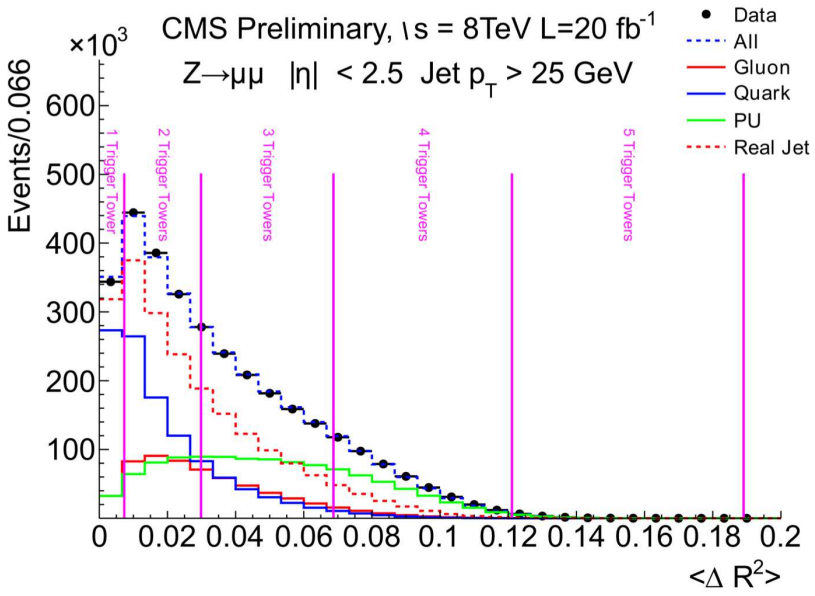
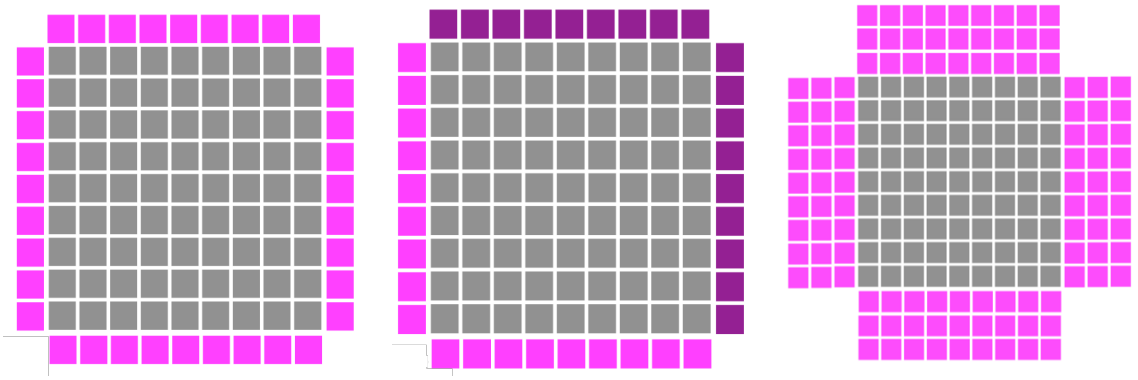


Figure 5.7: The p_{T} weighted distance of particles clustered in a jet with respect to the central jet axis, $\langle \Delta R^2 \rangle = \sum_i \Delta R_i^2 p_{\mathrm{T}i}^2 / \sum_i p_{\mathrm{T}i}^2$, where the different particles in a jet are indexed with i and are a distance of ΔR away from the jet centre. The number of trigger towers that the distances corresponds to are shown in pink. Calculated for jets from a simulation of Z to $\mu\mu$ events [106]



(a) Strips considered for donut subtraction. **(b)** Example of median energy strips. **(c)** Strips considered for chunky donut subtraction.

Figure 5.8: Various configurations of TT strips around the Level-1 jet algorithm window used for donut subtraction

donut are used to calculate the PU energy density. This reduces the chance that energy from another jet will be counted as PU and also removes strips that have very little energy in them from a downward fluctuation in PU contamination. An example of the two strips that could be selected can be seen in purple in Fig. 5.8b.

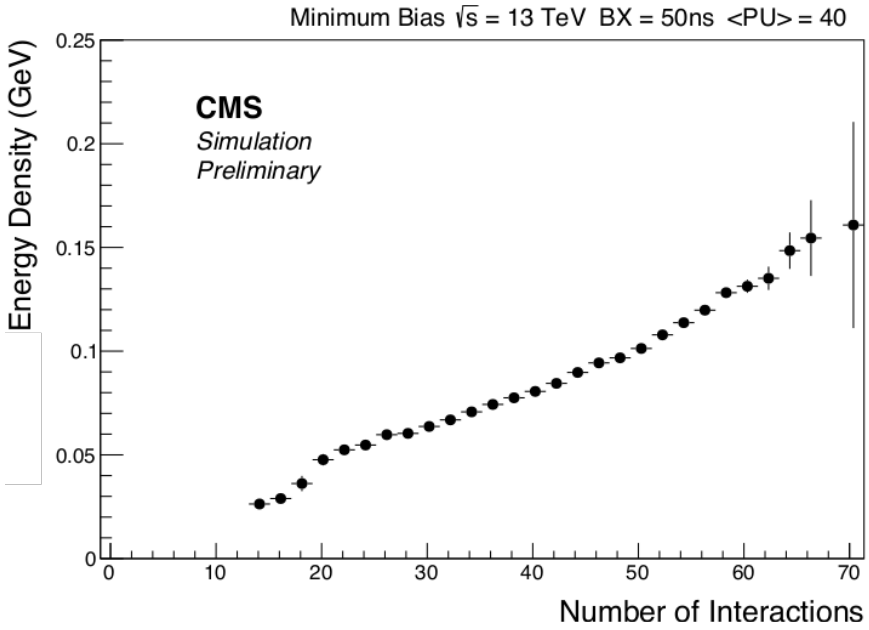


Figure 5.9: The energy density in the median two 3×1 TT strips of a chunky donut around Level-1 jets in the CMS calorimeters as a function of the number of simultaneous collisions. This is taken from minimum bias MC simulation at 13 TeV with a 50 ns bunch crossing time

One of the main issues with the donut subtraction algorithm is this sensitivity to fluctuations. It is mitigated by taking the median energy two strips but can be further reduced by increasing the area covered by the strips. The rings of TTs can be extended to be three towers wide, known as a *chunky* donut and illustrated in Fig. 5.8c. This is particularly effective in reducing the fluctuations in the positions of PU particles with respect to the jet in consideration.

To confirm that the energy in the median two strips of the chunky donut is a good measure of the PU in the event, it is plotted against the number of interactions for a minimum bias MC sample in Fig. 5.9. There is a good correlation that appears to pass through the origin, implying it is indeed a good measure of PU.

In the implementation of the Level-1 jet finding algorithm in the upgrade hardware, the TTs that make up the donut are already available in memory. This means donut subtraction has a very low latency penalty. This presents a significant advantage over a global PUS.

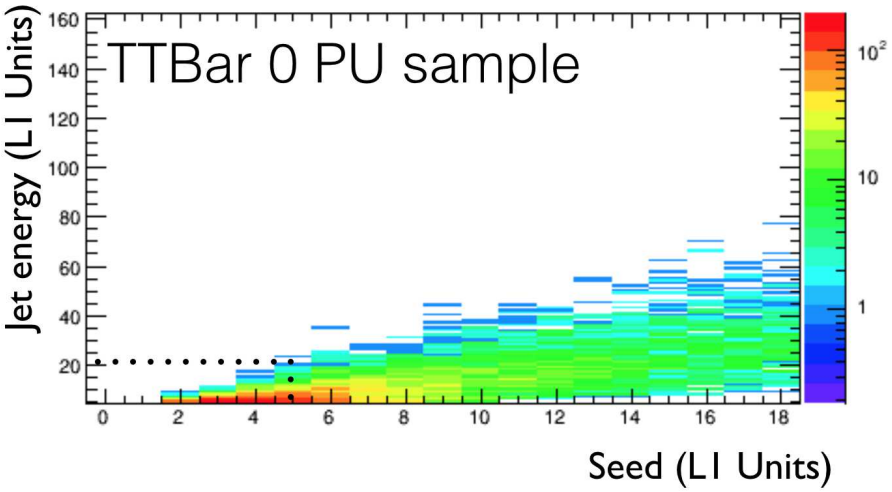


Figure 5.10: The Level-1 jet energy vs seed threshold for a simulated sample of top quark pair production events with no overlaid PU. The units of energy are *L1-units*, which correspond to 0.5 GeV each. A seed threshold of 2.5 GeV only removes up to 10 GeV jets from the hard scatter.

5.3.4 Jet seed threshold and zero suppression

Donut subtraction helps to remove the effects of PU on the reconstructed high energy jets, but is less successful at removing the soft jets that are purely from PU interactions. A simple way of reducing the number of soft jets is by introducing an energy threshold on the TT that can form a jet. The TT that is considered for a jet candidate in the algorithm outlined in Section 5.2.1 is required to be above a certain energy, known as the seed threshold. This is very easy to implement in hardware, and has the potential to save latency as it reduces the number of jets that need to be made. A disadvantage is that it can kill soft jets that originate from a primary vertex. It also does not account for any η dependence. For the studies in Sec. 5.5 a seed of 2.5 GeV was chosen as a benchmark that appeared to kill PU jets without removing jets above 10 GeV from a zero PU $t\bar{t}$ MC simulation. This is demonstrated in Fig. 5.10. As the Level-1 hardware measures energy in units corresponding to 0.5 GeV (*L1-units*), this is denoted as *Seed 5*.

To remove the effects of noise on the Level-1 trigger inputs, a *zero-suppression* is performed that requires TTs to have an energy above 0.5 GeV before they are considered to have any energy at all. As PU is expected to result in much smaller energy deposits than interesting physics processes, the effect of increasing the energy that is suppressed to zero when building the jets also acts as a form of PUS. This form

of PUS is considered in Sec. 5.5 and is named Tower Suppression (TSup). Despite being very easy to implement in hardware this algorithm does not adapt well to different PU conditions and can reduce the energy resolution of the jets.

5.4 Level-1 jet energy calibration

To obtain Level-1 jets that correspond as closely as possible to the true physics objects that they represent, their energy must be corrected. The varying response of the HCAL and ECAL as a function of jet p_{T} and η necessitates a calibration that depends on these variables. The need for calibration in the Level-1 trigger is exacerbated by the coarse level of information available compared to that in offline reconstruction.

The calibration is performed using QCD dijet MC simulation with an average PU of 40. Samples were generated with a range of generator scales from 10 to 600 GeV. These energies characterise the p_{T} of the leading jet and ensure a wide range of jet energies are available. From this sample, *generator jets* are clustered from the truth-level particles, the final result of the procedure described in Sec. 4.8. Clustering is performed with the anti-k_T algorithm with $R = 0.4$, after removing any muon and neutrino particles. This ensures that the Level-1 jets are just calibrated based on the particles that are deposited in the calorimeters. The Level-1 jet algorithm is performed on the simulated Level-1 trigger calorimeter inputs. Each Level-1 jet is matched to the generator jet that is closest in $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ to the central TT of the jet. If a generator level jet cannot be found within $\Delta R < 0.3$, the Level-1 jet in question is ignored. The p_{T} of the Level-1 jet (L1 p_{T}) after PUS is compared to the p_{T} of the generator jet (GEN p_{T}) and the response is defined as the ratio of these two quantities, L1 p_{T}/GEN p_{T}.

The distribution of the response is plotted in bins of generator jet p_{T} for each of the matched jet pairs. A Gaussian function is fit to the response to obtain an estimate for the mean response and standard deviation in a 2 GeV bin of generator jet p_{T}. The response is inverted to provide a corrective scale factor for the Level-1 jets. It is fit to a calibration function as a function of the L1 p_{T} in eight bins of η, up to $|\eta| < 3.0$ [108]. The function has the form:

$$\langle {p_{\mathrm{T}}}^{L1} / {p_{\mathrm{T}}}^{GEN} \rangle^{-1} = (p_0 + \frac{p_1}{(\log {p_{\mathrm{T}}}^{L1}) + p_2} + p_3 \exp(-p_4(\log {p_{\mathrm{T}}}^{L1} - p_5)^2)), \quad (5.2)$$

where p_n are free parameters to be found with a χ^2 minimisation fit, ${p_\mathrm{T}}^{L1}$ is the Level-1 jet p_T and ${p_\mathrm{T}}^{GEN}$ is the matched generator jet p_T. This parameterisation provides a multiplicative correction that can be applied to Level-1 jets as they are produced online where their original p_T, ${p_\mathrm{T}}^{raw}$, can be corrected to ${p_\mathrm{T}}^{corr}$ via:

$$ {p_\mathrm{T}}^{corr} = {p_\mathrm{T}}^{raw} \langle {p_\mathrm{T}}^{L1} / {p_\mathrm{T}}^{GEN} \rangle^{-1}. \tag{5.3} $$

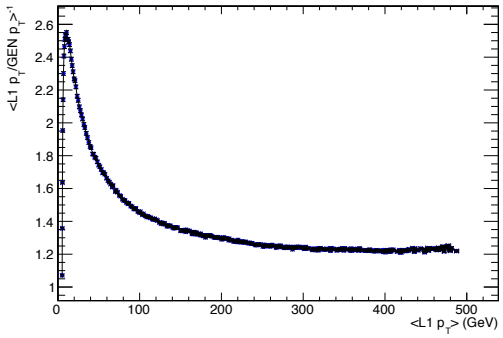
The inverse response as a function of Level-1 jet p_T for Level-1 jets produced with chunky donut subtraction and a central seed of 2.5 GeV are shown in Fig. 5.11.

To test the calibration procedure a closure test is performed by applying the calibration function to the simulation that was used to derive it. The mean response as a function of generator jet p_T in the range $|\eta| < 1.4$ is shown in Fig. 5.12 as an example. The calibration procedure produces a flat response at unity for jets with $p_\mathrm{T} > 50$ GeV. Below this value the matching procedure starts to break-down, as low p_T Level-1 jets are less likely to be matched to a corresponding generator jet. The Level-1 jets that are matched to a generator jet are more likely to have a higher value of p_T, as higher p_T jets are more likely to be identified. This skews the response upwards as the closure test approaches low values of p_T. From observing a series of such closure tests it was concluded that it was reasonable to consider Level-1 jets with a $p_\mathrm{T} > 20$ GeV, as in this case the corrected response was reasonably compatible with unity.

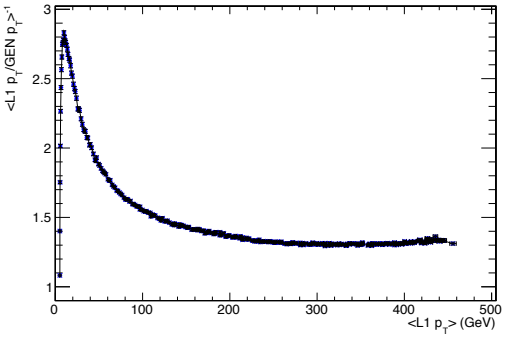
In future upgrades of the LHC it is proposed to increase the PU to ~ 140. This will provide a much higher instantaneous luminosity, but presents a significant challenge for reconstruction. To help inform this upgrade, the Level-1 jets were produced and calibrated on a sample with an average PU of 140. The resulting jet response as a function of generator jet p_T and Level-1 jet η can be seen in Fig. 5.13. The calibration was performed on jets with a seed of 2.5 GeV and chunky donut subtraction. The response is within 10% of unity up to $|\eta| < 2.5$ and for Level-1 jets with $p_\mathrm{T} > 50$ GeV. This suggests that the calibration and PUS procedure scales reasonably well with the number of PU interactions.

5.5 Performance of the upgraded algorithm

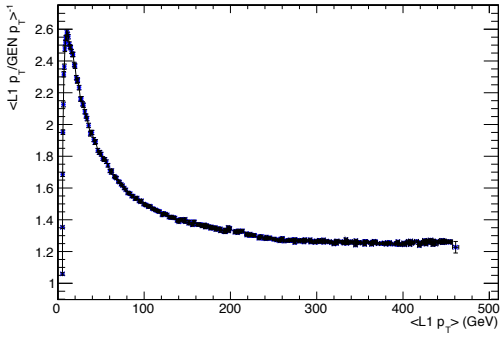
The performance of the jet algorithm with the various methods of PUS described in Sec. 5.3 is tested on 13 TeV MC simulation with an average PU of 40. A comparison is



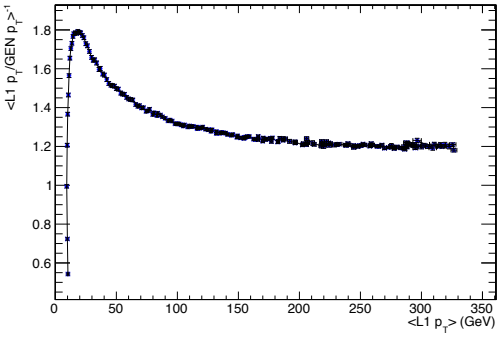
(a) $0.00 < \eta < 0.75$



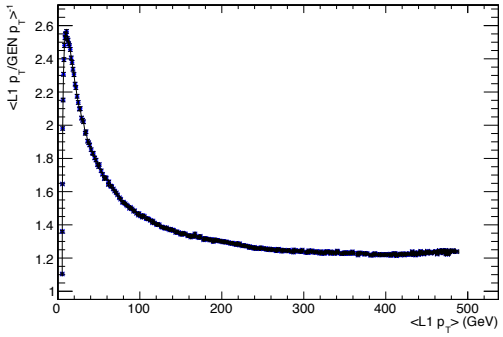
(b) $0.75 < \eta < 1.50$



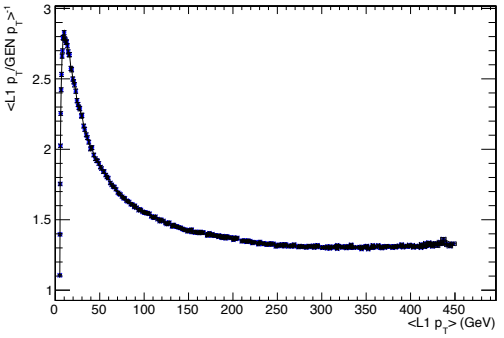
(c) $1.50 < \eta < 2.25$



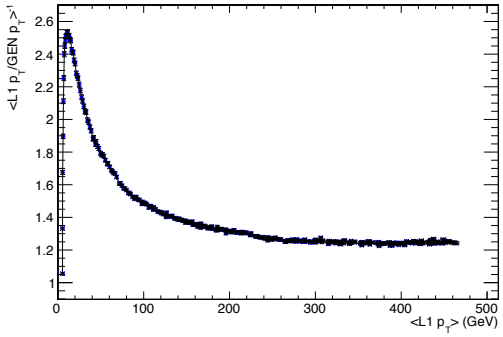
(d) $2.25 < \eta < 3.00$



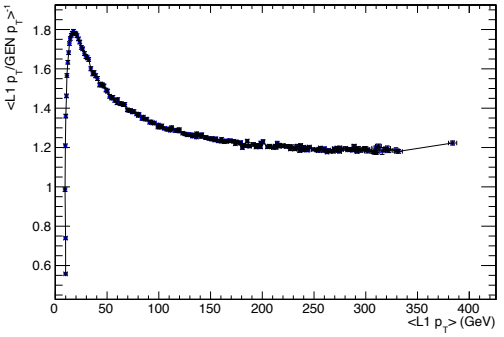
(e) $-0.75 < \eta < 0.00$



(f) $-1.50 < \eta < -0.75$



(g) $-2.25 < \eta < -1.50$



(h) $-3.00 < \eta < -2.25$

Figure 5.11: Calibration fits across all η ranges as a function of Level-1 jet p_{T}

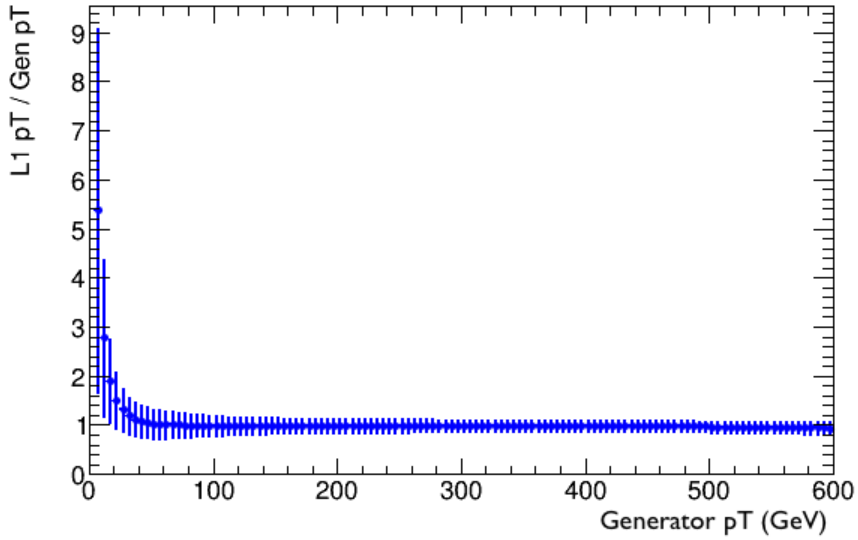


Figure 5.12: Closure test of the calibration procedure, checking the response as a function of generator jet p_{T} in the $|\eta| < 1.4$ range.

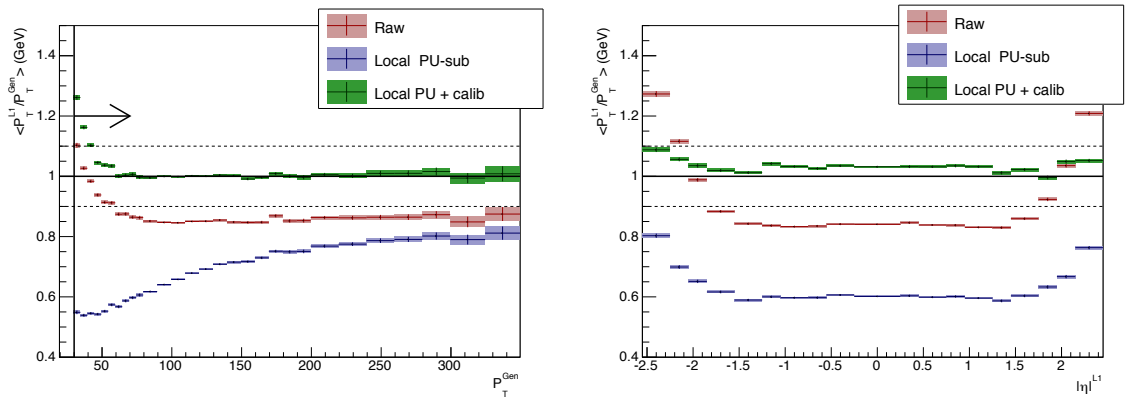


Figure 5.13: The ratio of the p_{T} of Level-1 jets with matched generator jets before and after they are calibrated. This is carried out at a PU of 140 and, demonstrates the effectiveness of the calibration in these extreme conditions.

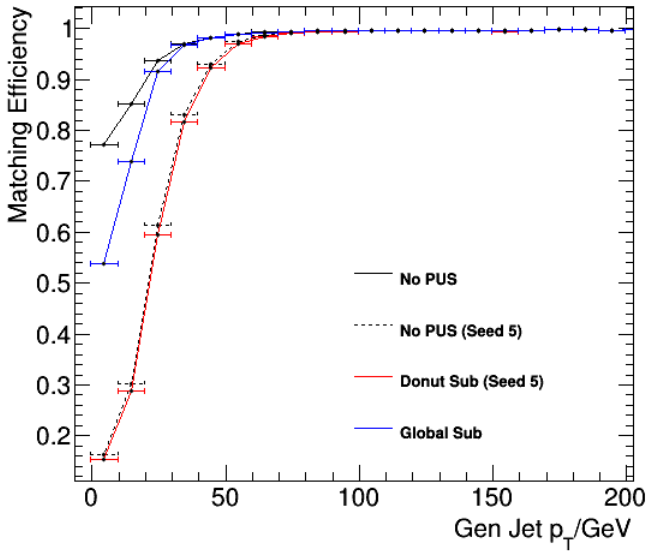
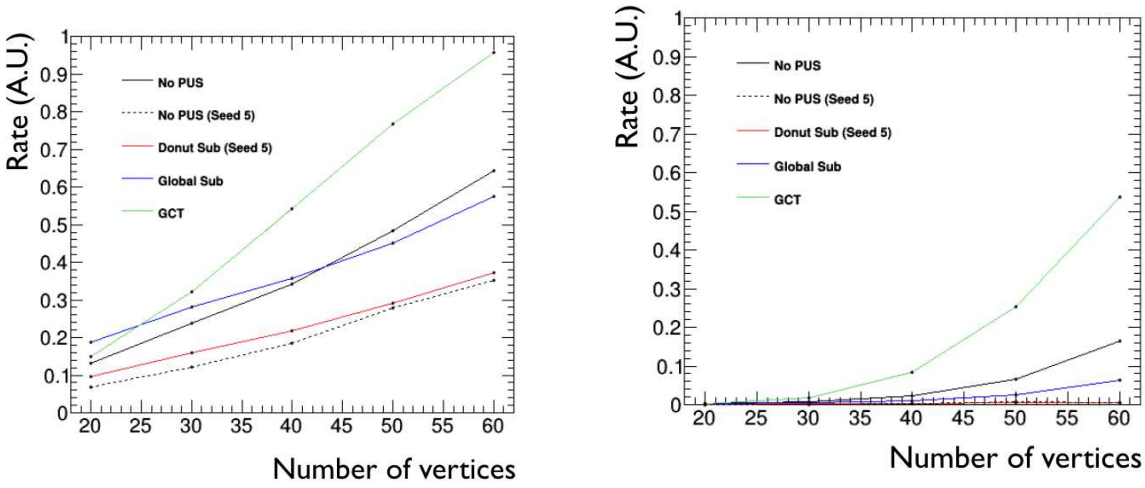


Figure 5.14: The efficiency with which a generator jet has a corresponding L1 jet within a radius, $R = 0.5$. This is carried out for all Gen jets in 70 000 simulated $t\bar{t}$ production events.

made to the jets produced by the GCT as it was in 2012. In some cases the performance of the interim calorimeter trigger that was run in the early stage of 2015 during the commissioning of the upgraded trigger is plotted with label *UCT*. To investigate rates a minimum bias sample was used, this sample only consists of overlaid PU events and has no generator jets. To test physics performance a sample of simulated $t\bar{t}$ events was used, which contains generator jets that originate from the hard interaction. The Level-1 jets are calibrated for each type of PUS with the method as outlined in Section 5.4.

Fig. 5.14 shows the efficiency of generator jets, produced as described in Sec. 5.4, being matched to a Level-1 jet as a function of their p_T. If a generator jet has a corresponding Level-1 jet within a radius $R = 0.5$ it is counted as matched. This radius is chosen as the distance from the centre of a L1 jet to the corner of its 9×9 square. A representative selection of the different PUS algorithms are shown. All of the upgrade Level-1 jet algorithms have close to 100% efficiency above 70 GeV. The application of a seed threshold significantly reduces the efficiency for low p_T jets, but the donut subtraction does not remove any extra signal jets.



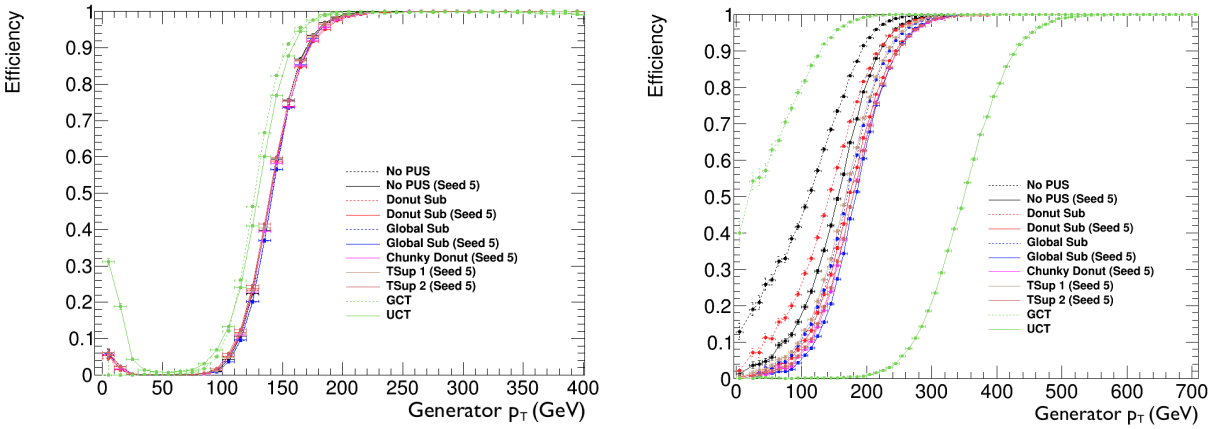
(a) Single-jet trigger $p_{\mathrm{T}} > 30$ GeV

(b) Quad-jet trigger $p_{\mathrm{T}} > 30$ GeV

Figure 5.15: The relative rate of events with $p_T > 30$ GeV leading (a) and fourth leading (b) jets after different PUS algorithms with and without seed thresholds for 70000 zero bias events as a function of the number of reconstructed vertices.

A key goal of PUS is to reduce the dependence of the Level-1 trigger acceptance rate on the number of PU interactions. To quantify this for a representative selection of algorithms, the rate against the number of reconstructed vertices is plotted for the minimum bias sample, Fig. 5.15. The rate is defined as the fraction of minimum bias events that pass a Level-1 jet p_T cut of 30 GeV. This cut is applied to a different jet rank depending on the type of trigger to be tested. For the *single-jet* trigger the cut is applied to the lead jet, for a *quad-jet* trigger the cut is applied to the fourth leading jet. The different forms of PUS clearly reduce the dependence of the rate on the number of interactions. The upgrade algorithm performs significantly better than the GCT, where the jet trigger rates are incredibly sensitive to PU. Of the algorithms investigated, the seed of 2.5 GeV, denoted *Seed 5*, has the most dramatic effect due to its ability to kill soft PU jets. The ρ-area subtraction, denoted *Global Sub*, also performs well. Applying donut subtraction without a seed threshold does not make much difference for this particular PUS benchmark. This is likely due to the fact that donut subtraction works well in correcting the resolution of jets contaminated with soft particles from PU vertices, rather than removing low p_{T} jets that originate from PU.

To characterise the energy resolution of the Level-1 jets with different PUS algorithms, the efficiency turn-on curves are plotted in Fig. 5.16. The lead jet efficiency



(a) Lead jet $p_T^{L1} > 150$ GeV efficiency

(b) $H_T{}^{L1} > 200$ GeV efficiency

Figure 5.16: Efficiency after a specific cut for Level-1 leading jets (a) and Level-1 H_T (b) as a function of the corresponding generator quantity.

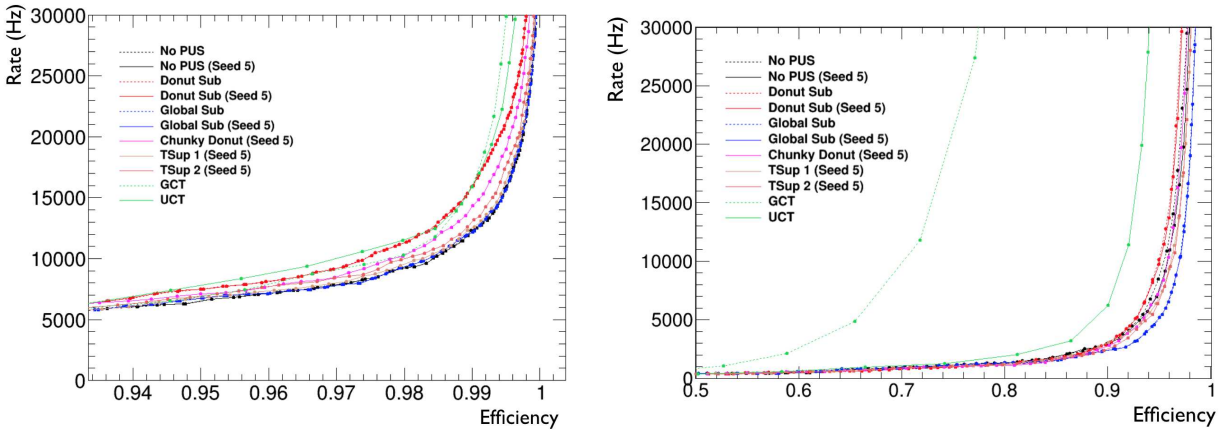
turn-on, Fig. 5.16a, is made by matching the generator jets to Level-1 jets, requiring $\Delta R < 0.5$ and taking the ratio of the matched generator jet p_T distribution without a cut on the Level-1 jet with the matched generator jet distribution after a cut on the Level-1 jets. Most algorithms show a similar performance for a single-jet trigger. The effects of PU are less relevant when just considering one high energy object.

More difference is observed when all jets with $p_T > 20$ GeV are summed to find the Level-1 H_T of the event. The plot in Fig. 5.16b is made by taking the ratio of generator H_T (also made from 20 GeV jets), with and without a cut on the Level-1 H_T. When considering a scalar sum of jets the effects of pileup add up and the different algorithms have a more significant separation. It is clear that performing PUS helps to significantly improve the energy resolution of Level-1 H_T.

In the turn-on curves, a change in offset along the x-axis results from a difference in jet energy scale with respect to the generator quantity. As the GCT jets do not have any calibration applied and the UCT has a different type of calibration, this can differ significantly. However, this does not give all the relevant information when deciding on a final algorithm to increase the overall performance of the trigger. The most important consideration is the trade-off between the efficiency of the algorithm at picking out a particular physics signature for a given rate. This does not depend on the specific energy scale of the jets relative to the generator quantities.

To better characterise the different Level-1 jet algorithms independently of energy scale, the efficiency of selecting events with a particular generator level jet quantity is plotted against the rate from a minimum bias sample. In each case, a range of thresholds for the equivalent Level-1 quantity are scanned through and the rate and efficiency for the thresholds are plotted. The rate is calculated for an instantaneous luminosity of 7×10^{33} cm^{-2}s^{-1} with a 50 ns LHC bunch crossing interval. Figure 5.18 shows the rate of the Level-1 trigger in Hz against the efficiency of selecting events for jet triggers. In Fig. 5.17a the efficiency of selecting $t\bar{t}$ events containing a generator jet with $p_{\mathrm{T}} > 150$ GeV is plotted against the minimum bias rate for an equivalent Level-1 threshold. The rate and efficiency are plotted for a range of cuts on the lead Level-1 jet for all the jet algorithms discussed in this chapter. In Fig. 5.17b the efficiency of selecting events with a fourth leading generator jet with $p_{\mathrm{T}} > 50$ GeV are plotted for a range of cuts on the fourth Level-1 jet. As was observed in the turn-on curves, the efficiencies for the lead jet triggers do not depend so much on the type of PUS, although there is a moderate improvement for all the upgrade algorithms over that used in the GCT. The difference becomes more evident when considering the efficiency of selecting a fourth leading jet. The lower p_{T} fourth leading jets show a more significant improvement with the upgraded jet algorithm. The application of PUS has some effect but it is still not significant when just considering relatively high p_{T} and low multiplicity jets. The performance favours some form of seed threshold, with the global PUS performing slightly better than the local variants.

Considering the jet triggers is important to understand the behaviour of the upgraded algorithm for individual jets. However, energy sum triggers are typically more widely used when searching for signatures of BSM physics with lots of hadronic activity and missing energy. This is due to the fact that they give a total measure of the hadronic energy scale of the event and a strong indication of the presence of any weakly interacting particles. The performance of the H_{T} and $\not{H}_{\mathrm{T}}$ triggers is therefore of significant interest and shown in Fig. 5.18. In Fig. 5.18a the efficiency of selecting $t\bar{t}$ events with a scalar sum of $p_T > 20$ GeV generator jets that is greater than 200 GeV for a range of thresholds on the Level-1 H_{T} is plotted against the minimum bias rate. When many jets are summed in this way the application of PUS makes a significant difference. Applying a seed threshold on the selected jets performs well. Applying the chunky donut or ρ-area PUS methods on top of a seed threshold obtains the best result.



(a) Lead jet $p_T > 150$ GeV

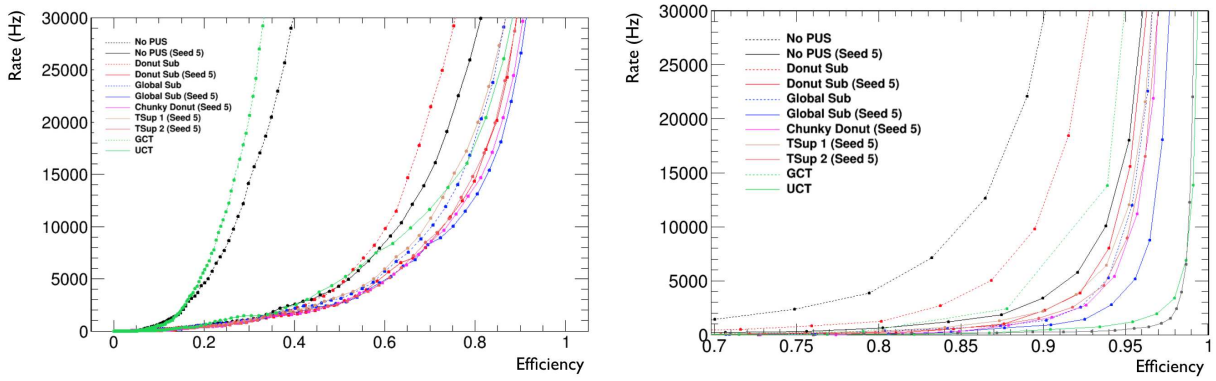
(b) Fourth leading jet $p_T > 50$ GeV

Figure 5.17: The normalised minimum bias rate (in Hz) against efficiency for a variety of thresholds of jet triggers made from jets with various PUS schemes. Based on $t\bar{t}$ and minimum bias MC simulation.

Figure 5.18b shows a similar rate and efficiency curve for the $\not{H}_{\mathrm{T}}$ variable. For reference on this plot the $\not{E}_{\mathrm{T}}$ rate and efficiency, labelled *MET*, is also shown. As a vector sum of jets is not made in the interim trigger, the curve with the UCT label is a form of $\not{E}_{\mathrm{T}}$ constructed from RCT regions. As $\not{E}_{\mathrm{T}}$ uses the full granularity TT information without any clustering it has an improved resolution, but loses robustness as the inputs are not calibrated and are more susceptible to influence from detector noise. Of the rate and efficiency curves constructed from a vector sum of jets, it is again observed that a seed threshold with chunky donut or ρ-area PUS give the best performance.

5.5.1 Conclusions

After investigating the performance of the upgrade trigger algorithms with simulation, it can be concluded that the upgrade algorithm presents a significant improvement over the algorithm used in the GCT during Run 1 of the LHC. The application of PUS appears most useful when summing jets to make trigger decisions. This is due to the fact that PU has the biggest effect on the energy of lower p_{T} jets from the primary vertex and is most likely to produce low p_{T} jets. Overall, it can be inferred that the application of a seed threshold effectively removes jets that originate from PU. It is then possible to successfully correct the energy of jets originating from the hard scatter with both local and global PUS methods. The chunky donut algorithm performs most



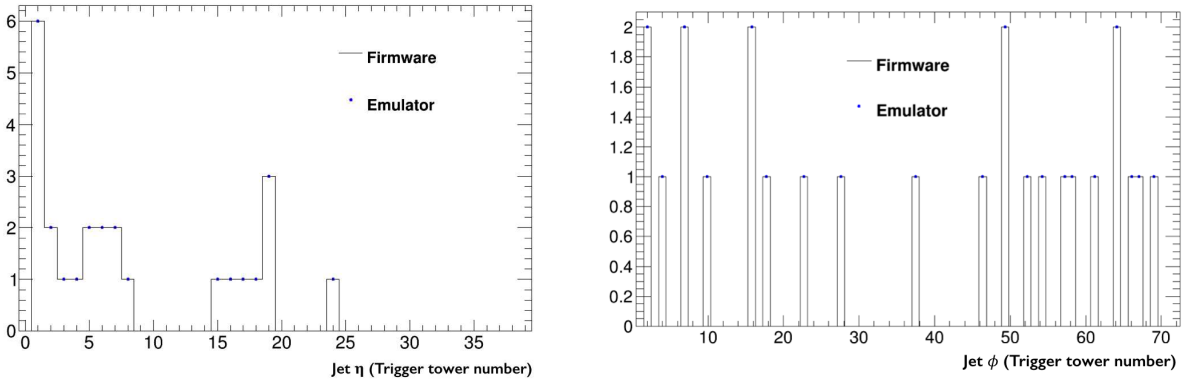
(a) $H_{\mathrm{T}} > 200$ GeV **(b)** $\not{H}_{\mathrm{T}} > 200$ GeV

Figure 5.18: The normalised minimum bias rate (in Hz) against efficiency for a variety of thresholds of energy sum triggers made from jets with various PUS schemes. Based on $t\bar{t}$ and minimum bias MC simulation.

effectively for local PUS. The ρ-area form of PUS seems to perform slightly better, but comes with significant latency penalties when implemented in hardware. This is particularly true when ρ-area PUS is applied with a seed, as two jet collections must be made, those with a seed and those without for the ρ calculation. Local forms of PUS are also more robust against localised detector faults, which can lead to issues with the trigger inputs. With these considerations, it is concluded that the 9×9 TT jets with a 2.5 GeV seed threshold and chunky donut PUS is the most promising candidate for an upgrade algorithm.

5.6 Firmware emulation and testing

After choosing a promising jet finder algorithm, it must be implemented within the firmware of the FPGA boards that will be used within the Level-1 trigger. To provide a reference for validation of the algorithm in firmware, a version is implemented in software, known as the *emulator*. The software is written to behave in a way that is very close to the implementation in the FPGA hardware. This emulator is then included in the collaboration wide *CMS software* and can be used to process data and simulation to produce an emulated Level-1 trigger output. This is particularly useful when testing the Level-1 trigger in simulation, for example when measuring trigger efficiencies for a particular physics analysis. It is also used to check that the Level-1 hardware is performing as expected during the online data collection.



(a) (b)

Figure 5.19: Demonstration of agreement between the trigger algorithm implemented in firmware and emulated in software for simulation studies

In the early stages of the firmware implementation, the jet finding was validated against the algorithm in the emulator. A collection of simulated events were processed by both the firmware within an FPGA and the emulator. After checking several of the jets were correctly reconstructed by hand, the output of the firmware and emulator were compared. As is shown in Fig. 5.19, the emulator and firmware implementations both find jets in exactly the same positions in the detector. These plots have a limited number of jets but demonstrate a first step towards a more comprehensive validation of the emulator. This kind of testing was extensively used to check the behaviour of the algorithm in the firmware and to ensure the trigger decisions were properly emulated in simulation.

The chunky donut 9×9 TT jet finder algorithm has been fully implemented into the upgrade hardware and successfully tested. It was commissioned throughout the data collection during 2015 and has been used to make the Level-1 trigger decisions when collecting data during Run 2 from 2016 onwards.

Chapter 6

Analysis strategy and event selection

Given the theoretical motivation for BSM physics presented in Chapter 2, the next chapters of this thesis describe a search for SUSY with the CMS detector in $\sqrt{s} = 13$ TeV proton collisions produced by the LHC. The analysis is designed to target SUSY produced through a strong force proton interaction that decays hadronically via SM particles to a weakly interacting LSP. The typical SUSY models considered in this search are those with gluino or squark pair production. However, it is possible to generalise the search to look for any BSM signatures that have a weakly interacting particle with significant momentum in the final state.

As SUSY has a significant number of free parameters, it can manifest itself in a wide range of decay topologies. For this reason the analysis is aimed to be as inclusive as possible, covering as much parameter space as is feasible. This is dictated by an ability to predict the backgrounds from SM processes with minimal uncertainties. The search then attempts to find statistically significant excesses in the data counts above those predicted for the various backgrounds. To maximise the chance of seeing BSM phenomena, events are typically required to have significant hadronic activity and missing momentum.

6.1 Challenges for a BSM search with jets and $\not{E}_{\mathrm{T}}$

The main challenge for a BSM search is in limiting and understanding the backgrounds, while maximising acceptance to a wide range of potential BSM signals. Within the analysis presented in this chapter, the backgrounds are considered in two separate categories, either as coming from a QCD multijet process or from another SM process

with neutrinos in the final state. As the search can look for generic signatures that are indicative of BSM processes, considerations of signal modelling are not as important as constraining and understanding these backgrounds.

6.1.1 The QCD multijet background

The most abundant SM process in pp-collisions at the LHC is QCD multijet production through strong force interactions. This background is orders of magnitude greater than other processes, as is evident from the total LHC cross section in Fig. 3.2. The majority of the QCD background consists of events with multiple jets, most commonly two, produced in a balanced configuration, with no significant missing momentum, $\not{E}_{\mathrm{T}}$ or $\not{H}_{\mathrm{T}}$. A requirement on these variables therefore significantly reduces the background. As the abundance of such events is so large, however, mismeasurement of the jets can lead to a significant QCD background passing such requirements. The fake missing energy can be introduced in a variety of ways:

- Detector effects leading to inefficiencies, such as jets falling into regions of the detector that are uninstrumented or have a reduced response due to hardware issues.
- Fake additional energy being introduced. This can occur through detector malfunctions, for example from problems in the readout electronics. It can also be introduced through PU that is attributed to the primary vertex. Additionally, beam interactions that produce particles outside the detector can introduce fake $\not{E}_{\mathrm{T}}$ signatures.
- Issues at the reconstruction stage, such as under or over-corrections of jet energies, or through the reconstruction algorithms missing physics objects that were produced.
- The energy of physics objects can be lost if their total energy is below a specific selection threshold.

Various techniques are applied to reduce the impact of these effects, as described throughout this chapter. Due to the nature of these effects, which are often time dependent, they cannot be fully accounted for in simulation. This fact, coupled with the difficulty in accurately simulating QCD processes, due to uncertainties in the

theoretical calculations, makes constraining the multijet background a significant challenge.

It is also possible for the QCD multijet background to be produced in association with genuine $\not{E}_{\mathrm{T}}$. This most often occurs in the case of *heavy flavour* QCD, where a $b\bar{b}$ quark pair are produced and at least one the b-quarks decays through the production of a neutrino. This is subdominant to the mismeasured multijet background, but must still be accounted for.

6.1.2 Backgrounds from SM processes with genuine $\not{E}_{\mathrm{T}}$

The other class of SM backgrounds is those produced with genuine $\not{E}_{\mathrm{T}}$. These processes only occur in the SM through the electroweak production of neutrinos. Due to the mass scale of the W and Z bosons, they are commonly produced in association with jets of significant momenta. The dominant processes that constitute this background are $\mathrm{W}+\mathrm{jets}$, $\mathrm{Z}+\mathrm{jets}$ and $t\bar{t}$, with residual backgrounds from other, lower cross section, vector-boson or top production processes.

As a hadronic analysis does not look for leptons in the final state, vetoing isolated leptons can help to reduce the background from the $\mathrm{W}+\mathrm{jets}$ and $t\bar{t}$ processes. In these cases, a lepton is always produced in association with the neutrino that results in the significant $\not{E}_{\mathrm{T}}$. Such events will only present a problem when the leptons are not reconstructed, or do not appear to be isolated. Additionally, if the W boson decays via a τ-lepton that subsequently decays hadronically, the event will not be removed by the lepton veto. This is a comparatively small background, however.

The major irreducible background in BSM searches with $\not{E}_{\mathrm{T}}$ comes from $\mathrm{Z} \to \nu\bar{\nu}$ processes. They produce a pure missing energy signature with no associated leptons to veto. Understanding this background to a high degree of accuracy is of great importance in hadronic BSM searches. However, due to the better simulation of electroweak processes compared to QCD processes, the simulation can be used to help understand this background with smaller uncertainties than the QCD background.

6.2 The α_{T} analysis

The search for SUSY described within this thesis revolves around suppression of the QCD multijet background to negligible levels through the use of topological variables, including α_{T}, described in Sec. 6.3. With negligible QCD, the large uncertainties that are usually prevalent when predicting this background are greatly reduced. The use of these topological variables also provides robustness against a wide range of mismeasurement effects. These effects are particularly relevant when looking at early data, when there is a limited understanding of potential detector issues. The remaining background from electroweak processes is then estimated in a *data-driven* way, minimising the uncertainties from the use of simulation.

During Run 1 of the LHC, the α_{T} analysis was used to search for SUSY. Datasets of 4.98 fb^{-1} at $\sqrt{s} = 7$ TeV and 18.5 fb^{-1} at $\sqrt{s} = 8$ TeV have been analysed, setting limits on SUSY production within the context of simplified models [109–113]. The analysis presented within this thesis is performed on the data collected during Run 2 of the LHC at $\sqrt{s} = 13$ TeV. The analysis has been redesigned and optimised to perform well at the higher centre of mass energy. Significant improvements have been made to improve the sensitivity of the analysis and in characterising the various backgrounds. The first iteration of a result with 2.3 fb^{-1} of data can be found at [114].

6.3 QCD multijet suppression with topological variables

As discussed in Sec. 6.1.1, a pure requirement on the $\not{E}_{\mathrm{T}}$ of an event is not always enough to remove the QCD multijet background. To effectively remove it, the mismeasurement that leads to the fake $\not{E}_{\mathrm{T}}$ signature must be identified and used to remove mismeasured events. This can be achieved by making use of topological variables, which use specific properties of the topology of an event to decide whether the $\not{E}_{\mathrm{T}}$ comes from mismeasurement or a weakly interacting particle.

6.3.1 The α_{T} variable

The dimensionless variable, α_{T}, was originally proposed as the α variable [115], but changed to a transverse quantity to make it more suitable for use in a hadron collider

[116, 117]. It is intrinsically robust against jet energy mismeasurements in multijet systems. For a dijet system, α_{T} is defined as:

$$\alpha_T = \frac{E_T^{j_2}}{M_T}, \tag{6.1}$$

where $E_T^{j_2}$ is the transverse energy of the lower energy jet and M_T is the invariant mass of the dijet system, defined as:

$$M_T = \sqrt{(\Sigma E_T^{j_i})^2 - (\Sigma p_x^{j_i})^2 - (\Sigma p_y^{j_i})^2}, \tag{6.2}$$

where $E_T^{j_i}$ is the transverse energy of jet, j_i and the x and y components of the transverse momentum are $p_x^{j_i}$ and $p_y^{j_i}$ respectively.

In the case that the event in question is a perfectly measured dijet event, $E_T^{j_1} = E_T^{j_2}$ and both jets are back-to-back in ϕ. This results in a value of $\alpha_{\mathrm{T}} = 0.5$ when the momentum of each jet is large in comparison with its mass, as is usually the case in QCD multijet events. If the jets are still back-to-back but one of them is mismeasured, this will result in a value of $\alpha_{\mathrm{T}} < 0.5$. However, in the case that the two jets are recoiling from a genuine source of $\not{E}_{\mathrm{T}}$, they will not be back-to-back and generally $\alpha_{\mathrm{T}} > 0.5$.

For events with more than two jets, a *pseudo-dijet* system is formed by summing jets vectorially to combine them. The system chosen is one that minimises ΔE_T, the difference between the E_T of each pseudo-jet, where E_T is the scalar sum of the transverse energies of all the jets in each pseudo-jet. This form of clustering is chosen as it forms the most balanced configuration, making the pseudo-dijet event appear as close to an event with no mismeasurement as possible. It leads to a generalised form of α_T:

$$\alpha_T = \frac{\Sigma E_T^{j_i} - \Delta E_T}{2\sqrt{(\Sigma E_T^{j_i})^2 - \not{H}_T^2}}. \tag{6.3}$$

In the case that there is no missing energy, $\Delta E_T = \not{H}_{\mathrm{T}} = 0$ and $\alpha_{\mathrm{T}} = 0.5$. However, if the energy of the jets is mismeasured, the value of ΔE_T will be very close to $\not{E}_{\mathrm{T}}$, resulting in $\alpha_{\mathrm{T}} < 0.5$. When the jets are recoiling from genuine $\not{E}_{\mathrm{T}}$, $\Delta E_T \sim 0$ and $\alpha_{\mathrm{T}} > 0.5$.

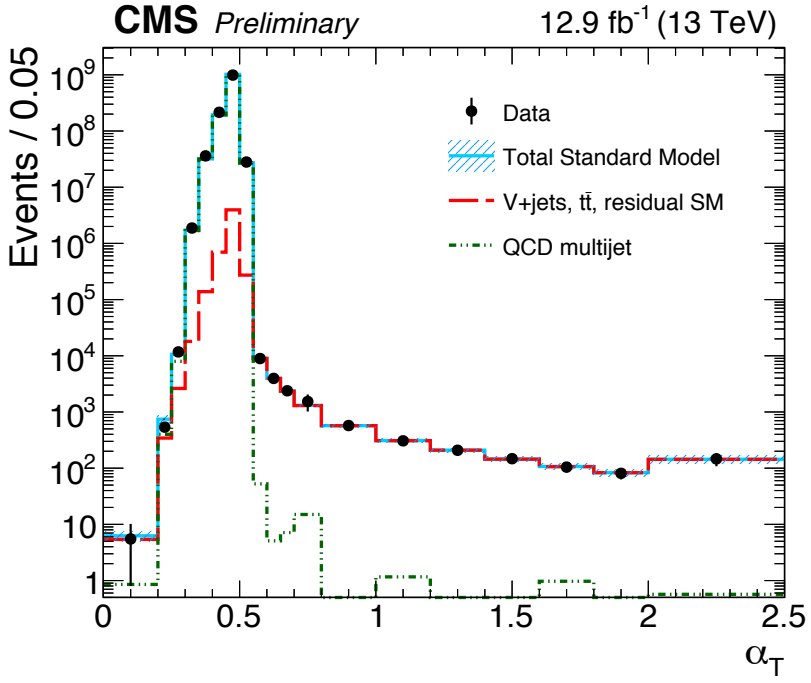


Figure 6.1: The α_T distribution for events with $H_T > 200$ GeV that pass a pre-selection criteria (Sec. 6.6) when $\alpha_T < 0.55$ and a signal selection criteria (Sec. 6.7) when $\alpha_T > 0.55$. The green dotted line shows the expected multijet QCD background that can be removed with an appropriate cut on α_T

The ability of an α_T requirement to remove the QCD multijet background is evident in Fig. 6.1. By choosing an appropriate cut above 0.5 it is possible to reduce the multijet background to a negligible level.

6.3.2 The $\Delta\phi^*_{\text{min}}$ variable

To further suppress the QCD multijet background after applying an α_T cut, the biased-$\Delta\phi$, $\Delta\phi^*_{\text{min}}$, variable is introduced. This is a variation on the $\Delta\phi$ variable, which is usually defined as the minimum azimuthal angle between the $\not{E}_T$ and any jet in the event. When calculating $\Delta\phi^*_{\text{min}}$, each jet is compared to the $\not{H}_T$, but the $\not{H}_T$ is recalculated as if the jet were not present in the event. This leads to the definition:

$$\Delta\phi^*_{\text{min}} = \min(\Delta\phi(\vec{p^{j_i}_T}, \vec{\not{H}^{j_i}_T})), \tag{6.4}$$

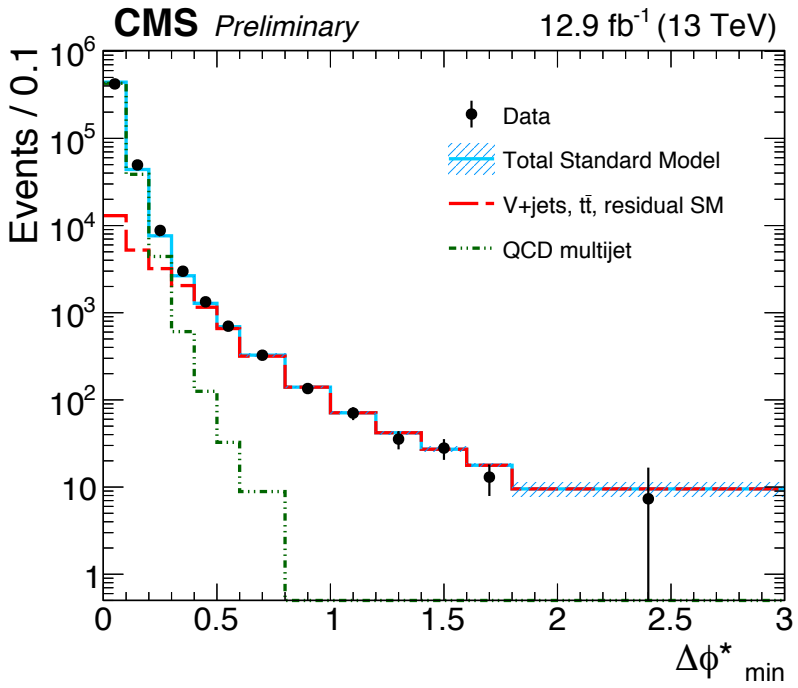


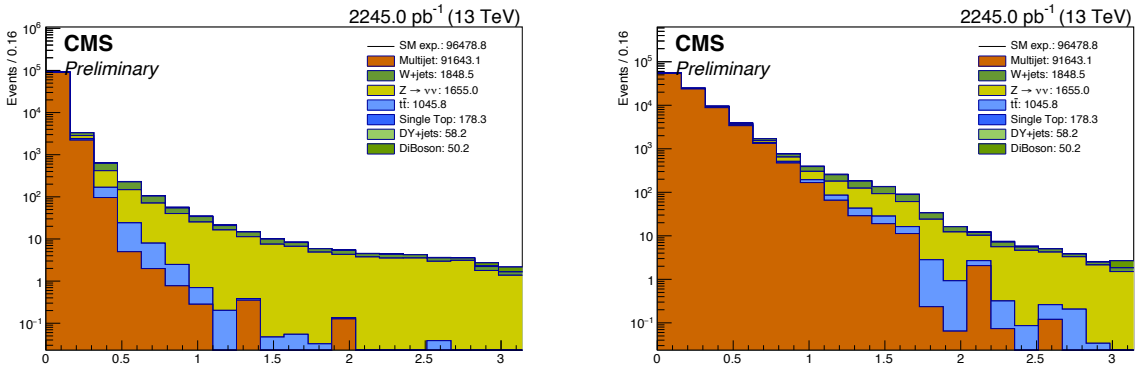
Figure 6.2: The $\Delta\phi^*_{\text{min}}$ distribution for events with $H_T > 800$ GeV that pass a pre-selection criteria (Sec. 6.6). The green dotted line shows the expected multijet QCD background that can be removed with an appropriate cut on $\Delta\phi^*_{\text{min}}$

considered over every jet, j_i, and where $\vec{\not{H}}_{\text{T}}^{\,j_i} = \vec{\not{H}}_{\text{T}} + \vec{p}_T^{\,j_i}$. The removal of the probe jet from the event adds robustness to over-measurement, as well as under-measurement, of the jet energies.

The $\Delta\phi^*_{\text{min}}$ variable acts to remove events in which the $\not{E}_{\text{T}}$ is collinear with the jets, a typical feature of mismeasurement. However, it also helps to remove the heavy flavour QCD multijet background, where a $b\bar{b}$ pair is produced and decays leptonically, producing neutrinos. These neutrinos are a genuine source of $\not{E}_{\text{T}}$ that are typically boosted along the jet direction. In these cases the $\Delta\phi^*_{\text{min}}$ helps to reject the QCD background that may have passed an α_{T} cut. This can happen in events with more than two jets, where the jet collinear to the $\not{E}_{\text{T}}$ is combined into a pseudojet in the α_{T} calculation in a way that masks the fact that there is $\not{E}_{\text{T}}$ along a jet axis. The $\Delta\phi^*_{\text{min}}$ distribution for multijet events and the remaining SM backgrounds is shown in Fig. 6.2. A requirement on $\Delta\phi^*_{\text{min}}$ above ~ 0.5 will significantly reduce any QCD multijet background.

To further demonstrate the efficacy of the $\Delta\phi^*_{\text{min}}$ variable, Fig. 6.3 shows a comparison of the abilities of the $\Delta\phi^*_{\text{min}}$ variable to control the QCD multijet background with

a similar jet-$\not\!\!H_T$ angular variable $\Delta\phi(j_{1234},\not\!\!H_T)_{\text{min}}$, the minimum azimuthal separation between the $\not\!\!H_T$ -vector and the leading four jets. The $\Delta\phi^*_{\text{min}}$ variable exhibits a distribution that is more sharply peaked for the QCD multijet background at low values and faster falling than $\Delta\phi(j_{1234},\not\!\!H_T)_{\text{min}}$. This demonstrates the ability of $\Delta\phi^*_{\text{min}}$ to provide a better control of the QCD background while retaining acceptance of events with genuine $\not\!\!H_T$, in this case represented by the V+jets, $t\bar{t}$ and other residual SM backgrounds with genuine $\not\!\!E_T$.



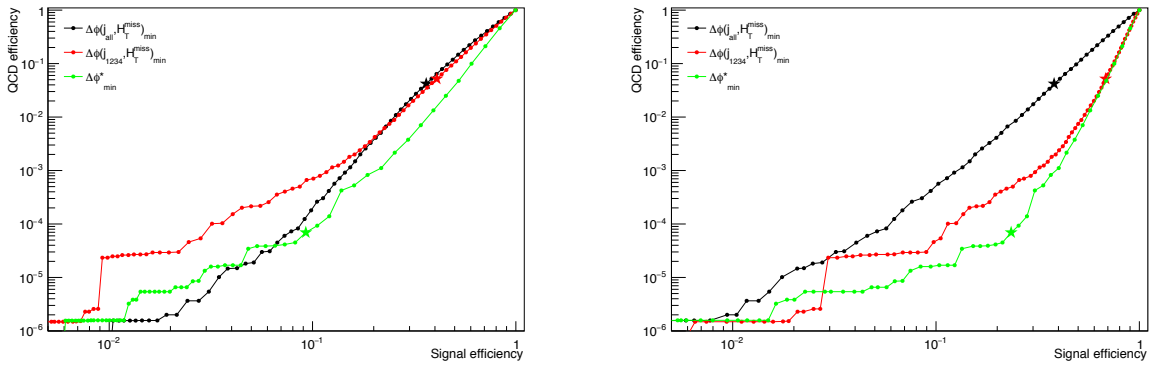
(a) $\Delta\phi^*_{\text{min}}$ distribution.

(b) $\Delta\phi(j_{1234},\not\!\!H_T)_{\text{min}}$ distribution.

Figure 6.3: $\Delta\phi^*_{\text{min}}$ and $\Delta\phi(j_{1234},\not\!\!H_T)_{\text{min}}$ distributions of MC simulation of the dominant analysis backgrounds after analysis selections for $H_T > 800$ GeV.

This is further demonstrated in Fig. 6.4, where the efficiency of retaining processes with genuine $\not\!\!H_T$ is plotted against the QCD background efficiency for a series of $\Delta\phi^*_{\text{min}}$, $\Delta\phi(j_{1234},\not\!\!H_T)_{\text{min}}$ and $\Delta\phi(j_{all},\not\!\!H_T)_{\text{min}}$ requirements, where $\Delta\phi(j_{all},\not\!\!H_T)_{\text{min}}$ considers all jets rather than just the leading four. The points corresponding to a requirement of 0.5 are highlighted as stars on the plot. As the general analysis strategy involves reducing the QCD multijet background to a negligible level while maximising signal acceptance, these plots demonstrate that this is most achievable with a cut on $\Delta\phi^*_{\text{min}}$. For the same threshold requirement of 0.5 on each variable, the $\Delta\phi^*_{\text{min}}$ variable provides an efficiency for multijet events that is approximately three orders of magnitude lower than for $\Delta\phi(j_{1234},\not\!\!H_T)_{\text{min}}$ at the cost of approximately a factor 3 reduction in signal efficiency. The threshold on $\Delta\phi^*_{\text{min}}$ required to give an approximately equivalent suppression of QCD achieved with the $\Delta\phi(j_{1234},\not\!\!H_T)_{\text{min}}$ variable is larger than 1.5, at the cost of a loss of a factor 5 in signal acceptance with respect to $\Delta\phi^*_{\text{min}}$. This also holds true in the extreme case of a high jet multiplicity signal model. Despite performing similarly to $\Delta\phi^*_{\text{min}}$ for separating the non-multijet from multijet

backgrounds, the $\Delta\phi(j_{all}, \not{H}_{\mathrm{T}})_{\mathrm{min}}$ variable performs worse for a high jet multiplicity signal model than $\Delta\phi^*_{\mathrm{min}}$.



(a) Acceptance of SM backgrounds with genuine $\not{E}_{\mathrm{T}}$ vs QCD acceptance

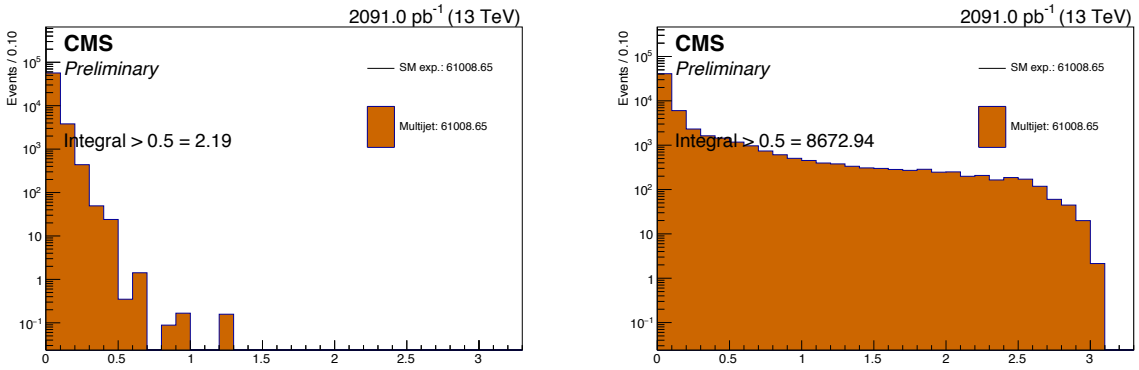
(b) High jet multiplicity SUSY model acceptance vs QCD acceptance

Figure 6.4: $\Delta\phi^*_{\mathrm{min}}$, $\Delta\phi(j_{1234}, \not{H}_{\mathrm{T}})_{\mathrm{min}}$ and $\Delta\phi(j_{all}, \not{H}_{\mathrm{T}})_{\mathrm{min}}$ efficiency for simulation of processes with genuine $\not{E}_{\mathrm{T}}$ vs QCD multijet background efficiency. The stars correspond to efficiencies with a cut of 0.5 on each variable. A generic case of non-multijet process efficiency is considered in (a). In (b) we consider an uncompressed SUSY model where gluinos are produced and decay via four tops to a pair of LSPs.

Additionally, the $\Delta\phi^*_{\mathrm{min}}$ variable displays robustness in the presence of severe event mismeasurement, which is not present in $\Delta\phi(j_{1234}, \not{H}_{\mathrm{T}})_{\mathrm{min}}$. A mismeasurement is simulated by artificially lowering the p_{T} of the jet that minimises the azimuthal separation variable to 41 GeV. Due to the removal of the probe jet from the computation of $\Delta\phi^*_{\mathrm{min}}$, the distribution of angular separation Fig. 6.5a is remains unchanged under severe mismeasurement. The $\Delta\phi(j_{1234}, \not{H}_{\mathrm{T}})_{\mathrm{min}}$ variable is sensitive to such mismeasurement as both $\not{H}_{\mathrm{T}}$ and the rank of the leading four jets are affected, resulting in a broader distribution with increased leakage, shown in Fig. 6.5b.

6.3.3 The missing energy ratio $\not{H}_{\mathrm{T}}/\not{E}_{\mathrm{T}}$

After requirements are made on the α_{T} and $\Delta\phi^*_{\mathrm{min}}$ variables, the majority of events with $\not{H}_{\mathrm{T}}$ introduced from jet mismeasurement are removed. However, this does not take account of events in which jets fall below their reconstruction threshold, detailed for this analysis in Sec. 6.4. If several jets fall just below threshold they can fake a significant $\not{H}_{\mathrm{T}}$. There is a much lower threshold on the particles that go into the $\not{E}_{\mathrm{T}}$ calculation than the jets that go into the $\not{H}_{\mathrm{T}}$ threshold. Therefore, a requirement on



(a) QCD $\Delta\phi^*_{\text{min}}$ distribution with mismeasurement. **(b)** QCD $\Delta\phi(j_{1234}, \not{H}_{\text{T}})_{\text{min}}$ distribution with mismeasurement.

Figure 6.5: $\Delta\phi^*_{\text{min}}$ and $\Delta\phi(j_{1234}, \not{H}_{\text{T}})_{\text{min}}$ distributions of QCD multijet simulation after analysis selections for $H_{\text{T}} > 800$ GeV in the case of severe mismeasurement. The total number of events that pass a $\Delta\phi > 0.5$ selection of the respective quantity are indicated.

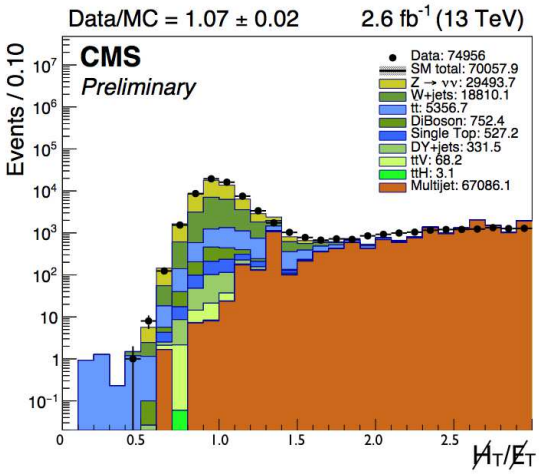


Figure 6.6: The $\not{H}_{\text{T}}/\not{E}_{\text{T}}$ distribution for MC simulation with 2.3 fb^{-1} of $\sqrt{s} = 13$ TeV data overlaid.

the ratio of these two variables helps to remove the background from physics objects that fall below threshold. The extent to which this can remove the remaining QCD multijet background after the α_{T} and $\Delta\phi^*_{\text{min}}$ cuts is shown in Fig. 6.6, where a typical requirement is $\not{H}_{\text{T}}/\not{E}_{\text{T}} < 1.25$.

6.4 Physics objects

The analysis makes use of physics objects that are reconstructed with the algorithms described in Chapter 4. As there are many different input parameters for each of the algorithms, the specific details relevant to the analysis are described in this section. Jets are reconstructed to characterise the hadronic activity within the event. Muons, electrons and photons are reconstructed for a signal region veto and for selecting the *control regions* that are used for background estimation, detailed in Sec. 6.8. *Isolated tracks* are additionally reconstructed to provide an extra type of lepton veto when the leptons are not fully reconstructed.

6.4.1 Jets

Jets are defined as sets of PF candidates clustered by the anti-k_T jet clustering algorithm with a distance parameter of 0.4 and they are cross-cleaned of leptons. Additionally, CHS is applied, charged hadrons that can be traced back to PU vertices are not clustered. The four-vectors of the jets are then corrected with the procedure outlined in Sec. 4.5.3.

The *loose* working point jet quality criteria are chosen, defined by the selections listed in Tab. 6.1. All jets are required to contain at least part of their energy within the ECAL and HCAL and must consist of at least one of each different type of particle. This is effective at removing fake jets that are the result of detector noise.

Jets are also b-tagged with the *medium* working point of the algorithm described in Sec. 4.5.4. This is obtained with a cut of > 0.800 on the algorithm discriminator variable. This results in a gluon/light-quark mis-tag rate of ~ 1 % (where *light* means u, d and s quarks), a charm-quark mis-tag rate of ~ 10 % and a b-quark b-tag efficiency of about 60 %.

6.4.2 Muons

Two types of selection are made on muons depending on whether they are used to carry out the veto in the signal region or used for selecting one of the control regions. Tighter criteria are used for the control region, muons are required to be well isolated

Table 6.1: The *loose* jet ID requirements.

Variable	cut	notes
$-3.0 < \eta_{\text{jet}} < 3.0$		
Neutral Hadron Fraction	< 0.99	-
Neutral Electromagnetic Fraction	< 0.99	-
Number of constituents	> 1	-
Charged Hadron Fraction	> 0	only for $\lvert\eta_{\text{jet}}\rvert < 2.4$
Charged Multiplicity	> 0	only for $\lvert\eta_{\text{jet}}\rvert < 2.4$
Charged Electromagnetic Fraction	< 0.99	only for $\lvert\eta_{\text{jet}}\rvert < 2.4$
$\lvert\eta_{\text{jet}}\rvert > 3.0$		
Neutral Electromagnetic Fraction	< 0.90	-
Number of Neutral Particles	> 10	-

and be reconstructed with the global algorithm, as detailed in Sec. 4.4. The muon must consist of a track with at least 5 inner layer hits, a good fit χ^2, good compatibility with the primary vertex and significant muon chamber hits. The isolation in the control region uses the relative isolation algorithm defined in Sec. 4.6 with effective-area correction, it is required that $I^{\text{rel}} < 0.15$.

For the purpose of vetoing muons in the signal region, a looser working point is used, which provides $\sim$98% efficiency. The muon is just required to be classified within the PF algorithm and reconstructed with the global or tracker muon algorithms. The mini-isolation algorithm is utilised with effective-area PU correction $I^{\text{rel}}_{\text{mini}} < 0.2$. The reconstructed muon must have a $p_{\text{T}} > 10$ GeV and be within $|\eta| < 2.1$.

6.4.3 Photons

Photons are identified according to a *tight* working point definition ($\sim$ 71 % efficiency) and required to be well isolated. A PF-based isolation which considers neutral and charged hadrons separately is used with a cone size $\Delta R < 0.3$ and effective-area corrections are applied to remove the effects of pileup. Table 6.2 summarises the isolation selection used. There are additional requirements on the ratio of HCAL and ECAL deposits and the kinematics of the deposits within the ECAL. The reconstructed photon must have a $p_{\text{T}} > 25$ GeV and be within $|\eta| < 2.5$.

Table 6.2: Photon isolation criteria (*tight* working point). The energy of particles within the isolation cone must be less than the value in the column, where p_T^γ is the p_T of the photon in GeV.

Categories	Barrel	EndCap
PF charged hadron isolation	1.66 GeV	1.04 GeV
PF neutral hadron isolation	$0.14 + e^{0.0028 \times {p_\mathrm{T}}^\gamma + 0.5408}$	$3.89 + 0.0172 \times {p_\mathrm{T}}^\gamma$
PF photon isolation	$1.40 + 0.0014 \times {p_\mathrm{T}}^\gamma$	$1.40 + 0.0091 \times {p_\mathrm{T}}^\gamma$

6.4.4 Electrons

In order to veto electrons a *loose* working point definition ($\sim$90% efficiency) is used. There are requirements on the ratio of HCAL and ECAL deposits, criteria to remove photons that have converted within the tracker, a minimum number of tracks required and kinematic requirements on the deposit within the ECAL. Electrons are also required to be isolated with the effective-area corrected mini-isolation algorithm. Isolated electrons are defined by $I^{\mathrm{rel}}_{\mathrm{mini}} < 0.1$. The reconstructed electron must have a $p_\mathrm{T} > 10$ GeV and be within $|\eta| < 2.5$.

6.4.5 Isolated tracks

A single isolated track comprises a charged PF candidate with $p_\mathrm{T} > 10\,\mathrm{GeV}$, $\Delta z(\mathrm{track}, \mathrm{PV}) < 0.05\,\mathrm{cm}$ and with a relative isolation smaller than 0.1, where the isolation is determined from the sum of the p_T of the charged PF candidates within $\Delta R < 0.3$. It is useful for identifying W bosons through their leptonic decays, when the lepton is not fully reconstructed. Single prong decays of the tau lepton can also be identified: $\tau \to \mathrm{h}^\pm + \mathrm{n}\pi^0\,\nu$.

6.4.6 Energy sums

The $\not{E}_\mathrm{T}$ is reconstructed as described in Sec. 4.7 and the Type-I $\not{E}_\mathrm{T}$ energy correction is applied. The $\not{E}_\mathrm{T}$ is used in the definition of the transverse mass of potential W bosons, M_T, which is in turn used as part of the selection criteria that define the single muon control sample (Sec. 6.8). It is also used for the $\not{H}_\mathrm{T}/\not{E}_\mathrm{T}$ cleaning filter, described in Sec. 6.3.3.

The H_T and $\not{H}_T$ energy sums are constructed with the jets outlined in Sec. 6.4.1, which are subject to the kinematic requirements of $p_T > 40$ GeV and $|\eta| < 3$.

6.5 Trigger strategy

As SUSY models have a lot of freedom in how they are manifest, a guiding principle of the analysis is to maintain acceptance of as much phase space as possible. Despite many SUSY models being ruled out at $O(100$ GeV$)$ energies, models with compressed spectra are still poorly constrained in this regime. The trigger strategy therefore revolves around maintaining as low an energy threshold as possible given the very high rate of QCD multijet processes. To select for events with significant hadronic activity and missing energy, triggers are used that rely on the H_T and $\not{H}_T$ of the events. Additionally, including topological variables, such as α_T, in the HLT allows events to be collected with H_T as low as 200 GeV.

So the rate of events that is processed by the HLT is not too high, HLT triggers are seeded with a Level-1 trigger requirement that is based on H_T or $\not{E}_T$. The upgrade of the hadronic trigger algorithms, described in Chapter 5, has allowed low thresholds to be maintained in the challenging Run 2 environment. The exact requirements made are given in Tab. 6.3.

To collect events that are of general use to hadronic analyses, the H_T of each event is reconstructed within the HLT. This is carried out with a custom form of PF, that is designed to work in a way that is quick enough given the time the trigger has to make a decision. To further help with this timing constraint, the H_T is calculated with jets clustered from only calorimeter deposits and a looser constraint is made, reducing the load on the PF reconstruction. This allows for all events that pass $H_T > 800$ GeV, calculated by the HLT, to be collected. When carrying out full online reconstruction, this results in a $\sim$100% efficiency for events with an offline $H_T > 900$ GeV.

To be able to efficiently collect events with lower values of H_T, the α_T variable is calculated within the HLT. A requirement is made on both α_T and H_T for a series of different H_T thresholds. These requirements are chosen to limit the rate of the α_T-H_T triggers to an acceptable level, while maintaining low acceptance of events with low H_T.

There are also general purpose missing energy triggers that require both an $\not{H}_T$ and $\not{E}_T$ threshold to be surpassed within the HLT. Including events selected by these triggers helps to increase the efficiency across the phase space. These variables, however, are quite sensitive to running conditions and mismeasurement, so are not as robust against the HLT rate as the α_T-H_T triggers. The full list of triggers used to collect data for the signal region of the α_T analysis can be seen in Tab. 6.3.

The events in the control regions are collected with muon or photon triggers, where the HLT requires at least one of these objects above a certain energy threshold. Due to the lower rate of events containing leptons, these triggers are at full efficiency for their requirements in the analysis. The specific triggers used are also listed in Tab. 6.3.

Table 6.3: Trigger thresholds of the Level-1 hardware trigger and HLT for the hadronic signal region and the leptonic control regions.

Analysis region	Level-1 requirements	HLT requirements
Signal	$H_T > 240$ GeV or $\not{E}_T > 70$ GeV	$H_T > 200$ GeV, $\alpha_T > 0.57, \not{E}_T > 90$ GeV
	$H_T > 240$ GeV or $\not{E}_T > 70$ GeV	$H_T > 250$ GeV, $\alpha_T > 0.55, \not{E}_T > 90$ GeV
	$H_T > 240$ GeV or $\not{E}_T > 70$ GeV	$H_T > 300$ GeV, $\alpha_T > 0.53, \not{E}_T > 90$ GeV
	$H_T > 240$ GeV or $\not{E}_T > 70$ GeV	$H_T > 350$ GeV, $\alpha_T > 0.52, \not{E}_T > 90$ GeV
	$H_T > 240$ GeV or $\not{E}_T > 70$ GeV	$H_T > 400$ GeV, $\alpha_T > 0.51, \not{E}_T > 90$ GeV
	$H_T > 240$ GeV	$H_T > 800$ GeV
	$\not{E}_T > 70$ GeV	$\not{E}_T > 90$ GeV or $\not{H}_T > 90$ GeV
μ + jets	${p_T}^{\mu} > 20$ GeV	${p_T}^{\mu} > 22$ GeV
$\mu\mu$ + jets	${p_T}^{\mu} > 20$ GeV	${p_T}^{\mu} > 22$ GeV
γ + jets	$H_T > 240$ GeV or ${p_T}^{\gamma} > 30$ GeV	$H_T > 800$ GeV or ${p_T}^{\gamma} > 175$ GeV

6.6 Pre-selection

To ensure that all events considered by the analysis have significant hadronic activity and missing energy, a pre-selection is carried out on all events considered in the signal and control regions. It is described within this section.

Events are required to contain at least one $p_T > 100$ GeV jet, with all other jets being considered if they have $p_T > 40$ GeV and are well reconstructed in the central region, $|\eta| < 3$. If any jets fall outside the η range, the event is vetoed. This ensures there is no significant energy deposited within the region of the detector with no tracker. As most BSM physics scenarios result in the production of particles at a high mass scale, they

are more likely to deposit their energy centrally than many of the SM backgrounds. Therefore, this *forward jet veto* does not significantly effect the signal efficiency for most models. The QCD multijet background consists of many soft scattering events, which typically deposit a significant proportion of energy within the forward region. As this is the case, the forward jet veto also helps to reduce the multijet background.

Significant hadronic activity is selected by requiring $H_T > 200$ GeV. This cut is primarily motivated by the trigger threshold, it is the lowest H_{T} value that can be reached with a reasonable trigger rate and efficiency. To ensure there is significant missing energy, it is required that $\not{H}_{\mathrm{T}} > 130$ GeV. This threshold is chosen as roughly equivalent to the magnitude of the $\not{H}_{\mathrm{T}}$ requirement made by the α_{T} cuts on dijet events within the signal region when there is no significant mismeasurement.

Additionally, to remove events with significant energy deposits that have not been reconstructed as jets, a requirement is made on the ratio $\not{H}_{\mathrm{T}}/\not{E}_{\mathrm{T}} < 1.25$, as motivated in Sec. 6.3.3. When this requirement is made for the control regions, the $\not{E}_{\mathrm{T}}$ is reconstructed ignoring the leptons or photons that define the particular region. This ensures a fair comparison with $\not{H}_{\mathrm{T}}$ and allows the lepton to simulate the missing momentum in the signal region.

To ensure that beam conditions or reconstruction effects have not produced a spurious $\not{E}_{\mathrm{T}}$, extra filters are applied. The filters used are designed to remove instances of fake $\not{E}_{\mathrm{T}}$ while retaining a high efficiency for real physics events, they are described in the following [91].

- The primary vertex must be well reconstructed, within $|z| \leq 24$ cm of the proton collision region and $d_{xy} < 2$ cm from the beam line
- Beam interactions exterior to the detector, *beam halo* effects, are removed by vetoing events with CSC and calorimeter energy deposits consistent with interactions from particles outside the detector
- Noise, caused by particle interactions with the read-out system and dead regions within the HCAL, are removed with the *HBHE noise and isolation filters*
- Anomalous signals within the ECAL endcap supercrystals are removed with the *ECAL Endcap SC Noise filter* and *ECAL Trigger Primitive (TP) filter*
- Events with misidentified straight tracks that are reconstructed to have a large p_{T} are removed by the *bad track filter*

Finally, two additional vetoes are applied to ensure events are purely hadronic with minimal misreconstruction. Any events that contain a single isolated track (SIT) with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and within $|\eta| < 2.5$ is vetoed. This reduces the background from single pronged τ decays and misreconstructed leptons. Within the muon control regions this track is ignored if it is within $\Delta R < 0.02$ of an identified lepton. Additionally, any events with jets that do not pass the loose requirements described in Sec. 6.4.1 are vetoed.

A summary of the pre-selection requirements is shown in Tab. 6.4.

Table 6.4: Summary of the pre-selection criteria.

Selection	Requirement
"MET filters"	Primary Vertex, CSC Beam Halo, HBHE Noise and Isolation, ECAL Endcap SC Noise, ECAL TP, bad track filter
Jet acceptance	$p_{\mathrm{T}} > 40\,\mathrm{GeV}$, $\|\eta\| < 3.0$
Lead jet acceptance	$p_{\mathrm{T}} > 100\,\mathrm{GeV}$, $\|\eta\| < 2.4$
Forward jet veto	$p_{\mathrm{T}} > 40\,\mathrm{GeV}$, $\|\eta\| > 3.0$
H_{T} requirement	$> 200\,\mathrm{GeV}$
$\not{H}_{\mathrm{T}}$ requirement	$> 130\,\mathrm{GeV}$
$\not{H}_{\mathrm{T}}/\not{E}_{\mathrm{T}}$ requirement	< 1.25
Single isolated track veto	$p_{\mathrm{T}} > 10\,\mathrm{GeV}$, $\|\eta\| < 2.5$

6.7 The signal region

The signal region selection is chosen to be inclusive for potential BSM signatures in the jets and missing momentum final state, while minimising the effects of mismeasurement and contributions from SM processes with genuine $\not{E}_{\mathrm{T}}$. To reduce the background from decays via W bosons, a veto is made on any leptons that are identified as described in Sec. 6.4. To maintain a fully hadronic final state any events containing reconstructed photons are also vetoed.

After the pre-selection requirements, there is still expected to be a significant multijet background in the signal region. This is reduced to a negligible level with a H_{T} dependent cut on the α_{T} of the event. Along with this, a cut is made on the $\Delta\phi^{*}_{\mathrm{min}}$ variable to ensure any multijet background that passes the α_{T} cut is removed. A summary of these cuts can be seen in Tab. 6.5. The α_{T} cuts are loosened as a function of H_{T} in correspondence with the trigger thresholds. At lower values of H_{T} the high

trigger rate requires a higher α_T threshold, due to the fact that the multijet background falls off as a function of H_T. In the case that events have just one identified jet, the value of α_T is undefined. In these *monojet* events there is no α_T cut. The $\Delta\phi^*_{min}$ variable is also undefined with $p_T < 40$ GeV jets. To get around this issue and provide multijet discrimination in monojet events, the p_T threshold on jets used to calculate the $\Delta\phi^*_{min}$ variable is reduced to 25 GeV. In the case that there are no other jets between 25 and 40 GeV the event is allowed into the signal region.

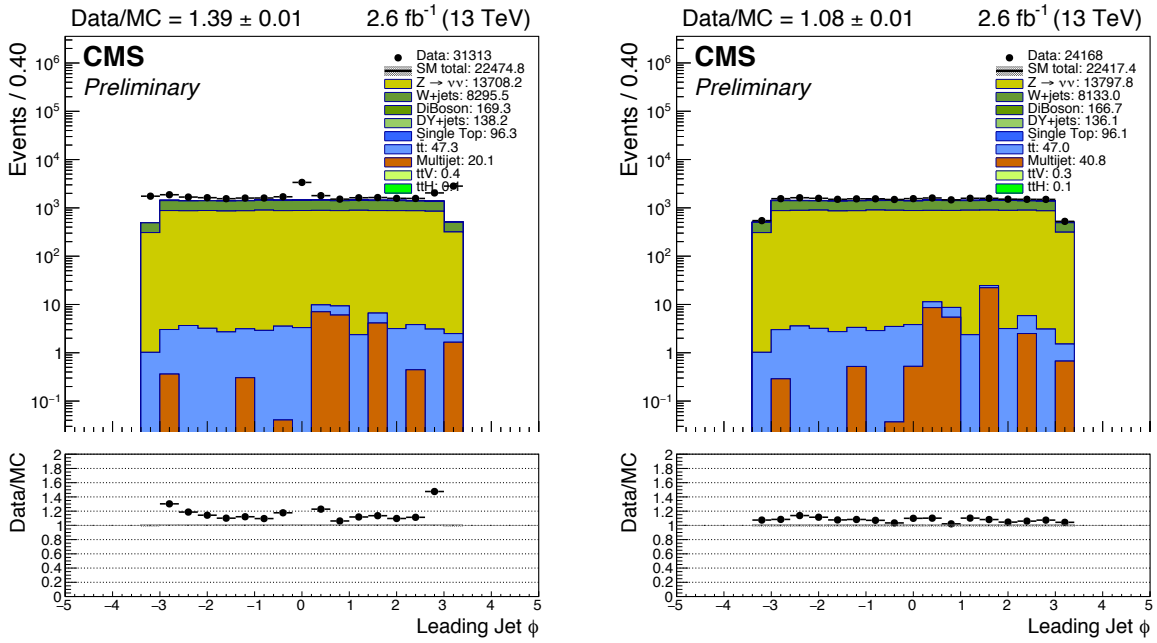
It is observed that the beam halo $\not{E}_T$ filter, described in Sec. 6.6, is ineffective in removing all cases of mismeasurement from the signal region. A requirement that the charged hadron fraction (CHF) of the lead jet is above 10% is enough to negate this effect. This cut is demonstrated in Fig. 6.7. Due to the fact that the proton beam is steered within the horizontal plane of the LHC, beam halo effects manifest themselves at ϕ values of 0 and π. The application of the CHF cut removes the spikes in the leading jet ϕ distribution at these values.

H_T (GeV)	200	250	300	350	400	500	600	>800
α_T threshold	0.65	0.60	0.55	0.53	0.52	0.52	0.52	–
$\Delta\phi^*_{min}$ threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table 6.5: The α_T and $\Delta\phi^*_{min}$ thresholds versus lower bound of H_T bin. For all H_T bins satisfying $H_T > 800$ GeV, no α_T cut is applied. No α_T requirement is imposed in the case that there is only one reconstructed jet.

6.8 The control regions

To carry out a data-driven estimation of the SM backgrounds, several control regions are defined. They are chosen to be orthogonal to the signal region by inverting at least one of the requirements. To estimate processes with genuine $\not{E}_T$, control regions are populated with events that contain at least one muon or photon. The muons or photons are then ignored when calculating any of the event level variables, such as α_T or $\Delta\phi^*_{min}$, to simulate the signature left by neutrinos or cases in which the lepton is not properly reconstructed. Extrapolation from these control regions with guidance from simulation is then used to estimate the background counts in the signal region, as described in Chapter 7. The pre-selection criteria are applied to each control region, which ensures they are in a similar phase-space to the signal region. As the requirement



(a) No CHF cut

(b) Lead jet CHF>0.1

Figure 6.7: Distributions in the signal region of the jet ϕ direction (a), and jet ϕ direction after applying a requirement of CHF > 0.1. The large excess in data at charged hadron fractions close to zero and $\phi = 0, \pi$ is consistent with beam halo effects, and is effectively suppressed by the aforementioned selection.

of an isolated lepton or photon already selects electroweak processes and rejects QCD multijet processes, there is no need for an α_{T} or $\Delta\phi^{*}_{\mathrm{min}}$ requirement in these control regions. Removing these requirements significantly increases the number of events available within each control region. This reduces statistical uncertainties when they are used for background estimation.

6.8.1 The μ + jets control region

The selection criteria for the μ + jets control region are chosen to select events containing W bosons that are produced in association with jets and decay to a muon and a neutrino. Exactly one muon that passes the control region selection criteria, described in Sec. 6.4.2, with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.1$ is required. The muon is required to be separated from other jets by $\Delta R(\mu, \mathrm{jet}) > 0.5$. If any other leptons or photons are present in the event then it is vetoed, as in the signal region.

To help ensure events containing W bosons are selected, a requirement is also made on the transverse mass of the W candidate. It is approximated by taking the transverse mass between the $\not{E}_{\mathrm{T}}$ and the muon, and is then required to be compatible with the W mass, $30 < M_T(\mu, \not{E}_{\mathrm{T}}) < 125$ GeV. This effectively suppresses any residual QCD multijet background that passes the muon requirement.

6.8.2 The $\mu\mu$ + jets control region

The selection criteria for the $\mu\mu$ + jets control region are chosen to select events containing Z bosons that are produced in association with jets and decay to a pair of muons. They are selected to be kinematically similar to the main irreducible background in the signal region, $\mathrm{Z} \to \nu\bar{\nu}$. Exactly two muons are required that pass tight selection criteria with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.1$. They are also required to be separated from jets as in the μ + jets control region. Again, if any other leptons or photons are present the event is vetoed.

To maximise the selection of events containing Z bosons the invariant mass of the two muons is required to be compatible with the Z mass. Events are used if they are within a 25 GeV mass window, $66.2 < M_{\mu_1\mu_2} < 116.2$ GeV.

6.8.3 The γ + jets control region

The γ + jets control region is defined to provide a sample of photon events with significant hadronic energy. Events with high energy photons have similar kinematic properties to events with Z bosons, which allows this control region to augment the prediction of the $\mathrm{Z} \to \nu\bar{\nu}$ background. To ensure the mass of the Z boson has a negligible effect when making this assumption, a single photon is required with $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 1.45$. This selection also maintains a high trigger efficiency, the photon trigger rate is significantly greater than the muon rate so has a higher p_{T} threshold. To ensure the photon did not originate from a hadronic jet, events with $\Delta R(\text{jet}, \gamma) < 1.0$ are vetoed. Events containing any photons or lepton other than the single selected photon are vetoed. Photon samples are more easily contaminated with misidentification of QCD multijet events than the muon samples. To mitigate this and maintain a phase space that is close to the signal region, the α_{T} requirements, outlined in Tab. 6.5, are also made for the γ + jets control region. Also, due to the high p_{T}

threshold on the photon, this control region is only used in the regions of the analysis with $H_{\mathrm{T}} > 400\,\mathrm{GeV}$.

6.8.4 The hadronic control regions

While the above control regions are useful for estimating the SM backgrounds with genuine $\not{E}_{\mathrm{T}}$, some hadronic control regions are defined to quantify the residual QCD multijet background that remains in the signal region. To define these control regions, all the signal region requirements are made except for the requirement on $\Delta\phi^{*}_{\mathrm{min}}$ or $\not{H}_{\mathrm{T}}/\not{E}_{\mathrm{T}}$ being inverted. The use of these control regions is discussed in Sec. 7.4.

6.9 Event categorisation

To help to separate signal from background for a wide range of different signal hypotheses, events are categorised based on the number of jets, n_{jet}, their total hadronic energy, H_{T}, the magnitude of the missing hadronic energy, $\not{H}_{\mathrm{T}}$, and the number of jets tagged as a b-quark, n_{b}. These particular variables are chosen because many BSM models predict excesses in the tails of the distributions that are not predicted by the SM. For example, many SUSY models favour decays via top quarks, resulting in a significant number of b-tagged jets and high n_{jet} multiplicities. All BSM models targeted by this search also produce significant $\not{H}_{\mathrm{T}}$.

Along with straightforward binning in the categorisation variables, events are also split based on their jet topology. They are categorised as either *symmetric*, *asymmetric* or *monojet* depending on the p_{T} of the second leading jet. In the case that the second leading jet has $p_{\mathrm{T}} > 100$ GeV the event is defined as symmetric. This category of events targets the typical SUSY pair production scenarios when there is a significant mass splitting between the SUSY parent and the LSP. However, in the case that there is a compressed spectrum, a small mass splitting, there is unlikely to be significant energy in the SM decay products of the SUSY system. Sensitivity to these events can be gained by selecting events with significant ISR or FSR that have recoiled from the invisible SUSY system. This topology is targeted by the monojet and asymmetric jet category which require the second jet to have $p_{\mathrm{T}} < 40$ GeV or $40 < p_{\mathrm{T}} < 100$ GeV respectively.

A summary of the event categorisation for the n_{b}, n_{jet} and H_{T} variables is shown in Tab. 6.6. This is mirrored in the control and signal regions. The $\not{H}_{\mathrm{T}}$ dimension is treated slightly differently, with a variable binning based on H_{T}. This will be discussed further in Sec. 7.3.2.

Table 6.6: Summary of the n_{jet}, n_{b}, H_{T} binning.

Variable	Binning
n_{jet}	1 (mono-jet) , $2, 3, 4, 5 \geq 6$
n_{b}	$0, 1, 2, \geq 3$
Symmetric:	
H_{T} (GeV)	200–250, 250–300, 300–350, 350–400, 400–500, 500–600, 600–800, >800
Asymmetric:	
H_{T} (GeV)	200–250, 250–300, 300–350, 350–400, 400–500, 500–600, >600
Monojet:	
H_{T} (GeV)	200–250, 250–300, 300–350, 350–400, 400–500, 500–600, >600

Chapter 7

Background prediction

7.1 The dataset

The dataset used by this version of the analysis comprises 12.9 ± 0.8 fb^{-1} of $\sqrt{s} =$ 13 TeV proton-proton collisions produced by the LHC and recorded by CMS during the first few months of 2016. The data for this analysis are collected with the triggers described in Sec. 6.5.

7.2 Simulated event samples

A selection of simulated MC samples are used to help estimate the SM backgrounds and possible signal yields within the analysis. The samples are generated with the methods discussed in Sec. 4.8. The majority of samples, including the signal model samples, are simulated with MADGRAPH5 [93, 118] at LO accuracy: QCD multijet samples, Drell-Yan + jets, W + jets, Z + jets and γ + jets. Using the same generator at NLO accuracy, the s-channel single top, $t\bar{t}$W and $t\bar{t}$Z events are simulated. Additionally, the POWHEG [119, 120] generator is used at NLO accuracy to describe the t-channel and tW-channel production of single top quarks. Diboson (WW, WZ, ZZ) production is simulated using PYTHIA 8.2 [95]. This generator is also used to simulate the decays of the sparticles within the signal models. The cross-sections used for total event yield normalisation are calculated at NLO and next-to-next-to-leading order (NNLO) precision [93, 120–125]. The detector response is simulated using GEANT4 [97].

As the simulations cannot fully recreate all the effects within data, a series of corrections are made on the event yields predicted by the simulation. They are described in the rest of this section.

7.2.1 Pileup reweighting

When the MC samples are generated, the number of PU collisions are simulated at a chosen rate that is expected to be close to that obtained in data. As the simulation is produced before the data is collected and the true PU distribution varies with time, the actual distribution cannot be accurately simulated. To correct for this effect the MC events are reweighted to ensure their PU profile, characterised by the number of reconstructed vertices, matches that in data. The efficacy of these correction factors can be seen in Fig. 7.1.

7.2.2 Scale factors

There are often small differences between the efficiencies of finding physics objects that are simulated to those that are observed in data. The difference between these efficiencies is measured using a known data sample, such as $Z \to \mu\mu$ events, to measure

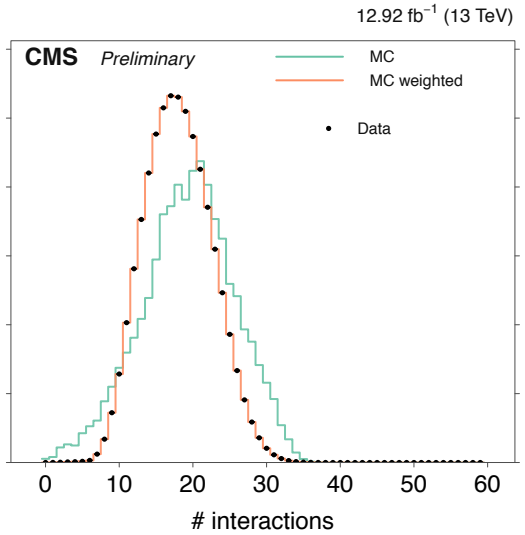


Figure 7.1: The distribution of the number of interactions per bunch crossing in MC simulation before and after PU reweighting, compared to the distribution measured in data.

muon efficiencies. A scale factor is then applied to the MC to reweight the simulated events in a way that means they more accurately simulate the efficiency of finding various physics objects. These scale factors are typically determined as a function of p_{T} and η, two quantities in which the detector response can vary significantly. These scale factors are used to correct lepton, photon and b-tagging efficiencies. As the simulation and data generally agree well, these scale factors are typically small, at most a few percent [126, 127].

7.2.3 Top p_{T} reweighting

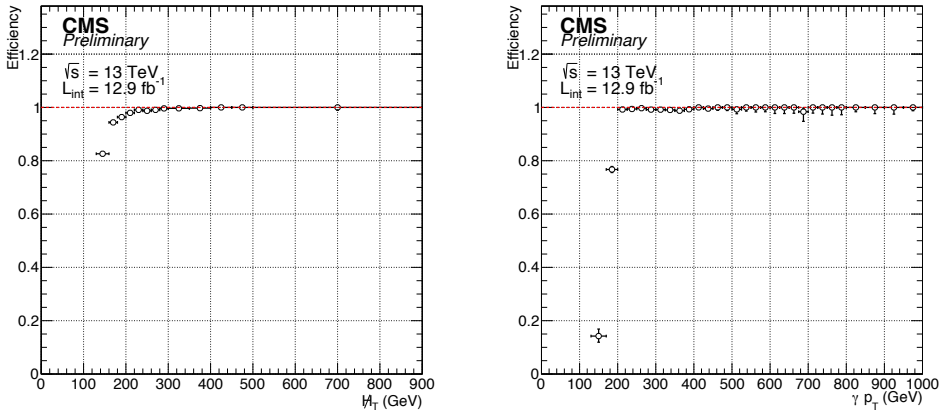
The $t\bar{t}$ simulation predicts a softer top-quark p_{T} spectrum than is observed in data [128]. This is corrected with a scale factor that is dependent on the p_{T} of the $t\bar{t}$ system. The scale factor was derived from the 8 TeV LHC result and is in the range 0-15%.

7.2.4 Trigger efficiencies

To account for inefficiencies in the trigger selection when collecting data, the MC sample event yields are corrected. For the muon control samples the trigger is simulated and the difference between the measured efficiency in data and simulation is corrected with a dedicated scale factor. For the γ + jets control sample the trigger efficiency is measured in a hadronic event sample and the MC is corrected as a function of the photon p_{T} and η. Within the signal region, the efficiencies are measured as a function of $\not{H}_{\mathrm{T}}$ using events that pass a reference electron trigger. As one of the analysis goals is to maintain low thresholds to H_{T} and $\not{H}_{\mathrm{T}}$ within the signal region, there are some inefficiencies at low values of H_{T} and $\not{H}_{\mathrm{T}}$. The trigger efficiency for a representative bin can be seen in Fig. 7.2. Additionally, the efficiency for one of the γ + jets control sample bins is also shown.

7.2.5 Cross-section corrections

The events considered within this analysis are selected to have a high H_{T} and high $\not{E}_{\mathrm{T}}$. The total number of events predicted by the MC simulation within this phase space does not agree with that observed in data. To correct this, normalisation correction factors are derived in *sidebands* after all the other corrections have been applied. This



(a) Signal region symmetric jet category $400 < H_{\mathrm{T}} < 600$ GeV

(b) γ + jets region over all H_{T}

Figure 7.2: Signal trigger efficiency in the $\not{H}_{\mathrm{T}}$ dimension measured with a muon sample (a) and γ + jets trigger efficiency measured with a hadronic sample (b).

is particularly important in the case of the γ + jets simulation, as the cross section for this process is only calculated to LO accuracy. For the other processes the cross section is calculated to at least NLO accuracy.

As will be described in Sec. 7.3, the backgrounds in the analysis are mostly predicted from data yields in the control regions. The simulation is not used to set the exact yields, however it is used to determine the admixture of the different processes in each of the analysis regions. It is for this reason that these corrections are important.

To derive the correction in the γ + jets control region, a sideband is defined with the α_{T} variable. The same selection is applied as for the γ + jets control region, other than the requirement that $0.50 < \alpha_{\mathrm{T}} < 0.52$. Within the μ + jets and $\mu\mu$ + jets control regions a sideband in $\not{H}_{\mathrm{T}}$ is used, requiring $100 < \not{H}_{\mathrm{T}} < 130$ GeV. A maximum likelihood fit across all the sidebands is performed, where the yields for the W + jets, γ + jets, $\mathrm{Z} \to \mu\mu$ and $t\bar{t}$ processes are modelled as Poisson distributions. Within the fit they are each allowed to be scaled by a free parameter. The fit then finds the best scale factor for each sample to produce the best data-simulation agreement within the sideband. These correction factors can then be applied to the simulation throughout the rest of the analysis. A summary of the correction factors can be seen in Table 7.1.

Table 7.1: Cross section corrections for SM processes determined from data sidebands.

SM process	Control sample	Data sideband	Correction
γ + jets	γ + jets	$0.50 < \alpha_T < 0.52$	1.33 ± 0.03
W + jets	μ + jets	$100 < \not{H}_T < 130$GeV	1.13 ± 0.01
Z + jets	$\mu\mu$ + jets	$100 < \not{H}_T < 130$GeV	0.99 ± 0.02
$t\bar{t}$	μ + jets, $\mu\mu$ + jets	$100 < \not{H}_T < 130$GeV	0.86 ± 0.01

The distributions of data and MC simulated events for key analysis variables in the major analysis regions are shown in Appendix A.1. All of these plots are made after all the corrections described in this section have been applied.

7.3 Background estimation for processes with genuine $\not{E}_T$

The accurate determination of the SM backgrounds is of utmost importance when searching for the indications of BSM phenomena. So as not to rely too heavily on the modelling of the backgrounds in simulation, a *data-driven* approach is utilised. As this analysis is carried out with the requirement of significant hadronic activity, mismodelling effects are particularly prevalent due to the difficulties in simulating strong-force interactions to a high degree of accuracy.

Given that the analysis selections are chosen to reduce the QCD multijet background to a negligible level, the significant SM backgrounds that must be predicted are those with genuine $\not{E}_T$ in their final state. The data-driven method used for predicting these backgrounds is described in this section.

7.3.1 Transfer factor method

The control samples, defined in Sec. 6.8, are chosen to provide a collection of events in data that are in a phase space that is close to that of the signal region. The control samples are split into (H_T, n_{jet}, n_b) bins in a way that is equivalent to the signal region, as introduced in Sec. 6.9. Each bin in a control region can then be extrapolated to the signal region through the use of *transfer factors (TFs)* that are derived from simulation. Each TF is defined as a ratio of the yields obtained from MC simulation for the same

bin of the signal region and a given control sample:

$$\mathrm{TF}(H_\mathrm{T}, n_\mathrm{jet}, n_\mathrm{b}) = \frac{N_\mathrm{MC}^\mathrm{signal}(H_\mathrm{T}, n_\mathrm{jet}, n_\mathrm{b})}{N_\mathrm{MC}^\mathrm{control}(H_\mathrm{T}, n_\mathrm{jet}, n_\mathrm{b})}, \tag{7.1}$$

where $N_\mathrm{MC}^\mathrm{signal}$ is the total simulated event yield for all background processes in the signal region and $N_\mathrm{MC}^\mathrm{control}$ is the equivalent event yield in a particular control region.

Making use of these TFs, predictions of background counts from SM processes, $N_\mathrm{pred}^\mathrm{signal}$, can then be made made based on the various control samples:

$$N_\mathrm{pred}^\mathrm{signal}(H_\mathrm{T}, n_\mathrm{jet}, n_\mathrm{b}) = \frac{N_\mathrm{MC}^\mathrm{signal}(H_\mathrm{T}, n_\mathrm{jet}, n_\mathrm{b})}{N_\mathrm{MC}^\mathrm{control}(H_\mathrm{T}, n_\mathrm{jet}, n_\mathrm{b})} \times N_\mathrm{obs}^\mathrm{control}(H_\mathrm{T}, n_\mathrm{jet}, n_\mathrm{b}), \tag{7.2}$$

where $N_\mathrm{obs}^\mathrm{control}$ is the number of events that are observed within the bin of a control region.

When constructing the TFs, the MC expectations for the following SM processes are considered: W + jets, $t\bar{t}$ + jets, $\mathrm{Z} \to \nu\bar{\nu}$ + jets, DY + jets, γ + jets, single top + jets production via the s, t, and tW-channels, $WW+$ jets, $WZ+$ jets, and $ZZ+$ jets, and $t\bar{t}\mathrm{V}$ or $t\bar{t}\mathrm{H}$.

The TFs account for differences in cross sections, acceptance, reconstruction efficiencies and kinematic requirements between the signal and control regions. One big advantage of using TFs is that any mismodelling effects that are consistent within a particular bin across relevant MC samples will cancel out. It is seen that the mismodelling of hadronic effects typically varies as a function of H_T and n_jet. An example of this mismodelling is visible in Fig. 7.3, the disagreement in the data and simulation changes as a function of the H_T or n_jet. The fact that the control samples are binned within these variables helps to negate this issue, which is cancelled out within the TF in each bin.

A systematic uncertainty is assigned to each TF to account for theoretical uncertainties and other mismodelling effects that do not cancel in the ratio. Details of the specific systematic uncertainties are given in Sec. 7.5.

Within the analysis, the TF method is used with different control regions to predict two different categories of background, the $\mathrm{Z} \to \nu\bar{\nu}$ and all the other remaining backgrounds. The $\mathrm{Z} \to \nu\bar{\nu}$ background is predicted with the $\mu\mu$ + jets, γ + jets and μ

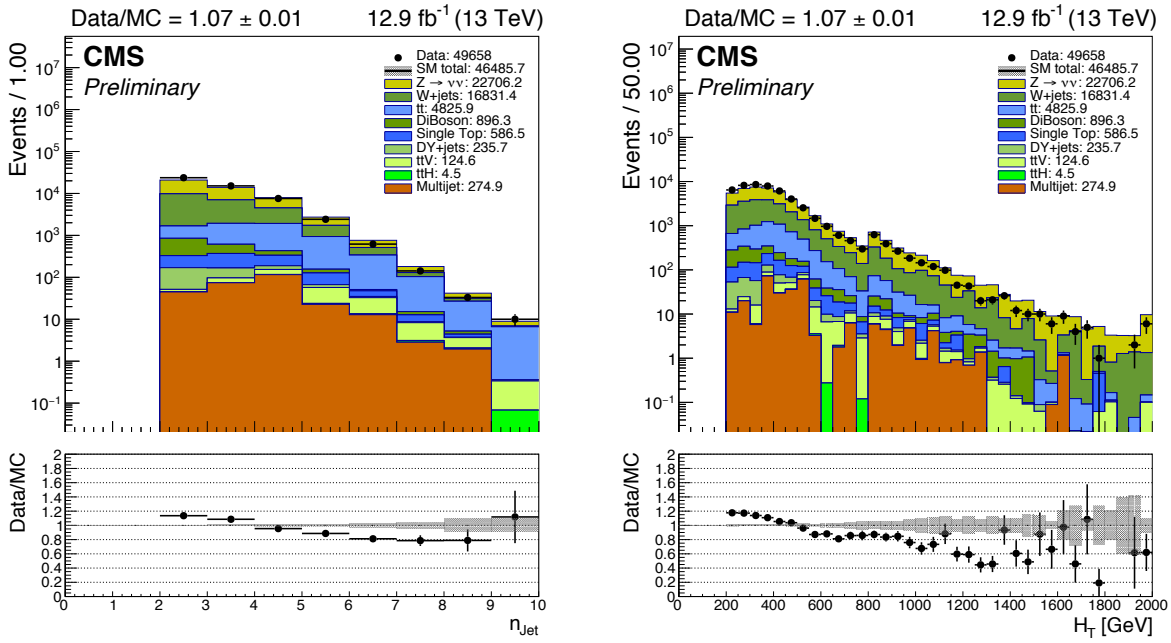


Figure 7.3: The H_T and n_{jet} distributions in all categories of the signal region that have a symmetric jet topology

+ jets control samples. The W, $t\bar{t}$ and other residual backgrounds are predicted with the μ + jets control sample. This method is ultimately implemented within a fitting procedure that is defined formally by the likelihood model described in Sec. 7.6. This procedure allows the appropriate systematic uncertainties to be included in a way that takes account of their correlations across all samples and bins.

To summarise the fitting procedure, the observation in each bin of the signal region is modelled with a Poisson distribution that has a mean of the sum of SM expectations (and a potential signal contribution). The components of this SM expectation are related to the expected yields in the control samples via the TFs. The observations in each bin of the control samples are similarly modelled as Poisson-distributed about the expected yields for each control sample. In this way, for a given bin, the observed yields in the signal and control samples are connected via the transfer factors derived from simulation. This allows multiple control samples to modify the predicted yields in the signal region within the constraints of their uncertainties.

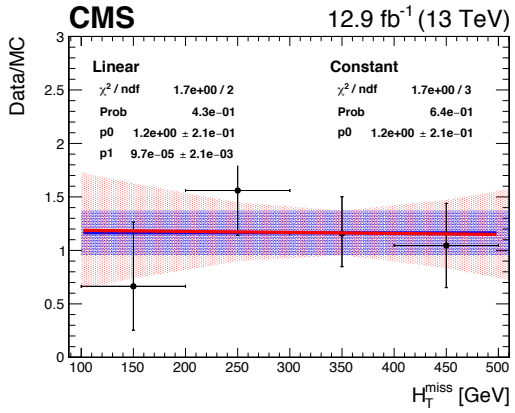
7.3.2 The $\not{H}_{\mathrm{T}}$ dimension

The TF method is used to estimate the total background counts in each of the (H_{T},n_{jet},n_{b}) bins. However, to maximise the sensitivity to BSM signatures with large missing momentum, events in the analysis are also categorised based on their $\not{H}_{\mathrm{T}}$. Due to limitations on the number of events available in each control sample, it is disadvantageous to add another binning dimension. This would result in statistical uncertainties of the control region event counts dominating the background prediction. Instead, the shape of the $\not{H}_{\mathrm{T}}$ distribution is taken directly from simulation for each of the (H_{T},n_{jet},n_{b}) bins in the signal region. This is justified by the fact that binning in these variables helps to isolate hadronic mismodelling effects within each bin, which then cancel in the TFs.

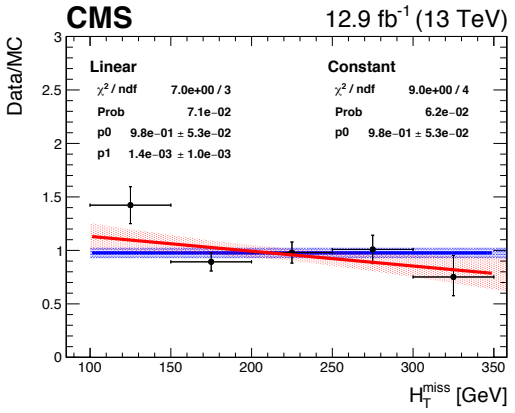
To make sure that this approach is valid, the assumption that the $\not{H}_{\mathrm{T}}$ dimension is well modelled in each bin must be tested. This testing is carried out within the γ + jets, μ + jets and $\mu\mu$ + jets control regions. Within each analysis bin of each of the control regions, the shape of the $\not{H}_{\mathrm{T}}$ distribution in data is compared to the distribution in MC simulation. The normalisation can be disregarded as it is set through the TF method. Provided the modelling is reasonable, the ratio of the normalised shapes is expected to have a value of one consistently across the $\not{H}_{\mathrm{T}}$ dimension, in each of the (H_{T},n_{jet},n_{b}) bins. Any difference is used to derive a systematic uncertainty on the $\not{H}_{\mathrm{T}}$ shape as described in Sec. 7.5.3.

An example of the data/MC ratio as a function of the $\not{H}_{\mathrm{T}}$ for two representative bins is shown in Fig. 7.4. The result of a constant and linear fit and their respective uncertainties are drawn on the plots in blue and red respectively. To ensure that a flat hypothesis of the ratio is valid across all bins, the significance of the deviation from zero of the linear parameters of the data/MC fits is shown in Fig. 7.5. As they are compatible with no linear dependence, the use of simulation to determine the $\not{H}_{\mathrm{T}}$ shape in each (H_{T},n_{jet},n_{b}) category is considered valid.

The sizes of the $\not{H}_{\mathrm{T}}$ bins within the signal region are chosen to contain enough data counts within each of the control regions to perform the validation. Taking $\not{H}_{\mathrm{T}}$ bin widths of 50 GeV is found to be enough to satisfy this condition, while not being limited by the measured $\not{H}_{\mathrm{T}}$ resolution. In most cases the $\not{H}_{\mathrm{T}}$ bins are naturally bounded from above by the maximum value of H_{T} in a particular bin. However, the highest H_{T} bin is unbounded from above, the final $\not{H}_{\mathrm{T}}$ bin is therefore limited to sit at $\not{H}_{\mathrm{T}} > 800$ GeV ($\not{H}_{\mathrm{T}} > 600$ GeV) for symmetric (asymmetric) topologies to ensure the requisite statistics. All of the bins used are listed in Appendix A.2.



(a) γ + jets, $n_{\mathrm{b}} = 1$, $n_{\mathrm{jet}} = 2$ symmetric category and H_{T} 400-500GeVbin



(b) $\mu\mu$ + jets, $n_{\mathrm{b}} = 0$, $n_{\mathrm{jet}} = 3$ asymmetric category and H_{T} 300-350GeVbin

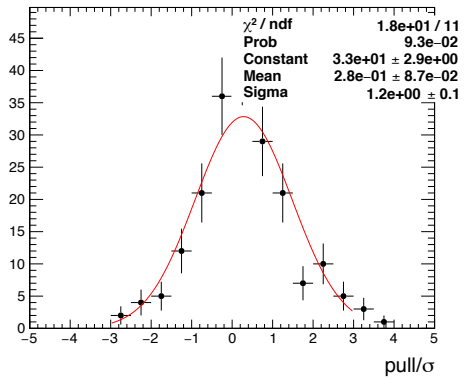
Figure 7.4: The data/MC distribution against $\not{H}_{\mathrm{T}}$ (denoted H_T^{miss}) for two representative categories in the γ + jets (a) and $M_{\mathrm{inv}}^{j1,j2}$(b) control regions. Linear and constant fits are made to the ratio and their p-value and fit parameters are shown on the plot.

7.4 Background estimation for QCD multijet processes

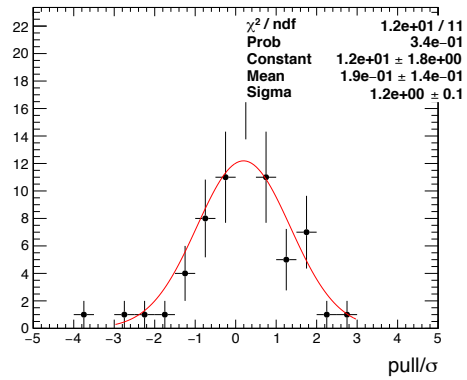
As discussed extensively in Chapter 6, the strategy of the analysis is based around reducing the QCD multijet background to a negligible level. However, it is necessary to verify that the multijet contribution within the signal region is indeed small. This is confirmed with the method described in this section. To remain sensitive to small contributions in the signal region from BSM phenomena, it is imperative to keep the uncertainties on the level of QCD contamination small. The method therefore utilises data-driven techniques, similar to those used for the estimation of processes with genuine $\not{E}_{\mathrm{T}}$ described in Sec. 7.3.1.

7.4.1 QCD-enriched sidebands

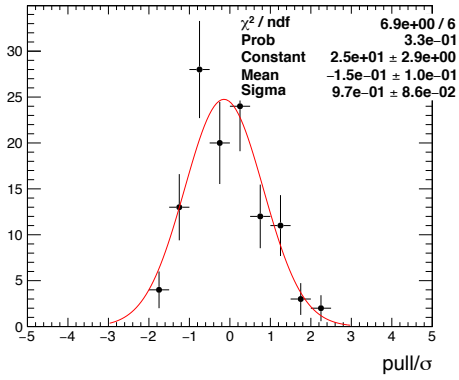
To be able to carry out a data-driven estimate, three data sidebands to the signal region are defined. They are chosen to be heavily QCD contaminated, but still close in phase space to the signal region. To achieve this, the signal region selection is applied with the requirements on two of the variables used to remove the multijet background inverted. They are defined by the requirement $\not{H}_{\mathrm{T}}/\not{E}_{\mathrm{T}} > 1.25$, $\Delta\phi^*_{\mathrm{min}} < 0.5$ and both these requirements, as illustrated in Table 7.2.



(a) μ + jets



(b) γ + jets



(c) $\mu\mu$ + jets

Figure 7.5: The distribution of the significance of the deviation from zero of the linear parameters of the data/MC fits. The pull is defined as the value of the parameter divided by its error. As all these pulls for each of the control regions are statistically compatible with zero, a linear hypothesis for the $\not{H}_T$ data/MC ratio is valid.

For the core QCD multijet background prediction, the $\not{H}_T/\not{E}_T$ sideband is used to estimate the QCD yields in the signal region through an extrapolation with a TF. To then carry out a validation of the transfer factor, the $\Delta\phi^*_{\text{min}}$ and double sidebands are used.

Table 7.2: Definition of sidebands used in the determination of the QCD background contributions in the signal region.

	$\Delta\phi^*_{\text{min}} < 0.5$	$\Delta\phi^*_{\text{min}} > 0.5$
$\not{H}_T/\not{E}_T > 1.25$	**A** Double sideband	**B** $\not{H}_T/\not{E}_T$ sideband
$\not{H}_T/\not{E}_T < 1.25$	**C** $\Delta\phi^*_{\text{min}}$ sideband	**D** Signal region

7.4.2 The method

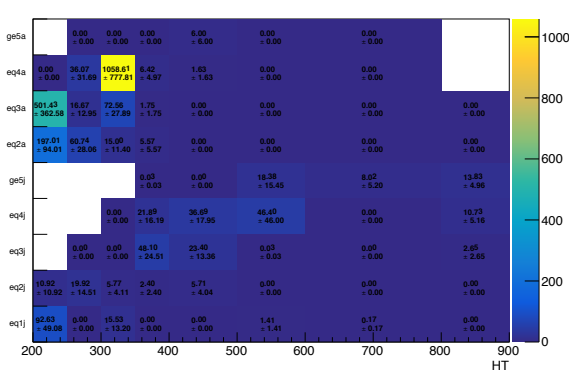
The method makes use of the ratio of simulated QCD counts in the signal region to the $\not{H}_T/\not{E}_T$ sideband, R, per (H_T,n_{jet}) bin. This is essentially a TF constructed from the QCD simulation. Each ratio is used as a multiplier on the predicted QCD counts per bin in the $\not{H}_T/\not{E}_T$ data sideband. These data counts are collected with the signal region triggers described in Sec. 6.5.

The prediction of the number of QCD multijet events in the sideband, $\mathcal{Q}$, is carried out with a maximum likelihood fit analagous to that described in Sec. 7.6, but in the $\not{H}_T/\not{E}_T$ sideband and with the total number of QCD events free to be determined. In this fit, the electroweak backgrounds within the sideband are determined with the method described in Sec. 7.3 with μ + jets, $\mu\mu$ + jets and γ + jets control regions that have the same selection as the control regions described in Sec. 6.8, apart from an inverted $\not{H}_T/\not{E}_T$ cut. All relevant systematic uncertainties on the TFs, described in Sec. 7.5 are taken into account. After the contribution of the electroweak backgrounds is estimated, the remaining data counts are attributed to QCD. In this prediction, all counts and predictions are inclusive in n_b and $\not{H}_T$. The product of these predicted QCD counts and the ratio, $\mathcal{Q} \times R$, provides an estimate of the level of QCD multijet events in each (H_T,n_{jet}) bin of the signal region.

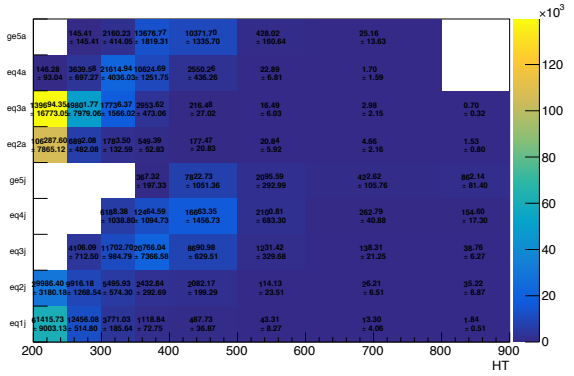
The number of counts from simulated QCD multijet events satisfying and failing the requirement $\not{H}_T/\not{E}_T < 1.25$, R, are summarised in Fig. 7.6a, 7.6b, and 7.6c. Figure 7.6d shows the expected counts from non-multijet backgrounds in the $\not{H}_T/\not{E}_T$ sideband, predicted with the TF method.

Figure 7.7a shows the observed counts in the $\not{H}_T/\not{E}_T$ sideband, and Fig. 7.7b shows the QCD counts in the sideband predicted by the maximum likelihood fit. Figure 7.7c shows the predicted counts for the multijet contribution in the signal region bins, which are obtained from the products of R and $\mathcal{Q}$, summarised in Figs. 7.6c and 7.7b. Finally, Fig. 7.7d shows the ratios of predicted multijet counts with respect to the expected counts from the other backgrounds with genuine $\not{E}_T$ in the signal region, predicted with the procedure defined in Sec. 7.3.

The predictions summarised in Fig. 7.7d show that the H_T-dependent α_T thresholds defined in Table 6.5 and the requirement of $\Delta\phi^*_{min} > 0.5$ suppress the QCD multijet contamination in all bins of the signal region to the percent-level or smaller with

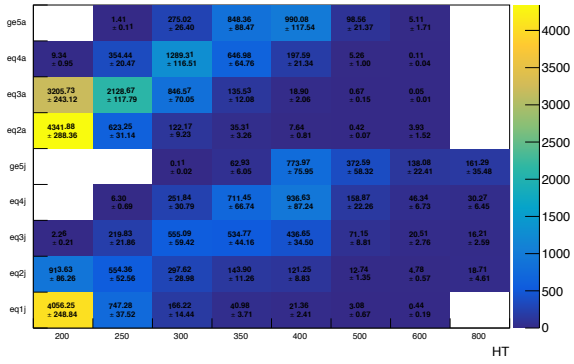


(a) Simulated QCD events in signal region.



(b) Simulated QCD events in $\not{H}_T/\not{E}_T$ sideband.

(c) Ratio R for simulated QCD events.

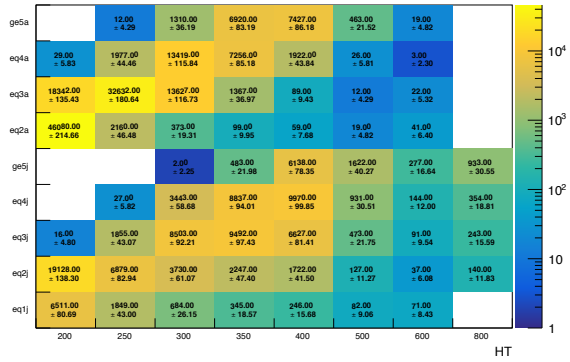


(d) Predicted electroweak backgrounds in the $\not{H}_T/\not{E}_T$ sideband.

Figure 7.6: Expected number of QCD multijet events determined from simulation, binned according to n_{jet} and H_T, that (a) satisfy and (b) fail the requirement $\not{H}_T/\not{E}_T < 1.25$. The bins are labelled as described in App. A.3. Also shown in (c) is the ratio R for QCD multijets, again determined from simulation. Finally, (d) shows the expected number of non-multijet events (V+jets and $t\bar{t}$, plus other residual non-multijet backgrounds) that fail the $\not{H}_T/\not{E}_T < 1.25$ requirement, predicted using the TF method and binned according to n_{jet} and H_T.

respect to the total non-multijet background. These predicted multijet events are included as a background contribution to the likelihood model described in Sec 7.6.

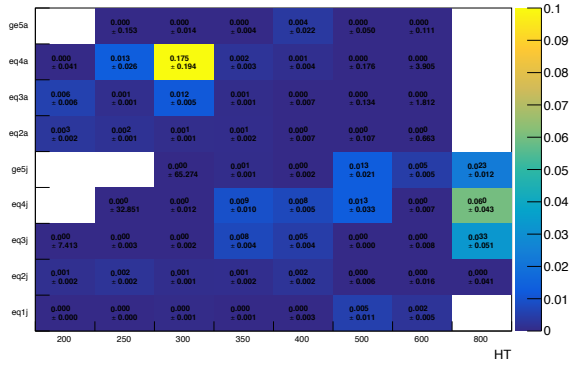
As the prediction of QCD in the signal region is carried out inclusively over n_b and $\not{H}_T$, the QCD shapes for these variables are taken from the non-multijet simulation and normalised to the QCD counts. A lack of statistics in the QCD simulation led to the adoption of this approach. Within uncertainties, the level of agreement is deemed acceptable given the small total QCD contribution to the signal region.



(a) Binned data counts in $\not{H}_T/\not{E}_T$ sideband.

(b) Predicted QCD counts in $\not{H}_T/\not{E}_T$ sideband, $\mathcal{Q}$.

(c) QCD multijet predictions in the signal region.



(d) Ratio of predicted multijet and non-multijet yields in the signal region.

Figure 7.7: The number of events observed in the $\not{H}_T/\not{E}_T > 1.25$ sideband, binned according to n_{jet} and H_T are shown in (a). The bins are labelled as described in App. A.3. In (b) these yields are corrected by subtracting the expected electroweak component. Shown in (c) is the result of multiplying the observed multijet events predicted in (b) by the translation factor from the sideband to the signal region determined with simulation (shown in Fig. 7.6). This gives a data driven expectation of the quantity of multijet background events in the signal region. Finally, (d), shows the ratio of expected multijet background events in the signal region divided by non-multijet backgrounds. The multijet background is therefore shown to be below the percent level.

7.4.3 Validation

Despite using a predominantly data-driven method, the prediction of the QCD contamination relies on the ratio of QCD counts, R, that is derived with simulation. This ratio is validated with data in a QCD enriched sideband, where the full signal region selection is used other than an inversion of the $\Delta\phi^*_{\text{min}}$ cut to $\Delta\phi^*_{\text{min}} < 0.5$. In this

sideband a data driven estimation of the QCD counts is carried out in two regions, the $\Delta\phi^*_{\text{min}}$ sideband with $\not{H}_{\text{T}}/\not{E}_{\text{T}}$ values less than 1.25 and the double sideband with values greater than 1.25. These predictions are again made with a maximum likelihood fit, analogous to that described in Sec. 7.4.2. This fit estimates the counts in the non-multijet backgrounds with μ + jets, $\mu\mu$ + jets and γ + jets control regions and all relevant systematic errors. The remaining data counts in each sideband are then attributed to QCD. With this estimation of QCD it is possible to derive a data driven ratio of QCD counts with $\not{H}_{\text{T}}/\not{E}_{\text{T}} < 1.25$ and those with $\not{H}_{\text{T}}/\not{E}_{\text{T}} > 1.25$, $R^{data}_{\Delta\phi^*_{\text{min}}<0.5}$. By taking MC counts in the $\Delta\phi^*_{\text{min}}$ sideband it is also possible to calculate the $\not{H}_{\text{T}}/\not{E}_{\text{T}}$ simulation ratio, $R_{\Delta\phi^*_{\text{min}}<0.5}$.

To validate the ratio R, it is assumed that if the simulation of the ratio in the $\Delta\phi^*_{\text{min}}$ sideband agrees with that derived from data, the simulated ratio that is not in the sideband is valid. Any disagreement is covered by a systematic error on the signal region QCD prediction. The ratio of $R_{\Delta\phi^*_{\text{min}}<0.5}$ and $R^{data}_{\Delta\phi^*_{\text{min}}<0.5}$ in H_{T} and n_{jet} bins is shown in Fig. 7.8. Bins in which there are insufficient statistics in data or simulation to make the calculation are left out. This plot illustrates that a fully correlated systematic of 100% taken on the predicted QCD contamination in the signal region should cover any disagreement between simulation and data.

7.5 Systematic uncertainties

After defining data driven methods for estimating the SM backgrounds in the analysis, it is necessary to take account of all sources of systematic uncertainties on these background predictions. In this section the sources of systematic uncertainty in the analysis are outlined with the different methods used for estimating them.

Due to the exceptional treatment of the $\not{H}_{\text{T}}$ dimension when carrying out background estimation, the systematics are split into two types. There are those that affect the total number of events in each (H_{T},n_{jet},n_{b}) bin (integrating over $\not{H}_{\text{T}}$), which are described in Sections 7.5.1 and 7.5.2. Most of these systematics are treated as uncertainties on the TFs used to predict the SM backgrounds with genuine $\not{E}_{\text{T}}$. There is also one other systematic uncertainty added to take account of uncertainties in the QCD multijet background prediction, described in Sec. 7.4.3. Additionally, systematic uncer-

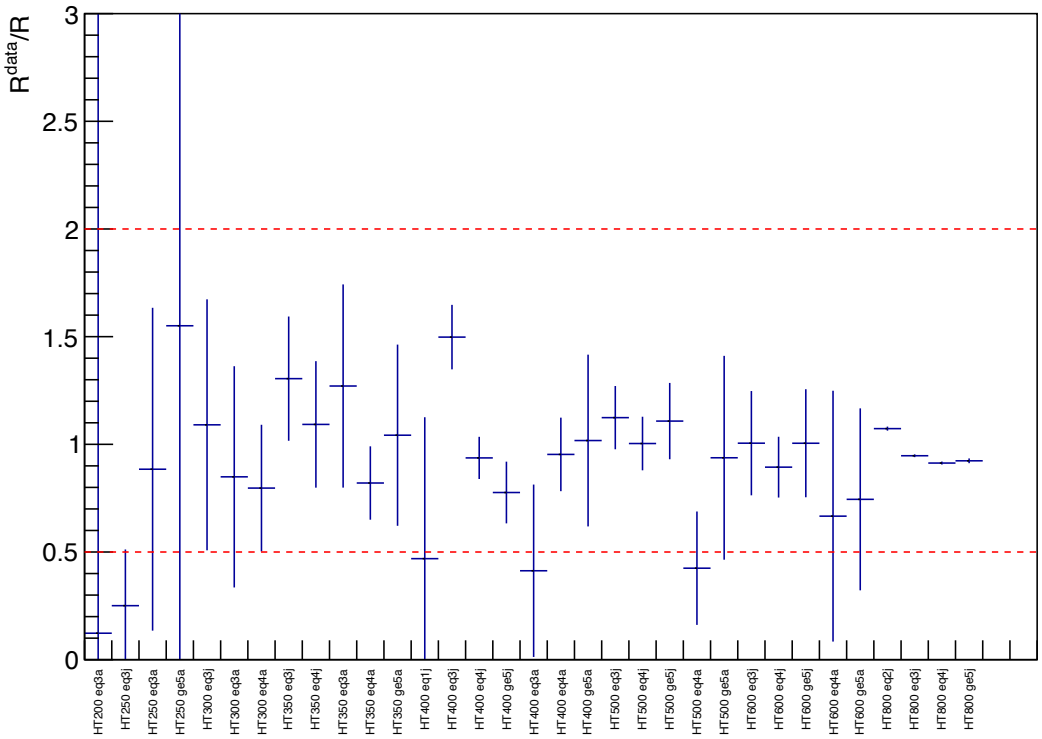


Figure 7.8: Ratio of the measurement of R, the pass/fail ratio for the $\not{H}_T/\not{E}_T$ selection, from data and Monte Carlo in the $\Delta\phi^*_{\text{min}} < 0.5$ sideband in (H_T, n_{jet}) bins. Dotted red lines demonstrate that disagreement is covered by a 100% systematic uncertainty on the ratio. The bins are labelled as described in App. A.3.

tainties are included that encode the limited knowledge on how the events distribute in the $\not{H}_T$ dimension, described in Sec. 7.5.3.

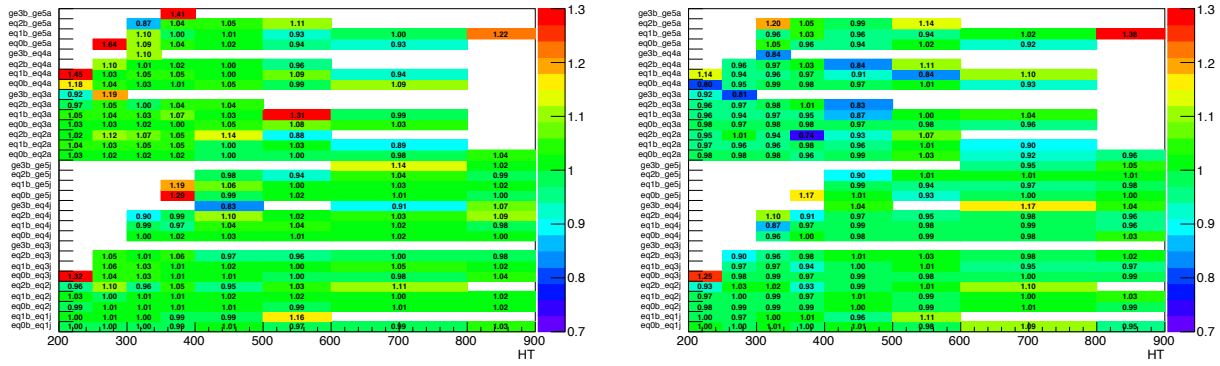
There are two approaches that are used to derive uncertainties from different sources. There are uncertainties associated with the correction factors that are applied to the simulation, which allows the systematic uncertainties to be derived by varying the simulation (Sec. 7.2.2). However, these sort of systematics only encode known and simulated sources of systematic uncertainty. To be able to account of unknown sources, additional data-driven uncertainties are derived with the use of the control samples (Sec. 7.5.2). After their descriptions below, a summary of all the uncertainties is given in Tab. 7.3.

7.5.1 Uncertainties derived from simulation

A set of corrections are applied to simulation that are described in Sec. 7.2. There is an uncertainty associated with each of these corrections, based on the way in which they are derived and uncertainties in any theoretical calculations that may be relevant. These uncertainties are propagated to each of the TFs which are of interest for the background prediction, namely: $\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$, $\mu\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$, $\gamma + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ and $\mu + jets \rightarrow t\bar{t} + \mathrm{W}$.

Jet energy scale

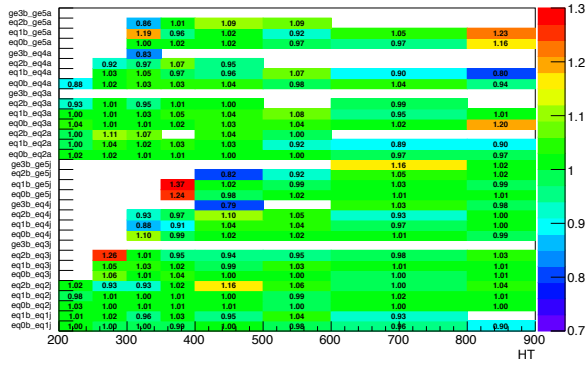
The effect of varying the jet energy corrections (JECs) by their uncertainty in the μ + jets and $\mu\mu$ + jets control regions on the TFs is investigated. As the H_T and jet multiplicity binning is mirrored in signal and control regions, the effect of jet energy scale on the TFs is expected to be small. However, the jet energy scale can still have an effect due to jets moving in and out of acceptance (above and below 40 GeV). The relative change in the TFs is presented as a function of H_T and jet category in Fig. 7.9-7.12. These plots show the change in TFs for the μ + jets, $\mu\mu$ + jets and γ + jets control samples as they are used to predict the $\mathrm{Z} \rightarrow \nu\bar{\nu}$ and the W+$t\bar{t}$ backgrounds. The plots for the other sources of systematics can be found in the Appendix A.4. The changes are typically in the range of 1-15%.



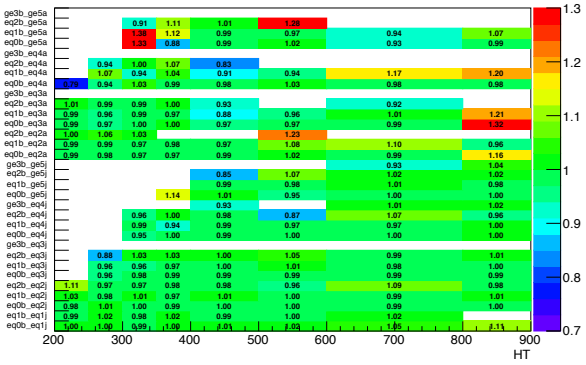
(a) JEC up variation

(b) JEC down variation

Figure 7.9: The relative change in the $\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ transfer factors when varying JEC in MC within its uncertainties, as a function of H_T (GeV) and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure. The bins are labelled as described in App. A.3.

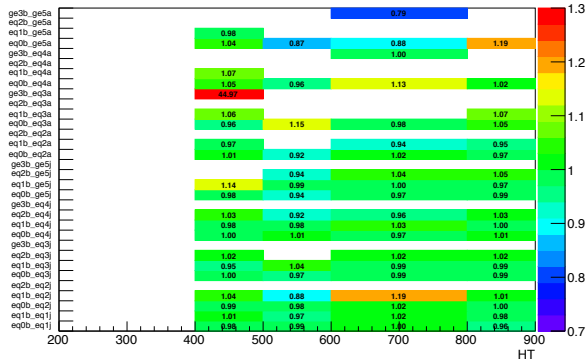


(a) JEC up variation

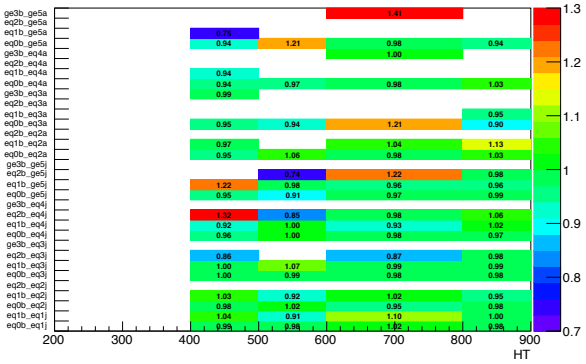


(b) JEC down variation

Figure 7.10: The relative change in the $\mu\mu + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying JEC in MC within its uncertainties, as a function of H_T (GeV) and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure. The bins are labelled as described in App. A.3.



(a) JEC up variation

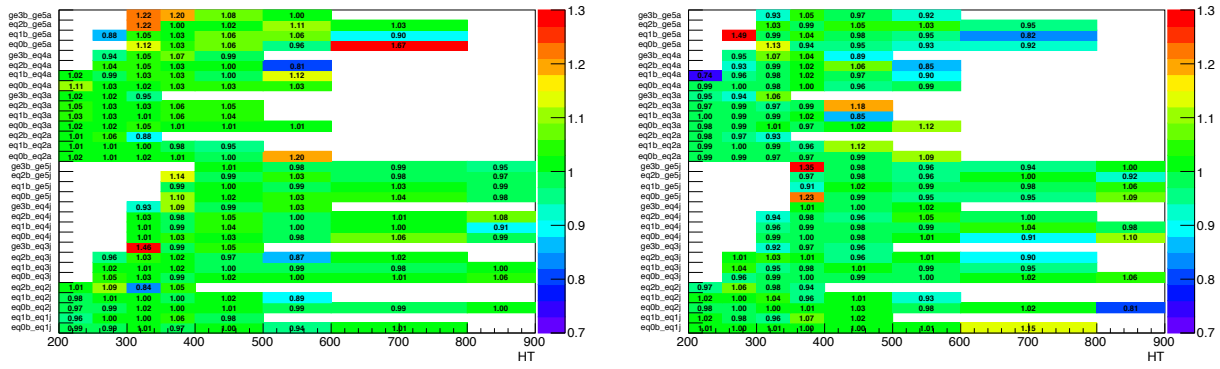


(b) JEC down variation

Figure 7.11: The relative change in the $\gamma + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying JEC in MC within its uncertainties, as a function of H_T (GeV) and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure. The bins are labelled as described in App. A.3.

B-tagging efficiency

The scale factors that take account of the b-tagging efficiencies and misidentification between simulation and data also have associated uncertainties. Since no extrapolation is performed in the background prediction across different n_b multiplicities, these uncertainties are expected to only have a small effect. The scale factors associated with the identification of b and c jets are varied together (since their measurements are



(a) JEC up variation **(b)** JEC down variation

Figure 7.12: The relative change in the $\mu + jets \rightarrow t\bar{t}$+W transfer factors when varying JEC in MC within its uncertainties, as a function of H_{T} (GeV) and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure. The bins are labelled as described in App. A.3.

correlated), while those associated with the misidentification of light jets are varied separately. The relative change in the TFs is presented as a function of H_{T} and jet category in Fig. A.13-A.20. They are typically in the range of 1-5%.

Lepton and photon efficiencies

Events from W and $t\bar{t}$ processes can enter the signal region when one of the leptons is not identified, by falling out of acceptance or not being properly reconstructed. There is an uncertainty associated with these acceptance and reconstruction effects that must be taken account of.

The uncertainties on the trigger, lepton identification and isolation efficiencies are varied by $\pm 1\sigma$ and their effects on the yield of W and $t\bar{t}$ event vetoes is studied. The procedure is repeated separately for muons and electrons. There is found to be at most a 2% difference in the W and $t\bar{t}$ yields in the signal region when these uncertainties are varied. This is propagated as an uncertainty to the TFs and taken as correlated across all the signal region bins.

Along with the considerations of the uncertainties on the simulated acceptance of leptons, uncertainties in the scale factors can directly effect yields in the muon control regions. The change in the TFs with the upwards and downwards variations of the

scale factors of the muons used to define the single and double muon control regions can be seen in Figs. A.21 to A.23. They are typically in the range 0-3%.

Photon trigger uncertainty

To take account of uncertainties in the photon trigger efficiencies, the effect of varying them on the γ + jets TF is investigated. To make a conservative estimate of the systematic, the uncertainty on this correction is taken as the size of the inefficiency. The relative change in the γ + jets transfer factor is presented in Fig. A.24, the variation is typically in the range 0-3%.

Signal trigger uncertainty

The uncertainties on the signal trigger efficiency measurements must also be taken into account. This is most relevant in the low H_{T} and $\not{H}_{\mathrm{T}}$ regions of the analysis, where the triggers are not fully efficient. As it is possible to measure the efficiency with both muon and electron reference triggers, a systematic is taken as the difference in the measurement between the two. The relative change in transfer factors is presented in Fig. A.25-A.28. The variation is typically in the range 0-5%.

Top p_T reweighting

To account for the uncertainty on the top p_{T} reweighting, an uncertainty on this correction is taken as the difference of the correction from unity. The relative change in transfer factors is presented in Fig. A.29-A.31. The variation is typically in the range 0-15%.

QCD contamination in the γ + jets control sample

Due to the greater prevalence of QCD multijet background in the γ + jets control sample, which is expected to be at the $\sim$5% level, an uncertainty is applied on this contamination. An arbitrarily large variation of $\pm$100% on the number of simulated QCD events is chosen and leads to a systematic variation on the TFs of at most 5% in the majority of bins. This value is small enough to remain subdominant to other

uncertainties and is found to be covered in the data-driven study using the photon control region, described in Sec. 7.5.2.

PU reweighting

There are uncertainties in the minimum bias cross section which is used to calculate the scale factors for PU reweighting. This 5% uncertainty is propagated and the relative change in the transfer factors under this variation is small (1-5%) and shown in each analysis bin in Fig. A.32-A.35.

7.5.2 Uncertainties derived from data-driven tests

To be able to take account of sources of systematic uncertainty that are not included in the variation of known scale factors, an additional procedure to derive uncertainties through a data driven method is defined. This allows assumptions and extrapolations within the analysis to be covered with an uncertainty that does not rely on potential limitations in the simulation modelling. To carry out this data-driven uncertainty estimation process, a suite of *closure tests* are performed. In each test the number of events in a given data control (sub-)sample is predicted using events from another data control (sub-)sample through a corresponding TF. The tests are defined in a way that tests a particular assumption made within the analysis. The agreement between the predicted and observed yields is expressed as the ratio $(N_{\text{obs}} - N_{\text{pred}})/N_{\text{pred}}$ while considering only the statistical uncertainties on N_{pred} and N_{obs}, where N_{pred} is the number of predicted events in the (sub-)sample and N_{obs} the number observed. This allows one to define a level of *closure*, which encapsulates the statistical significance of a deviation in the ratio from zero. The significance with which this closure deviates from zero can then be used to approximate the systematic uncertainty associated to the assumption the procedure is testing within the analysis.

These closure tests are performed separately for each n_{jet} category, as a function of H_{T}. The systematic uncertainty in each H_{T} bin is derived by summing in quadrature the ratio $(N_{\text{obs}} - N_{\text{pred}})/N_{\text{pred}}$ with its statistical error, after merging the n_{jet} categories into their symmetric and asymmetric topologies. Pairs of H_{T} bins are merged when the $\mu\mu$+jets sample is used, in order to increase the statistical power of the sample. Since the uncertainties derived with this approach are statistical in nature, these systematics are considered uncorrelated in each H_{T} bin and event topology category.

Extrapolation in α_T and $\Delta\phi^*_{min}$

There is no $\Delta\phi^*_{min}$ cut made in any of the control regions and no α_T cut in the muon control regions. As this is the case, it is implicitly assumed that the simulation correctly models the distribution of both of these variables when the TFs are constructed. To provide an estimate of the systematic uncertainty related to this assumption, dedicated closure tests are carried out. The tests are carried out in the μ + jets control region, using data yields with an α_T or $\Delta\phi^*_{min}$ requirement to predict the data yield of events with this requirement inverted. Explicitly for the α_T closure tests, the number of predicted events, $N_{pred}^{\alpha_T>x}$, is:

$$N_{pred}^{\alpha_T>x} = \frac{N_{MC}^{\alpha_T>x}}{N_{MC}^{\alpha_T<x}} \times N_{obs}^{\alpha_T<x} \tag{7.3}$$

where $N_{obs}^{\alpha_T<x}$ is the number of events that fail the α_T requirement x, $N_{MC}^{\alpha_T<x}$ is the number of simulated events that fail the requirement and $N_{MC}^{\alpha_T>x}$ is the number of events that pass the requirement. The closure is then defined as:

$$(N_{obs}^{\alpha_T>x} - N_{pred}^{\alpha_T>x})/N_{pred}^{\alpha_T>x}, \tag{7.4}$$

where $N_{obs}^{\alpha_T>x}$ is the number of observed events that pass the α_T requirement.

The results of the tests are shown in Fig. 7.13 as a function of H_T and n_{jet}. The grey band is the systematic uncertainty propagated through the analysis, taken as uncorrelated per each H_T bin and jet topology (symmetric/asymmetric). The systematic derived from these tests is in the range $4 - 32\%$.

The $\Delta\phi^*_{min}$ and α_T closure tests are testing the same kind of extrapolation, in a topological $\not{E}_T$ based variable. As this is the case the contribution to the systematic error is taken from only the α_T closure tests for bins with $H_T < 800\,\text{GeV}$ and from the $\Delta\phi^*_{min}$ tests for bins with $H_T > 800\,\text{GeV}$.

Modelling of the W/Z ratio

To validate the use of $W(\rightarrow \mu\nu)$ + jets and $t\bar{t}$ dominated μ + jets events to predict the $Z \rightarrow \nu\bar{\nu}$ background, tests are performed in data using single-muon and double-muon control regions. The events in the μ + jets control region are used to predict events in

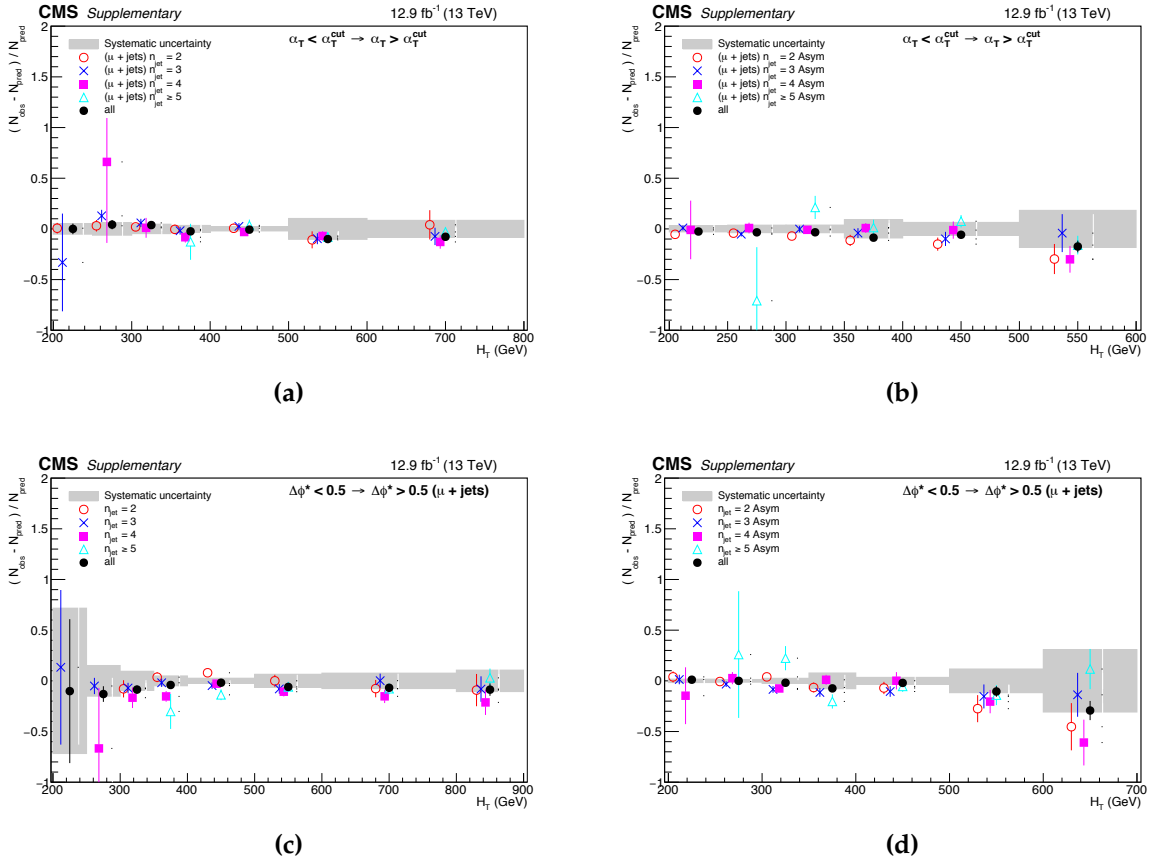
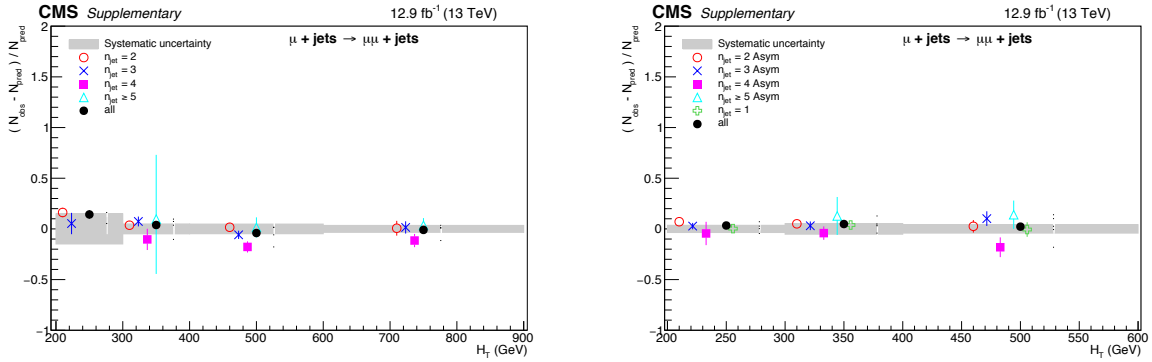


Figure 7.13: Data-driven tests probing the α_T (top row) and $\Delta\phi^*_{\text{min}}$ (bottom row) extrapolation for each n_{jet} category (open symbols) overlaid on top of the systematic uncertainty estimates used for each of the seven H_T bins (shaded bands). The symmetric (asymmetric) jet topologies are shown in the left (right) plot.

the $\mu\mu$ + jets control region, using transfer factors from simulation. These tests target the modelling of the W/Z ratio in simulation and also indirectly test muon acceptance effects, which are expected to be sub-dominant and whose uncertainties are already addressed elsewhere.

The results are shown in Fig. 7.14 as a function of H_T and n_{jet}. The grey band is the systematic uncertainty propagated through the analysis, taken as un-correlated per each H_T bin and jet topology (symmetric/asymmetric). The systematic derived from these tests is in the range $3-20\%$.



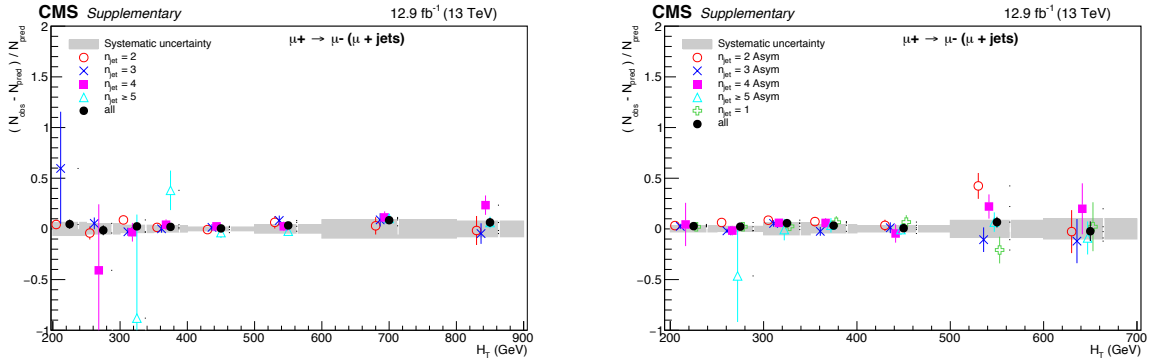
(a) **(b)**

Figure 7.14: Data-driven tests probing the use of the μ + jets control sample to predict the $Z \to \nu\bar{\nu}$ background for each n_{jet} category (open symbols) overlaid on top of the systematic uncertainty estimates used for each of the seven H_{T} bins (shaded bands). The symmetric (asymmetric) jet topologies are shown in the left (right) plot.

Modelling of the W/Z acceptance due to polarisation effects

As the kinematics of W^+ and W^- decays are subtly different, a data-driven test is introduced to check the modelling of this effect in simulation. In this study, carried out with events in the μ + jets control region, yields of μ^+ events are used to predict the yields of μ^- events. The production mechanism of W bosons from pp-collisions means high p_T W-bosons are predominantly left handed [129]. For high p_T bosons, this implies that W^+ decays to the left handed neutrino along its direction of motion while the lepton is pointing backward. The opposite behaviour is expected for the W^-. The lepton is therefore more boosted (and the neutrino less boosted) in W^+ decays than W^- decays. This leads to a larger number of W^+ decays in the single lepton control regions (which relies on the lepton p_T for acceptance) than in the signal region (which relies on the neutrino p_T for acceptance).

The results are shown in Fig. 7.15 as a function of H_{T} and n_{jet}. The grey band is the systematic uncertainty propagated through the analysis, taken as un-correlated per each H_{T} bin and jet topology (symmetric/asymmetric). The systematic derived from these tests is in the range $3 - 12\%$.



(a) **(b)**

Figure 7.15: Data-driven tests probing the W polarisation effects. These are shown for each n_{jet} category (open symbols) overlaid on top of the systematic uncertainty estimates used for each of the seven H_{T} bins (shaded bands). The symmetric (asymmetric) jet topologies are shown in the left (right) plot.

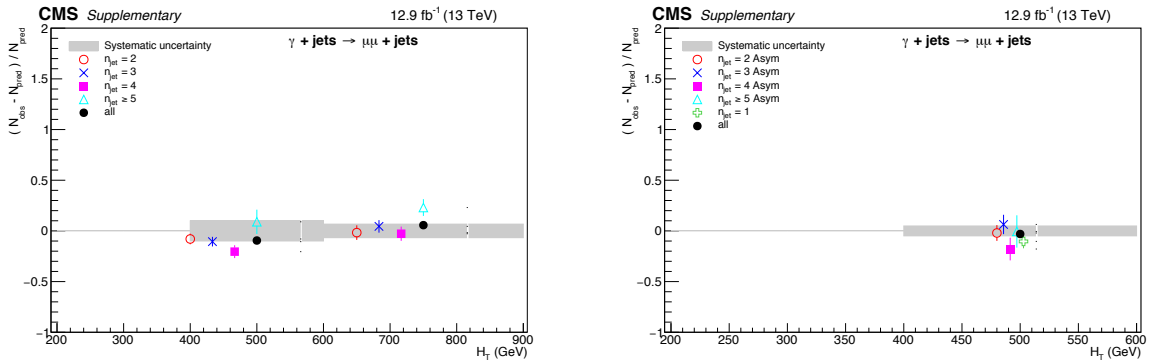
Modelling of the Z/γ ratio

To validate the use of γ + jets events to predict the $\text{Z} \to \nu\bar{\nu}$ background, tests are performed in data using the photon and double-muon control regions. The events in the γ + jets control are used to predict events in the $\mu\mu$ + jets control regions, again using transfer factors from simulation. These tests target the modelling of the Z/γ ratio in simulation and also indirectly test muon/photon acceptance effects, which are expected to be sub-dominant and whose uncertainties are already addressed elsewhere.

The results are shown in Fig. 7.16 as a function of H_{T} and n_{jet}. The grey band is the systematic uncertainty propagated through the analysis, taken as un-correlated per each H_{T} bin and jet topology (symmetric/asymmetric). The systematic derived from these tests is in the range $7 - 15\%$.

Modelling of the W/$t\bar{t}$ admixture

The 0 b-tag $\to$ 1 b-tag data-driven tests in the μ + jets control region probe the sensitivity of the TFs to the relative admixture of events from the W + jets and $t\bar{t}$ processes, since they utilise a W-enriched sample to predict a $t\bar{t}$-enriched sample. These tests also indirectly probe the modelling of the b-tagging efficiency, although this systematic effect is expected to be smaller and is already addressed by the dedicated



(a) **(b)**

Figure 7.16: Data-driven tests probing the Z/γ ratio for each n_{jet} category (open symbols) overlaid on top of the systematic uncertainty estimates used for each of the seven H_{T} bins (shaded bands). The symmetric (asymmetric) jet topologies are shown in the left (right) plot.

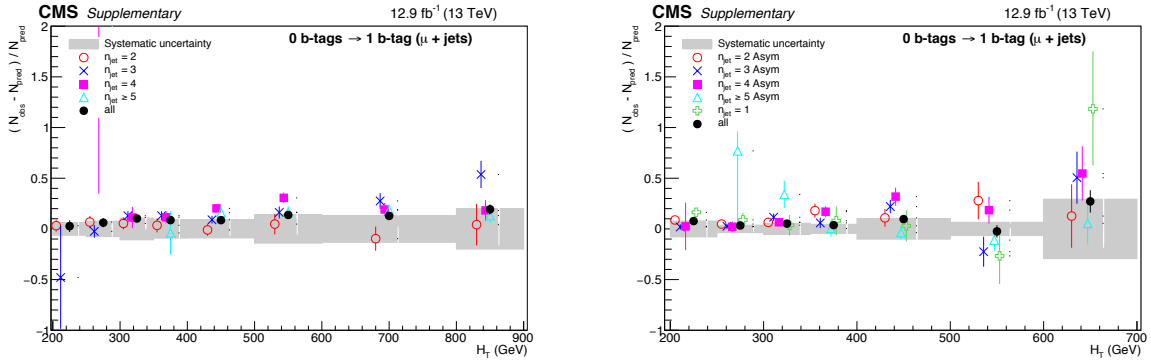
study presented in Sec. 7.5.1. These tests can slightly overestimate the uncertainty, as the admixture changes little between the μ + jets sample and the signal region, given that no extrapolation between different b-tag multiplicities is performed in the estimation of the background.

The results are shown in Fig. 7.17 as a function of H_{T} and n_{jet}. The grey band is the systematic uncertainty propagated through the analysis, taken as un-correlated per each H_{T} bin and jet topology (symmetric/asymmetric). The systematic derived from these tests is in the range $4-25\%$.

7.5.3 Uncertainties in the $\not{H}_{\text{T}}$ dimension

Along with the uncertainty on the TFs, described above, additional uncertainties in the shape of the $\not{H}_{\text{T}}$ distribution that is taken from simulation are derived. For all the known sources of systematic uncertainty, discussed in Sec. 7.5.1, the effect of the $\pm 1\sigma$ variation is propagated as an uncertainty that varies the $\not{H}_{\text{T}}$ shape.

There is also an uncertainty associated with the fact that the $\not{H}_{\text{T}}$ shape is taken directly from simulation. To estimate this, a data-driven method is devised, based on the validation of the $\not{H}_{\text{T}}$ shapes described in Sec. 7.3.2. Within the μ + jets, $\mu\mu$ + jets and γ + jets control regions the data/MC ratio of the $\not{H}_{\text{T}}$ shape is constructed for each (H_{T},n_{jet},n_{b}) bin. The agreement between the two distributions is parameterised with



(a) **(b)**

Figure 7.17: Data-driven tests probing the W and $t\bar{t}$ admixture in each n_{jet} category (open symbols) overlaid on top of the systematic uncertainty estimates used for each of the seven H_{T} bins (shaded bands). The symmetric (asymmetric) jet topologies are shown in the left (right) plot.

a linear orthogonal polynomial. This function is chosen as it allows one to make a linear fit with just one parameter, as the total overall normalisation is preserved. To derive the uncertainty for each of the background sources, a simultaneous fit of the $\not{H}_{\text{T}}$ shapes is performed across all the relevant control regions. This gives a value for the fit parameter, p, which should be compatible with flat, i.e. $p \approx 0$. The uncertainty is then taken as the quadrature sum of the value of p with its one sigma deviation. This allows alternative $\not{H}_{\text{T}}$ shapes to be derived based on the $\pm 1\sigma$ variation of this uncertainty. These alternative shapes are used to encode the data-driven uncertainty when it is propagated to the final result. An example of the kind of fit that occurs in a particular bin was shown in Fig. 7.4. The uncertainty of the linear fit is shown as a red shaded region.

The size of the systematic uncertainty for the data-driven orthogonal polynomial variation and the sources of known systematic uncertainties for one of the extremal analysis categories is shown in Fig. 7.18. The upwards and downward variation for each of the labelled sources are plotted for a normalised $\not{H}_{\text{T}}$ shape. The data-driven systematic dominates over other sources.

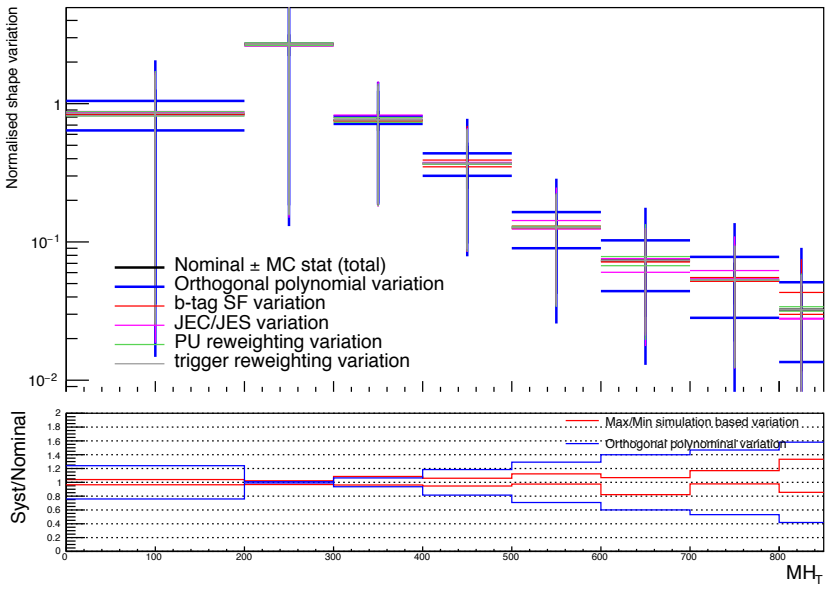


Figure 7.18: The systematic variation of the normalised $\not{H}_{\mathrm{T}}$ (denoted MH_T) distribution for an array of uncertainties derived from simulation and a data driven *orthogonal polynomial variation* in the extremal analysis category: H_{T} $800 - \infty$, $n_{\mathrm{jet}} \geq 5$, $n_{\mathrm{b}} \geq 2$.

Table 7.3: Summary of the systematics on the transfer factors considered in the analysis, with representative ranges of uncertainties and the correlation assummed, for the predictions of the $t\bar{t}$, W and $Z \to \nu\bar{\nu}$ background components.

Systematic	Method	Relative uncertainty on transfer factor				Correlation model
		$\mu + jets \to$ $Z \to \nu\bar{\nu}$	$\mu\mu + jets \to$ $Z \to \nu\bar{\nu}$	$\gamma + jets \to$ $Z \to \nu\bar{\nu}$	$\mu + jets \to$ $t\bar{t} + W$	
$\alpha_T/\Delta\phi^*_{\text{min}}$ extrapolation	data-driven tests	$3-30\%$	$3-30\%$	-	$3-30\%$	un-correlated across H_T/jet top.
W/Z ratio	data-driven tests	$4-15\%$	-	-	-	un-correlated across H_T/jet top.
Z/γ ratio	data-driven tests	-	-	$6-11\%$	-	un-correlated across H_T/jet top.
W/$t\bar{t}$ admixture	data-driven tests	-	-	-	$4-30\%$	un-correlated across H_T/jet top.
W polarisation	data-driven tests	$2-10\%$	-	-	$2-10\%$	un-correlated across H_T/jet top.
Jet energy scale	MC variations	$1-5\%$	$1-5\%$	$1-5\%$	$1-5\%$	fully correlated
B-tagging efficiency b and c jets	MC variations	$1-3\%$	$1-3\%$	$1-3\%$	$1-3\%$	fully correlated
B-tagging efficiency light jets	MC variations	$1-3\%$	$1-3\%$	$1-3\%$	$1-3\%$	fully correlated
Pileup weights	MC variations	$0-2\%$	$0-2\%$	$0-2\%$	$0-2\%$	fully correlated
Top p_T weights	MC variations	$1-30\%$	$1-10\%$	-	$1-10\%$	fully correlated
Lepton scale factor	MC variations	$1-3\%$	$1-3\%$	-	$1-3\%$	fully correlated
Signal trigger efficiency	MC variations	$1-2\%$	$1-2\%$	$1-2\%$	$1-2\%$	fully correlated
Photon trigger efficiency	MC variations	-	-	$1-2\%$	-	fully correlated

7.6 The likelihood model

To carry out a full interpretation of the results of the analysis with appropriate treatment of systematic uncertainties, a likelihood model is constructed. Events are categorised based on their (H_{T},n_{jet},n_{b}) bin, labelled ${H_{\mathrm{T}}}^{\mathrm{cat}}$, in both the signal region and control regions. Additionally, the signal region is categorised based on the $\not{H}_{\mathrm{T}}$ bins discussed in Sec. 7.3.2. These bins are represented with an additional index, i. Given that $n_{\mathrm{had},i}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}$ is the number of observed events, $b_{\mathrm{had},i}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}$ is the number of predicted background events and $s_{\mathrm{had},i}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}$ is the expected number of signal events, the likelihood function in each ${H_{\mathrm{T}}}^{\mathrm{cat}}$ category is defined as the product of poisson distributions for each $\not{H}_{\mathrm{T}}$ bin:

$$\mathcal{L}_{\mathrm{had}}^{{H_{\mathrm{T}}}^{\mathrm{cat}}} = \prod_i \mathrm{Poisson}(n_{\mathrm{had},i}^{{H_{\mathrm{T}}}^{\mathrm{cat}}} \mid b_{\mathrm{had},i}^{{H_{\mathrm{T}}}^{\mathrm{cat}}} + s_{\mathrm{had},i}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}). \tag{7.5}$$

For each control region, indexed j, an independent likelihood function can be constructed, given $n_{\mathrm{CR},j}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}$ is the number of observed events, $b_{\mathrm{CR},j}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}$ is the number of predicted background events and $s_{\mathrm{had},j}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}$ is the expected signal contamination in the control region:

$$\mathcal{L}_{\mathrm{CR,j}}^{{H_{\mathrm{T}}}^{\mathrm{cat}}} = \mathrm{Poisson}(n_{\mathrm{CR,j}}^{{H_{\mathrm{T}}}^{\mathrm{cat}}} \mid b_{\mathrm{CR,j}}^{{H_{\mathrm{T}}}^{\mathrm{cat}}} + s_{\mathrm{CR,j}}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}). \tag{7.6}$$

Due to the way the control regions are selected, the expected signal contamination is usually found to be negligible.

This results in a total likelihood that is the product over all the ${H_{\mathrm{T}}}^{\mathrm{cat}}$ bins and all the control regions. It can be defined as:

$$\mathcal{L} = \prod_{{H_{\mathrm{T}}}^{\mathrm{cat}}} (\mathcal{L}_{\mathrm{had}}^{{H_{\mathrm{T}}}^{\mathrm{cat}}} \times \prod_{\mathrm{j}} \mathcal{L}_{\mathrm{CR,j}}^{{H_{\mathrm{T}}}^{\mathrm{cat}}}) \tag{7.7}$$

7.6.1 Incorporation of systematic uncertainties

The background yields in the signal region are connected to the yields in the control regions through the use of TFs, as discussed in Sec. 7.3.1. To incorporate this into the fit of the likelihood model, the yields of the $\mathrm{Z} \to \nu\bar{\nu}$ and combined W+$t\bar{t}$ backgrounds in the signal region are correlated with the yields in the control regions. This is

implemented through a single floating parameter within the fit, which alters the background prediction in the signal region based on the control region yields and the TF for each category. With this implementation, the systematic uncertainties on each TF can be properly taken into account when making the background prediction in the signal region.

The data-driven uncertainties on the TFs, described in Sec. 7.5.2, are incorporated into the likelihood model as Gaussian distributed nuisance parameters that act on the floating parameter. They depend on the background process and the control region used and are uncorrelated between each of the ${H_{\mathrm{T}}}^{\mathrm{cat}}$ bins.

The uncertainties on the TFs determined from simulation, described in Sec. 7.5.1, are included as *shape* uncertainties on the TFs. This means they are fully correlated across all of the ${H_{\mathrm{T}}}^{\mathrm{cat}}$ bins but their magnitude is different depending on the bin. Additionally, an uncertainty on the expected signal contribution can also be taken account of with shape uncertainties. These extra uncertainties depend on the signal model in question and are discussed in Sec. 8.2.

The uncertainties derived from simulation can also cause bin migration of events within the $\not{H}_{\mathrm{T}}$ dimension of the signal region. This is incorporated as a *template* uncertainty, i.e. alternative $\not{H}_{\mathrm{T}}$ distributions for the $\pm 1\sigma$ variation for each of the sources of uncertainty.

To take account of the uncertainties in the $\not{H}_{\mathrm{T}}$ distribution, explained in Sec. 7.5.3, additional template uncertainties are introduced that are decorrelated in n_{jet} and H_{T}. Their $\pm 1\sigma$ is determined by the uncertainty on the linear fit used when determining the magnitude of the $\not{H}_{\mathrm{T}}$ uncertainty.

Finally, the uncertainty from the limited statistical power of the MC samples used is incorporated as an additional nuisance parameter per ${H_{\mathrm{T}}}^{\mathrm{cat}}$ and $\not{H}_{\mathrm{T}}$ bin. This is taken as uncorrelated across all bins.

The full likelihood for the hadronic signal region including the systematics and free floating parameters, with ${H_{\mathrm{T}}}^{\mathrm{cat}} \equiv J$, can be written as:

$$\mathcal{L}^{J}_{\mathrm{had}} = \prod_{i} \mathcal{L}^{J,i}_{\mathrm{had}} = \prod_{i} \mathrm{Pois}(n^{J,i}_{\mathrm{had}} | \sum_{j \in J} b^{j,i}_{\mathrm{Z}\to\nu\bar{\nu}\,,\mathrm{had}} \times \phi^{j}(\mu\mu \to (\mathrm{Z} \to \nu\bar{\nu})\,) \times a^{j} \times \rho^{j,i}_{\mathrm{Z}\to\nu\bar{\nu}\,,\mathrm{had}} + b^{j,i}_{\mathrm{W}+t\bar{t},\mathrm{had}} \times \phi^{j}(\mu \to (\mathrm{W} + t\bar{t})) \times a^{j} \times \rho^{j,i}_{\mathrm{W}+t\bar{t},\mathrm{had}} + b^{j,i}_{\mathrm{QCD,had}} \times \omega^{j,i}_{\mathrm{QCD,had}} + r \times s^{j,i}_{\mathrm{had}} \times \rho^{j,i}_{s,\mathrm{had}}), \tag{7.8}$$

where $b^{j,i}_{\mathrm{Z}\to\nu\bar{\nu}\ /W+t\bar{t},\mathrm{had}}$ are the predicted number of events from simulation for the electroweak backgrounds; $b^{j,i}_{\mathrm{QCD,had}}$ are the predicted number of events for the QCD multijet component; the a^j parameters are the parameters that can float in the fit; ϕ^j contains the systematic uncertainties on the transfer factors from the data-driven tests; $\rho^{j,i}$ contains the systematics from variations in simulation, the systematics derived from the control regions on the $\not{H}_{\mathrm{T}}$ shape and the uncertainty from the limited number of simulated events; r is the unconstrained 'signal strength' parameter and $\omega^{j,i}_{\mathrm{QCD,had}}$ contains the uncertainties on the QCD multijet component. The systematics enter into the likelihood models for the control regions in a similar way.

7.6.2 Fitting

The fit of the likelihood model to find the final result is carried out in two steps. Firstly, the predicted background yields in the signal region, $b^{H_{\mathrm{T}}\ \mathrm{cat}}_{\mathrm{had}}$, are determined with a fit in only the control regions. This is followed by a full fit using the full likelihood model, Eq. 7.7, taking account of all the correlations between the control regions and signal region, along with the statistical uncertainties from the finite number of events in the control region. Within the fit the likelihood is profiled against all nuisance parameters, which allows the determination of limits on various signal models. This is discussed further in Sec. 8.2.

Chapter 8

Results and interpretations

The analysis described in the previous Chapters has been carried out on the $12.9 \pm 0.8\,\text{fb}^{-1}$ dataset that was introduced in Sec. 7.1, the results are described in this chapter. The SM backgrounds have been predicted using the likelihood fit discussed in Sec. 7.6. The data has then been compared to the expected background yields and limits are set on the production of an array of different SUSY models. This result has been released publicly at [130].

8.1 Results

The total predicted SM yields for each of the (H_{T},n_{jet},n_{b}) bins, integrated over the $\not{H}_{\text{T}}$ dimension, are shown in Figs. 8.1, 8.2 and 8.3 for the monojet, asymmetric and symmetric jet categories respectively. In the top panel of each figure, the data counts with a representative statistical uncertainty are shown as black circles with error bars. The coloured histogram shows the result of the SM background prediction with the TF methods described in Chapter 7, the uncertainty of this prediction is represented with a shaded box (CR-only fit uncertainty). The predictions are split into the $\text{Z} \to \nu\bar{\nu}$, QCD multijet and other remaining SM backgrounds. On the bottom panel of each plot, the significance of the deviation of the data from the predicted SM background is plotted. The red circles show the deviation of the control region only background fit (Eq. 7.6), while the blue circles show the deviation of the fit that includes the signal region (Eq. 7.7). The deviation is represented as a *pull* that is defined as the number of observed events, minus the number of predicted events, divided by the 1σ uncertainty.

Along with the background prediction in each of the (H_T,n_{jet},n_b) categories, simulation is used to predict the $\not{H}_T$ shapes. The result of this prediction for a series of representative bins is shown in Fig. 8.4. The data yields and statistical errors are displayed by the black markers. The $\not{H}_T$ shape taken from simulation normalised based on the results from the control region only fit is displayed as the green histogram.

Overall, no significant deviation from the SM backgrounds is observed in the data. The results are well described by a SM-only hypothesis. There are a few large pulls, $\sim 3\sigma$, that are observed in the control-region only fit. However, after a fit including the signal region is carried out the pulls are significantly reduced. This suggests that such effects are properly covered by systematic uncertainties.

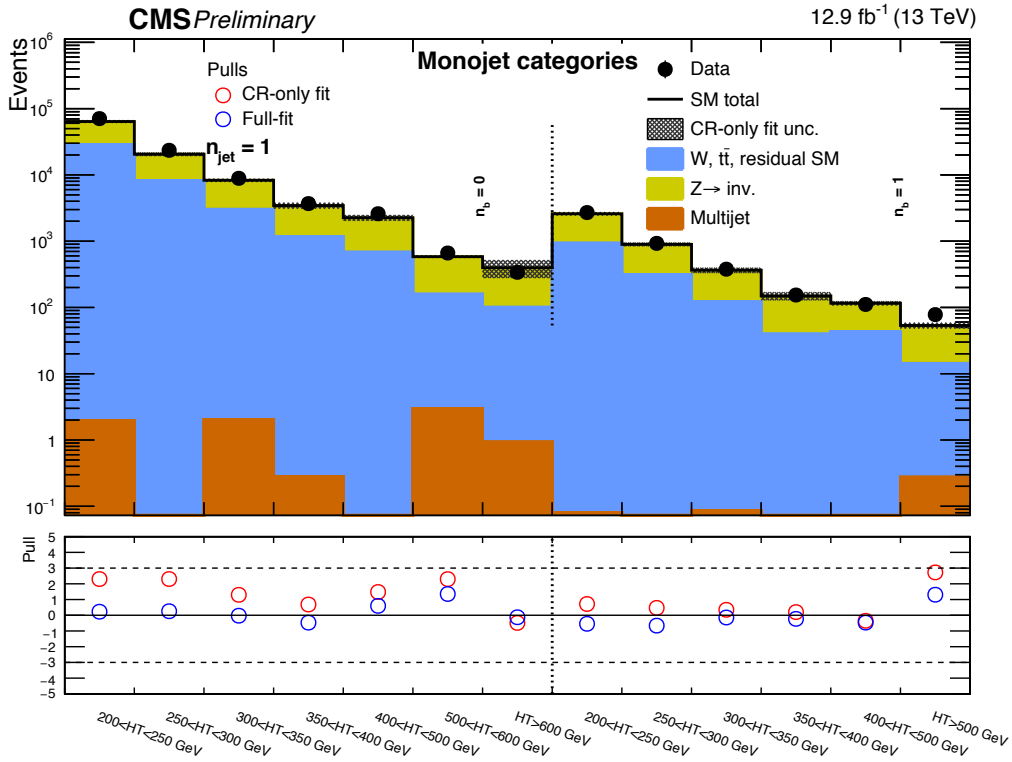


Figure 8.1: The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (black histogram with shaded band) as a function of n_b and H_T for the monojet topology ($n_{jet} = 1$) in the signal region. Under this is the significance of deviations (pulls) observed in data with respect to the SM expectations from the fit with only the control regions (red circles) and a full fit including the signal region (blue circles).

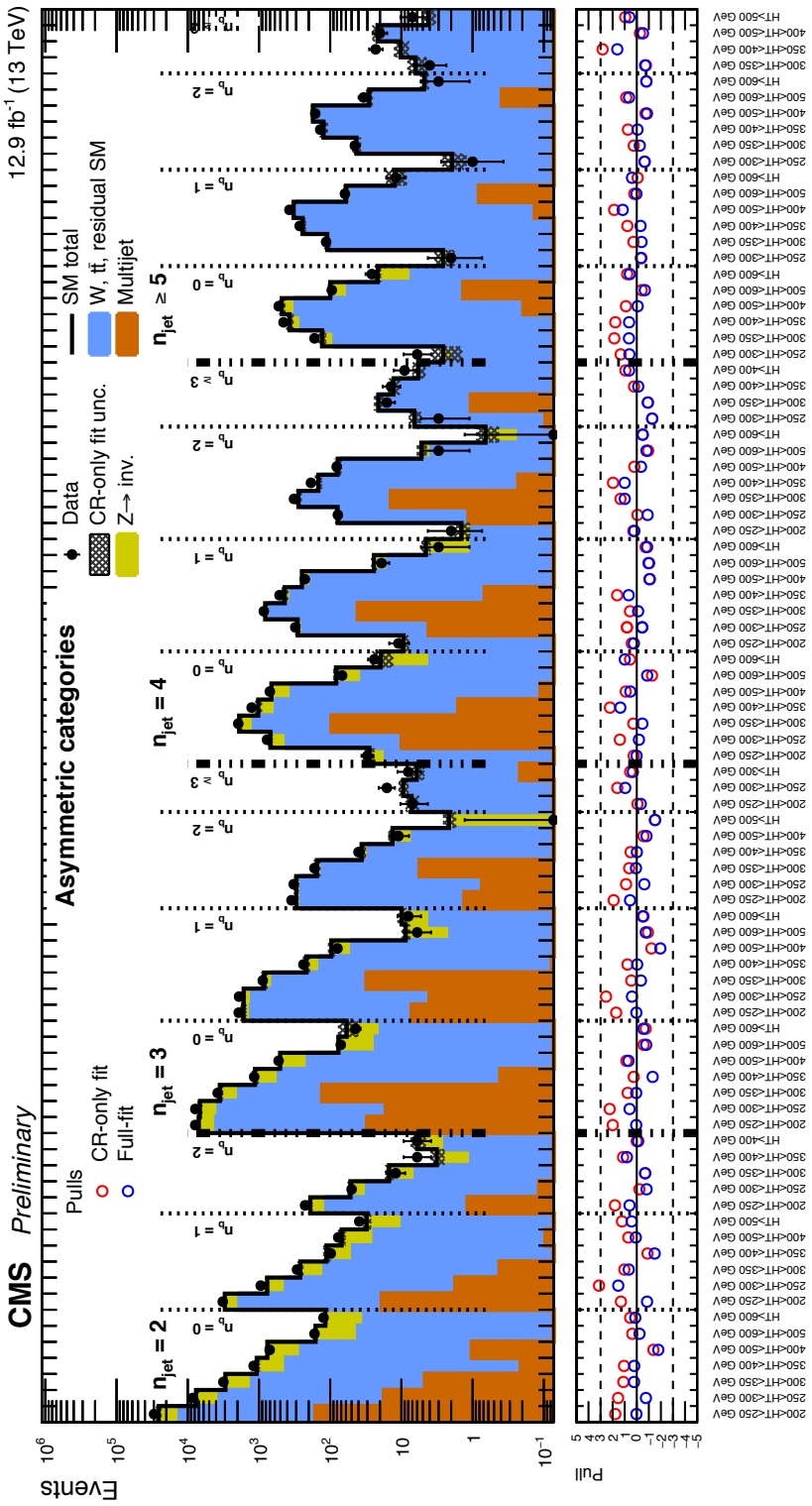


Figure 8.2: The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (black histogram with shaded band) integrated over $\not{H}_T$ as a function of n_{jet}, n_b and H_T for the asymmetric topology in the signal region. Under this is the significance of deviations (pulls) observed in data with respect to the SM expectations from the fit with only the control regions (red circles) and a full fit including the signal region (blue circles).

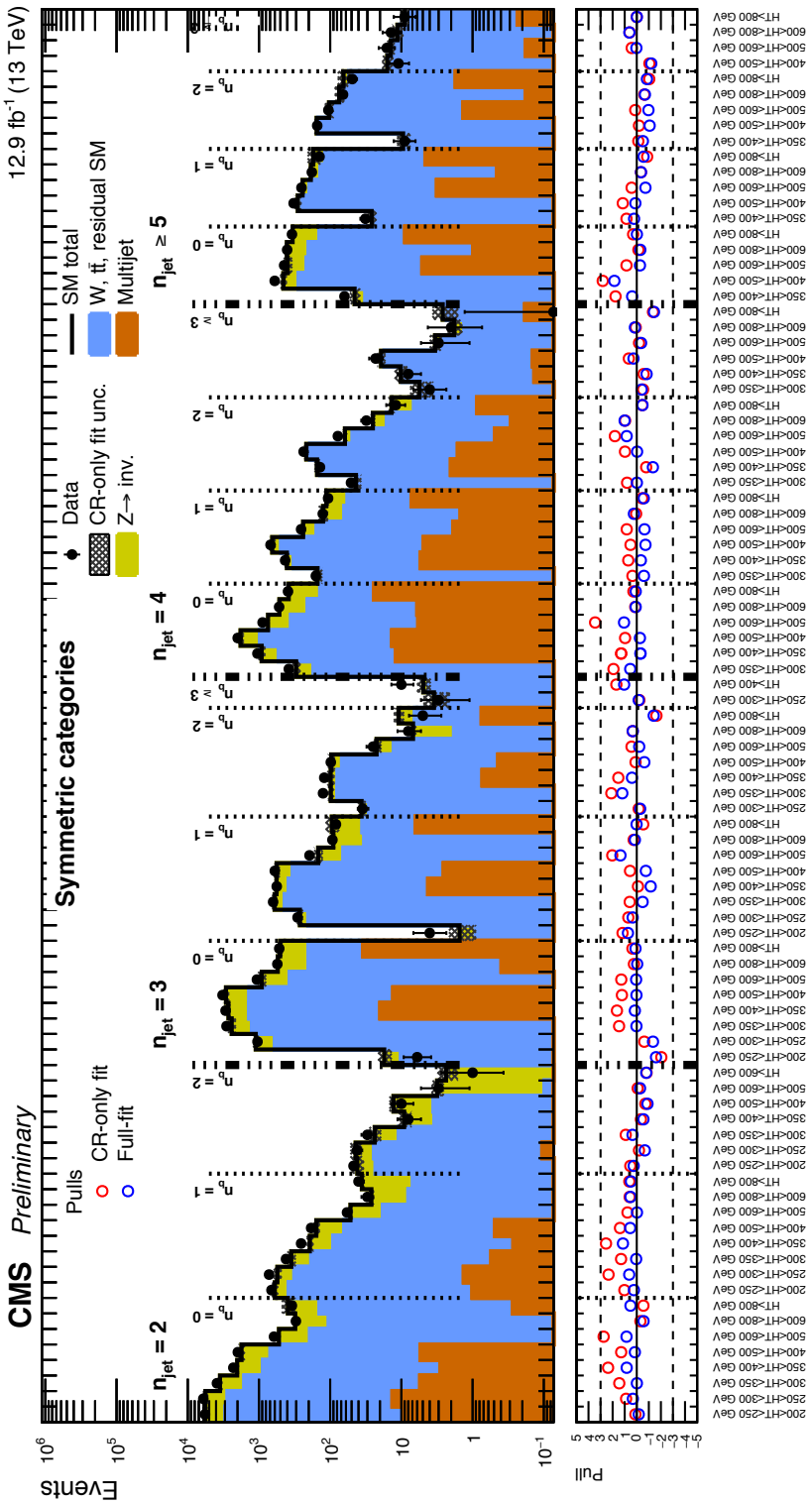
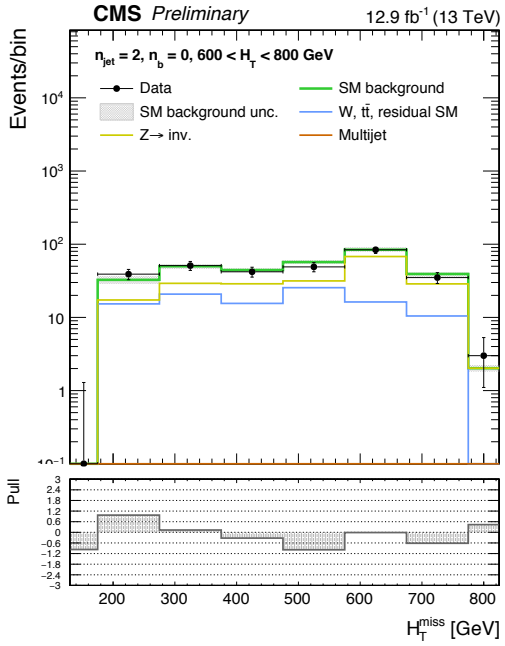
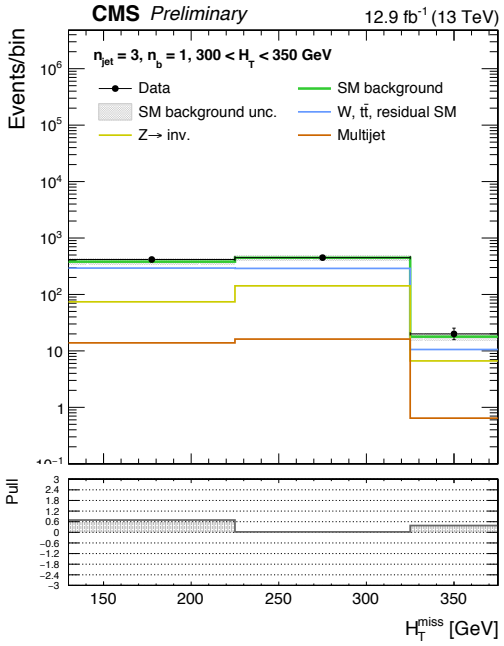


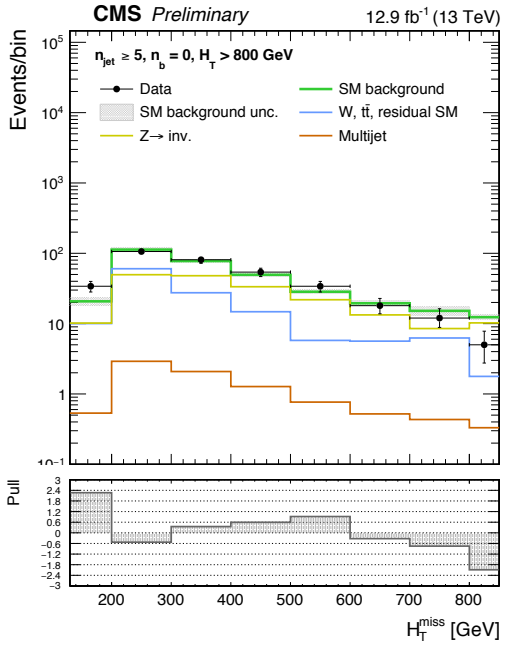
Figure 8.3: The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (black histogram with shaded band) integrated over $\not{H}_T$ as a function of n_{jet}, n_b and H_T for the symmetric topology in the signal region. Under this is the significance of deviations (pulls) observed in data with respect to the SM expectations from the fit with only the control regions (red circles) and a full fit including the signal region (blue circles).



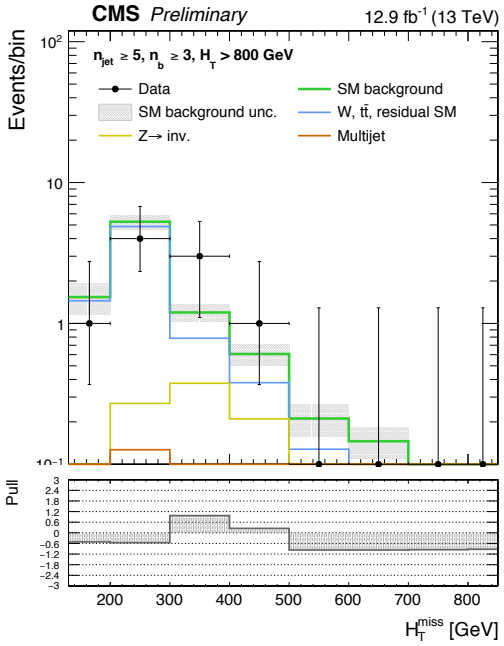
(a) Symmetric topology, medium H_{T}, low n_{jet}



(b) Asymmetric topology, low H_{T}



(c) Symmetric topology, high n_{jet} and H_{T}



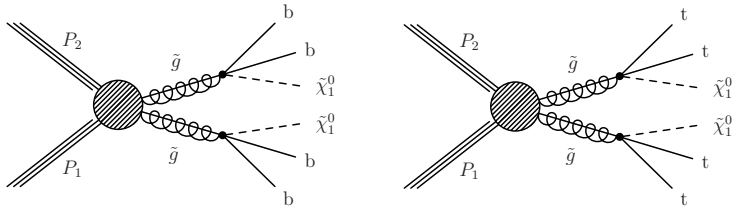
(d) Symmetric topology, high n_{jet}, n_{b} and H_{T}

Figure 8.4: The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (green histogram with shaded band) as a function of $\not{H}_{\mathrm{T}}$ for events in the signal region for four representative signal region categories. The final bin of each histogram is an overflow bin. Under this is the significance of deviations (pulls) observed in data with respect to the SM expectations.

8.2 Interpretation of the results

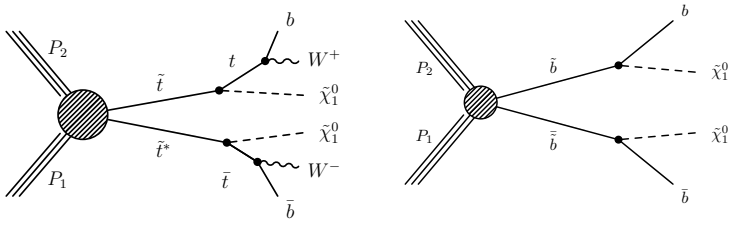
As there is no excess observed in the data above that expected from the SM background production, limits are set on the parameters of possible SUSY models. For the interpretation of the results, the simplified models discussed in Sec. 2.2.3 are utilised. Four particular models are used to interpret the results of the analysis. The models are named based on their production topology, with gluino pair produced models being prefixed with *T1* and squark pair produced models being prefixed with *T2*. Two varieties of the gluino pair produced models are shown in Fig. 8.5, in which the gluino decays via third generation quarks. Their corresponding squark pair produced models are shown in Fig. 8.6.

The predicted yields for each of these signal models are taken directly from MC simulation. To be able to understand which mass range of these models is excluded by the result, a large number of different simulations are produced for a range of gluino, squark and neutralino masses, $m_{\tilde{g}}$, $m_{\tilde{q}}$ and $m_{\tilde{\chi}_1^0}$ respectively. For each of these mass points it is considered whether the model is compatible with the observed results or not. If it is not, it is deemed to be excluded to a particular degree of certainty.



(a) T1bbbb **(b)** T1tttt

Figure 8.5: Feynman diagram of simplified models in which gluinos are pair produced and decay to an LSP via third generation squarks.



(a) T2tt **(b)** T2bb

Figure 8.6: Feynman diagram of simplified models in which stops or sbottoms are pair produced and decay to an LSP via third generation squarks.

8.2.1 Uncertainties on signal models

Sources of systematic uncertainty are propagated to the predicted yields from the MC simulation of the signal models. These uncertainties are considered across all the (H_T,n_{jet},n_b) categories, as well as the $\not{H}_T$ shapes. Most of the uncertainties are determined in the same way as those in the SM background prediction, Sec. 7.5.1. Due to the fact that the signal model yields come from simulation, an additional uncertainty on the total integrated luminosity is included. Additionally, an extra correction factor is applied to the number of jets from initial state radiation. This also has an associated uncertainty. Each of these uncertainties must be determined for each independent simplified model. A summary of the uncertainties for a representative T2bb simplified model and their correlation across all categories is shown in Tab. 8.1. The *normalisation* uncertainties just effect the total yield in each (H_T,n_{jet},n_b), while the *shape* uncertainties include variations in the $\not{H}_T$ shape.

Table 8.1: The magnitude of uncertainties in the signal model yields in the case of a T2bb simplified model.

Systematic source	Type	Correlated	Typical magnitude (%)
Luminosity	Normalisation	Yes	6.2
Monte Carlo statistics	Norm. + shape	No	1–50
Jet energy scale	Norm. + shape	Yes	3–10
b-tag efficiency scale factors	Norm. + shape	Yes	5–40
Lepton scale factors	Normalisation	Yes	1–5
Pile-up	Norm. + shape	Yes	0–5
Trigger efficiency	Norm. + shape	Yes	0–4
Initial state radiation	Norm. + shape	Yes	1–20
Modelling of $\not{H}_T$	Normalisation	Yes	1–5

8.2.2 Exclusion limits

For each of the mass points and all the simplified models introduced in Sec. 8.2, an upper limit at a 95% confidence level (CL) on the cross section to produce the pair of sparticles considered in the model is determined. This limit is produced under a background plus signal hypothesis with a modified frequentist approach. The potential contribution of the simplified models to each of the analysis bins in the signal region is considered. The statistical approach uses a one-sided profile likelihood ratio

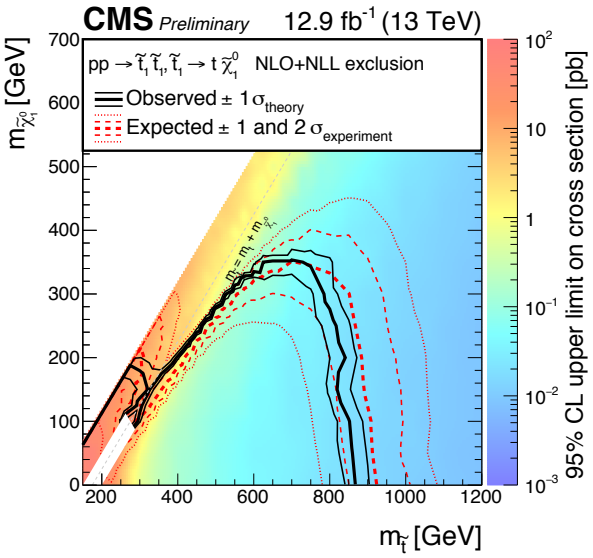
as the test statistic. The limit is set using the CL_S criterion [131,132] and the asymptotic formulae [133] are used to approximate the distributions of the test statistics under the relevant hypothesis (either background only or signal plus background).

The results of the exclusion of two gluino models (T1tttt and T1bbbb) and two squark models (T2tt and T2bb) are shown in Fig. 8.2. The 95% CL upper limit on the cross section is plotted as a two dimensional histogram with a colour scale for a range of different sparticle and LSP masses. The theoretical cross sections for the models are determined with NLO accuracy and there is an assumption of a 100% branching ratio to the final state of each model. Based on this, observed (black) and expected (red) exclusion contours are plotted with their $\pm 1\sigma$ uncertainty bands, encapsulating the experimental uncertainty for the limit and the theoretical uncertainties of the signal model cross section. For the T2tt model, the low mass region is blanked out when the mass of the stop, $m_{\tilde{t}}$, is close to the mass of the top quark and the total $\not{E}_{\mathrm{T}}$ in the event is small. This is due to issues in the $\not{E}_{\mathrm{T}}$ reconstruction of the simulated SUSY model events in this region of parameter space.

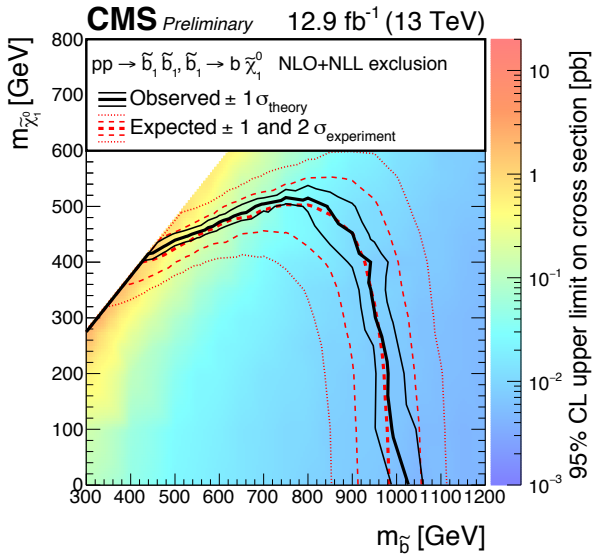
In gluino mediated models, the gluino mass is excluded up to 1775 GeV and the LSP, $\tilde{\chi}^0$, up to 1175 GeV. In the squark mediated models, the sbottom mass is excluded up to 1025 GeV and the stop mass is excluded up to 875 GeV with LSP exclusions of 525 GeV and 350 GeV respectively. The strongest observed exclusions are summarised in Tab. 8.2. This presents a significant improvement on some of the limits found with the 19.5 fb^{-1} result made with $\sqrt{s} = 8$ TeV proton collisions from Run 1. In the best cases the gluino masses were excluded at < 1400 GeV and the sbottom masses to < 800 GeV [43].

The limit curves have a generic structure that is similar for each model. The maximum exclusion of the gluino or squark is fairly consistent across low LSP masses, resulting in a vertical limit on the right of the plot. The mass to which this line extends is generally limited by statistical uncertainties from the small total counts in the highest H_{T} and $\not{H}_{\mathrm{T}}$ bins. The limit here is insensitive to systematic uncertainties of a magnitude that is typical within this analysis. However, the systematic uncertainties do have more of a bearing on the limit along the top of the exclusion curve. The weakening of the limit here is caused by kinematic considerations, which are relevant when the mass of the LSP is close to the mass of the squark or gluino. Some of the more populated bins contribute more significantly to these limits, the magnitude and correlation of systematic uncertainties can therefore change the shape of the curve in this location. The type of systematic uncertainty that has the largest bearing depends

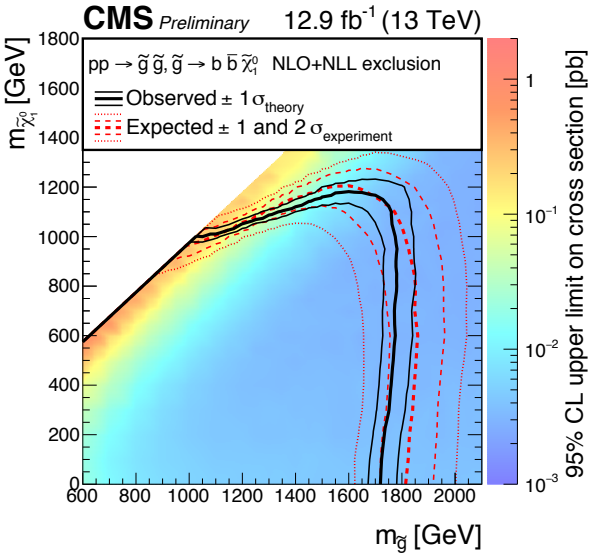
on the model in question. For example, SUSY models with a high number of jets in the final state are more sensitive to the jet energy scale uncertainties. As the SUSY models that are investigated in this analysis typically have high values of n_{b}, n_{jet} and $\not{H}_{\mathrm{T}}$, the systematic uncertainties on these variables are expected to dominate.



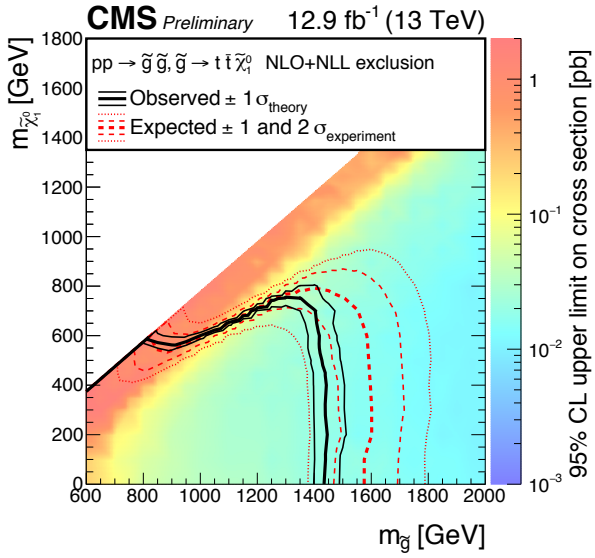
(a) T2tt: Upper limit on the cross section in the $(m_{\mathrm{stop}}, m_{\mathrm{LSP}})$ plane



(b) T2bb: Upper limit on the cross section in the $(m_{\mathrm{sbottom}}, m_{\mathrm{LSP}})$ plane



(c) T1bbbb: Upper limit on the cross section in the $(m_{\mathrm{Gluino}}, m_{\mathrm{LSP}})$ plane



(d) T1tttt: Upper limit on the cross section in the $(m_{\mathrm{Gluino}}, m_{\mathrm{LSP}})$ plane

Figure 8.7: The 95% observed upper limit on the cross section (histogram), with the expected (dotted red line) and observed (black line) exclusion contours. Shown for a selection of SUSY models discussed in Sec. 8.2.

Table 8.2: A summary of the strongest observed (expected) mass exclusions for the simplified models introduced in Sec. 8.2. The limit on the mass of the relevant gluino or squark and the LSP, $\tilde{\chi}_1^0$, are quoted and all have uncertainties of $\pm$ 25GeV.

Production mode	Squark	Strongest obs. (exp.) mass exclusion [GeV]	
		Gluino or squark	$\tilde{\chi}^0$
Gluino-mediated	Bottom	1775 (1850)	1175 (1200)
Gluino-mediated	Top	1450 (1600)	750 (800)
Direct	Bottom	1025 (975)	525 (500)
Direct	Top	875 (925)	350 (350)

Chapter 9

Conclusion

The collection of a significant dataset of $\sqrt{s} = 13$ TeV pp-collision data by the LHC during the early stages of Run 2 has presented the best opportunity in recent years to explore BSM physics at a new energy frontier. With all the particles of the SM discovered, but a series of serious questions unanswered, there is a strong motivation to search for BSM physics. One of the most promising BSM theories, natural SUSY, is now well within reach of the particle physics community. A search for this in the sensitive all-hadronic final state has been carried out and presented within this thesis.

The start of Run 2 saw the LHC move into a new and challenging regime. With a higher collision energy and the promise of a significant increase in PU, the rate of production of high energy physics processes increased substantially. This presented a big challenge for the Level-1 hardware trigger. The hardware itself has been upgraded and a new Level-1 jet algorithm with dynamic PUS has been devised and implemented. This algorithm has had significant success in meeting the challenges of Run 2. It has been commissioned and is currently in operation, allowing low energy thresholds on the types of hadronic events that are collected. This is particularly important for the analyses at CMS that target SUSY.

A search for SUSY in the all-hadronic final state has been described and the results of the search with 12.9 ± 0.8 fb^{-1} of Run 2 collision data has been presented. The analysis revolves around suppressing the dominant QCD multijet background with requirements made on the α_{T} and $\Delta\phi^{*}_{\mathrm{min}}$ topological variables. A data-driven prediction of the multijet background demonstrates that it is reduced to a negligible level. The remaining backgrounds mainly constitute SM processes that produce neutrinos, a source of genuine $\not{E}_{\mathrm{T}}$. These backgrounds are predicted through a data-driven method that makes use of muon and photon control samples. The predicted background yields

in the signal region are obtained by extrapolating from the control regions through transfer factors made from simulation. Appropriate systematic uncertainties on the magnitude of these transfer factors and the shape of the $\not{H}_{\mathrm{T}}$ dimension, which is taken directly from simulation, are calculated. This allows the final result to be obtained through a maximum-likelihood fit for a series of signal hypotheses.

The search has found that the observed data within the signal region is compatible with the expected yields from the SM backgrounds. As no evidence for BSM phenomena is observed, limits are set on the production of a series of key SUSY models. These results extend beyond the limits set during Run 1 of the LHC. The gluino mass is excluded up to 1775 GeV, the sbottom mass is excluded up to 1025 GeV and the stop mass is excluded up to 875 GeV with model dependent exclusions of the LSP in the range of a few hundred GeV.

With the results of hadronic searches for BSM physics not observing any evidence of non-SM phenomena, the case for natural SUSY is starting to weaken. If it were to exist in one of its most commonly conceived forms, it would be expected that the high energy run of the LHC would start to observe evidence of the existence of such SUSY particles. It is still possible that with a larger dataset and a better understanding of the systematic effects in the data, SUSY could present itself at the LHC. However, the best opportunity so far for a glimpse of natural SUSY has resulted in a null result.

There is clear evidence for the existence of DM in the universe, with a weakly interacting massive particle remaining a favoured candidate. The hadronic SUSY searches are in a prime position to reinterpret their results from the perspective of these sorts of generic DM models. The LHC will therefore remain one of the most powerful tools in probing the energy frontier beyond the SM.

Now is a pivotal moment in the history of particle physics. As the LHC continues to collect high energy collision data and the understanding of the data collected by CMS matures, there will be more concrete conclusions about the nature of BSM physics. It will be possible to finally discount the existence of favoured forms of natural SUSY in the next couple of years of LHC data. There is now the potential to observe any manner of other, more difficult to observe, BSM scenarios that may exist at the new energy frontier. If nothing presents itself by the end of Run 2, the field will then have to reassess exactly how best to search for BSM phenomena and how to move into the future.

Appendix A

Appendices

A.1 Characterisation of the signal and control regions

Extra data-MC comparisons of the key analysis variables in each of the signal and control regions of the analysis are included in this section

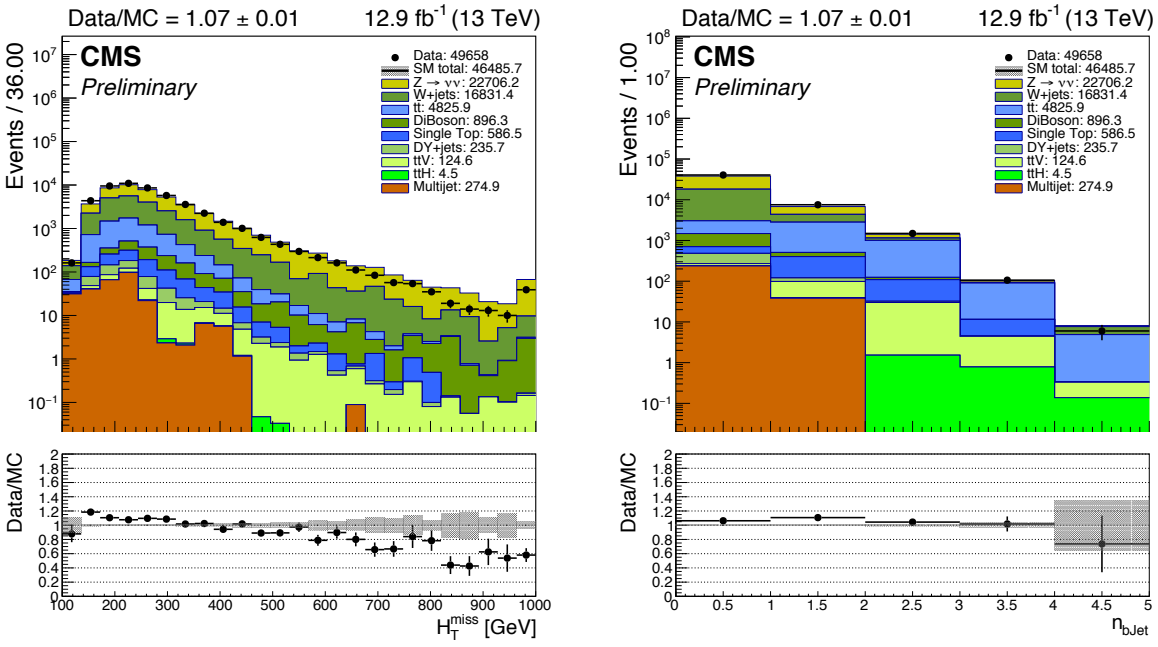


Figure A.1: Key analysis variables for hadronic signal region (symmetric n_{jet} bins)

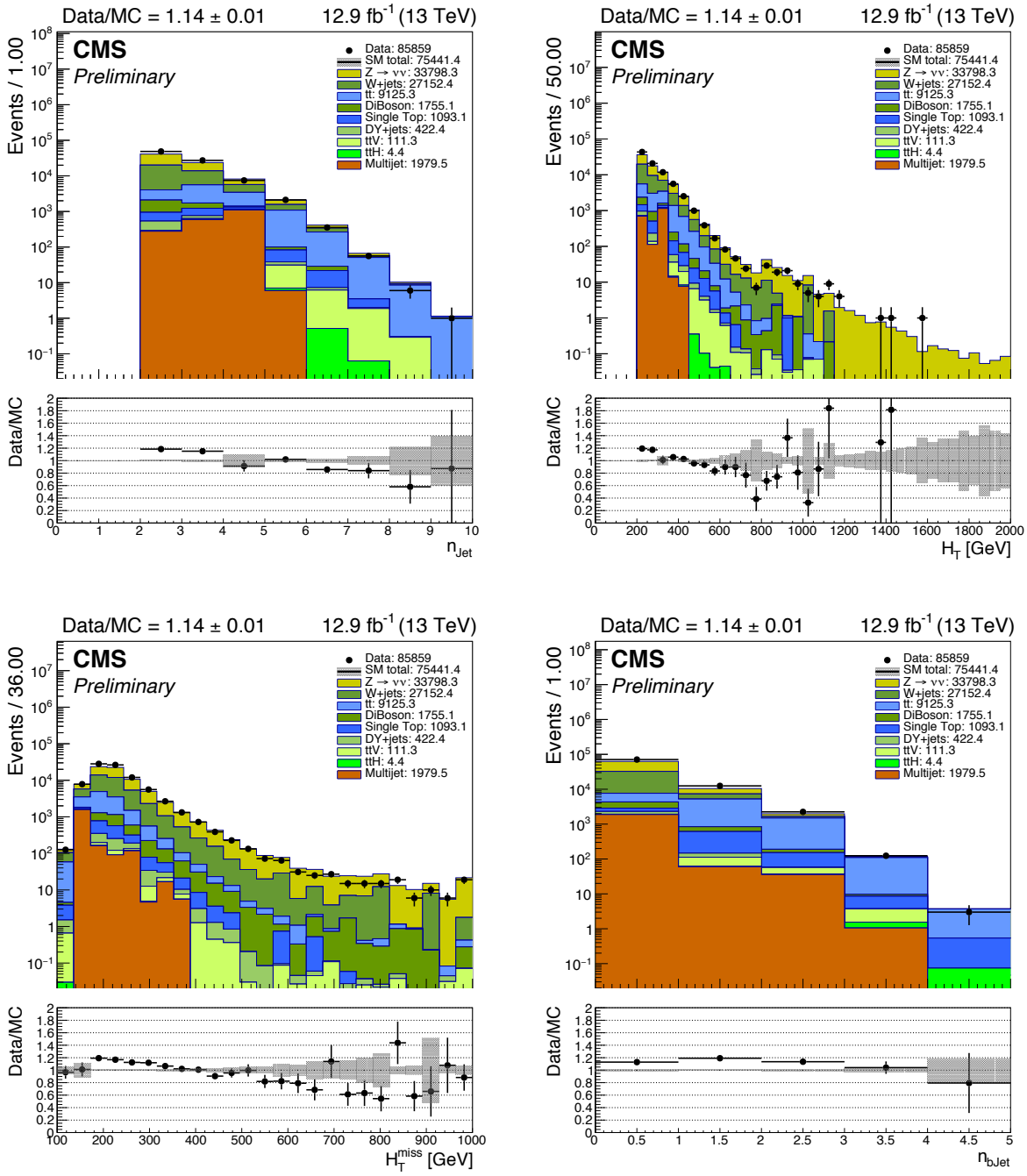


Figure A.2: Key analysis variables for hadronic signal region (asymmetric n_{jet} bins)

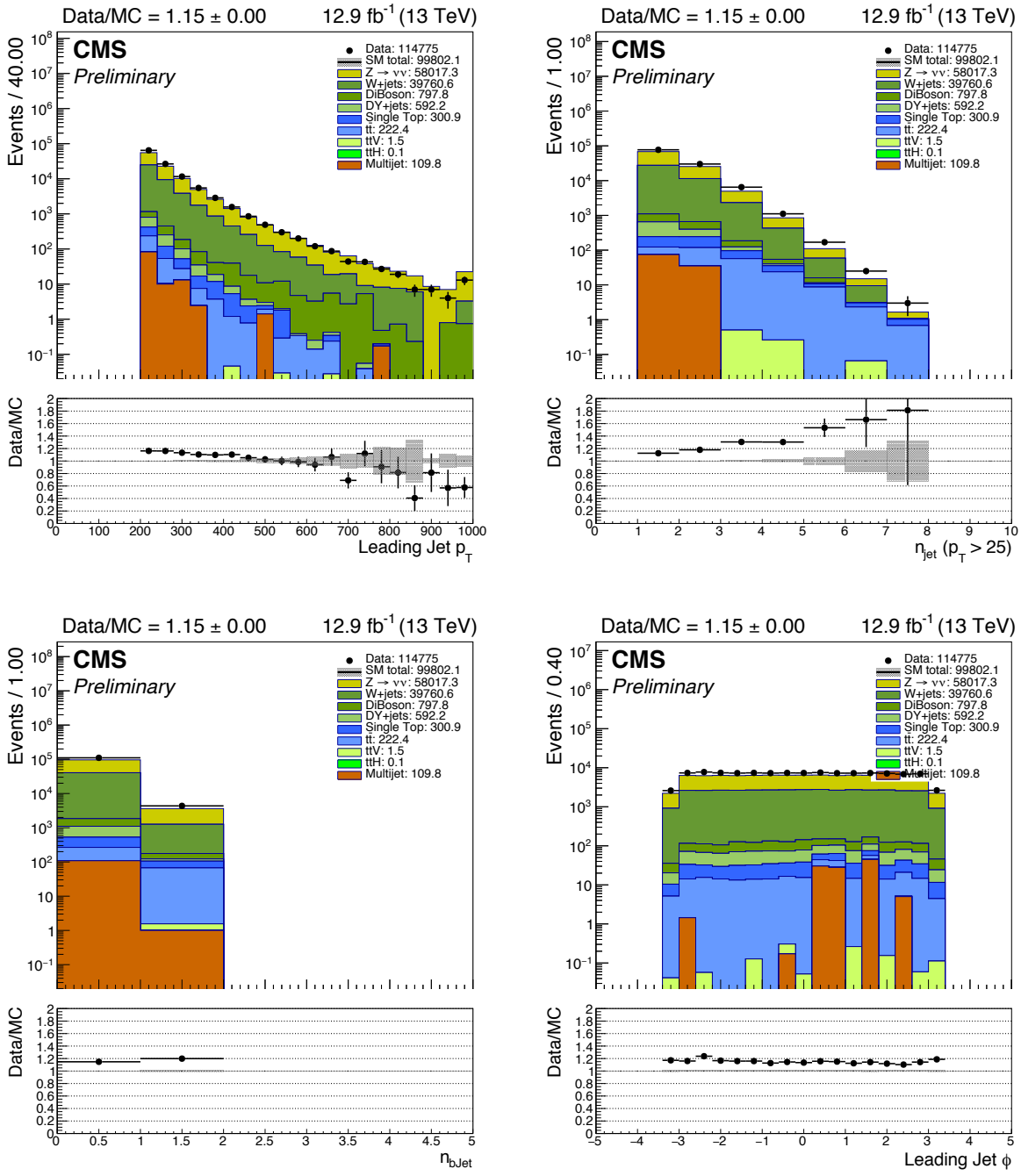


Figure A.3: Key analysis variables for hadronic signal region (monojet bins)

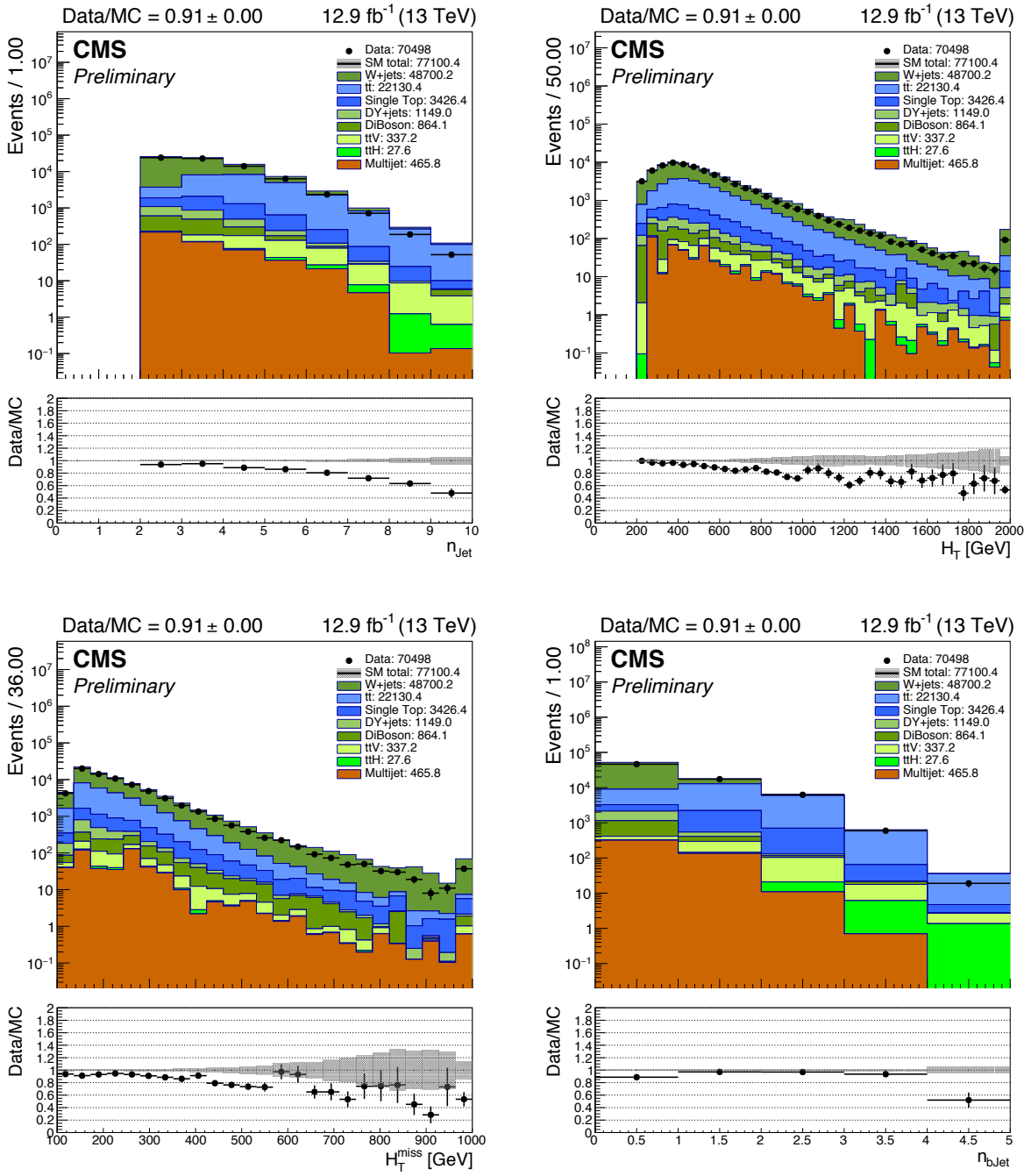


Figure A.4: Key analysis variables for single muon control region (symmetric n_{jet} bins)

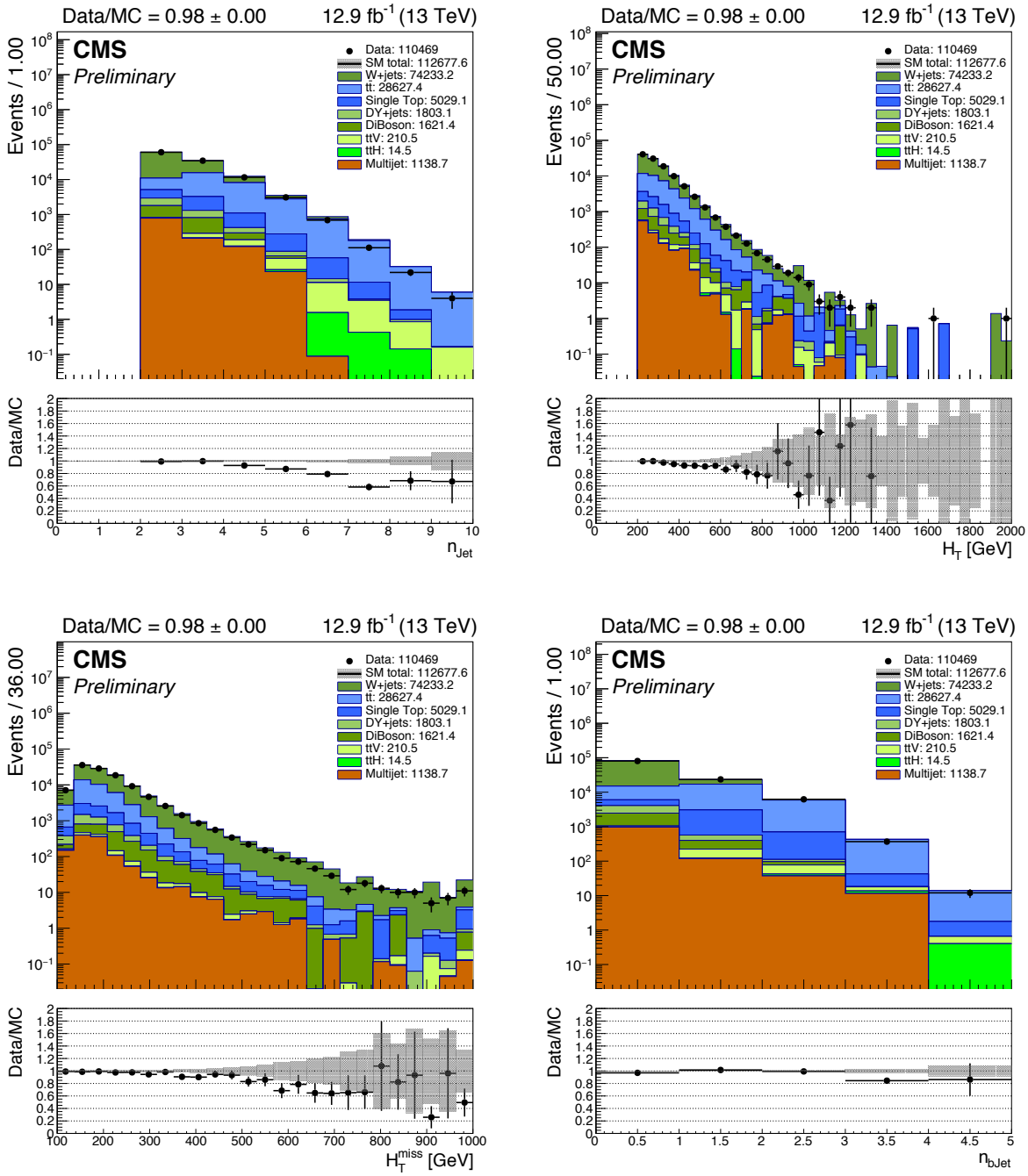


Figure A.5: Key analysis variables for single muon control region (asymmetric n_{jet} bins)

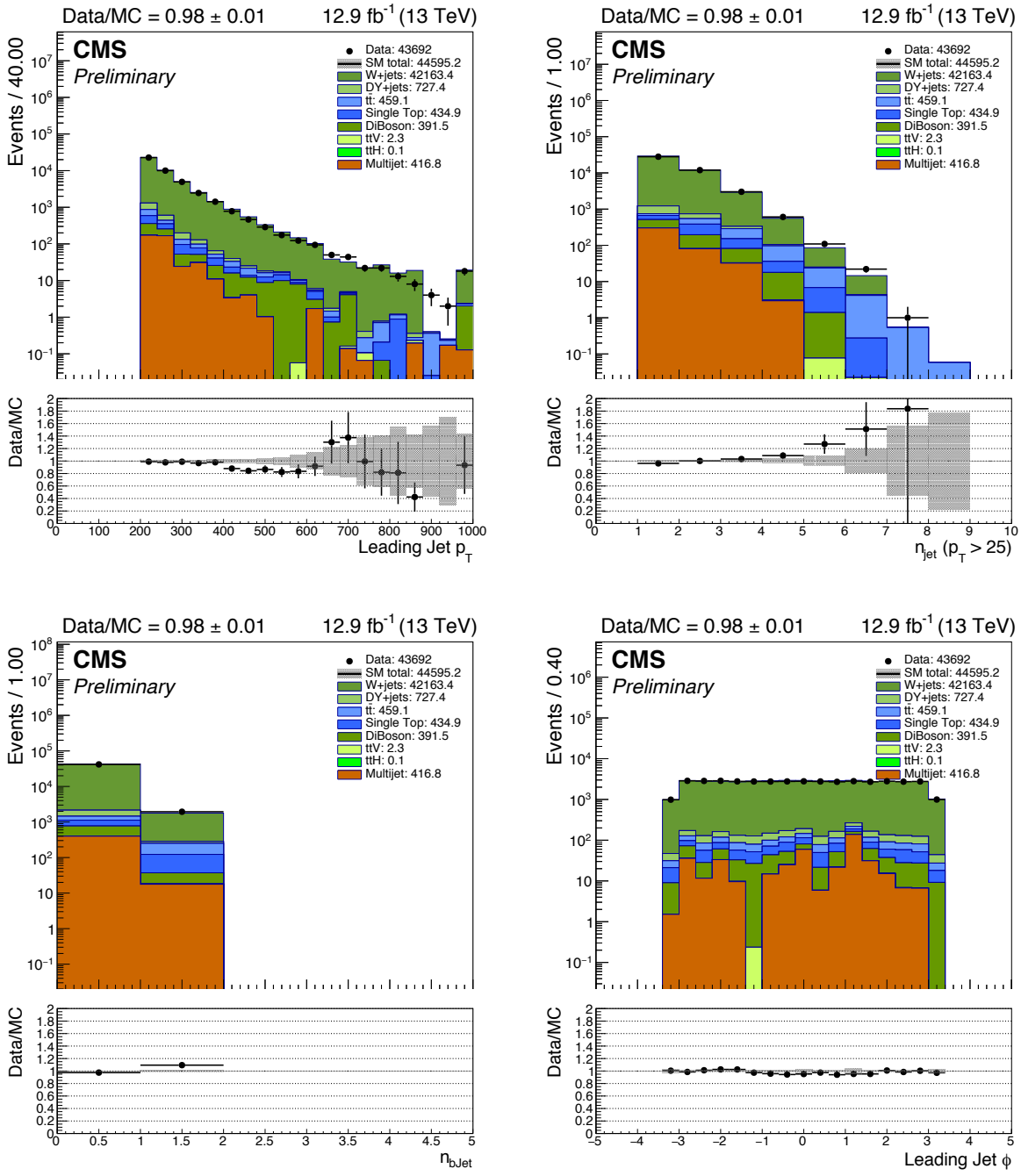


Figure A.6: Key analysis variables for single muon control region (monojet bins)

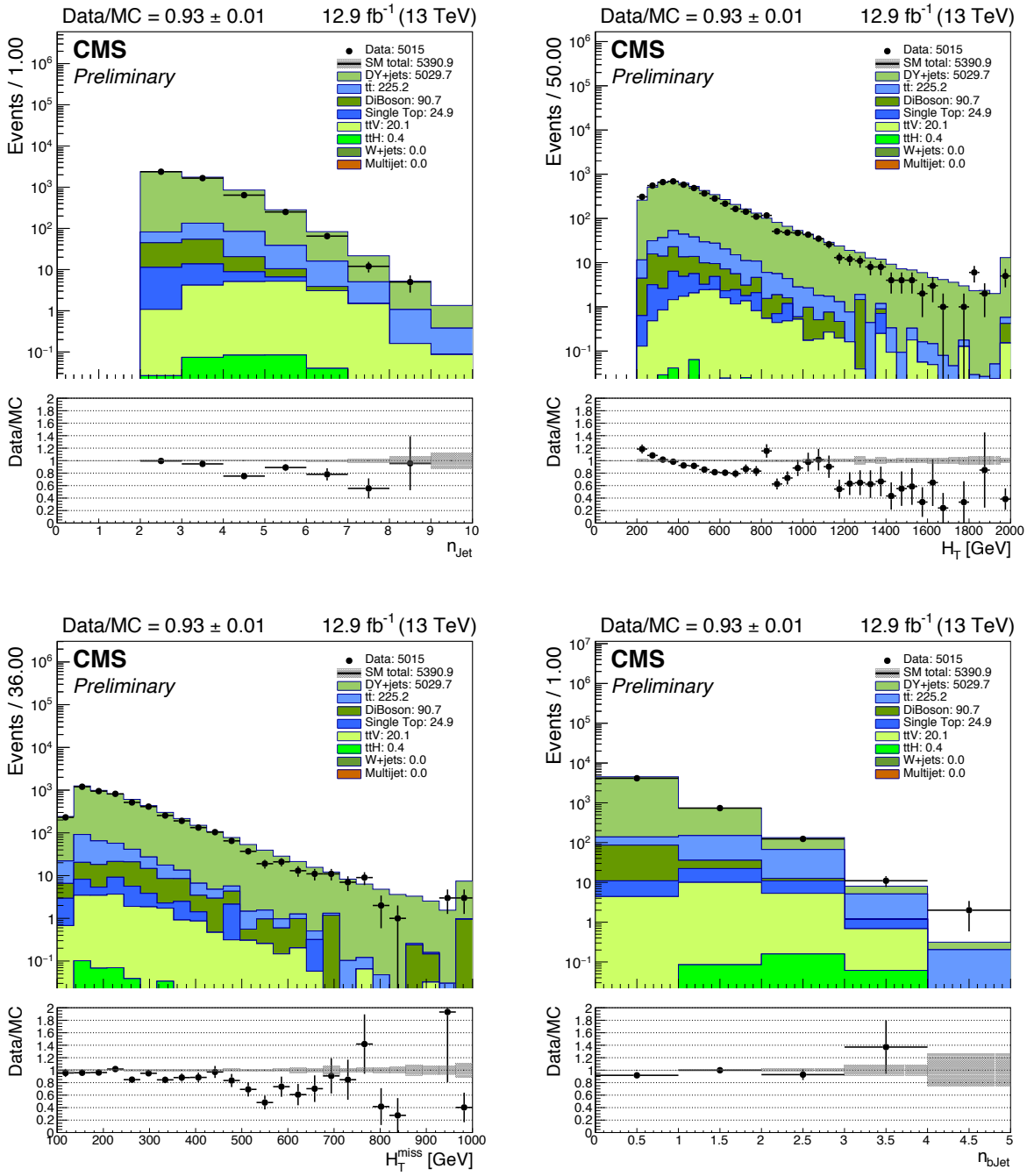


Figure A.7: Key analysis variables for double muon control region (symmetric n_{jet} bins)

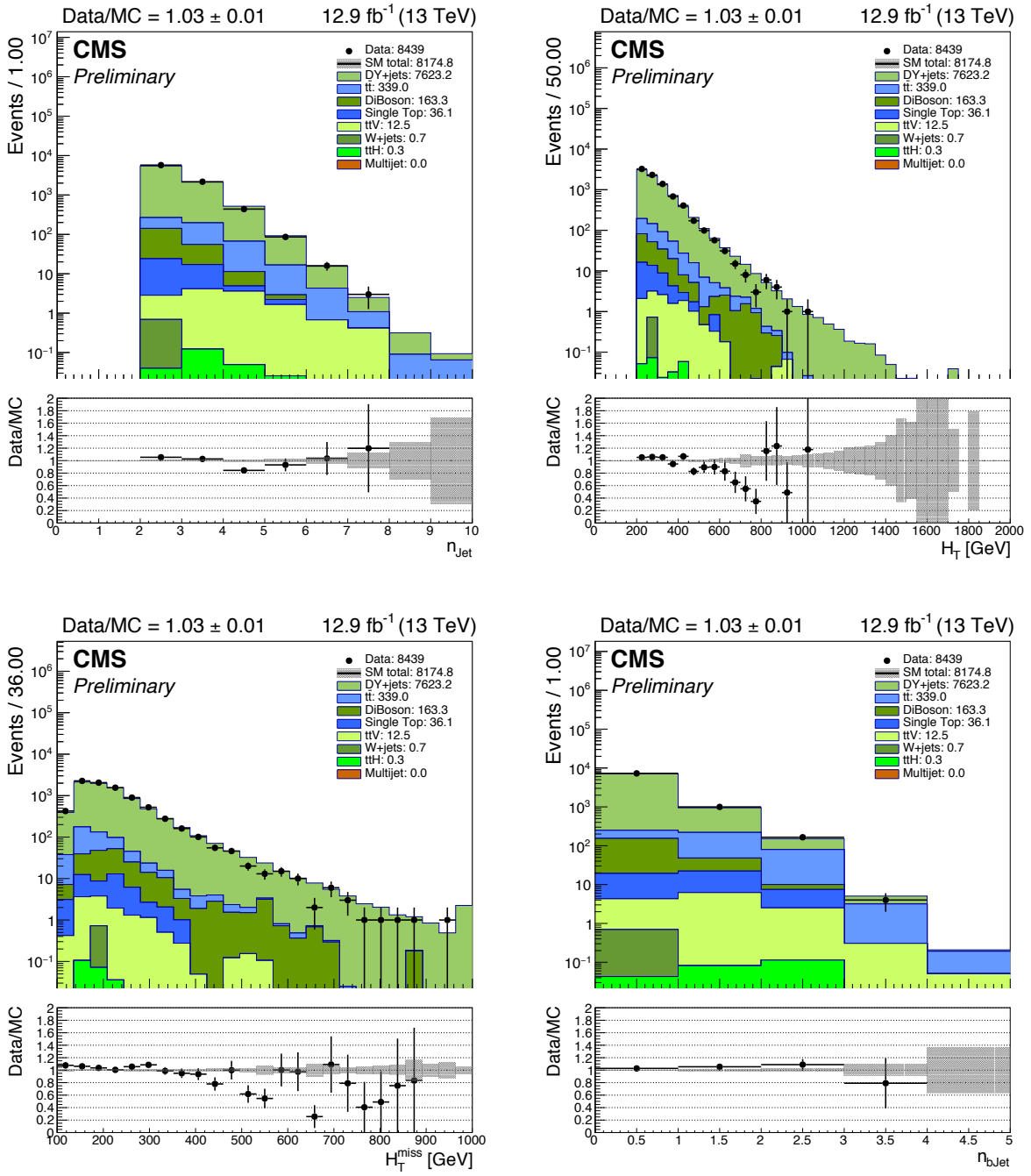


Figure A.8: Key analysis variables for double muon control region (asymmetric n_{jet} bins)

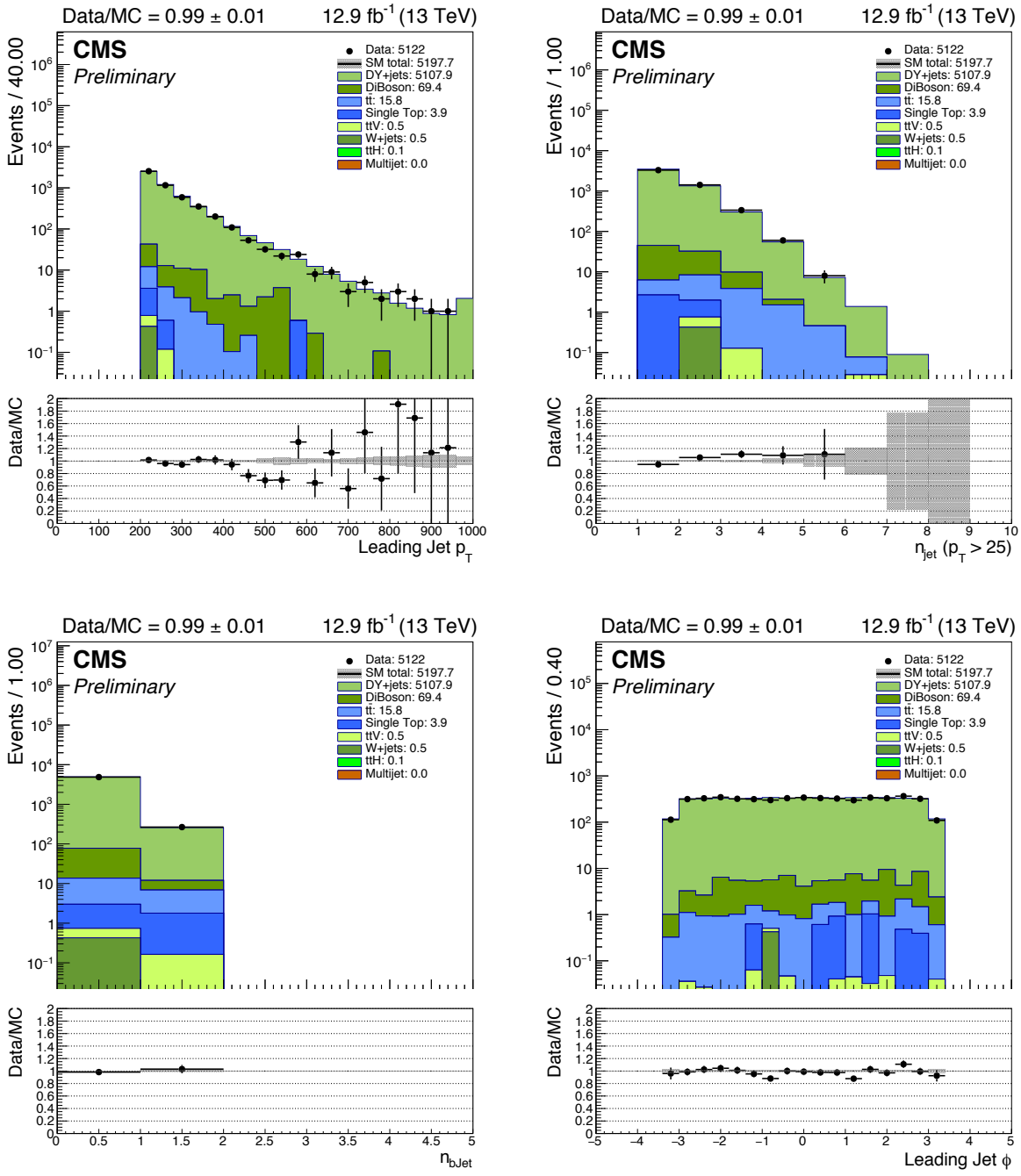


Figure A.9: Key analysis variables for double muon control region (monojet bins)

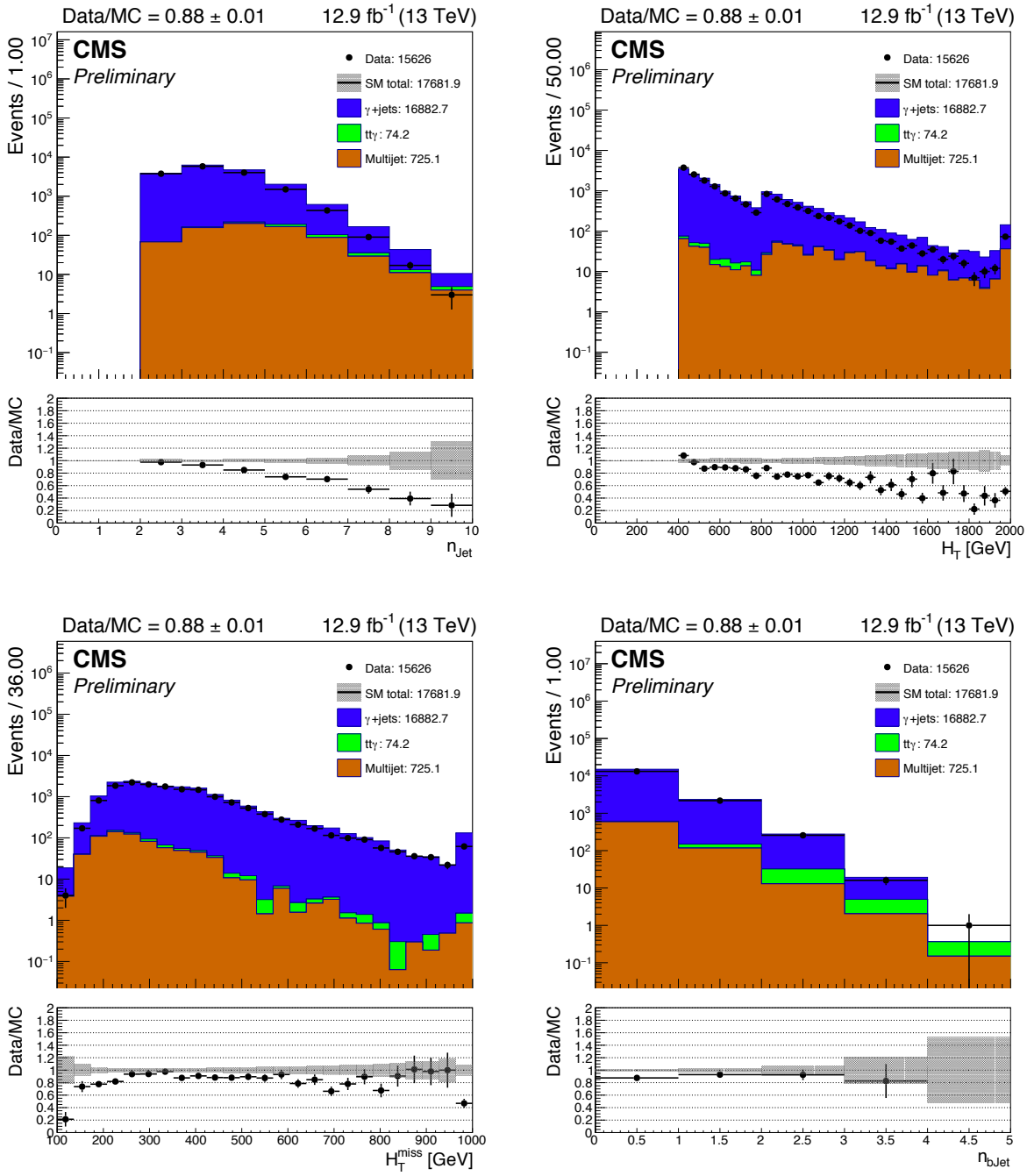


Figure A.10: Key analysis variables for single photon control region (symmetric n_{jet} bins)

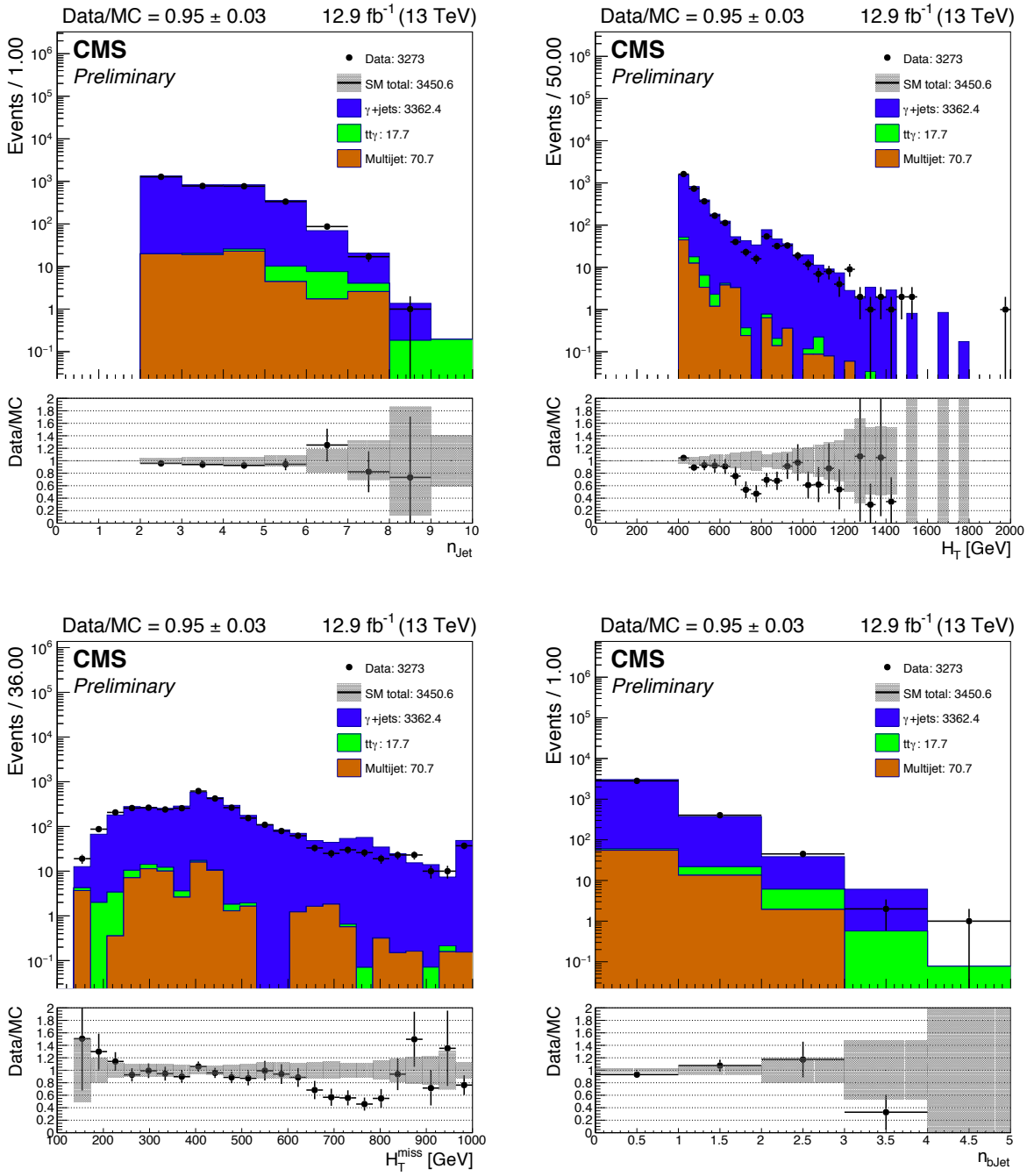


Figure A.11: Key analysis variables for single photon control region (asymmetric n_{jet} bins)

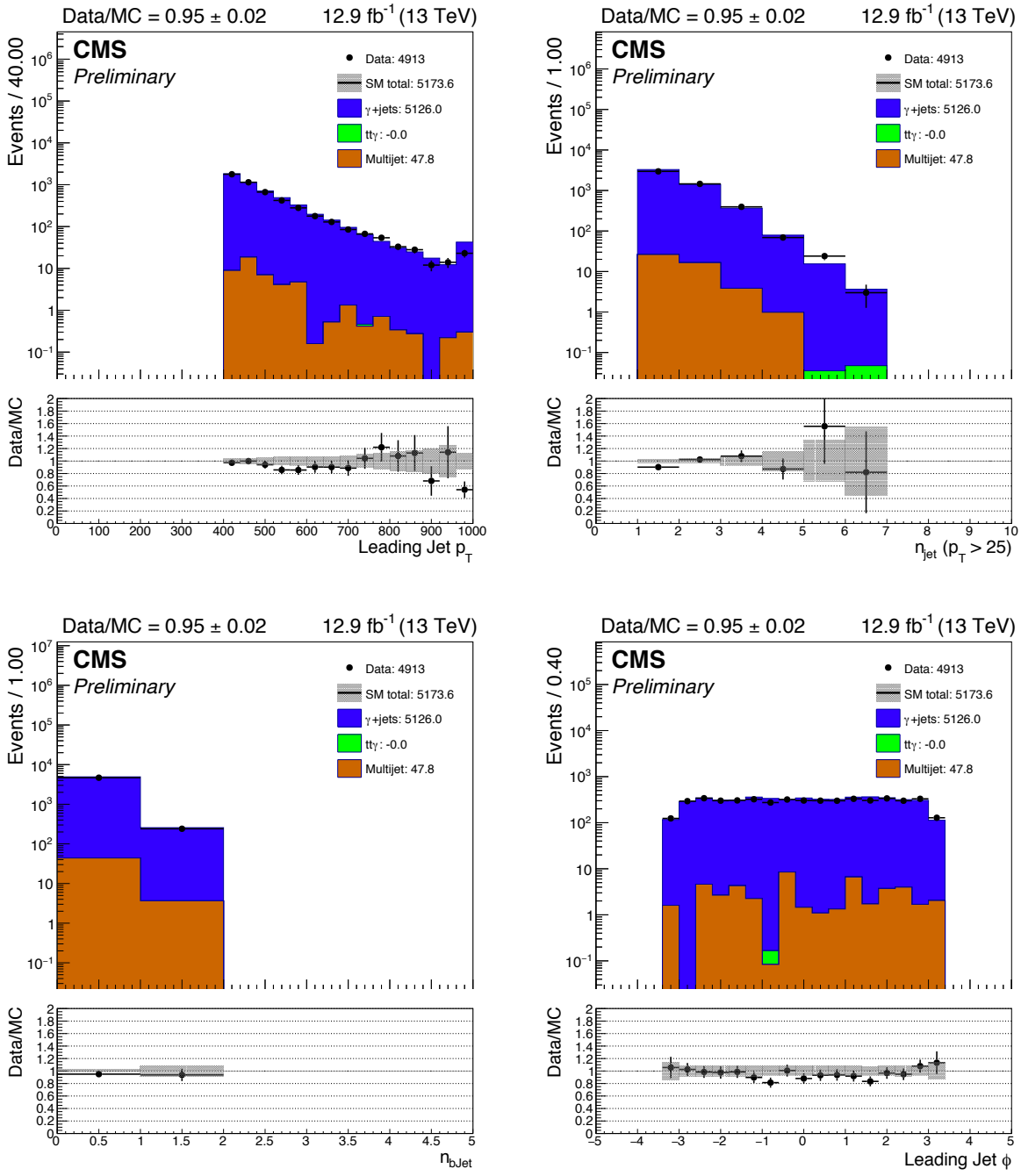


Figure A.12: Key analysis variables for single photon control region (monojet bins)

A.2 Binning of $\not{H}_{\mathrm{T}}$ dimension

In Tab. A.1-A.8 the binning of the $\not{H}_{\mathrm{T}}$ dimension for all the categories used in the analysis is presented, for the asymmetric and symmetric topologies respectively. Where the binning is not specified no template is used. For the monojet category no template is used.

Table A.1: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet} = 2$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}} = 2$, $n_{\mathrm{b}} = 0$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 175, 225, 275, 325, 375, 425, 475
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525, 575
	$600 < H_T < 800$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525, 575, 625, 675, 725, 775
	$H_T > 800$ GeV	130, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} = 2$, $n_{\mathrm{b}} = 1$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 175, 225, 275, 325, 375, 425, 475
	$500 < H_T < 600$ GeV	130, 225, 275, 325, 375, 425, 475, 525, 575
	$600 < H_T < 800$ GeV	130, 225, 275, 325, 375, 425, 475, 525, 575, 675, 725, 775
	$H_T > 800$ GeV	130, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} = 2$, $n_{\mathrm{b}} = 2$	$200 < H_T < 250$ GeV	-
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	-
	$350 < H_T < 400$ GeV	130, 250, 300, 350
	$400 < H_T < 500$ GeV	-
	$500 < H_T < 600$ GeV	130, 225, 325, 375, 475, 525, 575
	$H_T > 600$ GeV	-

Table A.2: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet} = 3$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}} = 3$, $n_{\mathrm{b}} = 0$	$200 < H_T < 250$ GeV	130, 225
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 175, 225, 275, 325, 375, 425, 475
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525, 575
	$600 < H_T < 800$ GeV	130, 225, 275, 325, 375, 425, 475, 525, 575, 625, 675, 725, 775
	$H_T > 800$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} = 3$, $n_{\mathrm{b}} = 1$	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 175, 225, 275, 325, 375, 425, 475
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525, 575
	$600 < H_T < 800$ GeV	130, 225, 275, 325, 375, 425, 475, 525, 575, 625, 675, 725, 775
	$H_T > 800$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} = 3$, $n_{\mathrm{b}} = 2$	$250 < H_T < 300$ GeV	130, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 175, 225, 275, 325, 375, 425
	$500 < H_T < 600$ GeV	130, 200, 250, 300, 425, 475, 525, 575
	$600 < H_T < 800$ GeV	130, 200, 250, 350, 400, 450, 500, 550, 600, 650, 700, 750
	$H_T > 800$ GeV	-
$n_{\mathrm{jet}} = 3$, $n_{\mathrm{b}} \geq 3$	$250 < H_T < 300$ GeV	-
	$350 < H_T < 400$ GeV	-
	$H_T > 400$ GeV	-

Table A.3: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet} = 4$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}} = 4$, $n_{\mathrm{b}} = 0$	$250 < H_T < 300$ GeV	-
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400, 450
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525, 575
	$600 < H_T < 800$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750
	$H_T > 800$ GeV	130, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} = 4$, $n_{\mathrm{b}} = 1$	$250 < H_T < 300$ GeV	-
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 150, 200, 250, 300, 350
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400, 450
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525
	$600 < H_T < 800$ GeV	130, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750
	$H_T > 800$ GeV	130, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} = 4$, $n_{\mathrm{b}} = 2$	$250 < H_T < 300$ GeV	-
	$300 < H_T < 350$ GeV	130, 150, 200, 250, 300
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400, 450
	$500 < H_T < 600$ GeV	130, 200, 250, 300, 350, 400, 450, 500
	$600 < H_T < 800$ GeV	130, 225, 275, 325, 375, 425, 475, 525, 575, 625, 675, 725
	$H_T > 800$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} = 4$, $n_{\mathrm{b}} \geq 3$	$350 < H_T < 400$ GeV	-
	$400 < H_T < 500$ GeV	-
	$500 < H_T < 600$ GeV	130, 300, 400
	$600 < H_T < 800$ GeV	-
	$H_T > 800$ GeV	-

Table A.4: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet} \geq 5$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}} \geq 5$, $n_{\mathrm{b}} = 0$	$350 < H_T < 400$ GeV	130, 150, 200, 250, 300, 350
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400, 450
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525
	$600 < H_T < 800$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750
	$H_T > 800$ GeV	130, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} \geq 5$, $n_{\mathrm{b}} = 1$	$350 < H_T < 400$ GeV	130, 150, 200, 250, 300
	$400 < H_T < 500$ GeV	130, 175, 225, 275, 325, 375, 425
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 475, 525
	$600 < H_T < 800$ GeV	130, 225, 275, 325, 375, 425, 475, 525, 575, 625, 675, 725
	$H_T > 800$ GeV	130, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} \geq 5$, $n_{\mathrm{b}} = 2$	$350 < H_T < 400$ GeV	130, 175, 225, 275
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350
	$500 < H_T < 600$ GeV	130, 200, 250, 300, 350, 400, 450
	$600 < H_T < 800$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650
	$H_T > 800$ GeV	130, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}} \geq 5$, $n_{\mathrm{b}} \geq 3$	$400 < H_T < 500$ GeV	-
	$500 < H_T < 600$ GeV	-
	$600 < H_T < 800$ GeV	130, 225, 275, 325, 375, 425, 475, 525, 575
	$H_T > 800$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550, 650, 700, 800

Table A.5: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet}^{asym} = 2$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}}^{\mathrm{asy}} = 2\,, n_{\mathrm{b}} = 0$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 225, 275
	$300 < H_T < 350$ GeV	130, 275, 325
	$350 < H_T < 400$ GeV	130, 325, 375
	$400 < H_T < 500$ GeV	130, 375, 425, 475
	$500 < H_T < 600$ GeV	130, 525, 575
	$H_T > 600$ GeV	130, 600, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} = 2\,, n_{\mathrm{b}} = 1$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 225, 275
	$300 < H_T < 350$ GeV	130, 275, 325
	$350 < H_T < 400$ GeV	130, 375
	$400 < H_T < 500$ GeV	130, 425, 475
	$H_T > 500$ GeV	-
$n_{\mathrm{jet}}^{\mathrm{asy}} = 2\,, n_{\mathrm{b}} = 2$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 225, 275
	$300 < H_T < 350$ GeV	130, 325
	$350 < H_T < 400$ GeV	-
	$H_T > 400$ GeV	-

Table A.6: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet}^{asym} = 3$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}}^{\mathrm{asy}} = 3\,, n_{\mathrm{b}} = 0$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 225, 275, 325, 375, 425, 475
	$500 < H_T < 600$ GeV	130, 425, 475, 525, 575
	$H_T > 600$ GeV	130, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} = 3\,, n_{\mathrm{b}} = 1$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 225, 275, 325, 375, 425, 475
	$500 < H_T < 600$ GeV	130, 450, 500, 550
	$H_T > 600$ GeV	130, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} = 3\,, n_{\mathrm{b}} = 2$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 250, 300, 350
	$400 < H_T < 500$ GeV	130, 275, 325, 375, 425
	$H_T > 500$ GeV	130, 575, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} = 3\,, n_{\mathrm{b}} \geq 3$	$200 < H_T < 250$ GeV	-
	$250 < H_T < 300$ GeV	-
	$H_T > 300$ GeV	-

Table A.7: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet}^{asym} = 4$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}}^{\mathrm{asy}} = 4\,, n_{\mathrm{b}} = 0$	$200 < H_T < 250$ GeV	130, 175, 225
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325, 375
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400, 450
	$500 < H_T < 600$ GeV	130, 250, 300, 350, 400, 450, 500, 550
	$H_T > 600$ GeV	130, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} = 4\,, n_{\mathrm{b}} = 1$	$200 < H_T < 250$ GeV	-
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 175, 225, 275, 325
	$350 < H_T < 400$ GeV	130, 150, 200, 250, 300, 350
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400, 450
	$500 < H_T < 600$ GeV	130, 250, 300, 350, 400, 450, 500, 550
	$H_T > 600$ GeV	130, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} = 4\,, n_{\mathrm{b}} = 2$	$200 < H_T < 250$ GeV	-
	$250 < H_T < 300$ GeV	130, 175, 225, 275
	$300 < H_T < 350$ GeV	130, 150, 200, 250, 300
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400
	$500 < H_T < 600$ GeV	130, 325, 375, 425, 475, 525
	$H_T > 600$ GeV	-
$n_{\mathrm{jet}}^{\mathrm{asy}} = 4\,, n_{\mathrm{b}} \geq 3$	$250 < H_T < 300$ GeV	-
	$300 < H_T < 350$ GeV	-
	$350 < H_T < 400$ GeV	130, 175, 225, 275
	$H_T > 400$ GeV	-

Table A.8: The $\not{H}_{\mathrm{T}}$ binning for the jet category $n_{jet}^{asym} \geq 5$.

Jet category	H_T bin	$\not{H}_{\mathrm{T}}$ template binning (GeV)
$n_{\mathrm{jet}}^{\mathrm{asy}} \geq 5$, $n_{\mathrm{b}} = 0$	$250 < H_T < 300$ GeV	-
	$300 < H_T < 350$ GeV	130, 150, 200, 250, 300
	$350 < H_T < 400$ GeV	130, 150, 200, 250, 300, 350
	$400 < H_T < 500$ GeV	130, 175, 225, 275, 325, 375, 425
	$500 < H_T < 600$ GeV	130, 200, 250, 300, 350, 400, 450, 500, 550
	$H_T > 600$ GeV	130, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} \geq 5$, $n_{\mathrm{b}} = 1$	$250 < H_T < 300$ GeV	-
	$300 < H_T < 350$ GeV	130, 150, 200, 250, 300
	$350 < H_T < 400$ GeV	130, 175, 225, 275, 325
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350, 400
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425, 500
	$H_T > 600$ GeV	130, 225, 275, 325, 375, 450, 500, 550, 600, 650, 700, 750, 800
$n_{\mathrm{jet}}^{\mathrm{asy}} \geq 5$, $n_{\mathrm{b}} = 2$	$250 < H_T < 300$ GeV	-
	$300 < H_T < 350$ GeV	130, 150, 200, 250
	$350 < H_T < 400$ GeV	130, 150, 200, 250, 300
	$400 < H_T < 500$ GeV	130, 150, 200, 250, 300, 350
	$500 < H_T < 600$ GeV	130, 175, 225, 275, 325, 375, 425
	$H_T > 600$ GeV	130, 275, 325, 375, 450, 500, 600, 700
$n_{\mathrm{jet}}^{\mathrm{asy}} \geq 5$, $n_{\mathrm{b}} \geq 3$	$300 < H_T < 350$ GeV	-
	$350 < H_T < 400$ GeV	130, 150, 200, 250
	$400 < H_T < 500$ GeV	-
	$H_T > 500$ GeV	-

A.3 Bin labels key

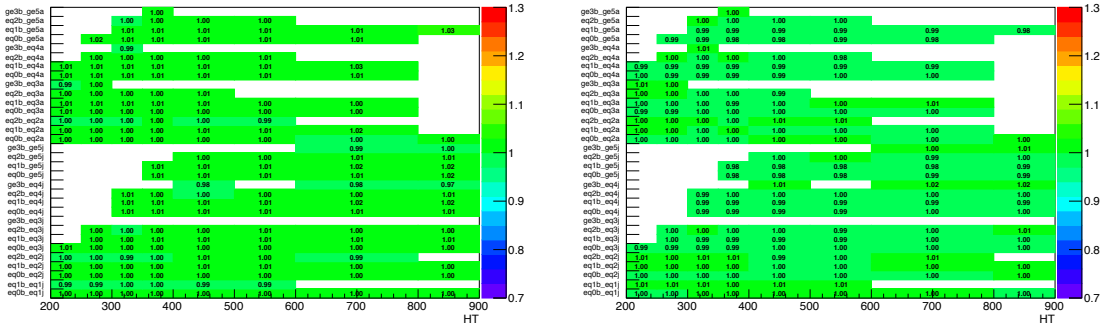
The n_{jet} categories are labelled with four letter strings that indicate the number of jets and the topology, *a* for asymmetric and *j* for symmetric. The n_{b} categories contain the letter *b*. The number of jets is represented by the number and is prefixed by either *eq* corresponding to $=$, or *ge* corresponding to $\geq$.

The H_{T} bins are labelled based on their lower bin edge in GeV. This bin extends up to the next H_{T} bin, with the exception of 800 which is open ended for the symmetric category or 600 which is open ended for the asymmetric and mono-jet categories.

As an example, *HT250 eq4j* would correspond to the $250 < H_{\text{T}} < 300$ GeV bin with $n_{\text{jet}} = 4$ and a symmetric topology. Alternatively, *HT600 ge5a* would correspond to the $600 < H_{\text{T}} < \infty$ GeV bin with $n_{\text{jet}} \geq 5$ and an asymmetric topology.

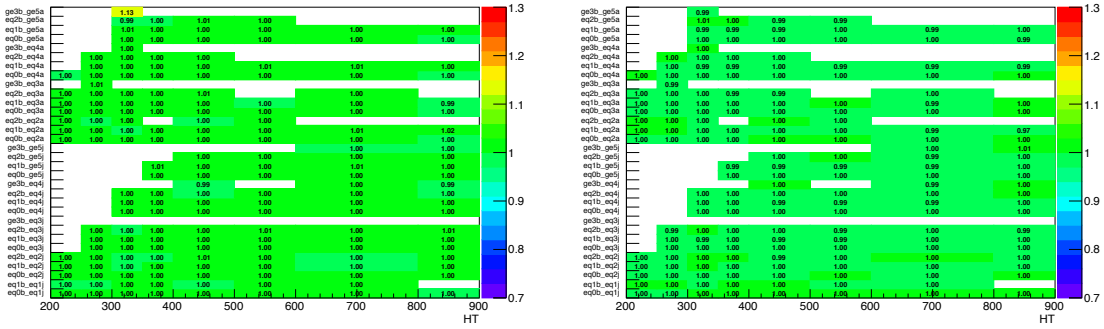
A.4 Variation in transfer factors from known systematic uncertainties

The variations of the transfer factors after variations of known sources of systematic uncertainty, as discussed in Sec. 7.5.1. These are shown for all relevant transfer factors from the γ + jets, μ + jets and $\mu\mu$ + jets control samples. The plots are labelled as described in the key in Appendix A.3.



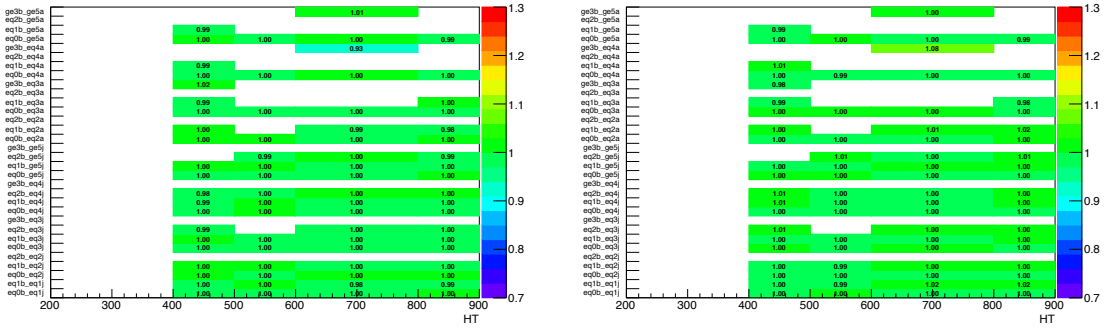
(a) b-tag SF (heavy) up variation **(b)** b-tag SF (heavy) down variation

Figure A.13: The relative change in the $\mu + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying b-tag SF for heavy jets in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



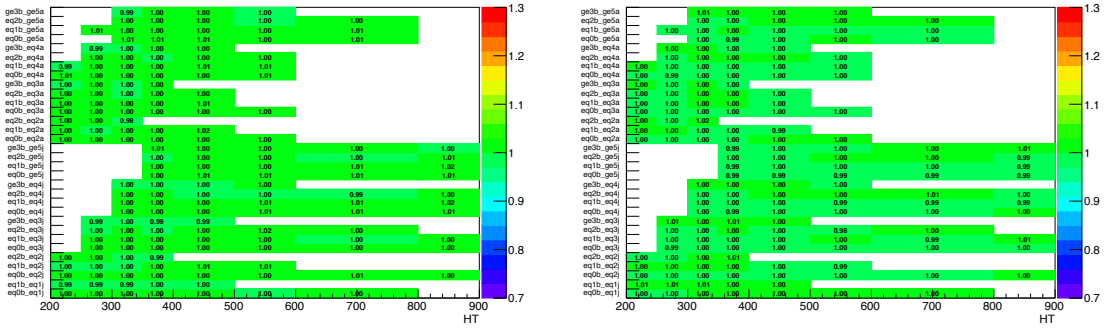
(a) b-tag SF (heavy) up variation **(b)** b-tag SF (heavy) down variation

Figure A.14: The relative change in the $\mu\mu + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying b-tag SF for heavy jets in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



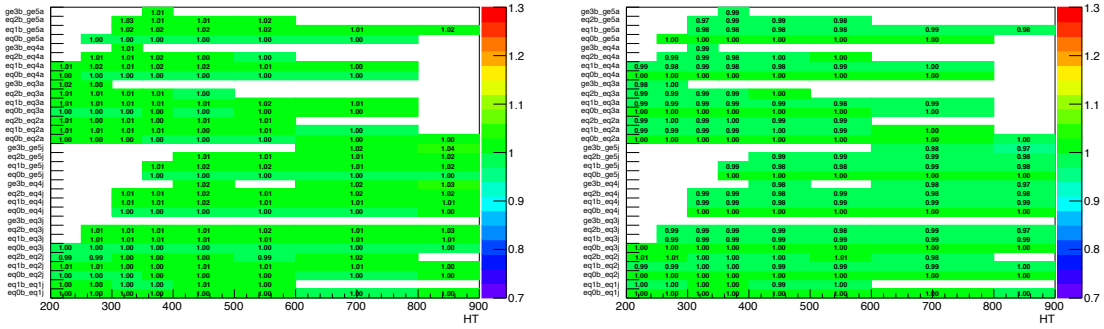
(a) b-tag SF (heavy) up variation **(b)** b-tag SF (heavy) down variation

Figure A.15: The relative change in the $\gamma + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying b-tag SF for heavy jets in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



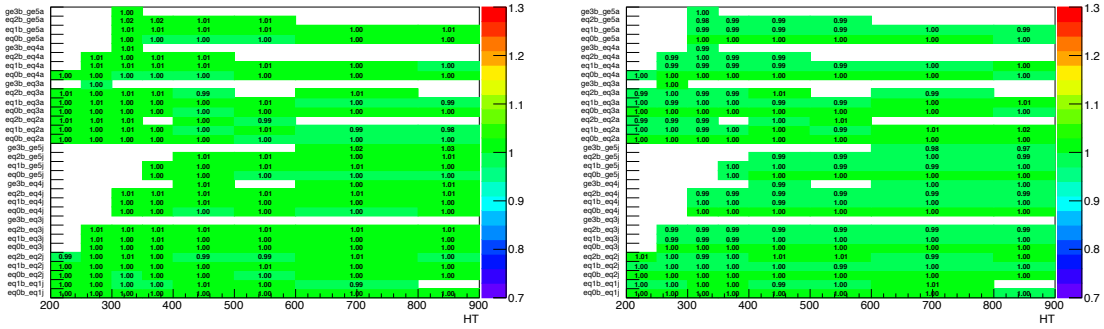
(a) b-tag SF (heavy) up variation **(b)** b-tag SF (heavy) down variation

Figure A.16: The relative change in the $\mu + jets \rightarrow \mathrm{tt} + \mathrm{W}$ transfer factors when varying b-tag SF for heavy jets in MC within its uncertainties, as a function of H_{T} and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



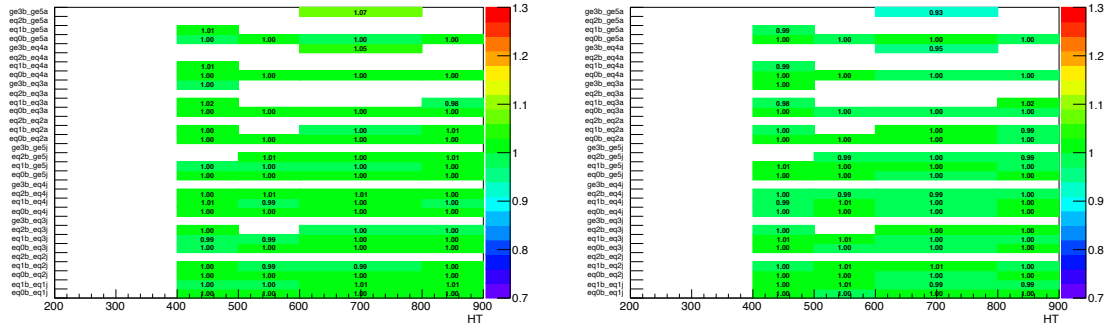
(a) b-tag SF (light) up variation **(b)** b-tag SF (light) down variation

Figure A.17: The relative change in the $\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ transfer factors when varying b-tag SF for light jets in MC within its uncertainties, as a function of H_{T} and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



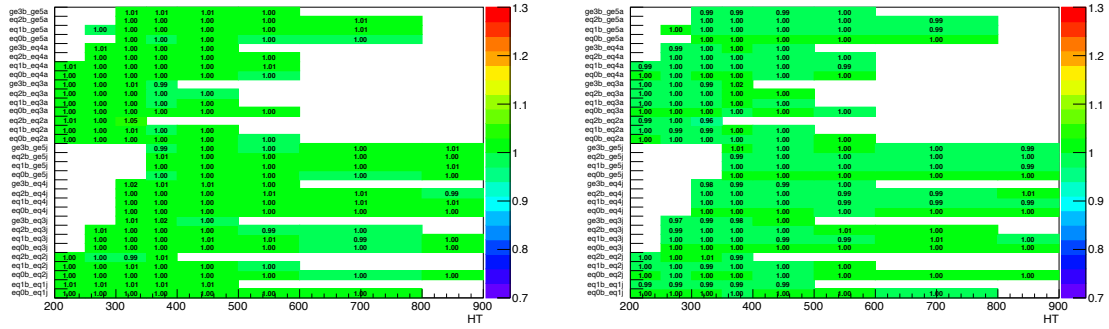
(a) b-tag SF (light) up variation **(b)** b-tag SF (light) down variation

Figure A.18: The relative change in the $\mu\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ transfer factors when varying b-tag SF for light jets in MC within its uncertainties, as a function of H_{T} and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



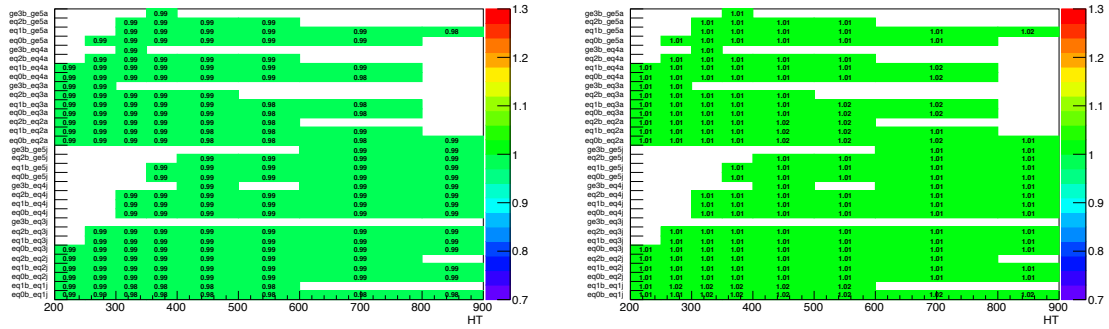
(a) b-tag SF (light) up variation **(b)** b-tag SF (light) down variation

Figure A.19: The relative change in the $\gamma + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying b-tag SF for light jets in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



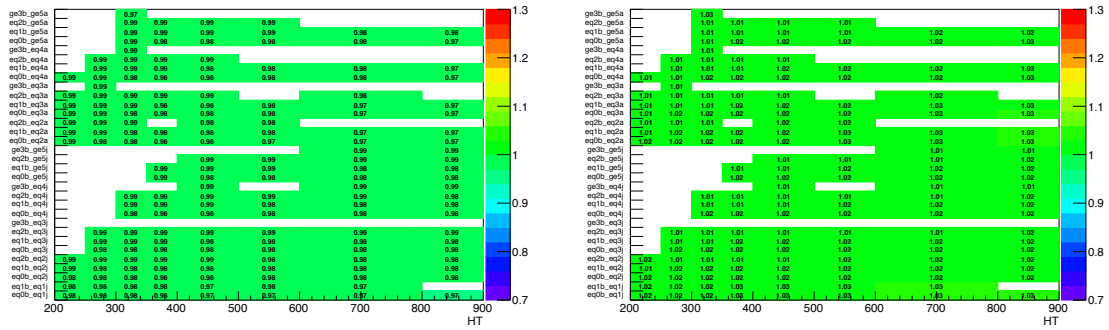
(a) b-tag SF (light) up variation **(b)** b-tag SF (light) down variation

Figure A.20: The relative change in the $\mu + jets \rightarrow \mathrm{tt} + \mathrm{W}$ transfer factors when varying b-tag SF for light jets in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



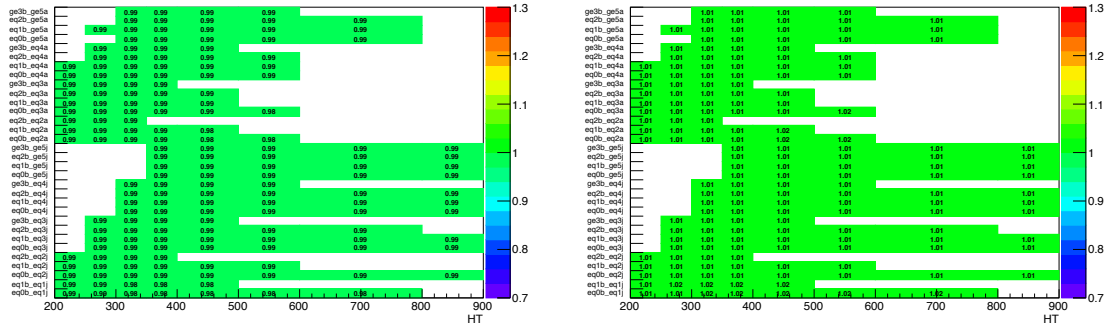
(a) muon scale factor up variation **(b)** muon scale factor down variation

Figure A.21: The relative change in the $\mu + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying muon scale factor in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



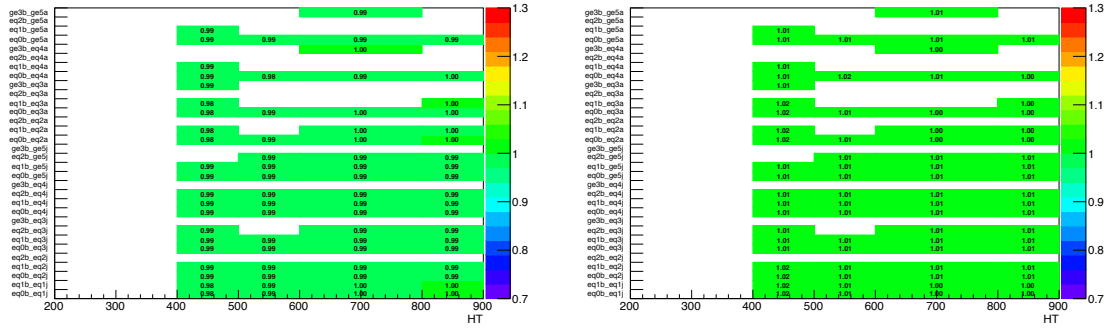
(a) muon scale factor up variation **(b)** muon scale factor down variation

Figure A.22: The relative change in the $\mu\mu + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying muon scale factor in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



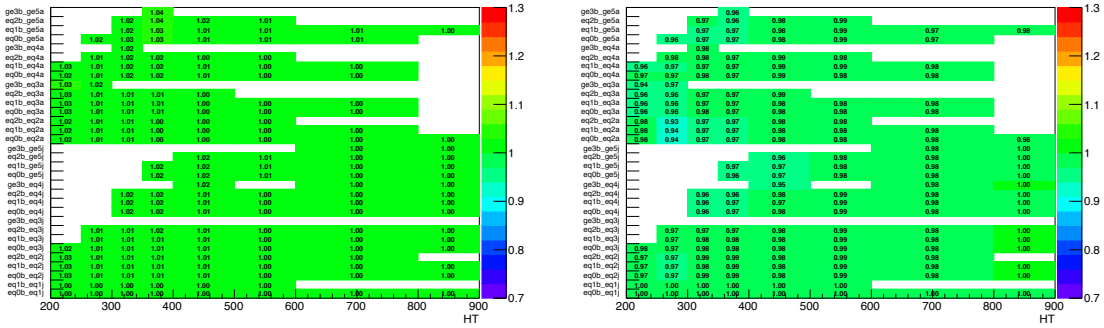
(a) muon scale factor up variation **(b)** muon scale factor down variation

Figure A.23: The relative change in the $\mu + jets \rightarrow \mathrm{tt} + \mathrm{W}$ transfer factors when varying muon scale factor in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



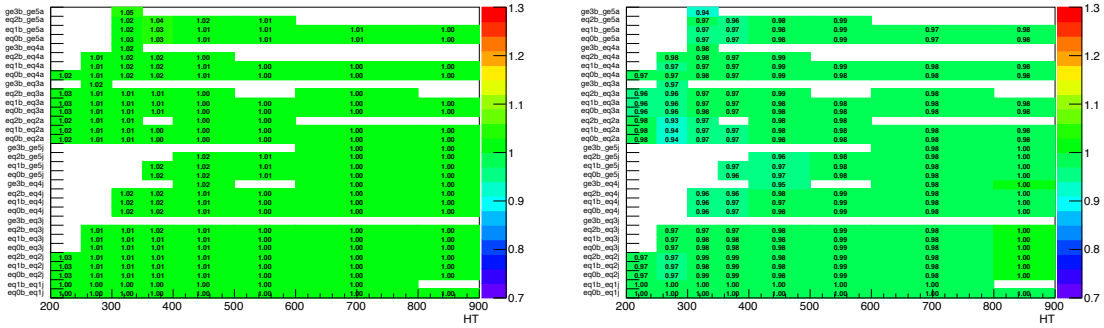
(a) Photon trigger weight up variation **(b)** Photon trigger weight down variation

Figure A.24: The relative change in the $\gamma + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying photon trigger weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



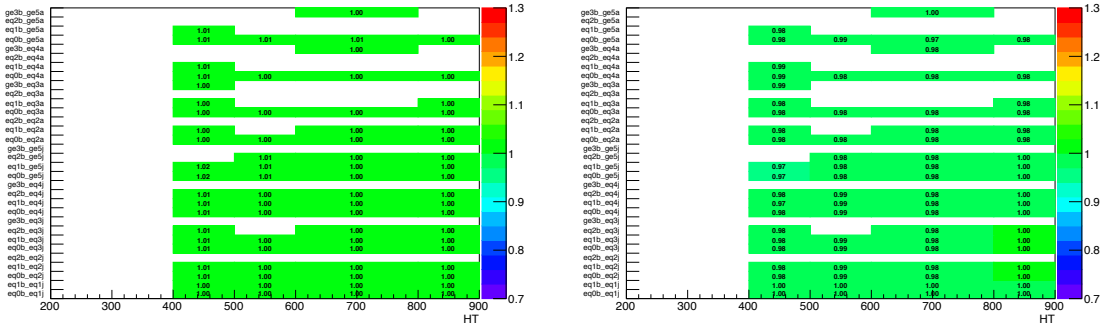
(a) trigger weight up variation **(b)** trigger weight down variation

Figure A.25: The relative change in the $\mu + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying trigger weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



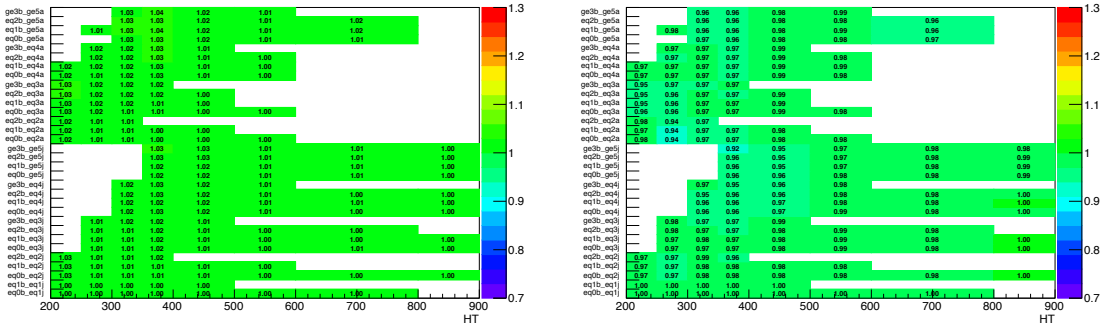
(a) trigger weight up variation **(b)** trigger weight down variation

Figure A.26: The relative change in the $\mu\mu + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying trigger weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



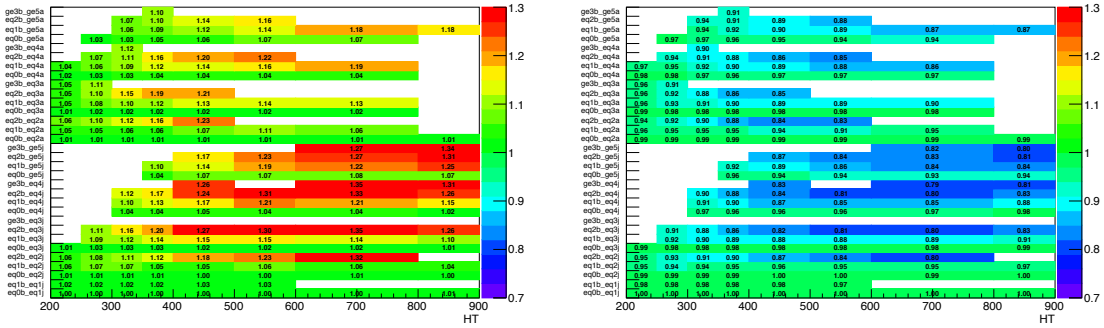
(a) trigger weight up variation **(b)** trigger weight down variation

Figure A.27: The relative change in the $\gamma + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying trigger weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



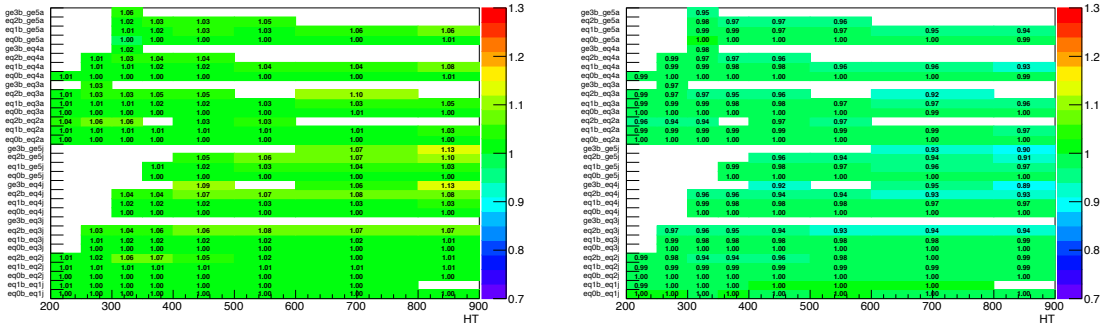
(a) trigger weight up variation **(b)** trigger weight down variation

Figure A.28: The relative change in the $\mu + jets \rightarrow \mathrm{tt} + \mathrm{W}$ transfer factors when varying trigger weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



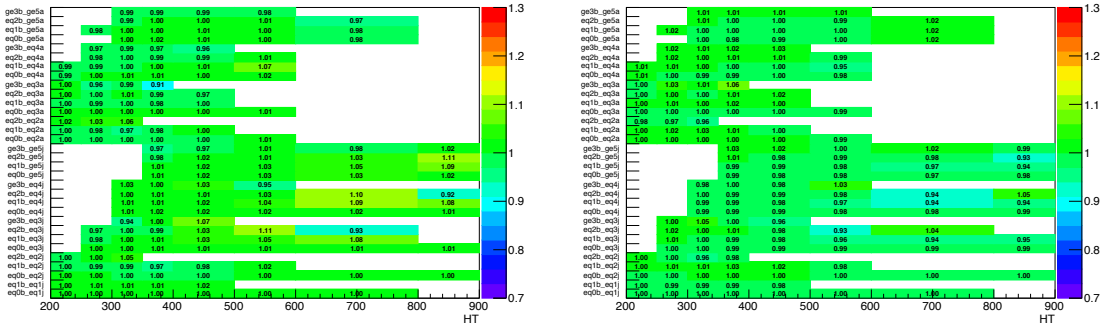
(a) top p_T weight up variation **(b)** top p_T weight down variation

Figure A.29: The relative change in the $\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ transfer factors when varying top p_T weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



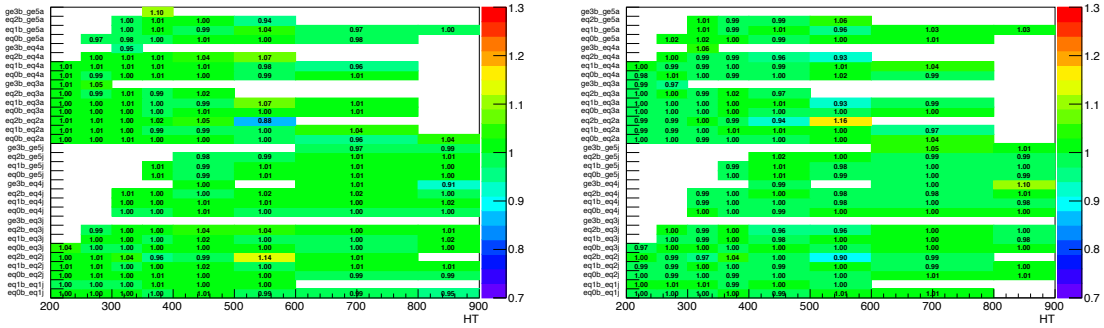
(a) top p_T weight up variation **(b)** top p_T weight down variation

Figure A.30: The relative change in the $\mu\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ transfer factors when varying top p_T weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



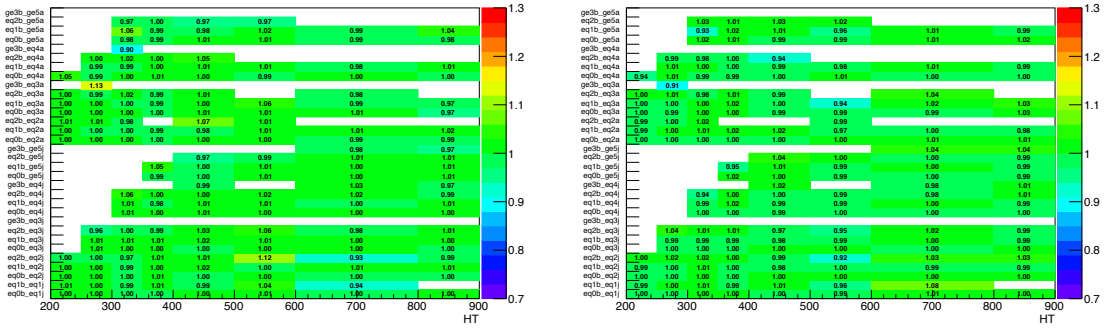
(a) top p_T weight up variation **(b)** top p_T weight down variation

Figure A.31: The relative change in the $\mu + jets \rightarrow \mathrm{tt} + \mathrm{W}$ transfer factors when varying top p_T weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



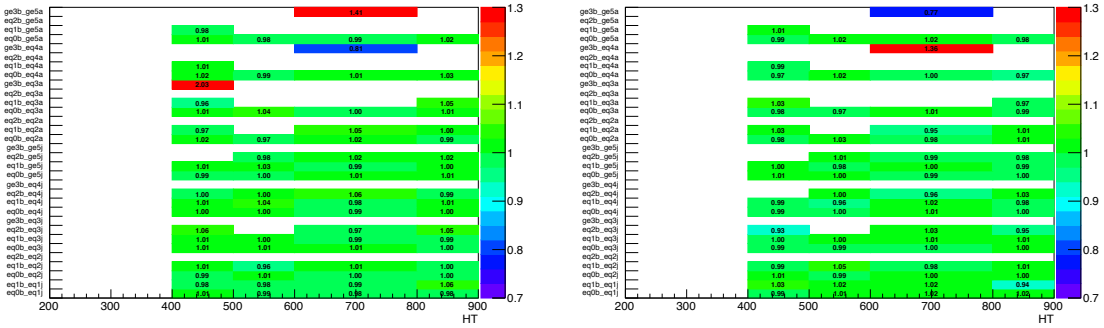
(a) PU weight up variation **(b)** PU weight down variation

Figure A.32: The relative change in the $\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ transfer factors when varying PU weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



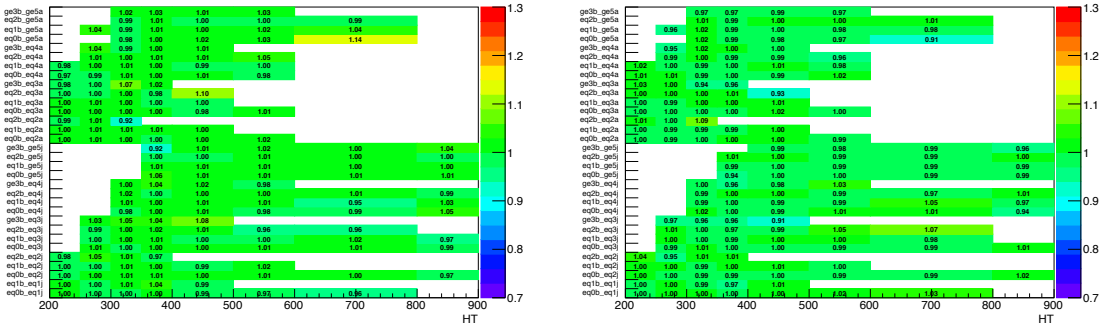
(a) PU weight up variation **(b)** PU weight down variation

Figure A.33: The relative change in the $\mu\mu + jets \rightarrow (\mathrm{Z} \rightarrow \nu\bar{\nu})$ transfer factors when varying PU weight in MC within its uncertainties, as a function of H_T and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



(a) PU weight up variation **(b)** PU weight down variation

Figure A.34: The relative change in the $\gamma + jets \rightarrow (Z \rightarrow \nu\bar{\nu})$ transfer factors when varying PU weight in MC within its uncertainties, as a function of H_{T} and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.



(a) PU weight up variation **(b)** PU weight down variation

Figure A.35: The relative change in the $\mu + jets \rightarrow \mathrm{tt} + \mathrm{W}$ transfer factors when varying PU weight in MC within its uncertainties, as a function of H_{T} and jet category. Variations corresponding to $+1\sigma$ (-1σ) are shown in the left (right) figure.

Bibliography

[1] A. Einstein and M. Grossmann, "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie", *Zeitschrift für Mathematik und Physik* **63** (1914).

[2] C. M. Will, "The Confrontation between General Relativity and Experiment", *Living Rev. Rel.* **17** (2014) 4, doi:10.12942/lrr-2014-4, arXiv:1403.7377.

[3] A. Salam and J. C. Ward, "Electromagnetic and weak interactions", *Phys. Lett.* **13** (1964) 168–171, doi:10.1016/0031-9163(64)90711-5.

[4] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

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

[6] S. Weinberg, "Ultraviolet divergences in quantum theories of gravitation", in *General Relativity: An Einstein Centenary Survey*, pp. 790–831. 1980.

[7] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[8] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[9] 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.

[10] 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.

[11] J. C. Kapteyn, "First Attempt at a Theory of the Arrangement and Motion of the Sidereal System", *Astrophys. J.* **55** (1922) 302–328, doi:10.1086/142670.

[12] J. H. Oort, "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems", *Bull. Astron. Inst. Netherlands* **6** (1932) 249–287.

[13] M. Markevitch, A. H. Gonzalez, D. Clowe et al., "Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56", *Astrophys. J.* **606** (2004) 819–824, doi:10.1086/383178, arXiv:astro-ph/0309303.

[14] J. P. Dietrich, N. Werner, D. Clowe et al., "A filament of dark matter between two clusters of galaxies", *Nature* **487** (July, 2012) 202–204, doi:10.1038/nature11224, arXiv:1207.0809.

[15] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters", *Astron. Astrophys.* **594** (2016) A13, doi:10.1051/0004-6361/201525830, arXiv:1502.01589.

[16] G. Hinshaw et al., "Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results", *The Astrophysical Journal Supplement Series* **180** (2009), no. 2, 225.

[17] S. Weinberg, "The Cosmological Constant Problem", *Rev. Mod. Phys.* **61** (1989) 1–23, doi:10.1103/RevModPhys.61.1.

[18] Supernova Search Team Collaboration, "Observational evidence from supernovae for an accelerating universe and a cosmological constant", *Astron. J.* **116** (1998) 1009–1038, doi:10.1086/300499, arXiv:astro-ph/9805201.

[19] S. P. Martin, "A Supersymmetry primer", doi:10.1142/9789812839657_0001, arXiv:hep-ph/9709356. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[20] D. J. Griffiths, "Introduction to elementary particles; 2nd rev. version". Physics textbook. Wiley, New York, NY, 2008.

[21] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C40** (2016), no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.

[22] SNO Collaboration, "Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino

Observatory", *Phys. Rev. Lett.* **87** (Jul, 2001) 071301, doi:10.1103/PhysRevLett.87.071301.

[23] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.

[24] E. Noether, "Invariant variation problems", *Transport Theory and Statistical Physics* **1** (January, 1971) 186–207, doi:10.1080/00411457108231446, arXiv:physics/0503066.

[25] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (Jun, 1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[26] LHCb Collaboration, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.* **115** (2015) 072001, doi:10.1103/PhysRevLett.115.072001, arXiv:1507.03414.

[27] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.

[28] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[29] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132, doi:10.1016/0031-9163(64)91136-9.

[30] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[31] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (1964) 585, doi:10.1103/PhysRevLett.13.585.

[32] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156.

[33] T. W. B. Kibble, "Symmetry breaking in non-Abelian gauge theories", *Phys. Rev.* **155** (1967) 1554, doi:10.1103/PhysRev.155.1554.

[34] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the Unified Model of Elementary Particles", *Progress of Theoretical Physics* **28** (1962), no. 5, 870, doi:10.1143/PTP.28.870.

[35] G. Jungman, M. Kamionkowski, and K. Griest, "Supersymmetric dark matter", *Phys. Rept.* **267** (1996) 195–373, doi:10.1016/0370-1573(95)00058-5, arXiv:hep-ph/9506380.

[36] M. H. Poincaré, "Sur la dynamique de l'électron", *Rendiconti del Circolo Matematico di Palermo (1884-1940)* **21** (1906), no. 1, 129–175, doi:10.1007/BF03013466.

[37] S. Coleman and J. Mandula, "All Possible Symmetries of the S Matrix", *Phys. Rev.* **159** (Jul, 1967) 1251–1256, doi:10.1103/PhysRev.159.1251.

[38] NOBELPRIZE.ORG, "Running coupling constants", 2017. https://www.nobelprize.org/nobel_prizes/physics/laureates/2004/phypub4highen.jpg.

[39] C. Csaki, "The Minimal supersymmetric standard model (MSSM)", *Mod. Phys. Lett.* **A11** (1996) 599, doi:10.1142/S021773239600062X, arXiv:hep-ph/9606414.

[40] Super-Kamiokande Collaboration Collaboration, "Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ in 0.31 megaton $\cdot$ years exposure of the Super-Kamiokande water Cherenkov detector", *Phys. Rev. D* **95** (Jan, 2017) 012004, doi:10.1103/PhysRevD.95.012004.

[41] G. R. Farrar and P. Fayet, "Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry", *Phys. Lett.* **B76** (1978) 575–579, doi:10.1016/0370-2693(78)90858-4.

[42] A. Ali, "New aspects of high-energy proton-proton collisions". Springer, 1988.

[43] CMS Collaboration, "Summary of comparison plots in simplified models spectra for the 8 TeV dataset", 2016. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SUSYSMSSummaryPlots8TeV.

[44] J. Alwall, P. Schuster, and N. Toro, "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.

[45] 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.

[46] H1 Collaboration, "A Search for selectrons and squarks at HERA", *Phys. Lett.* **B380** (1996) 461–470, doi:10.1016/0370-2693(96)00640-5, arXiv:hep-ex/9605002.

[47] J. Butterworth and H. K. Dreiner, "R-parity violation at HERA", *Nucl. Phys.* **B397** (1993) 3–34, doi:10.1016/0550-3213(93)90334-L, arXiv:hep-ph/9211204.

[48] S. Braibant, "SUSY searches at LEP", in *Proceedings, 38th Rencontres de Moriond on QCD and High-Energy Hadronic Interactions: Les Arcs, France, March 22-29, 2003*. 2003. arXiv:hep-ex/0305058.

[49] CDF, D0 Collaboration, "SUSY searches at the Tevatron", *EPJ Web Conf.* **28** (2012) 09006, doi:10.1051/epjconf/20122809006, arXiv:1202.0712.

[50] O. S. Bruning, P. Collier, P. Lebrun et al., "LHC Design Report Vol.1: The LHC Main Ring", Technical Report CERN-2004-003-V1, CERN-2004-003, CERN-2004-003-V-1, (2004).

[51] O. Buning, P. Collier, P. Lebrun et al., "LHC Design Report. 2. The LHC infrastructure and general services", Technical Report CERN-2004-003-V-2, CERN-2004-003, (2004).

[52] M. Benedikt, P. Collier, V. Mertens et al., "LHC Design Report. 3. The LHC injector chain", Technical Report CERN-2004-003-V-3, CERN-2004-003, (2004).

[53] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[54] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[55] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[56] STFC, "CERN Accelerator Complex", 2016. http://www.stfc.ac.uk/research/particle-physics-and-particle-astrophysics/large-hadron-collider/cern-accelerator-complex/.

[57] W. J. Stirling. private communication. http://www.hep.ph.ic.ac.uk/ ∼ wstirlin/plots/plots.html.

[58] CMS Collaboration, "CMS physics: Technical design report", Technical Report CERN-LHCC-2006-001, CMS-TDR-008-1, (2006).

[59] CMS Collaboration, V. Karimäki, M. Mannelli, P. Siegrist et al., "The CMS tracker system project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[60] CMS Collaboration, "CMS: The electromagnetic calorimeter. Technical design report", Technical Report CERN-LHCC-97-33, CMS-TDR-4, (1997).

[61] P. Adzic, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *Journal of Instrumentation* **2** (2007), no. 04, P04004.

[62] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[63] CMS HCAL Collaboration, "Design, performance, and calibration of CMS hadron-barrel calorimeter wedges", *Eur. Phys. J.* **C55** (2008) 159–171, doi:10.1140/epjc/s10052-008-0573-y.

[64] CMS Collaboration, "CMS, the Compact Muon Solenoid. Muon technical design report", Technical Report CERN-LHCC-97-32, (1997).

[65] CMS Collaboration, "CMS reconstruction improvement for the muon tracking by the RPC chambers", *PoS* **RPC2012** (2012) 045, doi:10.1088/1748-0221/8/03/T03001, arXiv:1209.2646. [JINST8,T03001(2013)].

[66] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", Technical Report CERN-LHCC-2013-011. CMS-TDR-12, (Jun, 2013).

[67] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems",. https://cds.cern.ch/record/706847.

[68] CMS Collaboration, G. L. Bayatyan, M. Della Negra, A. Herve et al., "CMS computing: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 2005. Submitted on 31 May 2005.

[69] CMS Collaboration, S. Cittolin, A. Rácz, and P. Sphicas, "CMS The TriDAS

Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project". Technical Design Report CMS. CERN, Geneva, 2002.

[70] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[71] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth.* **A262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.

[72] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239, doi:10.1109/5.726788.

[73] W. Waltenberger, "Adaptive vertex reconstruction", Technical Report CERN-CMS-NOTE-2008-033, (2008).

[74] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, (Apr, 2009).

[75] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", Technical Report CMS-PAS-PFT-10-001, 2010.

[76] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", Technical Report CMS-PAS-PFT-10-002, CERN, Geneva, (2010).

[77] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Journal of Instrumentation* **10** (2015), no. 06, P06005.

[78] W. Adam, R. Frühwirth, A. Strandlie et al., "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", Technical Report CMS-NOTE-2005-001, CERN, Geneva, (Jan, 2005).

[79] T. C. Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *Journal of Instrumentation* **7** (2012), no. 10, P10002.

[80] G. P. Salam, "Towards jetography", *The European Physical Journal C* **67** (2010),

no. 3, 637–686, doi:10.1140/epjc/s10052-010-1314-6.

[81] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", *Journal of High Energy Physics* **2008** (2008), no. 04, 063.

[82] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *The European Physical Journal C* **72** (2012), no. 3, 1896, doi:10.1140/epjc/s10052-012-1896-2.

[83] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[84] CMS Collaboration Collaboration, "Jet algorithms performance in 13 TeV data", Technical Report CMS-PAS-JME-16-003, CERN, Geneva, (2017).

[85] CMS Collaboration, "Jet performance in CMS", *PoS* **EPS-HEP2013** (2013) 433.

[86] CMS Collaboration, "Jet energy scale and resolution performances with 13 TeV data", Technical Report CMS-DP-2016-020, (Jun, 2016).

[87] T. C. Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation* **6** (2011), no. 11, P11002.

[88] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett.* **B659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[89] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **86** (Jul, 2012) 010001, doi:10.1103/PhysRevD.86.010001.

[90] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, (2016).

[91] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *Journal of Instrumentation* **10** (2015), no. 02, P02006.

[92] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[93] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[94] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[95] T. Sjöstrand, S. Ask, J. R. Christiansen et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[96] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[97] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[98] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[99] G. Hall, D. Newbold, M. Pesaresi et al., "A time-multiplexed track-trigger architecture for CMS", *Journal of Instrumentation* **9** (2014), no. 10, C10034.

[100] R. Frazier, S. Fayer, G. Hall et al., "A demonstration of a Time Multiplexed Trigger for the CMS experiment", *Journal of Instrumentation* **7** (2012), no. 01, C01060.

[101] CMS Collaboration, "Technical Proposal for the Upgrade of the CMS detector through 2020", *CERN-LHCC-2011-006* (2011).

[102] J.-B. Sauvan, "Performance and upgrade of the CMS electromagnetic calorimeter trigger for Run II", Technical Report CMS-CR-2014-068, CERN, (May, 2014).

[103] L. Mastrolorenzo, "The CMS Level-1 Tau algorithm for the LHC Run II", in *ICHEP 2014: 37th International Conference on High Energy Physics, 2-9 Jul 2014, Valencia (Spain)*. 2014.

[104] M. Cacciari, J. Rojo, G. P. Salam et al., "Jet reconstruction in heavy ion collisions", *The European Physical Journal C* **71** (2011), no. 1, 1539, doi:10.1140/epjc/s10052-011-1539-z.

[105] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.

[106] CMS Collaboration, "Pileup Jet Identification", Technical Report CMS-PAS-JME-13-005, CERN, Geneva, (2013).

[107] M. Cacciari, J. Rojo, G. P. Salam et al., "Jet Reconstruction in Heavy Ion Collisions", *Eur. Phys. J.* **C71** (2011) 1539, doi:10.1140/epjc/s10052-011-1539-z, arXiv:1010.1759.

[108] J. Brooke, B. Mathias, A. Tapper et al., "Calibration and Performance of the Jets and Energy Sums in the Level-1 Trigger", Technical Report CMS-IN-2013/006, (2013).

[109] CMS Collaboration, "Search for Supersymmetry at the LHC in Events with Jets and Missing Transverse Energy", *Phys. Rev. Lett.* **107** (2011) 221804, doi:10.1103/PhysRevLett.107.221804, arXiv:1109.2352.

[110] CMS Collaboration, "Search for Supersymmetry in pp Collisions at 7 TeV in Events with Jets and Missing Transverse Energy", *Phys. Lett.* **B698** (2011) 196–218, doi:10.1016/j.physletb.2011.03.021, arXiv:1101.1628.

[111] 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 alphaT", *JHEP* **01** (2013) 077, doi:10.1007/JHEP01(2013)077, arXiv:1210.8115.

[112] 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 $\sqrt{s} = 8$ TeV", *Eur. Phys. J.* **C73** (2013), no. 9, 2568, doi:10.1140/epjc/s10052-013-2568-6, arXiv:1303.2985.

[113] CMS Collaboration, "Search for top squark pair production in compressed-mass-spectrum scenarios in proton-proton collisions at sqrt(s) = 8 TeV using the alphaT variable", *Phys. Lett. B* (2016) doi:10.1016/j.physletb.2017.02.007, arXiv:1605.08993. doi: 10.1016/j.physletb.2017.02.007.

[114] CMS Collaboration, "Search for new physics in final states with jets and missing transverse momentum in $\sqrt{s}$ = 13 TeV pp collisions with the α_{T} variable", *Accepted by: Eur. Phys. J. C.* (2015). https://arxiv.org/abs/1611.00338.

[115] L. Randall and D. Tucker-Smith, "Dijet Searches for Supersymmetry at the LHC", *Phys. Rev. Lett.* **101** (2008) 221803, doi:10.1103/PhysRevLett.101.221803, arXiv:0806.1049.

[116] CMS Collaboration, "SUSY searches with dijet events", Technical Report CMS-PAS-SUS-08-005, CERN, 2008. Geneva, (Oct, 2008).

[117] CMS Collaboration, "Search strategy for exclusive multi-jet events from supersymmetry at CMS", Technical Report CMS-PAS-SUS-09-001, CERN, 2009. Geneva, (Jul, 2009).

[118] J. Alwall, R. Frederix, S. Frixione et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[119] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[120] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J.* **C71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[121] S. Alioli, P. Nason, C. Oleari et al., "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, doi:10.1007/JHEP02(2010)011,10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: JHEP02,011(2010)].

[122] R. Gavin, Y. Li, F. Petriello et al., "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388–2403, doi:10.1016/j.cpc.2011.06.008, arXiv:1011.3540.

[123] T. Melia, P. Nason, R. Rontsch et al., "W+W-, WZ and ZZ production in the POWHEG BOX", *JHEP* **11** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.

[124] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[125] R. Gavin, Y. Li, F. Petriello et al., "W Physics at the LHC with FEWZ 2.1", *Comput. Phys. Commun.* **184** (2013) 208–214, doi:10.1016/j.cpc.2012.09.005, arXiv:1201.5896.

[126] CMS Collaboration Collaboration, "Muon Identification and Isolation efficiency on full 2016 dataset", Technical Report CMS-DP-2017-007, (Mar, 2017).

[127] CMS Collaboration Collaboration, "Performance of b-Tagging Algorithms in Proton Collisions at 13 TeV using the 2016 Data", Technical Report CMS-DP-2016-042, (Jul, 2016).

[128] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur. Phys. J.* **C75** (2015) 542, doi:10.1140/epjc/s10052-015-3709-x, arXiv:1505.04480.

[129] Z. Bern et al., "Left-Handed W Bosons at the LHC", *Phys. Rev.* **D84** (2011) 034008, doi:10.1103/PhysRevD.84.034008, arXiv:1103.5445.

[130] CMS Collaboration, "An inclusive search for new phenomena in final states with one or more jets and missing transverse momentum at 13 TeV with the AlphaT variable", Technical Report CMS-PAS-SUS-16-016, CERN, Geneva, (2016).

[131] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instr. and Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2.

[132] A. L. Read, "Presentation of search results: the CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[133] G. Cowan, K. Cranmer, E. Gross et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0.

List of figures

List of tables

List of Acronyms

CMS Compact Muon Solenoid

LHC Large Hadron Collider

SM Standard Model

SUSY supersymmetry

LSP lightest supersymmetric particle

GR General Relativity

DM dark matter

CL confidence level

ISR initial state radiation

FSR final state radiation

GUT Grand Unified Theory

LS1 Long Shutdown 1

LINAC2 Linear Accelerator 2

PSB Proton Synchrotron Booster

PS Proton Synchrotron

SPS Super Proton Synchrotron

PU pileup

OOTPU out of time pileup

PUS pileup subtraction

BSM beyond the SM

QED quantum electrodynamics

ECAL electromagnetic calorimeter

HCAL hadron calorimeter

EB ECAL barrel

EE ECAL endcaps

HB hadron barrel

HE hadron endcaps

HF hadron forward

HO hadron outer

L1T Level-1 Trigger

GT Global Trigger

GCT Global Calorimeter Trigger

RCT Regional Calorimeter Trigger

TT Trigger Tower

TP Trigger Primitive

HLT high-level trigger

TMT Time-Multiplexed Trigger

CSC cathode strip chamber

DT drift tube

RPC resistive plate chamber

FPGA field programmable gate array

PV primary vertex

CTF combinatorial track finder

PF particle flow

MC Monte Carlo

CHS charged hadron subtraction

CHF charged hadron fraction

SIT single isolated track

CSVv2 Combined Secondary Vertex version 2

LO leading order

MSSM Minimal Supersymmetric Standard Model

SMS simplified model spectra

NLO next-to-leading order

NNLO next-to-next-to-leading order

PDF parton distribution function

QCD quantum chromodynamics

CKM Cabibbo-Kobayashi-Maskawa

PMNS Pontecorvo-Maki-Nakagawa-Sakata

CMB Cosmic Microwave Background

CP charge parity

RF radio frequency

TF transfer factor